

NORTHWESTERN UNIVERSITY

Model-driven elucidation of synthetic and immune signaling mechanisms

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Interdisciplinary Biological Sciences

By

Joseph J. Muldoon

EVANSTON, ILLINOIS

September 2020

© Copyright by Joseph J. Muldoon 2020

All Rights Reserved

ABSTRACT

This dissertation focuses on the development of quantitative approaches for characterizing endogenous signaling pathways and designing new pathways in mammalian cells. I demonstrate how mathematical descriptions that are formulated to explain gene expression patterns can also serve as a powerful springboard for deeper analyses into the properties and functions of pathways. Throughout, a recurring theme is that by measuring the behaviors of single cells and building models that directly incorporate these observations, we are better equipped (1) to uncover fundamental biological mechanisms and (2) to achieve genetic engineering design goals.

To the first point, I investigated the macrophage response to pathogenic stimulus. Heterogeneity is a hallmark of this cell type, but whether or how this variation relates to protective immune functions is not well understood. By integrating single-cell tracking and dynamical systems modeling approaches, I identified a previously unrecognized form of intercellular coordination that we termed *quorum licensing*. I found that macrophages track the history of their density, and then in a manner independent of previous explanations for how cytokine production is amplified in this system, the cells preemptively decide on the proportion of the population that will become highly activated in response to an inflammatory cue. This behavior involves coordinating heterogeneous cellular activation states in a way that generates a nonlinear response at the population level. The role of this newly defined collective decision-making strategy might be to both amplify inflammatory responses and limit them within sites of injury.

To the second point, cells have a vast capacity to be repurposed, and engineered cell-based devices are finding applications in the targeted treatment of diseases. Because the genetic components such as synthetic receptors and transcription factors for building cellular functions are nascent, quantitative principles governing their effective integration are very much needed. I developed statistical and dynamical modeling approaches to elucidate mechanisms by which a variety of genetic parts operate. I subsequently demonstrated an approach for predictively implementing complex genetic programs in mammalian cells. These efforts, and the collaborations they entail, comprise part of a transformation in the field of mammalian synthetic biology from a reliance upon biophysical intuition to the utilization of model-guided interpretation and design. By imparting greater specificity to activation and robust performance to heterogeneity, in the long-term these studies can inform more effective and sophisticated cell-based devices.

ACKNOWLEDGEMENTS

There are many people who I would like to acknowledge for their support throughout my graduate school journey. Early on, I benefitted from the mentorship of Yishan Chuang and Danny Wells, grew from working with Taylor Dolberg, Patrick Donahue, Kelly Sarnese, and Jessica Yu, and also learned from Nickey Daringer, Justin Finkle, Rachel Hartfield, Michelle Hung, Albert Xue, and Andrew Younger. Subsequently, I learned from and worked alongside Narasimhan Balakrishnan, Sebastian Bernasek, Jon Boucher, Slava Butkovich, Jason Cain, Will Corcoran, Simrita Deol, Katie Dreyer, Hailey Edelstein, Taylor Gunnels, Josh Levy, Roxi Mitrut, and Devin Stranford, and comrades in co-advisement Alex Prybutok and Kate Dray. We share a camaraderie and lasting memories.

It has been an honor to work with tremendous undergraduate and master's students including Juliana Feng, Amy Hong, Vis Kandula, Suchitra Sankaranarayan, and Parth Shah, and to work alongside other outstanding undergraduate students including Everett Allchin, Joe Draut, Anthony Kang, and members of NU iGEM teams. Each day, what made this journey so meaningful, in addition to the science itself, was the experience of doing science together. I am so proud of you all.

I formed enduring friendships in Evanston and Chicago. I am grateful to the GeneMods co-founders Weston Kightlinger, Isaac Larkin, and Sarah Stainbrook, to the subsequent coordinators, and to the CSB community; my colleagues in IBiS and ChBE, including Nicolle Bonar, Alex Karge, Do Soon Kim, Ben Parker, Soo Ro, Sarah Stainbrook, and Gokay Yamankurt; my friends outside of lab including John Bostick, Phil Espe, Noah Ford, Parker Gaims, Will Higginbotham, Josh Hills, David Miller, Amelia Plunk, and Ado Rivera; and Bob Hasty, the NU Philharmonia, and others with whom I made music.

I am grateful for the recent opportunity to join a project to develop SARS-CoV-2 diagnostic tests. The circumstances in which this dissertation was written, and in which everyone has been operating, speak to the need for solutions. Many laboratories have pivoted or redoubled their efforts to develop diagnostics, vaccines, and treatments; perhaps a silver lining of these times will be the biomedical advances that result from these efforts. I thank Julius Lucks and other PIs for their leadership, and Kate Dray, Jithin George, Kirsten Jung, Grant Rybnicky, Sasha Shirman, Matt Verosloff, and others for collaboration.

I thank the faculty and staff of the IBiS program, including Cathy Prullage, Deborah Klos Dehring, and Traci Galbaugh for their help in navigating graduate school; Keith Tyo for an exciting rotation project; Bill Miller for leading the Biotechnology Training Program/Cluster while I was a member; the Kellogg Management for Scientists and Engineers program for an unforgettable learning experience; my committee members Erik Andersen, Evan Scott, and Rosemary Braun for their insight and feedback; and my scientific mentors and advisors prior to graduate school including David Auble, Melissa Carver, Marcus Gillis, Jay Hirsh, Kimberly Kelly, Keith Kozminski, Jason Papin, Kunal Poorey, Janaki Purushe, Manolito Torralba, and Ramya Viswanathan for the experiences that shaped my path.

To my advisors Josh Leonard and Neda Bagheri, I am grateful for the chance to join both labs. Thank you for fostering an environment that enabled me to explore a variety of interests, for believing in me throughout, and for supporting my development as a scientist. To Neda, thank you for bringing me on board at a stage when I still had much to learn, including about all things statistical, computational, and big-data. To Josh, thank you for creating a setting where I could test out and run with ideas and grow as a scientist and mentor. The opportunity and encouragement you provided to work with mentees was a defining experience for me. Arriving at the end of this chapter in life is bittersweet, but in starting the next chapter I feel ready because of the experiences and knowledge I gained during my time here at NU.

Lastly, I am fortunate to have a loving family that has supported me, as we all have supported each other, in many aspects of life. My parents John and Sylvie are people of tremendous strength and wisdom, for whom I cannot fully convey my admiration in words. They raised me with the values to be able to lead a meaningful life and have always been a sounding board. With my brothers Michael and Stephen, we have retained our bond despite the physical distance between Illinois and Virginia. I thank Nora and my extended family for supporting me growing up and through my education. I am grateful to so many people who have helped me pursue my dreams.

ABBREVIATIONS

ABA	abscisic acid
ABI1	Abl interactor 1
AD	activation domain
ARE	AU-rich element
AUROC	area under the receiver operator characteristic
AUPR	area under the precision-recall
BANJO	Bayesian network inference with Java objects
Bcl-3	B-cell lymphoma 3-encoded protein
BFA	brefeldin A
BMM	bone marrow-derived macrophages
CAF1	carbon catabolite repressor 4-associated factor 1
CAR	chimeric antigen receptor
CD11b	cluster of differentiation molecule 11B; integrin alpha M
CD28	cluster of differentiation 28
CMV	cytomegalovirus (promoter)
COMET	composable mammalian elements of transcription
CORR	correlation (abbreviation)
CRISPR	clustered regularly interspaced short palindromic repeats
CV	coefficient of variation
DAMP	damage-associated molecular pattern
DBD	DNA-binding domain
DFB	double feedback (motif)
DMSO	dimethyl sulfoxide (vehicle for rapamycin treatment)
DREAM	dialogue on reverse engineering and assessment methods
DsDed	DsRed-Express2 R95K mutant
DsRed	DsRed-Express2 (abbreviation)
DUSP1	dual specific phosphatase 1
EBFP2	enhanced blue fluorescent protein 2
ECD	ectodomain
EF1 α	elongation factor-1 alpha (promoter)
EGFP	enhanced green fluorescent protein
eIF2	eukaryotic initiation factor 2
eIF4E	eukaryotic translation initiation factor 4E
ERK	extracellular-signal-regulated kinase
ERS	edge rank score
ES	edge score
EtOH	ethanol (vehicle for rapalog treatment and abscisic acid treatment)
EYFP	enhanced yellow fluorescent protein
F4/80	epidermal growth factor-like module-containing mucin-like hormone receptor-like 1
FACS	flow-activated cell sorting
FBD	feedback dominance
F.D.	fold difference
FF	feedforward (motif)
FFFB	feedforward with feedback (motif)
FGFR1	fibroblast growth factor receptor 1
FGFR4	fibroblast growth factor receptor 4
FI	fan-in (motif)
FKBP	FK506 binding protein
FN	false negative
FP	false positive
FRB	FKBP-rapamycin binding
GA3-AM	gibberellic acid acetoxymethyl ester
GAI	gibberellin insensitive

GENIE3	gene network inference with ensemble of trees
GID1A	gibberellin insensitive dwarf 1A
GNW	GeneNetWeaver
gp41-1	(split intein)
GpA	glycophorin A
GS	glycine-serine (linker)
HuR	Hu-antigen R
IKK	nuclear factor of kappa light polypeptide gene enhancer in B cells inhibitor kinase
IKKK	nuclear factor of kappa light polypeptide gene enhancer in B cells inhibitor kinase kinase
IκB	nuclear factor of kappa light polypeptide gene enhancer in B cells inhibitor
IL-10	Interleukin 10
IL-10R	Interleukin 10 receptor
intC	split intein C-terminal fragment
intN	split intein N-terminal fragment
IRF	interferon regulatory factor
IW	inferred weight
JNK	c-Jun N-terminal kinase
KC	C-X-C motif ligand 1
LPS	lipopolysaccharide
MAPK	p38 mitogen activated protein kinase
mCherry	(a monomeric red fluorescent protein)
MCP-1	monocyte chemoattractant protein 1
MD-2	lymphocyte antigen 96
MEFL	molecules of equivalent fluorescein
MEPB	molecules of equivalent pacific blue
MEPTR	molecules of equivalent PE-Texas Red
MESA	modular extracellular sensor architecture
MFI	mean fluorescence intensity
MIMO	multi-input multi-output
MK2	mitogen activated protein kinase-activated protein kinase 2
mKate2	(a monomeric far-red fluorescent protein)
MIDER	mutual information distance and entropy reduction
MIP-1α	macrophage inflammatory protein 1α
MIP-1β	macrophage inflammatory protein 1β
MyD88	myeloid differentiation primary response gene 88
NF-κB	nuclear factor of kappa light polypeptide gene enhancer in B cells
NW	null weight
ODE	ordinary differential equation
PABP	poly(A)-binding protein
PAMP	pathogen-associated molecular pattern
PC	protease chain
PCA	principal component analysis
PCR	polymerase chain reaction
PEST	(a degradation tag)
PKC	protein kinase C
PMA	phorbol 12-myristate 13-acetate
PRS	protease recognition sequence
PYL1	pyrabactin resistance-like
QL	quorum licensing
QS	quorum sensing
RANTES	regulated on activation, normal T cell expressed and secreted; C-C motif ligand 5
RaZFa	rapamycin-activated zinc finger activator
RNAPII	RNA polymerase II
SFB	single feedback (motif)
sgRNA	single guide RNA

STAT3	signal transducer and activator of transcription 3
sTNFR	soluble tumor necrosis factor receptor
TACE	tumor necrosis factor alpha-converting enzyme
TAM	tumor-associated macrophage
TC	target chain
TEVp	tobacco etch virus protease
TF	transcription factor
TIC	TLR4-inhibitory complex
TIGRESS	trustful inference of gene regulation using stability selection
TIR	Toll/IL-1 receptor (domain)
TLR4	Toll-like receptor 4
TMD	transmembrane domain
TN	true negative
TNF	tumor necrosis factor
TNFR1	tumor necrosis factor receptor 1
TP	true positive
TRIF	Toll/interleukin-1 receptor-domain-containing adapter-inducing interferon beta
TTP	tristetraprolin
UFB	upstream feedback (motif)
UTR	untranslated region
VP16	VP16 activation domain
VP64	VP64 activation domain
VPR	VP64-p65-Rta activation domain
WT	wild type
ZF	zinc finger
ZFa	zinc finger activator
ZFi	zinc finger inhibitor

TABLE OF CONTENTS

ABSTRACT	3
ACKNOWLEDGEMENTS	4
ABBREVIATIONS	6
TABLE OF CONTENTS	9
CHAPTER 1. Introduction and background	15
1.1 Dynamical modeling in cell biology	16
1.2 Immune cell heterogeneity and coordination	18
1.3 Engineering mammalian cellular functions	20
CHAPTER 2. Collective decision-making in macrophage activation	25
2.1 Abstract	26
2.2 Introduction	26
2.3 Materials and Methods	28
2.4 Results	32
2.5 Discussion	46
2.6 Acknowledgements	49
CHAPTER 3. Modeling gene expression by synthetic transcription factors	50
3.1 Abstract	51
3.2 Introduction	51
3.3 Materials and Methods	52
3.4 Results	52

3.5 Discussion	10
3.6 Acknowledgements	60
CHAPTER 4. Design-driven engineering of mammalian genetic programs	61
4.1 Abstract	62
4.2 Introduction	63
4.3 Materials and Methods	63
4.4 Results	64
4.5 Discussion	67
4.6 Acknowledgements	79
CHAPTER 5. A genome-scale knockout screen for negative regulators of NF- κ B activity	80
5.1 Abstract	81
5.2 Introduction	82
5.3 Materials and Methods	82
5.4 Results	85
5.5 Discussion	88
5.6 Acknowledgements	90
CHAPTER 6. Conclusions	92
REFERENCES	93
APPENDIX 1. Vectors and cell types	100
APPENDIX 2. Model equations for macrophage quorum licensing	119
APPENDIX 3. Modeling gene expression heterogeneity	125
APPENDIX 4. Multiplexing receptors	143
APPENDIX 5. Model equations for transcription factors	146
	161

APPENDIX 6. Model equations for genetic components	11 169
APPENDIX 7. Identification of determinants of network inference algorithm performance	173
APPENDIX 8. New ligand-sensing for receptors	184
APPENDIX 9. New intracellular cargo for receptors	187

LIST OF TABLES AND FIGURES**CHAPTER 1**

- Figure 1.1 Parameter fitting
- Figure 1.2 Cellular heterogeneity
- Figure 1.3 Considerations and principles for engineering mammalian cell-based devices

CHAPTER 2

- Figure 2.1 The TNF response to LPS is heterogeneous and requires intercellular communication
- Figure 2.2 Cell density modulates the heterogeneity of macrophage activation
- Figure 2.3 A dynamical population model explains intracellular and extracellular signaling and regulation
- Figure 2.4 Mechanisms of TNF regulation differ in phenotypic and functional consequences
- Figure 2.5 Primary macrophages regulate TNF production through quorum licensing

CHAPTER 3

- Figure 3.1 COMET
- Figure 3.2 A model for COMET-mediated gene regulation
- Figure 3.3 Tuning transcription through ZF mutants and AD variants
- Figure 3.4 Transcriptional inhibition
- Figure 3.5 Composing Boolean logic

CHAPTER 4

- Figure 4.1 Logical evaluation is enabled by transcriptional and post-translational regulation
- Figure 4.2 Compact multi-output logic is attained through functional modularity
- Figure 4.3 Analog behaviors are constructed by using TFs that play multiple roles
- Figure 4.4 Sensors can be linked to genetic programs to make signaling cascades

CHAPTER 5

- Figure 5.1 The TIC hypothesis

- Table 5.1 Prior NF- κ B screens
- Figure 5.2 Screen design and implementation
- Figure 5.3 Screen results
- Figure 5.4 Validation of candidate regulators of NF- κ B activity

CHAPTER 6

- Figure 6.1 Quorum licensing

APPENDIX 1

- Table A1.1 Vectors
- Table A1.2 Cell types

APPENDIX 2

- Figure A2.1 Primary cell experiments
- Figure A2.2 Secreted factors
- Table A2.1 State variables
- Table A2.2 Parameters
- Table A2.3 Stimulus-specific and perturbation-specific parameters
- Table A2.4 Ordinary differential equations

APPENDIX 3

- Figure A3.1 Distribution of gene expression in a population

APPENDIX 4

- Figure A4.1 Multiplexed receptors and hybrid promoter implementation
- Figure A4.2 A model that accounts for cell variation to explain heterogeneous promoter activity
- Figure A4.3 A dynamical model that links receptor signaling to promoter activity
- Figure A4.4 Level-matching between receptor signaling and the promoter
- Figure A4.5 Promoter engineering to improve performance
- Figure A4.6 Receptor engineering to improve performance

APPENDIX 5

- Table A5.1 ZFa and fitted parameters for x6-C promoters

Table A5.2 Fitted parameters for modifications to promoter architecture and ZFa domains

APPENDIX 6

Table A6.1 Model parameters for genetic programs

APPENDIX 7

Figure A7.1 Evaluating performance of network inference

Figure A7.2 Network confidence varies across algorithms

Table A7.1 Network motifs

Table A7.2 Logic gates

APPENDIX 8

Figure A8.1 Development of gibberellin-MESA

Figure A8.2 Development of abscisic acid-MESA

APPENDIX 9

Figure A9.1 Split intein mutagenesis

Figure A9.2 Functional test of split intein cargo signaling

Figure A9.3 Adapting MESA for inhibitory signaling

CHAPTER 1. Introduction and background

Part of a version of this chapter was previously published as:

Muldoon J.J.*, Donahue P.S.*, Dolberg T.B.*, Leonard J.N. Building with intent: technologies and principles for engineering mammalian cell-based therapies to sense and respond. *Current Opinion in Biomedical Engineering* **4**, 127–133 (2017).¹ *Equal contributions

This chapter begins with an overview of dynamical modeling with a focus on how I have used it. The second section on innate immune cell biology provides a background for **Chapters 2 and 5**, and the third section on mammalian cell-based devices relates to **Chapters 3 and 4**.

1.1 Dynamical modeling in cell biology

Models: In a broad sense, a model can be thought of as an idealized representation of some aspect of reality. Any modeling project begins with a question or a purpose, such as to organize information, define and test hypotheses, identify important elements, or make predictions about new scenarios. In the biological sciences, computational models have become immensely valuable in helping to derive insights from data and shape research directions. In my experience, some of the goals at the outset of a project can often be achieved directly through the deliberate process of model construction, even prior to any simulations. Thus, at its core, the application of modeling in the biological sciences provides a framework for distilling the fundamental essence of a question and illuminating a path forward to address it.

There are several well-known aphorisms that have been attributed to various sources on this topic, and at the risk of contributing to their frequent quotation, I think the lessons to be learned from these statements remain at least as meritorious now as ever. Box and Draper said that "all models are wrong, but some are useful,"² meaning that the goal is not to represent all of reality exactly, but rather to make useful approximations within a defined scope. A conceptual or mathematical model need not be true or complete to be useful. Naisbitt said that "we are drowning in information but starved for knowledge,"³ emphasizing the need for ways to extract useful information from large amounts of data—a sentiment that seems all the more salient in the omics age. Eigen said that "a theory has only the alternative of being right or wrong. A model has a third possibility: it may be right, but irrelevant,"⁴ meaning that modeling should be pursued in a manner that aligns with the ultimate goals of a study. This idea has been central to my work in making the design and implementation of genetic programs in mammalian cells more practical. More humorously, Brewster, Einstein, Eigen, Berra, and others have said that "in theory there is no difference between theory and practice, while in practice there is."⁵ At the end of the day, I hope that my work will lead to practical advances, and there will certainly still be room for refinement of the models.

Dynamical modeling: The primary framework I have employed across different projects involves ordinary differential equations (ODEs), which can be made to represent dynamic processes using various levels of mechanistic granularity⁶. In particular, physicochemical modeling describes biochemical reactions using state variables for the components and parameters for the kinetics. Mass action kinetics are used, assuming determinism (rather than stochasticity), high numbers of molecules (as is known to be generally the case for the systems investigated here), and that transport is more rapid within than between cellular compartments^{7,8}.

The process of model development differs on a case-by-case basis, but some general ideas seem to hold (**Figure 1.1**). First, equations are formulated based on prior knowledge, hypotheses, and assumptions. This step is often associated with turning cartoons depicting a hypothesis as to how a system works into more precise and useful cartoons, and then into the equations. Second, simulations are conducted by formulating variations on the original system of equations corresponding to the applicable perturbative conditions (e.g., differences in ligand treatments and the timing and dosing), providing initial values for the state variables, and assigning values to the parameters. Third, the simulations are compared to a set of experimental observations, and a search is conducted to determine whether there exist sets of parameter values that provide a suitable fit to the data based on defined criteria in the form of objective functions. Where there are multiple datasets or features of datasets to which the simulations are compared, multi-objective optimization is used^{9,10}. In practice, this whole process involves multiple iterations in which candidate models are proposed, tested, and assessed. If a suitable fit is obtained, then there are several types of analysis that can be conducted. For example, if there are several models representing competing

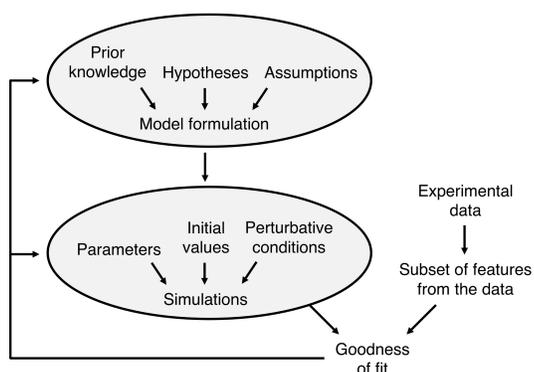


Figure 1.1. Parameter fitting. High-level schematic for parameterizing a model to data.

hypotheses, then differences in their goodness of fit can be used to identify certain mechanisms as being more explanatory for the experimental observations. As another example, a trained model can be applied to simulate untested scenarios, either to identify specific regimes within a system (e.g., for improving the performance of a genetic program) or to validate the model's predictive capacity more broadly.

Parameter estimation¹¹ is a rich area of study for which many algorithms have been developed^{12,13}. Despite these advances, perhaps surprisingly, in systems that are complex and high-dimensional (with many free parameters), it can still be a challenge to find a solution that yields a good fit. The other side of this coin is that after conducting a parameter search, there are often many parameter sets that yield similarly good fits. These models are called sloppy (as compared to stiff) and are common where nonlinear interactions are involved¹⁴. Sloppiness arises from correlations in sensitivities of ODE solutions to parameters and does not necessarily reflect a model's quality or utility¹⁵. Because some parameters are inevitably unconstrained, it can be useful to focus on model predictions and behaviors rather than on finding a definitively best parameter set, and this is the perspective that I have taken in characterizing endogenous pathways and in building new ones, for which background is provided in **Sections 1.2 and 1.3**, respectively.

1.2 Immune cell heterogeneity and coordination

Macrophages in innate immunity: Innate immune cells play a tremendous range of roles such as phagocytosing pathogens, regulating tissue homeostasis, and recruiting and presenting antigens to adaptive immune cells. A deeper understanding of these roles is coming into focus with new technologies and tools to gather, integrate, and make inferences from data¹⁶. Within the innate immune system, the mononuclear phagocyte lineage comprises monocytes, macrophages, and other cells that migrate or reside throughout the body and are responsible for trophic, inflammatory, and protective activities¹⁷. Here I focus on macrophages, which become activated in response to cues like growth factors, cytokines, and pathogen-associated molecular patterns (PAMPs) and functionally polarize, i.e., differentiate into specific phenotypes to carry out context-dependent activities.

This polarization is typically categorized as follows: M1 or classically activated macrophages are microbicidal, pro-inflammatory, and antigen-presenting; M2 or alternatively activated macrophages are anti-inflammatory and promote wound healing; and tumor-associated macrophages (TAMs) are M2-like with

some M1 characteristics. Many genes are involved in polarization at the level of signal transduction, gene expression, metabolism, and cytokine production¹⁸. Interestingly, it has recently become evident that the M1–M2 axis is an oversimplification. Xue et al. analyzed macrophage transcriptomes in response to a panel of stimuli, and through a machine learning and network analysis, found that polarization was better described by a high-dimensional phenotypic space¹⁹. This study underscores how much remains to be uncovered on the fundamental biology of macrophage activation.

Heterogeneity and communication: Single-cell analyses have revealed striking heterogeneity at the RNA²⁰⁻²² and protein^{23,24} levels among genetically identical immune cells, including macrophages and other cell types following treatment with the same cues. It is possible that these differences could arise from underlying variation in the expression of key regulatory genes. This explanation would indicate a role for extrinsic noise, which describes different initial conditions that take deterministically different trajectories—as opposed to intrinsic noise, which describes similar initial conditions that deviate due to stochasticity²⁵. Additionally, it has been proposed that heterogeneity might confer advantageous properties to immune cell populations²⁶, although specific ways in which this might play out are still being elucidated. Example types of cellular heterogeneity that could potentially be uncovered are illustrated in **Figure 1.2**.

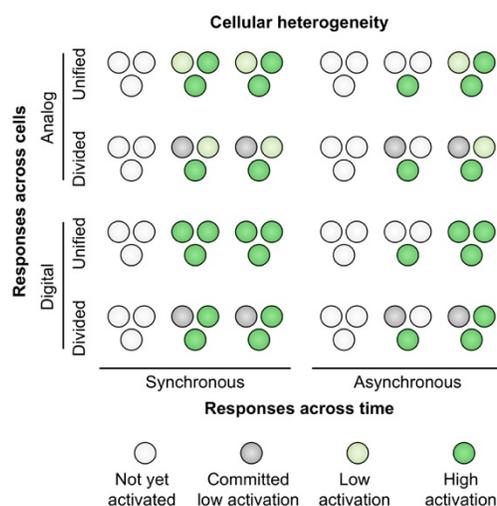


Figure 1.2. Cellular heterogeneity. Variation comes in different varieties. The cartoon illustrates scenarios that differ in whether cells respond in a graded or all-or-none manner, whether the decision to activate is unanimous or split, and whether activation is simultaneous or staggered. Each circle represents a cell, and its activation state is color-coded. The three close circles represent a cell population that changes over time (three diagrams from left to right).

An important concept that goes hand in hand with heterogeneity is intercellular communication. Immune cells need to coordinate their response to threats, and this line of communication is generally established by secreting factors such as cytokines that other immune cells sense and respond to. Several studies have documented immune cell behaviors and likened them to bacterial quorum sensing (QS)—a phenomenon in which bacteria secrete a molecule that acts as a proxy for cell density, and once a threshold concentration of that molecule is reached (indicating high cell density), target gene expression is induced²⁷. In this way, the activities of individual cells are connected to each other. For immune cells, density appears to be a recurring and crucial regulator of activity, and there are also activities that regulate cell density^{28,29}. Some of these instances differ from the QS concept and can be more directly ascribed to paracrine signaling, e.g., canonical explanations for the macrophage tumor necrosis factor (TNF) response to lipopolysaccharide (LPS)^{30,31}. However, other instances are QS-like. As one example, hair follicle regeneration involves the recruitment of macrophages, and this process occurs above a threshold amount of injury to the hair follicles³². As another example, in the pathogen-induced accumulation of monocytes *in vivo*, the monocytes produce nitric oxide, and when high nitric oxide levels are reached (corresponding to high cell density), the population undergoes a coordinated metabolic shift towards lower cytokine production and ultimately the resolution of inflammation³³. In these examples, as in bacterial QS, cells sense a secreted molecule to "get on the same page" and act similarly. An exciting frontier is the interplay between cellular heterogeneity and intercellular communication. As I show in **Chapter 2**, macrophages employ a form of coordination that leads them to activate *differently* in response to a threat, but in such a way that amplifies the population-level outcome in cases where many macrophages are present.

1.3 Engineering mammalian cellular functions

Cell-based therapies are a rapidly advancing medical frontier. Cells can carry out tasks that are inaccessible pharmacologically, such as directed trafficking, therapeutic production *in situ*, and cell-mediated killing. One promising approach, as evidenced by the growing number of trials for chimeric antigen receptor (CAR) T-cell therapies³⁴, is to genetically engineer cells to perform defined sense-and-respond behaviors. Currently, there is a need for tools and principles to make these types of behaviors more sophisticated and reliable. Here, I outline at a high level several challenges and forward-looking possibilities that may be broadly

applicable to mammalian cellular engineering, and some of these topics (particularly relating to increasing sophistication and reliability) can be aided by the advances described in **Chapters 3 and 4**.

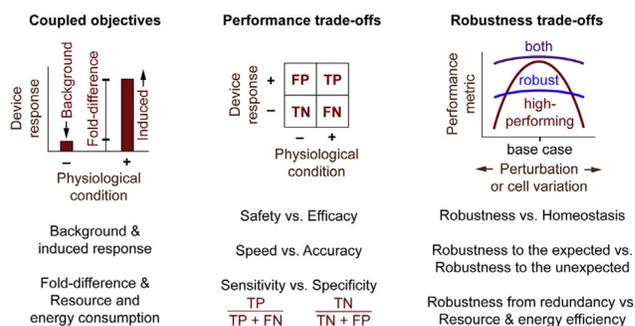
Design priorities and trade-offs: In considering the steps that go into developing a cell-based device, it becomes apparent that this process necessarily involves a prioritization of multiple objectives. This prioritization, whether explicit or implicit, could affect the development of a prototype to a final product. Currently, it is unclear how the decisions that are made in early-stage research funnel biomolecular designs and constrain outcomes in subsequent preclinical and clinical evaluations. Moreover, the choices that improve performance in a lab setting do not necessarily guarantee robustness³⁵: the ability to perform a function under wide-ranging conditions that the cells would encounter in a recipient.

In analyzing situations with multiple objectives, it is useful to consider trade-offs that might arise as a result of prioritization. In this context, trade-offs include the accuracy versus rapidity with which a device responds to a cue³⁶, and the sensitivity of a device to the cue versus the specificity with which the device responds only to that particular cue. Each of these trade-offs can shift the balance between device safety and efficacy (**Figure 1.3a**). These concepts have already proven useful for CAR T-cell therapies: small molecule-inducible kill switches³⁷ and activation-ON switches³⁸ for improved safety; requiring multiple cues to trigger an activation event to confer improved specificity³⁹; and varying the cell dose or route of administration, or the affinity of the receptor for the target ligand, to achieve decreased on-target off-tissue activation⁴⁰. New therapies will benefit from a more defined understanding of the relationships between initial prioritization, prototype evaluation, and success in translation.

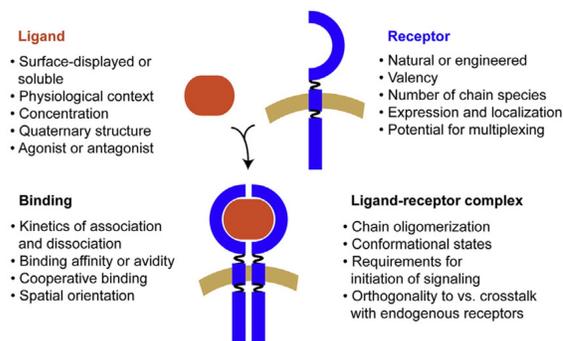
Extracellular sensor properties: Sense-and-respond programs can be engineered using natural or synthetic cell surface receptors. Key choices include the type of ligand-binding domain employed (e.g., native receptor domain, single-chain variable fragment, or nanobody⁴¹) and its size, valency, stability, orientation, and ligand-binding mechanism and kinetics (**Figure 1.3b**). Experience from CAR T cells indicates that tuning ligand-binding affinity can improve how and under what conditions a device activates⁴⁰. A challenge is that although physiological ligand concentrations can be estimated for some species in vivo⁴², local ligand concentrations in vivo are generally less understood and vary between milieus.

Consequences of intercellular variation: There now exist various methods to implement genetic programs in cells, and elucidating the relationship between implementation method and cellular device

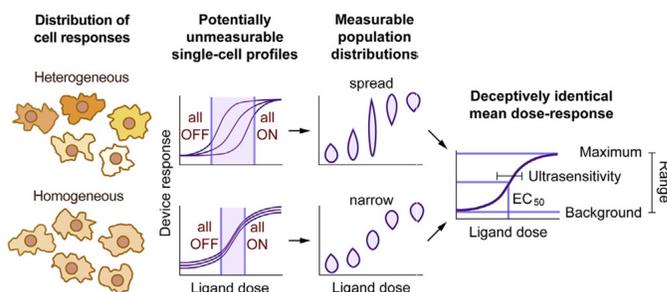
a Design priorities and trade-offs



b Extracellular sensor properties



c Consequences of intercellular variation



d Coordination of cellular functions

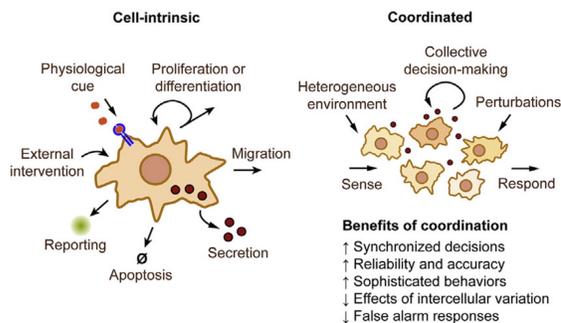


Figure 1.3. Considerations and principles for engineering mammalian cell-based devices.

a Designing a cell-based device involves balancing trade-offs for various characteristics, which may each be desirable but cannot be realized simultaneously. To illustrate, we consider a cell that is intended to become activated only under a specified physiological condition, by expressing an engineered receptor for sensing a cue and producing a measurable readout in response. Fold difference is calculated as the ligand-induced readout divided by the ligand-independent (background) readout. Sensitivity and specificity are defined using classification terms: T = true, F = false, P = positive, N = negative. These performance metrics differ from robustness, which is the extent to which performance is maintained under perturbations or cell variation. While an ideal device is both high-performing and robust, in practice these characteristics may come at some expense of one another.

b The target ligand imposes constraints on how a receptor can be made to form a productive signaling complex. For a ligand with multiple subunits, better performance may be realized using a receptor that has a corresponding valency, oligomerizes, or undergoes cooperative binding; however, these features can also introduce geometric or kinetic tuning requirements.

c Intercellular variation can impact the design, development, and performance of a cell-based device. Such variation is related to how endogenous and exogenous genes are expressed and other sources of biological noise. Metrics such as EC₅₀ (half maximal effective concentration of ligand) or ultrasensitivity (the steepness of the dose-response) that are derived from a mean profile of a heterogeneous population do not represent all cells. Therefore, single-cell analysis is an important part of characterization.

d Strategies for intercellular coordination could potentially diminish the effects of intercellular variation and enhance the robustness of engineered functions.

performance is an area of active investigation. A key choice is whether to integrate a genetic payload genomically or to maintain it extra-genomically and potentially transiently. Each delivery method has advantages and disadvantages. Lentiviral and retroviral vectors carry payloads up to ~10 kb, integrate pseudorandomly in the genome, and can become silenced in certain cell types depending on the integration site⁴³. Transposon and transposase systems deliver larger payloads but are less efficient than lentiviral and retroviral vectors and integrate randomly (with a less biased pattern of integration)⁴⁴. Recombinase-based landing pads provide site-specific integration and more homogeneous gene expression, but thus far the required insertion of the initial landing pad precludes the use of this technology in primary cells⁴⁵, and it is not yet clear whether a given safe harbor site for targeting genomic integration is effective across applications⁴⁶. Targeted integration by Cas9-mediated DNA editing can confer site specificity, and sgRNAs are relatively straightforward to design, but integration efficiency remains a challenge⁴⁷. Extra-genomic artificial chromosomes are emerging vectors that are maintained at a single copy, replicate during mitosis, and carry payloads of up to many megabases⁴⁸. Finally, plasmid transfection, RNA replicons⁴⁹, and non-integrating viral vectors are transient delivery options. Each method confers its own profile of variation in gene expression and risk of gene silencing, and how such variation can impact device characterization and performance is a topic of ongoing investigation.

Intercellular variation in device performance can be attributed to external sources, such as sample preparation, cell density, and cell-to-cell contact, as well as internal sources like cell cycle asynchrony, unequal inheritance in cell division, and stochastic fluctuations in gene expression^{50,51}. While design goals may be framed using digital metaphors like circuits, ultimately cells operate within a continuum of variation in both component doses and signals, and such variation can limit the extent to which a device reliably performs an engineered function (**Figure 1.3c**)⁵². It has been suggested that if the factors that contribute to variation are correlated, then distributions of cell behavior will be skewed by outlier cells⁵³. Indeed, it was recently shown that cells expressing high levels of exogenous components produced outlier responses that dominated the average behavior of the population⁵⁴. More broadly, such effects may influence iterative tuning to bias design choices toward those in which many cells do not exhibit the desired functionality. Therefore, precise characterizations for ligand detection limits, dose response profiles, and parameter sensitivities require individual-cell resolved analyses⁵⁵. Rigorous consideration of how intercellular variation

impacts device performance, and perhaps minimization of such variation in the manufacturing process⁵⁶, may lead to devices that perform better at translational stages.

Coordination of cellular functions: The desired function of a cell-based device is often specified as a cell-intrinsic input-output relationship. However, in vivo, cells operate as populations and interact with other cell types in complex and dynamic environments. As a result, it may not be obvious how many cells are required to achieve a desired outcome, or what fraction of the population will exhibit an intended function. Engineering intercellular coordination (e.g., orthogonal to native signaling) is a promising strategy to enhance the reliability or synchrony of programmed behaviors and thus mask the effects of inevitable cell variation and unexpected perturbations (**Figure 1.3d**). For example, negative paracrine feedback could be used to confer adaptive responses to sustained stimuli, and positive paracrine feedback could be used to provide more digital (“yes or no”) commitment to population-level decisions^{57,58}.

Recent studies suggest that more sophisticated functional programs can be realized by multiplexing different receptors⁵⁴ or different types of receptors in a cell⁵⁹, or potentially by distributing the job of sensing across multiple cell subpopulations. Non-receptor components have been distributed in this way, allowing subpopulations to exchange metabolites⁶⁰ and to coordinate transgene expression in a manner dependent on the subpopulation densities and for how long the subpopulations are in communication⁶¹. Whether such coordination will increase or decrease the challenges of designing and implementing a therapeutic product remains to be seen, but this type of approach may eventually augment the safety and efficacy of cell-based devices.

CHAPTER 2. Collective decision-making in macrophage activation

A version of this chapter was previously published as:

Muldoon J.J., Chuang Y., Bagheri N.[^], Leonard J.N.[^] Macrophages employ quorum licensing to regulate collective activation. *Nat Commun* **11**, 878 (2020).⁶² [^]Co-corresponding

2.1 Abstract

This chapter represents a central focus of my graduate study. In addition to pursuing a fundamental question on how macrophages coordinate their activation, I developed a computational skillset that was key to addressing topics of dynamics, heterogeneity, and robustness that arose in and enabled other projects (e.g., **Chapters 3–4, Appendixes 3–6**). In particular, identifying a role for gene expression noise on quorum licensing was subsequently the inspiration for developing a way to account for the effects of gene expression variation on the performance of synthetic genetic components in other settings.

Macrophage-initiated inflammation is tightly regulated to eliminate threats such as infections while suppressing harmful immune activation. However, individual cells' signaling responses to pro-inflammatory cues are heterogeneous, with subpopulations emerging with high or low activation states. Here, we use single-cell tracking and dynamical modeling to develop and validate a revised model for LPS-induced macrophage activation that invokes a mechanism we term quorum licensing. The results show that bimodal phenotypic partitioning of macrophages is primed during the resting state, dependent on cumulative history of cell density, predicted by extrinsic noise in transcription factor expression, and independent of canonical LPS-induced intercellular feedback in the TNF response. Our analysis shows how this density-dependent coupling produces a nonlinear effect on collective TNF production. We speculate that by linking macrophage density to activation, this mechanism could amplify local responses to threats and prevent false alarms.

2.2 Introduction

In responding to external cues, cells are faced with many options, but by sharing information a population of cells can make more effective decisions than can each individual alone⁶³. These decisions are generally mediated by secreted products. Bacteria use QS molecules to coordinate when and whether to form biofilms, and social amoeba secrete cyclic AMP to coordinate their aggregation. In each case, a proxy is used to convey information about the local number of cells available to coordinate the task.

In immunology, an established example of coordination is the amplification of the response to a perceived threat. During infection, cells such as macrophages provide an immediate line of defense by initiating inflammation to eliminate invading microbes⁶⁴. This process often begins when the bacterial

membrane component and pro-inflammatory cue LPS is sensed by Toll-like receptor 4 (TLR4). Prior to TLR4 activation, the transcription factor NF- κ B is sequestered in the cytoplasm by Inhibitor of κ B (I κ B). Upon activation, I κ B kinase (IKK) targets I κ B for degradation^{65,66}, allowing NF- κ B to translocate to the nucleus. There, NF- κ B induces the transcription of genes such as TNF, a cytokine that mediates inflammation and pathogen clearance^{16,67,68}. Other pathways downstream of TLR4 increase TNF production by stabilizing the mRNA and promoting translation and cleavage of the pro-protein for secretion. Extracellular TNF then signals through TNF Receptor 1 (TNFR1), driving NF- κ B activation in positive feedback^{31,69-72}, and this is a general explanation for how macrophages and other cells amplify their response to LPS.

The regulation of TNF has multiple layers. Recently, it was found that when the concentration of LPS exceeds a certain threshold, the induced signaling through NF- κ B drives the transcription of its own RelA subunit in a process termed the feedback dominance (FBD) switch, producing *intracellular* positive feedback on NF- κ B expression and activity⁷³. Other pathways act to constrain the response to LPS and ensure its eventual resolution: cell-intrinsic regulators (those with intracellular origins) include microRNAs and mRNA-binding proteins that decrease *Tnf* stability and translation⁷⁴, as well as I κ B^{65,75} and various inhibitors of IKK^{66,69,76} induced by NF- κ B in negative feedback; cell-extrinsic regulators (those with extracellular origins) include interleukin 10 (IL-10), in that IL-10 signaling via the IL-10 receptor (IL-10R) antagonizes NF- κ B activity and destabilizes *Tnf* stability and translation. In combination, these interlocking positive and negative motifs confer the functional plasticity necessary for immune cells to balance pathogen clearance with harmful side effects such as cytotoxicity and tissue damage⁷⁷.

Given the many facets of the regulation of NF- κ B and TNF, computational models have proven valuable for elucidating the properties of these systems and the roles of individual components. Early models explicated intracellular signaling^{65,66,78-80}, and subsequent models included newly appreciated mechanisms such as intercellular feedback^{31,69,71,81-84}. Recent studies have incorporated cell heterogeneity by attributing observed differences in gene expression either to stochastic fluctuations⁸⁵⁻⁸⁷ or to variation in initial values⁸⁸, kinetic parameters^{73,89-91}, or timing of signaling events⁹². A key consideration for understanding signaling and regulation in macrophages, in particular, is that these cells characteristically exhibit broad phenotypic heterogeneity^{19,23,93}. It has been proposed that this variation could have important

functional consequences, such as to broaden the repertoire of responses to stimuli⁹⁴, propagate or restrain coordinated actions, or convert digital single-cell decisions into analog population-level ones²⁶. While these ideas are interesting, specific mechanisms by which such heterogeneity might confer functional gain are not well understood.

In this study, we investigate the intriguing observation that when macrophages are treated with LPS, cell subpopulations emerge with high and low activation states. We propose a revised model in which macrophages use a process that we term quorum licensing to link the history of their density to the proportion of cells that become highly activated. This investigation provides new insights into how populations of macrophages use density information to regulate their collective activation.

2.3 Materials and Methods

RAW cell culture. RAW 264.7 cells were cultured in complete DMEM medium containing 1% DMEM (Gibco #31600091), 0.35% w/v D-glucose (Sigma #50-99-7), 0.37% w/v sodium bicarbonate (Fisher #S233-500), 10% heat-inactivated FBS (Gibco #16140071), 4 mM L-glutamine (Gibco #25030081), and 100 U ml⁻¹ penicillin and 100 µg ml⁻¹ streptomycin (Gibco #15140122) in tissue culture-treated dishes (Corning) at 37°C in 5% CO₂. To passage, medium was aspirated, and cells were washed in PBS, incubated in PBS-EDTA (5 mM EDTA in PBS pH 7.4) (37°C, 5 min), detached by gentle scraping, pelleted by centrifugation in 50 ml conical tubes (125×g, 5 min), and resuspended in fresh medium and plated. Reporter cells were cultured under the same conditions. For higher density passaging (**Figure 2.2g**), reporter cells were grown to cover a large majority of surface area of dishes, and this condition was maintained for several days leading up to the functional experiment. RAW 264.7 cells were a gift from David Segal (NIH), and reporter cells⁷³ were a gift from Iain Fraser (NIH). Cell lines were not authenticated further.

Confocal microscopy. Experiments were conducted using reporter cells and a Zeiss inverted Axio Observer Z1 confocal microscope with a custom environmental control chamber for CO₂, humidity, and temperature control. Data were exported, and fluorescence was quantified using ImageJ software⁹⁵: at each time point, the mean signal for each cell was quantified and the mean background signal based upon multiple sampled regions without cells was subtracted. The resulting value was multiplied by the area enclosed by the plasma membrane in the image slice to determine EGFP-RelA signal and mCherry signal.

A similar method was applied using the area within the nuclear membrane to determine nuclear EGFP-RelA signal. For conditions at high density, 30 cells were quantified; cells were not quantified if they divided or exited the field of view during the timecourse. For low density, 20 cells were quantified; due to the low number of trackable cells within the field of view in this case, some traces start or end within the timecourse.

Generation of monoclonal lines. The original reporter cell line was generated⁷³ via lentiviral integration of the two reporters encoded in a single vector—using a low multiplicity of infection, such that the large majority of cells underwent at most one integration⁹⁶—followed by fluorescence-activated cell sorting for EGFP-positive clones, clonal expansion, and functional screening for a clone exhibiting LPS-inducible EGFP-RelA nuclear translocation and mCherry expression. In the current study, seven monoclonal sublines (A–G) were generated from the parental line by limiting dilution and clonal expansion.

L929 cell culture. L929 fibroblasts (ATCC) were cultured in RPMI-1640 medium (Gibco #11875093, containing L-glutamine) supplemented with 10% heat-inactivated FBS, 100 U ml⁻¹ penicillin, and 100 µg ml⁻¹ streptomycin in tissue culture-treated dishes at 37°C in 5% CO₂. For passaging, medium was aspirated, and cells were washed in PBS, incubated in trypsin-EDTA (Gibco #25300054) (37°C, 5 min), detached by tapping the dish and pipetting in fresh medium, pelleted by centrifugation in 50 ml conical tubes (150×g, 5 min), resuspended in fresh medium, and plated. L929-conditioned medium was prepared by filtering supernatant at two or three days after the previous passage, by which time the cells had covered a large majority of the surface area of dishes. Conditioned medium was stored at –20°C and thawed for differentiating primary cells.

Bone marrow harvest and primary cell culture. C57BL/6 male mice (Jackson Labs; 5–10 weeks old) were sacrificed, and bone marrow cells were harvested from femurs and tibias⁹⁷. All of the animals were handled according to the animal protocol (#IS00003438), which was approved by the Northwestern University Institutional Animal Care and Use Committee and complies with all relevant ethical regulations for animal testing and research. Cells were cultured in differentiation medium containing complete RPMI-1640 medium (as used for culturing L929 cells) supplemented with 10% L929-conditioned medium in non-treated 10 cm dishes (Corning #CLS430591) (4×10^4 monocytic cells ml⁻¹, 10 ml per dish) at 37°C in 5% CO₂. At three days post-harvest, medium was aspirated without removing the partly adherent cells and replaced with differentiation medium containing fresh complete RPMI-1640 medium supplemented with

15% L929-conditioned medium. Cells were cultured for four more days, by which time the cells had become adherent BMM.

BMM surface staining. At seven days post-harvest, BMM were stained for surface markers of differentiation and assayed by flow cytometry. Medium was aspirated, and cells were washed in PBS, incubated in FACS buffer (FB; 5 mM EDTA and 0.1% BSA in PBS) (4°C, >10 min), detached by tapping the dish firmly and pipetting in FB, aliquoted as 2.5×10^5 cells per FACS tube, and pelleted by centrifugation (400×g, 5 min). Supernatant was decanted, paraformaldehyde (PFA; 4% in PBS, 30 µl) was added, and tubes were flicked to mix and incubated (4°C, 20 min). 1 ml FB was added, tubes were flicked to mix and centrifuged, and supernatant was decanted; this step was performed a total of three times. To block, normal mouse serum (Sigma #M5905, 10 µl) was added, and tubes were flicked to mix and incubated (room temperature—approximately 22°C, 15 min). Staining was conducted with primary conjugated antibodies (BD Biosciences): PE rat anti-CD11b (#553311, 0.04 µg) and Alexa Fluor 647 rat anti-mouse F4/80 (#565854, 0.04 µg). Isotype controls were PE rat IgG2b, κ isotype control (#553989, 0.04 µg) and Alexa Fluor 647 rat IgG2a, κ isotype control (#557690, 0.04 µg). Compensation control samples were prepared using anti-surface marker antibodies separately, and a no-antibody control sample was prepared. Tubes were flicked to mix and incubated (4°C, 1 h). 1 ml FB was added, tubes were flicked to mix and centrifuged, and supernatant was decanted; this step was performed a total of three times. Several drops of FB were added, and tubes were covered in foil and kept on ice until flow cytometry.

RAW functional experiments. High density (3.3×10^5 cells ml⁻¹) and low density (4.1×10^4 cells ml⁻¹) conditions used 1.35 ml of cell culture per well of a 6-well plate. The very low density condition for reporter cells (5.2×10^3 cells ml⁻¹) used 8 ml of cell culture in a 10 cm dish. Ligand treatments included recombinant mouse IL-10 (R&D Systems #417-ML, 10 ng ml⁻¹ except as indicated in **Figure 2.1b**) at -12 hps, recombinant mouse sTNFR (R&D Systems #763208, 8.3 µg ml⁻¹) at -1 hps, *E. coli* 055:B5 LPS (Sigma-Aldrich, varied doses) or PMA (Cayman Chemical #10008014, varied doses) at 0 hps (36 h post-plating), and BFA (2 µg ml⁻¹, Sigma-Aldrich #B5936) at 1 or 2 hps. Medium was not exchanged during ligand treatments. Cells were harvested at the time indicated for each experiment (3, 6, or 12 hps).

BMM functional experiments. At seven days post-harvest, medium was aspirated, and cells were washed with PBS, incubated in PBS-EDTA (>10 min), and detached by firmly tapping plates and pipetting

in PBS-EDTA. Cells were centrifuged in 50 ml conical tubes (400×g, 5 min), supernatant was discarded, and cells were resuspended in complete RPMI-1640 medium without L929-conditioned medium and re-plated at varying densities: high density in non-treated 6-well plates (Falcon #351146) (3.3×10^5 cells ml⁻¹, 1.35 ml per well) and 1/4th, 1/8th, and 1/16th of high density in non-treated 10 cm dishes (8 ml). Ligand treatments included IL-10 (10 ng ml⁻¹) at -12 hps, sTNFR (8.3 μg ml⁻¹) at -1 hps, LPS (varied doses) at 0 hps, and BFA (2 μg ml⁻¹) at 1 hps. Medium was not exchanged during ligand treatments. Cells were harvested at 3 hps. Conditions were carried out in biological replicates indicated in figure legends.

TNF staining for RAW cells and BMM. Medium was aspirated, and cells were washed with PBS, incubated in FB (37°C, 5 min for RAW cells; 4°C, >10 min for BMM), detached from plates (by gentle scraping for RAW cells; by pipetting for BMM), pipetted into two FACS tubes per sample (one for the anti-TNF stain and one for the isotype control stain), and pelleted by centrifugation (150×g, 5 min for RAW cells; 400×g, 5 min for BMM). Supernatant was decanted, PFA was added, and tubes were flicked to mix and incubated (4°C, 20 min). 1 ml FB was added, tubes were flicked to mix and centrifuged, and supernatant was decanted; this step was performed a total of three times. 1 ml permeabilization wash buffer (PWB; 0.5% saponin and 0.1% BSA in PBS) was added, tubes were flicked to mix and centrifuged, and supernatant was decanted; this step was performed a total of two times. To block, normal mouse serum (10 μl) was added, and tubes were flicked to mix and incubated (room temperature, 20 min). PE-conjugated rat anti-mouse TNF antibody (BD Bioscience #554419, 0.1 μg) or PE-conjugated rat isotype control antibody (BD Bioscience #554685, 0.1 μg) was added, and tubes were flicked to mix and incubated (4°C, 1 h). 1 ml PWB was added, tubes were flicked to mix and centrifuged, and supernatant was decanted; this step was performed a total of three times. Several drops of FB were added, and tubes were covered in foil and kept on ice until flow cytometry.

Preparation of reporter cells for flow cytometry. Medium was aspirated, and reporter cells were washed with PBS, incubated in FB (37°C, 5 min), detached from plates by gentle scraping, pipetted into FACS tubes, and pelleted by centrifugation (150×g, 5 min). Cells were fixed to prevent loss of reporter signal: supernatant was decanted, PFA was added, and tubes were flicked to mix and incubated (4°C, 20 min). 1 ml FB was added, tubes were flicked to mix and centrifuged, and supernatant was decanted; this step was performed a total of three times. Several drops of FB were added, and tubes were covered in foil

and kept on ice until flow cytometry.

Flow cytometry. Samples were run on a BD LSRII or BD LSR Fortessa flow cytometer using the FITC channel for EGFP-RelA, PE-Texas Red channel for mCherry, PE channel for TNF intracellular staining and for CD11b surface staining, and APC channel for F4/80 surface staining. Data were gated using FlowJo software on live (FSC-A vs. SSC-A) and single-cell (FSC-A vs. FSC-H) bases, and fluorescence values were exported from FlowJo and imported into MATLAB for further analysis.

Secretion assay. For the BMM experiment in **Figure 2.5**, supernatant for each condition was collected at 3 hps in chilled Eppendorf tubes, clarified by centrifugation (14,000×g, 4°C, 10 min), pipetted into new chilled tubes, and stored (−80°C). Clarified supernatants were later thawed on ice, and analytes were measured using the Bio-Plex Pro mouse cytokine 23-plex Group I assay (Bio-Rad #M60009RDPD), a Bio-Plex 200 instrument, and Milliplex Analyst software. Each biological replicate was assayed in technical duplicate. Protein concentrations were determined based on a standard curve as described by the manufacturer-provided protocol and values for Lot #64209360. This assay reflects analytes that were secreted between the time of plating and time of BFA treatment. For the condition with sTNFR treatment, in which extracellular TNF is bound by sTNFR, low TNF signal is potentially attributable to an inaccessibility for binding by assay detection reagents.

2.4 Results

TNF expression is heterogeneous and varies with cell density: Macrophage phenotypic heterogeneity has been observed in several studies^{19,23,93}, and non-genetic heterogenous activation has been described in the widely used model cell line RAW 264.7^{73,93}. We selected the RAW 264.7 model system to investigate how perturbations that modulate the response to LPS affect the heterogeneity with which macrophages become activated, as represented by expression of TNF (**Figure 2.1a**). Pre-treatment of cells with IL-10, prior to treatment with LPS, diminished the average intracellular TNF protein expression measured at 3 h post-stimulation (hps), although TNF distributions across IL-10 doses were broad and overlapping (**Figure 2.1b**). TNF expression also was not highly correlated with flow cytometric proxies for cell size, suggesting the heterogeneity was not due to cell cycle asynchrony alone.

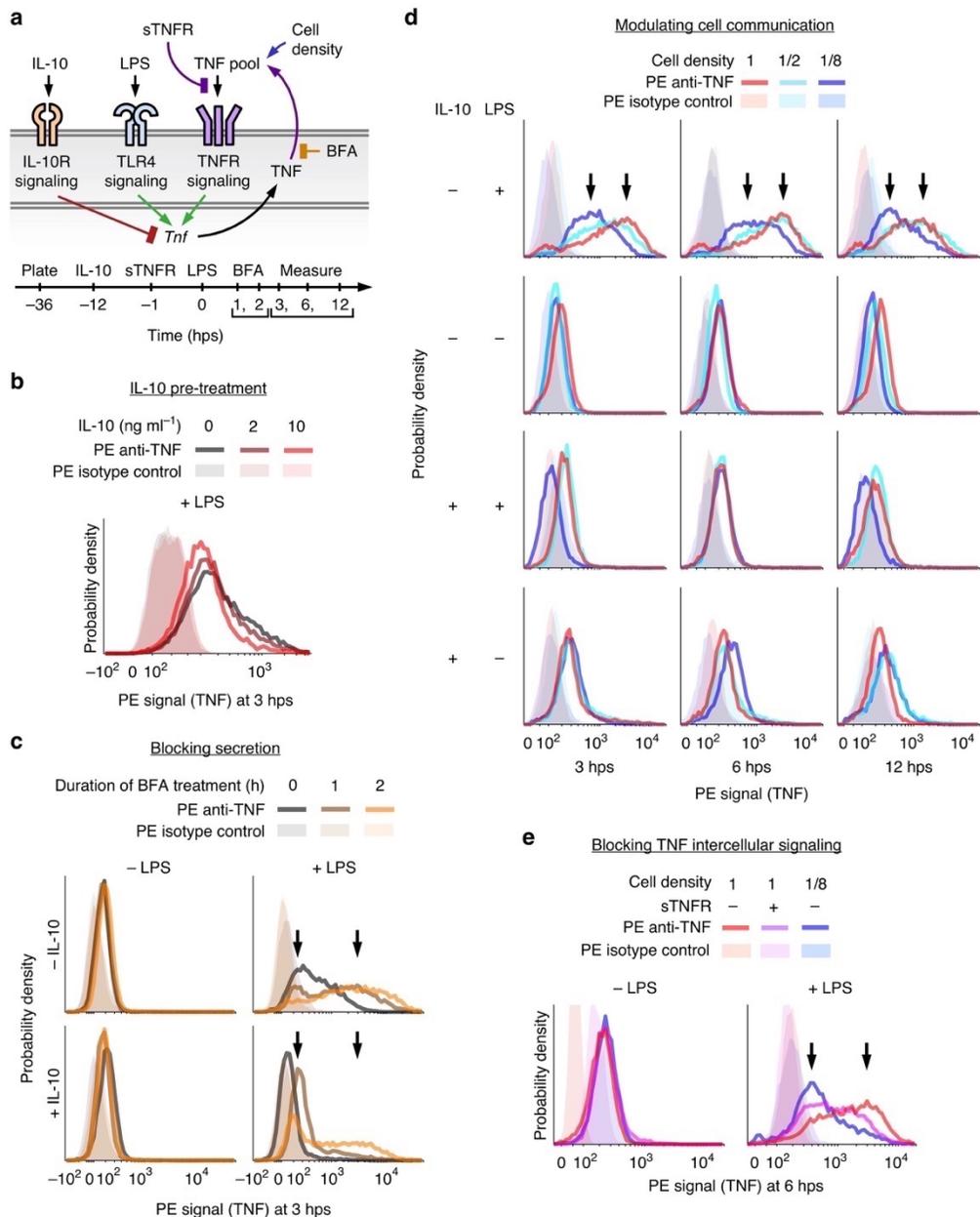


Figure 2.1. The TNF response to LPS is heterogeneous and requires intercellular communication. **a** The diagram summarizes the perturbations and stimuli applied to investigate TNF expression and intercellular communication (hps, hours post-stimulation with LPS). LPS activates TLR4 signaling, which induces TNF expression. IL-10 pretreatment activates IL-10R signaling, which inhibits LPS-induced TNF expression. Secreted TNF activates TNFR signaling, which induces TNF further through intercellular feedback. BFA prevents secretion, causing TNF to accumulate intracellularly. Varying the cell density modulates the concentrations of secreted factors such as TNF. sTNFR binds extracellular TNF and prevents TNFR signaling. **b** IL-10 pretreatment diminishes LPS induced TNF expression. **c** TNF expression is heterogeneous with high-expressing and low-expressing subpopulations. After pretreatment with IL-10 (10 ng ml⁻¹) and/or treatment with LPS (100 ng ml⁻¹), cells were treated with BFA for 1 or 2 h. Arrows in **c–e** indicate low and high modes of the TNF distributions. **d** The full TNF response to LPS requires intercellular communication. **e** Intercellular feedback through secreted TNF is necessary for the full response.

To obtain a more direct readout of TNF production, we applied brefeldin A (BFA) to inhibit anterograde transport from the endoplasmic reticulum to the Golgi apparatus and prevent secretion. We reasoned that if variation in TNF secretion were a main source of heterogeneity, then BFA would diminish heterogeneity, and if gene regulation were the main source, then BFA would exaggerate it. Without LPS, BFA had no appreciable effect on intracellular TNF, indicating no detectable basal TNF production. However, when added after stimulation with LPS, BFA led to wide-ranging accumulation of on average several-fold more TNF cell⁻¹ h⁻¹ (**Figure 2.1c**). Unexpectedly, while most cells accumulated more TNF with longer BFA treatment, some cells accumulated little or no TNF over time. When cells were pre-treated with IL-10 prior to LPS, TNF accumulation was less than when treated with LPS only, as expected, yet TNF accumulation was still wide-ranging for most cells and low for others. Therefore, blocking secretion unmasked substantial hidden variation in TNF production and showed that the cell population includes both high (wide-ranging) and low responders to LPS regardless of IL-10 pre-treatment.

Since LPS-induced TNF intercellular signaling is known to contribute to NF- κ B activity^{31,69-72}, we examined the effect of intercellular communication on TNF production. To modulate communication in a manner that is not biased toward or against specific secreted factors, we varied cell density at plating (full, half, and one-eighth of previously used conditions) as a general handle for tuning the magnitude of coupling between cells. Without IL-10 and with LPS, a cell density-dependent effect was evident (**Figure 2.1d**). At 3 hps, average TNF expression correlated with density. At 6 hps, half and full density were TNF-high while one-eighth density remained low, and by 12 hps expression had decreased for each case. All of the distributions were heterogeneous, but at high density intracellular TNF remained skewed toward high expression over time and at low density it remained skewed toward low expression. With IL-10 or without LPS, little to no intracellular TNF was detectable. Thus, the full response to LPS requires intercellular communication, and cell density-associated effects on TNF production persist over time.

To investigate whether secreted TNF sustains its own LPS-induced expression, cells were pre-treated with excess soluble TNF receptor (sTNFR) to titrate extracellular TNF from binding cell surface receptors. As TNF is bound by sTNFR, TNFR signaling is blocked⁷¹. With LPS, cells at high density with sTNFR pre-treatment expressed TNF at an intermediate level on average—less than at high density without sTNFR, and more than at low density without sTNFR—and the distribution remained heterogeneous

(**Figure 2.1e**). Therefore, although TNF intercellular feedback is required for full TNF production as expected, this mechanism does not explain the wide-ranging expression or the distinct high and low activation states observed. Thus, an additional explanation is required.

The activation states' proportions depend on cell density: To investigate the phenomena described above, we examined regulation upstream of the TNF protein using a previously validated clonal macrophage cell line with two genomically integrated reporters⁷³. Such a reporter system enables one to resolve the dynamics and heterogeneity of individual cell signaling responses. The first reporter is a fusion of enhanced green fluorescent protein (EGFP) and RelA (the p65 subunit of NF- κ B) driven by the *Rela* promoter. In the resting cell state, EGFP-RelA is sequestered primarily in the cytoplasm. Upon activation such as through TLR4 signaling, this protein translocates to the nucleus and induces the transcription of NF- κ B target genes. Above a sufficient LPS dose, EGFP-RelA also induces the expression of endogenous RelA (and of EGFP-RelA) via an intracellular positive feedback loop termed FBD⁷³. Thus, EGFP-RelA tracks both the localization and expression of NF- κ B. The second reporter is a fusion of mCherry and a destabilizing PEST tag driven by the *Tnf* promoter. *mCherry* RNA lacks the *Tnf*-specific 3' UTR and is decoupled from *Tnf*-specific post-transcriptional regulation. These features make the mCherry protein a proxy for transcription from the *Tnf* promoter, rather than for downstream TNF protein expression, after accounting for the time delay for mCherry translation and maturation⁹⁸.

We utilized this reporter system to examine RelA expression and localization and *Tnf* promoter activity under the perturbations used above (**Figure 2.2a**). In all cases without LPS, EGFP-RelA and mCherry expression were low. With LPS, the mCherry distribution shifted from unimodal to bimodal, consistent with the observed expression of endogenous TNF (**Figure 2.1c**). At the three high density conditions most cells were mCherry-high, and at low density a greater proportion was mCherry-low. EGFP-RelA mirrored the pattern for mCherry. We also observed that average levels of TNF and mCherry ranked differently across conditions. In particular, cells at high density with IL-10 and LPS treatment had greater mCherry expression than did cells at low density with LPS treatment (**Figure 2.2a**), yet this order was reversed for TNF expression (**Figure 2.1**). The difference could be due to post-transcriptional downregulation of *Tnf* via IL-10R signaling⁹⁹⁻¹⁰², which would diminish TNF protein more than *Tnf* transcription, and mCherry is a proxy for the latter.

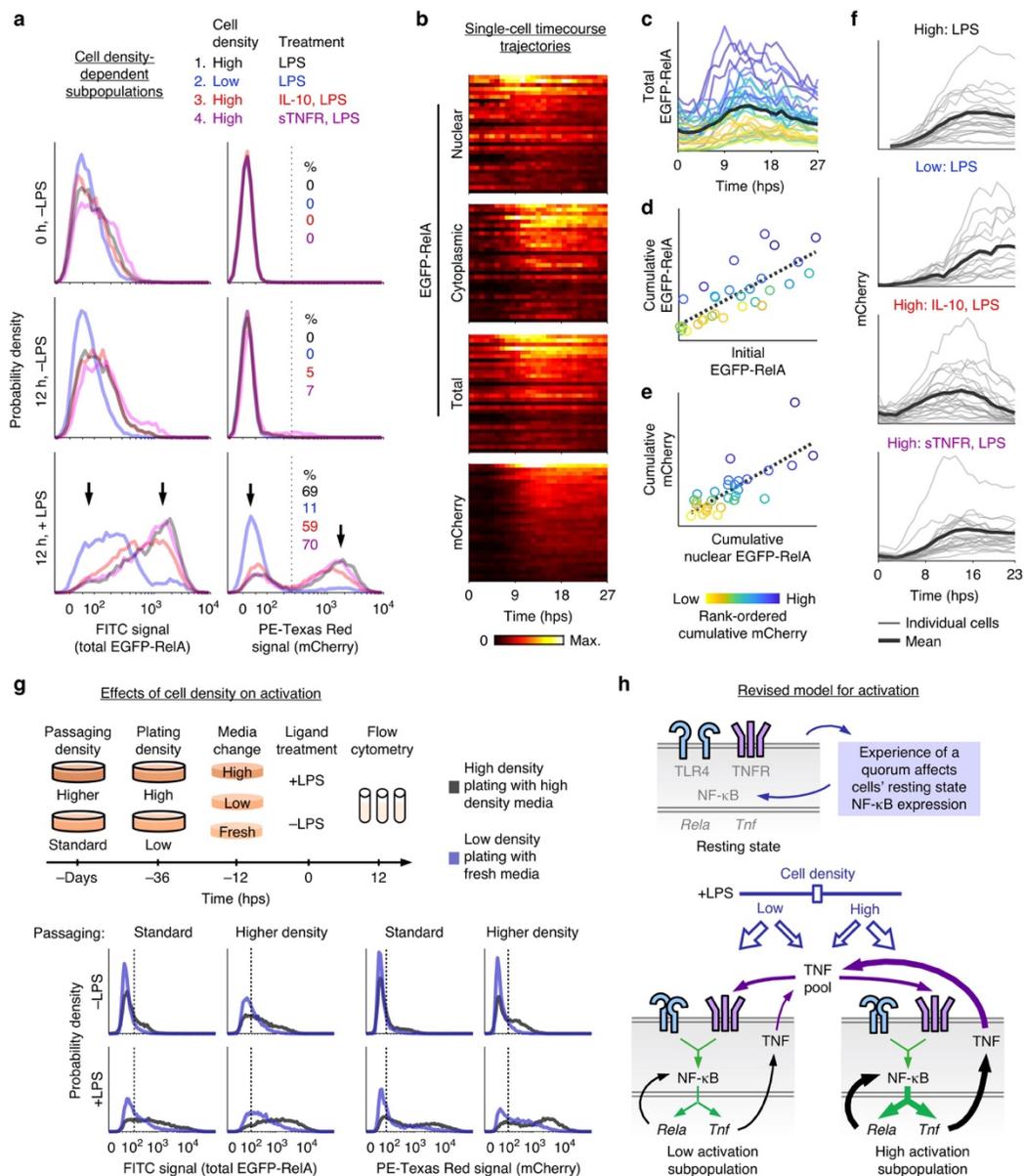


Figure 2.2. Cell density modulates the heterogeneity of macrophage activation. **a** Reporter protein fluorescence was measured by flow cytometry for the indicated cell densities, time points, and ligand treatments. Percentages of highly activated cells were determined using a threshold (dotted vertical line) at the nadir between the two modes (arrows) of mCherry distributions. **b** Reporter trajectories for $n = 30$ cells at high density after treatment with sTNFR and LPS. Cells are ordered by cumulative mCherry expression and color-coded by fluorescence magnitude within heat maps. **c** Single-cell trajectories for total EGFP-RelA expression. The mean is in bold. **d–e** Relationship between initial and cumulative total EGFP-RelA ($R^2 = 0.61$, one-tailed permutation test $p = 2 \times 10^{-7}$) and between cumulative nuclear EGFP-RelA and cumulative mCherry ($R^2 = 0.59$, one-tailed permutation test $p = 3 \times 10^{-7}$). Dotted lines are linear fits, and axes are linearly scaled. In **c–e**, color-coding denotes rank-ordered cumulative mCherry expression. **f** Single-cell mCherry trajectories, with the mean in bold. Values are in a.u. specific to each panel. **g** Effect of culture density-associated conditions on macrophage activation heterogeneity. Fluorescence units are comparable within each reporter protein and passaging density. Dotted vertical lines distinguish low and high activation. **h** Revised conceptual model for macrophage activation with differently activated cell density-dependent subpopulations.

Altogether, macrophage activation was bimodal, and subpopulation proportions varied with cell density. This observation held under perturbations to TNF-regulating pathways such as TNFR and IL-10R signaling. The dependence of the decision to become highly activated on cell density has some resemblance to bacterial QS, in that the phenotypes of individual cells are determined by information shared by the population. However, an important distinction is that in QS essentially all cells become activated if a threshold concentration of QS molecule is surpassed²⁷, whereas here the *proportion* of highly activated cells increases with density. Therefore, because only a fraction of the cells become licensed to reach a state of high activation, the population exhibits an analog response rather than a digital one. To distinguish these two phenomena, we refer to the macrophage behavior as quorum licensing (QL).

Single-cell analysis supports a revised model for activation: We next investigated whether heterogeneous activation was due to variation in the magnitude and/or timing of the response to LPS. Variation in magnitude would indicate a role for intrinsic or extrinsic noise, due to stochastic fluctuations or to deterministic outcomes of variation in initial (pre-LPS) conditions, respectively. Variation in timing could indicate a domino effect where early high-expressing cells activate other cells. To track the dynamics of EGFP-RelA expression and localization and mCherry expression, cells were stimulated and monitored over a one-day timecourse using confocal laser-scanning microscopy. Quantification of individual trajectories showed that *Tnf* promoter activation varied primarily in magnitude rather than timing (**Figure 2.2b**). For EGFP-RelA (at high density with sTNFR and LPS), the mean expression increased and peaked at 14 ± 4 hps (\pm standard deviation among cells), after peak nuclear signal (10 ± 6 hps) and around peak cytoplasmic signal (15 ± 4 hps). The peak signal was greater in the cytoplasm than in the nucleus (1.4 ± 0.4 fold), and depletion of the nuclear portion mid-timecourse coincided with cytoplasmic accumulation. The expression of mCherry was initially low and then increased and peaked at 18 ± 5 hps. A small subpopulation of cells expressed very little mCherry or EGFP-RelA (consistent with flow cytometry observations for TNF and the reporters), and reporter expression appeared unrelated to which cells were in physical contact. Intriguingly, although high induction of both reporters co-occurred in the same cells, the bimodality in TNF expression (measured at 3 hps in **Figure 2.1c**) was evident *before* the observed increase in EGFP-RelA (**Figure 2.2b**), even considering the ~ 1 h chromophore maturation time. Since target gene expression would be expected to increase *after* an increase in the expression of a transcriptional regulator, the observed sequence

indicates that the FBD switch cannot be causal for the early TNF burst observed. Instead, these events are conditionally independent, i.e., regulated by an upstream process and not by each other.

The microscopy analysis revealed additional dynamical features of the single-cell responses. For nuclear-localized EGFP-RelA, while the average profile underwent an overall increase and subsequent decrease, some cells showed multiple peaks of varied amplitudes in agreement with a recent study⁹¹. There was also variation in an apparent nuclear reservoir of transcription factor, in agreement with another study⁹⁰. For total EGFP-RelA, pre-LPS expression varied and was correlated with post-LPS cumulative expression (**Figure 2.2c–d**). The predictive power of this initial condition indicates a role for extrinsic noise (variation initially present in the system) in determining the post-FBD amount of transcription factor. Cumulative nuclear EGFP-RelA and cumulative mCherry were also correlated, as expected for a transcription factor and target gene (**Figure 2.2e**). These outcomes indicate that extrinsic noise in RelA expression propagates to activity at the *Tnf* promoter. To further examine promoter activity, we quantified mCherry trajectories under each perturbation (**Figure 2.2f**). At high density with LPS, mCherry increased for 16 ± 2 hps, indicating continued transcription after the intracellular TNF protein showed a decrease (**Figure 2.1d**). This protein decrease is consistent with known post-transcriptional mechanisms that downregulate *Tnf*^{f4}, and such simultaneously opposing transcriptional and post-transcriptional regulation represents a type of control that has been compared to operating both the throttle and brake pedals of a vehicle¹⁰³. Cells at low density took more time to reach peak mCherry signal compared to at high density. Cells pretreated with IL-10 increased in mCherry for 12 ± 3 hps and then decreased toward basal levels. Cells treated with sTNFR responded to LPS similarly to cells without this antagonist. Together, the measurements from flow cytometry and microscopy show how TNF is differently regulated under each perturbation.

Another phenomenon related to cell density was that EGFP-RelA levels differed between low and high density both with and without LPS stimulation (**Figure 2.2a**). This difference suggested that secreted factors might affect cells' resting states in a way that predicts the response to LPS. To more carefully investigate how high density-associated conditions impact the patterns described above, we exposed cells to combinations of standard or higher density passaging (for at least three days), low or high density plating (at 36 h pre-LPS), and different conditioned media (fresh, low density, or high density; at 12 h pre-LPS). In general, exposure to more high density-associated conditions increased both the basal and induced

reporter expression (**Figure 2.2g**). For cells that were passaged at higher density and later provided conditioned medium from cells cultured at high density, the outcomes were similar regardless of the cell density at plating (which determines the number of cells present), indicating that sustained exposure to high density-associated secreted factors was sufficient to prime and enable full LPS-induced activation. For the highest-density combination of conditions, reporter bimodality was prominent and right-shifted, even prior to LPS stimulation (**Figure 2.2g**). The effect of passaging, which took place several days prior to plating at 36 h pre-LPS, indicates that the effects of culture density are heritable across cell generations. These observations motivate our proposal of a revised model for macrophage activation by QL, in which modifications to the extracellular milieu that occur during the resting state modulate the propensity for cells to become highly activated (**Figure 2.2h**).

A new computational model explains heterogeneous activation: To integrate our observations with prior knowledge on macrophage activation, we developed a dynamical model for the intracellular and intercellular signaling network. We reasoned that such a framework could enable us to investigate how TNF is regulated and whether heterogeneity confers advantages to a population. Key aspects of model development were to include the most essential components, concisely portray biochemical processes, identify salient features of the data, fit parameters, and evaluate the extent to which simulations could explain the observations. To start, we examined studies on NF- κ B and TNF^{31,71-73,79,84,90,91,99-101,104-109} and synthesized this information to produce a preliminary system of ordinary differential equations representing a cell that can inducibly express, secrete, and sense TNF. We reduced the model to decrease complexity¹¹⁰, and proposed, evaluated, and refined network topologies and their corresponding formulations (sets of equations) to produce the network depicted in **Figure 2.3a**.

Extrinsic noise featured prominently in the data. Therefore, we hypothesized that it might be possible to train a homogeneous model (a one-cell model) based on mean flow cytometry and microscopy data and subsequently to incorporate heterogeneity among cells. Parameters were split into two groups for separate rounds of fitting using parameter sweeps, multi-objective optimization, and an evolutionary algorithm (**Appendix 2**). The first round used a focused model of NF- κ B activation with cell-intrinsic influences: TLR4 signaling; NF- κ B activation, nucleocytoplasmic translocation, and inactivation; I κ B expression; and FBD. A fit to data for sTNFR and LPS (to exclude the influence of TNFR signaling) resulted

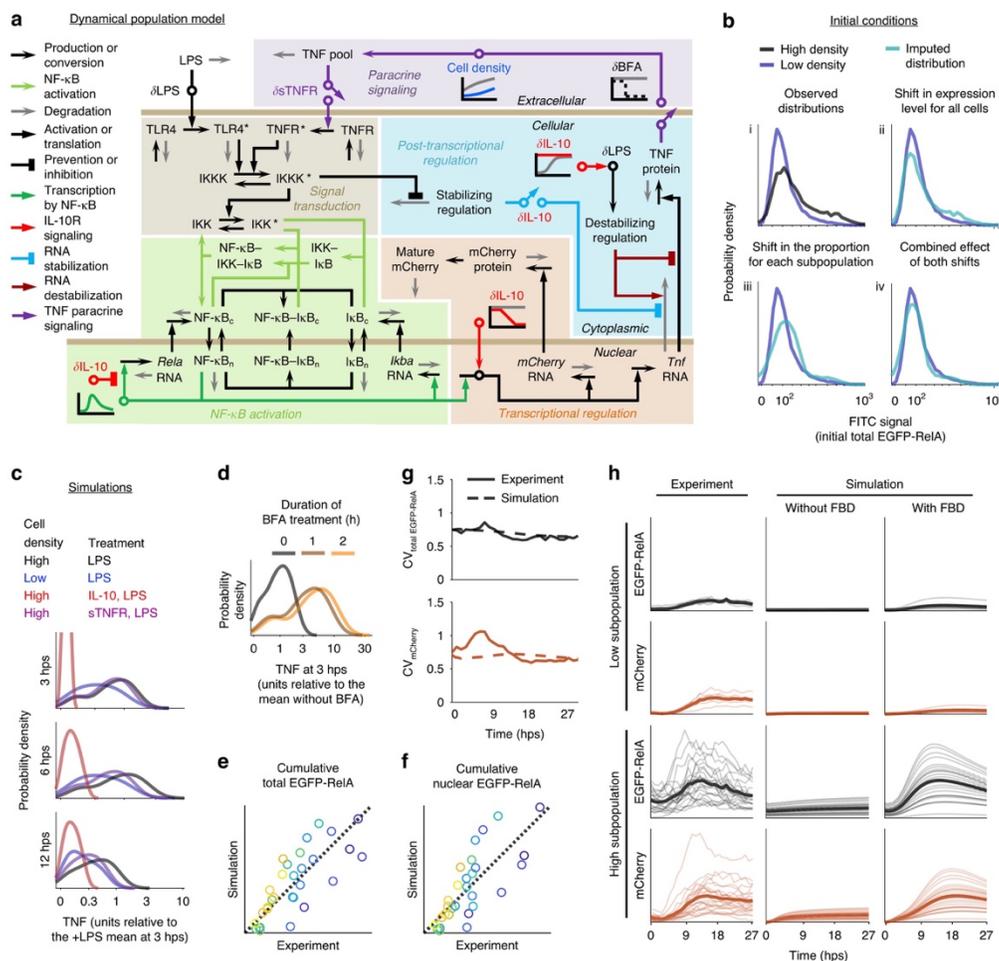


Figure 2.3. A dynamical population model explains intracellular and extracellular signaling and regulation. **a** The diagram summarizes variables, reactions, and mechanisms in the model of macrophage activation. Symbols: bold non-italicized text for variables; horizontal bars for compartment boundaries; n for nuclear and c for cytoplasmic; asterisks for activated receptors or kinases; delta symbol and circles for perturbation-specific effects; diagonal arrows for a perturbation's effect; and graphs for time-dependent processes. **b** Generation of comparable distributions of initial values. The observed distributions in (i) are in non-comparable units. To initiate simulations with NF- κ B distributions that vary between high and low cell density and that match experimental observations for EGFP-RelA, we impute from an observed high density distribution (black) a low density distribution (teal) that matches the observed low density distribution (blue). The transform shifts the distribution (ii) and adjusts the proportion of cells in high vs. low states using a Gaussian mixture model of two populations fit to the high density distribution (iii), such that the combined transform (iv) generates an imputed distribution matching the observed low-density distribution (blue) with units relatable to high density (black). **c–d** Simulated TNF distributions from the calibrated model match experimental trends for cell density and ligand conditions and BFA conditions. All cells have the topology in **a** and are heterogeneous in initial value and transcription of *Rela* RNA and initial value of cytoplasmic NF- κ B-I κ B, derived from **b**. **e–h** Comparison of simulated and experimental outcomes for $n = 30$ cells at high density with sTNFR and LPS treatment. **e–f** Cumulative total EGFP-RelA ($R^2 = 0.61$, one-tailed permutation test $p = 3 \times 10^{-7}$) and cumulative nuclear EGFP-RelA ($R^2 = 0.60$, one-tailed permutation test $p = 4 \times 10^{-7}$), calculated using time-integrated values. The dotted diagonal is the identity line, and axes are linearly scaled. **g** Coefficient of variation (CV) in mCherry and total EGFP-RelA expression over time. **h** Trajectories were grouped post hoc by high or low activation. Mean values are in bold. Simulations are shown with and without FBD.

in a family of parameter subsets with similar outcomes. The second round used the full model, incorporating *Tnf* transcriptional, post-transcriptional, and translational regulation; mCherry expression; effects of IL-10; TNF secretion and its inhibition by BFA; population growth; and TNFR signaling. Each perturbation was simulated by modifying equations to capture its mode of action. A simultaneous fit to all of the data, in which some parameters were constrained to subsets from the first round and others underwent a free search, yielded a suitable outcome. We next adapted the model to represent a population coupled by TNF intercellular feedback, in which each cell varies in the basal state with extrinsic noise in NF- κ B. Initial values for NF- κ B were assigned based on EGFP-RelA confocal microscopy measurements (**Figure 2.2d**). Additionally, we observed that EGFP-RelA distributions differed between high and low density by a shift in values and a shift between the two activation modes (**Figure 2.3b**), and applying a transformation in silico to the initial values at high density could produce a distribution in comparable units for initial values at low density. In summary, the model was trained on homogeneous post-LPS data and initialized with heterogeneous pre-LPS data.

Remarkably, the use of varied pre-LPS EGFP-RelA levels, and the transformation of this distribution between high and low density conditions, enabled resulting simulations to capture the heterogeneity in the data, including intracellular TNF expression across perturbations and over time (**Figure 2.3c**, compare to **Figure 2.1d–e**) and TNF accumulation following treatment with BFA (**Figure 2.3d**, compare to **Figure 2.1c**). These simulations also accounted for the majority of the variation in cumulative transcription factor expression and localization (**Figure 2.3e–f**). Furthermore, they closely tracked the distributions and trajectories of reporter expression (**Figure 2.3g–h**) and supported the conclusion that both the high and low activation subpopulations underwent FBD (**Figure 2.3h**). These findings support our strategy of first training the model on mean data and then introducing heterogeneity in a way that incorporates how the density-dependent pre-LPS state predicts the response to LPS.

Model validated by observations from a distinct test dataset: We sought to validate the model by testing whether it could predict responses to conditions not included in model development or fitting. To this end, we simulated populations at different densities and LPS doses. FBD was included for LPS treatment at or above 1 ng ml⁻¹ (our estimate for a threshold at which this mechanism is active, based on the original study⁷³). To obtain a test dataset, reporter cells were assayed under the same conditions

(Figure 2.4a). Model predictions were broadly in agreement with the data: distributions for EGFP-RelA and mCherry had more rightward shifts with higher LPS doses, and two subpopulations were present with density-dependent proportions across doses. At lower doses and densities, the high-activation subpopulation had diminished reporter expression, resulting in more overlap between subpopulations.

To test if any of the above observations were specific to TLR4 signaling, we evaluated the same conditions but for the stimulus used phorbol 12-myristate 13-acetate (PMA), a membrane-permeable small molecule that activates protein kinase C (PKC), which then activates IKK and by extension NF- κ B. As was observed with LPS stimulation, PMA-induced expression of both reporters was bimodal with cell density dependence, and mCherry expression was ligand dose-dependent (**Figure 2.4a**). Therefore, QL-associated activation of the *Tnf* promoter via NF- κ B does not require that NF- κ B be activated via TLR4. Interestingly, unlike trends observed with LPS, EGFP-RelA expression following PMA addition differed little from the case without PMA, which indicates that FBD was induced by NF- κ B activation via TLR4 but not via PKC.

Simulations elucidate roles for TNF regulatory mechanisms: We next applied the validated model to examine roles of TNF regulatory mechanisms. First, we evaluated the robustness of simulations to global parameter variation by sampling parameter values from distributions with different coefficients of variation (CV) centered on fitted values. Variation was tolerated to some extent—consistent with a general feature of systems biology models—sloppy parameter sensitivities¹⁴, in which various combinations of values yield similar fits. However, for variation beyond a CV of ~ 0.1 – 0.2 the outcomes began to diverge, suggesting sensitivity to certain mechanisms. We then used the model to explore how varying the effect magnitudes of six individual mechanisms (#1–6 in **Figure 2.4b**) would affect key readouts: the mean cumulative expression of NF- κ B and mCherry (phenotypic consequences) and the total cumulative secretion of TNF (a functional consequence). The mCherry and TNF readouts were sensitive to transcriptional induction by NF- κ B (#2), as expected. The TNF readout was the most sensitive to TLR4 signal transduction (#1), stabilizing regulation (#3), and destabilizing regulation (#4). Notably, FBD (#6) was the only mechanism to which the NF- κ B readout was sensitive, indicating an apparent buffer from other processes, and all three readouts were affected by FBD, indicating FBD has wide-reaching effects. Also notable was that TNFR signal transduction (#5) affected the mCherry and TNF readouts only modestly, suggesting that even if cells were more tightly coupled by greater TNFR signal transduction, there would

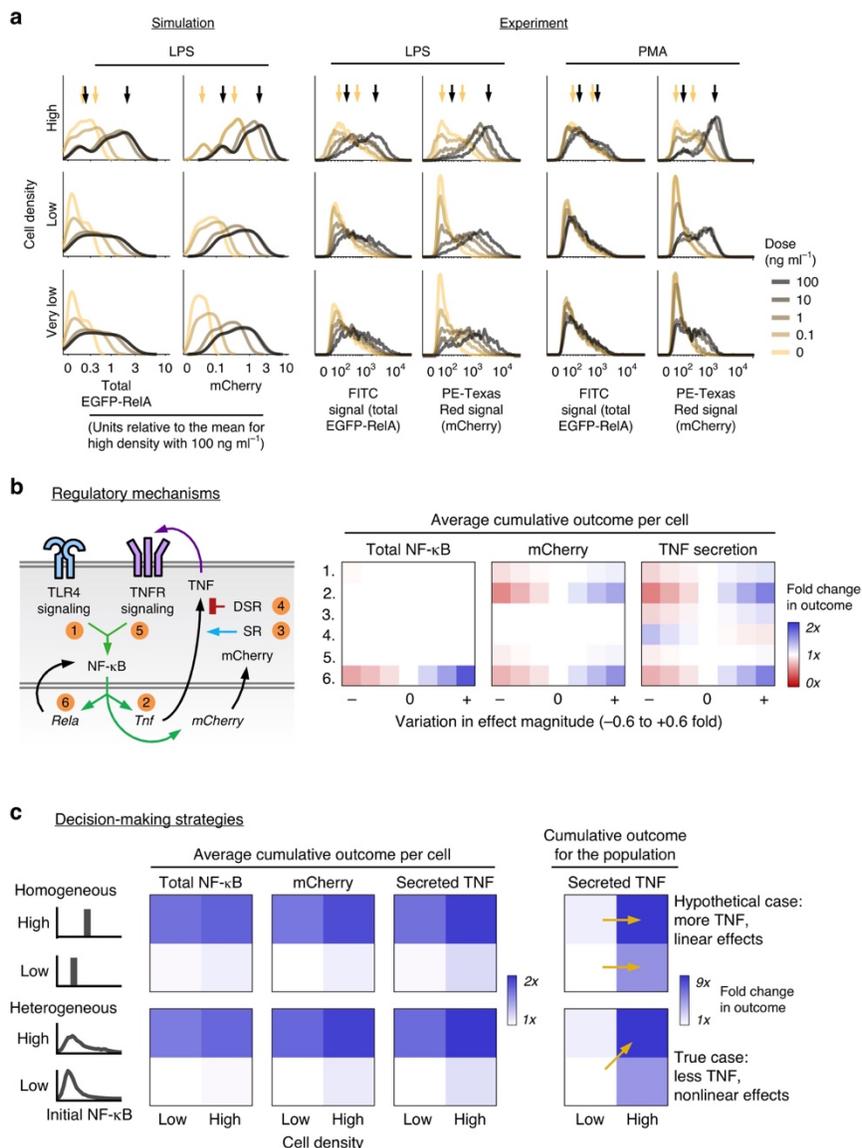


Figure 2.4. Mechanisms of TNF regulation differ in phenotypic and functional consequences. **a** Predicted and measured distributions of reporter expression across cell densities and doses of LPS or PMA at 12 hps. Very low density is 1/8th of low density. Arrows denote modes of the distributions for high density without stimulus or with 100 ng ml^{-1} of LPS or PMA. For simulations: low and very low density used the same initial values; mCherry distributions without stimulus were obtained without FBD and without intercellular feedback; and FBD was applied for LPS doses at and above 1 ng ml^{-1} . **b** Simulated outcomes after individually varying the effect magnitudes of mechanisms that regulate TNF, with perturbations numbered and depicted in the diagram. Outcomes were assessed using the 12 hps time-integrated amounts of total EGFP-RelA, mCherry, and secreted TNF (cumulative secreted flux) per cell at high density. Conditions corresponding to the zero on the x-axis indicate base case effect magnitudes ($1\times$ on the color scale). Abbreviations: stabilizing regulation (SR) and destabilizing regulation (DSR). **c** Comparison of activation for homogeneous (hypothetical) or heterogeneous (observed) initial distributions of transcription factor expression. For each readout, a value of $1\times$ was set for the outcome given a heterogeneous population at low density with low initial values. Homogeneous initial values were set to the mean of corresponding heterogeneous distributions. In the right panel, the effect of the number of cells is incorporated into the total amount of secreted TNF. Arrows indicate comparisons from the main text, and key trends are noted on the right.

not be a large functional gain. Overall, the analysis showed that the most consequential mechanisms were FBD and transcriptional induction by NF- κ B.

To investigate whether the observed heterogeneity could produce outcomes that differ from a hypothetical homogeneous population, we examined scenarios of high and low cell density in combination with high and low resting state levels of NF- κ B. Homogeneous cells were assigned the mean value (without extrinsic noise) of NF- κ B derived from the corresponding heterogeneous distribution, and LPS-induced outcomes were assessed by the three readouts. Unexpectedly, across scenarios, the homogeneous population was slightly more activated than was the heterogeneous one (**Figure 2.4c** left panel). Within both populations, initial conditions had a larger effect on the readouts for cells than did the number of cells (depicted by a greater distinction between rows than columns in the heatmaps). For total TNF secretion, both factors were important, and the coupling of initial NF- κ B to cell density produced a greater fold change in TNF than would be the case if these factors were independent (**Figure 2.4c** right panel, diagonal arrow vs. horizontal arrows). This coupling was also more consequential than the distinction between heterogeneity and homogeneity. Therefore, we conclude that it is not heterogeneity per se that is important, but rather that heterogeneity is regulated in a density-dependent manner; density controls the partitioning of a population of macrophages into high and low responders, yielding an effective strategy for driving population-level activation nonlinearly with cell density.

Primary macrophages employ quorum licensing: To determine whether QL also occurs in primary cells, we generated murine bone marrow-derived macrophages (BMM). Cells were plated at different densities, subjected to combinations of stimuli, and assayed for intracellular TNF expression at 3 hps. We observed that BMM exhibit the key features of QL: LPS-induced TNF expression was bimodal; the proportion of cells in the high activation mode increased with cell density; and TNF expression was bimodal even when cells were pre-treated with sTNFR or IL-10 (**Figure 2.5a**). This outcome is consistent with observations in RAW cells and reporter cells (**Figures 2.1–2.2**) that QL is independent of canonical TNF-regulating pathways including LPS-induced TNF intercellular feedback.

Lastly, to investigate how cell density might impact the products secreted by BMM, we used a multiplexed assay to evaluate cytokines and chemokines in the supernatant (**Figure 2.5b**). At high cell density, there was a large boost in secreted TNF. This nonlinear scaling of collective TNF secretion with

cell density supports our proposed explanation that this is a central functional consequence of QL. Additionally, the assay ruled out several analytes—those that were not basally secreted and thus cannot coordinate the population to be poised for activation, and those that did not trend with density—as candidate mediators of QL. Of the twenty-three analytes evaluated, only the five chemokines examined (KC, MCP-1, MIP-1 α , MIP-1 β , and RANTES) were detected in the basal state and generally increased with cell density. Although these factors exhibit several necessary features of a QL mediator, chemokine receptor signaling activates multiple intracellular pathways, and evaluation of causality requires substantial further investigation. Nonetheless, these observations demonstrate that QL occurs in primary macrophages, mediating a nonlinear relationship between cell density and population-level TNF production.

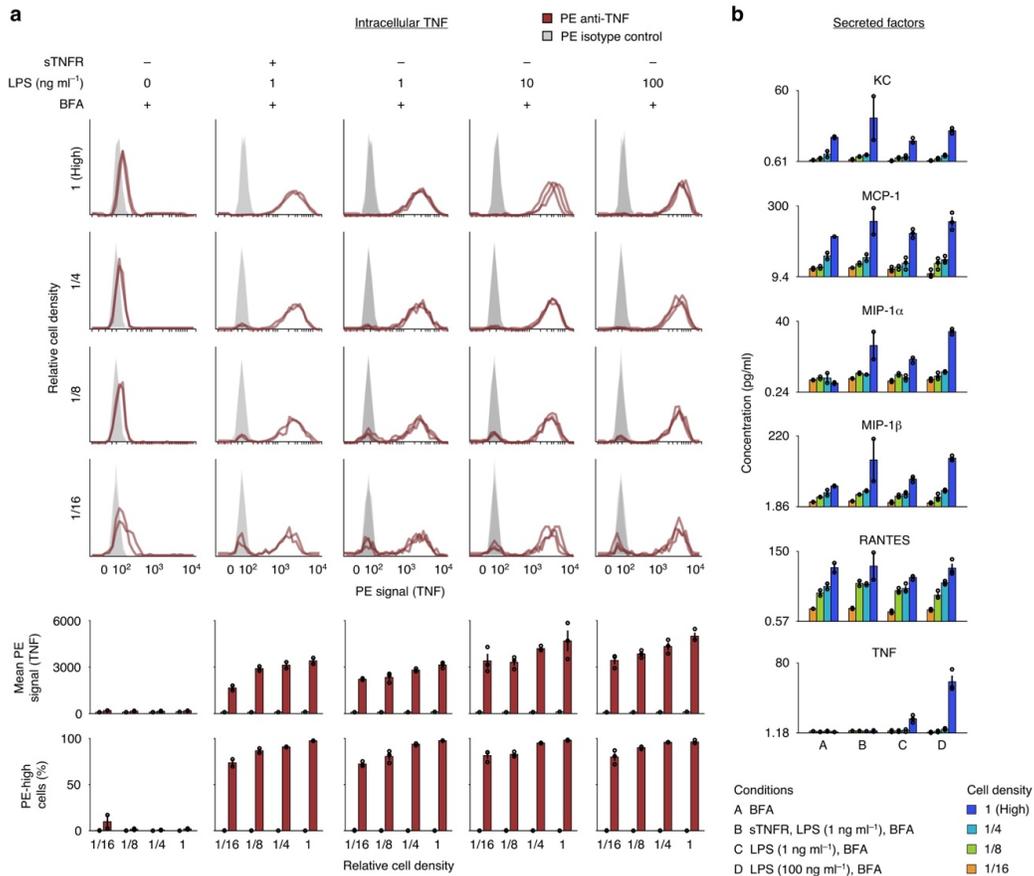


Figure 2.5. Primary macrophages regulate TNF production through quorum licensing. **a** Bone marrow-derived macrophages exhibit cell density-dependent bimodality in LPS-induced TNF expression. The first two conditions (columns) are in biological duplicate and the other three conditions are in biological triplicate. **b** Secreted factors were measured in cell culture supernatants. The minimum value on each y-axis is the observed lower limit of detection of the assay. Bar graphs represent the mean of the biological replicates and S.E.M.

2.5 Discussion

This study explores the observation that within a population of genetically identical macrophages, a potent activating stimulus (LPS) drives high expression of TNF in only a subset of cells. The observation of TNF heterogeneity is consistent with a recent study that showed cellular responses vary widely in response to such cues³⁰. In examining the regulation that underlies this variation, we observed heterogeneity in transcription factor expression and localization, *Tnf* promoter activation, and TNF expression. Furthermore, the measured distributions showed states for high (wide-ranging) and low activation. This bimodality was tunable by modulating conditions related to cell density, indicating a role for intercellular communication potentially via secreted proteins, metabolites, or extracellular vesicles¹¹¹. Using single-cell tracking, dynamical modeling, and primary cell experiments, we proposed, developed, and validated a revised model for activation by QL, in which a population's experience of density over time impacts the extent to which cells become poised for activation.

Although bacterial QS provides a useful conceptual reference point, QL more closely resembles other types of phenotypic bimodality, which have been observed in native and synthetic contexts, across species, and stem from various mechanisms. Some of these mechanisms do not require intercellular communication. For example, continuous variation in protein expression in a population can produce digital outcomes for kinase activity in cells due to activation thresholds¹¹², and intracellular positive feedback on synthetic promoters can support stochastic activation through transcriptional bursting¹¹³. Other instances of phenotypic bimodality do use intercellular communication. For example, in T cells, secretion of the cytokine IL-2 feeds back through IL-2R signaling and promotes further IL-2 expression. As the sensing involves capturing extracellular IL-2, feedback is confined locally and only some cells sense high amounts¹¹⁴; TNF intercellular signaling similarly involves competitive uptake of TNF¹¹⁵. In the case of a synthetic genetic circuit in yeast with topological similarity to TNF intercellular signaling, two modes of activation were observed with proportions that varied with cell density⁵⁸. Mathematical modeling analysis revealed that if cells rely on autocrine signaling more at low density than at high density, and if at low density few cells activate through autocrine signaling alone, then at high density a greater proportion of cells should activate as the balance shifts toward paracrine signaling via a shared cue. QL could belong to this same family of bimodal activation behaviors. It could even comprise a broader phenomenon—many genes are

regulated by pathways that overlap with those that regulate expression of TNF¹⁰¹, and single-cell RNA-seq analyses have revealed bimodality in hundreds of immune response genes^{20,22}. In addition, macrophages were recently found to use a quorum-based mechanism to *resolve* inflammatory processes via production of nitric oxide³³. Investigating the various mechanisms by which heterogeneity and coordination interact represents an exciting avenue for future investigation.

Our study suggests nuanced interpretations for several phenomena. First, it was recently shown that NF- κ B can induce RelA in intracellular positive feedback termed the FBD switch⁷³. We found that this effect persists over the long-term and coincides with sustained *Tnf* promoter activity. However, these dynamics differ from those of intracellular TNF protein, which undergoes a burst (in which bimodality in its production is evident) before FBD takes effect, and then begins to decline before mCherry reaches peak expression. Second, phenotypic variation among genetically identical cells is often attributed to stochasticity (intrinsic noise). However, in this system, the predictive power of the pre-LPS state indicates a substantial role for extrinsic noise. If additional species were measured to more fully characterize the pre-LPS state, we anticipate that the multivariate initial conditions would provide further predictive power as to how individual cells are poised to respond to LPS. Third, we observed an inherited density-dependent propensity for activation involving an accumulation of transcription factor in the resting state. This result suggests that one way in which a population can become and remain at least temporarily poised for high activation is for cells to stockpile a transcription factor (or conceivably other molecules), and in the absence of continued high density-associated conditions, these levels would eventually decrease. Whatever the underlying mechanisms are for maintaining long-term activation propensity, a strategy that enables cells to calibrate their response to future threats by integrating their experience of the surroundings over time (including information about whether other cells have been recruited to that site) could provide functional utility.

A fourth insight is that even though the role of TNF in intercellular communication is established^{31,69-72}, we found that as an intercellular signal, TNF had a modest influence on FBD and *Tnf* promoter activity. Intercellular TNF signaling had a larger effect on TNF protein expression, consistent with known mechanisms through which TNFR signaling enhances TNF expression through post-transcriptional regulation¹⁰⁷. Since chronic TNF production is implicated in various diseases, therapeutic strategies have focused on blocking TNF from binding surface receptors by administering anti-TNF antibodies or sTNFR¹¹⁶

or by administering antagonists of TLR4 or associated proteins¹¹⁷. While we did not attribute QL to specific mediators, if these factors can be identified in future work, they could represent targets for immunomodulation. For example, the BMM secretion assay showed that chemokines accumulate with cell density prior to LPS treatment—a necessary feature of any candidate mediator. It has been shown that chemokines mediate leukocyte recruitment to sites of infection¹¹⁸⁻¹²¹, chemokine receptor signaling regulates NF- κ B activity¹²²⁻¹²⁴, and NF- κ B itself induces the expression of certain chemokines¹²⁵. While the observations from the secretion assay are intriguing, there are many chemokines and they signal through overlapping pathways, and thus our findings identify an opportunity for future investigation as to whether these molecules have a role in QL.

Looking beyond the scope of this investigation, it is interesting to speculate whether QL may be adaptive for immune function. Such a mechanism could provide a way for a cell population to preemptively coordinate a response to microbial incursion prior to LPS-induced intercellular signaling. As an example, we consider a scenario in which a wound is experienced, resulting in microbial incursion. Since tissue damage can immediately trigger local sterile inflammation (via damage-associated molecular patterns, DAMPs)¹²⁶, macrophages among other cells could locally accumulate and become primed for high activation through QL. Then, should replication of invading microbes produce more pathogen-associated molecular patterns (PAMPs) such as LPS, the local inflammatory response would escalate. Another consequence would be to limit potent inflammation to local environments. In the scenario posed above—a wound experiencing an infection—some DAMPs could travel to sites that are remote from the wound, even if the microbes are limited to the wound site. QL could act to limit the most potent macrophage-mediated responses to sites characterized by *both* the presence of DAMPs and the local accumulation of macrophages, nonlinearly driving local cytokine production. If macrophage recruitment could be coupled to activation in this way, then QL would enhance the specificity of this potent but potentially harmful facet of innate immune function. Conversely however, for conditions such as sarcoidosis, fibrosis, and atherosclerosis^{127,128} characterized by abnormal accumulation of macrophages and related cells, QL would be maladaptive by supporting chronic inflammation. These ideas each comprise compelling avenues for future investigation, building upon the insights gained here into mechanisms linking heterogenous immune cell activation to intercellular communication.

2.6 Acknowledgements

I thank Yishan Chuang for mentorship and collaboration; Daniel Wells, William Kath, and Aushra Abouzeid for guidance on model fitting; Michelle Hung, Kelly Sarnese, Taylor Dolberg, Patrick Donahue, and Devin Stranford for guidance on experiments; Keith MacRenaris and Jessica Hornick for assistance with microscopy; Patrick Donahue and Hailey Edelstein for animal protocol management, handling, and dissections; Ziyin Huang for assistance with the secretion assay; Sean Allen for reagents; and Sarah Stainbrook, Gokay Yamankurt, and my lab colleagues for useful discussions.

CHAPTER 3. Modeling gene expression by synthetic transcription factors

A version of this chapter was previously published as:

Donahue P.S., Draut J.W.*, Muldoon J.J.*, Edelstein H.I.*, Bagheri N., Leonard J.N. The COMET toolkit for composing customizable genetic programs in mammalian cells. *Nat Commun* **7**, 779 (2020).¹²⁹ *Equal contributions

3.1 Abstract

Advances in the design of synthetic genetic programs have enabled the engineering of cells with new sense-and-respond capabilities. As increasingly sophisticated functions are developed for fundamental research and medical applications, there remains a need for cells to carry out programs precisely and as intended. Achieving this level of control requires molecular components that are well-characterized, tunable, and can be predictably arranged. To this end, I collaborated with colleagues on the development of the Composable Mammalian Elements of Transcription (COMET) toolkit, comprising zinc finger transcription factors (ZF-TFs) and cognate promoters for regulating gene expression. I developed concise mathematical models that explain the experimental data, and which provide a basis to understand mechanisms for tuning gene expression based on modular features of the TFs (zinc finger, mutant variant, and activation domain) and promoters (number, spacing, and arrangement of binding sites). The analysis shows how COMET benefits from several properties, including certain dose response characteristics, direct relationships between model parameters and physical aspects of TFs and promoters, and a unique dual mechanism for inhibitory TFs. By applying a formal approach to define rules and parameters that govern the activities of these genetic parts, I identified distinct advantageous properties for this system, which as investigated in **Chapter 4** enable the design of complex genetic programs in mammalian cells.

3.2 Introduction

The construction of genetic programs has emerged as a powerful approach for investigating cell signaling¹³⁰ and for engineering cell-based devices^{1,131}. These systems often make use of transcriptional regulation, and for research on eukaryotic cells, TetR^{132,133} and Gal4¹³⁴ are established TFs. However, there remains a need to be able to program the activation and inhibition of gene expression with greater flexibility and precision, and to be able to use greater numbers of TFs that are orthogonal to each other and still interoperable with other genetic technologies such as upstream sensors. Recently, various families of TFs have been investigated for programmable sequence-specific binding capabilities¹³⁵⁻¹³⁹. Among these are zinc finger (ZF)-TFs^{140,141}, which also have the benefit of being relatively small, thus affording more space under constraints such as cargo limits of certain gene delivery methods. The COMET toolkit was developed by colleagues with these properties in mind, by drawing from designed ZF-TFs¹⁴² and modifying or

appending other domains onto these proteins to confer new functions. I developed mathematical models to provide mechanistic insights into the performance characteristics of these genetic parts and to help enable their use in future applications.

3.3 Materials and Methods

Methods are described in **Appendix 5**.

3.4 Results

Elucidating mechanisms of COMET gene expression: My colleagues obtained ZFs that were previously characterized in yeast by Khalil et al.¹⁴³ and investigated the use of these domains in inducing transcription in mammalian cells (**Figure 3.1**). Each TF has a ZF DNA-binding domain (DBD) fused to an activation domain (AD) such as VP16, which forms a ZF activator (ZFa) that recruits RNA polymerase II (RNAPII) to initiate transcription¹⁴⁴. A cognate promoter was produced for each ZFa by placing a ZF binding site upstream of a minimal promoter that has been shown to confer low background and inducible expression^{145,146}.

To help elucidate the mechanisms by which this system operates, I developed a mathematical model. As summarized in **Figure 3.2**, this process involved considering steps involved in gene expression, writing equations to encapsulate these steps (writing equations is tantamount to formulating a hypothesis

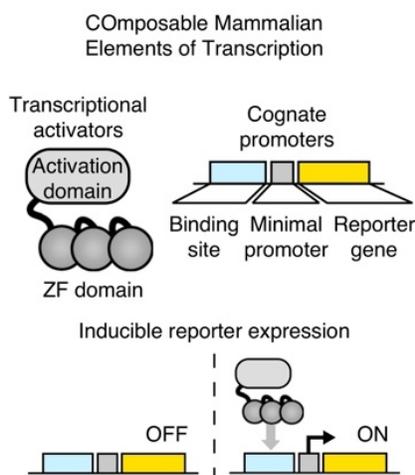


Figure 3.1. COMET. The schematic shows the features of the TFs and promoters.

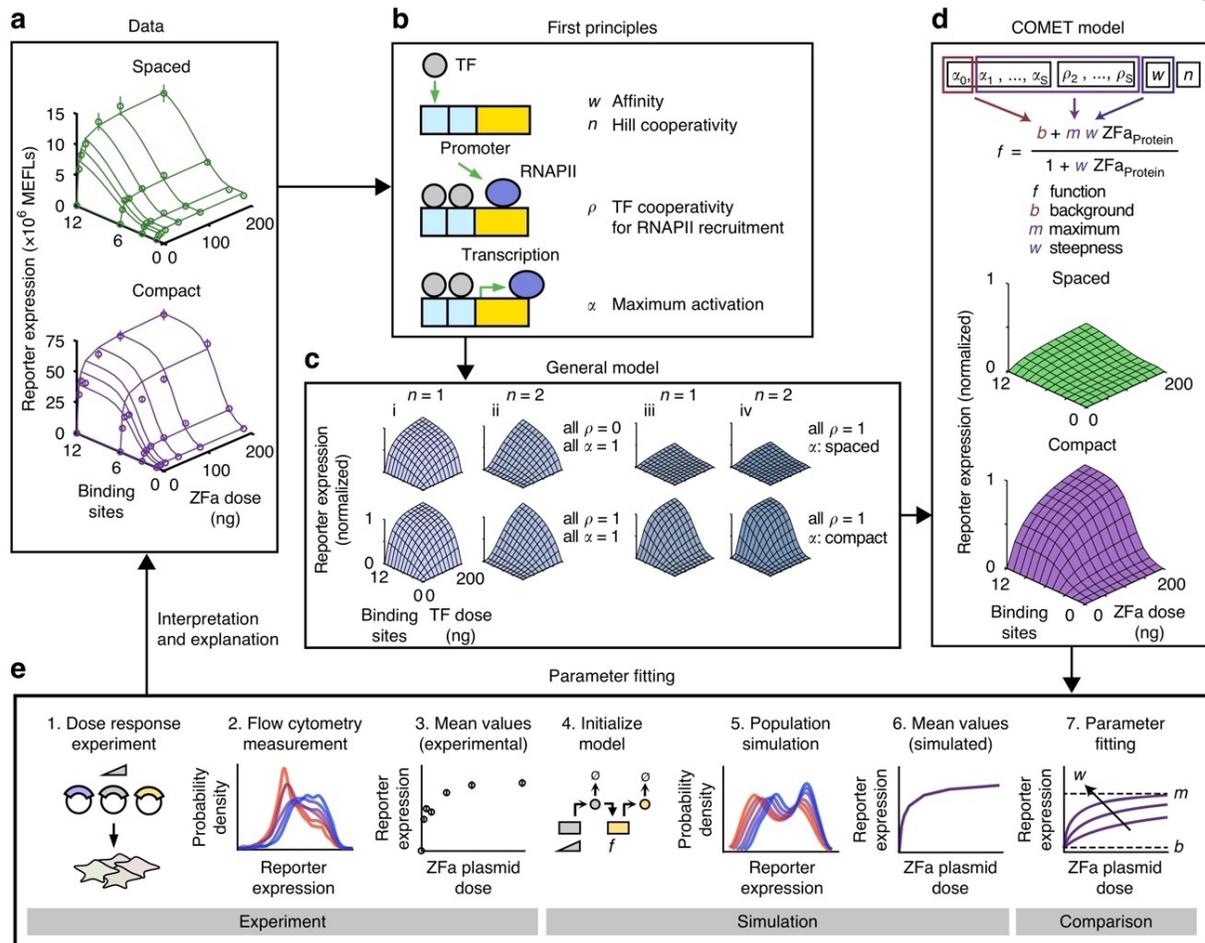


Figure 3.2. A model for COMET-mediated gene regulation. This figure summarizes the process of model development, refinement, and fitting. **a** The COMET model (model outputs are represented by the lines on each plot) explains experimentally observed trends (circles) for reporter expression as a function of ZFa dose and promoter features. This model uses a fitted response function for ZFa-induced gene expression (discussed in **b–e**) and simulates a cell population to account for variation in gene expression; lines depict the average outcome for the population. The experiment was conducted in biologic triplicate. Error bars represent the S.E.M. **b** We started with a detailed model of transcriptional activation in which reporter expression depends on TF concentration, a metric related to TF-DNA-binding affinity (w), TF-DNA-binding cooperativity ($n = 1$ for non-cooperative, $n > 1$ for cooperative), RNAPII recruitment cooperativity by each multiple-TF configuration at a promoter ($\rho = 0$ for non-cooperative, $\rho > 0$ for cooperative), and maximum promoter activation by each configuration ($0 \leq \alpha \leq 1$). **c** This model yielded four types of landscapes (i–iv) under different assumptions, and two representative examples of each type are shown. COMET most closely resembles (iii). **d, e** A model that represents ZFa-induced reporter expression by a response function was used to fit the data in **a** (the workflow for parameter estimation is depicted in **e**). The terms in this concise model can be related to terms in the mechanistic model. Landscapes in **c, d** are simulations of a single cell (homogenous model), and those in **a** are simulated mean values for a heterogeneous population. The outputs of this final fitted model are represented alongside experimental data in **a**.

as to how gene expression operates), identifying a formulation consistent with experimental observations, and simplifying this representation by removing details not required to describe observed trends in order to generate a concise model. Then, the last step was fitting parameters for the concise model to data to quantitatively describe experimental observations. I hypothesized that this process should generate a set of experimentally grounded parameters representing interpretable features of TF-promoter activity. Throughout, the goal was not to predict TF or promoter sequences *de novo*, but rather to describe and provide insight into observed trends. The explanatory value of such a model often exceeds insights that are accessible by intuition alone, and ultimately this framework could be used to design new genetic functions based upon COMET parts.

I initiated this process by using first principles to produce a detailed model with features of transcriptional control¹⁴⁷ including physical and functional interactions between the promoter, TFs, and proteins like RNAPII (**Figure 3.2b, Appendix 5**). This detailed model relates transcriptional output to TF concentration, TF-DNA binding affinity, TF-DNA binding cooperativity, RNAPII recruitment cooperativity, and maximum promoter activation. I then generated a series of theoretical landscapes analogous to the experimental landscapes in (**Figure 3.2a**), varying parameters across a biologically reasonable range, and observed that the landscapes fell within one of four categories defined with respect to the concavity and sigmoidicity of cross-sections along each axis (**Figure 3.2c**). The experimental data most closely resembled case (iii), indicating that TF-DNA binding is non-cooperative, but RNAPII recruitment is cooperative, and the maximum transcription rate (at a high ZFa dose) increases with both the number and compactness of binding sites. Therefore, the enhanced potency of the compact promoters stems from the cooperative recruitment of transcriptional machinery.

Based upon the observed ZFa dose response profiles (**Figure 3.2a**) and these insights, I proposed a concise response function to represent the rate of transcription (f) as a function of ZFa dose with three parameters: background (TF-independent) transcription (b), a steepness metric (w) related to TF-DNA-binding affinity, and a metric for maximum transcription (m) (**Figure 3.2d**). As indicated, the three parameters in this concise response function can be related to the additional parameters in the original detailed representation. For a given ZFa-promoter combination, m is experimentally determined and is based upon the number and spacing of binding sites in the promoter, and b is determined based on reporter

expression without ZFa; w can be fit to ZFa dose response data by our previously developed method that improves parameter estimation by accounting for variation in gene expression⁵⁴ (**Figure 3.2e, Appendix 3**; fitted parameters are in **Appendix 5**). Simulated data from the calibrated model provided close agreement with the experimental data, demonstrating that a concise representation can be used to analyze and describe COMET-mediated gene expression.

Comparison of the calibrated model and experimental data confirmed two trends that hold across conditions. First, the dependence of relative reporter output on binding site number is independent of the dose of ZFa plasmid when the output is scaled to its maximum value in each binding site series. Second, the dependence of relative reporter output on ZFa dose is independent of the number of binding sites when the output is scaled to its maximum value in each dose series. Thus, inducible gene expression follows a pattern that holds across various promoter designs and that is captured by a concise model. The occurrence of these similar patterns, when paired with the properties elucidated by the model, makes ZFa-induced gene expression readily interpretable—a desirable feature for a transcriptional toolkit.

Tuning transcription through protein engineering: We next investigated strategies for tuning gene expression through two approaches for TF engineering: altering the affinity of the ZF for the DNA and altering the strength of the AD. For the first strategy, four arginine residues in the ZF that interact with the DNA backbone were mutated (**Figure 3.3a**); mutations to alanine ablate favorable charge interactions and have been shown to decrease inducible target gene expression in yeast^{143,148,149}. As intended, ZFa-mediated gene expression decreased with an increasing number of substitutions (**Figure 3.3**). Interestingly, while changing the promoter architecture affected only the maximum transcription (m) (**Figure 3.2**), ZF mutations affected both the maximum transcription and relative steepness of the dose response curve (m and w). Additionally, changes in these values were correlated, revealing an axis along which ZFa R-to-A mutations tune TF strength.

For another tuning strategy, two ADs were tested in place of VP16 (VP64¹⁵⁰ and VPR¹⁵¹; **Figure 3.3c**) to investigate whether these stronger ADs could increase gene expression conferred by TFs with low-affinity ZFs. Comparison of the dose response for the weakest-binding ZFa mutant (AAAA) with each AD to the VP16 ZFa bearing a wild-type (WT) ZF domain (**Figure 3.3d**) shows that as AD strength increased, both m and w increased. Although the two domains of a ZFa are physically modular, since they affect the

same parameters in the response function, the domains are functionally intertwined. In summary, the observations show how gene expression can be tuned through selection of physical features—ZF domain choice, mutations that affect DNA binding affinity, AD choice, and the number, spacing, and arrangement of binding sites in the promoter—and together this ensemble of designs provides a variety of realizable response profiles (**Figure 3.3e**).

Design of inhibitory TFs: Inhibitors comprise a key component of a versatile TF toolkit. My colleagues hypothesized that removing the AD from the ZFa would result in an inhibitor that binds DNA without inducing transcription (ZF inhibitor, ZFi), and that inhibition could be made more potent by incorporating a bulky domain to sterically block ZFa from binding adjacent sites in the promoter or to block

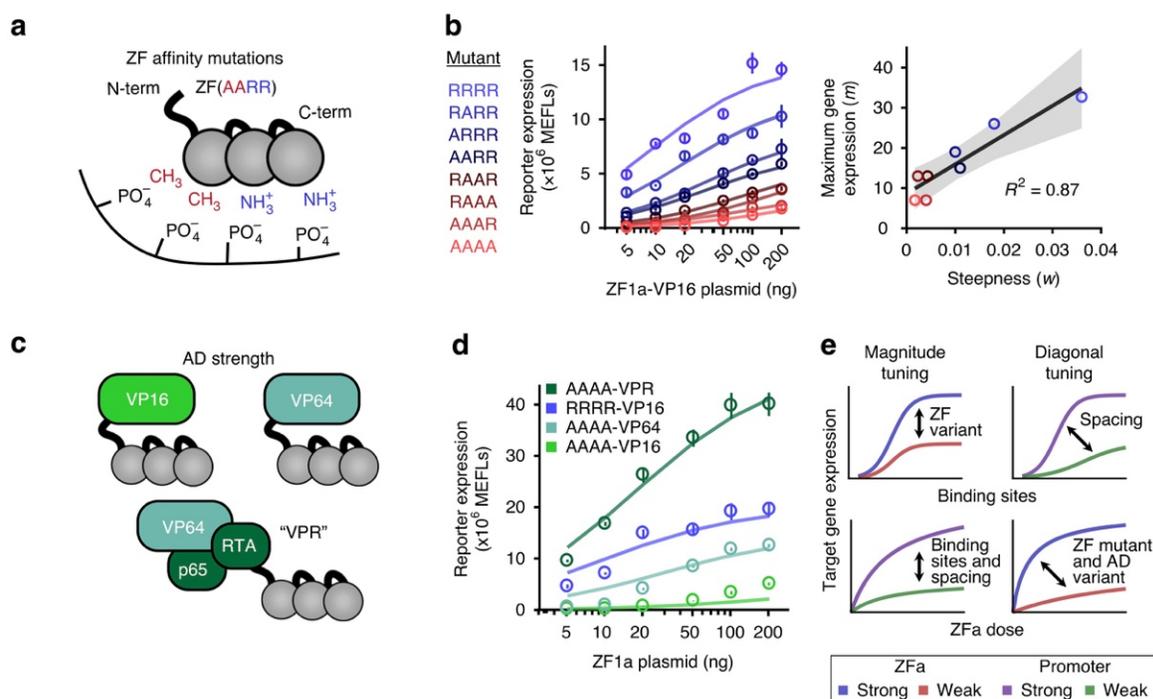


Figure 3.3. Tuning transcription through ZF mutants and AD variants. **a** The cartoon illustrates arginine-to-alanine (R-to-A) mutations in the ZF domain, which decrease the DNA-binding affinity. **b** Left: ZF mutations modulate the steepness and the maximum of the ZFa dose response profile. Circles represent experimental data and solid lines represent fitted response function models. Right: correlation between m and w parameters across mutants. The regression line is $m = 7.3 \times 10^2 w + 8.6$, and the shaded region is the 95% confidence interval (one-tailed permutation test $p < 0.001$). **c** The cartoon depicts evaluated ADs. **d** The choice of AD affects the steepness and the maximum of the dose response. Circles represent experimental data and solid lines represent fitted response function models. **e** The cartoon summarizes expected trends in output gene expression that result from tuning each modular feature of the ZFa and promoters. These design choices can produce either a vertical shift or diagonal shift in response profiles with respect to the number of binding sites and the dose of ZFa. Experiments were conducted in biologic triplicate. Error bars depict S.E.M.

the recruitment of RNAPII or associated factors (**Figure 3.4a**). To test this hypothesis, DsRed-Express2 (abbreviated throughout as DsRed) was fused to the ZF domain. To help understand the mechanism of transcriptional inhibition, I considered that within each cell, promoters occupy an ensemble of states that depend on the promoter architecture and the ZFa and ZFi that are present (**Figure 3.4b**). As the relative dose of ZFi to ZFa increases, the distribution of the ensemble should shift toward states that are more inhibited; a trend towards more inhibition should also occur by increasing the relative DNA binding affinity of the ZFi versus that of the ZFa. Given our understanding of ZFa-mediated transcriptional activation, I speculated that the inhibitors should act via a dual mechanism with these properties: (i) competitive inhibition: since each site in the promoter can accommodate at most one TF, the binding of an inhibitor should prevent the binding of an activator; and (ii) decreased cooperativity: since inhibitors intersperse between activators, the spacing between activators should widen, and the effective m should resemble that of a promoter with lower cooperativity.

To experimentally test the proposed dual mechanism of inhibition, dose responses were conducted for the ZFi and ZFi-DsRed inhibitors using the ZF1x6-S and ZF1x6-C promoters, with ZFa dose held constant (**Figure 3.4c**). When ZFi was applied to the compact promoter, reporter expression matched the concise model for competitive inhibition alone. However, for the other three cases, observed reporter expression began to deviate with increasing doses of inhibitor, and by high doses it showed complete loss of cooperative RNAPII recruitment. The inhibitor dose at which the experiment began to deviate from the model was lower for ZFi-DsRed compared to ZFi and for spaced promoters compared to compact promoters. At intermediate doses of inhibitor, reporter expression ramped down toward single-site promoter behavior (**Figure 3.4b** middle column, **Figure 3.4c** dotted lines), and by high doses the ramp down was complete (**Figure 3.4b** right column). The highest dose of ZFi-DsRed, used with the compact promoter, resulted in a profound 400-fold decrease in reporter expression. To further examine the case where the employed inhibitor did not disrupt cooperative RNAPII recruitment (i.e., ZFi used with the x6-C promoter), we paired a panel of ZFi varying in DNA-binding affinity with a reduced-affinity ZFa mutant (**Figure 3.4d**). For all cases examined, ZFi-mediated inhibition was still predicted by competitive inhibition alone. Thus, the compact promoter is more capable of cooperative RNAPII recruitment than is the spaced promoter, and ZFi is a weaker inhibitor than is ZFi-DsRed, such that the dual inhibition mechanism applies to three of the

four types of inhibitor-promoter pairings evaluated, and the pairing most responsive to inhibition is ZFi-DsRed with a compact promoter. The mechanism by which cooperative transcriptional machinery recruitment renders the compact promoter architecture highly activatable by a ZFa also causes such promoters to be substantially inhibited through disruption of this mechanism by a ZFi-DsRed.

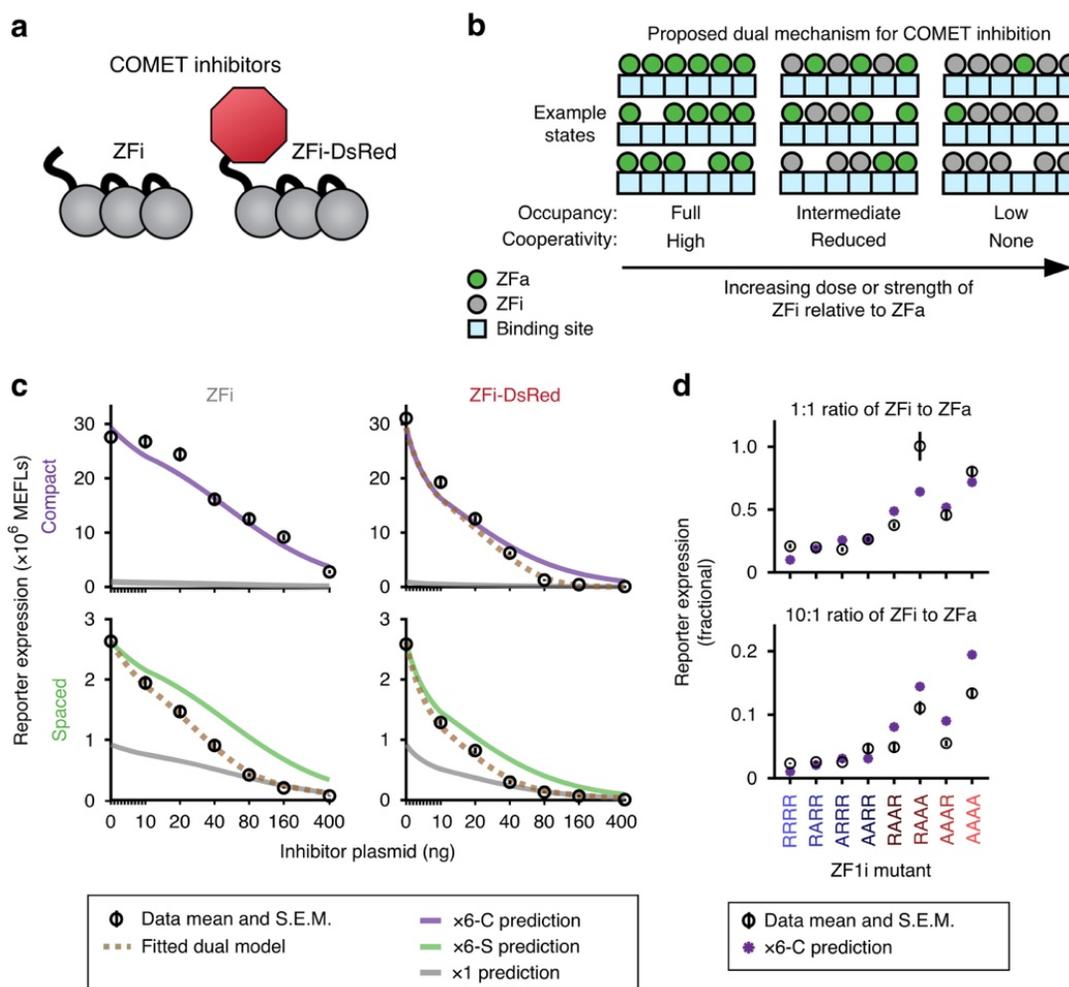


Figure 3.4. Transcriptional inhibition. **a** The schematic depicts two types of inhibitors that were evaluated. **b** The cartoon summarizes the proposed conceptual model of ZFi-mediated inhibition. Within each cell, a promoter can occupy states with different configurations of ZFa and ZFi. Several example states are shown for three conditions of increasing dose or strength of inhibitor (i.e., DNA-binding affinity) relative to activator. **c** ZFi and ZFi-DsRed differ from standard competitive inhibitors. Predictions for competitive inhibition alone, for various promoter configurations, are depicted with solid lines. COMET inhibitors track the dotted lines, which represent fits to the dual mechanism model, except in the case of ZFi paired with x6-C, which tracks the competitive inhibition-only prediction. Each condition uses ZF1a at a dose of 40 ng. X-axes are scaled linearly from 0 to 10 ng and logarithmically above 10 ng. **d** Measured and predicted reporter expression were compared for a panel of ZFi mutants. Each condition uses ZF1a (RAAR) at a dose of 40 ng and the ZF1x6-C compact promoter. Experiments were conducted in biologic triplicate. Error bars represent the S.E.M.

Implementing Boolean logic: Finally, we explored whether COMET could be used to encode Boolean logic functions within individual promoters. Colleagues pursued a strategy of hybrid promoters with alternating sites for combinations of ZFa to implement AND logic (**Figure 3.5a**), with the hypothesis that cooperative activation on compact promoters would occur only when both species of ZFa are present. Synergistic activation from closely arranged binding sites has been used to make AND gates in mammalian cells¹⁵², but arranging sites in alternating patterns does not necessarily guarantee this outcome⁵⁴. For a three-pair hybrid promoter, AND gate behavior was observed even at low ZFa levels; 5 ng of each plasmid encoding ZF2a and ZF3a together produced more reporter expression than did 200 ng of plasmid encoding either ZFa alone (**Figure 3.5b**). The steep OFF-ON transition along the perimeter of the landscape is due to the effective transition between x3-S and x6-C architectures—an advantageous behavior of COMET that differs from previously reported AND gates utilizing tTA and Gal4 (**Figure 3.5c, Appendix 5**)⁵⁴.

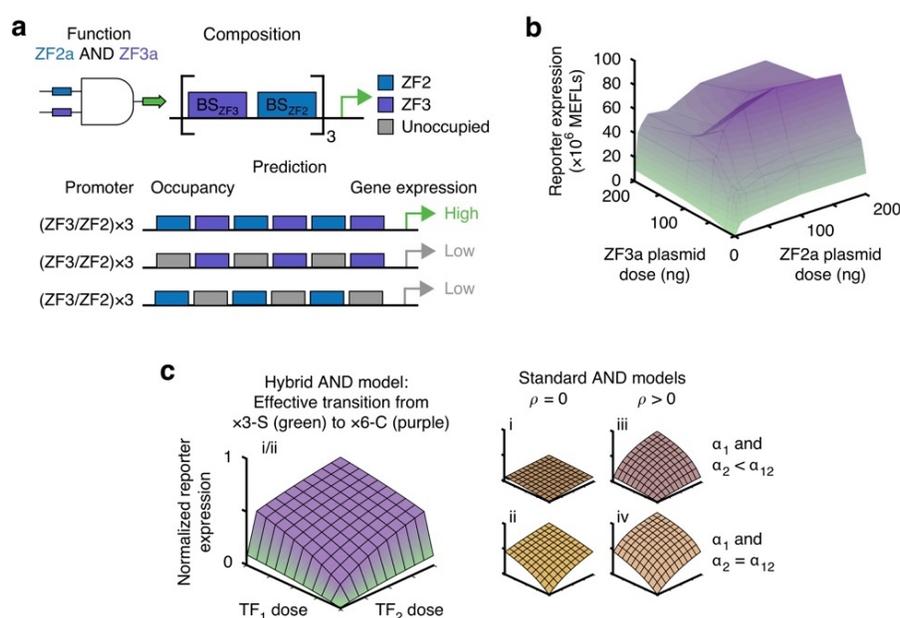


Figure 3.5. Composing Boolean logic. **a** The cartoon summarizes a strategy for single-layer, promoter-based logic gates with ZF-TFs. We hypothesized that AND gate promoters could be designed by using multiple repeats of a paired ZF3/ZF2 motif. Full occupancy of this promoter by both ZF2a and ZF3a mimics a fully occupied x6-C promoter, and partial occupancy (with either ZFa alone) mimics an x3-S promoter. Thus, there is a large increase in gene expression when the promoter is occupied by two types of ZFa compared to one type. **b** Two-input dose response for the AND gate with three repeats of paired binding sites. The landscape is shaded from green to purple to facilitate visualization in the z-axis direction. **c** A theoretical model of COMET AND behavior is compared with other models of transcriptional AND gates; the latter vary in whether activators have multiplicative cooperativity (ρ) and whether maximum activation (α) is equivalent for TFs individually and together.

3.5 Discussion

The engineering of cellular functions often involves multiple iterations of the design-build-test-learn cycle. In prokaryotes, this process has been streamlined by the development of large libraries of well characterized genetic parts and computational tools such as Cello¹⁵³. COMET similarly provides a set of TFs and promoters, and the characterization of these components and their tunable features provided a foundation for a mathematical model. I used the model to elucidate mechanisms by which activators and inhibitors operate at promoters and fitted parameters to describe how these activities vary across the design choices examined. This integrated approach helped us move from the identification of general qualitative trends (e.g., increasing the number of binding sites in a promoter generally increases inducible gene expression) to arrive at quantitative and mechanistic understanding as to how design choices affect TF-promoter activity. This insight could not have been deduced from prior knowledge, including biophysical intuition or characterization of similar TFs and promoters in yeast¹⁴³. Whether the design rules elucidated here enable model-driven predictive design is an important question that is investigated in **Chapter 4**.

In nature, transcription is regulated by the combined activities that take place across promoters, enhancers, insulators, and other non-coding elements. Despite the precision with which TF activities must be coordinated, there is a growing appreciation that TF binding and gene expression are often leaky¹⁵⁴. Findings that binding site sequences are often degenerate and that individual TF binding events are often low-affinity suggest that multiple TFs and binding sites might be not only useful, but necessary, to confer specificity to gene expression. Indeed, enhancers generally contain groups of binding sites for various TFs. In some cases with multiple binding sites, cooperativity can arise, either from TF-DNA binding, TF-TF interactions at nearby sites, or from ways in which multiple TFs act to recruit co-activators or co-repressors. It is thought that cooperativity could confer certain advantages, e.g., buffering against variation in TF levels: as long as a TF is present in a sufficient amount, this would be sufficient to induce gene expression¹⁵⁴.

Several mechanisms can produce cooperativity in TF binding. TFs can bind to each other and increase their affinity for a DNA sequence; TFs with low affinity for each other can bind to DNA and form more stable TF-TF associations; and the binding of one TF to DNA can alter the shape of the DNA in a way that increases the energetic favorability of a second TF binding to the DNA¹⁵⁵. In a study on the TF Gal4, a reporter readout was measured from integrated constructs containing one or two binding sites compared

to the wild type yeast sequence with four sites. Synergistic reporter induction was evident with as few as two sites, and footprinting analysis showed that TF binding to DNA was cooperative and trended with the reporter readout. Although synergy was characteristic of the low-affinity binding sites, the outcomes were only additive for the high-affinity sites. Thus, cooperativity stemmed from TF-DNA binding if the interaction was weak, and transcriptional activity was related to the number of TF molecules bound¹⁵⁶. Using **Figure 3.2c** as a guide, the case of high-affinity sites resembles case (i) and the case of low-affinity sites resembles case (ii).

TFs based on ZF domains coordinate many functions in nature¹⁵⁷, and different characteristics with respect to cooperativity have been identified. One study reported cooperative binding to DNA for the progesterone receptor¹⁵⁸. Another study reported that DNA binding could be either positively or negatively cooperative, or neither, depending on the specific DNA element, for the estrogen receptor¹⁵⁹. In a recent study that implemented synthetic ZF-TFs in yeast, cooperativity was generated by using promoters containing multiple binding sites and a system to load multiple TFs onto a co-expressed protein scaffold; Hill coefficients fitted to dose response data were higher when the binding sites and scaffold repeats were extended¹⁶⁰. These studies show that there are many ways in which transcriptional activation properties are tunable, including with ZF-TFs, and that tunability occurs in nature and also can be engineered.

In the COMET study, a key insight is that promoter strength was explained by the cooperative recruitment of transcriptional machinery, rather than TF-DNA binding, and in a way that was tunable by altering the number of and spacing between binding sites. This promoter-based cooperativity is useful in that it confers both low background expression and high fold induction, enables single-layer logic gates, and can be potently modulated by DsRed-ZFi to inhibit ZFa-mediated expression. For these reasons and more, we expect that this toolkit will provide useful capabilities for designing mammalian cellular functions.

3.6 Acknowledgements

I thank Patrick Donahue for the invitation to be a part of this study, and for camaraderie in projects that stemmed from this collaboration; Hailey Edelstein and Joe Draut for collaboration; Juliana Feng and Vis Kandula for assistance with cloning; and Amy Hong, Cameron McDonald, Justin Finkle, and Sebastian Bernasek for useful discussions on the computational modeling.

CHAPTER 4. Design-driven engineering of mammalian genetic programs

This chapter is in preparation as:

Muldoon J.J., Kandula V., Hong M., Donahue P.S., Boucher J.D., Bagheri N., Leonard J.N. Design-driven engineering of mammalian genetic programs. *In preparation*.¹⁶¹

4.1 Abstract

Genetically engineering cells to perform customizable functions is an emerging field with a range of technological and translational applications, however it generally remains challenging to engineer these functions systematically in mammalian cells. To address this need, here we develop a method enabling accurate genetic program design using high-performing genetic parts and predictive computational models. We build multi-functional proteins integrating transcriptional and post-translational control, validate models for these mechanisms, implement digital and analog processing, develop a strategy for topological circuit compaction, and effectively link genetic circuits with sensors to carry out multi-input evaluations. The functional modularity and compositional versatility of these parts enable one to readily encode a given design objective in numerous ways. This approach substantially elevates our capacity to predictively design sophisticated mammalian cellular functions.

4.2 Introduction

Early demonstrations of genetically engineered functions in mammalian cells show great promise for applications in cell differentiation, regenerative medicine, and cell-based therapies. These applications generally require precise control of gene expression and the capability to sense and respond to external cues¹⁶²⁻¹⁶⁶. Despite the growing availability of parts for doing so, assembling parts to compose customized genetic programs remains laborious and challenging and often requires iterative experimental tuning or down-selection to identify functional configurations. This process limits the scope of programs that one can feasibly compose and debug, and likely the performance of even functional programs. Thus, there is a need for a more systematic and precise design process.

Model-guided design has been demonstrated in the composition of some functions, including transcriptional logic in bacteria¹⁵³ as well as logical and analog behaviors in yeast¹⁶⁰, but this type of approach remains relatively new in mammalian systems. Associated challenges include the availability of appropriate parts, suitably descriptive computational models that enable predictions using these parts, and computational and conceptual tools that enable the identification of designs that function robustly despite inherent biological variability. In this study, we sought to address these challenges by developing a model-

driven process that enables one to propose a tractable set of candidate circuits for construction and testing and applied this framework to implement a variety of logical, analog, and sense-and-respond behaviors.

4.3 Materials and Methods

Plasmid cloning and purification: Plasmids were designed in SnapGene (GSL Biotech LLC), and primers were ordered from Integrated DNA Technologies. Several domains were sourced from Donahue, et al.¹²⁹, but prior to the COMET study: VP16 and ZF domains are from Khalil, et al.¹⁴³, VP64 is from Chavez, et al. (Addgene #63798)¹⁵¹, FRB and FKBP are from Daringer, et al. (Addgene #58876, #58877)¹⁶⁷, and DsRed-Express2 is a gift from David Schaffer.

Split inteins are from Hermann, et al. (Addgene #51267, #51268)¹⁶⁸. The ABA-binding domains PYL1 and ABI1¹⁶⁹⁻¹⁷³ from Gao, et al.¹⁷⁴ were utilized to make ABA-ZFa. The PEST tag is from the mouse ornithine decarboxylase gene¹⁷⁵. Two types of plasmid backbones are used: pcDNA (pPD005, Addgene #138749), which was modified from Thermo Fisher Scientific #V87020 as described by Donahue, et al.¹²⁹; and a series of transcription unit positioning vectors (TUPVs), which are derived from the modified pcDNA and previously published by Donahue, et al.¹²⁹, and based upon the mMoClo system from Duportet, et al.⁴⁵. Insulator sequences in TUPVs are from Bintu, et al. (Addgene #78099)¹⁷⁶.

Cloning was performed primarily using standard PCR, restriction, and ligation methods (reagents from New England Biolabs and Thermo Fisher Scientific), and in some cases through Golden Gate assembly, followed by transformation into chemically competent TOP10 *E. coli* (Thermo Fisher Scientific). Transformed *E. coli* were grown on LB/Ampicillin agar plates at 37°C, colonies were picked and grown in liquid LB/Ampicillin cultures, plasmid DNA was isolated (E.Z.N.A. plasmid mini kit, Omega Bio-tek), and DNA inserts were sequence-verified (ACGT, Inc.). Plasmids were prepared using polyethylene glycol-based extraction as described previously¹²⁹. DNA purity and concentration were measured using a Nanodrop 2000 (Thermo Fisher Scientific).

Mammalian cell culture: HEK293FT cells were cultured in complete DMEM medium containing 1% DMEM powder (Gibco #31600091), 0.35% w/v D-glucose (Sigma #50-99-7), 0.37% w/v sodium bicarbonate (Fisher #S233-500), 10% heat-inactivated FBS (Gibco #16140071), 4 mM L-glutamine (Gibco #25030081), and 100 U ml⁻¹ penicillin and 100 µg ml⁻¹ streptomycin (Gibco #15140122) in tissue culture-

treated 10 cm dishes (Corning # 500001672) at 37°C in 5% CO₂. To passage, medium was aspirated, and cells were washed in PBS, incubated in trypsin-EDTA (Gibco #25300054; 37°C, 5 min), detached by tapping the dish, and resuspended in fresh medium and plated. This cell line tested negative for *Mycoplasma* using the MycoAlert Mycoplasma detection kit (Lonza #LT07-318).

Transfection: Cells were plated in 24-well plates (Corning #3524; 3×10^5 cells ml⁻¹, 0.5 ml per well) and transfected after adhering to the plates, typically between 8–14 h after plating. Transfections were carried out using the calcium phosphate protocol¹²⁹: plasmids are mixed together in defined amounts, CaCl₂ (2 M, 15% v/v) is added, and this solution is pipetted dropwise into an equal volume of 2x HEPES-buffered saline (500 mM HEPES, 280 mM NaCl, 1.5 mM Na₂HPO₄); the solution is gently pipetted four times, and three minutes later it is vigorously pipetted 20 times and added dropwise onto plated cells. In this study, DNA doses are reported in plasmid mass (ng) per well of cells or gene copies per well of cells. In each transfection experiment, “empty vector” (pPD005) was included in the transfection mix to maintain a consistent total mass of DNA per well. At one day after plating, medium was aspirated and replaced with fresh medium. In some experiments, the fresh medium contained vehicle or ligand. In **Figure 4.3**, the vehicle was 0.1% DMSO (v/v in cell culture) and the ligand was 100 nM rapamycin in 0.1% DMSO. In **Figure 4.4a,b**, the vehicle was 0.1% EtOH and the ligand was either 100 nM rapalog (Takara #AP21697) or 100 μM abscisic acid (ABA; Goldbio #21293-29-8) in 0.1% EtOH. In **Fig. 4d,e**, the vehicle was 0.2% EtOH, and the ligand conditions included 100 nM rapalog, 100 μM ABA, or both ligands in 0.2% EtOH.

Flow cytometry: Samples were prepared for flow cytometry generally at 40–48 h post-transfection. For each well, medium was aspirated, five drops of PBS were added, PBS was aspirated, and two drops of trypsin-EDTA were added. Cells were incubated (37°C, 5 min), plates were tapped to detach cells, and four drops of cold (4°C) DMEM were added. The contents of each well were pipetted up and down several times to detach cells and pipetted into FACS tubes containing FACS buffer (FB; 2 ml; PBS pH 7.4, 5 mM EDTA, 0.1% w/v BSA). Tubes were centrifuged (150×g, 5 min), liquid was decanted, and two drops of FB were added. Samples were kept on ice and wrapped in foil, and then run on a BD LSR Fortessa special order research product using the following configuration: Pacific Blue channel with 405 nm excitation laser and 450/50 nm filter for EBFP2; FITC channel with 488 nm excitation laser and 505LP 530/30 nm filter for

EYFP; and PE-Texas Red channel with 552 nm excitation laser and 600LP 610/20 nm filter for mKate2. Approximately 10^4 live single-cell events were collected per sample.

Flow cytometry data analysis: Flow cytometry data were analyzed using FlowJo software (FlowJo, LLC) to gate on single-cell (FSC-A vs. FSC-H) and live (FSC-A vs. SSC-A) bases, compensated using compensation control samples, and gated as transfection-positive. The mean reporter signal in MFI was obtained for each sample. UltraRainbow Calibration Particles (Spherotech #URCP-100-2H) were run in each flow cytometry experiment. Beads were gated on an FSC-A vs. FSC-H basis, the nine bead subpopulations of varying intensities were identified, and the mean MFI for each subpopulation in the FITC channel and PE-Texas Red channel was obtained. These values in combination with manufacturer-supplied MEFL and MEPTR values for each subpopulation were used to fit a regression line with y-intercept equal to zero. The mean and S.E.M. for the three biological replicates were calculated. Autofluorescence background signal was subtracted using samples transfected with the transfection control marker, and error was propagated. MFI values were converted to MEFL or MEPTR using the slope of the regression line, and error was propagated. Histograms in supplementary figure panels represent reporter signal in MFI.

Nomenclature: Genes are named by their protein domains in order from N-terminus to C-terminus. Domains are generally connected by flexible linkers comprising glycine and serine. Several abbreviations are used: ZFa is an AD-ZF for any choice of AD and ZF; similarly, RaZFa is an AD-FRB and FKBP-ZF, and ABA-ZFa is an AD-PYL1 and ABI1-ZF. DsRed refers to wild type DsRed-Express2, and DsDed is an DsRed-Express2 R95K mutant. We use a streamlined nomenclature that differs from that used in the original COMET report¹²⁹, in that inhibitors do not use ZFi notation: ZFi is now termed ZF, and DsRed-ZFi is now termed DsRed-ZF.

The constitutive promoters used are CMV and EF1 α . The inducible promoters used are COMET promoters, which are named as “[ZF domain]x[number of binding sites]-[binding site arrangement]”. For example, ZF1x6-C has six compact sites for ZF1. There are two non-standard cases: ZF1/2x6-C has six compact overlapping sites for ZF1 and ZF2 (up to six sites occupied, and up to six per ZF); (ZF2/ZF6)x3 has six compact sites alternating between ZF2 and ZF6 (up to six sites occupied, and up to three per ZF).

Statistical analysis: Each sensor in **Fig. 4.4a,b** was assessed using a one-tailed Welch’s unpaired *t*-test, with the null hypothesis that reporter signal was equal with and without ligand treatment. Genetic

programs in **Fig. 4d,e** were assessed using a three-factor ANOVA and Tukey's honest significant difference (HSD) test, with the null hypothesis that reporter signal was equal across the two input types, four topologies, and four input combinations. Effects were considered significant if $p < 0.05$, and additionally for the HSD test if the comparisons had an adjusted $p < 0.05$.

Computational method details—Described further in **Appendix 6**. The modeling approach accounts for variation in gene expression using the constrained sampling method^{54,129} (which is used here to describe the distribution of gene expression observed when cells are harvested via trypsin digest). Some parameter values are from the COMET study¹²⁹ and others are newly estimated or fitted (**Table A6.1**).

Diagrams: Genetic programs for digital operations are depicted using genetic diagrams and electronic diagrams. The former represents each promoter, protein, and regulatory interaction, and the latter represents the logic underlying these interactions.

4.4 Results

Biological parts for integrating transcriptional and post-translational control of gene expression: The strategy that we pursued for genetic program design was uniquely enabled by the development of the COMposable Mammalian Elements of Transcription (COMET): a toolkit of transcription factors (TFs) and promoters with tunable properties enabling precise and orthogonal control of gene expression¹²⁹. These TFs comprise a zinc finger (ZF) DNA-binding domain and a functional domain, e.g., VP16 and VP64 are activation domains (AD) that with a ZF form an activator (ZFa). The promoters contain binding sites (e.g., ZF1x6-C has six compactly arranged ZF1 sites), and transcriptional activity can be evaluated by measuring the signal from an inducible reporter gene. We also developed mathematical models to characterize how protein domain and promoter choices affect TF-promoter activity. Here, we investigated whether these biological parts and descriptive computational tools could be adapted and applied to achieve predictive genetic program design.

Although COMET includes many parts for transcriptional regulation, we hypothesized that genetic program design would be further facilitated by introducing a mechanism for regulation at the post-translational level (**Figure 4.1a,b**). To investigate this possibility, we evaluated new parts based on split inteins: complementary domains that fold and *trans*-splice to covalently ligate flanking domains (exteins)¹⁷⁷.

Figure 4.1 Logical evaluation is enabled by transcriptional and post-translational regulation. (a,b) Cartoons depict (a) the genetic components and (b) their arrangement and use in simulations to produce intended functions. Transcription is mediated by COMET TFs, which here are modified with split inteins to incorporate post-translational regulation via splicing. Genetic parts that carry out specified activities and that can be described mathematically should enable the predictive customization of mammalian cellular functions. In the schematics, circles are protein domains, arrows are splicing or regulation, yellow highlighting indicates the inputs, and the red node is the output. (c–j) A panel of logic gates was designed, simulated, and experimentally evaluated. Synthetic digital logic in cells is inherently analog, and component doses were selected to examine this behavior and underscore particular features; e.g., in c, reporter signal decreases at a high intC-ZF1 dose because unspliced intC-ZF1 inhibits ZFa-mediated transcription. In the electronic diagrams that represent the logical evaluations (teal background), lines denote splicing or regulation. Processes that have a modest effect within the dose range examined, and that because of fundamentally analog behavior do not carry out a fully digital function, are denoted by dotted lines. In the mechanistic diagrams (blue background), the purple bent arrows are promoters, and the black arrows are splicing and regulation. Yellow highlighting indicates the components for which dose is varied (in gene copies). Simulation and experimental results are presented in heatmaps that indicate how the two inputs affect reporter output (mKate2 signal in MEPTRs); color-coding denotes the mean reporter signal from three biological replicates, scaled by maximum value in each heatmap. Simulations in c are a fit to the data, and subsequent panels (d–j) are predictions. (k) Some of the motifs that were used in the gate designs confer sharp ON/OFF transitions in reporter output. For example, a standard activation dose response was not ultrasensitive, but layering two inhibitors in a cascade did produce ultrasensitivity (Hill coefficient $n > 1$).

We selected the split intein gp41-1 for its rapid splicing kinetics¹⁷⁸. To test an application of this mechanism, we appended an AD to the gp41-1 N-terminal fragment (intN) and a ZF to the C-terminal fragment (intC). These parts were used to construct an AND gate in which a reporter gene was induced only when both fragments were present (**Figure 4.1c**), demonstrating that COMET-mediated gene expression can be controlled using split inteins. We next incorporated this reconstitution mechanism into our modeling framework by modifying ordinary differential equations from the original study, which concisely represent transcriptional regulation, and fitting newly introduced parameters to the data. We also extended the model to incorporate split inteins onto two types of inhibitors: ZF, which competes with ZFa for binding site occupancy in the promoter; and ZF fused to DsRed-Express2 (abbreviated as DsRed-ZF), which through a dual mechanism of inhibition also reduces the cooperativity of ZFa-mediated RNAPII recruitment at multi-site promoters¹²⁹. Here, we introduced an R95K mutation to ablate the DsRed chromophore¹⁷⁹, yielding a non-fluorescent inhibitor we termed DsDed-ZF. The extended model accurately recapitulated the component dose-dependent performance of the AND gate (**Figure 4.1c, lower**), providing verification that this extension can describe split intein-based circuits.

Model-guided design of genetic programs: As a first test of the predictive capacity of the revised model, we designed a panel of genetic logic gates. Our modeling objective was to identify promising designs that implement a given function. Given this objective, we opted to not include additional model complexity that might be required to predict all aspects of circuit behavior. To date, TFs based on ZFs^{129,141}, TALEs¹⁸⁰⁻¹⁸³, dCas9^{184,185}, and other proteins¹⁵² have been used to implement transcriptional logic in mammalian cells, and some of these studies make use of splicing^{141,180,184}. Logic has also been implemented using RNA-binding proteins¹⁸⁶, proteases^{187,188}, and designed protein-binding domains¹⁸³. However, we would still benefit from a way to precisely predict behaviors that can be realized using directly corresponding genetic parts, and so we set out to test whether we could use our model to guide the design of a panel of gates.

Throughout, simulations employed a statistical model for gene expression variation, which we have previously shown to be important in accounting for the effect of cellular variation on how an engineered function is carried out across a cell population^{54,129}. To make an IMPLY gate, the AND gate was modified by appending DsDed to intC-ZF1 and co-expressing a VP64-ZF1 activator. Experimental outcomes (i.e., reporter signal readout across component doses) were consistent with the prediction that readout would be low only if DsDed-intC-ZF1 were present in sufficient excess over its splicing partner, VP64-intN, to function as an inhibitor (**Figure 4.1d**). To make a NAND gate, an inhibitor was split into DsDed-intN and intC-ZF1 and co-expressed with an activator. Outcomes were consistent with the prediction that readout would be low only with sufficient reconstitution of the inhibitor (**Figure 4.1e**). These cases demonstrate that model-guided design can identify effective topologies as well as the precise relationship between input component levels and circuit outcomes.

An ideal framework would confer the versatility to achieve a given objective in various ways. We speculated that in all, the properties of COMET, the use of splicing, and potentially the layering of regulators in series might provide a sufficient basis for this capability. As a test, we compared four designs for a NIMPLY gate utilizing different mechanisms. The first two designs used inhibition mediated by ZF1 (**Figure 4.1f**) or DsDed-ZF1 (**Figure 4.1g**). The third used splicing of an VP64-intC-ZF1 activator to a DsDed-ZF1 inhibitor, such that the readout would be high only if VP64-intC-ZF1 were in sufficient excess of its splicing partner DsDed-intN (**Figure 4.1h**). The fourth used a double inversion cascade, in which an upstream inhibitor prevented a downstream inhibitor from acting on the reporter (**Figure 4.1i**); this represents a

variation on a topology that was previously examined in a bacterial context¹⁸⁹. All four designs produced NIMPLY as predicted. We next tested whether the above properties could be utilized simultaneously, and indeed we were able to build an AND gate by splitting the cascade upstream inhibitor into DsDed-intN and intC-ZF10 (**Figure 4.1j**). Unlike ZFa-mediated activation, the cascade exhibited ultrasensitivity (Hill coefficient $n = 2.8$)—a signal transformation in which a small change in input yields a large change in output, and high output is produced only with sufficient input. Ultrasensitivity buffered the circuit against low inputs, such that the output remained low for input levels that in the ZFa case would have produced half-maximal activation (**Figure 4.1k**).

Across the panel, five of the eight gates exhibited a goodness of prediction metric (comparing the simulated and observed outcomes, Q^2) of at least 90%, indicating a high predictive capacity across the dose response landscapes. Even for the gate with the lowest Q^2 (IMPLY), the model correctly predicted the trend across most input dose combinations. Altogether, these results demonstrate the feasibility of model-guided design of genetic logic gates in mammalian cells, and that the choice of parts and mechanism employed to implement a given function yield predictable performance characteristics.

Compression of circuit design using functional modularity: We next investigated whether predictive design could be extended beyond individual gates to generate multi-input multi-output (MIMO) gates. A putative advantage of using orthogonal parts like COMET TFs and promoters is that these parts may be combined without disrupting their functions. However, simply appending modules can lead to inefficient and cumbersome designs, and thus a central focus of our approach was genetic compactness as well as performance. Ultimately, enhancing compactness and minimizing redundancy in the design process might eliminate failure modes and facilitate implementations within the constraints of gene delivery vehicles. Although genetic compression—reducing the number of components required to implement a given specification—has been investigated in other ways such as through recombinase-mediated DNA rearrangement¹⁹⁰, here we sought a previously unattainable form of *topological* compaction by protein multi-tasking (**Figure 4.2a**). We hypothesized that because our genetic parts operate through direct interactions without relying on long-range mechanisms such as chromatin modification, they might exhibit functional modularity, i.e., domains could be concatenated and retain their functions. This property would be of great

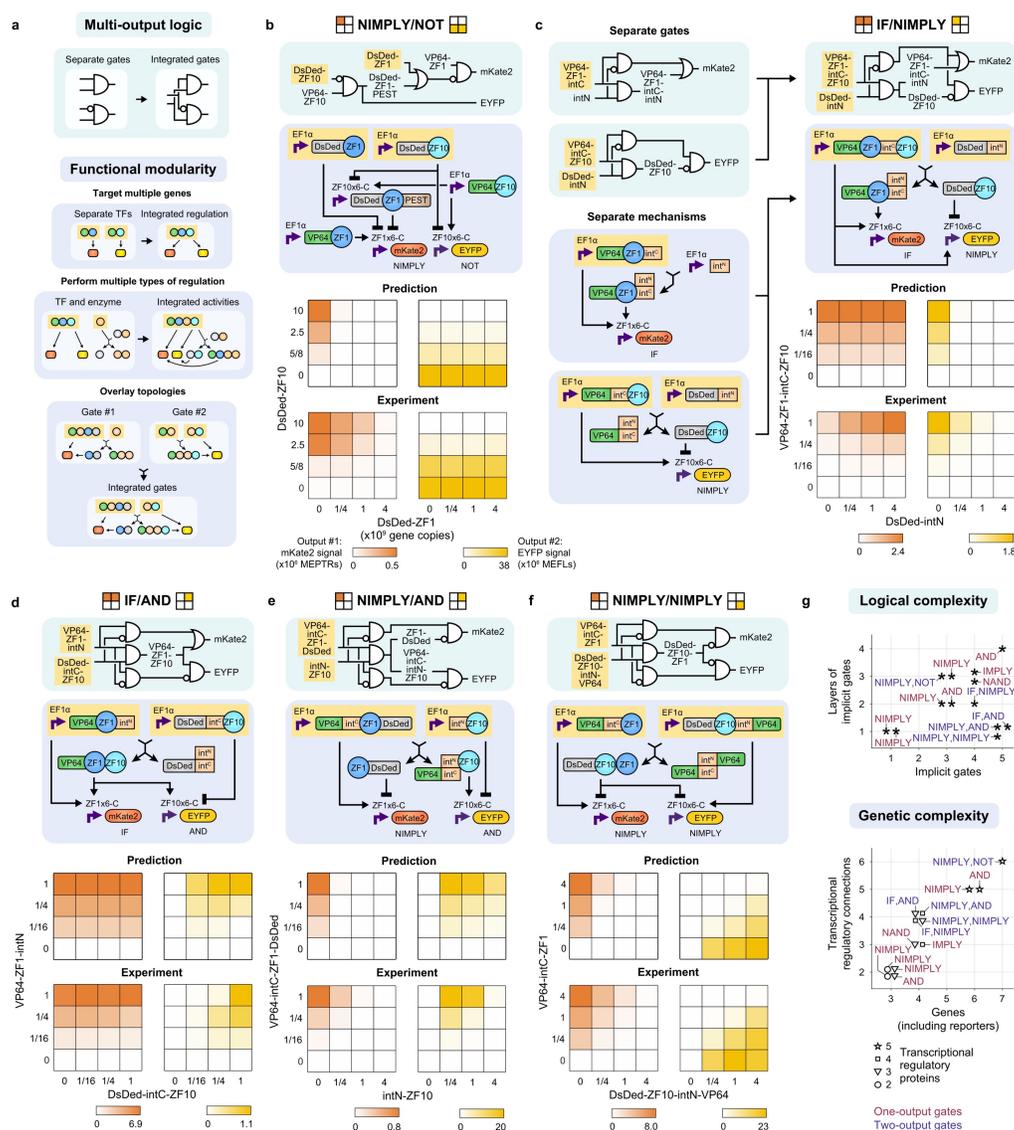


Figure 4.2. Compact multi-output logic is attained through functional modularity. (a) A strategy for multi-output logic is proposed by using multi-tasking proteins that retain the functions of their constituent domains. The cartoons depict the use of multiple DNA-binding domains on a TF to regulate multiple genes, the embedding of a split intein fragment within a functioning TF to enzymatically alter its activity, and the merging of features from multiple genetic programs to enable their compact simultaneous implementation. (b–f) A panel of multi-input-multi-output gates was designed, simulated, and experimentally evaluated. As an example, c) is deconstructed to show how separate topologies containing proteins that have some domains in common and are amenable to the appending of additional domains can be overlaid. In the plots, color-coding denotes the mean mKate2 and EYFP reporter signal from three biological replicates, scaled by maximum value in each heatmap. (g) These plots summarize the complexity of the gates that were designed and validated in Figure 4.1 (red) and Figure 4.2 (purple), with complexity defined based on the size and depth of their logical representation (upper) or based on the numbers of genes, regulatory connections, and regulatory proteins employed (lower). The expanded toolkit of genetic parts and model-guided approach was successful for building circuits spanning a range of attributes, which suggests that this design process could be executed reliably for many future objectives.

utility by enabling multi-tasking proteins that act at multiple promoters or in both transcriptional and post-translational roles, and that could be arranged to execute multiple functions in an efficient fashion.

To investigate these questions, we designed and tested a panel of MIMO gates. As a base case, we simply appended a NIMPLY gate and a NOT gate (**Figure 4.2b**). This success demonstrates the potential for composite functions, but it brings no efficiency relative to the individual gates. To test the possibility of topological compaction, we first designed an IF/NIMPLY gate, hypothesizing that VP64-ZF1-intC-ZF10 would act as a bispecific activator (on two promoters) and interact with an inert DsDed-intN to produce a VP64-ZF1-intC/intN activator and a DsDed-ZF10 inhibitor (**Figure 4.2c**). The second gate, IF/AND, used an activator and an inhibitor to produce a bispecific activator and an insert protein, through essentially the inverse mechanism of that in the IF/NIMPLY gate (**Figure 4.2d**). Third, a NIMPLY/AND gate used a VP64-intC-ZF1-DsDed activator and an intN-ZF10 inhibitor to invert their respective activities. We hypothesized that the former protein would act as an activator, in that DsDed would not preclude VP64 from conferring transcriptional activation (**Figure 4.2e**). Lastly, a NIMPLY/NIMPLY gate used two activators to produce a bifunctional inhibitor and an inert protein. We note that if this circuit had used the same readout for both reporters it would be a XOR gate (**Figure 4.2f**). Overall, the predictions explained most of the variance in the outcomes, and several cases were in close agreement ($\geq 90\%$ Q^2). Minor deviations are potentially attributable to effects such as differences in stability for different proteins; however, we chose not to add such effects because increasing model complexity could lead to overfitting. Moreover, this choice did not preclude model-guided identification of high-performing designs.

A surprising finding was that when we examined performance at the single-cell level, some population-level outcomes were driven by cell subpopulations. In some circuits, subpopulations induced one reporter or the other, but not both, and thus population outcomes were driven by shifts in subpopulation frequencies. In other circuits, this task distribution was not apparent. Although neither behavior was an explicitly designed feature, both types of behavior were predicted by simulations. Altogether, the gates described in **Figures 4.1,4.2** span a wide range of attributes including the number of genes, regulatory connections, and regulatory proteins (**Figure 4.2g**). The successful development of these circuits without the need for additional tuning demonstrates that this framework may be well-suited to overcoming complexity-associated barriers with mammalian genetic program design.

Implementation of analog signal processing: Although digital logic has many uses, biology also employs analog signal processing for many purposes, and we next examined whether our tools could be employed in this way. The first property that we sought to implement was ultrasensitivity, which is desirable in engineering sharp activation^{191,192} and is observed in the natural control of processes like cell growth, division, and apoptosis¹⁹³. The second property was bandpass concentration filtering, in which an output is produced only within a range of input values^{187,194}. Bandpasses are salient for both natural and synthetic spatial patterning¹⁹⁵. To develop a strategy for implementing these properties, we made use of existing mechanistic insights. Previously, we determined that ZFa-mediated activation is cooperative at the level of transcription initiation, and in comparing promoter architectures, maximal transcription increased with the number and compactness of binding sites¹²⁹. This COMET promoter feature confers both high inducibility and potent inhibition. We also deduced that TF *binding* to promoters is generally non-cooperative, so transcriptional output is not ultrasensitive ($n = 1$) to ZFa dose. To construct systems exhibiting ultrasensitivity ($n > 1$), we examined several strategies in which the output is inhibited only at low activator doses. The first design paired a VP16-intN input with intC-ZF. We reasoned that at low VP16-intN doses, intC-ZF-mediated inhibition would dominate, and at high doses, transactivation by reconstituted VP16-ZF would dominate. We also tested this concept with the addition of DsDed-ZF to promote more relative inhibition at low input doses. However, the increase in ultrasensitivity was modest for these cases. We reasoned this was due to insufficient inhibition at low activator doses based on decreased stability conferred by the intC domain to the inhibitor.

In comparison to a ZFa base case (**Figure 4.3a**) ($n = 1.0$), a VP16-ZF input with DsDed-ZF was more ultrasensitive ($n = 1.8$) (**Figure 4.3b**). This led us to consider a vehicular analogy: the circuits with DsDed-ZF are akin to applying the brake (inhibition) while applying the accelerator (activation), but a more effective approach might be to release the brake as the accelerator is applied. To realize this concept and circumvent stability effects, we used a chemically responsive COMET TF (RaZFa) in which rapamycin-induced heterodimerization domains FRB and FKBP are fused to an AD and a ZF, respectively. In the presence of rapamycin (which in this scenario is not an input, but rather an environmental species), heterodimerization of VP16-FRB and FKBP-ZF converts FKBP-ZF (brake) into RaZFa (accelerator), which

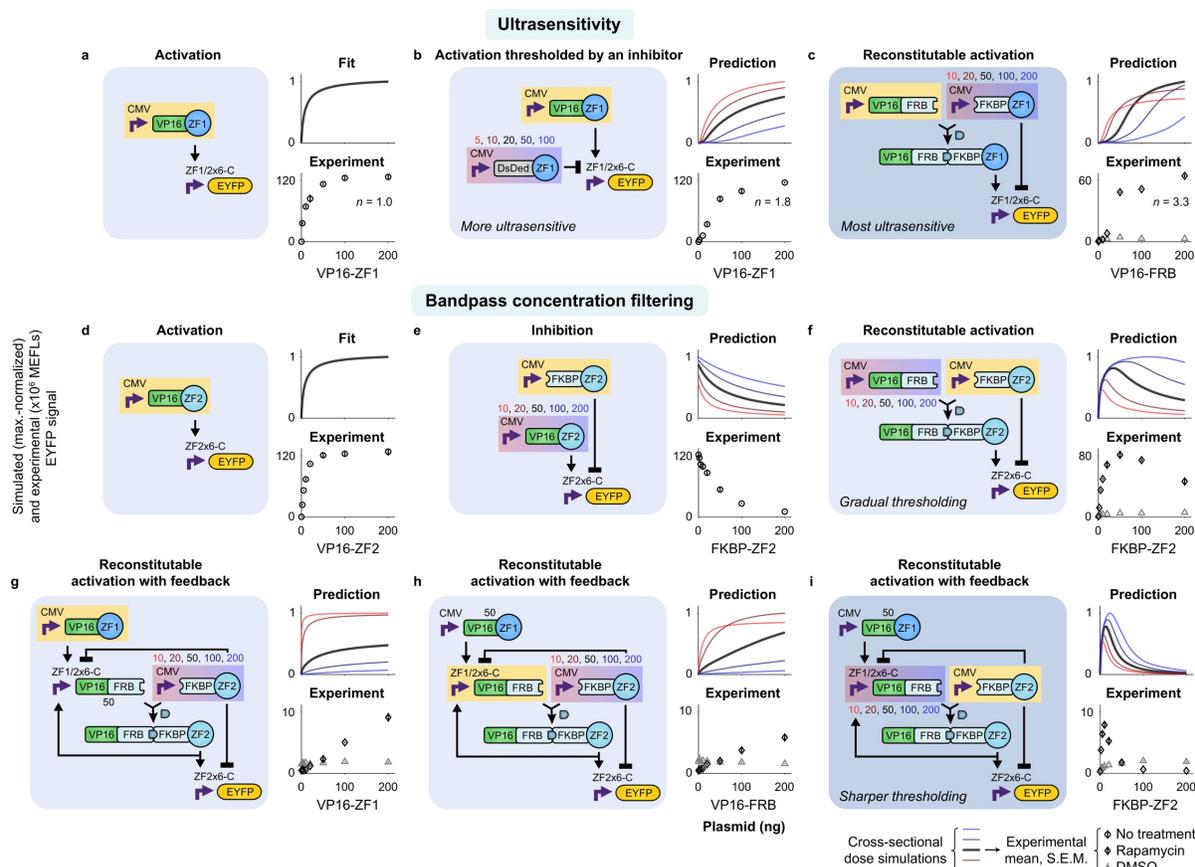


Figure 4.3. Analog behaviors are constructed by using TFs that play multiple roles. Reconstitutable TFs have dose response properties that are conducive to analog signal processing. Simulated and experimentally observed responses are shown relating to (a–c) ultrasensitivity and (d–i) bandpass concentration filtering. Several designs were evaluated for the ability to meet these objectives. To implement ultrasensitivity, the Hill coefficient (n) was most effectively increased through a strategy of removing an inhibitor in the process of producing an activator (c). To implement bandpass concentration filtering, a tighter upper threshold was best achieved through a similar strategy that also included additional regulation: moderate levels of FKBP-ZF act primarily to reconstitute RaZFa, and high levels of FKBP-ZF act to inhibit the reporter and VP16-FRB (i). Simulations in a and d are fitted to data, and the other panels are predictions. The prediction plots present simulations for how output gene expression varies with dose of the component highlighted in yellow; each plot includes a set of responses varying with the component highlighted in red-to-blue gradation. Doses for the x-axes and above the varied component in the diagrams are in plasmid ng. Each experimental plot corresponds to the simulated condition with the dark line (for the middle dose of the varied component). The ZF1/2x6-C promoter has six partially overlapping ZF1 and ZF2 sites. DMSO is the vehicle for rapamycin, which is used here as an environmental species (not an input). The simulations with RaZFa correspond to conditions with rapamycin treatment. Experiment plots represent the mean and S.E.M. of EYFP reporter signal from three biological replicates.

induces the reporter. With rapamycin, the response of this circuit to VP16-FRB input indeed exhibited greater ultrasensitivity ($n = 3.3$), consistent with the prediction (**Figure 4.3c**). Thus, in this system, ultrasensitivity can arise through cascades (**Figure 4.1**) or reconstitution (**Figure 4.3**), and neither mechanism requires the cooperativity in TF-DNA binding that is often associated with ultrasensitive responses.

We next investigated circuits to implement bandpass concentration filtering, which can be used to produce a response within a range of input signal. Our strategy was to use mechanisms that inhibit reporter output only at high doses of activator input, and the predictions were based on a fitted ZFa base case (**Figure 4.3d**). We hypothesized that although FKBP-ZF is necessary for RaZFa-mediated activation, excess FKBP-ZF would be inhibitory. We confirmed that FKBP-ZF acted as an inhibitor (**Figure 4.3e**), and implemented a test circuit with VP16-FRB and rapamycin; as predicted, the response to FKBP-ZF input showed a peak in output, but no sharp upper threshold (**Figure 4.3f**). Based on these results, we designed a new topology for a sharper bandpass; of the motifs in this design, simulations suggested the two paths of negative regulation from FKBP-ZF were the most important. With VP16-ZF or VP16-FRB as the input, responses were activating (**Figure 4.3g,h**), demonstrating cross-sectional predictive capacity for a complex genetic system. With FKBP-ZF as the input, we expected that at zero dose there would be no activation, at moderate doses there would be, and at high doses excess FKBP-ZF would decrease reconstitution (by inhibiting induction of VP16-FRB) and inhibit the reporter. The outcome closely matched the prediction of a bandpass with a sharp upper threshold (**Figure 4.3i**). These results demonstrate a design capability for not only digital gates, but also analog behaviors, which can be realized using chemically responsive parts and non-canonical topologies.

Integrating genetic circuits with sensors to build sense-and-respond functions: While the predictive design of genetic programs is a substantial technical advance, translating this capability to other technologies will require integrating genetic circuits with native or synthetic parts that sense and modulate the state of the cell or its environment. A recurring challenge associated with this compositional goal is level-matching the output of a sensor to the input requirements of a downstream circuit^{54,196}. We investigated whether our designed circuits could overcome this challenge and be effectively linked to sensors without requiring trial-and-error tuning. In support of the feasibility of adding an upstream sensor

layer, simulations suggested that ZFa could be arranged in series without prohibitively driving up background or dampening induced signal. Thus, we examined two classes of synthetic sensors (transmembrane and intracellular) for which we hypothesized signaling could be made COMET-compatible. For the first, we selected the modular extracellular sensor architecture (MESA)—a self-contained receptor and signal transduction system that transduces ligand binding into orthogonal regulation of target genes^{167,197}. In this mechanism, ligand-mediated dimerization of two transmembrane proteins called the target chain (TC) and protease chain (PC) promotes PC-mediated proteolytic *trans*-cleavage of a TC-bound TF. We explored several strategies for building COMET-compatible MESA based on a recent scaffold¹⁹⁸ and the parts developed in the current study. The best performance was observed using rapalog-inducible MESA that release ZFa for activating signaling or DsDed-ZF for inhibitory signaling (representing a new function for this receptor), and the ZFa format was carried forward.

For the second sensor, we built a new TF—ABA-ZFa, which is analogous to RaZFa—by fusing the abscisic acid (ABA)-binding domains PYL1 and ABI1¹⁷⁴ to an AD and a ZF, respectively. We observed that both sensors displayed excellent performance with reporter induction upon ligand treatment (**Figure 4.4a,b**). For Rapa-MESA-ZF6a (ZF6a was selected for its high potency stemming from cooperative transcriptional activation¹²⁹), the ligand-inducible fold difference in signal was ~200x, several fold higher than for recent receptors using tTA¹⁹⁸ and that signal through a distinct mechanism¹⁹⁹, and thus the highest for this system to date. For ABA-ZF2a (ZF2a was also selected for its high potency), ligand-independent signal was unobservable and induced signal was high, yielding essentially perfect performance. Thus, both sensors have a low off state and a high on state and benefit from the advantageous property of COMET promoter-based cooperativity.

We carried forward the two validated sensors and examined whether downstream circuits comprising genetic parts and designed topologies from this study could be seamlessly linked with the new input layer. To this end, we designed a panel of synonymous topologies that implement AND logic through different mechanisms (**Figure 4.4c**): (1) a hybrid promoter with alternating TF sites (based on a similar architecture from the COMET study), (2) splicing (as in **Figure 4.1c**), (3) splicing with DsDed (as in **Figures 4.1d, 4.2d** for tighter inhibition), and (4) and splicing with feedback (as in **Figure 4.3g-i**). All four topologies exhibited AND behavior when tested using ZFa as inputs (**Figure 4.4d**), demonstrating the versatility for

attaining a given objective in multiple ways. Moreover, when coupled to ligand-activated sensors, these circuits still conferred AND behavior, and there was no substantial loss of performance (i.e., fold induction with two ligands remained much greater than with each one individually) in carrying out this more complex sensing function (**Figure 4.4e**). A comparison across the designs provides some insights. The hybrid promoter in topology 1 was high-performing, and the splicing topologies in 2–4 generally yielded improvement, despite the additional cascade layer, by reducing one-input signals not only to near zero-input levels, but also to near reporter-only background. Of the topologies examined, 2 and 3 were the most effective at producing a high two-input signal and low signals in other cases. These results demonstrate that genetic programs can be designed by a model-driven process that uses directly corresponding parts and varied mechanisms, and then these programs can be readily linked to different classes of sensors for multiplexed evaluation of ligand inputs without incurring a performance hit.

4.5 Discussion

In this study, we generated regulators integrating transcriptional and post-translational control and validated a modeling framework for accurate genetic program design and prediction. Experimental observations of dose response behavior closely matched simulations, even in scenarios employing new proteins, including those with many domains, and new topologies, including those with many interacting components. The gates spanned a range of logical complexity and genetic complexity (**Figure 4.2g**), and the capability to construct these functions represents an advance in mammalian synthetic biology. Throughout, no trial-and-error was needed to arrive at the specified design goals. Once the base case parts had been characterized, no additional parameterization was needed to simulate new regulators. Since the mechanisms employed for binding, splicing, activation, and inhibition can be described by concise formalisms, no fundamental revamping was needed to make predictions from an originally descriptive model. Lastly, even though a relatively small set of protein domains was utilized, we were able to combine the domains in many ways, and no large libraries were needed to produce the variety of behaviors observed.

Key properties that enabled sophisticated design included antagonistic bifunctionality²⁰⁰, in which a component can exert opposing effects on a target gene depending on the other components in the circuit, and functional modularity, which enabled multiple activities to be encoded in proteins. Sophisticated design

was also enabled through split genetic parts, including those that splice or dimerize. Split parts are conducive to encoding both digital (**Figures 4.1,4.2**) and analog (**Figure 4.3**) functions. It is also worth noting that split parts shift some of the regulation from the transcriptional level to protein-protein interactions, which could increase the speed of signal processing and in the genomic context might reduce the risk of gene silencing. Another benefit, relating to circumventing cargo limitations of gene delivery vehicles, is that large programs could be distributed across multiple vectors²⁰¹ such that a function is reconstituted only in cells receiving all vectors.

Finally, we demonstrated that COMET TFs can be fused to MESA to produce best-in-class receptors and that these are interoperable with downstream programs (**Figure 4.4b,e**); seamless level-matching was achieved through the cooperative potency of the released TF aligning with the responsiveness of the promoter in the downstream layer. Altogether, these attributes, in combination with the many ways in which components can be arranged to regulate each other, greatly expand the genetic program design space. We conclude that the new components and quantitative approaches developed here should be effective for engineering mammalian cellular functions.

4.6 Acknowledgements

I thank Vis Kandula and Amy Hong for tremendous effort and insight throughout the project; Patrick Donahue and Jon Boucher for collaboration; Taylor Dolberg for assistance with experiments; Tae-Eun Kim for testing split intein-based TFs; Kate Dray, Joe Draut, and Juliana Feng for assistance with cloning; Kate Dray, Everett Allchin, and Chris Coleman for model discussions; and my lab colleagues for discussions.

CHAPTER 5. A genome-scale knockout screen for negative regulators of NF- κ B activity

This chapter is in preparation as:

Muldoon J.J.*, Sankaranarayan S.*, Clutter M.R., Bagheri N., Leonard J.N. *In preparation*.²⁰² *Equal contributions

5.1 Abstract

Immune cells such as monocytes and macrophages use the pattern recognition receptor TLR4 to sense pathogen-related or damage-related cues and initiate signaling to regulate antimicrobial and inflammatory responses. TLR4 was of central importance in the macrophage quorum licensing study, in that it mediated the response to LPS by signaling to activate the transcription factor NF- κ B to induce TNF expression (**Chapter 2**). In previous work²⁰³, the Leonard lab investigated structural perturbations to TLR4 and proposed that the standard model describing the *initiation* of signaling needs refinement. The revised model involves a putative TLR4-inhibitory complex (TIC) that retains the receptor in a signaling-incompetent state prior to ligand binding. Here, we set out to evaluate the TIC hypothesis, starting with a genome-scale CRISPR/Cas9 knockout screen in THP-1 monocytes for constitutive negative regulators of the activity of NF- κ B. The results provide a set of genes that in future work can be investigated to elucidate TLR4-dependent or TLR4-independent roles in regulating NF- κ B activity and for identifying TIC.

5.2 Introduction

Toll-like receptors (TLRs) mediate the recognition of microbial patterns and tissue damage²⁰⁴⁻²⁰⁶. Aberrant TLR signaling is implicated in autoimmunity and cancer²⁰⁷⁻²⁰⁹, and the modulation of TLR signaling is a therapeutic target^{210,211}. Each TLR has a ligand-binding ectodomain, a transmembrane domain (TMD), and an intracellular Toll/IL-1 receptor (TIR) domain. In the canonical model for the initiation of signaling, ligand binding to ectodomains (ECDs) leads to receptor dimerization and the recruitment of downstream mediators to dimerized TIR domains. Although downstream mediators have been identified, the mechanism by which receptor dimerization induces signaling is not well understood²¹². Among the TLRs, TLR4 induces inflammation in response to the lipid A component of LPS from bacteria. Excessive signaling, however, leads to harmful inflammation, and in the extreme, septic shock^{204,207,208,210}. TLR4 also senses a variety of non-canonical ligands²¹³⁻²¹⁹, though the mechanism and utility of this capability remain unclear.

TLR4 is a structurally and mechanistically distinct TLR. The adaptor protein myeloid differentiation factor 2 (MD-2)²²⁰ binds to the TLR4 ECD and facilitates TLR4-MD-2 cell surface colocalization²²¹. When the MD-2 hydrophobic core binds lipid A, MD-2 exposes an interface that binds the ECD of another TLR4, leading to formation of a heterotetrameric TLR4-MD-2 that is stabilized by ECD-ECD interactions²²². The

TLR4 ECD is involved in preventing constitutive signaling, and in the steric occlusion model, ECDs sterically occlude TIR dimerization in the absence of ligand²²³. However, several observations challenge this model. For example, it was shown that the deletion of nearly the entire ECD did not induce constitutive signaling whereas deletion of a slightly larger portion did²²³. Additionally, the insertion of a flexible amino acid linker between the TMD and ECD decreased LPS-inducible signaling but did not lead to constitutive signaling²²⁴.

A recent study in the Leonard lab provides a basis for a revised model of TLR4 signaling. Daringer et al.²⁰³ generated and studied structural perturbations to the receptor and made the following observations: (i) transmembrane-anchored TIR domains (without ECDs) were sufficient to initiate constitutive signaling; (ii) sequestering TIR domains inflexibly at the membrane by deleting the native 10 aa intracellular linker prevented signaling; and (iii) introducing a long intracellular linker did not elevate signaling in the absence of ligand, and signaling remained LPS-inducible; this outcome is inconsistent with the steric occlusion model, because this long linker when fully extended would span nearly twice the distance to bridge the widest part of the ECD²²². Furthermore, signaling from this receptor variant was LPS-inducible, which does not support a mechanism in which ligand-induced receptor dimerization or conformational rearrangement of ECDs leads to an intracellular conformational change that alleviates steric inhibition of TIR dimerization.

The above observations led to a hypothesis that in the absence of ligand, TLR4 is retained in a signaling-incompetent conformation through interactions between TLR4 and one or more as-of-yet unidentified species termed the TLR4-inhibitory complex (TIC) (**Figure 5.1**). TIC would bind the C-terminal region of the TLR4 ECD, TIR domain, and one or more interaction partners, which could be adaptor proteins and/or other molecules. If TIC-bound TIR domains were held in this conformation, then upon ligand-binding-induced ECD dimerization and/or rearrangement, interactions with TIC would be disrupted and TIR domains would dimerize and initiate signaling. TIC would retain the TIR domains in a membrane-proximal location,

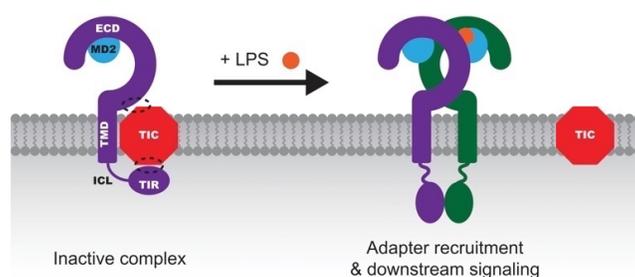


Figure 5.1. The TIC hypothesis. Illustration from Daringer, et al. of the hypothesis in which a TLR4 inhibitory complex negatively regulates the initiation of TLR4 signaling in the absence of ligand.

which prevents signaling, or otherwise preclude interactions between TIR domains and adapter proteins. This model could explain why TLR4 signaling is inducible by non-canonical ligands: if induction of signaling requires disruption of TIC binding, rather than ligand-guided formation of a specific conformation, then there could exist multiple mechanisms for inducible signaling.

Given the unknown of what constitutes TIC, we set out to evaluate the TIC hypothesis using a genome-scale screen. The use of libraries is an established approach for phenotype-to-genotype screens, and in combination with next-generation sequencing they can be implemented in pooled format. CRISPR/Cas9 has been adapted for this purpose to identify genes that regulate various cellular processes²²⁵⁻²²⁷. Here we used CRISPR/Cas9 screening to identify regulators of basal NF- κ B activity in a THP-1 monocyte reporter cell line that we generated. We note that several screens have been conducted previously for regulators of NF- κ B activity, generally differing from our approach, as described in **Table 5.1**.

Table 5.1. Prior NF- κ B screens.

Study	Description
Chew, et al. ²²⁸	Genome-scale pooled knockdown screen by transfecting siRNAs into an A549 cell line containing an integrated NF- κ B-inducible luciferase reporter and treating with or without TNF.
Teo, et al. ²²⁹	Genome-scale arrayed gain-of-function screen by transfecting cDNAs and an NF- κ B-inducible luciferase reporter into the 293T cell line and treating with TNF.
Gewurz, et al. ²³⁰	Genome-scale arrayed knockdown screen by transfecting siRNAs into an HEK293 cell line containing an integrated NF- κ B-inducible GFP reporter and a Tet-inducible LMP1 gene.
Lee, et al. ²³¹	Genome-scale pooled insertional inactivation screen by transducing the KBM7 cell line with a gene-trap retrovirus that has an NF- κ B-inducible blasticidin resistance gene and treating with blasticidin to select for resistant cells.
Olarerin-George, et al. ²³²	Middle-scale arrayed screen by transfecting miRNA mimics (several hundred) into an HEK293 cell line containing an integrated NF- κ B-inducible luciferase reporter and treating with/without TNF.
Feldman, et al. ²³³	Middle-scale imaging-based pooled screen for perturbations that affect NF- κ B nucleocytoplasmic translocation by transducing a lentiviral sgRNA library into a HeLa cell line containing integrated p65-mNeon and inducible Cas9, and treated with IL-1 β or TNF; screens were also conducted in other cell lines.
Verma, et al. ²³⁴	Genome-scale arrayed knockdown screen by transfecting siRNA into macrophages (differentiated from the THP-1 monocyte cell line using PMA) containing an integrated NF- κ B-inducible luciferase reporter and treating with LPS or TNF.
Zablocki-Thomas, et al. ²³⁵	Genome-scale pooled screen by lentivirally transducing the GeCKO v2 sgRNA library into a KBM7 cell line containing an integrated NF- κ B-inducible GFP, constitutive TLR3, and constitutive Cas9, with poly(I:C) treatment.
Covarrubias, et al. ²³⁶	Genome-scale pooled screen by lentivirally transducing a customized sgRNA library into immortalized murine bone marrow-derived macrophages containing an integrated NF- κ B-inducible GFP reporter and constitutive Cas9 ²³⁷ and treating with/without LPS.
Sfikas, et al. ²³⁸	Middle-scale arrayed screen by transfecting siRNA (several hundred) into a U2OS osteosarcoma cell line containing an integrated NF- κ B-inducible luciferase and treating with etoposide or ionizing radiation to induce DNA damage.

5.3 Materials and Methods

Vector cloning and purification: A lentiviral vector containing an NF- κ B-inducible EGFP and a CMV-driven Cas9 fused via T2A peptide to a hygromycin resistance gene was generated to provide a readout for the screen. To construct the vector, an NF- κ B-responsive cassette (NF- κ B-ELAM-EGFP) was PCR-amplified from the NEE vector²³⁹ and inserted into a vector (pPD094) containing a CMV-driven mCherry fused via T2A peptide to a hygromycin resistance gene. Cas9 was PCR-amplified from lentiCas9-Blast (Addgene #52962) and inserted into the vector to replace mCherry. Cloning was conducted using standard restriction and ligation methods (New England Biolabs and Thermo Fisher Scientific), followed by transformation into chemically competent TOP10 *E. coli* (Thermo Fisher Scientific). Cells were plated onto LB/ampicillin agar plates and incubated at 37°C, and colonies were picked and grown in LB/ampicillin medium at 37°C in a shaking incubator. Vector was isolated using an E.Z.N.A. plasmid mini kit (Omega Bio-Tek), and inserts were sequence-verified using DNA sequencing services through ACGT, Inc. Vector was prepared using polyethylene glycol-based extraction. DNA purity and concentration were measured using a Nanodrop 2000 (Thermo Fisher Scientific).

HEK293FT cell culture: HEK293FT cells were cultured in complete DMEM: 1% DMEM powder (Gibco #31600091), 0.35% w/v D-glucose (Sigma #50-99-7), 0.37% w/v sodium bicarbonate (Fisher #S233-500), 10% heat-inactivated FBS (Gibco #16140071), 4 mM L-glutamine (Gibco #25030081), and 100 U ml⁻¹ penicillin and 100 μ g ml⁻¹ streptomycin (Gibco #15140122). Cells were grown in tissue culture-treated 10 cm dishes (Corning #500001672) at 37°C in 5% CO₂. To passage, medium was aspirated, and cells were washed in PBS, incubated in trypsin-EDTA (Gibco #25300054; 37°C, 5 min), detached by tapping the dish, and resuspended in fresh medium and plated.

Lentiviral packaging and harvest: Lentivirus was produced by transfecting HEK293FT cells using the calcium phosphate method. Cells were plated (3×10^5 cells ml⁻¹, 10 ml) in fresh medium in 10 cm dishes. At ~8 h after plating, by which time the cells had adhered, the reporter vector (10 μ g), a plasmid for constitutive expression of EYFP (1 μ g) as a transfection control, and the second-generation lentiviral packaging plasmids pSPAX2 (8 μ g) and MD2G (3 μ g) were mixed with 2M CaCl₂ (150 μ l) and sterile water (for a total volume of 1 ml) and pipetted dropwise into sterile 2 \times HEPES-buffered saline (500 mM HEPES, 280 mM NaCl, 1.5 mM Na₂HPO₄; 1 ml). This mixture was pipetted four times gently, incubated for 3 minutes,

pipetted many times vigorously, and added dropwise onto cell culture. Cells were incubated overnight, and the next day medium was aspirated and replaced with fresh medium. At 28 h after the medium change, EYFP expression was confirmed by fluorescence microscopy, the medium was pipetted from cell culture into conical tubes and centrifuged ($500 \times g$, 2 min, 4°C), and the lentivirus-containing supernatant was filtered through a sterile $0.45 \mu\text{m}$ filter and stored at -80°C .

THP-1 monocyte cell culture: Monocytes were cultured in complete RPMI-1640 medium: RPMI-1640 (ATCC #30-2001 or Gibco #A1049101), 10% FBS, and 100 U ml^{-1} penicillin and $100 \mu\text{g ml}^{-1}$ streptomycin. Cells were grown in suspension in tissue culture-treated 10 cm dishes and incubated at 37°C and 5% CO_2 . Cells were passaged at a density of 10^5 – 10^6 cells ml^{-1} by diluting with fresh complete medium, or by centrifuging in 50 ml conical tubes ($150 \times g$, 5 min), removing the medium, resuspending the cells in fresh medium, and plating.

Cell line generation: THP-1 cells were transduced with the lentiviral reporter by spinoculation. Cells (10^5 cells ml^{-1} , 2 ml) and viral suspension (4 ml) were combined in 50 ml conical tubes. A no-virus control was included and was brought to 6 ml with complete RPMI-1640 medium. Polybrene (Sigma-Aldrich #TR-1003; $6 \mu\text{l}$, for 0.1% v/v in the mixture) was added. Following centrifugation ($1200 \times g$, 1 h, 25°C), the supernatant was aspirated, and cells were resuspended in complete RPMI-1640 medium (2 ml) and plated in a tissue culture-treated 6-well plate (Corning). After two days, each sample was split into two wells, and to one of these wells, hygromycin B (Thermo Fisher Scientific #400053; $300 \mu\text{g ml}^{-1}$) was added. Cells were cultured for one week (passaging with 50% fresh medium and 50% conditioned medium) to select for transduced cells. Non-transduced cells were observed not to survive the antibiotic treatment. Reporter cells were subsequently cultured using a maintenance concentration of $200 \mu\text{g ml}^{-1}$ hygromycin B.

Reporter validation by analytical flow cytometry: To test for reporter inducibility, THP-1 cells and reporter cells were plated in 6-well plates and treated with *E. coli* 055:B5 LPS (Sigma-Aldrich; 100 ng ml^{-1}) or not treated in biologic triplicate. 24 h later, samples were prepared for flow cytometry. The contents of each well were pipetted into FACS tubes and centrifuged ($150 \times g$, 5 min), the liquid was decanted, and two drops of FACS buffer (PBS pH 7.4, 5 mM EDTA, 0.1% BSA) were added. Samples were kept on ice and wrapped in foil, and then run on a BD LSR Fortessa special order research product. EGFP signal was measured in the FITC channel with a 488 nm excitation laser and a 505LP 530/30 nm filter. Approximately

10^4 live single-cell events were collected per sample. Data were analyzed using FlowJo software (FlowJo, LLC) to gate on single-cell (FSC-A vs. FSC-H) and live (FSC-A and SSC-A) bases. The mean reporter signal in MFI for each sample was converted to units of Molecules of Equivalent Fluorescein (MEFLs) using a calibration curve generated from UltraRainbow Calibration Particles (Spherotech #URCP-100-2H).

Library transduction: The Brunello library contains sgRNA lentiviral vectors for the pooled knockout of 19,114 genes in human cells. For each screening replicate, reporter cells were independently transduced (1 h centrifugation at $1000 \times g$, followed by 4–5 h in a 37°C culture incubator) in 12-well plates with $8 \mu\text{g ml}^{-1}$ polybrene and an amount of lentivirus corresponding to an MOI of approximately 0.3. Enough cells were transduced to achieve coverage of approximately 400–500 cells per sgRNA in the library. Transduced cells were selected by treating with puromycin (Invivogen #NC9138918, $1 \mu\text{g ml}^{-1}$) for four days. Brunello-transduced reporter cells were cultured for an additional four days in the absence of puromycin prior to FACS.

Flow-activated cell sorting: At approximately one week after library transduction, cells were sorted using a BD FACS Aria four-laser special order research product. Gates were drawn to identify live single-cell events (based on forward scatter and side scatter) and to distinguish EGFP signal (using the FITC channel with a 488 nm excitation laser and a 505LP 525/30 nm filter). The mid gate was drawn to cover the 5% of cells with the highest EGFP signal, except for the 0.25% of cells with the highest signal which were covered by the hi gate. Approximately $1\text{--}2 \times 10^6$ mid cells and $5\text{--}7 \times 10^5$ hi cells were collected.

Sample preparation, NGS, and bioinformatic analysis: Pre-sort, GFP-mid sorted, and GFP-hi sorted cell populations from each replicate were stored at -80°C until genomic DNA extraction was performed. Genomic DNA was extracted from cell pellets using a Quick-DNA kit (Zymo #D4068 or #D4075). Library preparation was performed by PCR using *Ex Taq* polymerase (Takara #RR001A) with up to $10 \mu\text{g}$ of genomic DNA per reaction. A mix of P5 stagger primers and P7 primers with a unique barcode for pooled sample deconvolution were used to amplify the sgRNA-containing region from genomic DNA²⁴⁰. PCR products were analyzed by agarose gel electrophoresis, and amplicons of the expected size were isolated using Zymoclean gel DNA recovery kit (Zymo #D4008). Samples were sequenced at NUSeq using an Illumina NextSeq500 instrument with a 75 bp single-end protocol to generate approximately 10^8 reads. Data were processed using CRISPRCloud2 by comparing the mid post-sort to pre-sort and the hi post-sort to

pre-sort. This step amalgamated the data across the three replicate screens and yielded two types of ranked lists of hits: barcode-centric, which considers each barcode separately, and gene-centric, which incorporates the results across the four barcodes for each gene. Thus, there were four lists: mid barcode-centric, mid gene-centric, hi barcode-centric, and hi gene-centric. In each list, the hits contained associated metrics such as \log_2 fold enrichment and p -value. The mid lists were more enriched in low p -values among the top hits for \log_2 fold enrichment, suggesting more reliable outcomes (consistent with obtaining a greater number of cells for mid than for hi during FACS), and the barcode-centric lists distinguished which specific sgRNAs might be effectively employed in follow-up validation, so we chose to focus on the mid barcode-centric list.

Prioritization of candidate regulators: We analyzed the 454 hits (439 genes) from the mid barcode-centric list that showed greater than two-fold enrichment. Genes were annotated based on prior knowledge from Uniprot and literature sources by criteria: localization (if known, then if a membrane protein, and if so, then if integral membrane, peripheral membrane, or lipid-anchored; or if not membrane-associated) and function (if known, then if a regulator of NF- κ B activity; or, if unknown, then if suspected to be a regulator). Of the hits that were known or suspected regulators of NF- κ B activity, 100 barcodes (96 genes) were known or suspected to be membrane-associated, and 40 barcodes (35 genes) were known or suspected not to be membrane-associated.

5.4 Results

The first goal was to produce a cell line with which to conduct a screen for genes that constitutively negatively regulate NF- κ B activity. Starting with the THP-1 monocyte cell line, which expresses TLR4 and can signal through this receptor, we generated a reporter line by lentivirally transducing a vector containing a constitutive Cas9 fused via T2A peptide to a hygromycin resistance gene and an NF- κ B-inducible GFP reporter. This line was subsequently transduced with the Brunello sgRNA library (**Figure 5.2a**). The logic of the screen is that reporter signal would be expected if a negative regulator of NF- κ B were knocked out, and not in the absence of knockout or if unrelated genes were knocked out (**Figure 5.2b**). The reporter line exhibited approximately 50-fold LPS-inducible reporter signal, indicating that TLR4 signaling could be distinguished by this readout (**Figure 5.2c**). Screens were conducted in biologic triplicate by transducing

the library, isolating cells using FACS that in the basal state expressed EGFP, and identifying hits through sgRNA barcode sequencing (**Figure 5.2d**). In this process, two post-sort populations were collected: the mid gate covered the 5% of cells with the highest signal, except for the 0.25% with the highest signal, which were covered by the hi gate. Bioinformatic analysis using CRISPRCloud²²⁴¹, in which results from the three replicate screens were combined, yielded barcode and gene lists with associated fold enrichment and p -values for the comparison of mid post-sort to pre-sort and of hi post-sort to pre-sort.

To identify candidate regulators independent of potential sgRNA-to-sgRNA variation in knockout efficiency for each gene, we took a barcode-centric approach, i.e., post-sort enrichment of each barcode was treated independently of other barcode outcomes, rather than amalgamating outcomes for each gene. The hits displayed up to several fold enrichment and spanned several orders of magnitude in p -values

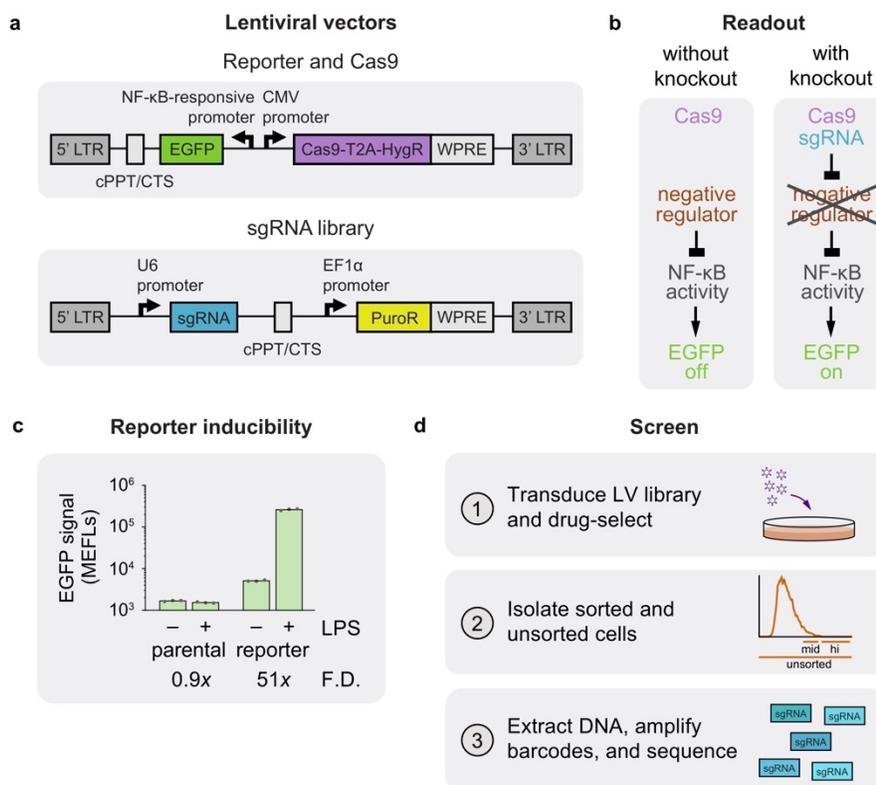


Figure 5.2. Screen design and implementation. **a** Diagram of the vectors for the reporter and Cas9 and for the sgRNA library. **b** Strategy for identifying negative regulators of NF- κ B activity through gene knockouts. **c** THP-1 monocytes transduced with the reporter and Cas9 vector inducibly express EGFP. Fluorescence was measured at one day after LPS treatment (100 ng ml⁻¹) (one-tailed Welch's t -test, $p = 4 \times 10^{-4}$ for the reporter line). Conditions are in biologic triplicate, and bars indicate the mean. F.D. is the fold difference in signal. **d** Summary of the methodology for screening and identification of candidate regulators. Library transduction and FACS were conducted in biologic triplicate.

based on the enrichment across replicate screens (**Figure 5.3**). We chose to focus on the mid results, because, compared to the hi results, the barcodes that were most enriched were more likely to have low p -values (< 0.05). 454 hits (439 genes) had greater than two-fold enrichment. The hits were annotated based on prior knowledge of protein subcellular localization and potential role in regulating NF- κ B activity. 140 barcodes (131 genes) were known or suspected regulators of NF- κ B activity. Going forward, I recommend that the fuller set, particularly of the 439 genes, is also reviewed by others to identify any other promising hits; e.g., there are numerous general RNA-binding proteins that were not carried over to this stage.

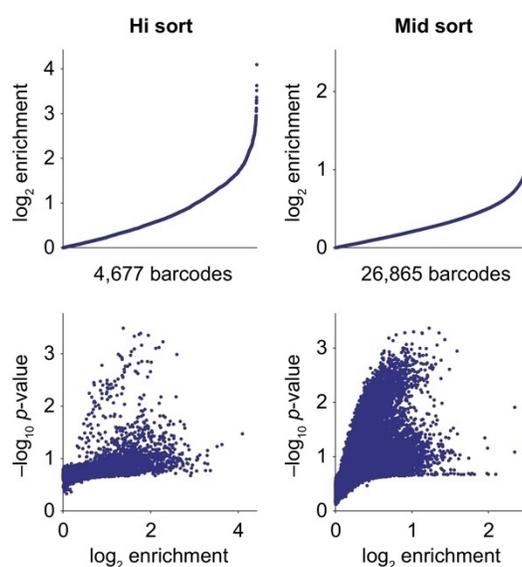


Figure 5.3. Screen results. Waterfall plots and volcano plots for barcodes that were enriched in the post-sort populations compared to the pre-sort population. The hi and mid results encompass 4,224 genes and 15,921 genes, respectively.

5.5 Discussion

The screen results provide a list of candidate genes that can be investigated in validation experiments for roles in regulating basal NF- κ B activity, and potentially TLR4 signaling in the form of TIC. There are several ways in which individual-gene arrayed validation experiments could proceed. In one version, the monocyte reporter line would be transduced with a panel of sgRNA lentiviruses for individual gene knockouts; an observed increase in reporter signal compared to in an sgRNA-less control transduction would validate a gene as a negative regulator. In another version, experiments could use CRISPRi knockdown, and/or be

carried out with primary human monocytes. In any case, it could be useful to examine prior knowledge such as transcriptomic profiling data that would indicate whether a gene is transcribed in the cell type of interest, prior to investigating the gene's role.

Another route to validation that could more directly address the TIC hypothesis is illustrated in **Figure 5.4**. Freestyle 293-F cells, which do not express TLR4 or MD-2, could be transduced with the reporter from **Figure 5.1a**, and following selection, lipofected with or without TLR4/MD-2 and with the sgRNA panel. After ~1 week, reporter signal would be assayed by flow cytometry. The expectation is that: for the sgRNA-less control or for the knockout of a gene that is not a regulator of basal NF- κ B activity, reporter expression would remain low regardless of whether TLR4/MD-2 is present; for knockout of a negative regulator that is not TIC, reporter expression would increase regardless of TLR4/MD-2; and for knockout of TIC, reporter expression would increase only if TLR4/MD-2 is present.

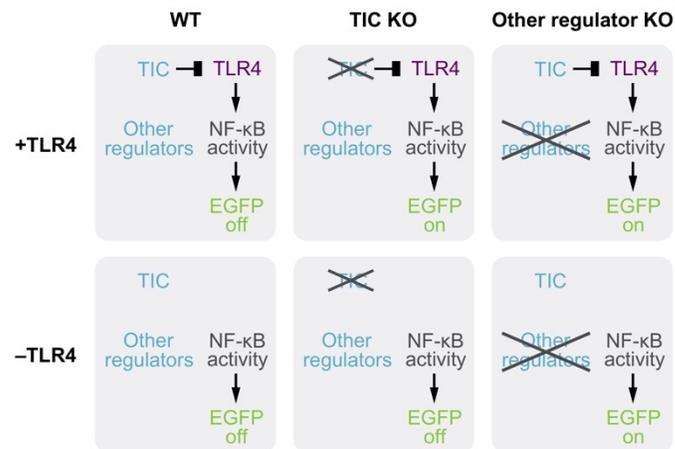


Figure 5.4. Validation of candidate regulators of NF- κ B activity. The cartoon depicts the proposed methodology for validating candidate regulators by arrayed knockout screening.

Below, genes from the set of 131 are categorized into different rationales for potential subsequent investigation. This first set includes non-membrane associated proteins in relation to annotated roles with respect to NF- κ B activity.

- Known negative regulators: STK11, CMIP, SIK3, RASSF1, UBA7, CCDC122, TRAFD1, TOLLIP, NISCH, TNFAIP3, PRKAA2

- Potential regulators: NLRP3, FYB, MAPKAPK5, APOD, NUDT4, MAPK11, HECTD1, USP45, YWHAB, JOSD1, LRRC71, BCL7C, SBK3
- Known regulators, and potentially a positive or negative role is known: AIRE, SH3BGR, MAP3K2
- Known positive regulators, which is interesting because the screen indicated a negative role: NRAS, SERPINB4, PTPN11, COPS8, GIMAP6, GUCY1A3

This second set includes membrane-associated (integral, peripheral, lipid-anchored) proteins.

- Known negative regulators: CYLD, RAMP1, CALCOCO2, RHOH, PIK3CG
- Potential regulators: PPAP2A, GNGT1, KCNMB2, TMEM210, GPR151, AQP2, BACE2, GPR84, TBC1D9, SLC1A2, CNR2, MR1, GPR182, COMT, ADGRE3, TMEM161A, SLC4A10, DRP2, HEPH, P2RY8, HTR6, TMEM185B, KCND3, SI, OR8G2, SLC44A2, PRND, SNAP23, RND1, RERGL, MAPKAP1, SLCO4A1, FRS3, LPPR2, VNN3, AQP1, PCDHGA4, SLC44A3, TVP23C, ZNRF2, SLC35F4, JPH3, PTPRD, HPCAL1, ANKRA2, ANTXR1, CMTM5, LRRTM3, CSPG5, OR10Q1, MCOLN2, GPR85, OPRM1, OR5C1, OR52A5, CDH7, TM9SF1, ADCY6, ABCA7, CRTAM, TMEM110, EMC4, DNAJC25-GNG10, NSMF, ROR2, GPC2, KCNN2, APLP1, GCGR, TRHDE, SLC24A2, OR52K1, RNFT2, NEURL1, OR14J1, EFNA3, RRNAD1, TSPAN14, EPB41L2, PLEKHA1, PLIN4, CLIC3, ABCC6, SLC22A18, XKRX
- Known regulators, and potentially a positive or negative role is known: CDCP1, KLRG1, PCP2
- Known positive regulators, which is interesting because the screen indicated a negative role: MST1R, IFNAR1, LRRC4

Lastly, it may be worthwhile to examine genes that are known negative regulators that did not appear in the list of 439 genes, such as NFKBIA, NFKBIB, NFKBIE, TNFAIP3.

5.6 Acknowledgements

I thank Suchitra Sankaranarayan and Matt Clutter for collaboration on an exciting and challenging project, Nichole Daringer, Kelly Sarnese, and Nikita Divekar for early experimental studies, and Josh Leonard for the mentorship and leadership in grant writing that enabled this investigation.

CHAPTER 6. Conclusions

This chapter highlights key themes, ongoing research, and potential future directions for the QL study, the set of studies on genetically programming mammalian cellular functions, and the TIC study.

Quorum licensing: Inflammation is the body's protective response to eliminate a perceived threat. When immune cells like macrophages sense an infection, they secrete factors that recruit other cells to clear the microbes. Cells need to tightly regulate the amplitude and duration of this response; hyporesponsiveness carries the risk of microbes persisting, leading to sepsis, and hyperresponsiveness carries toxic side effects, leading to chronic diseases in which the tissue does not return to homeostasis. Recently, it has become evident that this highly regulated response involves cell-to-cell heterogeneity in the production of cytokines. **Chapter 2** elucidates an example of this phenomenon in which cell density modulates the proportion of cells that exhibit a high TNF response to LPS through a mechanism we termed quorum licensing (**Figure 6.1**). Intriguingly, growing evidence suggests there are many factors that shape cell-to-cell variation. For example, desynchronization of the molecular clock contributes to heterogeneous responses to LPS; the expression of certain clock genes was found to influence the proportion of cells that produce high levels of IL-12p40 (though this mechanism did not affect TNF production)²⁴². There are accumulating reports of genes that are expressed bimodally upon immune cell activation, and it might be that there are many contributors that act in different ways on different cytokines, rather than a single overarching explanation.

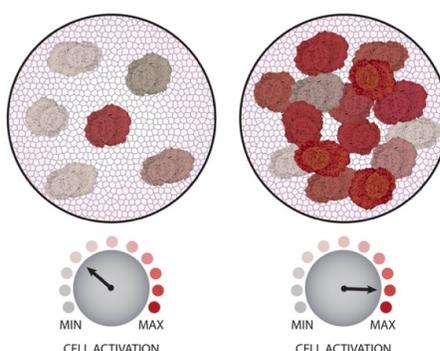


Figure 6.1. Quorum licensing. In this newly recognized form of immune coordination, macrophages sense how many of them are present, and then decide on how many should become highly activated.

A metaphor that I suggest for imagining the above scenario is that immune cells each have a dashboard for controlling the production of pathogen-responsive proteins. When one considers a standard diagram depicting the transduction of extracellular information about a pathogen across the cell membrane by receptors, through the cytoplasm by signaling proteins, and into the nucleus by transcription factors, what one is really doing is zooming in on one section of this dashboard. However, there are also other controls that exist, which may not be depicted in the diagram and might operate separately from canonical pathways. e.g., IL-12 and TNF are both LPS-inducibly expressed, yet there is also circadian control of IL-12 but apparently not TNF, and QL control of at least TNF. Furthermore, our ability to hypothesize as-of-yet undiscovered controls might not be well-supported by standard choices for visual representation. Indeed, the features of QL illustrated in **Figure 2.2h** and **Figure 6.1** are not readily incorporated using the pathway notation in **Figure 2.3a**. Going forward, it may be that some of the additional knobs and dials on the dashboard will continue to be illuminated through serendipitous observations. Alternatively, perhaps these controls will be more systematically uncovered by creative examination of new independent variables, technologies for integrating single-cell measurements of pathway activity with global cell profiles²⁴³, and the use of whole pathogens rather than purified ligands as stimuli²⁴⁴.

In considering the implications of the QL study, it is worth noting the densities of macrophages *in vivo* vary widely by tissue, an organism's health, and perturbations used in studies that report these values²⁴⁵⁻²⁴⁷. Additionally, cell culture density has been shown to affect the phenotype (e.g., expression of surface markers of differentiation) and function (e.g., secretion of cytokines like TNF) of bone marrow-derived macrophages (BMM)²⁴⁸. Although we used a standard protocol for BMM differentiation and cell culture densities that are within the general range of values reported for *in vivo* studies, we avoid making assumptions or comparisons on the expected effects between *in vitro* and *in vivo* densities. Rather, the main takeaway is that cell density has an effect that might provide functional utility by enabling cells to calibrate their responses to future threats.

Our observation that certain chemokines (KC most prominently, and perhaps also others that were not measured) trended with BMM cell density in the absence of stimulus suggests a potential avenue for future investigation into factor(s) that mediate QL. If it turns out that a single chemokine is responsible, then assays for TNF production from a panel of cells that have various chemokine knockout genotypes or from

cells that have undergone blockade pre-treatment such as with a panel of soluble recombinant receptors against each chemokine could be useful. Furthermore, cells could be pre-treated with recombinant chemokines individually to test whether any of these conditions shifts the proportion of high-TNF-producing cells. However, if multiple chemokines are involved, which seems plausible, then alternative strategies would be necessary. For example: assay at the RNA level or secreted protein level for many chemokines from non-stimulated cell cultures at different densities, identify the chemokines that exhibit a profile consistent with that of a QL mediator, and then take subsequent steps to narrow down this list. These steps could be used to identify the pathways that cells use to carry out the QL phenomenon.

Achieving engineering design goals: Chapters 3 and 4 and Appendices 3–6, 8, and 9 encompass several projects in which I developed new methods and tools to modulate gene expression and build new signaling pathways. There has been great progress on these fronts for organisms such as bacteria and yeast, and more recently for mammalian systems, but for a variety of applications, we are still in need of high-performing parts for programming cellular functions as well as quantitative approaches to guide how we choose parts and link them together to carry out operations that work as intended. It would be useful if we had the tools and the conceptual and formal models to enable us to engineer biological systems with a reliability more like that of electronic systems, or at least using a more formal design process.

Through several collaborations, I have focused on two areas in mammalian cellular engineering: sensing cues using receptors, and processing signals using transcription factors and enzymatic reactions. The impetus for the approaches that I developed comes from a study on multiplexing MESA receptors to sense two ligands and signal to a downstream promoter (**Appendix 4**)⁵⁴. We found that although a panel of engineered promoters exhibited AND gate behavior as intended when paired with intracellular TFs, it was difficult to level-match receptor signaling output with promoter input requirements to achieve AND gate behavior. I developed models to describe the activity of the promoters and receptors, and in considering how to train the models to experimental data, noted that the cells were highly varied in their activation levels; with the QL study fresh in mind, I considered that this variation might be important to understand. I decided to simulate not a single representative cell, but rather a population of cells that resembled the data. To account for cellular variation, I developed an algorithm that provides general way to model a population of cells (**Appendix 3**), by extending statistical features of gene expression distributions for any specified

number, dose, and type of regulation of genes in a circuit. This framework enabled an analysis of not only population-average outcomes, which is what is experimentally most accessible and can be compared, but also inferred representative single cells. This strategy is similar to the QL study, in that there is one dynamical model that describes a mechanism, and variation is introduced using a statistical distribution.

In applying this framework, I found that reporter expression for the population mean was consistently greater than that for a cell expressing the mean amount of each component. Thus, our perception of typical behavior is skewed (generally overestimated) by outlier cells. More broadly, any genetic circuit with nonlinear behaviors might be biased in this way, so a model that captures variation could improve the interpretation of those cases too. I did a subsequent analysis and identified effects that come about in linking parts together for the full system, and identified properties that we could change to improve performance, but overall it was clear that what we really needed was a different set of parts and descriptions to more effectively program cell behaviors. This conclusion supported and was motivation for subsequent studies in which my colleagues and I developed the COMET toolkit of tunable TFs and promoters for customizable gene expression¹²⁹, high-performing next-generation MESA receptors¹⁹⁸, and genetic parts and computational tools for predictively designing genetic programs¹⁶¹.

COMET: This toolkit benefits from a relatively large number of characterized TFs and promoters, a modular design that facilitates the addition of new parts, and a methodology for incorporating these parts into a concise modeling framework. It might be that the parameter values that were fitted in this study are specific to the DNA delivery method and cell type in which the characterizations were conducted. However, fundamental mechanisms of transcription are maintained across contexts, so it is possible that certain trends will extend across cell types and delivery methods, e.g., the general rank order of values for the parameter w , which describes the steepness of a dose response and is related to TF-DNA binding affinity.

COMET TFs confer dose response properties that differ from those of tTA and Gal4 (used in the multiplexing study in **Appendix 4**) and thus appear better suited for use with hybrid promoters (**Figures 3.5c, Figure 4.4c**) and synthetic receptors (**Figure 4.4d**). It is notable that the multiplexing study (**Appendix 4**), COMET study (**Chapter 3**), and genetic programs study (**Chapter 4**) each make use of single-layer transcriptional AND gates in which two TFs converge at a multi-site promoter, yet the outcomes were very different. Perhaps there is a difference in the stage in which cooperativity takes effect, such as TF-DNA

binding for the former study and post-binding for the latter two. However, the key factor could have been level-matching. The input requirements of the COMET hybrid promoters appear to be sufficiently level-matched to the signaling from sensors such as a small molecule-responsive TF and MESA receptor to enable an exceptional AND gate (**Figure 4.4c,d**) and potentially other operations. The improved outcome for the VP64-ZF2 AND VP64-ZF6 gate (**Figure 4.4c**) compared to the VP16-ZF2 AND VP16-ZF3 gate (**Figures 3.5c**) suggest that ZF3 is a non-ideal DBD. We also know from the COMET characterization that ZF2 and ZF6 have high m values with VP16 and should be more ideal than ZF3, however the switch in AD choice from VP16 to VP64 might have also been important in achieving the observed improvement. Overall, these outcomes suggest that the use of COMET hybrid promoters to carry out logical operations is more effective when using potent TFs.

The advances from this study have enabled several subsequent ongoing projects. In one project, the original model has been revamped to more formally described chemical responsiveness conferred by the use of small molecule-binding domains on transcriptional activators. Extensions may also be taken to characterize chemically responsive inhibitors, which represent a complementary approach for external and temporal control of gene expression. In another project, some of the principles from the original model are being applied to design cell-based devices for sensing hypoxic conditions and secreting therapeutic payloads within solid tumors. Here, because the genetic programs will be genomically integrated rather than transiently delivered, the framework with fitted b , m , and w parameters would not be used, and instead the utility of these values could be to guide the choice of potent TFs. However, we do know from the COMET study that the cooperativity boost from increasing the number of binding sites in a promoter does transfer to the genome, which bodes well for applications involving genomic integration, including those that modulate cooperativity through inhibitors utilizing the dual mechanism of inhibition.

Genetic programs: This investigation involved a synthesis of technologies and conceptual advances from multiple projects^{54,129,198} that generated new genetic parts and explanatory models, and for the first time demonstrated how these parts could be predictively combined to make cells carry out high-performing functions. These functions included digital logic, topologically compact multi-output logic, analog operations with chemical responsiveness, and interfacing of genetic programs with sensors to produce signaling cascades. Thus, in a sense, this project closed the loop on the multiplexing study. I suggest that

future directions could include continued development of receptors for new ligands (in particular, cytokines), testing circuits in a genomic context (with insertion mediated by either recombination, viral vector, or Cas9), and validating functional outputs such as secreted proteins with therapeutic roles. Additionally, it would be ideal if circuits were designed in such a way to bypass complications associated with heterogeneity, and on this topic a preliminary unpublished investigation of COMET in the genome bodes well.

The broad capability for accurate design-driven engineering represents a key advance in mammalian synthetic biology, a field that has grown substantially in recent years. With more designed genetic parts becoming available, and a growing understanding of how to effectively use these parts, applications spanning directed cell differentiation (such as into tissues and organoids), cell-based implants for detecting disease biomarkers or augmenting homeostasis, and cell-based immunotherapies are coming into reach^{163,164,249}. Principles from **Chapter 4** are enabling ongoing research directions including the design of cell-based devices that use split intein-based parts and feedback, with potential applications in closed-loop metabolic homeostasis and eventually treating metabolic disorders.

There are several benefits of using split intein-based parts and small molecule-reconstitutable parts. One benefit relates to shifting regulation from the transcriptional level to protein-protein interactions, which would be expected to increase the speed of a cell-based device's response (as protein-protein interactions are generally on a faster timescale than the interactions in gene expression) and could bypass the issue of gene silencing²⁵⁰. Another benefit relates to gene delivery. To accommodate the limited cargo capacities of viral vectors, one could distribute a large genetic program across multiple vectors in such a way that its function is reconstituted only in cells that receive all of the vectors. A last benefit is that split parts are inherently conducive to encoding logical operations. As an example that incorporates the above ideas, the following hypothetical genetic program would be expected to carry out a four-input AND gate, in which one of the inputs is externally supplied ABA: genes for VP64-PYL1, ABI1-intN, and intC-ZF2 are each regulated by a respective condition-specific promoter, and an output transgene is regulated by a ZF2x12-C promoter. Alternatively, a synonymous design could use ZF2-PYL1, ABI1-intN, and intC-VP64 as the inducible genes, which again highlights the flexibility demonstrated in the genetic programs study for proposing a tractable panel of candidate topologies for achieving a design goal.

TIC: If TIC is identified through the validation experiments described at the end of **Chapter 5**, then the next step would be to characterize TIC. It would be useful to perform a clonal all-allele knockout of TIC and evaluate its role in signaling through both arms of the TLR4 pathway (MyD88 and TRIF). One option would be to use cell lines with and without TIC that induce the expression of one reporter in response to NF- κ B activation (which is downstream of both MyD88 and TRIF) and another reporter in response to Interferon Regulatory Factor (IRF) activation (which is TRIF-specific). An additional line of investigation would be to use RNA-Seq to characterize whether and how the role of TIC modulates downstream gene expression with and without LPS treatment.

It would also be useful to investigate whether TIC interacts physically with TLR4, such as through co-immunoprecipitation assays. HA-tagged TLR4 and 3xFLAG-tagged TIC could be co-expressed following transient transfection in HEK293FT cells. TLR4 would be precipitated from cell lysates using anti-HA-antibody-conjugated beads, washed, and denatured, and the released precipitate could be analyzed by anti-FLAG Western blot. The reciprocal assay could also be conducted: precipitating by anti-FLAG-antibody-conjugated beads and detecting by anti-HA Western blot. The expectation is that TIC and TLR4 would co-precipitate without LPS treatment. The specific way in which the physical interaction would be disrupted with LPS treatment does not have a pre-defined expectation, because in the proposed model, LPS treatment need not necessarily disrupt all physical interactions. Lastly, TIC and TLR4 could be fluorescently tagged, and microscopy could be used to evaluate their co-localization in cells with and without LPS treatment. These steps would elucidate a putatively important layer of immune regulation.

Concluding thoughts: With rapidly advancing technologies to interrogate and engineer cells, in combination with new models and algorithms to extract meaning from datasets, there is tremendous potential to improve our understanding of cell signaling and ultimately to improve human health. In graduate school, I have developed rigorous ways to characterize gene regulation as well as genetic parts and principles to design new functions in mammalian cells. Throughout, I have been able to untangle complex and non-intuitive problems by developing a skillset that bridges experimental and computational approaches. In this journey, I have had the opportunity to pursue projects working with excellent mentors, collaborators, and mentees, to whom I am very grateful.

REFERENCES

- 1 Muldoon, J. J., Donahue, P. S., Dolberg, T. B. & Leonard, J. N. Building with intent: technologies and principles for engineering mammalian cell-based therapies to sense and respond. *Curr Opin Biomed Eng* **4**, 127–133 (2017).
- 2 Box, G. E. P. & Draper, N. R. *Empirical model-building and response surfaces*. (John Wiley & Sons, 1987).
- 3 Naisbitt, J. *Megatrends: ten new directions transforming our lives*. (Warner Books, 1982).
- 4 Eigen, M. *The physicist's conception of nature*. 594–632 (1973).
- 5 Brewster, B. in *The Yale Literary Magazine* Vol. 47 (Tuttle, Morehouse & Taylor, 1882).
- 6 Janes, K. A. *et al.* Cue-signal-response analysis of TNF-induced apoptosis by partial least squares regression of dynamic multivariate data. *J Comput Biol* **11**, 544–561 (2004).
- 7 Aldridge, B. B., Burke, J. M., Lauffenburger, D. A. & Sorger, P. K. Physicochemical modelling of cell signalling pathways. *Nat Cell Biol* **8**, 1195–1203 (2006).
- 8 Le Novère, N. Quantitative and logic modelling of molecular and gene networks. *Nat Rev Genet* **16**, 146–158 (2015).
- 9 Deb, K., Pratap, A., Agarwal, S. & Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans Evol Comp* **6**, 182–197 (2002).
- 10 Deb, K. *Multi-objective optimization*. 403–449 (Springer, 2014).
- 11 van Riel, N. A. W. Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments. *Brief Bioinform* **7**, 364–374 (2006).
- 12 Ashyraliyev, M., Fomekong-Nanfack, Y., Kaandorp, J. A. & Blom, J. G. Systems biology: parameter estimation for biochemical models. *FEBS J* **276**, 886–902 (2009).
- 13 Sun, J., Garibaldi, J. M. & Hodgman, C. Parameter estimation using metaheuristics in systems biology: a comprehensive review. *IEEE/ACM Trans Comput Biol Bioinform* **9**, 185–202 (2012).
- 14 Gutenkunst, R. N. *et al.* Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol* **3**, 1871–1878 (2007).
- 15 Tönsing, C., Timmer, J. & Kreutz, C. Cause and cure of sloppiness in ordinary differential equation models. *Phys Rev E* **90**, 023303 (2014).
- 16 Zak, D. E., Tam, V. C. & Aderem, A. Systems-level analysis of innate immunity. *Annu Rev Immunol* **32**, 547–577 (2014).
- 17 Geissmann, F., Gordon, S., Hume, D. A., Mowat, A. M. & Randolph, G. J. Unravelling mononuclear phagocyte heterogeneity. *Nat Rev Immunol* **10**, 453–460 (2010).
- 18 Lawrence, T. & Natoli, G. Transcriptional regulation of macrophage polarization: enabling diversity with identity. *Nat Rev Immunol* **11**, 750–761 (2011).

- 19 Xue, J. *et al.* Transcriptome-based network analysis reveals a spectrum model of human macrophage activation. *Immunity* **40**, 274–288 (2014).
- 20 Shalek, A. K. *et al.* Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498**, 236–240 (2013).
- 21 Jaitin, D. A. *et al.* Massively parallel single-cell RNA-Seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776–779 (2014).
- 22 Shalek, A. K. *et al.* Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **510**, 363–369 (2014).
- 23 Lu, Y. *et al.* Highly multiplexed profiling of single-cell effector functions reveals deep functional heterogeneity in response to pathogenic ligands. *Proc Natl Acad Sci USA* **112**, E607–E615 (2015).
- 24 Han, Q. *et al.* Polyfunctional responses by human T cells result from sequential release of cytokines. *Proc Natl Acad Sci USA* **109**, 1607–1612 (2012).
- 25 Raser, J. M. & O'Shea, E. K. Noise in gene expression: origins, consequences, and control. *Science* **309**, 2010–2013 (2005).
- 26 Satija, R. & Shalek, A. K. Heterogeneity in immune responses: from populations to single cells. *Trends Immunol* **35**, 219–229 (2014).
- 27 Jayaraman, A. & Wood, T. K. Bacterial quorum sensing: signals, circuits, and implications for biofilms and disease. *Annu Rev Biomed Eng* **10**, 145–167 (2008).
- 28 Antonioli, L., Blandizzi, C., Pacher, P., Guillemins, M. & Haskó, G. Rethinking communication in the immune system: the quorum sensing concept. *Trends Immunol* **40**, 88–97 (2019).
- 29 Postat, J. & Bousso, P. Quorum sensing by monocyte-derived populations. *Front Immunol* **10**, 2140 (2019).
- 30 Xue, Q. *et al.* Analysis of single-cell cytokine secretion reveals a role for paracrine signaling in coordinating macrophage responses to TLR4 stimulation. *Sci Signal* **8**, ra59 (2015).
- 31 Caldwell, A. B., Cheng, Z., Vargas, J. D., Birnbaum, H. A. & Hoffmann, A. Network dynamics determine the autocrine and paracrine signaling functions of TNF. *Genes Dev* **28**, 2120–2133 (2014).
- 32 Chen, C.-C. *et al.* Organ-level quorum sensing directs regeneration in hair stem cell populations. *Cell* **161**, 277–290 (2015).
- 33 Postat, J., Olekhovitch, R., Lemaître, F. & Bousso, P. A metabolism-based quorum sensing mechanism contributes to termination of inflammatory responses. *Immunity* **49**, 654–665 (2018).
- 34 Holzinger, A., Barden, M. & Abken, H. The growing world of CAR T cell trials: a systematic review. *Cancer Immunol Immunother* **65**, 1433–1450 (2016).
- 35 Kitano, H. Biological robustness. *Nat Rev Genet* **5**, 826–837 (2004).
- 36 Mehta, P., Lang, A. H. & Schwab, D. J. Landauer in the age of synthetic biology: energy consumption and information processing in biochemical networks. *J Stat Phys* **162**, 1153–1166 (2016).

- 37 Di Stasi, A. *et al.* Inducible apoptosis as a safety switch for adoptive cell therapy. *N Engl J Med* **365**, 1673–1683 (2011).
- 38 Wu, C. Y., Roybal, K. T., Puchner, E. M., Onuffer, J. & Lim, W. A. Remote control of therapeutic T cells through a small molecule-gated chimeric receptor. *Science* **350**, aab4077 (2015).
- 39 Zhang, E. & Xu, H. A new insight in chimeric antigen receptor-engineered T cells for cancer immunotherapy. *J Hematol Oncol* **10** (2017).
- 40 Shi, H., Liu, L. & Wang, Z. Improving the efficacy and safety of engineered T cell therapy for cancer. *Cancer Lett* **328**, 191–197 (2013).
- 41 Kirchhofer, A. *et al.* Modulation of protein properties in living cells using nanobodies. *Nat Struct Mol Biol* **17**, 133–138 (2010).
- 42 Kelley, S. O. What are clinically relevant levels of cellular and biomolecular analytes? *ACS Sens* **2**, 193–197 (2017).
- 43 Suerth, J. D., Schambach, A. & Baum, C. Genetic modification of lymphocytes by retrovirus-based vectors. *Curr Opin Immunol* **24**, 598–608 (2012).
- 44 Singh, H., Huls, H., Kebriaei, P. & Cooper, L. J. N. A new approach to gene therapy using Sleeping Beauty to genetically modify clinical-grade T cells to target CD19. *Immunol Rev* **257**, 181–190 (2013).
- 45 Duportet, X. *et al.* A platform for rapid prototyping of synthetic gene networks in mammalian cells. *Nucleic Acids Res* **44**, 2677–2690 (2014).
- 46 Papapetrou, E. P. & Schambach, A. Gene insertion into genomic safe harbors for human gene therapy. *Mol Ther* **24**, 678–684 (2016).
- 47 Chu, V. T. *et al.* Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat Biotechnol* **33**, 543–548 (2015).
- 48 Martella, A., Pollard, S. M., Dai, J. & Cai, Y. Mammalian synthetic biology: time for Big MACs. *ACS Synth Biol* **5**, 1040–1049 (2016).
- 49 Wroblewska, L. *et al.* Mammalian synthetic circuits with RNA binding proteins for RNA-only delivery. *Nat Biotechnol* **33**, 839–841 (2015).
- 50 Gough, A. *et al.* Biologically relevant heterogeneity: metrics and practical insights. *SLAS Discov* **22**, 213–237 (2017).
- 51 Loewer, A. & Lahav, G. We are all individuals: causes and consequences of non-genetic heterogeneity in mammalian cells. *Curr Opin Genet Dev* **21**, 753–758 (2011).
- 52 Morel, M., Shtrahman, R., Rotter, V., Nissim, L. & Bar-Ziv, R. H. Cellular heterogeneity mediates inherent sensitivity–specificity tradeoff in cancer targeting by synthetic circuits. *Proc Natl Acad Sci USA* **113**, 8133–8138 (2016).
- 53 Beal, J. Biochemical complexity drives log-normal variation in genetic expression. *Eng Biol* **1**, 55–60 (2017).

- 54 Hartfield, R. M., Schwarz, K. A., Muldoon, J. J., Bagheri, N. & Leonard, J. N. Multiplexing engineered receptors for multiparametric evaluation of environmental ligands. *ACS Synth Biol* **6**, 2042–2055 (2017).
- 55 Lin-Gibson, S., Sarkar, S., Elliott, J. & Plant, A. Understanding and managing sources of variability in cell measurements. *Cell & Gene Therapy Insights* **2**, 663–673 (2016).
- 56 Heathman, T. R. J. *et al.* The translation of cell-based therapies: clinical landscape and manufacturing challenges. *Regen Med* **10**, 49–64 (2015).
- 57 Helikar, T., Kochi, N., Konvalina, J. & Rogers, J. A. in *Systems biology for signaling networks* Ch. 12, 295–336 (Springer-Verlag, 2010).
- 58 Youk, H. & Lim, W. A. Secreting and sensing the same molecule allows cells to achieve versatile social behaviors. *Science* **343**, 1242782 (2014).
- 59 Roybal, K. T. *et al.* Engineering T cells with customized therapeutic response programs using synthetic Notch receptors. *Cell* **167**, 419–432 (2016).
- 60 Wang, W.-D., Chen, Z.-T., Kang, B.-G. & Li, R. Construction of an artificial intercellular communication network using the nitric oxide signaling elements in mammalian cells. *Exp Cell Res* **314**, 699–706 (2007).
- 61 Bacchus, W. *et al.* Synthetic two-way communication between mammalian cells. *Nat Biotechnol* **30**, 991–996 (2012).
- 62 Muldoon, J. J., Chuang, Y., Bagheri, N. & Leonard, J. N. Macrophages employ quorum licensing to regulate collective activation. *Nat Commun* **11**, 878 (2020).
- 63 Perkins, T. J. & Swain, P. S. Strategies for cellular decision-making. *Mol Syst Biol* **5** (2009).
- 64 Liu, G. & Yang, H. Modulation of macrophage activation and programming in immunity. *J Cell Physiol* **228**, 502–512 (2012).
- 65 Hoffmann, A., Levchenko, A., Scott, M. L. & Baltimore, D. The I κ B–NF- κ B signaling module: Temporal control and selective gene activation. *Science* **298**, 1241–1245 (2002).
- 66 Lipniacki, T., Paszek, P., Brasier, A. R., Luxon, B. & Kimmel, M. Mathematical model of NF- κ B regulatory module. *J Theor Biol* **228**, 195–215 (2004).
- 67 Xaus, J. *et al.* LPS induces apoptosis in macrophages mostly through the autocrine production of TNF- α . *Blood* **95**, 3823–3831 (2000).
- 68 Vandenbon, A., Teraguchi, S., Akira, S., Takeda, K. & Standley, D. M. Systems biology approaches to toll-like receptor signaling. *WIREs Syst Biol Med* **4**, 497–507 (2012).
- 69 Werner, S. L., Barken, D. & Hoffmann, A. Stimulus specificity of gene expression programs determined by temporal control of IKK activity. *Science* **309**, 1857–1861 (2005).
- 70 Purvis, J. E. & Lahav, G. Encoding and decoding cellular information through signaling dynamics. *Cell* **152**, 945–956 (2013).
- 71 Covert, M. W., Leung, T. H., Gaston, J. E. & Baltimore, D. Achieving stability of lipopolysaccharide-induced NF- κ B activation. *Science* **309**, 1854–1857 (2005).

- 72 Parameswaran, N. & Patial, S. Tumor necrosis factor- α signaling in macrophages. *Crit Rev Eukaryot Gene Expr* **20**, 87–103 (2010).
- 73 Sung, M.-H. *et al.* Switching of the relative dominance between feedback mechanisms in lipopolysaccharide-induced NF- κ B signaling. *Sci Signal* **7**, ra6 (2014).
- 74 Murray, P. J. & Smale, S. T. Restraint of inflammatory signaling by interdependent strata of negative regulatory pathways. *Nat Immunol* **13**, 916–924 (2012).
- 75 Kanarek, N., London, N., Schueler-Furman, O. & Ben-Neriah, Y. Ubiquitination and degradation of the inhibitors of NF- κ B. *Cold Spring Harb Perspect Biol* **2** (2010).
- 76 Vereecke, L., Beyaert, R. & van Loo, G. The ubiquitin-editing enzyme A20 (TNFAIP3) is a central regulator of immunopathology. *Trends Immunol* **30**, 383–391 (2009).
- 77 Hu, X., Chen, J., Wang, L. & Ivashkiv, L. B. Crosstalk among Jak-STAT, Toll-like receptor, and ITAM-dependent pathways in macrophage activation. *J Leukoc Biol* **82**, 237–243 (2007).
- 78 Joo, J., Plimpton, S., Martin, S., Swiler, L. & Faulon, J.-L. Sensitivity analysis of a computational model of the IKK–NF- κ B–I κ B–A20 signal transduction network. *Ann NY Acad Sci* **1115**, 221–239 (2007).
- 79 Basak, S., Behar, M. & Hoffmann, A. Lessons from mathematically modeling the NF- κ B pathway. *Immunol Rev* **246**, 221–238 (2012).
- 80 Williams, R. A., Timmis, J. & Qvarnstrom, E. E. Computational models of the NF- κ B signalling pathway. *Computation* **2**, 131–158 (2014).
- 81 Cheong, R., Hoffmann, A. & Levchenko, A. Understanding NF- κ B signaling via mathematical modeling. *Mol Syst Biol* **4** (2008).
- 82 Sharp, G. C., Ma, H., Saunders, P. T. K. & Norman, J. E. A computational model of lipopolysaccharide-induced nuclear factor kappa B activation: a key signalling pathway in infection-induced preterm labour. *PLoS One* **8**, e70180 (2013).
- 83 Pękaliski, J. *et al.* Spontaneous NF- κ B activation by autocrine TNF α signaling: a computational analysis. *PLoS One* **8**, e78887 (2013).
- 84 Maiti, S., Dai, W., Alaniz, R. C., Hahn, J. & Jayaraman, A. Mathematical modeling of pro- and anti-inflammatory signaling in macrophages. *Processes* **3**, 1–18 (2015).
- 85 Lipniacki, T., Paszek, P., Brasier, A. R., Luxon, B. A. & Kimmel, M. Stochastic regulation in early immune response. *Biophys J* **90**, 725–742 (2006).
- 86 Tay, S. *et al.* Single-cell NF- κ B dynamics reveal digital activation and analogue information processing. *Nature* **466**, 267–271 (2010).
- 87 Hughey, J. J., Gutschow, M. V., Bajar, B. T. & Covert, M. W. Single-cell variation leads to population invariance in NF- κ B signaling dynamics. *Mol Biol Cell* **26**, 583–590 (2015).
- 88 Lee, R. E. C., Walker, S. R., Savery, K., Frank, D. A. & Gaudet, S. Fold change of nuclear NF- κ B determines TNF-induced transcription in single cells. *Mol Cell* **53**, 867–879 (2014).
- 89 Hasenauer, J. *et al.* Identification of models of heterogeneous cell populations from population snapshot data. *BMC Bioinform* **12**, 1–15, doi:10.1186/1471-2105-12-125 (2011).

- 90 Kalita, M. K. *et al.* Sources of cell-to-cell variability in canonical Nuclear Factor- κ B (NF- κ B) signaling pathway inferred from single cell dynamic images. *J Biol Chem* **286**, 37741–37757 (2011).
- 91 Cheng, Z., Taylor, B., Ourthiague, D. R. & Hoffmann, A. Distinct single-cell signaling characteristics are conferred by the MyD88 and TRIF pathways during TLR4 activation. *Sci Signal* **8**, ra69 (2015).
- 92 Paszek, P. *et al.* Population robustness arising from cellular heterogeneity. *Proc Natl Acad Sci USA* **107**, 11644–11649 (2010).
- 93 Ravasi, T. *et al.* Generation of diversity in the innate immune system: macrophage heterogeneity arises from gene-autonomous transcriptional probability of individual inducible genes. *J Immunol* **168**, 44–50 (2002).
- 94 Altschuler, S. J. & Wu, L. F. Cellular heterogeneity: do differences make a difference? *Cell* **141**, 559–563 (2010).
- 95 Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671–675 (2012).
- 96 Fraser, I. *et al.* in *Protein Phosphatase Protocols* Vol. 365 *Methods in Molecular Biology* (ed G. Moorhead) Ch. 20, (Humana Press, 2007).
- 97 Zhang, X., Goncalves, R. & Mosser, D. M. in *Curr Protocol Immunol* Vol. 83 Ch. 14, (John Wiley & Sons, 2008).
- 98 Merzlyak, E. M. *et al.* Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nat Methods* **4**, 555–557 (2007).
- 99 Rajasingh, J. *et al.* IL-10-induced TNF- α mRNA destabilization is mediated via IL-10 suppression of p38 MAP kinase activation and inhibition of HuR expression. *FASEB J* **20**, E1393–E1403 (2006).
- 100 Brooks, S. A. & Blackshear, P. J. Tristetraprolin (TTP): interactions with mRNA and proteins, and current thoughts on mechanisms of action. *Biochim Biophys Acta* **1829**, 666–679 (2013).
- 101 Carpenter, S., Ricci, E. P., Mercier, B. C., Moore, M. J. & Fitzgerald, K. A. Post-transcriptional regulation of gene expression in innate immunity. *Nat Rev Immunol* **14**, 361–376 (2014).
- 102 Chuang, Y., Hung, M. E., Cangolese, B. K. & Leonard, J. N. Regulation of the IL-10-driven macrophage phenotype under incoherent stimuli. *Innate Immun* **22**, 647–657 (2016).
- 103 Palumbo, M. C. *et al.* Collective behavior in gene regulation: post-transcriptional regulation and the temporal compartmentalization of cellular cycles. *FEBS J* **275**, 2364–2371 (2008).
- 104 Keogh, B. & Parker, A. E. Toll-like receptors as targets for immune disorders. *Trends Pharm Sci* **32**, 435–443 (2011).
- 105 Gaba, A. *et al.* IL-10-mediated tristetraprolin induction is part of a feedback loop that controls macrophage STAT3 activation and cytokine production. *J Immunol* **189**, 2089–2093 (2012).
- 106 Gais, P. *et al.* TRIF signaling stimulates translation of TNF- α mRNA via prolonged activation of MK2. *J Immunol* **184**, 5842–5848 (2010).
- 107 Kafasla, P., Skliris, A. & Kontoyiannis, D. L. Post-transcriptional coordination of immunological responses by RNA-binding proteins. *Nat Immunol* **15**, 492–502 (2014).

- 108 Bonizzi, G. & Karin, M. The two NF- κ B activation pathways and their role in innate and adaptive immunity. *Trends Immunol* **25**, 280–288 (2004).
- 109 Bode, J. G., Ehltling, C. & Häussinger, D. The macrophage response towards LPS and its control through the p38MAPK–STAT3 axis. *Cell Signal* **24**, 1185–1194 (2012).
- 110 Radulescu, O., Gorban, A. N., Zinovyev, A. & Lilienbaum, A. Robust simplifications of multiscale biochemical networks. *BMC Syst Biol* **2** (2008).
- 111 Raposo, G. & Stoorvogel, W. Extracellular vesicles: exosomes, microvesicles, and friends. *J Cell Biol* **200**, 373–383 (2013).
- 112 Birtwistle, M. R. *et al.* Emergence of bimodal cell population responses from the interplay between analog single-cell signaling and protein expression noise. *BMC Syst Biol* **6** (2012).
- 113 To, T.-L. & Maheshri, N. Noise can induce bimodality in positive transcriptional feedback loops without bistability. *Science* **327**, 1142–1145 (2010).
- 114 Busse, D. *et al.* Competing feedback loops shape IL-2 signaling between helper and regulatory T lymphocytes in cellular microenvironments. *Proc Natl Acad Sci USA* **107**, 3058–3063 (2010).
- 115 Bagnall, J. *et al.* Quantitative analysis of competitive cytokine signaling predicts tissue thresholds for the propagation of macrophage activation. *Sci Signal* **11**, eaaf3998 (2018).
- 116 Bradley, J. R. TNF-mediated inflammatory disease. *J Pathol* **214**, 149–160 (2008).
- 117 Peri, F. & Piazza, M. Therapeutic targeting of innate immunity with Toll-like receptor 4 (TLR4) antagonists. *Biotechnol Adv* **30**, 251–260 (2012).
- 118 VanOtteren, G. M. *et al.* Compartmentalized expression of RANTES in a murine model of endotoxemia. *J Immunol* **154**, 1900–1908 (1995).
- 119 Standiford, T. J., Strieter, R. M., Lukacs, N. W. & Kunkel, S. L. Neutralization of IL-10 increases lethality in endotoxemia. *J Immunol* **155**, 2222–2229 (1995).
- 120 Kurihara, T., Warr, G., Loy, J. & Bravo, R. Defects in macrophage recruitment and host defense in mice lacking the CCR2 chemokine receptor. *J Exp Med* **186**, 1757–1762 (1997).
- 121 De Filippo, K. *et al.* Mast cell and macrophage chemokines CXCL1/CXCL2 control the early stage of neutrophil recruitment during tissue inflammation. *Blood* **121**, 4930–4937 (2013).
- 122 Wang, D. & Richmond, A. Nuclear factor- κ B activation by the CXC chemokine melanoma growth-stimulatory activity/growth-regulated protein involves the MEKK1/p38 mitogen-activated protein kinase pathway. *J Biol Chem* **276**, 3650–3659 (2001).
- 123 Yang, J. & Richmond, A. Constitutive I κ B kinase activity correlates with Nuclear Factor- κ B activation in human melanoma cells. *Cancer Res* **61**, 4901–4909 (2001).
- 124 Richmond, A. NF- κ B, chemokine gene transcription and tumour growth. *Nat Rev Immunol* **2**, 664–674 (2002).
- 125 Ciesielski, C. J., Andreakos, E., Foxwell, B. M. J. & Feldmann, M. TNF α -induced macrophage chemokine secretion is more dependent on NF- κ B expression than lipopolysaccharides-induced macrophage chemokine secretion. *Eur J Immunol* **32**, 2037–2045 (2002).

- 126 Chen, G. Y. & Nuñez, G. Sterile inflammation: sensing and reacting to damage. *Nat Rev Immunol* **10**, 826–837 (2010).
- 127 Laskin, D. L. & Pendino, K. J. Macrophages and inflammatory mediators in tissue injury. *Annu Rev Pharmacol Toxicol* **35**, 655–677 (1995).
- 128 Moore, K. J., Sheedy, F. J. & Fisher, E. A. Macrophages in atherosclerosis: a dynamic balance. *Nat Rev Immunol* **13**, 709–721 (2013).
- 129 Donahue, P. S. *et al.* The COMET toolkit for composing customizable genetic programs in mammalian cells. *Nat Commun* **11**, 779 (2020).
- 130 Bashor, C. J., Horwitz, A. A., Peisajovich, S. G. & Lim, W. A. Rewiring cells: synthetic biology as a tool to interrogate the organizational principles of living systems. *Annu Rev Biophys* **39**, 515–537 (2010).
- 131 Lim, W. A. Designing customized cell signalling circuits. *Nat Rev Mol Cell Biol* **11**, 393–403 (2010).
- 132 Gossen, M. & Bujard, H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci USA* **89**, 5547–5551 (1992).
- 133 Gossen, M. *et al.* Transcriptional activation by tetracyclines in mammalian cells. *Science* **268**, 1766–1769 (1995).
- 134 Kakidani, H. & Ptashne, M. GAL4 activates gene expression in mammalian cells. *Cell* **52**, 161–167 (1988).
- 135 Boch, J. *et al.* Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509–1512 (2009).
- 136 Moscou, M. J. & Bogdanove, A. J. A simple cipher governs DNA recognition by TAL effectors. *Science* **326**, 1501 (2009).
- 137 Li, Y. *et al.* Modular construction of mammalian gene circuits using TALE transcriptional repressors. *Nat Chem Biol* **11**, 207–213 (2015).
- 138 Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).
- 139 Stanton, B. C. *et al.* Systematic transfer of prokaryotic sensors and circuits to mammalian cells. *ACS Synth Biol* **3**, 880–891 (2014).
- 140 Hurt, J. A., Thibodeau, S. A., Hirsh, A. S., Pabo, C. O. & Joung, J. K. Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. *Proc Natl Acad Sci USA* **100**, 12271–12276 (2003).
- 141 Lohmueller, J. J., Armel, T. Z. & Silver, P. A. A tunable zinc finger-based framework for Boolean logic computation in mammalian cells. *Nucleic Acids Res* **40**, 5180–5187 (2012).
- 142 Maeder, M. L., *et al.* Rapid “open-source” engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol Cell* **31**, 294–301 (2008).
- 143 Khalil, A. S. *et al.* A synthetic biology framework for programming eukaryotic transcription functions. *Cell* **150**, 647–658 (2012).

- 144 Ptashne, M. How eukaryotic transcriptional activators work. *Nature* **335**, 683–689 (1988).
- 145 Hansen, J. *et al.* Transplantation of prokaryotic two-component signaling pathways into mammalian cells. *Proc Natl Acad Sci USA* **111**, 15705–15710 (2014).
- 146 Ede, C., Chen, X., Lin, M. Y. & Chen, Y. Y. Quantitative analyses of core promoters enable precise engineering of regulated gene expression in mammalian cells. *ACS Synth Biol* **5**, 395–404 (2016).
- 147 Alon, U. *An introduction to systems biology: design principles of biological circuits.* (Chapman and Hall/CRC, 2006).
- 148 Elrod-Erickson, M., Benson, T. E. & Pabo, C. O. High-resolution structures of variant Zif268-DNA complexes: implications for understanding zinc finger-DNA recognition. *Structure* **6**, 451–464 (1998).
- 149 Pavletich, N. P. & Pabo, C. O. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* **252**, 809–817 (1991).
- 150 Beerli, R. R., Segal, D. J., Dreier, B. & Barbas, C. F., 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 USA* **95**, 14628–14633 (1998).
- 151 Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nat Methods* **12**, 326–328 (2015).
- 152 Angelici, B., Mailand, E., Haefliger, B. & Benenson, Y. Synthetic biology platform for sensing and integrating endogenous transcriptional inputs in mammalian cells. *Cell Rep* **16**, 2525–2537 (2016).
- 153 Nielsen, A. A. K. *et al.* Genetic circuit design automation. *Science* **352**, aac7341 (2016).
- 154 Spitz, F. & Furlong, E. E. M. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* **13**, 613–626 (2012).
- 155 Morgunova, E. & Taipale, J. Structural perspective of cooperative transcription factor binding. *Curr Opin Struct Biol* **47**, 1–8 (2017).
- 156 Giniger, E. & Ptashne, M. Cooperative DNA binding of the yeast transcriptional activator GAL4. *Proc Natl Acad Sci USA* **85**, 382–386 (1988).
- 157 Stubbs, L., Sun, Y. & Caetano-Anolles, D. Function and evolution of C2H2 zinc finger arrays. *Subcell Biochem* **52**, 75–94 (2011).
- 158 Tsai, S. Y., Tsai, M.-J. & O'Malley, B. W. Cooperative binding of steroid hormone receptors contributes to transcriptional synergism at target enhancer elements. *Cell* **57**, 443–448 (1989).
- 159 Predki, P. F. & Sarkar, B. Cooperative interaction of oestrogen receptor 'zinc finger' domain polypeptides on DNA binding. *Biochem J* **205**, 805–810 (1995).
- 160 Bashor, C. J. *et al.* Complex signal processing in synthetic gene circuits using cooperative regulatory assemblies. *Science* **364**, 593–597 (2019).
- 161 Muldoon, J. J. *et al.* Design-driven engineering of mammalian genetic programs. (In preparation).
- 162 Fischbach, M. A., Bluestone, J. A. & Lim, W. A. Cell-based therapeutics: the next pillar of medicine. *Sci Trans Med* **5**, 179ps177 (2013).

- 163 Kitada, T., DiAndreth, B., Teague, B. & Weiss, R. Programming gene and engineered-cell therapies with synthetic biology. *Science* **359**, eaad1067 (2018).
- 164 Xie, M. & Fussenegger, M. Designing cell function: assembly of synthetic gene circuits for cell biology applications. *Nat Rev Mol Cell Biol* **19**, 507–525 (2018).
- 165 Slusarczyk, A. L., Lin, A. & Weiss, R. Foundations for the design and implementation of synthetic genetic circuits. *Nat Rev Genet* **13**, 406–420 (2012).
- 166 Lienert, F., Lohmueller, J. J., Garg, A. & Silver, P. A. Synthetic biology in mammalian cells: next generation research tools and therapeutics. *Nat Rev Mol Cell Biol* **15**, 95–107 (2014).
- 167 Daringer, N. M., Dudek, R. M., Schwarz, K. A. & Leonard, J. N. Modular extracellular sensor architecture for engineering mammalian cell-based devices. *ACS Synth Biol* **3**, 892–902 (2014).
- 168 Hermann, M. *et al.* Binary recombinase systems for high-resolution conditional mutagenesis. *Nucleic Acids Res* **42**, 3894–3907 (2014).
- 169 Park, S.-Y. *et al.* Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. *Science* **324**, 1068–1071 (2009).
- 170 Ma, Y. *et al.* Regulators of PP2C phosphatase activity function as abscisic acid sensors. *Science* **324**, 1064–1068 (2009).
- 171 Melcher, K. *et al.* A gate–latch–lock mechanism for hormone signalling by abscisic acid receptors. *Nature* **463**, 602–608 (2009).
- 172 Miyazono, K.-i. *et al.* Structural basis of abscisic acid signalling. *Nature* **462**, 609–614 (2009).
- 173 Yin, P. *et al.* Structural insights into the mechanism of abscisic acid signaling by PYL proteins. *Nat Struct Mol Biol* **16**, 1230–1236 (2009).
- 174 Gao, Y. *et al.* Complex transcriptional modulation with orthogonal and inducible dCas9 regulators. *Nat Methods* **13**, 1043–1049 (2016).
- 175 Loetscher, P., Pratt, G. & Rechsteiner, M. The C terminus of mouse ornithine decarboxylase confers rapid degradation on dihydrofolate reductase. *J Biol Chem* **17**, 11213–11220 (1991).
- 176 Bintu, L. *et al.* Dynamics of epigenetic regulation at the single-cell level. *Science* **351**, 720–724 (2016).
- 177 Vila-Perelló, M. & Muir, T. W. Biological applications of protein splicing. *Cell* **143**, 191–200 (2010).
- 178 Carvajal-Vallejos, P., Pallissé, R., Mootz, H. D. & Schmidt, S. R. Unprecedented rates and efficiencies revealed for new natural split inteins from metagenomic sources. *J Biol Chem* **287**, 28686–28696 (2012).
- 179 Baird, G. S., Zacharias, D. A. & Tsien, R. Y. Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. *Proc Natl Acad Sci USA* **97**, 11984–11989 (2000).
- 180 Lienert, F. *et al.* Two- and three-input TALE-based AND logic computation in embryonic stem cells. *Nucleic Acids Res* **41**, 9967–9975 (2013).
- 181 Gaber, R. *et al.* Designable DNA-binding domains enable construction of logic circuits in mammalian cells. *Nat Chem Biol* **10**, 203–208 (2014).

- 182 Lebar, T., Verbič, A., Ljubetič, A. & Jerala, R. Polarized displacement by transcription activator-like effectors for regulatory circuits. *Nat Chem Biol* **15**, 80–87 (2018).
- 183 Chen, Z. *et al.* De novo design of protein logic gates. *Science* **368**, 78–84 (2020).
- 184 Ma, D., Peng, S. & Xie, Z. Integration and exchange of split dCas9 domains for transcriptional controls in mammalian cells. *Nat Commun* **7**, 13056 (2016).
- 185 Kim, H., Bojar, D. & Fussenegger, M. A CRISPR/Cas9-based central processing unit to program complex logic computation in human cells. *Proc Natl Acad Sci USA* **116**, 7214–7219 (2019).
- 186 Ausländer, S., Ausländer, D., Müller, M., Wieland, M. & Fussenegger, M. Programmable single-cell mammalian biocomputers. *Nature* **487**, 123–127 (2012).
- 187 Gao, X. J., Chong, L. S., Kim, M. S. & Elowitz, M. B. Programmable protein circuits in living cells. *Science* **361**, 1252–1258 (2018).
- 188 Fink, T. *et al.* Design of fast proteolysis-based signaling and logic circuits in mammalian cells. *Nat Chem Biol* **15**, 115–122 (2018).
- 189 Hooshangi, S., Thiberge, S. & Weiss, R. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc Natl Acad Sci USA* **102**, 3581–3586 (2005).
- 190 Lapique, N. & Benenson, Y. Genetic programs can be compressed and autonomously decompressed in live cells. *Nat Nanotech* **13**, 309–315 (2018).
- 191 O'Shaughnessy, E. C., Palani, S., Collins, J. J. & Sarkar, C. A. Tunable signal processing in synthetic MAP kinase cascades. *Cell* **144**, 119–131 (2011).
- 192 Shopera, T. *et al.* Robust, tunable genetic memory from protein sequestration combined with positive feedback. *Nucleic Acids Res* **43**, 9086–9094 (2015).
- 193 Zhang, Q., Bhattacharya, S. & Andersen, M. E. Ultrasensitive response motifs: basic amplifiers in molecular signalling networks. *Open Biol* **3**, 130031 (2013).
- 194 Greber, D. & Fussenegger, M. An engineered mammalian band-pass network. *Nucleic Acids Res* **38**, e174 (2010).
- 195 Scholes, N. S. & Isalan, M. A three-step framework for programming pattern formation. *Curr Opin Chem Biol* **40**, 1–7 (2017).
- 196 Wang, Y.-H., Wei, K. Y. & Smolke, C. D. Synthetic biology: advancing the design of diverse genetic systems. *Annu Rev Chem Biomol Eng* **4**, 69–102 (2013).
- 197 Schwarz, K. A., Daringer, N. M., Dolberg, T. B. & Leonard, J. N. Rewiring human cellular input–output using modular extracellular sensors. *Nat Chem Biol* **13**, 202–209 (2017).
- 198 Edelstein, H. I. *et al.* Elucidation and refinement of synthetic receptor mechanisms. *bioRxiv* (2020).
- 199 Dolberg, T. B. *et al.* Computation-guided optimization of split protein systems. *bioRxiv* (2019).
- 200 Hart, Y. & Alon, U. The utility of paradoxical components in biological circuits. *Mol Cell* **49**, 213–221 (2013).

- 201 Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
- 202 Muldoon, J. J. & Sankaranarayanan, S., *et al.* Elucidating negative regulators of NF- κ B signaling through CRISPR-Cas9 genome-scale knockout screening. (In preparation).
- 203 Daringer, N. M., Schwarz, K. A. & Leonard, J. N. Contributions of unique intracellular domains to switchlike biosensing by Toll-like receptor 4. *J Biol Chem* **290**, 8764–8777 (2015).
- 204 Trinchieri, G. & Sher, A. Cooperation of Toll-like receptor signals in innate immune defence. *Nat Rev Immunol* **7**, 179–190 (2007).
- 205 Takeda, K., Kaisho, T. & Akira, S. Toll-like receptors. *Annu Rev Immunol* **21**, 335–376 (2003).
- 206 Iwasaki, A. & Medzhitov, R. Toll-like receptor control of the adaptive immune responses. *Nat Immunol* **5**, 987–995 (2004).
- 207 Savva, A. & Roger, T. Targeting toll-like receptors: promising therapeutic strategies for the management of sepsis-associated pathology and infectious diseases. *Front Immunol* **4**, 387 (2013).
- 208 Wang, J. Q., Jeelall, Y. S., Ferguson, L. L. & Horikawa, K. Toll-like receptors and cancer: MYD88 mutation and inflammation. *Front Immunol* **5**, 367 (2014).
- 209 Mills, K. H. TLR-dependent T cell activation in autoimmunity. *Nat Rev Immunol* **11**, 807–822 (2011).
- 210 Hennessy, E. J., Parker, A. E. & O'Neill, L. A. J. Targeting Toll-like receptors: emerging therapeutics? *Nat Rev Drug Discov* **9**, 293–307 (2010).
- 211 Fekonja, O., Avbelj, M. & Jerala, R. Suppression of TLR signaling by targeting TIR domain-containing proteins. *Curr Protein Pept Sci* **13**, 776–788 (2012).
- 212 Gay, N. J., Symmons, M. F., Gangloff, M. & Bryant, C. E. Assembly and localization of Toll-like receptor signalling complexes. *Nat Rev Immunol* **14**, 546–558 (2014).
- 213 Schmidt, M. *et al.* Crucial role for human Toll-like receptor 4 in the development of contact allergy to nickel. *Nat Immunol* **11**, 814–819 (2010).
- 214 Raghavan, B., Martin, S. F., Esser, P. R., Goebeler, M. & Schmidt, M. Metal allergens nickel and cobalt facilitate TLR4 homodimerization independently of MD2. *EMBO Rep* **13**, 1109–1115 (2012).
- 215 Smiley, S. T., King, J. A. & Hancock, W. W. Fibrinogen stimulates macrophage chemokine secretion through Toll-like receptor 4. *J Immunol* **167**, 2887–2894 (2001).
- 216 Park, J. S. *et al.* Involvement of toll-like receptors 2 and 4 in cellular activation by high mobility group box 1 protein. *J Biol Chem* **279**, 7370–7377 (2004).
- 217 Kim, S. *et al.* Signaling of high mobility group box 1 (HMGB1) through toll-like receptor 4 in macrophages requires CD14. *Mol Med* **19**, 88–98 (2013).
- 218 Marr, N. & Turvey, S. E. Role of human TLR4 in respiratory syncytial virus-induced NF- κ B activation, viral entry and replication. *Innate Immun* **18**, 856–865 (2012).
- 219 Georgel, P. *et al.* Vesicular stomatitis virus glycoprotein G activates a specific antiviral Toll-like receptor 4-dependent pathway. *Virology* **362**, 304–313 (2007).

- 220 Giannini, T. L. *et al.* Isolation of an endotoxin–MD-2 complex that produces Toll-like receptor 4-dependent cell activation at picomolar concentrations. *Proc Natl Acad Sci USA* **101**, 4186–4191 (2004).
- 221 Nagai, Y. *et al.* Essential role of MD-2 in LPS responsiveness and TLR4 distribution. *Nat Immunol* **3**, 667–672 (2002).
- 222 Park, B. S. *et al.* The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature* **458**, 1191–1195 (2009).
- 223 Panter, G. & Jerala, R. The ectodomain of the Toll-like receptor 4 prevents constitutive receptor activation. *J Biol Chem* **286**, 23334–23344 (2011).
- 224 Treeby, M., Vasl, J., Ota, P., Friedrich, J. & Jerala, R. Different functional role of domain boundaries of Toll-like receptor 4. *Biochem Biophys Res Commun* **381**, 65–69 (2009).
- 225 Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84 (2014).
- 226 Koike-Yusa, H., Li, Y., Tan, E.-P., Velasco-Herrera, M. D. C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat Biotechnol* **32**, 267–273 (2014).
- 227 Zhou, Y. *et al.* High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* **509**, 487–490 (2014).
- 228 Chew, J. *et al.* WIP1 phosphatase is a negative regulator of NF- κ B signalling. *Nat Cell Biol* **11**, 659–666 (2009).
- 229 Teo, H. *et al.* Telomere-independent Rap1 is an IKK adaptor and regulates NF- κ B-dependent gene expression. *Nat Cell Biol* **12**, 758–767 (2010).
- 230 Gewurz, B. E. *et al.* Genome-wide siRNA screen for mediators of NF- κ B activation. *Proc Natl Acad Sci USA* **109**, 2467–2472 (2012).
- 231 Lee, C. C., Carette, J. E., Brummelkamp, T. R. & Ploegh, H. L. A reporter screen in a human haploid cell line identifies CYLD as a constitutive inhibitor of NF- κ B. *PLoS One* **8**, e70339 (2013).
- 232 Olarerin-George, A. O., Anton, L., Hwang, Y. C., Elovitz, M. A. & Hogenesch, J. B. A functional genomics screen for microRNA regulators of NF- κ B signaling. *BMC Biol* **11** (2013).
- 233 Feldman, D. *et al.* Optical pooled screens in human cells. *Cell* **179**, 787–799 (2019).
- 234 Verma, S., De Jesus, P., Chanda, S. K. & Verma, I. M. SNW1, a novel transcriptional regulator of the NF- κ B pathway. *Mol Cell Biol* **39**, e00415-00418 (2019).
- 235 Zablocki-Thomas, L., Menzies, S. A., Lehner, P. J., Manel, N. & Benaroch, P. A genome-wide CRISPR screen identifies regulation factors of the TLR3 signalling pathway. *Innate Immun* (2020).
- 236 Covarrubias, S., *et al.* High throughput CRISPR screening identifies genes involved in macrophage viability and inflammatory pathways. *bioRxiv*.
- 237 Covarrubias, S. *et al.* CRISPR/Cas-based screening of long non-coding RNAs (lncRNAs) in macrophages with an NF- κ B reporter. *J Biol Chem* **292**, 20911–20920 (2017).

- 238 Sfikas, A. *et al.* Parallel siRNA screens to identify kinase and phosphatase modulators of NF- κ B activity following DNA damage. *bioRxiv*.
- 239 Leonard, J. N. *et al.* The TLR3 signaling complex forms by cooperative receptor dimerization. *Proc Natl Acad Sci USA* **105**, 258–263 (2008).
- 240 Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184–191 (2016).
- 241 Jeong, H.-H., Kim, S. Y., Rousseaux, M. W. C., Zoghbi, H. Y. & Liu, Z. Beta-binomial modeling of CRISPR pooled screen data identifies target genes with greater sensitivity and fewer false negatives. *Genome Res* **29**, 999–1008 (2019).
- 242 Allen, N. C. *et al.* Desynchronization of the molecular clock contributes to the heterogeneity of the inflammatory response. *Sci Adv* **12**, eaau1851 (2019).
- 243 Lane, K. *et al.* Measuring signaling and RNA-Seq in the same cell links gene expression to dynamic patterns of NF- κ B activation. *Cell Syst* **4**, 458–469 (2017).
- 244 Lane, Andres-Terre, M., Kudo, T., Monack, D. M. & Covert, M. W. Escalating threat levels of bacterial infection can be discriminated by distinct MAPK and NF- κ B signaling dynamics in single host cells. *Cell Syst* **8**, 183–196 (2019).
- 245 Lan, H., Nikolic-Paterson, D. J., Mu, W. & Atkins, R. C. Local macrophage proliferation in the progression of glomerular and tubulointerstitial injury in rat anti-GBM glomerulonephritis. *Kidney Int* **48**, 753–760 (1995).
- 246 Jenkins, S. J. *et al.* Local macrophage proliferation, rather than recruitment from the blood, is a signature of TH2 inflammation. *Science* **332**, 1284–1288 (2011).
- 247 Brown, N. J. *et al.* Fas death receptor signaling represses monocyte numbers and macrophage activation in vivo. *J Immunol* **173**, 7584–7593 (2004).
- 248 Lee, C. M. & Hu, J. Cell density during differentiation can alter the phenotype of bone marrow-derived macrophages. *Cell Biosci* **3**, 30 (2013).
- 249 Teixeira, A. P. & Fussenegger, M. Engineering mammalian cells for disease diagnosis and treatment. *Curr Opin Biomed Eng* **55**, 87–94 (2018).
- 250 Fitzgerald, M., Gibbs, C. & Deans, T. L. Rosa26 docking sites for investigating genetic circuit silencing in stem cells. *bioRxiv* (2019).
- 251 Sung, M.-H. *et al.* Switching of the relative dominance between feedback mechanisms in lipopolysaccharide-induced NF- κ B signaling. *Sci Signal* **7**, ra6 (2014).
- 252 Sobol, I. M. Uniformly distributed sequences with an additional uniform property. *USSR Comp Math Math+* **16**, 1332–1337 (1976).
- 253 Hartfield, R. M., Schwarz, K. A., Muldoon, J. J., Bagheri, N. & Leonard, J. N. Multiplexing Engineered Receptors for Multiparametric Evaluation of Environmental Ligands. *ACS synthetic biology* **6**, 2042–2055, doi:10.1021/acssynbio.6b00279 (2017).
- 254 Fedorov, V. D., Themeli, M. & Sadelain, M. PD-1- and CTLA-4-based inhibitory chimeric antigen receptors (iCARs) divert off-target immunotherapy responses. *Sci Trans Med* **5**, 215ra172 (2013).

- 255 Kloss, C. C., Condomines, M., Cartellieri, M., Bachmann, M. & Sadelain, M. Combinatorial antigen recognition with balanced signaling promotes selective tumor eradication by engineered T cells. *Nat Biotechnol* **31**, 71–75 (2012).
- 256 Roybal, K. T. *et al.* Precision tumor recognition by T cells with combinatorial antigen-sensing circuits. *Cell* **164**, 770–779 (2016).
- 257 Anderson, J. C., Voigt, C. A. & Arkin, A. P. Environmental signal integration by a modular AND gate. *Mol Syst Biol* **3**, 133 (2007).
- 258 Wang, B., Kitney, R. I., Joly, N. & Buck, M. Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nat Commun* **2**, 508 (2011).
- 259 Moon, T. S., Lou, C., Tamsir, A., Stanton, B. C. & Voigt, C. A. Genetic programs constructed from layered logic gates in single cells. *Nature* **491**, 249–253 (2012).
- 260 Brödel, A. K., Jaramillo, A. & Isalan, M. Engineering orthogonal dual transcription factors for multi-input synthetic promoters. *Nat Commun* **7**, 13858 (2016).
- 261 Kramer, B. P., Fischer, C. & Fussenegger, M. BioLogic gates enable logical transcription control in mammalian cells. *Biotechnol Bioeng* **87**, 478–484 (2004).
- 262 Marchisio, M. A. & Stelling, J. Automatic design of digital synthetic gene circuits. *PLoS Comput Biol* **7**, e1001083 (2011).
- 263 Win, M. N. & Smolke, C. D. Higher-order cellular information processing with synthetic RNA devices. *Science* **322**, 456–460 (2008).
- 264 Bonnet, J., Yin, P., Ortiz, M. E., Subsoontorn, P. & Endy, D. Amplifying genetic logic gates. *Science* **340**, 599–603 (2013).
- 265 Siuti, P., Yazbek, J. & Lu, T. K. Synthetic circuits integrating logic and memory in living cells. *Nat Biotechnol* **31**, 448–452 (2013).
- 266 Weinberg, B. H. *et al.* Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells. *Nat Biotechnol* **35**, 453–462 (2017).
- 267 Regot, S. *et al.* Distributed biological computation with multicellular engineered networks. *Nature* **469**, 207–211 (2011).
- 268 Tamsir, A., Tabor, J. J. & Voigt, C. A. Robust multicellular computing using genetically encoded NOR gates and chemical ‘wires’. *Nature* **469**, 212–215 (2011).
- 269 Ramalingam, K. I., Tomshine, J. R., Maynard, J. A. & Kaznessis, Y. N. Forward engineering of synthetic bio-logical AND gates. *Biochem Eng J* **47**, 38–47 (2009).
- 270 Schreiber, J., Arter, M., Lapique, N., Haefliger, B. & Benenson, Y. Model-guided combinatorial optimization of complex synthetic gene networks. *Mol Syst Biol* **12**, 899 (2016).
- 271 Schukur, L., Geering, B., Charpin-El Hamri, G. & Fussenegger, M. Implantable synthetic cytokine converter cells with AND-gate logic treat experimental psoriasis. *Sci Trans Med* **7**, 318ra201 (2015).
- 272 Sadowski, I., Ma, J., Triezenberg, S. & Ptashne, M. GAL4-VP16 is an unusually potent transcriptional activator. *Nature* **335**, 563–564 (1988).

- 273 Prindle, A. & Hasty, J. Making gene circuits sing. *Proc Natl Acad Sci USA* **109**, 16758–16759 (2012).
- 274 Davidsohn, N. *et al.* Accurate predictions of genetic circuit behavior from part characterization and modular composition. *ACS Synth Biol* **4**, 673–681 (2014).
- 275 Beal, J. *et al.* Model-driven engineering of gene expression from RNA replicons. *ACS Synth Biol* **4**, 48–56 (2015).
- 276 Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods* **10**, 973–976 (2013).
- 277 Maeder, M. L. *et al.* CRISPR RNA-guided activation of endogenous human genes. *Nat Methods* **10**, 977–979 (2013).
- 278 Siciliano, V. *et al.* Construction and modelling of an inducible positive feedback loop stably integrated in a mammalian cell-line. *PLoS Comput Biol* **7**, e1002074, doi:10.1371/journal.pcbi.1002074 (2011).
- 279 di Bernardo, D., Marucci, L., Menolascina, F. & Siciliano, V. Predicting synthetic gene networks. *Methods Mol Biol* **813**, 57–81, doi:10.1007/978-1-61779-412-4_4 (2012).
- 280 Corish, P. & Tyler-Smith, C. Attenuation of green fluorescent protein half-life in mammalian cells. *Protein Eng* **12**, 1035–1040 (1999).
- 281 Muldoon, J. J., Yu, J. S., Fassia, M.-K. & Bagheri, N. Network inference performance complexity: a consequence of topological, experimental, and algorithmic determinants. *Bioinform* **35**, 3421–3432 (2019).
- 282 Bansal, M., Belcastro, V., Ambesi-Impiombato, A. & di Bernardo, D. How to infer gene networks from expression profiles. *Mol Syst Biol* **3** (2007).
- 283 Bonneau, R. Learning biological networks: from modules to dynamics. *Nat Chem Biol* **4**, 658–664 (2008).
- 284 De Smet, R. & Marchal, K. Advantages and limitations of current network inference methods. *Nat Rev Microbiol* **8**, 717–729 (2010).
- 285 Marbach, D. *et al.* Revealing strengths and weaknesses of methods for gene network inference. *Proc Natl Acad Sci USA* **107**, 6286–6291 (2010).
- 286 Oates, C. J. & Mukherjee, S. Network inference and biological dynamics. *Ann Appl Stat* **6**, 1209–1235 (2012).
- 287 Ocone, A., Haghverdi, L., Mueller, N. S. & Theis, F. J. Reconstructing gene regulatory dynamics from high-dimensional single-cell snapshot data. *Bioinform* **31**, i89–i96 (2015).
- 288 Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**, 1083–1086 (2017).
- 289 Alexopoulos, L. G., Saez-Rodriguez, J., Cosgrove, B. D., Lauffenburger, D. A. & Sorger, P. K. Networks inferred from biochemical data reveal profound differences in Toll-like receptor and inflammatory signaling between normal and transformed hepatocytes. *Mol Cell Proteomics* **9**, 1849–1865 (2010).

- 290 Gu, J. & Xuan, Z. Inferring the perturbed microRNA regulatory networks in cancer using hierarchical gene co-expression signatures. *PLoS One* **8**, e81032 (2013).
- 291 Sass, S. *et al.* MicroRNA-target network inference and local network enrichment analysis identify two microRNA clusters with distinct functions in head and neck squamous cell carcinoma. *Int J Mol Sci* **16**, 30204–30222 (2015).
- 292 Volinia, S. *et al.* Reprogramming of miRNA networks in cancer and leukemia. *Genome Res* **20**, 589–599 (2010).
- 293 Lee, E., Chuang, H.-Y., Kim, J.-W., Ideker, T. & Lee, D. Inferring pathway activity toward precise disease classification. *PLoS Comput Biol* **4**, e1000217 (2008).
- 294 Wu, X., Jiang, R., Zhang, M. Q. & Li, S. Network-based global inference of human disease genes. *Mol Syst Biol* **4**, 189 (2008).
- 295 Ciaccio, M. F., Chen, V. C., Jones, R. B. & Bagheri, N. The DIONESUS algorithm provides scalable and accurate reconstruction of dynamic phosphoproteomic networks to reveal new drug targets. *Integr Biol* **7**, 776–791 (2015).
- 296 Gardner, T. S., di Bernardo, D., Lorenz, D. & Collins, J. J. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* **301**, 102–105 (2003).
- 297 Iorio, F., Saez-Rodriguez, J. & di Bernardo, D. Network based elucidation of drug response: from modulators to targets. *BMC Syst Biol* **7**, 139 (2013).
- 298 Korkut, A. *et al.* Perturbation biology nominates upstream–downstream drug combinations in RAF inhibitor resistant melanoma cells. *eLife* **4**, e04640 (2015).
- 299 Lecca, P. & Re, A. in *Cancer Gene Networks* Vol. 1513 *Methods in Molecular Biology* Ch. 8, 101–117 (2016).
- 300 Wildenhain, J. *et al.* Prediction of synergism from chemical-genetic interactions by machine learning. *Cell Syst* **1**, 383–395 (2015).
- 301 Chen, S. & Mar, J. C. Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data. *BMC Bioinform* **19**, 232 (2018).
- 302 Hache, H., Lehrach, H. & Herwig, R. Reverse engineering of gene regulatory networks: a comparative study. *EURASIP J Bioinform Syst Biol*, 617281 (2009).
- 303 Madhamshettiwar, P. B., Maetschke, S. R., Davis, M. J., Reverter, A. & Ragan, M. A. Gene regulatory network inference: evaluation and application to ovarian cancer allows the prioritization of drug targets. *Genome Med* **4**, 41 (2012).
- 304 Maetschke, S. R., Madhamshettiwar, P. B., Davis, M. J. & Ragan, M. A. Supervised, semi-supervised and unsupervised inference of gene regulatory networks. *Brief Bioinform* **15**, 195–211 (2013).
- 305 Ud-Dean, S. M. M. & Gunawan, R. Ensemble inference and inferability of gene regulatory networks. *PLoS One* **9**, e103812 (2014).
- 306 Zou, C. & Feng, J. Granger causality vs. dynamic Bayesian network inference: a comparative study. *BMC Bioinform* **10**, 122 (2009).

- 307 Bellot, P., Olsen, C., Salembier, P., Oliveras-Vergés, A. & Meyer, P. E. NetBenchmark: a bioconductor package for reproducible benchmarks of gene regulatory network inference. *BMC Bioinform* **16**, 312 (2015).
- 308 Tjårnberg, A., Morgan, D. C., Studham, M., Nordling, T. E. M. & Sonnhhammer, E. L. L. GeneSPIDER – gene regulatory network inference benchmarking with controlled network and data properties. *Mol Biosyst* **13**, 1304–1312 (2017).
- 309 Wang, M. *et al.* LegumeGRN: a gene regulatory network prediction server for functional and comparative studies. *PLoS Comput Biol* **8**, e67434 (2013).
- 310 Setty, Y., Mayo, A. E., Surette, M. G. & Alon, U. Detailed map of a cis-regulatory input function. *Proc Natl Acad Sci USA* **100**, 7702–7707 (2003).
- 311 Mangan, S. & Alon, U. Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci USA* **100**, 11980–11985 (2003).
- 312 Kalir, S., Mangan, S. & Alon, U. A coherent feed-forward loop with a SUM input function prolongs flagella expression in *Escherichia coli*. *Mol Syst Biol* **1**, 2005.0006 (2005).
- 313 Inoue, T. & Meyer, T. Synthetic activation of endogenous PI3K and Rac identifies an AND-gate switch for cell polarization and migration. *PLoS One* **3**, e3068 (2008).
- 314 Sudarsan, N. *et al.* Tandem riboswitch architectures exhibit complex gene control functions. *Science* **314**, 300–304 (2006).
- 315 Shea, M. A. & Ackers, G. A. The OR control system of bacteriophage lambda: a physical-chemical model for gene regulation. *J Mol Biol* **181**, 211–230 (1985).
- 316 Ackers, G. A., Johnson, A. D. & Shea, M. A. Quantitative model for gene regulation by λ phage repressor. *Proc Natl Acad Sci USA* **79**, 1129–1133 (1982).
- 317 Bintu, L. *et al.* Transcriptional regulation by the numbers: models. *Curr Opin Genet Dev* **15**, 116–124 (2005).
- 318 Ronen, M., Rosenberg, R., Shraiman, B. I. & Alon, U. Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. *Proc Natl Acad Sci USA* **99**, 10555–10560 (2002).
- 319 Bar-Even, A. *et al.* The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* **50**, 4402–4410 (2011).
- 320 Hargrove, J. L., Hulse, M. G. & Beale, E. G. The kinetics of mammalian gene expression. *Bioessays* **13**, 667–674 (1991).
- 321 Huynh-Thu, V. A., Irrthum, A., Wehenkel, L. & Geurts, P. Inferring regulatory networks from expression data using tree-based methods. *PLoS One* **5**, e12776 (2010).
- 322 Breiman, L. Random Forests. *Machine Learning* **45**, 5–32 (2001).
- 323 Haury, A.-C., Mordelet, F., Vera-Licona, P. & Vert, J.-P. TIGRESS: trustful inference of gene regulation using stability selection. *BMC Syst Biol* **6**, 145 (2012).

- 324 Hartemink, A. J., Gifford, D. K., Jaakkola, T. S. & Young, R. A. in *Pac Symp Biocomput* (eds R. Altman *et al.*) 422–433 (New Jersey, 2001).
- 325 Yu, J., Smith, V. A., Wang, P. P., Hartemink, A. J. & Jarvis, E. D. Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinform* **20**, 3594–3603 (2004).
- 326 Villaverde, A. F., Ross, J., Morán, F. & Banga, J. R. MIDER: network inference with mutual information distance and entropy reduction. *PLoS One* **9**, e96732 (2014).
- 327 Cantone, I. *et al.* A yeast synthetic network for in vivo assessment of reverse-engineering and modeling approaches. *Cell* **137**, 172–181 (2009).
- 328 Marbach, D., Schaffter, T., Mattiussi, C. & Floreano, D. Generating realistic in silico gene networks for performance assessment of reverse engineering methods. *J Comput Biol* **16**, 229–239 (2009).
- 329 Wolpert, D. H. & Macready, W. G. No free lunch theorems for optimization. *IEEE Trans Evol Comp* **1**, 67–82 (1997).
- 330 Dassa, B., London, N., Stoddard, B. L., Schueler-Furman, O. & Pietrokovski, S. Fractured genes: a novel genomic arrangement involving new split inteins and a new homing endonuclease family. *Nucleic Acids Res* **37**, 2560–2573 (2009).
- 331 Pan, D. *et al.* An intein-mediated modulation of protein stability system and its application to study human cytomegalovirus essential gene function. *Sci Rep* **6**, 26167 (2016).
- 332 Gramespacher, J. A., Stevens, A. J., Nguyen, D. P., Chin, J. W. & Muir, T. W. Intein zymogens: conditional assembly and splicing of split inteins via targeted proteolysis. *J Am Chem Soc* **139**, 8074–8077 (2017).
- 333 Böcker, J. K., Dörner, W. & Mootz, H. D. Light-control of the ultra-fast Gp41-1 split intein with preserved stability of a genetically encoded photo-caged amino acid in bacterial cells. *Chem Commun* **55**, 1287–1290 (2019).
- 334 Bachmann, A.-L. & Mootz, H. D. An unprecedented combination of serine and cysteine nucleophiles in a split intein with an atypical split site. *J Biol Chem* **290**, 28792–28804 (2015).
- 335 Beyer, H. M., Mikula, K. M., Li, M., Wlodawer, A. & Iwaï, H. The crystal structure of the naturally split gp41-1 intein guides the engineering of orthogonal split inteins from cis-splicing inteins. *FEBS J* **287**, 1886–1898 (2019).
- 336 Shah, N. H., Eryilmaz, E., Cowburn, D. & Muir, T. W. Naturally split inteins assemble through a “capture and collapse” mechanism. *J Am Chem Soc* **135**, 18673–18681 (2013).

APPENDIX 1. Vectors and cell types

Table A1.1 Vectors

Addgene submissions: ° COMET kit. △ COMET individual. * Genetic programs. ^ Receptor mechanisms. # mMoClo kit.

	ID	Promoter	Gene	Notes	
MESA protease mutagenesis					
	pJM001	CMV	PC FKBP-CD28-TEVp-intermediate	Intermediate	
^	pJM002	CMV	PC FKBP-CD28-TEVp-AIP(M)		
^	pJM003	CMV	PC FKBP-CD28-TEVp-AIP(A)		
^	pJM004	CMV	PC FKBP-CD28-TEVp-AIP(K)		
^	pJM005	CMV	PC FKBP-CD28-TEVp-AIP(Y)		
^	pJM006	CMV	PC FKBP-CD28-TEVp-PRS		
^	pJM007	CMV	PC FKBP-CD28-TEVp-PRS(M)		
^	pJM008	CMV	PC FKBP-CD28-TEVp-PRS(A)		
^	pJM009	CMV	PC FKBP-CD28-TEVp-PRS(K)		
^	pJM010	CMV	PC FKBP-CD28-TEVp-PRS(Y)		
Outdated TUPV backbones					
	pJM400	EF1α	EBFP2-P2A-BlastR (TUPV1)	Use pJM450 instead	
	pJM401	EF1α	EBFP2-P2A-BlastR (TUPV2)	Use pJM451 instead	
	pJM402	EF1α	EBFP2-P2A-BlastR (TUPV3)	Use pJM452 instead	
	pJM403	EF1α	EBFP2-P2A-BlastR (TUPV4)	Use pJM453 instead	
	pJM404	EF1α	EBFP2-P2A-BlastR (TUPV5)	Use pJM454 instead	
	pJM405	EF1α	EBFP2-P2A-BlastR (TUPV6)	Use pJM455 instead	
	pJM406	EF1α	EBFP2-P2A-BlastR (TUPV7)	Use pJM456 instead	
	pJM407	EF1α	EBFP2-P2A-BlastR (TUPV8)	Use pJM457 instead	
	pJM408	EF1α	EBFP2-P2A-BlastR (TUPV9)	Use pJM458 instead	
First set of TUPVs and IVs for COMET-LP					
	pJM409	CMV	mKate2		
*	pJM410	ZF1x6-C	mKate2		
	pJM411	ZF1x6-S	mKate2		
	pJM412	tRE3GV	mKate2		
	pJM413	EF1α	mKate2		
	pJM414	tRE3GV	3xFLAG-NLS-VP64-ZF1(RRRR)-P2A-EYFP	Do not use	
	pJM415	tRE3GV	3xFLAG-NLS-VP16-ZF1(RRRR)-P2A-EYFP	Do not use	
	pJM416	tRE3GV	3xFLAG-NLS-VP64-ZF1(AAAA)-P2A-EYFP	Do not use	
	pJM417	tRE3GV	3xFLAG-NLS-VP16-ZF1(AAAA)-P2A-EYFP	Do not use	
	pJM418	EF1α	3xFLAG-NLS-FKBP-ZF1(RRRR)-P2A-EYFP	Do not use	
	pJM419	EF1α	3xFLAG-NLS-VP64-ZF1(RRRR)-P2A-EYFP	Do not use	
	pJM420	EF1α	3xFLAG-NLS-VP16-ZF1(RRRR)-P2A-EYFP	Do not use	
	pJM421	EF1α	rtTA3		
#	pJM422	EF1α	rtTA3		
	pJM423	EF1α	NES-FRB-VP64		
	pJM427	Integration vector with pJM410, pJM414, pJM422, pJM403, pPD614, pPD630			Do not use
	pJM428	Integration vector with pJM410, pJM415, pJM422, pJM403, pPD614, pPD630			Do not use
	pJM429	Integration vector with pJM410, pJM416, pJM422, pJM403, pPD614, pPD630			Do not use
	pJM430	Integration vector with pJM410, pJM417, pJM422, pJM403, pPD614, pPD630			Do not use
	pJM431	Integration vector with pJM411, pJM415, pJM422, pJM403, pPD614, pPD630			Do not use
	pJM432	Integration vector with pJM410, pJM418, pJM423, pJM403, pPD614, pPD630			Do not use
	pJM433	Integration vector with pJM410, pJM419, pJM402, pPD613, pPD630			Do not use
	pJM434	Integration vector with pJM410, pJM420, pJM402, pPD613, pPD630			Do not use
	pJM435	Integration vector with pJM412, pJM421, pJM402, pPD613, pPD630			
	pJM436	Integration vector with pJM413, pJM401, pPD612, pPD630			
	pJM437	Integration vector with pJM409, pJM401, pPD612, pPD630			
	pJM442	Integration vector with pJM410, pJM401, pPD612, pPD630			
TUPV backbones					
#	pJM450	EF1α	EBFP2-P2A-BlastR (TUPV1)		
#	pJM451	EF1α	EBFP2-P2A-BlastR (TUPV2)		
#	pJM452	EF1α	EBFP2-P2A-BlastR (TUPV3)		
#	pJM453	EF1α	EBFP2-P2A-BlastR (TUPV4)		
#	pJM454	EF1α	EBFP2-P2A-BlastR (TUPV5)		
#	pJM455	EF1α	EBFP2-P2A-BlastR (TUPV6)		
#	pJM456	EF1α	EBFP2-P2A-BlastR (TUPV7)		
#	pJM457	EF1α	EBFP2-P2A-BlastR (TUPV8)		
#	pJM458	EF1α	EBFP2-P2A-BlastR (TUPV9)		
Second set of TUPVs and IVs for COMET-LP					
	pJM460	tRE3GV	3xFLAG-NLS-VP64-ZF1(RRRR)		
	pJM461	tRE3GV	3xFLAG-NLS-VP16-ZF1(RRRR)		
	pJM462	tRE3GV	3xFLAG-NLS-VP64-ZF1(AAAA)		

	pJM463	tRE3GV	3xFLAG-NLS-VP16-ZF1(AAAA)		
	pJM464	EF1 α	NLS-FKBP-ZF1(RRRR)		
*	pJM465	EF1 α	3xFLAG-NLS-VP64-ZF1		
Δ	pJM466	EF1 α	3xFLAG-NLS-VP16-ZF1		
#	pJM467	tRE3GV	EYFP		
	pJM468	EF1 α	EYFP		
#	pJM469	EF1 α	rtTA3		
	pJM470	EF1 α	NES-VP64-FRB		
	pJM471	Integration vector with pJM410, pJM460, pJM467, pJM469, pJM404, pPD615, pPD630			
	pJM472	Integration vector with pJM410, pJM461, pJM467, pJM469, pJM404, pPD615, pPD630			
	pJM473	Integration vector with pJM410, pJM462, pJM467, pJM469, pJM404, pPD615, pPD630			
	pJM474	Integration vector with pJM410, pJM463, pJM467, pJM469, pJM404, pPD615, pPD630			
	pJM475	Integration vector with pJM411, pJM461, pJM467, pJM469, pJM404, pPD615, pPD630			
	pJM476	Integration vector with pJM410, pJM464, pJM468, pJM470, pJM404, pPD615, pPD630			
	pJM477	Integration vector with pJM410, pJM465, pJM468, pJM403, pPD614, pPD630			
	pJM478	Integration vector with pJM410, pJM466, pJM468, pJM403, pPD614, pPD630			
First set of TFs and FPs for genetic programs					
	pJM500	EF1 α	EYFP		
	pJM501	EF1 α	EBFP2		
*	pJM502	ZF1x6-C	EYFP		
*	pJM503	EF1 α	3xFLAG-NLS-ZF1		
	pJM504	EF1 α	3xFLAG-NLS-ZF1		
*	pJM505	EF1 α	3xFLAG-NLS-DsRed-ZF1		
*	pJM506	EF1 α	3xFLAG-NLS-DsDed-ZF1		
*	pJM507	EF1 α	3xFLAG-NLS-DsDed-ZF10		
	pJM508	EF1 α	3xFLAG-NLS-VP64-Intermediate	Intermediate	
	pJM509	EF1 α	3xFLAG-NLS-VP64-intN (with incomplete, non-enzymatic intN)	Do not use	
	pJM510	EF1 α	3xFLAG-NLS-DsDed-intN (with incomplete, non-enzymatic intN)	Do not use	
	pJM511	ZF10x6-C	3xFLAG-NLS-DsDed-ZF1		
*	pJM512	EF1 α	3xFLAG-NLS-VP64-ZF1		
	pJM513	EF1 α	3xFLAG-NLS-VP64-ZF1		
	pJM514	EF1 α	3xFLAG-NLS-VP64-ZF1		
	pJM515	EF1 α	3xFLAG-NLS-VP64-ZF1		
*	pJM516	EF1 α	intC-ZF1-NLS-HA		
	pJM517	ZF10x6-C	3xFLAG-NLS-VP64-ZF1		
	pJM518	EF1 α	3xFLAG-NLS-VP64-ZF10		
	pJM519	EF1 α	3xFLAG-NLS-VP64-ZF10		
*	pJM520	EF1 α	3xFLAG-NLS-VP64-ZF10		
	pJM521	EF1 α	3xFLAG-NLS-VP64-ZF10		
*	pJM522	EF1 α	intC-ZF10-NLS-HA		
	pJM523	EF1 α	3xFLAG-NLS-VP64-ZF1 NLS-HA		
*	pJM524	EF1 α	3xFLAG-NLS-VP64-intC-ZF1-NLS-HA		
*	pJM525	EF1 α	3xFLAG-NLS-DsDed-intC-ZF1-NLS-HA		
	pJM526	EF1 α	3xFLAG-NLS-VP64-LL-ZF1-NLS-HA		
	pJM527	EF1 α	3xFLAG-NLS-VP64-LL-ZF10-NLS-HA		
	pJM528	EF1 α	3xFLAG-NLS-VP64-intC-ZF10-NLS-HA		
*	pJM529	EF1 α	3xFLAG-NLS-DsDed-intC-ZF10-NLS-HA		
uORFs and IVs					
	pJM532	Integration vector with pJM410, pJM511, pJM513, pJM520, pJM454, pPD615, pPD630			
	pJM535	Integration vector with pJM410, pJM525, pJM452, pPD613, pPD630			
	pJM543	EF1 α	ACC_uORF ACC_3xFLAG-NLS-VP64-ZF1		
	pJM544	EF1 α	TTT_uORF GAA_3xFLAG-NLS-VP64-ZF1		
	pJM545	EF1 α	ACC_uORF ACC_3xFLAG-NLS-VP64-ZF10		
	pJM546	EF1 α	ACC_uORF ACC_3xFLAG-NLS-DsDed-intC-ZF1-NLS-HA		
	pJM547	Integration vector with pJM410, pJM511, pJM513, pJM545, pJM454, pPD615, pPD630			
	pJM548	Integration vector with pJM410, pJM511, pJM544, pJM545, pJM454, pPD615, pPD630			
	pJM549	Integration vector with pJM410, pJM511, pJM543, pJM545, pJM454, pPD615, pPD630			
	pJM550	Integration vector with pJM410, pJM546, pJM452, pPD613, pPD630			
Second set of TFs and FPs for genetic programs					
	pJM551	EF1 α	3xFLAG-NLS-DsDed-ZF1-PEST		
	pJM552	EF1 α	3xFLAG-NLS-DsDed-ZF10-PEST		
*	pJM553	ZF10x6-C	3xFLAG-NLS-DsDed-ZF1-PEST		
*	pJM554	EF1 α	3xFLAG-NLS-VP64-intN		
*	pJM555	EF1 α	3xFLAG-NLS-DsDed-intN		
	pJM556	EF1 α	3xFLAG-NLS-VP64-intN_K43A		
	pJM557	EF1 α	3xFLAG-NLS-VP64-intN_K45A		
	pJM558	EF1 α	3xFLAG-NLS-VP64-intN_E52A		
	pJM559	EF1 α	3xFLAG-NLS-VP64-intN_fiveA		
	pJM560	EF1 α	intC-ZF1-NLS-HA_E102A		

	pJM561	EF1α	intC-ZF1-NLS-HA_E104A	
	pJM562	EF1α	intC-ZF1-NLS-HA_sixA	
	pJM563	EF1α	3xFLAG-NLS-VP64-intN_K43A_K45A	
	pJM564	EF1α	3xFLAG-NLS-VP64-intN_K45A_E52A	
	pJM565	EF1α	intC-ZF1-NLS-HA_E98A_E104A	
	pJM566	EF1α	intC-ZF1-NLS-HA_E102A_E104A	
	pJM567	EF1α	3xFLAG-NLS-VP64-intN-ZF10_K45A_E52A	
	pJM568	EF1α	3xFLAG-NLS-DsDed-intN-ZF10_K45A_E52A	
	pJM569	EF1α	NLS-ZF10-intC-ZF1-NLS-HA_E102A_E104A	
	pJM570	EF1α	NLS-intC-ZF1-NLS_E102A_E104A	
*	pJM571	ZF2x6-C	3xFLAG-NLS-VP64-intN	
*	pJM572	ZF6x6-C	intC-ZF1-NLS-HA	
*	pJM573	ZF6x6-C	3xFLAG-NLS-DsDed-intC-ZF1-NLS-HA	
*	pJM574	ZF2x6-C	3xFLAG-NLS-DsDed-ZF1	
*	pJM575	ZF6x6-C	3xFLAG-NLS-DsDed-ZF10	
*	pJM576	ZF2x6-C	3xFLAG-NLS-VP64-intC-ZF1-NLS-HA	
*	pJM577	ZF6x6-C	3xFLAG-NLS-DsDed-intN	
	pJM578	ZF2x6-C	3xFLAG-NLS-VP64-ZF1	
	pJM579	ZF2x6-C	3xFLAG-NLS-VP64-ZF10	
*	pJM580	ZF6x6-C	3xFLAG-NLS-VP64-ZF1	
	pJM581	ZF6x6-C	3xFLAG-NLS-VP64-ZF10	
*	pJM582	EF1α	3xFLAG-NLS-VP64-ZF2	
*	pJM583	EF1α	3xFLAG-NLS-VP64-ZF6	
*	pJM584	ZF2x6-C	mKate2	
*	pJM585	ZF6x6-C	mKate2	
	pJM586	ZF10x6-C	mKate2	
*	pJM587	ZF10x6-C	EYFP	
*	pJM588	EF1α	3xFLAG-NLS-VP64-intC-ZF1-DsDed-NLS-HA	
*	pJM589	EF1α	intN-ZF10-NLS-HA	
*	pJM590	EF1α	3xFLAG-NLS-VP64-ZF1-intC-ZF10-NLS-HA	
*	pJM591	EF1α	3xFLAG-NLS-VP64-ZF1-intN	
*	pJM592	EF1α	3xFLAG-NLS-DsDed-ZF10-intN-VP64	
	pJM593	ZF1x5-C/ZF2x1	3xFLAG-NLS-VP64-intN	
	pJM594	ZF1x5-C/ZF6x1	intC-ZF1-NLS-HA	
*	pJM595	ZF6x6-C	intC-ZF1-ZF2-NLS-HA	
	pJM596	ZF1x5-C/ZF6x1	intC-ZF1-ZF2-NLS-HA	
*	pJM597	ZF(2/6)x3	mKate2	
First set of MESA receptors for genetic programs				
*	pJM600	EF1α	PC 3xFLAG-FKBP-FGFR4-TEVp	
	pJM601	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-tTA	
	pJM602	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-intermediate	Intermediate
	pJM603	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-NLS-VP64-intN	Do not use
	pJM604	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-intC-ZF1-NLS	Do not use
	pJM605	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-NLS-DsDed-intN	Do not use
	pJM606	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-intC-ZF10-NLS	Do not use
	pJM607	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-NLS-VP64-intC-ZF10-NLS	Do not use
	pJM608	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-NLS-VP64-ZF1	Do not use
	pJM609	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-NLS-DsDed-ZF10	Use pJM620 instead
	pJM610	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-NLS-DsDed-ZF1	Use pJM621 instead
	pJM611	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(G)-NLS-VP64-ZF1	Do not use
*	pJM612	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)-NLS-VP64-ZF1	
	pJM613	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GSG)-NLS-VP64-ZF1	
	pJM614	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)x2-NLS-VP64-ZF1	
	pJM615	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)-NLS-VP64-intN	Do not use
	pJM616	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)-intC-ZF1-NLS	Do not use
	pJM617	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)-NLS-DsDed-intN	Do not use
	pJM618	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)-intC-ZF10-NLS	Do not use
	pJM619	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)-NLS-VP64-intC-ZF10-NLS	Do not use
	pJM620	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)-NLS-DsDed-ZF10	
*	pJM621	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)-NLS-DsDed-ZF1	
	pJM622	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)-NLS-VP64-intN_K43A	Do not use
	pJM623	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)-NLS-VP64-intN_K45A	Do not use
	pJM624	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)-NLS-VP64-intN_E52A	Do not use
	pJM625	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)-NLS-VP64-intN_fiveA	Do not use
	pJM626	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)-intC-ZF1-NLS_E102A	Do not use
	pJM627	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)-intC-ZF1-NLS_E104A	Do not use

pJM628	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)-intC-ZF1-NLS_sixA	Do not use
pJM629	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-Link18-intN-3xFLAG-NLS-VP64-ZF1	Do not use
pJM630	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-Link18-intN-3xFLAG-NLS-VP64-ZF1_K45A	Do not use
pJM631	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-Link18-intN-3xFLAG-NLS-VP64-ZF1_K43A_K45A	Do not use
pJM632	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-Link18-intN-3xFLAG-NLS-VP64-ZF1_K45A_E52A	Do not use
pJM633	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)-intC-ZF1-NLS_E98A_E104A	Do not use
pJM634	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)-intC-ZF1-NLS_E102A_E104A	Do not use
Membrane-tethered competitors for split inteins			
pJM650	EF1α	3xFLAG-Link40-PDGFR-intermediate	Do not use
pJM651	EF1α	3xFLAG-Link40-PDGFR-Link6-beta3	Do not use
pJM652	EF1α	3xFLAG-Link40-PDGFR-Link12-beta3	Do not use
pJM653	EF1α	3xFLAG-Link40-PDGFR-Link18-beta3	Do not use
pJM654	EF1α	3xFLAG-Link40-PDGFR-Link6-beta6	Do not use
pJM655	EF1α	3xFLAG-Link40-PDGFR-Link12-beta6	Do not use
pJM656	EF1α	3xFLAG-Link40-PDGFR-Link18-beta6	Do not use
Second set of MESA receptors for genetic programs			
pJM657	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)x3-NLS-VP64-ZF1	
pJM658	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)x4-NLS-VP64-ZF1	
pJM659	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)x5-NLS-VP64-ZF1	
pJM660	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)x6-NLS-VP64-ZF1	
pJM661	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)x2-intC-ZF1-NLS_E102A_E104A	Do not use
pJM662	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)x3-intC-ZF1-NLS_E102A_E104A	Do not use
pJM663	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)x4-intC-ZF1-NLS_E102A_E104A	Do not use
pJM664	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)x5-intC-ZF1-NLS_E102A_E104A	Do not use
pJM665	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)x6-intC-ZF1-NLS_E102A_E104A	Do not use
pJM666	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)-NLS-VP64-intN-ZF10_K45A_E52A	Do not use
pJM667	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)-NLS-DsDed-intN-ZF10_K45A_E52A	Do not use
pJM668	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)-NLS-ZF10-intC-ZF1-NLS_E102A_E104A	Do not use
pJM669	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)-NLS-intC-ZF1-NLS_E102A_E104A	Do not use
pJM670	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)-NLS-VP64-ZF2	
* pJM671	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)-NLS-VP64-ZF6	
pJM672	EF1α	TC 3xFLAG-FRB-FGFR4-PRS(M)-(GS)-NLS-VP64-ZF10	
synNotch lentiviral vectors from Addgene			
pJM700	SFFV	GFP-PDGFR	Addgene #79129
pJM701	pGK	Myc-LaG17nanobody-coreNotch-Gal4-VP16	Addgene #79127
synNotch in TUPVs			
pJM702	EF1α	GFP-PDGFR	
pJM703	EF1α	Myc-LaG17nanobody-coreNotch-Gal4-VP16	
pJM704	EF1α	Myc-LaG17nanobody-coreNotch-intermediate	Intermediate
Intermediate plasmids for Gib-MESA and ABA-MESA			
pJM750	CMV	PC intermediate	Intermediate
pJM751	CMV	TC intermediate	Intermediate
Gib-MESA PC with varying linker lengths			
^ pJM752	CMV	PC 3xFLAG-GID1-Link30-FGFR4-TEVp	
^ pJM753	CMV	PC 3xFLAG-GID1-Link50-FGFR4-TEVp	
^ pJM754	CMV	PC 3xFLAG-GID1-Link70-FGFR4-TEVp	
^ pJM755	CMV	PC 3xFLAG-GID1-Link90-FGFR4-TEVp	
^ pJM756	CMV	PC 3xFLAG-GAI-Link30-FGFR4-TEVp	
^ pJM757	CMV	PC 3xFLAG-GAI-Link50-FGFR4-TEVp	
^ pJM758	CMV	PC 3xFLAG-GAI-Link70-FGFR4-TEVp	
^ pJM759	CMV	PC 3xFLAG-GAI-Link90-FGFR4-TEVp	
Gib-MESA TC with varying linker lengths			
^ pJM760	CMV	TC 3xFLAG-GID1-Link30-FGFR4-PRS(M)-tTA	
^ pJM761	CMV	TC 3xFLAG-GID1-Link50-FGFR4-PRS(M)-tTA	
^ pJM762	CMV	TC 3xFLAG-GID1-Link70-FGFR4-PRS(M)-tTA	
^ pJM763	CMV	TC 3xFLAG-GID1-Link90-FGFR4-PRS(M)-tTA	
^ pJM764	CMV	TC 3xFLAG-GAI-Link30-FGFR4-PRS(M)-tTA	
^ pJM765	CMV	TC 3xFLAG-GAI-Link50-FGFR4-PRS(M)-tTA	
^ pJM766	CMV	TC 3xFLAG-GAI-Link70-FGFR4-PRS(M)-tTA	
^ pJM767	CMV	TC 3xFLAG-GAI-Link90-FGFR4-PRS(M)-tTA	
Gib-MESA TC with varying linker types			
pJM768	CMV	TC 3xFLAG-GID1-LinkHelical-PRS(M)-tTA	
pJM769	CMV	TC 3xFLAG-GID1-LinkIlgG4hinge-PRS(M)-tTA	
pJM770	CMV	TC 3xFLAG-GID1-LinkIlgG4hingeCH3-PRS(M)-tTA	
pJM771	CMV	TC 3xFLAG-GID1-LinkIlgG4hingeCH2CH3-PRS(M)-tTA	
pJM772	CMV	TC 3xFLAG-GID1-Link70CD33SS-PRS(M)-tTA	

ABA-MESA PC with varying linker lengths			
^	pJM773	CMV	PC 3xFLAG-ABI1-Link30-FGFR4-TEVp
^	pJM774	CMV	PC 3xFLAG-ABI1-Link50-FGFR4-TEVp
^	pJM775	CMV	PC 3xFLAG-ABI1-Link70-FGFR4-TEVp
^	pJM776	CMV	PC 3xFLAG-ABI1-Link90-FGFR4-TEVp
^	pJM777	CMV	PC 3xFLAG-PYL1-Link30-FGFR4-TEVp
^	pJM778	CMV	PC 3xFLAG-PYL1-Link50-FGFR4-TEVp
^	pJM779	CMV	PC 3xFLAG-PYL1-Link70-FGFR4-TEVp
^	pJM780	CMV	PC 3xFLAG-PYL1-Link90-FGFR4-TEVp
ABA-MESA TC with varying linker lengths			
^	pJM781	CMV	TC 3xFLAG-ABI1-Link30-FGFR4-PRS(M)-tTA
^	pJM782	CMV	TC 3xFLAG-ABI1-Link50-FGFR4-PRS(M)-tTA
^	pJM783	CMV	TC 3xFLAG-ABI1-Link70-FGFR4-PRS(M)-tTA
^	pJM784	CMV	TC 3xFLAG-ABI1-Link90-FGFR4-PRS(M)-tTA
^	pJM785	CMV	TC 3xFLAG-PYL1-Link30-FGFR4-PRS(M)-tTA
^	pJM786	CMV	TC 3xFLAG-PYL1-Link50-FGFR4-PRS(M)-tTA
^	pJM787	CMV	TC 3xFLAG-PYL1-Link70-FGFR4-PRS(M)-tTA
^	pJM788	CMV	TC 3xFLAG-PYL1-Link90-FGFR4-PRS(M)-tTA
ABA-MESA-COMET			
	pJM789	EF1 α	PC 3xFLAG-ABI1-Link50-FGFR4-TEVp
	pJM790	EF1 α	TC 3xFLAG-PYL1-Link90-FGFR4-PRS(M)-VP64-ZF1
	pJM791	EF1 α	TC 3xFLAG-PYL1-Link90-FGFR4-PRS(M)-VP64-ZF2
	pJM792	EF1 α	TC 3xFLAG-PYL1-Link90-FGFR4-PRS(M)-VP64-ZF6
	pJM793	EF1 α	TC 3xFLAG-PYL1-Link90-FGFR4-PRS(M)-VP64-ZF10
Gib-MESA PC with varying TMDs			
^	pJM794	CMV	PC 3xFLAG-GID1-Link90-CD28-TEVp
^	pJM795	CMV	PC 3xFLAG-GID1-Link90-GpA-TEVp
^	pJM796	CMV	PC 3xFLAG-GID1-Link90-FGFR1-TEVp
^	pJM797	CMV	PC 3xFLAG-GID1-Link90-Valine-TEVp
Gib-MESA TC with varying TMDs			
^	pJM798	CMV	TC 3xFLAG-GAI-Link70-CD28-PRS(M)-tTA
^	pJM799	CMV	TC 3xFLAG-GAI-Link70-GpA-PRS(M)-tTA
^	pJM800	CMV	TC 3xFLAG-GAI-Link70-FGFR1-PRS(M)-tTA
^	pJM801	CMV	TC 3xFLAG-GAI-Link70-Valine-PRS(M)-tTA
ABA-MESA PC with varying TMDs			
^	pJM802	CMV	PC 3xFLAG-ABI1-Link50-CD28-TEVp
^	pJM803	CMV	PC 3xFLAG-ABI1-Link50 GpA-TEVp
^	pJM804	CMV	PC 3xFLAG-ABI1-Link50-FGFR1-TEVp
^	pJM805	CMV	PC 3xFLAG-ABI1-Link50-Valine-TEVp
ABA-MESA TC with varying TMDs			
^	pJM806	CMV	TC 3xFLAG-PYL1-Link90-CD28-PRS(M)-tTA
^	pJM807	CMV	TC 3xFLAG-PYL1-Link90-GpA-PRS(M)-tTA
^	pJM808	CMV	TC 3xFLAG-PYL1-Link90-FGFR1-PRS(M)-tTA
^	pJM809	CMV	TC 3xFLAG-PYL1-Link90-Valine-PRS(M)-tTA
Other vectors used in the genetic programs study			
◇	pPD005	CMV	n/a
◇	pPD100	CMV	3xFLAG-NLS-VP16-ZF1
	pPD133	CMV	EYFP
◇	pPD189	CMV	3xFLAG-NLS-VP16-ZF2
	pPD193	CMV	EBFP2
◇	pPD270	ZF1/2x6-C	EYFP
◇	pPD290	ZF1x6-C	EYFP
◇	pPD303	CMV	3xFLAG-NLS-VP64-ZF1
◇	pPD341	CMV	3xFLAG-NLS-FKBP-ZF1
◇	pPD353	CMV	3xFLAG-NLS-VP16-FRB
◇	pPD545	ZF2x6-C	EYFP
*	pPD1122	CMV	3xFLAG-NES-PYL1-VP64
*	pKD011	CMV	3xFLAG-NLS-ZF2-ABI
*	pJB001	CMV	3xFLAG-NLS-FKBP-ZF2
*	pJB002	CMV	3xFLAG-NLS-DsDed-ZF1
*	pJB003	ZF1/2x6-C	3xFLAG-NLS-VP16-FRB
*	pJB004	CMV	3xFLAG-NLS-VP16-intN
*	pJB005	CMV	3xFLAG-NLS-intC-ZF2
Vectors used in the CRISPR study			
	pSS005		Lentiviral vector containing NF- κ B-inducible EGFP and CMV-driven Cas9-T2A-HygroR
Other vectors used in the receptor mechanisms study			
^	pPD806	CMV	PC 3xFLAG-FKBP-FGFR4-TEVp
^	pPD815	CMV	TC 3xFLAG-FRB-FGFR4-PRS(M)-tTA

Table A2.2 Cell types

Cell type	Notes
RAW 264.7 macrophages	Provided by David Segal (NIH)
RAW 264.7 reporter macrophages	Monoclonal line; generated by Sung, et al. ⁷³ by transducing with a lentivirus containing NF- κ B-inducible EGFP-RelA and <i>Tnf</i> promoter-driven mCherry-PEST
RAW 264.7 reporter macrophage sublines	Monoclonal sublines; generated by limiting dilution and clonal expansion of RAW 264.7 reporter macrophages
L929 fibroblasts	From ATCC
bone marrow-derived macrophages	Differentiated from bone marrow-derived monocytes from C57BL/6 mice
HEK293FT cells	For transfections and for lentivirus production
THP-1 monocytes	Parental line (not used for the library generation in the CRISPR screen)
THP-1 reporter monocytes	Polyclonal line; transduced with a lentivirus containing NF- κ B-inducible EGFP and CMV-driven Cas9-T2A-HygroR

APPENDIX 2. Model equations for macrophage quorum licensing

This appendix corresponds to **Chapter 2**, and a version of this appendix was previously published as:

Muldoon J.J., Chuang Y., Bagheri N.[^], Leonard J.N.[^] Macrophages employ quorum licensing to regulate collective activation. *Nat Commun* **11**, 878 (2020).⁶² [^]Co-corresponding

Overview

A computational model was developed to incorporate the new findings with prior knowledge on macrophage activation. The scope is focused processes to which the new findings most directly pertain. This section describes prior knowledge and the model formulation, development, parameterization, and analysis.

Model formulation based upon prior knowledge

Cellular mechanisms and their representation in the model are described below. Salient state variables are in parentheses.

TLR4 signaling ($\dot{x}_1, \dot{x}_2, \dot{x}_5, \dot{x}_6, \dot{x}_{EC2}$): *Mechanism:* LPS activates TLR4 signaling through two pathways named after the adaptor proteins MyD88 (Myeloid differentiation primary response gene 88) and TRIF (Toll/interleukin-1 receptor (TIR)-domain-containing adapter-inducing interferon- β)⁹¹. While these pathways differ in certain components and downstream effects, they overlap in NF- κ B activation. MyD88 signaling involves the formation of a multi-subunit protein complex at the plasma membrane called the Myddosome, whose activity is induced and terminated rapidly. TRIF signaling requires TLR4 internalization to endosomes; signaling is initially delayed as activated TLR4 begins to reversibly shuttle from the plasma membrane to endosomes, but it is longer-lasting than MyD88 signaling. Signaling from activated receptors terminates after maturation of early endosomes to late endosomes. *Model:* LPS has an initial value of 1 a.u., corresponding to a dose of 100 ng ml⁻¹. The inactive (TLR4) and active (TLR4*) forms of the receptor are assigned initial values of 0.1 and 0 a.u., respectively. Since LPS is in large molar excess of the receptor, its loss over time can be represented simply by first-order degradation. TLR4 is synthesized constitutively in its inactive form and undergoes first-order degradation. Receptor activation depends on LPS dose, not cell density. For the signaling cascade, we developed a reduced complexity version of a previously published model⁹¹, e.g., that does not distinguish MyD88-mediated and TRIF-mediated signaling. As a general principle, we used a minimal number of variables to represent the most salient processes. During model development, we observed that reducing the granularity of certain mechanisms had little impact on the dynamics of interest such as NF- κ B translocation.

NF- κ B activation ($\dot{x}_7, \dot{x}_8, \dot{x}_{10}, \dot{x}_{11}, \dot{x}_{13}, \dot{x}_{14}, \dot{x}_{15}, \dot{x}_{16}, \dot{x}_{17}, \dot{x}_{18}$): *Mechanism:* NF- κ B is a dimer composed from five subunits: RelA (p65), cRel, RelB, p50, and p52⁷⁹. In TLR4 signaling and in TNFR signaling, IKKK

phosphorylates IKK (I κ B kinase), which phosphorylates I κ B, targeting I κ B for degradation. The de-sequestered NF- κ B can then translocate to the nucleus and induce target gene transcription. NF- κ B is eventually re-sequestered by I κ B, and NF- κ B–I κ B translocates to the cytoplasm. This activation and inactivation comprise one cycle of NF- κ B nucleocytoplasmic translocation⁹¹. *Model*: we reduced and modified a portion of a previous model⁹¹ while retaining NF- κ B oscillatory behavior. Since RAW reporter cells express a functional EGFP-p65, and p65/p50 is the primary NF- κ B dimer, we represent EGFP-p65-containing and native p65/p50 dimers as the same NF- κ B variable. The variable I κ B represents the I κ B α gene product¹⁰⁴.

RelA feedback dominance switching (\dot{x}_9, \dot{x}_{11}): *Mechanism*: RAW cells treated with LPS above a certain dose threshold enter a positive feedback loop in which NF- κ B induces *Rela* expression²⁵¹. This feedback dominance (FBD) switch counters the negative feedback from the induction of I κ B by NF- κ B. Since the switch requires *de novo* expression of the TF Ikaros, it takes effect starting several hours post-LPS. LPS at 100 ng ml⁻¹ is well above the dose threshold. *Model*: NF- κ B undergoes the FBD switch in all cells at and above a presumed dose of 1 ng ml⁻¹. A time-dependent function was formulated for NF- κ B-induced activity at the *Rela* promoter, based on previously published timecourse CHIP data²⁵¹ for NF- κ B localization at the *Rela* promoter in RAW cells.

Early regulation of Tnf translation ($\dot{x}_{24}, \dot{x}_{25}$): *Mechanism*: in the resting cell state, *Tnf* mRNA lacks a poly(A) tail and is not translated¹⁰¹. After LPS treatment, the mRNA is polyadenylated, allowing poly(A)-binding protein (PABP) to pseudo-circularize the mRNA, which increases ribosome recycling for rapid translation. TRIF signaling also promotes translation, by activating p38 mitogen activated protein kinase (MAPK), which activates MAP kinase-activated protein kinase 2 (MK2) to phosphorylate eukaryotic translation initiation factor 4E (eIF4E), which binds to the 5' mRNA cap and recruits the 40S ribosomal subunit. TRIF signaling also dephosphorylates eukaryotic initiation factor 2 (eIF2), which de-represses translation by recruiting the 60S subunit¹⁰¹. *Model*: Since TNF production requires LPS treatment, the initial values of *Tnf* mRNA and TNF protein are set to zero regardless of the initial value of NF- κ B.

Tnf post-transcriptional regulation ($\dot{x}_{19}, \dot{x}_{20}, \dot{x}_{24}, \dot{x}_{25}$): *Mechanism*: *Tnf* mRNA is regulated post-transcriptionally through AU-rich elements (AREs) in its 5' UTR, with binding sites for over 20 proteins¹⁰¹. Some proteins such as Tristetraprolin (TTP) destabilize the mRNA by recruiting deadenylases and

degradation factors, and others such as Hu-antigen R (HUR) stabilize the mRNA by competing with destabilizing proteins for occupancy¹⁰¹. In unstimulated macrophages, TTP is expressed at low levels¹⁰⁵. Shortly after LPS treatment, kinases including p38 and Erk are activated and *Tnf* mRNA is stabilized¹⁰⁶. However, TLR4 and TNFR signaling also induce TTP expression via p38 and ERK signaling¹⁰⁷ (and IL-10R signaling also induces TTP expression, via STAT3¹⁰⁵). TTP binds to *Tnf* mRNA and leads to its destabilization¹⁰⁰. The outcome is a limited-duration burst in TNF expression¹⁰⁶. *Model*: To capture the limited-duration burst in TNF expression while limiting model complexity, stabilizing regulation (SR) is represented by one variable and destabilizing regulation (DSR) is represented by another. SR becomes active after LPS (downstream of TLR4* and TNFR* via IKKK*), and it decreases in activity over time. SR slows the degradation of *Tnf* mRNA. DSR becomes active after a delay. DSR increases *Tnf* mRNA degradation and suppresses TNF translation.

Action of IL-10 pre-treatment through STAT3 (represented through a parameter, not state variables): *Mechanism*: Bcl-3 is a nuclear-localized protein that dimerizes with p50 or p52 and binds NF- κ B-responsive promoters¹⁰⁸. Bcl-3 expression is induced by STAT3 in IL-10R signaling. *Model*: To limit model complexity, rather than introducing variables for IL-10R signaling, STAT3 activation, or Bcl-3 interactions, we found that representing the effect of IL-10 on FBD via a fitted parameter was sufficient to capture the observed decrease in reporter expression.

Action of IL-10 pre-treatment through MAPKs (\dot{x}_{19} , \dot{x}_{20}): *Mechanism*: MAPKs regulate the initial response to LPS and the resolution¹⁰⁹. Their effects on downstream targets are complex and have been described as *incoherent*—having seemingly opposing effects. Signaling via MyD88 and TRIF activates p38, extracellular-signal-regulated kinases 1 and 2 (ERK1/2), and c-Jun N terminal kinase (JNK). MK2, a phosphorylation target of p38, regulates *Tnf* and *Il10* mRNA stability by: (a) preventing recruitment of the adenylyase CCR4-associated factor 1 (CAF1), thereby preventing proteins like TTP from destabilizing target mRNAs, and (b) inducing TTP expression. IL-10R signaling also regulates *Tnf*, by activating STAT3 (which induces TTP) and increasing expression of dual specific phosphatase 1 (DUSP1, which dephosphorylates p38 and inhibits late-phase p38 activity)¹⁰⁹. In macrophages, IL-10R signaling destabilizes inflammatory cytokine mRNAs like *Tnf* that contain 3' UTR AU-rich elements (AREs), by: (a) repressing LPS-induced activation of p38 MAPK, and (b) inhibiting expression of HuR, a protein that stabilizes mRNAs by binding

AREs⁹⁹. *Model*: IL-10 treatment decreases stabilizing regulation and increase destabilizing regulation of *Tnf* mRNA. Effects of the MAPKs are represented by SR downstream of IKKK activation, and effects of TTP are represented by DSR. IL-10 pre-treatment prevents SR, and it activates DSR by 0 hps.

TNFR activation ($\dot{x}_3, \dot{x}_4, \dot{x}_5, \dot{x}_6, \dot{x}_{EC1}$): *Mechanism*: Extracellular TNF binds TNFR1, and the receptor-ligand complex is internalized⁷². Adaptor proteins activate IKKK and MAPK signaling, leading to NF-kB activation. *Model*: the TNF pool has an initial value of 0 a.u. The inactive (TNFR) and active (TNFR*) forms of the receptor are assigned initial values of 0.1 and 0 a.u., respectively, analogous to the TLR4 receptor. TNFR* and TLR4* converge at IKKK activation, and thus overlap in regulating *Tnf*.

Cell density (\dot{x}_{EC1}): *Mechanism*: Cell density affects the proportion of highly activated cells and the amount of secreted TNF. *Model*: To account for population growth over time, the rate at which secreted TNF contributes to the extracellular TNF pool is multiplied by a time-dependent function. The multiplier is applied to the pool, not individual cells. The proportion of highly activated cells is based on experimental observations.

Blockade of TNFR signaling (\dot{x}_3, \dot{x}_4): *Mechanism*: At 1 h pre-LPS, cells were treated with soluble TNF receptor (sTNFR), which competes with surface TNFR for TNF. The dose of sTNFR was chosen based on a prior study such that it would be in molar excess of secreted TNF⁷¹. *Model*: sTNFR blockade prevents TNFR activation. As a result, following LPS treatment, downstream nodes such as IKKK are activated to a lesser extent than they would be with paracrine feedback.

Secretion ($\dot{x}_{25}, \dot{x}_{EC1}$): *Mechanism*: p38 modulates the activity of proteins involved in endocytic trafficking to enhance cytokine secretion during the LPS response. At the cell surface, TNF precursor is released in soluble form following cleavage by TNF α -converting enzyme (TACE), which is activated by an LPS-activated lipid hydrolase¹⁰⁹. *Model*: since simulations begin at 0 hps, TNF secretion is assigned a constant rate parameter.

Blockade of secretion ($\dot{x}_{25}, \dot{x}_{EC1}$): *Mechanism*: Brefeldin A (BFA) prevents secretion involving Golgi transport. After BFA treatment, TNF is no longer secreted, and it can accumulate intracellularly. *Model*: A time-dependent step-down function represents prevention of TNF secretion. Since some TNF is secreted prior to BFA, some paracrine signaling occurs, though to a lesser extent than without BFA.

Model formulation for other cellular processes

Receptor activation: Receptors are synthesized constitutively in an inactive form and are initially at steady state. Receptors become activated through a second-order reaction with the extracellular cue, and initiate the downstream cascade. Certain rate constants differ for TLR4 and TNFR.

Kinase cascade: Kinases are present at a constant level. They are initially inactive, and become activated by a second-order reaction with an upstream node. Basal deactivation is first-order. For example, TLR4* and IKKK react to convert IKKK to IKKK*, which eventually returns to IKKK (due to the action of phosphatases).

Translocation: Translocation between the cytoplasm and nucleus is first order. Rate constants differ by species (NF- κ B, I κ B, and NF- κ B-I κ B) and direction of movement.

Transcription: Transcription was formulated using fractional activation. Terms for transcriptional activators are in both the numerator and denominator. Terms for inhibitory effects are in the denominator.

Translation: Translation is generally treated as first order with mRNA. However, *Tnf* translation has additional regulation.

Secretion: There are 30 cells in the model, and the sum of their TNF secretion contributes to the extracellular pool.

Degradation: Degradation is generally treated as first order. Cases involving regulated degradation are nonlinear.

Model development, parameterization, and implementation

Before parameterizing the full model, we started with a model of cell-intrinsic effects on NF- κ B that includes TLR4 signaling, regulation of NF- κ B activation, and RelA and I κ B expression. Following LPS treatment, TLR4 is activated to TLR4*, which activates IKKK to IKKK*, which activates IKK to IKK*. In the cytoplasm, IKK* associates with I κ B or NF- κ B-I κ B to form IKK-I κ B or NF- κ B-IKK-I κ B, respectively. IKK-I κ B and NF- κ B also associate to form NF- κ B-IKK-I κ B, and IKK targets I κ B for degradation. (Formally, degradation occurs regardless of whether NF- κ B is complexed, but for the purpose of model reduction we represented this process for NF- κ B-IKK-I κ B only. This decision did not have a noticeable impact on NF- κ B activity.) The reaction releases NF- κ B and IKK, and NF- κ B can then enter the nucleus and induce transcription of

Rela and *Ikba*. NF- κ B, I κ B, and NF- κ B-I κ B translocate between the nucleus and cytoplasm. Since NF- κ B-I κ B exits the nucleus with a much faster rate constant than that with which it enters, its translocation is treated as unidirectional. As with the representation of the effect of IKK on I κ B, this model reduction had no discernable impact. During model development, we encountered many such instances where complexity could be reduced for reasons such as separation of timescales, redundant pathway effects, or negligible reaction fluxes.

To conduct an broad search of parameter space for fitting the model, we tested many parameter sets using a Sobol sequence—a pseudorandom number list that uniformly samples the unit hypercube in the limit of the sequence²⁵². This initial sweep was followed by multi-objective optimization using a genetic algorithm with many generations, each comprising the following steps:

1. Evaluation: quantify goodness of fit for each parameter set based on the deviation of simulated outcomes from experimental data.
2. Selection: identify parameter sets that yield the best fits in the current generation. Stringent criteria were applied to eliminate sets for which the population-mean simulated outcomes fell outside of any specified windows of acceptable deviation from the population-mean experimental data.
3. Repopulation: replicate the selected sets to restore the population size.
4. Mutation: introduce random variation to the parameter values, drawing from Gaussian distributions centered on current values. Simulated annealing was used, in which coefficients of variation were decreased after many generations, to narrow in on solutions.

The algorithm yielded a family of similar-performing four-parameter sets for the NF- κ B module. These sets were carried forward and sampled during the fitting of the remaining eight parameters in the full model, and this fit yielded the homogeneous (one-cell) model. Fitting, simulations, and analysis were conducted in MATLAB.

To investigate the NF- κ B module independent of TNF intercellular feedback, the homogeneous model was run under various conditions. The results depict how, following TLR4 activation, NF- κ B enters the nucleus and induces target gene transcription. Damped oscillations are attributable to the negative

feedback between NF- κ B activation and I κ B expression and the time required for *de novo* I κ B expression. We find that depending on the LPS dose, FBD switch, and initial value of inactive NF- κ B, simulations yield qualitatively distinct trajectories for total and nuclear NF- κ B, consistent with wide-ranging outcomes observed by confocal microscopy.

A population of 30 cells was generated using equivalent intracellular state variables and reactions, which all had access to LPS and the extracellular TNF pool. Heterogeneity was introduced by assigning differences to the basal transcription rate of NF- κ B RNA, initial value of NF- κ B RNA, and initial value of cytoplasmic NF- κ B-I κ B (corresponding to 0 hps). These values are proportional to confocal microscopy measurements for the 30 cells at high density. Values for cells at low density, in equivalent units, were obtained as described for **Figure 2.3b**.

Model parameterization: cell growth

RAW cell density was monitored over time for different cell densities and treatment conditions. These data were used to fit a logistic model for time-dependent growth.

Model formulation: representing the experimental perturbations

Perturbations were modeled as described below. Parameters are denoted by k for kinetic processes and w for weights in transcriptional regulation. Variables are abbreviated as: NF κ Bn, nuclear NF- κ B; Tnfm, *Tnf* mRNA; TNF, intracellular TNF; TNFpool, extracellular TNF pool; SR, stabilizing regulation on *Tnf* mRNA; DSR, destabilizing regulation on *Tnf* mRNA.

Terms containing δ indicate a treatment or perturbation that is either present or absent for each scenario to be analyzed.

$$\delta_{IL10} = \begin{cases} no\ IL10 \rightarrow 0 \\ IL10 \rightarrow 1 \end{cases} \quad (A.21)$$

$$\delta_{sTNFR} = \begin{cases} no\ sTNFR \rightarrow 0 \\ sTNFR \rightarrow 1 \end{cases} \quad (A.22)$$

$$\delta_{LPS} = \begin{cases} no\ LPS \rightarrow 0 \\ LPS \rightarrow 1 \end{cases} \quad (A.23)$$

$$\delta_{\text{BFA}} = \begin{cases} \text{no BFA} \rightarrow 0 \\ \text{BFA} \rightarrow 1 \end{cases} \quad (\text{A2.4})$$

Time-dependent cell density (ρ) is described by a logistic equation. Units are relative to high cell density (1 a.u.) at the time of LPS treatment ($t = 0$). From a fit to data (**Supplementary Fig. 3c**) with constraints for eight-fold difference at plating and for equal density in the limit of time, the horizontal asymptote for maximum density (r_1) is 1.99 a.u., the rate of logistic growth (r_2) is 0.0365 h^{-1} , and τ_{density} is – 0.247 h for high density and 62.6 h for low density.

$$\rho(t, \tau_{\text{density}}) = \frac{r_1}{1 + e^{-r_2 \cdot (t - \tau_{\text{density}})}} \quad (\text{A2.5})$$

Total TNF secretion by cells $i=1:N$, accounting for population growth and sTNFR pre-treatment, is:

$$\rho \cdot (1 - \delta_{\text{sTNFR}}) \cdot k_{\text{secretion}} \cdot \sum_i^N [\text{TNF}_i] \quad (\text{A2.6})$$

The effect of FBD on *Rela* transcription is represented by a unitless time-dependent function F_{FBD} , which was formulated and fitted based on timecourse CHIP data²⁵¹ for RelA localization at the *Rela* promoter post-LPS.

$$F_{\text{FBD}} = \frac{1}{6} \cdot \delta_{\text{LPS}} \cdot t^3 \cdot e^{-t} \quad (\text{A2.7})$$

Rela transcription involves basal transcription, inducible transcription including the effect of FBD, and the inhibitory effect of IL-10 pre-treatment.

$$k_{\text{tx}} \frac{w_{\text{basalTxNFkB}} + F_{\text{FBD}} \cdot w_{\text{NFkBtxNFkB}} \cdot [\text{NFkBn}]}{1 + F_{\text{FBD}} \cdot w_{\text{NFkBtxNFkB}} \cdot ([\text{NFkBn}] + w_{\text{IL10txFBD}} \cdot \delta_{\text{IL10}})} \quad (\text{A2.8})$$

Tnf and *mCherry* transcription are each induced by nuclear NF-κB.

$$F_{\text{NFkB_Tnf}} = w_{\text{NFkBtxTnf}} \cdot [\text{NFkBn}] \quad (\text{A2.9})$$

An additional consideration with IL-10 pre-treatment is that we experimentally observed a decrease in *Tnf* promoter activity beginning after some delay post-LPS. This can be described by a time-dependent

ramp down function (F_{IL10_Tnf}) starting at time τ_{IL10_1} and ending at time τ_{IL10_2} , where H is the Heaviside function.

$$F_{IL10_Tnf} = 1 - \delta_{IL10} \cdot \left(1 - \frac{\max(t - \tau_{IL10_1}, 0)}{\tau_{IL10_2} - \tau_{IL10_1}}\right) \cdot \left(1 - \frac{1}{1 + H(t - \tau_{IL10_2})}\right) \quad (A2.10)$$

Transcription from the *Tnf* promoter, incorporating IL-10's effect, is:

$$\delta_{LPS} \cdot k_{tx} \cdot \frac{F_{NFkB_Tnf}}{1 + F_{NFkB_Tnf}} \cdot F_{IL10_Tnf} \quad (A2.11)$$

For post-transcriptional regulation of *Tnf* mRNA, there are two opposing effects: stabilizing regulation slows the degradation, and destabilizing regulation promotes the degradation.

$$F_{stabilize} = \frac{1}{1 + w_{stabilize} \cdot [SR]} \quad (A2.12)$$

$$F_{destabilize} = \frac{w_{destabilize} \cdot [DSR] \cdot (w_{maxdestabilize} - 1)}{1 + w_{destabilize} \cdot [DSR]} \quad (A2.13)$$

With these effects, *Tnf* mRNA degradation is:

$$k_{degTnfm} \cdot [Tnfm] \cdot (F_{stabilize} + F_{destabilize}) \quad (A2.14)$$

Destabilizing regulation also acts to repress translation of the TNF protein:

$$F_{repress_tl} = \frac{1 + w_{destabilize} \cdot [DSR]}{1 + w_{maxdestabilize} \cdot w_{destabilize} \cdot [DSR]} \quad (A2.15)$$

Translational repression is represented as:

$$k_{tl} \cdot [Tnfm] \cdot F_{repress_tl} \quad (A2.16)$$

sTNFR pre-treatment blocks TNFR signaling. Its effect is a zero multiplier on TNFR activation.

$$(1 - \delta_{sTNFR}) \cdot k_{activateTNFR} \cdot [TNFR] \cdot [TNFpool] \quad (A2.17)$$

BFA prevents secretion starting at the time of BFA treatment (τ_{BFA}).

$$\delta_{\text{BFA}}(t, \tau_{\text{BFA}}) = 1 - H(t - \tau_{\text{BFA}}) \quad (\text{A2.18})$$

BFA's effect is represented by a step-down multiplier on TNF secretion for each cell.

$$k_{\text{secretion}} \cdot \delta_{\text{BFA}} \cdot [\text{TNF}] \quad (\text{A2.19})$$

State variables, parameters^{31,69,84,90,91,251}, stimulus-specific and perturbation-specific parameters, and ODEs are provided in **Tables A2.1–A2.4**.

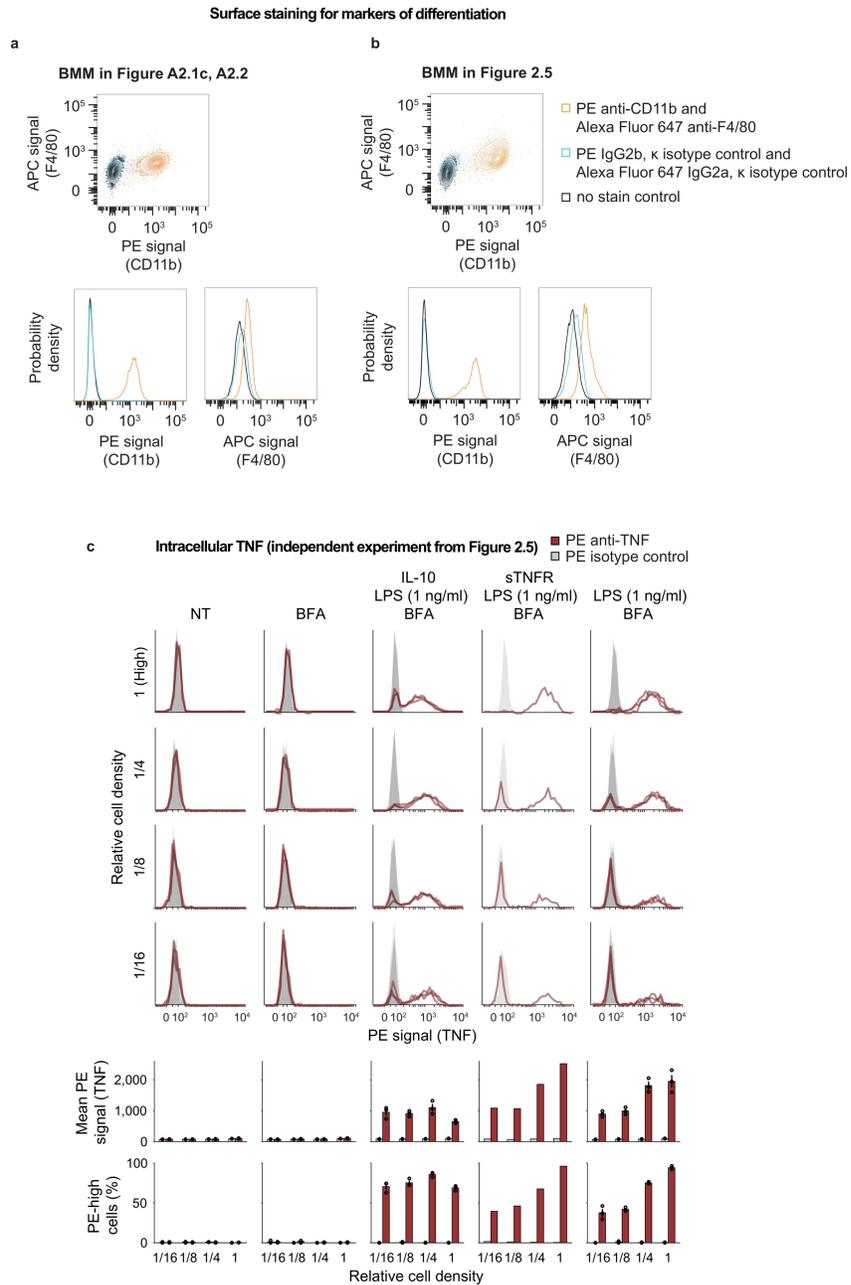


Figure A2.1. Primary cell experiments. **a–b** Surface staining for markers of differentiation. At seven days after each of the two bone marrow harvests, differentiation to macrophages was assessed by surface staining for CD11b and F4/80. A signal that is distinguishable from the isotype control was observed for both harvests. **c** An independent experiment from **Figure 2.5** varying the plating density and treatment conditions shows cell density-dependent bimodality in LPS-inducible TNF expression. Treatment conditions (columns) were sampled in biological triplicate, except for the fourth condition (with sTNFR), which was sampled using one biological replicate. Bar graphs represent the mean of the biological replicates and S.E.M.

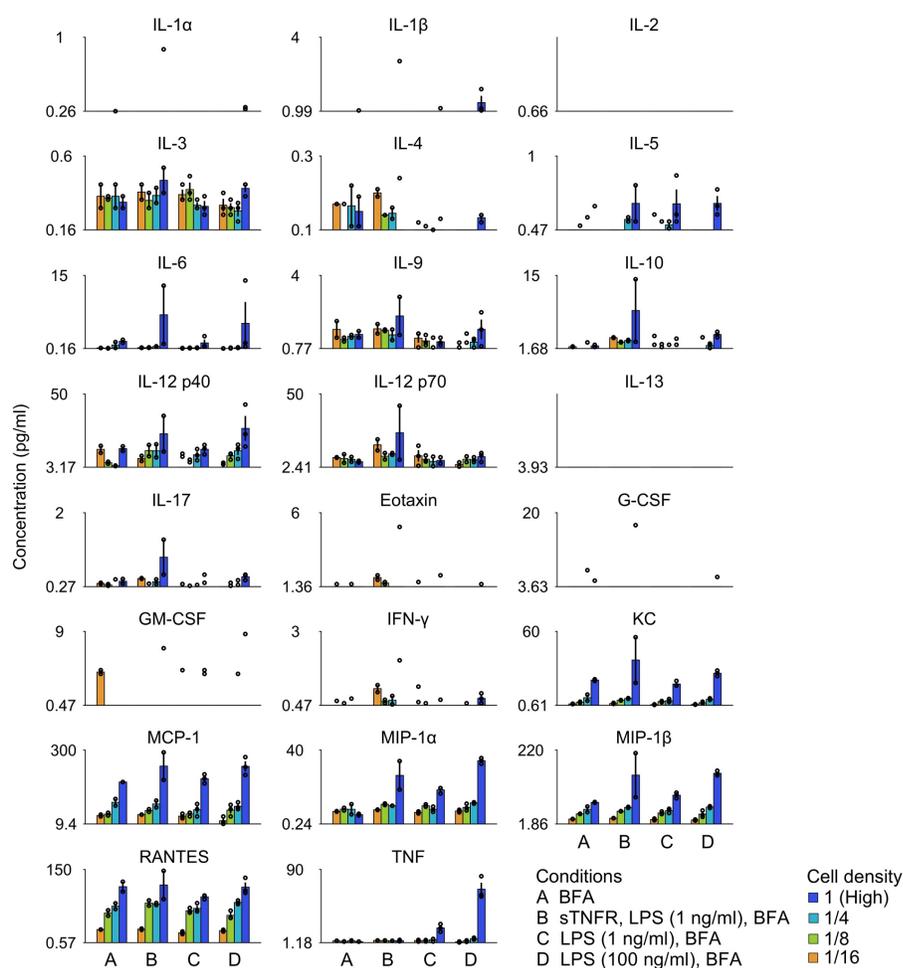


Figure A2.2. Secreted factors. Secreted factors in the supernatant were evaluated by multiplexed assay (23 analytes). These cell culture supernatants correspond to the experiment in **Figure 2.5**. Bars graphs represent the mean of the biological replicates and S.E.M. The minimum value on each y-axis is the observed lower limit of detection of the assay. For each analyte-density-treatment combination, if at least one biological replicate was below the limit of detection, then the mean for the set of replicates was considered to be below the limit.

Table A2.1. State variables

# ^a	Names	Descriptions	Initial values (a.u.)
\dot{X}_1	TLR4	Inactive TLR4	0.1
\dot{X}_2	TLR4*	Active TLR4	0
\dot{X}_3	TNFR	Inactive TNFR	0.1
\dot{X}_4	TNFR*	Active TNFR	0
\dot{X}_5	IKKK	Inactive	0.1
\dot{X}_6	IKKK*	Active	0
\dot{X}_7	IKK	Inactive	0.1
\dot{X}_8	IKK*	Active	0
\dot{X}_9	NFkBm	<i>Rela</i> mRNA	varies by cell ^a
\dot{X}_{10}	NFkBc	NF- κ B cytoplasmic	0
\dot{X}_{11}	NFkBn	NF- κ B nuclear	0
\dot{X}_{12}	IkBm	<i>Ikba</i> mRNA	0
\dot{X}_{13}	IkBc	IkB cytoplasmic	0
\dot{X}_{14}	IkBn	IkB nuclear	0
\dot{X}_{15}	NFkB_IkBc	NF- κ B-IkB cytoplasmic	varies by cell ^b
\dot{X}_{16}	NFkB_IkBn	NF- κ B-IkB nuclear	0
\dot{X}_{17}	IKK_IkB	IKK-IkB	0
\dot{X}_{18}	NFkB_IKK_IkB	NF- κ B-IKK-IkB	0
\dot{X}_{19}	Stabilizing_regulation	Stabilizing regulation	1 without IL-10; 0 with IL-10
\dot{X}_{20}	Destabilizing_regulation	Destabilizing regulation	0 without IL-10; max. with IL-10
\dot{X}_{21}	mCherrym	<i>mCherry</i> mRNA	0
\dot{X}_{22}	mCherry	mCherry protein	0
\dot{X}_{23}	mCherryf	mCherry protein folded	0
\dot{X}_{24}	Tnfm	<i>Tnf</i> mRNA	0
\dot{X}_{25}	TNF	TNF protein	0
\dot{X}_{EC1} ^c	TNFpool	Extracellular TNF protein	0
\dot{X}_{EC2} ^c	LPS	LPS stimulus	1 (corresponds to 100 ng/ml)

^a Variables #1 through 25 are intracellular, and the last two variables are extracellular. ODEs for one cell are in **Table A2.4**. A system of 30 cells has $25 \times 30 + 2 = 752$ ODEs.

^b Estimated steady-state initial value for *Rela* mRNA: $NFkBm = 0.007 * NFkB_IkBc$

^c Initial values of the inactive cytoplasmic complex, based on experimental quantification of initial EGFP-RelA from confocal microscopy, and normalized so that the mean is 0.1.

$NFkB_IkBc = [0.0022, 0.0046, 0.0065, 0.0068, 0.0085, 0.0146, 0.0355, 0.0437, 0.0460, 0.0465, 0.0619, 0.0667, 0.0668, 0.0676, 0.0826, 0.0940, 0.1023, 0.1151, 0.1180, 0.1320, 0.1417, 0.1533, 0.1543, 0.1594, 0.1689, 0.1775, 0.2052, 0.2216, 0.2418, 0.2540]$

Imputed values for low density:

$NFkB_IkBc = [0.0016, 0.0034, 0.0049, 0.0051, 0.0064, 0.0109, 0.0266, 0.0328, 0.0180, 0.0299, 0.0130, 0.0305, 0.0326, 0.0058, 0.0069, 0.0349, 0.0500, 0.0507, 0.0620, 0.0705, 0.0767, 0.0885, 0.0990, 0.1063, 0.1157, 0.1267, 0.1331, 0.1539, 0.1662, 0.1813]$

^d The two extracellular (EC) variables are appended to the end of the system of ODEs for one or more cells.

Table A2.2. Parameters

Names	Descriptions	Values	Units ^a	Sources
$k_{\text{synthesis_TLR4}}$	synthesis of TLR4	0.0185	$\mu\text{M h}^{-1}$	Calculated based on degradation rate
$k_{\text{activate_TLR4}}$	activation of TLR4	10	$\mu\text{M}^{-1} \text{h}^{-1}$	Estimated, to produce oscillatory translocations at 100 ng/ml LPS and lesser effects at lower doses
$k_{\text{deg_TLR4}}$	degradation of TLR4	0.185	h^{-1}	Estimated from a reduced model of Cheng et al. 2015 ⁹¹
$k_{\text{synthesis_TNFR}}$	synthesis of TNFR	0.0102	$\mu\text{M h}^{-1}$	Calculated based on degradation rate
$k_{\text{activate_TNFR}}$	activation of TNFR	^b 0.614 / nc	$\mu\text{M}^{-1} \text{h}^{-1}$	Fitted
$k_{\text{deg_TNFR}}$	degradation of TNFR	0.102	h^{-1}	From Werner et al. ⁶⁹ and used by Caldwell et al. ³¹
$k_{\text{TLR4_IKKK}}$	activation of IKKK by TLR4	0.588	$\mu\text{M}^{-1} \text{h}^{-1}$	Estimated in order to produce oscillatory translocations
$k_{\text{TNFR_IKKK}}$	activation of IKKK by TNFR	0.588	$\mu\text{M}^{-1} \text{h}^{-1}$	Assumed equal to $k_{\text{TLR4_IKKK}}$
$k_{\text{inactivate_IKKK}}$	inactivation of IKKK	150	h^{-1}	Cheng et al. 2015 ⁹¹
$k_{\text{activate_IKK}}$	activation of IKK	60000	$\mu\text{M}^{-1} \text{h}^{-1}$	Cheng et al. 2015 ⁹¹
$k_{\text{inactivate_IKK}}$	inactivation of IKK	19.4	h^{-1}	^c Fitted in the NF- κ B module
$k_{\text{assoc_IKK_IkbC}}$	association of IKK and IkbC	81	$\mu\text{M}^{-1} \text{h}^{-1}$	Cheng et al. 2015 ⁹¹
$k_{\text{assoc_IKK_NFkBikBc}}$	association of IKK and NF- κ B-IkbC	666	$\mu\text{M}^{-1} \text{h}^{-1}$	Cheng et al. 2015 ⁹¹
$k_{\text{deg_NFkBikBIKKc}}$	degradation of NF- κ B-Ikb-IKKc	432	h^{-1}	Cheng et al. 2015 ⁹¹
k_{tx}	max. transcription	1	$\mu\text{M h}^{-1}$	Assumed equal for all genes
$w_{\text{basaltx_NFkB}}$	basal transcription of NF- κ B mRNA	^d mean= 10^{-4}	μM	Estimated to maintain NF- κ B expression in absence of FBD
$w_{\text{NFkBtx_NFkB}}$	transcription of NF- κ B mRNA by NF- κ B	0.219	N/A	^c Fitted in the NF- κ B module
$w_{\text{IL10tx_FBD}}$	repression of transcription of NF- κ B mRNA by NF- κ B, by IL-10	Fitted: 7.27	μM	^c Fitted in the NF- κ B module
$k_{\text{deg_NFkBm}}$	degradation of NF- κ B mRNA	0.14	h^{-1}	Assumed 5 h half-life
k_{tl}	translation	15	h^{-1}	Cheng et al. 2015 ⁹¹
$k_{\text{import_NFkBc}}$	import of NF- κ B from the cytoplasm to the nucleus	324	h^{-1}	Cheng et al. 2015 ⁹¹
$k_{\text{export_NFkBn}}$	export of NF- κ B from the nucleus to the cytoplasmic	1.5	h^{-1}	Estimated to produce oscillatory translocations
$k_{\text{assoc_NFkB_Ikb}}$	association of NF- κ B and Ikb	1800	$\mu\text{M h}^{-1}$	Cheng et al. 2015 ⁹¹
$k_{\text{assoc_IkbIKKc_NFkBc}}$	association of Ikb-IKKc and NF- κ Bc	1800	$\mu\text{M h}^{-1}$	Cheng et al. 2015 ⁹¹
$k_{\text{deg_NFkB}}$	degradation of NF- κ B	0.36	h^{-1}	Matched to the estimate for Ikb in the resting state (where IKK is inactive); Sung et al. ⁷³
$w_{\text{NFkBtx_Ikb}}$	transcription of Ikb mRNA by NF- κ B	6.53	N/A	^c Fitted in the NF- κ B module
$k_{\text{deg_Ikbm}}$	degradation of Ikb mRNA	2.1	h^{-1}	Cheng et al. 2015 ⁹¹
$k_{\text{import_IkbC}}$	import of IkbC from the cytoplasm to the nucleus	1.08	h^{-1}	Cheng et al. 2015 ⁹¹
$k_{\text{export_IkbN}}$	export of IkbN from the nucleus to the cytoplasmic	0.72	h^{-1}	Cheng et al. 2015 ⁹¹
$k_{\text{deg_Ikb}}$	degradation of Ikb	4.2	h^{-1}	Cheng et al. 2015 ⁹¹
$k_{\text{export_NFkBikBn}}$	export of NF- κ B-IkbN	49.7	h^{-1}	Kalita et al. 2011 ⁹⁰
$\tau_{\text{u stabilize}}$	timing of stabilizing regulation ending	0.234	h^{-1}	Fitted
$\tau_{\text{u IKKK}}$	contribution of IKKK activity to stabilizing regulation	$3.96 * 10^5$	μM^{-1}	Fitted
$c_{\text{stabilize}}$	threshold IKKK activity for stabilizing regulation	$3.92 * 10^{-5}$	μM	Estimated; involved in stabilizing regulation, and is based on the values of the IKKK* state variable
$\tau_{\text{u destabilize}}$	timing of destabilizing regulation starting	7.3	h	Fitted
$w_{\text{NFkBtx_Tnf}}$	transcription of Tnf mRNA & mCherry mRNA by NF- κ B	1.99	N/A	Fitted
$k_{\text{deg_mCherrym}}$	degradation of mCherry mRNA	2.1	h^{-1}	Assumed equal to half-life of other short-lived mRNAs
$k_{\text{deg_mCherry}}$	degradation of mCherry	0.5	h^{-1}	Assumed half-life near 1 h; estimated based on avg. time to peak expression

K_{mature}	maturation (folding) of mCherry	0.7	h^{-1}	Assumed half-life for maturation of 1 h
$K_{\text{deg_Tnfm}}$	degradation of Tnf mRNA	2.1	h^{-1}	Assumed equal to half-life of other short-lived mRNAs
$W_{\text{stabilize}}$	effect of stabilizing regulation	1.7	μM^{-1}	Fitted
$W_{\text{maxdestabilize}}$	maximum stabilizing regulation	3	N/A	Estimated constant involved in destabilizing regulation
$W_{\text{destabilize}}$	effect of destabilizing regulation	0.747	μM^{-1}	Fitted
$K_{\text{secretion}}$	TNF secretion	4.17	h^{-1}	Fitted
$K_{\text{deg_TNF}}$	degradation of TNF (intracellular)	0.7	h^{-1}	^e See note
r_1	max. relative population density	1.99	RCD	^f Separately fitted
r_2	population growth rate	0.0365	h^{-1}	^f Separately fitted
$K_{\text{deg_TNFpool}}$	degradation of TNF pool	0.27	h^{-1}	Maiti et al. 2015 ⁸⁴
$K_{\text{deg_LPS}}$	degradation of LPS (in media)	0.058	h^{-1}	Assumed half-life of 12 h

- ^a Units are: concentration (μM), time in hours (h), relative cell density (RCD), number of cells (nc), or no units (N/A).
- ^b The parameter for TNFR activation was scaled by dividing by the number of cells. For $nc=30$, the value is $0.0206 \mu\text{M}^{-1} \text{h}^{-1}$.
- ^c Four parameters were first fit in the NF- κB module; the other eight were fit in the full model.
- ^d Varies by cell; equal to 0.001 times the initial concentration of total cellular NF- κB .
- ^e We found that given the data available, it would be challenging to estimate values for parameters representing both possible fates of the TNF protein (degradation and secretion). Since degradation is first order and has no downstream consequence in the model, whereas secretion does have a consequence in the activation of TNFR signaling, we chose to prioritize the estimate for secretion. To constrain the search for the free parameter for secretion, we set the value for intracellular degradation to that also used for extracellular degradation (which for simulated outcomes should affect the histograms in **Figures 2.3–2.4** in a linearly proportional manner, and not affect the conclusions).
- ^f Separately fitted by constrained nonlinear least squares optimization, prior to fitting the ODE model.

Table A2.3. Stimulus-specific and perturbation-specific parameters

Names	Descriptions	Values	Units
d_{LPS}	LPS treatment at 0 h	1 with; 0 without	N/A
d_{sTNFR}	sTNFR pre-treatment at -1 h	1 with; 0 without	N/A
d_{IL10}	IL-10 pre-treatment at -12 h	1 with; 0 without	N/A
τ_{uIL10_1}	time when transcription of Tnf begins to decrease due to IL-10	6	h
τ_{uIL10_2}	time when transcription of Tnf effectively ceases due to IL-10	18	h
τ_{uBFA}	time of BFA treatment	varies with; infinity without	h
$\tau_{\text{u density}}$	parameter for initial cell density at time of plating (at -36 h)	-0.247 high density; 62.6 low density	h

Table A2.4. Ordinary differential equations

#	ODEs	Reactions
\dot{X}_1	$k_{\text{synthesis_TLR4}}$ $- d_{\text{LPS}} * k_{\text{activate_TLR4}} * [\text{TLR4}] * [\text{LPS}]$ $- k_{\text{deg_TLR4}} * [\text{TLR4}]$	synthesis: $0 \rightarrow \text{TLR4}$ activation: $\text{TLR4} \rightarrow \text{TLR4}^*$ degradation: $\text{TLR4} \rightarrow 0$
\dot{X}_2	$d_{\text{LPS}} * k_{\text{activate_TLR4}} * [\text{TLR4}] * [\text{LPS}]$ $- k_{\text{deg_TLR4}} * [\text{TLR4}^*]$	activation: $\text{TLR4} \rightarrow \text{TLR4}^*$ degradation: $\text{TLR4}^* \rightarrow 0$
\dot{X}_3	$k_{\text{synthesis_TNFR}}$ $- k_{\text{activate_TNFR}} * [\text{TNFR}] * [\text{TNFpool}] * (1 - d_{\text{sTNFR}})$ $- k_{\text{deg_TNFR}} * [\text{TNFR}]$	synthesis: $0 \rightarrow \text{TNFR}$ activation: $\text{TNFR} \rightarrow \text{TNFR}^*$ degradation: $\text{TNFR} \rightarrow 0$
\dot{X}_4	$k_{\text{activate_TNFR}} * [\text{TNFR}] * [\text{TNFpool}] * (1 - d_{\text{sTNFR}})$ $- k_{\text{deg_TNFR}} * [\text{TNFR}^*]$	activation: $\text{TNFR} \rightarrow \text{TNFR}^*$ degradation: $\text{TNFR}^* \rightarrow 0$
\dot{X}_5	$- k_{\text{TLR4_IKKK}} * [\text{TLR4}^*] * [\text{IKKK}]$ $- k_{\text{TNFR_IKKK}} * [\text{TNFR}^*] * [\text{IKKK}]$ $+ k_{\text{inactivate_IKKK}} * [\text{IKKK}^*]$	activation: $\text{IKKK} \rightarrow \text{IKKK}^*$, via TLR4^* activation: $\text{IKKK} \rightarrow \text{IKKK}^*$, via TNFR^* inactivation: $\text{IKKK}^* \rightarrow \text{IKKK}$
\dot{X}_6	$k_{\text{TLR4_IKKK}} * [\text{TLR4}^*] * [\text{IKKK}]$ $+ k_{\text{TNFR_IKKK}} * [\text{TNFR}^*] * [\text{IKKK}]$ $- k_{\text{inactivate_IKKK}} * [\text{IKKK}^*]$	activation: $\text{IKKK} \rightarrow \text{IKKK}^*$, via TLR4^* activation: $\text{IKKK} \rightarrow \text{IKKK}^*$, via TNFR^* inactivation: $\text{IKKK}^* \rightarrow \text{IKKK}$
\dot{X}_7	$- k_{\text{activate_IKK}} * [\text{IKKK}^*] * [\text{IKK}]$ $+ k_{\text{inactivate_IKK}} * [\text{IKK}^*]$	activation: $\text{IKK} \rightarrow \text{IKK}^*$ inactivation: $\text{IKK}^* \rightarrow \text{IKK}$
\dot{X}_8	$k_{\text{activate_IKK}} * [\text{IKKK}^*] * [\text{IKK}]$ $- k_{\text{inactivate_IKK}} * [\text{IKK}^*]$ $- k_{\text{assoc_IKK_IkBc}} * [\text{IKK}^*] * [\text{IkBc}]$ $- k_{\text{assoc_IKK_NFkB_IkBc}} * [\text{IKK}^*] * [\text{NFkB_IkBc}]$ $+ k_{\text{deg_NFkB_IkBc_IKKc}} * [\text{NFkB_IKK_IkB}]$	activation: $\text{IKK} \rightarrow \text{IKK}^*$ inactivation: $\text{IKK}^* \rightarrow \text{IKK}$ association: $\text{IKK}^* + \text{IkBc} \rightarrow \text{IKK-IkBc}$ association: $\text{IKK}^* + \text{NFkB-IkBc} \rightarrow \text{NFkB-IkB-IKKc}$ degradation: $\text{NFkB-IkB-IKKc} \rightarrow \text{IKK} + \text{NFkBc}$
\dot{X}_9	$k_{\text{tx}} * (W_{\text{basaltx_NFkB}} + d_{\text{LPS}} * t^3 * \exp(-t) / 6 * W_{\text{NFkBtx_NFkB}} * [\text{NFkBn}]) / (1 + d_{\text{LPS}} * t^3 * \exp(-t) / 6 * W_{\text{NFkBtx_NFkB}} * ([\text{NFkBn}] + W_{\text{IL10tx_FBD}} * d_{\text{IL10}}))$ $- k_{\text{deg_NFkBm}} * [\text{NFkBm}]$	transcription: $0 \rightarrow \text{NFkBm}$ degradation: $\text{NFkBm} \rightarrow 0$
\dot{X}_{10}	$k_{\text{tl}} * [\text{NFkBm}]$ $- k_{\text{import_NFkBc}} * [\text{NFkBc}]$ $+ k_{\text{export_NFkBn}} * [\text{NFkBn}]$ $- k_{\text{assoc_NFkB_IkBc}} * [\text{NFkBc}] * [\text{IkBc}]$ $- k_{\text{assoc_IkBc_NFkBc}} * [\text{NFkBc}] * [\text{IKK_IkB}]$ $+ k_{\text{deg_NFkB_IkBc_IKKc}} * [\text{NFkB_IKK_IkB}]$ $- k_{\text{deg_NFkB}} * [\text{NFkBc}]$	translation: $0 \rightarrow \text{NFkBc}$ import: $\text{NFkBc} \rightarrow \text{NFkBn}$ export: $\text{NFkBn} \rightarrow \text{NFkBc}$ association: $\text{NFkBc} + \text{IkBc} \rightarrow \text{NFkB-IkBc}$ association: $\text{IkB-IKKc} + \text{NFkBc} \rightarrow \text{NFkB-IkB-IKKc}$ degradation: $\text{NFkB-IkB-IKKc} \rightarrow \text{IKK}^* + \text{NFkBc}$ degradation: $\text{NFkBc} \rightarrow 0$
\dot{X}_{11}	$k_{\text{import_NFkBc}} * [\text{NFkBc}]$ $- k_{\text{export_NFkBn}} * [\text{NFkBn}]$ $- k_{\text{assoc_NFkB_IkBn}} * [\text{NFkBn}] * [\text{IkBn}]$ $- k_{\text{deg_NFkB}} * [\text{NFkBn}]$	import: $\text{NFkBc} \rightarrow \text{NFkBn}$ export: $\text{NFkBn} \rightarrow \text{NFkBc}$ association: $\text{NFkBn} + \text{IkBn} \rightarrow \text{NFkB-IkBn}$ degradation: $\text{NFkBc} \rightarrow 0$
\dot{X}_{12}	$k_{\text{tx}} * W_{\text{NFkBtx_IkB}} * ([\text{NFkBn}]) / (1 + W_{\text{NFkBtx_IkB}} * [\text{NFkBn}])$ $- k_{\text{deg_IkBm}} * [\text{IkBm}]$	transcription: $0 \rightarrow \text{IkBm}$ degradation: $\text{IkBm} \rightarrow 0$
\dot{X}_{13}	$k_{\text{tl}} * [\text{IkBm}]$ $- k_{\text{import_IkBc}} * [\text{IkBc}]$ $+ k_{\text{export_IkBn}} * [\text{IkBn}]$ $- k_{\text{deg_IkB}} * [\text{IkBc}]$ $- k_{\text{assoc_NFkB_IkB}} * [\text{NFkBc}] * [\text{IkBc}]$ $- k_{\text{assoc_IKK_IkBc}} * [\text{IKK}^*] * [\text{IkBc}]$	translation: $\text{IkBm} \rightarrow \text{IkBc}$ import: $\text{IkBc} \rightarrow \text{IkBn}$ export: $\text{IkBn} \rightarrow \text{IkBc}$ degradation: $\text{IkBc} \rightarrow 0$ association: $\text{NFkBc} + \text{IkBc} \rightarrow \text{NFkB-IkBc}$ association: $\text{IKK}^* + \text{IkBc} \rightarrow \text{IKK-IkBc}$

\dot{X}_{14}	$k_{import_IkBc} * [IkBc]$ $- k_{export_IkBn} * [IkBn]$ $- k_{deg_IkB} * [IkBn]$ $- k_{assoc_NFkB_IkB} * [NFkBn] * [IkBn]$	import: $IkBc \rightarrow IkBn$ export: $IkBn \rightarrow IkBc$ degradation: $IkBn \rightarrow 0$ association: $NFkBn + IkBn \rightarrow NFkB-IkBn$
\dot{X}_{15}	$k_{export_NFkB_IkBn} * [NFkB_IkBn]$ $+ k_{assoc_NFkB_IkB} * [NFkBc] * [IkBc]$ $- k_{assoc_IKK_NFkB_IkBc} * [IKK^*] * [NFkB_IkBc]$	export: $NFkB-IkBn \rightarrow NFkB-IkBc$ association: $NFkBc + IkBc \rightarrow NFkB-IkBc$ association: $IKK^* + NFkB-IkBc \rightarrow NFkB-IKK-IkBc$
\dot{X}_{16}	$- k_{export_NFkB_IkBn} * [NFkB_IkBn]$ $+ k_{assoc_NFkB_IkB} * [NFkBn] * [IkBn]$	export: $NFkB-IkBn \rightarrow NFkB-IkBc$ association: $NFkBn + IkBn \rightarrow NFkB-IkBn$
\dot{X}_{17}	$k_{assoc_IKK_IkBc} * [IKK^*] * [IkBc]$	association: $IKK^* + IkBc \rightarrow IkB-IKKc$
	$- k_{assoc_IkB_IKKc_NFkBc} * [IKK_IkB] * [NFkBc]$	association: $IkB-IKKc + NFkBc \rightarrow NFkB-IkB-IKKc$
\dot{X}_{18}	$k_{assoc_IKK_NFkB_IkBc} * [IKK^*] * [NFkB_IkBc]$ $+ k_{assoc_IkB_IKKc_NFkBc} * [NFkBc] * [IKK_IkB]$ $- k_{deg_NFkB_IKK_IkB} * [NFkB_IKK_IkB]$	association: $NFkB-IkBc + IKK^* \rightarrow NFkB-IkB-IKKc$ association: $IkB-IKKc + NFkBc \rightarrow NFkB-IkB-IKKc$ degradation: $NFkB-IkB-IKKc \rightarrow IKK^* + NFkBc$
\dot{X}_{19}	$-(1 - d_{IL10}) * \tau_{Ustabilize} * [Stabilizing_regulation] * 1 / (1 + \tau_{IKK} * \max(0, [IKK^*] - C_{stabilize}))$	time-dependent function
\dot{X}_{20}	$1 - (1 - d_{IL10}) * (1 - 1 / (1 + \exp(-t - \tau_{Udestabilize})))$	time-dependent function
\dot{X}_{21}	$d_{LPS} * (k_{tx} * W_{NFkBtx_Tnf} * [NFkBn] / (1 + W_{NFkBtx_Tnf} * [NFkBn])) * (1 - d_{IL10} * (1 - (1 - \max(t - \tau_{U_{IL10_1}, 0}) / (\tau_{U_{IL10_2}} - \tau_{U_{IL10_1}})) * (1 - 1 / (1 + \exp(-99 * (t - \tau_{U_{IL10_2}}))))))$ $- k_{deg_mCherry} * [mCherry]$	transcription: $0 \rightarrow mCherry$; delayed effect of IL-10 pre-treatment degradation: $mCherry \rightarrow 0$
\dot{X}_{22}	$k_t * [mCherry]$ $- k_{deg_mCherry} * [mCherry]$	translation: $0 \rightarrow mCherry$ degradation: $mCherry \rightarrow 0$
\dot{X}_{23}	$k_{mature} * [mCherry]$ $- k_{deg_mCherry} * [mCherry]$	maturation: $mCherry \rightarrow mCherry$ degradation: $mCherry \rightarrow 0$
\dot{X}_{24}	$d_{LPS} * (k_{tx} * W_{NFkBtx_Tnf} * [NFkBn] / (1 + W_{NFkBtx_Tnf} * [NFkBn])) * (1 - d_{IL10} * (1 - (1 - \max(t - \tau_{U_{IL10_1}, 0}) / (\tau_{U_{IL10_2}} - \tau_{U_{IL10_1}})) * (1 - 1 / (1 + \exp(-99 * (t - \tau_{U_{IL10_2}}))))))$ $- k_{deg_Tnfm} * [Tnfm] * (1 / (1 + w_{stabilize} * [Stabilizing_regulation]) + (w_{maxdestabilize} - 1) * w_{destabilize} * [Destabilizing_regulation] / (1 + w_{destabilize} * [Destabilizing_regulation]))$	transcription: $0 \rightarrow TNFm$; delayed effect of IL-10 pre-treatment degradation: $TNFm \rightarrow 0$
\dot{X}_{25}	$k_t * [Tnfm] * ((1 + w_{destabilize} * [Destabilizing_regulation]) / (1 + w_{maxdestabilize} * w_{destabilize} * [Destabilizing_regulation]))$ $- k_{secretion} * [TNF] * (1 - (1 + \exp(-99 * (t - \tau_{BFA})))^{-1})$ $- k_{deg_TNF} * [TNF]$	translation: $0 \rightarrow TNF$ secretion: $TNF \rightarrow TNFpool$ degradation: $TNF \rightarrow 0$
\dot{X}_{EC1}	$(r1 / (1 + \exp(-r2 * (t - \tau_{Udensity})))) * (1 - (1 + \exp(-99 * (t - \tau_{BFA})))^{-1}) * k_{secretion} * SUM([TNF])$ $- k_{deg_TNFpool} * [TNFpool]$	secretion: $TNF \rightarrow TNFpool$ degradation: $TNFpool \rightarrow 0$
\dot{X}_{EC2}	$- k_{deg_LPS} * [LPS]$	degradation: $LPS \rightarrow 0$

APPENDIX 3. Modeling gene expression heterogeneity

The original methodology for this appendix was described and published in:

Hartfield R.M.*, Schwarz K.A.*, Muldoon J.J.*, Bagheri N., Leonard J.N. Multiplexing engineered receptors for multiparametric evaluation of environmental ligands. *ACS Synthetic Biology* **6**, 2042–2055 (2017).⁵⁴

*Equal contributions

refined and published in:

Donahue P.S., Draut J.W.*, Muldoon J.J.*, Edelstein H.I.*, Bagheri N., Leonard J.N. The COMET toolkit for composing customizable genetic programs in mammalian cells. *Nat Commun* **7**, 779 (2020).¹²⁹ *Equal contributions

and modified and is in preparation as:

Muldoon J.J., Kandula V., Hong M., Donahue P.S., Boucher J.D., Bagheri N., Leonard J.N. Design-driven engineering of mammalian genetic programs. *In preparation*.¹⁶¹

Overview: In developing models to investigate genetic programs, I account for two phenomena: cell heterogeneity using a statistical model (**Figure A3.1**), and gene regulation using a dynamical model.

Statistical model: Heterogeneity is represented by simulating genetic programs in a way that resembles their outcomes in cells, which vary in the expression of the components. The *in silico* population (**Z**) is an $N \times P$ matrix, where N is the number of cells ($n = 1:N$, for $N = 200$) and P is the number of plasmids ($p = 1:P$). Components that are encoded on separate plasmids are assigned separate columns. For example, the ZF1a gene is assigned one column and the reporter gene is assigned another column. **Z** is generated using the constrained sampling method²⁵³ using the following steps.

1. Specify parameters for the target marginal distribution of gene expression. Based on flow cytometry measurements of constitutively expressed fluorescent proteins from co-transfected plasmids (and FACS buffer-based harvest in flow cytometry preparation), the characteristic distribution for each protein was log-bimodal Gaussian, described by the parameters $\mu_1 = 1.95$, $\sigma_1 = 0.3$, $\mu_2 = 3.4$, and $\sigma_2 = 0.6$ a.u. on a log₁₀-scaled axis²⁵³. (Different parameters are used for trypsin-based harvest¹⁶¹.)

2. Specify a target correlation coefficient to model gene expression from co-transfected plasmids. A Pearson correlation of $r = 0.8$ was used based on observed correlations.

3. Based on the target correlation, specify a lower bound and upper bound of acceptable values. Values should be chosen that are close to the target, such as 0.765 and 0.835.

4. Generate a joint distribution using the parameters for the marginal distribution and the target correlation coefficient. This output is a candidate $N \times P$ matrix for population variation. Distributions can be generated using the multivariate normal random number generator in MATLAB.

5. Compute the correlation coefficient matrix ($P \times P$).

6. While any non-diagonal entries in the correlation coefficient matrix are outside of the range of acceptable values specified in step 3, repeat steps 4 and 5. (The time required to run this step greatly increases with the value of P .)

7. For the accepted matrix, normalize the values in each column to a mean of one to obtain the population matrix **Z**. It is useful to plot the generated distributions and correlations to confirm resemblance to the target outcomes.

Implementation: Dynamical models are run in a for-loop for each cell in a simulated population, and mean end-point simulated reporter protein level is calculated. Some figures use a single-cell (homogeneous) model; these cases forgo the heterogeneity for-loop and instead correspond to the mean-transfected cell, which is a scenario for the average protein expression from each transfected plasmid.

Each row (cell) and column (plasmid) in \mathbf{Z} is a scalar for the relative expression of a gene. The z value is used as a multiplier on the production term for each RNA species in dynamical models.

$$z_{i,p} \cdot k_{\text{tx}} \cdot \text{Species}_{\text{RNA}} \quad (\text{A3.1})$$

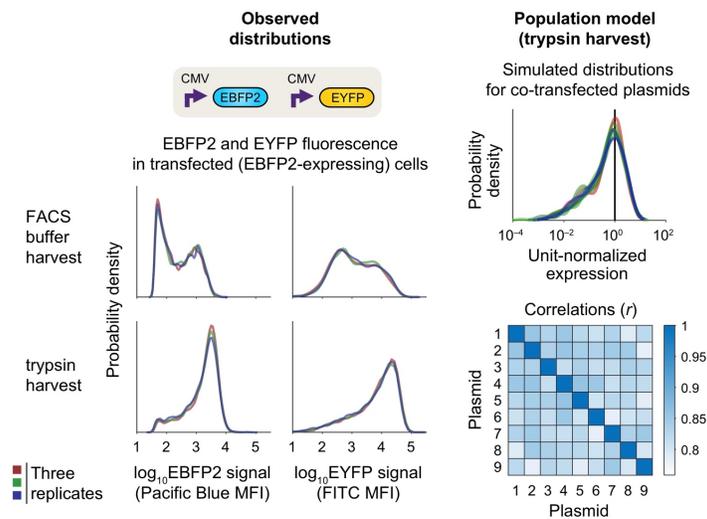


Figure A3.1. Distribution of gene expression in a population. A statistical model matching observed variation in gene expression was used to simulate heterogeneous cell populations. The flow cytometric distributions (collected in the same experiment) show EBFP2 signal and EYFP signal after gating on EBFP2-expressing (transfected) cells. The genetic programs study uses a trypsin harvest, and the MESA multiplexing study and COMET study use a FACS buffer harvest. The right panels show unit-normalized simulated distributions of gene expression and simulated correlations for co-expressed genes on transfected plasmids, all based on the generated population matrix for a trypsin harvest. Relative expression from each simulated plasmid species is similarly distributed across cells (upper right), and expression across plasmid species is similarly correlated ($r \sim 0.8$) within cells (lower right).

APPENDIX 4. Multiplexing receptors

A version of this appendix was previously published as:

Hartfield R.M.*, Schwarz K.A.*, Muldoon J.J.*, Bagheri N., Leonard J.N. Multiplexing engineered receptors for multiparametric evaluation of environmental ligands. *ACS Synthetic Biology* **6**, 2042–2055 (2017).⁵⁴

*Equal contributions

Summary: In this study, we investigated whether MESA receptors could be multiplexed. My colleagues developed hybrid promoters that exhibited AND gate activation by two TFs (tTA and Gal4) and evaluated these promoters when paired with receptors. The multiplexed system responded differently to cues applied individually and in combination, but the same synergy was not observed as when the promoters were characterized with soluble TFs. To improve our understanding of how these genetic components interface, I developed a mechanistic model incorporating the experimental observations. The analysis highlighted key factors that affect the receptors and promoters and enabled an in silico exploration of potential modifications towards improved performance. The methods and conclusions that I arrived at through this study represent an important precursor to several subsequent investigations^{129,161,198}.

Background: Engineered receptors have proven useful for building cell functions. For example, using chimeric antigen receptors (CARs), which induce signaling downstream of the T cell receptor, multiplexing strategies have been developed to achieve NOT logic as a safety switch²⁵⁴, or AND logic²⁵⁵ to reduce off-target activation. Layered AND logic cascades have also been used for activation precision²⁵⁶. Other non-receptor Boolean logic strategies include using multilayer transcriptional cascades²⁵⁷⁻²⁶⁰, hybrid promoters with multiple TF binding sites^{261,262}, post-transcriptional or post-translational regulation^{141,186,263}, DNA recombinases²⁶⁴⁻²⁶⁶, or distributing tasks across cells^{267,268}. Computational approaches have proven important for identifying effective designs, through case-by-case studies^{269,270} and tools for automated design^{153,262,270}. Going forward, receptors that perform logical operations could support numerous applications²⁷¹, and these innovations may be facilitated by integrating experimental and computational methods. Towards this goal, we investigated how and whether two MESA receptors could be multiplexed with their outputs converging at the same promoter.

Experimental investigation: As a strategy for multiplexing, my colleagues investigated a single-layer transcriptional gate (**Figure A4.1**) by designing promoters that could be activated by two TFs (tTA and Gal4) that can drive transcription from reporter constructs comprising 5–7 repeated DNA motifs (TetO and UAS, respectively) upstream of a minimal promoter. The hypothesis was that since multiple TF molecules must be recruited to induce transcription^{132,272}, some hybrid promoters containing both TetO and UAS sequences in certain patterns might exhibit AND logic by requiring the recruitment of both tTA and Gal4. As hypothesized, several of the promoters exhibited AND gate behavior when tested with and without

each TF co-transfected in HEK293FT cells. Two of these promoters (named H1 and H2) were carried forward and evaluated with: the two TFs at varied doses; one of two receptors (for sensing VEGF¹⁹⁷ and releasing tTA, and for sensing rapamycin¹⁶⁷ and releasing Gal4) at varied TC and PC doses and the soluble TF for the other receptor, and the reciprocal case, with and without ligand treatment; and both receptors with and without ligand treatment.

Ideal genetic circuit performance requires the output of upstream components to match the input requirements of downstream components²⁷³. We reasoned that this level-matching requires that the amounts of TFs released from receptor signaling match the amounts required for synergistic activation of the hybrid promoter. However, the results showed a lack of AND gate behavior when receptors were used in place of soluble TFs, suggesting that receptor signaling was insufficient to match promoter requirements. We hypothesized that modifications to receptor dosing could potentially improve outcomes. Given the large space of choices involved, and the challenge of gaining systematic understanding from empirical tuning alone, I next developed a computational model to facilitate interpretation of these observations, elucidate factors that affect receptor and hybrid promoter performance, and potentially identify an improved setup.

Model-guided characterization: Guiding our next approach was a key observation that only a small percentage of cells exhibited distinguishable promoter activation (were ON) with both TFs compared to with reporter alone. As a result, mean reporter expression was generally much greater for the ON subpopulation than for the whole population. The percentage of ON cells also generally varied with setup: experiments with two soluble TFs and the reporter (three components on three plasmids) had percentages that correlated with TF plasmid doses and reached about 20% at the highest doses; experiments with one

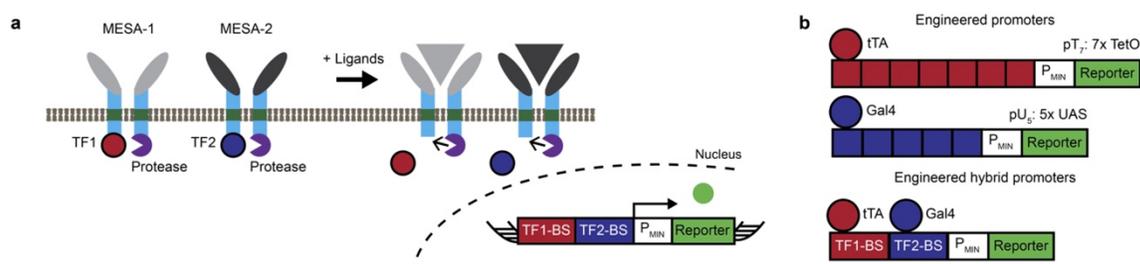


Figure A4.1 Multiplexed receptors and hybrid promoter implementation. **a** In the proposed strategy, two receptors each ligand-inducibly release a TF that enters the nucleus. **b** Hybrid promoters are regulated by both TF1 and TF2, and contain TetO (red) and UAS (blue) sites in different configurations (shown and evaluated in the published study). Single-TF promoters are illustrated for comparison.

receptor, one soluble TF, and the reporter (four components on four plasmids) and with two receptors and the reporter (five components on five plasmids) had successively lower percentages that still correlated with plasmid doses. I hypothesized that small ON percentages might arise from two types of intercellular variation: (i) each cell might receive a different number of molecules of each plasmid, which could restrict level-matching to a subset of cells, and/or (ii) cells might exhibit differences in transcription rate, translation rate, and/or transfection efficiency (the efficiency with which a plasmid, once taken up, enters the nucleus and becomes transcription-competent), which comprise sources of variation that are distinct from the amounts of plasmids received. Although these two types of variation cannot be readily distinguished in the data, we can treat them as together determining the effective initial conditions for the dose-dependent amount of each transfected gene that each cell can express. Based on these principles, I developed a model to investigate the effects of intercellular variation, and incorporated the combined effects of these two types of intercellular variation by assigning different amounts of each plasmid to each cell.

Heterogeneity is represented by modeling a population of cells in which key metrics of variability match those observed in experiments. I developed a method to generate *in silico* populations with the statistical features observed in a cotransfection experiment, for any specified number of plasmids (**Figure A4.2a, Appendix 3**); this representation is consistent with a recent analysis of how gene expression can be distributed in a cell population⁵³. From the *in silico* population, we can interrogate individual cells or calculate population-level metrics, such as mean reporter expression. From a principal component analysis (PCA), the first principal component explained 84–90% of the variation (depending on the number of cotransfected plasmids, from five to two, respectively) and corresponded to an axis along which plasmid amounts vary but their ratio is constant. Thus, in a cotransfection with equal amounts of two plasmids A and B, most cells will express similar amounts of each. Some cells will express more of plasmid A than plasmid B, or vice versa, which explains the remaining variation.

This result predicts that consolidating the genes encoding receptor chains and hybrid promoters onto fewer plasmids would not substantially affect intercellular variation or increase the percentage of ON cells. However, the system is sensitive to the number of different components. We consider the following idea: for a system of n different components, there exists an n -dimensional space that represents the amount of each component expressed per cell. In an experiment, each cell occupies a coordinate in this

space. There also exists in this space a functional region, which might be unknown a priori, corresponding to various combinations of amounts of each component that yield desirable ligand-inducible promoter activation, quantified by F.D. As n increases, such as by replacing one soluble TF for the two receptor chains, it is possible that the overlap in space between the functional region and the region occupied by a cell population will change. From this perspective, one way to frame the goal of implementing an engineered function that is robust to intercellular variation is to choose component doses (and system designs) that yield high overlap between the functional region and the region that is populated by cells. We define the

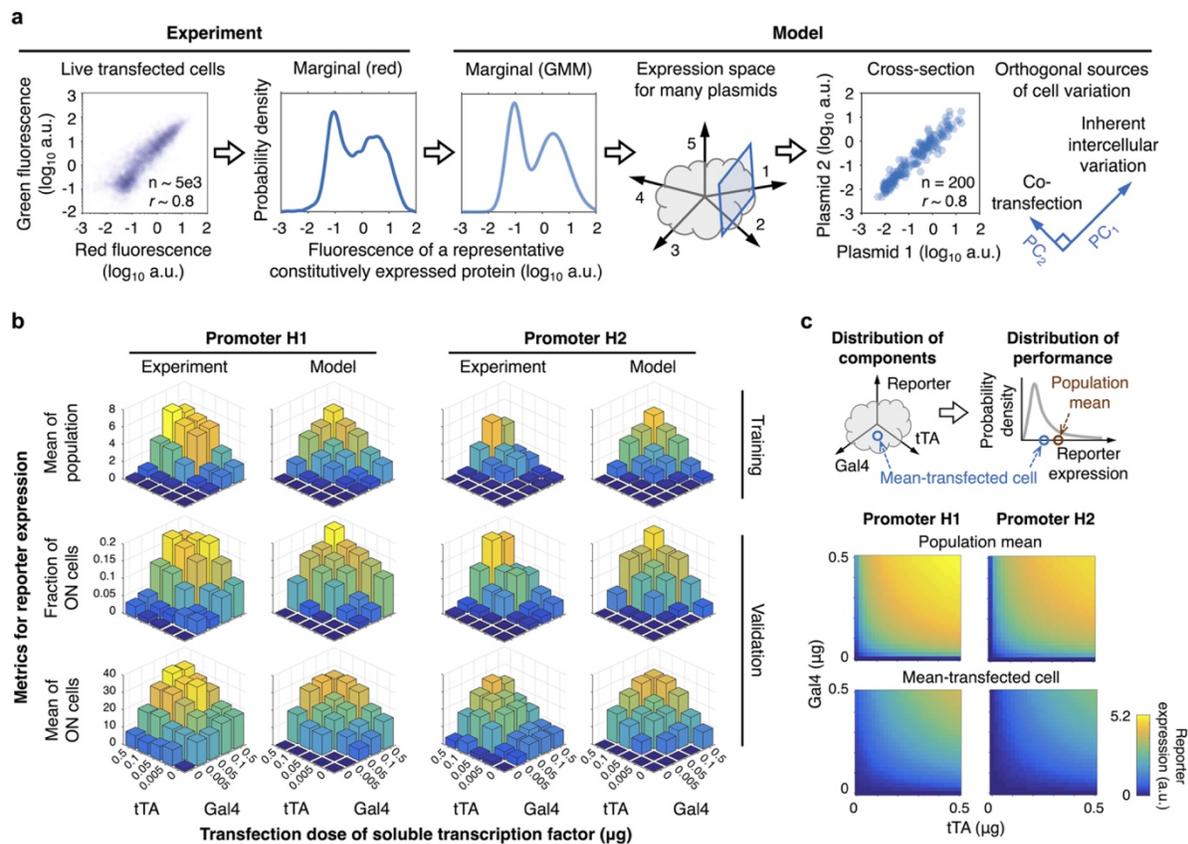


Figure A4.2. A model that accounts for cell variation to explain heterogeneous promoter activity. **a** A statistical model was formulated and trained on data to account for intercellular variation. The marginal distribution was modeled by a Gaussian mixture model. The resulting population exhibits the expected covariance between plasmids (inferred from constitutive expression of fluorescent proteins). PCA identified two sources of variability: the major contributor is inherent variation, and the minor contributor is variation due to cotransfection of multiple plasmids. The Pearson correlation coefficient (r) in the cross-section is 0.8 on a linear scale and 0.9 on a \log_{10} scale. **b** A dynamical model for TF expression and hybrid promoter activity was formulated and trained on mean average data for various tTA and Gal4 plasmid dose combinations. **c** The promoter model maps from a three-dimensional plasmid transfection distribution onto a one-dimensional reporter expression distribution. The distributions depict the reporter expression for hybrid promoters H1 and H2, when quantified for the population mean and the mean-transfected cell.

robustness of a system as the extent to which a performance metric (e.g., F.D.) is maintained as doses are varied. As robustness increases, more cells in a population exhibit the desired function. Robustness is therefore distinct from performance, which is the F.D. as calculated for a single cell in principle or as measured for a population mean average in practice. Having established a statistical and conceptual model for intercellular variation, I next addressed the mechanisms by which this system operates.

I started by developing a model for the hybrid promoters H1 and H2. For each promoter, inducible transcription was formulated as fractional activation f :

$$f = \frac{w_T \cdot \mathbf{tTA} + w_G \cdot \mathbf{Gal4} + w_T \cdot w_G \cdot \rho \cdot \mathbf{tTA} \cdot \mathbf{Gal4}}{1 + w_T \cdot \mathbf{tTA} + w_G \cdot \mathbf{Gal4} + w_T \cdot w_G \cdot \rho \cdot \mathbf{tTA} \cdot \mathbf{Gal4}} \quad (\text{A4.1})$$

Parameters w_T and w_G are responsiveness to tTA and Gal4, respectively; ρ is synergy; and tTA and Gal4 are simulated protein levels. Populations were initialized using the intercellular variation model, and parameters were fit to data (equations and parameter values are in the published study). The fitted values indicate that between the promoters, H1 is more responsive and H2 is more synergistic. H1 is 29x more responsive to Gal4 than tTA, H2 is 33x more responsive to Gal4 than tTA, the tTA response is 6.3x greater for H1 than H2, the Gal4 response is 5.5x greater for H1 than H2, and synergy is 21x greater for H2 than H1. Notably, although the calibration utilized only the mean reporter measurements, the model predicted trends in the observed heterogeneity, including for the fraction of ON cells and the mean reporter expression in this subpopulation (**Figure A4.2b**). These predictions validate our approach for describing intercellular variation and how variation affects system performance.

An important feature of the dose response landscapes was that the mean reporter expression was consistently greater than the reporter expression for a cell receiving the mean amount of each plasmid, i.e., the mean-transfected cell (**Figure A4.2c**). Cells that received greater than average plasmid amounts had even greater than average reporter expression, and reporter distributions were right-skewed. An implication is that the experimental characterization was in part driven by outlier cells that received high amounts of plasmids in each condition. Since characterization of not only the hybrid promoters, but also other genetic circuits, could be driven by outlier cells, we posit that a model-guided investigation of single-cell outcomes in which intercellular variation is represented explicitly might improve the interpretation of experimental results. Indeed, similar ideas have been used in characterizing other engineered genetic circuits^{274,275}.

Elucidating properties of multiplexed receptor performance: A model for multiplexed MESA

signaling was formulated at a level of granularity that includes interactions for various receptor complexes (Figure A4.3a). Key features based on prior knowledge, including receptor-ligand interactions and findings

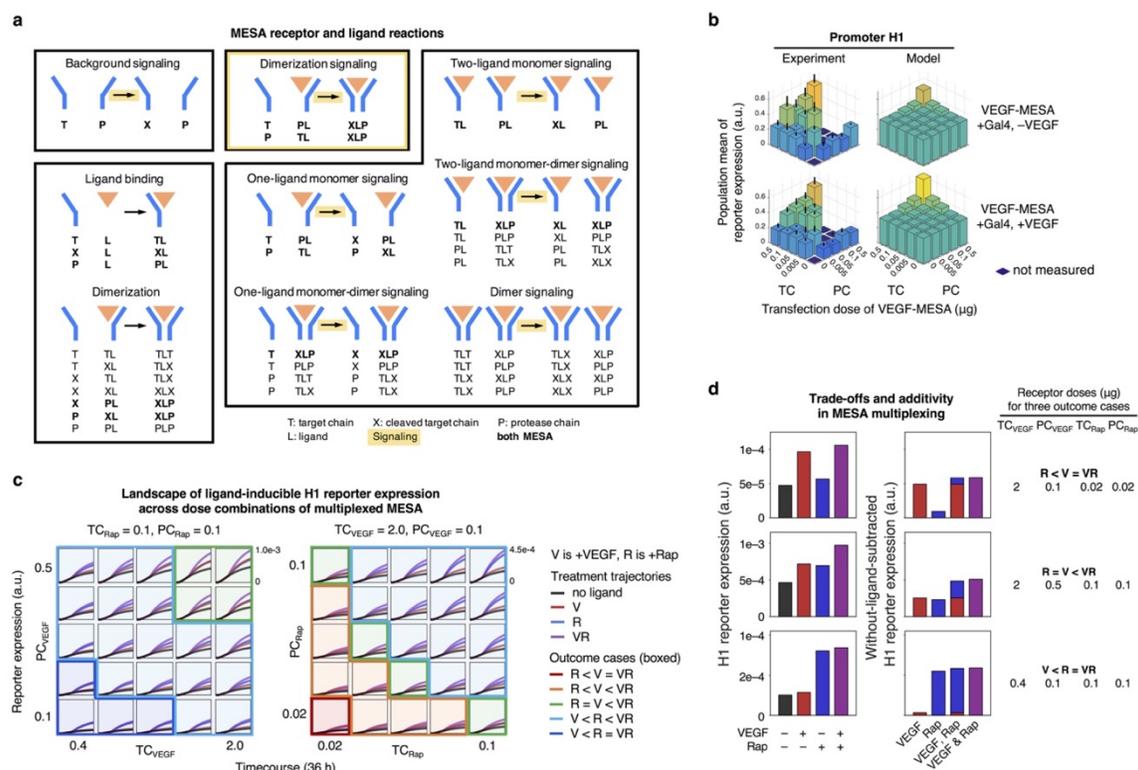


Figure A4.3. A dynamical model that links receptor signaling to promoter activity. **a** The illustration summarizes the 28 types of receptor reactions, which are grouped into nine categories (named) and four modalities (boxed). Reactions that occur for both MESA (Rap-MESA and VEGF-MESA) are bolded, categories that release a soluble TF are highlighted (yellow arrow), and the modality for canonical ligand-induced signaling is highlighted (yellow box). For the four modalities: (i) background signaling is the only one that occurs without ligand, (ii) ligand-binding and dimerization involve ligand but do not directly result in signaling, (iii) dimerization signaling is the canonical pathway, and (iv) the remaining categories involve, but are not directly mediated by, the ligand and are subject to crosstalk. **b** Data that were used to determine F.D. are compared to simulated outcomes for VEGF-MESA, soluble Gal4, and promoter H1, with and without VEGF treatment. Experimental data (originally in MFI) were linearly scaled to enable a more direct visual comparison with simulations. **c** Timecourse H1 reporter trajectories across TC and PC doses are shown for the mean-transfected cell with and without each ligand treatment (V, VEGF; R, Rap; VR, VEGF and Rap). In the left panel, VEGF-MESA doses are varied while Rap-MESA dose is constant, and in the right panel, Rap-MESA doses are varied while VEGF-MESA dose is constant. Simulations are grouped into five outcome cases (represented by box shading and outline color) based on the rank-ordered expression with each ligand treatment. **d** Three cases from **c** are examined in more detail. The left panel shows the absolute reporter expression, and the right panel shows ligand-induced reporter expression after the background (without ligand) is subtracted, to illustrate the additive ligand-induced response to these ligands. There is a trade-off for two-ligand-induced signaling, in which adjustments to the MESA plasmid dose that increase the F.D. for one ligand also decrease F.D. for the other ligand.

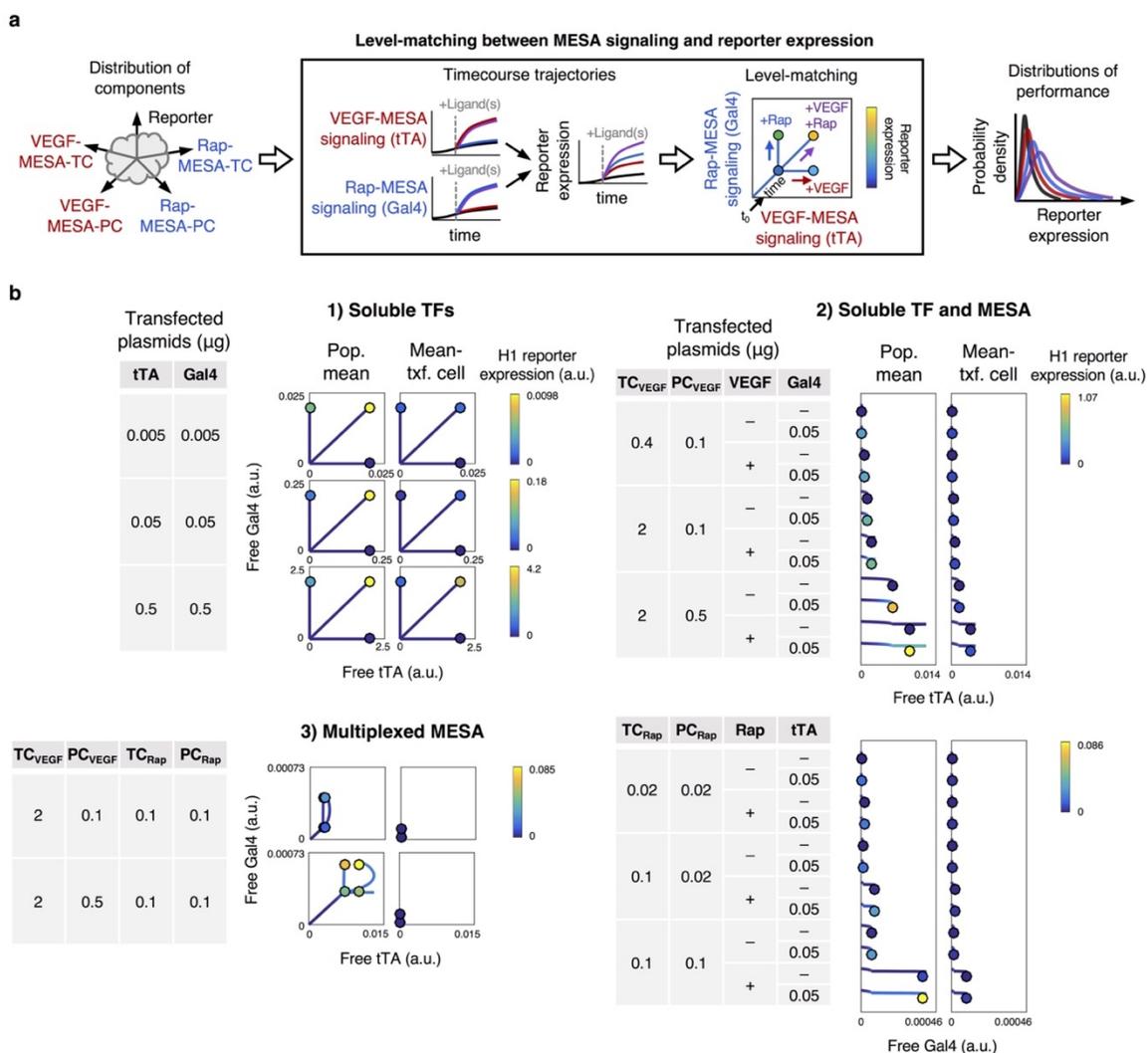


Figure A4.4. Level-matching between receptor signaling and the promoter. **a** Level-matching is depicted by yo-yo plots, which represent the trajectories of free TF and reporter variables without a time axis. Reporter expression across the timecourse (the string) and at the circled end point (the yo-yo, corresponding to the time of measurement) is color-coded. Each profile begins at the origin, proceeds through state space depending on plasmid doses and treatment with either, both, or no ligand, and ends at the circled coordinate. **b** Outcomes for each ligand treatment are shown for varied plasmid doses in three scenarios: (1) two soluble TFs, (2) one soluble TF and one receptor, and (3) two receptors. TFs are in comparable arbitrary units, and reporter expression is color-coded by reporter-specific a.u. Profiles in which the circled coordinate differs from the maximum coordinate along an axis indicate that the trajectory of the corresponding TF peaks and decreases during the timecourse. Diagonal lines indicate that the trajectories of both TFs are changing proportionately, curved lines indicate that both are changing and in a way that is not proportional, and vertical and horizontal lines indicate that one is changing while the other has reached a steady state. In the second scenario, only the TF released from MESA (and not the soluble TF) is plotted, and a slight downward curvature for the 36 h timecourse is shown for clarity. In the third scenario, ideal level-matching for AND gate functionality would be conferred by TF trajectories that lead to much higher reporter expression with both ligands compared to either or no ligand. In such a scenario, the upper-right yo-yo would be the only one of the four that is able to access the synergistic regime of the hybrid promoter.

from previous MESA studies^{167,197}, are: receptors are synthesized intracellularly, exocytosed to the cell surface, and degrade from both compartments; rapamycin ligand can diffuse intracellularly, but VEGF ligand cannot; VEGF-MESA can heterodimerize or homodimerize, and Rap-MESA can heterodimerize but not homodimerize; crosstalk in non-ligand mediated signaling is possible, because both receptors use the same PC protease and TC cleavage recognition sequence; and chains that recognize the same ligand can form stable dimers, but chains that recognize different ligands cannot. Calibration of this model to data provided estimates for the synthesis of receptors (relative to soluble proteins), background signaling, receptor degradation, ligand-binding to each receptor, and stable chain dimerization. The equations and parameter values are in the published study.

The model formulation makes relatively few assumptions and avoids overfitting peaks and valleys that deviate from the main observed trends. By smoothening (or discounting) individual outlier data points, such an analysis improves the overall interpretability of the dose-response landscape. When comparing observed vs. simulated ligand-inducible VEGF-MESA signaling in the presence of excess Gal4, the simulations were consistent with observed trends (**Figure A4.3b**). For a soluble TF and a MESA receptor (TC and PC) transfected at the same plasmid dose, the TF tended to contribute more to promoter activity. A model-guided interpretation for this outcome is that (i) rapid production of soluble TFs (relative to receptors) offsets rapid degradation, leading to high accumulation, and (ii) for the receptors, not all TCs are cleaved following ligand treatment. We next analyzed multiplexed receptors at different plasmid doses, and in doing so identified a trade-off: tuning receptor levels to increase the difference in two-ligand-induced reporter expression with respect to one ligand led to a decrease in the difference with respect to the other ligand. To highlight this effect, we grouped reporter trajectories into different outcome cases (**Figure A4.3c**). Examining three examples in more detail (**Figure 4.3d**) shows that across different receptor doses, reporter expression was additive: the two-ligand-induced increase in reporter expression above background equaled the sum of both single-ligand induced increases above background. The additivity indicates that promoter activity had a linear dependence on the TFs, rather than the synergy that would be expected for the AND gate promoters based upon characterizations with soluble TFs. These findings support the conclusion that the amounts of TFs released from receptors were below the amounts required to access promoter synergy, so the system exhibited an additive rather than synergistic response to ligands.

To better understand how multiplexed receptors interface with the hybrid promoter, we investigated the role of level-matching, i.e., the relationship between the amounts of TFs that are released by receptors under different ligand treatments, and the amounts that are required to activate promoter synergy only when both ligands are sensed. This relationship is challenging to observe directly, yet it is also key for explaining multiplexed receptor performance. Since nonlinear TF profiles from ligand-induced signaling depend on more components and are more complex than their linear counterparts from experiments with soluble TFs, we could not visualize level-matching with a dose-response landscape as in **Figure A4.2b**. Therefore, I opted to introduce a representation termed yo-yo plots, in which timecourse TF trajectories with various ligand combinations are represented without a time axis, and reporter expression across the timecourse (the string) and at the endpoint (the yo-yo) is color-coded (**Figure A4.4a**). Using this approach, I examined how free TFs were released over time under conditions in level-matching experiments (**Figure A4.4b**). The analysis confirmed the expectation that mean reporter outcomes were driven by outliers. Furthermore, it showed how outlier effects became magnified as each soluble TF was replaced by the two receptor chains. Across experiments, reporter expression for the population mean was consistently greater than for the mean-transfected cell, which poses a challenge to achieving level-matching by simply tuning component doses. Moreover, these analyses confirmed that across the MESA doses evaluated in experiments, the amounts of TFs released by receptors were below the levels required to induce synergistic activation of the H1 promoter. This case study illustrates how model-guided analysis of combined sensor and processor modules can identify quantitative properties that benefit and that limit system performance. This approach may also be inverted to guide the selection of components with properties that achieve performance goals.

Strategies to improve AND gate functionality: Given the bounds on AND gate performance, I examined whether alternative promoter and receptor properties could better achieve performance goals. In this prospective analysis, I considered parameter values that could reasonably be implemented, even if it is not yet possible to predict specific physical modifications that would result in specific new parameter values. One benefit of using a mechanistic model, compared to a more abstract formalism, is that the parameters do ultimately correspond to physical features that in future investigations could be tuned. For example, background signaling could be decreased by mutating the protease active site, protease cleavage

sequence, or transmembrane domain, or ligand binding could be modulated by mutating or replacing the ECD. Thus, an analysis of potentially realizable scenarios could guide subsequent investigations.

H1 served as a base case (promoter #1) for hypothetical promoters that vary in responsiveness to each TF and/or in synergy (**Figure A4.5a**). Properties of promoters (cases #3–9) are reported relative to promoter #2 (H1=), in which tTA responsiveness was set equal to Gal4 responsiveness. For a controlled comparison between promoters #3–9, I chose transcriptional weights that yielded dose-response landscapes that differ from the base case but resemble each other's maximal activity within the range of TF doses examined (**Figure A4.5b**). This range more closely matches the inferred range of TFs released from signaling (**Figure A4.4**), and synergistic activation for the new promoters occurs within this range.

To assess how each promoter affects MESA multiplexing, independent of a single-dose-specific implementations, I conducted a four-dimensional sweep of receptor doses. The outcome for the mean-transfected cell from each population is presented as one data point in the plot for each promoter (**Figure A4.5c**). F.D. for the two-ligand case relative to each one-ligand case is indicated by the position along the horizontal and vertical axes, and F.D. for the two-ligand case relative to the no-ligand case is color-coded. Perimeters define the performance bounds of the promoters, and the most ideal AND behavior is realized in the upper-right, where the two-ligand F.D. is greater than each one-ligand F.D. and the trade-off with respect to each ligand (described in **Figure A4.4d**) has been balanced. For each case, a selected ideal outcome is denoted by a box. As was observed with the soluble TF sweeps, simply setting the responsiveness of each TF to be the same (#2) conferred minimal changes. However, large improvements in AND functionality were realized by increasing promoter synergy either alone (#6) or in combination with TF responsiveness (#7-9). To interpret the selected ideal F.D. outcomes in **Figure A4.5c**, I examined reporter expression (**Figure A4.5d**). Compared to promoter #1, reporter expression with different ligand treatments increased by ~50x with promoter #2, and by up to another order of magnitude with promoters #3–9. However, improved AND gate functionality did not necessarily follow from changes that conferred the largest increases in expression (e.g., promoter #4). Rather, improvements arose from the largest *difference* in increased expression between the two-ligand case and one-ligand cases (e.g., promoter #6).

To evaluate the effects of intercellular variation, I calculated F.D. across a range of amounts of total plasmid received by transfected cells (**Figure A4.5e**). The maximum F.D. for each promoter was achieved

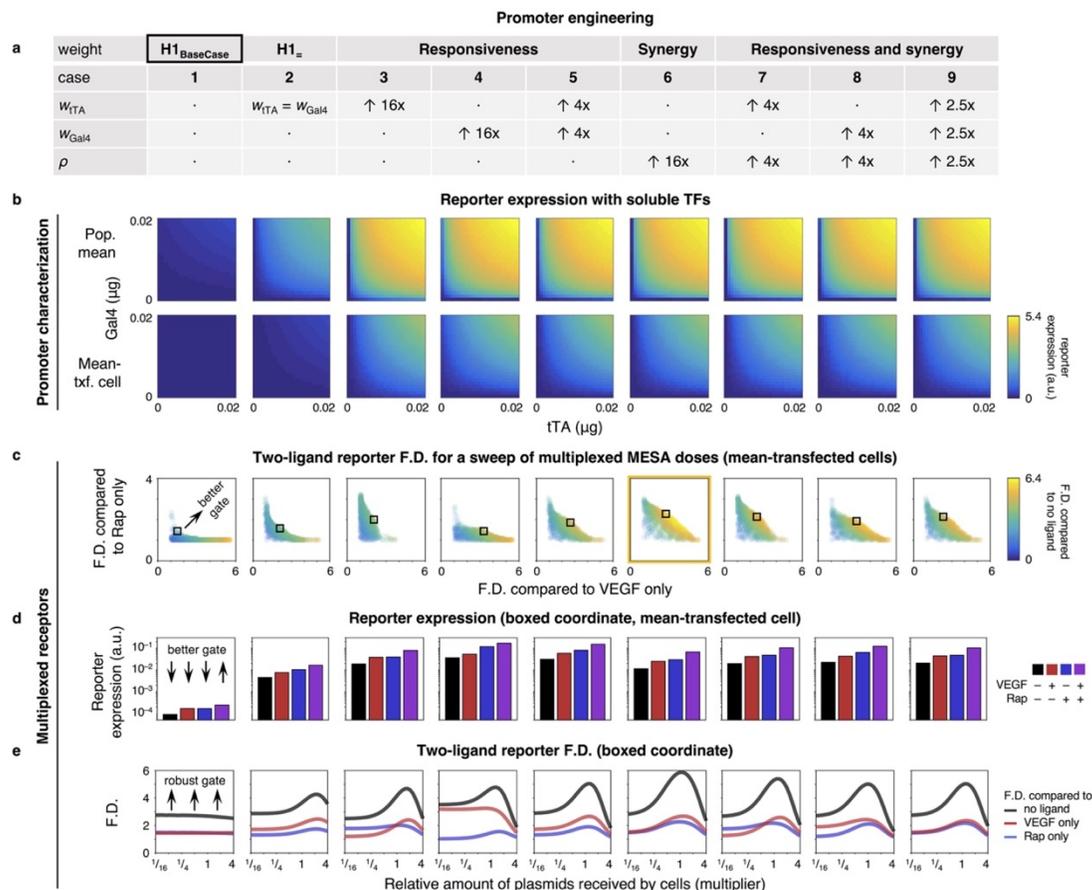


Figure A4.5. Promoter engineering to improve performance. **a** Hypothetical promoters that vary in responsiveness to each TF and/or in synergy were simulated. Multipliers for transcriptional weight parameters in cases #3–9 are in comparison to case #2, in which tTA responsiveness is set equal to Gal4 responsiveness. Case #2 is a responsiveness-balanced version of H1 (base case, #1). **b** Promoters were characterized by reporter expression for the population mean and the mean-transfected cell, using doses of soluble TFs that match the inferred range of TFs released in MESA signaling. **c** Multiplexed MESA performance was assessed by a sweep of 1,000 receptor plasmid dose combinations (0 to 0.5 μg per plasmid), each of which is represented by a point. Plots show three performance metrics—two-ligand-induced F.D. calculated with respect to: (i) VEGF treatment (x-axis), (ii) rapamycin treatment (y-axis), and (iii) no ligand (color-coded). Metrics are calculated for the mean-transfected cell. Better AND gates are realized towards the upper-right of each plot. All three metrics cannot be maximized simultaneously, as evidenced by the absence of outcomes in the upper-right corner, because choosing doses that maximize any one metric comes at the expense of decreasing one or both others. Therefore, the best AND gate requires each metric to be increased only to an extent, in a way that balances the trade-off with the others. A representative ideal instance for each case is indicated by a box and is examined further in **d** and **e**. The best promoter overall (#6) is outlined in yellow. **d** A comparison of reporter expression for instances identified by the boxes in **c** using the mean-transfected cell. **e** Effects of cell variation on the three metrics. X-axis numbers are multipliers for the relative amounts of plasmids received by cells (determined without the variance from the minor principal component that is due to cotransfection), such that a value of 1 is the mean-transfected cell. The multipliers 1/16, 1/4, 1, and 4 correspond to the 23rd, 45th, 62nd, and 85th percentiles, respectively, for plasmid transcription for cells in a population, as determined from the intercellular variation model. Each line represents the F.D. from simulations of increasing plasmid dose, from left to right. Greater robustness to intercellular variation (in context of the specific plasmid doses for each case) is indicated by increase in F.D. across a wider range of x-axis values.

by a subpopulation, and the location of this window differed between promoters. Interestingly, while the base case was the lowest performing, it was the most robust to variation in plasmid dose, with a relatively flat profile. Promoters #3–9 exhibited distinct maxima for F.D. at specific plasmid doses and exceeded the maximum F.D. of promoter #1, indicating that for these hypothetical promoters, obtaining transfected cells with intermediate amounts of plasmids (given the specified dose) would confer maximal performance. Alternatively, any strategy that reduces intercellular variation in expression levels, perhaps such as genomic integration of MESA expression constructs, could improve the performance of such promoters, although this appears to be less promising for H1. Altogether, this analysis provides new insights into how future promoters may be designed, evaluated, and utilized with receptors.

I next investigated how MESA may be modified to tune performance in multiplexing applications. Mirroring the approach for exploring promoter variations, hypothetical modifications to receptor kinetics and design were introduced individually and in combination in silico to generate a panel of distinct cases (**Figure**

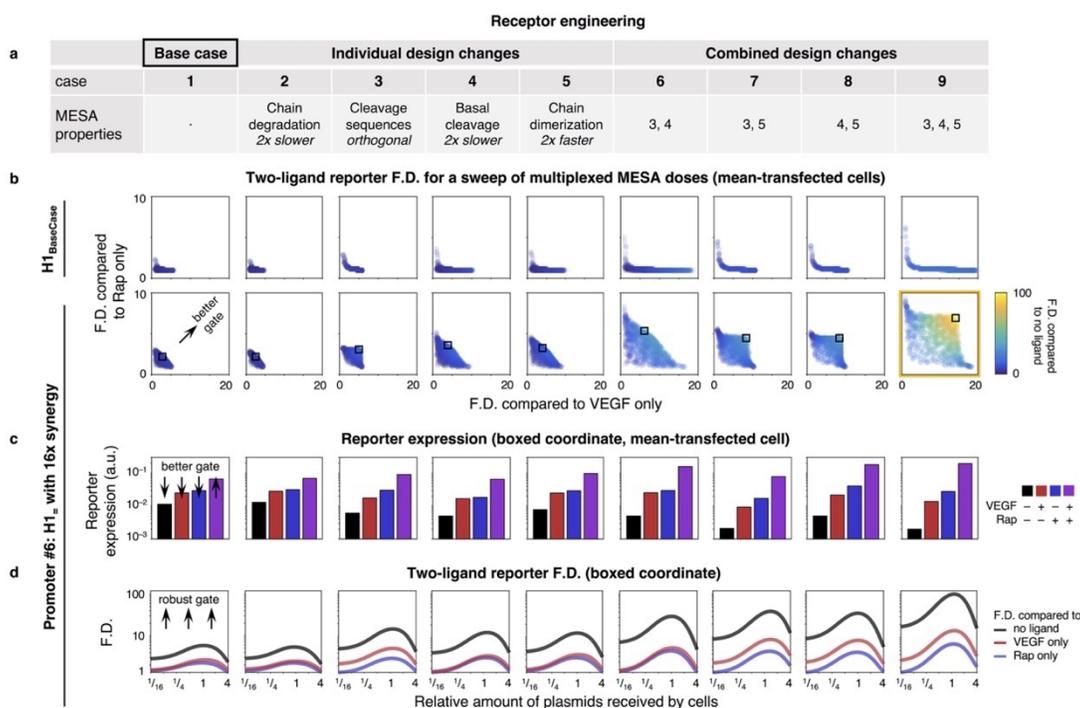


Figure A4.6. Receptor engineering to improve performance. **a** Hypothetical modifications to MESA receptor kinetics and design features were produced in silico in reference to the base case (#1). Cases #2–5 are modifications to #1, and #6–9 are combinations of the modifications in #3–5. **b** Multiplexing was assessed and plotted as described **Figure A4.5**. Outcomes are shown for two promoters: base case (promoter #1, upper row) and a high-performing promoter from **Figure A4.5** (promoter #6, lower row). For promoter #6, a representative ideal instance for each receptor modification is indicated by a box and is examined further in **c** and **d**. The best receptor overall (#9) is outlined in yellow.

A4.6a). For each case, system performance was compared to the base case, comprising the existing receptor, which was paired with either the base case promoter (promoter #1) or the best hypothetical promoter from **Figure A4.5** (promoter #6) (**Figure A4.6b**). Although making changes to the receptor while retaining the base case promoter had little impact on performance, larger improvements were possible with promoter #6. The greatest effects were conferred when multiple modifications were implemented together: orthogonal cleavage recognition sequences, slower basal cleavage, and faster ligand-induced receptor dimerization (receptor case #9). Examining the reporter expression shows how each modification affected F.D. (**Figure A4.6c**). Importantly, not all intuitively attractive changes improved system performance, and so this analysis helps identify promising strategies. For example, decreasing the receptor degradation rate in receptor case #2 could increase the amount of TF that can be released, but this increases the background and one-ligand expression more than it increases two-ligand expression. In contrast, modifications such as receptor cases #6 and #9 increase the two-ligand induced signaling while maintaining or driving down background. That is, the amounts of TFs released when both ligands are sensed can activate the new promoter synergistically, whereas TF levels in the presence of neither or only one ligand drive much less promoter activation. We also found that although engineering both the promoter and receptors in combination still resulted in a scenario in which maximal F.D. was observed within windows of plasmid dose (**Figure A4.6d**), AND behavior remained robust across about an order of magnitude in cell variation, representing a large subpopulation. These results help to identify potential strategies that could be explored for engineering receptors and promoters to achieve level-matching. The model developed in this study could be similarly utilized to prospectively evaluate the future use and design of MESA for other cell functions. More broadly, this approach highlights the utility of quantitative analyses that capture intercellular variation to guide the design of engineered systems that confer desired performance objectives.

Discussion: We investigated how MESA receptors can be multiplexed. AND gate logic served as a useful test case for level-matching with a downstream promoter. We anticipate that other types of logic may be achieved using a similar approach, other engineered systems may exhibit similar phenomenology, and the construction of such functions may be realized by using *in silico* analysis to examine the system components and constraints and to guide design choices. Furthermore, other logical programs might have less stringent requirements for design criteria. For example, an OR gate in which each MESA releases the

same TF to induce a single-TF promoter could require less tuning. Since the output of each receptor can in principle be exchanged, our results suggest that other regulators such as ZF-TFs¹⁴³ or dCas9-based TFs^{276,277} could also be used in receptor multiplexing to program mammalian cell functions.

Acknowledgements: I thank Rachel Hartfield and Kelly Sarnese for collaboration on this study.

APPENDIX 5. Model equations for transcription factors

This appendix corresponds to **Chapter 3**, and a version of this appendix was previously published as:

Donahue P.S., Draut J.W.*, Muldoon J.J.*, Edelstein H.I.*, Bagheri N., Leonard J.N. The COMET toolkit for composing customizable genetic programs in mammalian cells. *Nat Commun* **7**, 779 (2020).¹²⁹ *Equal contributions

Gene regulation is represented using a system of ODEs. The example below depicts a constitutively expressed ZFa inducing the expression of a reporter. Transcription of ZFa RNA scales linearly with plasmid dose. Transcription of reporter RNA depends on ZFa protein concentration via a response function f , which is described in subsequent sections. RNA degradation, protein translation, and protein degradation are represented as first-order processes. The k terms are fixed parameters that are either defined as equal to 1 unit or are based on a previous estimate: $k_{\text{transcription}} = 1$ arbitrary transcription unit, $k_{\text{degRNA}} = 2.7 \text{ h}^{-1}$ based on a previous study²⁷⁸, $k_{\text{translation}} = 1$ arbitrary translation unit, $k_{\text{degZFa}} = 0.35 \text{ h}^{-1}$ based on another study²⁷⁹, and $k_{\text{degReporter}} = 0.029 \text{ h}^{-1}$ based on another study²⁸⁰.

$$\frac{d\mathbf{ZFa}_{\text{RNA}}}{dt} = k_{\text{transcription}} \cdot \text{dose}_{\text{ZFa}} - k_{\text{degRNA}} \cdot \mathbf{ZFa}_{\text{RNA}} \quad (\text{A5.1})$$

$$\frac{d\mathbf{ZFa}_{\text{Protein}}}{dt} = k_{\text{translation}} \cdot \mathbf{ZFa}_{\text{RNA}} - k_{\text{degZFa}} \cdot \mathbf{ZFa}_{\text{Protein}} \quad (\text{A5.2})$$

$$\frac{d\mathbf{Reporter}_{\text{RNA}}}{dt} = k_{\text{transcription}} \cdot f(\mathbf{ZFa}_{\text{Protein}}) - k_{\text{degRNA}} \cdot \mathbf{Reporter}_{\text{RNA}} \quad (\text{A5.3})$$

$$\frac{d\mathbf{Reporter}_{\text{Protein}}}{dt} = k_{\text{translation}} \cdot \mathbf{Reporter}_{\text{RNA}} - k_{\text{degReporter}} \cdot \mathbf{Reporter}_{\text{Protein}} \quad (\text{A5.4})$$

Although the rate constants for transcription and translation for both the ZFa and reporter are set equal to 1 unit, these processes differ in living cells. As a result, 1 unit of ZFa RNA can correspond to a different number of molecules in a living cell than 1 unit of reporter RNA, and likewise for 1 unit of each protein. However, importantly, 1 unit for a given species (e.g., reporter protein) *can* be treated as equivalent across simulation conditions (e.g., ZFa plasmid doses), and these are the comparisons of interest in our analysis.

For a ZFa-inducible promoter, the response function f is defined as:

$$f = \frac{b + m \cdot w \cdot \mathbf{ZFa}_{\text{Protein}}}{1 + w \cdot \mathbf{ZFa}_{\text{Protein}}} \quad (\text{A5.5})$$

where b is a non-negative value for TF-independent (leaky or background) transcription; m is a unitless value for maximum activation (for ZF1a, $m \geq 1$) that depends on the number and spacing of binding sites and the TF; and w is a positive value related to the steepness of the ZFa dose response. The ZFa variable refers to the simulated protein level—this is a function of plasmid dose, but is in distinct units from and is not equivalent to plasmid dose.

The parameter m describes the maximum transcription that a specific ZFa can drive at a promoter with a specific number and spacing of binding sites. An m value of 1 is defined for ZF1a with a x1 promoter. We found that values for m vary with the numbers of binding sites (BS). This relationship can be approximated by sigmoid functions as shown below for ZF1a. The max argument ensures that m does not go below 1 and that it increases monotonically with the number of binding sites.

$$m_{\text{spaced}} = \max\left(\frac{8.5}{1 + e^{-0.48(BS-7.6)}}, 1\right) \quad (\text{A5.6})$$

$$m_{\text{compact}} = \max\left(\frac{41}{1 + e^{-0.98(BS-4.6)}}, 1\right) \quad (\text{A5.7})$$

For TFs that follow similar binding site-response behavior, sigmoids appear vertically stretched or squashed. This effect can be represented by changing the numerator value in the fraction for the m function.

Calibration: The model was implemented with modifications to RNA production terms to incorporate cell heterogeneity:

$$\frac{d\mathbf{ZFa}_{\text{RNA}}}{dt} = k_{\text{transcription}} \cdot z_{i,p\mathbf{ZFa}} \cdot \text{dose}_{\mathbf{ZFa}} - k_{\text{degRNA}} \cdot \mathbf{ZFa}_{\text{RNA}} \quad (\text{A5.8})$$

$$\frac{d\mathbf{ZFa}_{\text{Protein}}}{dt} = k_{\text{translation}} \cdot \mathbf{ZFa}_{\text{RNA}} - k_{\text{degZFa}} \cdot \mathbf{ZFa}_{\text{Protein}} \quad (\text{A5.9})$$

$$\frac{d\mathbf{Reporter}_{\text{RNA}}}{dt} = k_{\text{transcription}} \cdot z_{i,p\mathbf{Reporter}} \cdot f(\mathbf{ZFa}_{\text{Protein}}) - k_{\text{degRNA}} \cdot \mathbf{Reporter}_{\text{RNA}} \quad (\text{A5.10})$$

$$\frac{d\mathbf{Reporter}_{\text{Protein}}}{dt} = k_{\text{translation}} \cdot \mathbf{Reporter}_{\text{RNA}} - k_{\text{degReporter}} \cdot \mathbf{Reporter}_{\text{Protein}} \quad (\text{A5.11})$$

where z denotes the intracellular and intercellular variation, using values for the i^{th} cell and p^{th} plasmid. The model was run by iterating through each cell in the population (over a 42 h simulated duration corresponding to a typical experimental duration), and the population mean was calculated.

In experiments from which data were used to estimate parameters, a ZF1a dose response (0, 5, 10, 20, 50, 100, 200 ng plasmid) with a ZF1x6-C promoter-driven reporter (200 ng plasmid) was included as a fiducial marker for normalizing experiment-specific MEFLs to model-specific units that would be consistent across simulations. For each new ZFa, parameters can be estimated from dose response data using the following steps. First, data for the new ZFa are normalized to the within-experiment ZF1a series: to arrive at the m -equivalent units required for steps 2 and 3 below, divide the MEFL values for the new ZFa series by the mean of the MEFL values for the [5, 10, 20, 50, 100, 200] ng portion of the ZF1a series, and multiply by 22.4 (this value is determined from the ZF1a experiment in which m was originally defined).

Second, specify m for the new ZFa series using the maximum observed (or expected) reporter expression. Third, determine b from the data point for ZFa-independent reporter expression. Lastly, fit w by minimizing the sum of squares error between experimental data and simulated population means. The experimental series and simulated series should use the same ZFa plasmid doses, and they should be normalized equivalently such as by dividing by the mean reporter expression of the series. For cases of non-monotonic reporter expression, data points above the ZFa dose yielding maximum reporter expression should not be used to fit w , as the response function is intended to describe only the data from zero ZFa plasmid dose through the maximum reporter expression.

Standard models of transcription: Figure 3.2 compares the COMET model with standard models of transcription that use more parameters¹⁴⁷. Fractional activation f by a TF (y) with promoter affinity w and Hill cooperativity n for TF-DNA binding, at a promoter that has one binding site, exhibits leaky transcription α_0 , and can be maximally activated by the TF to an amount α , is represented as:

$$f = \frac{a_0 + a(wy)^n}{1 + (wy)^n} \quad (\text{A5.12})$$

This formulation can be extended to other scenarios. For two TFs (y_1 and y_2) with respective maximal activation α_1 and α_2 , a combined activation α_{12} , and TF cooperativity ρ for RNAP recruitment, at a promoter with one site per TF, the formulation is:

$$f = \frac{a_0 + a_1(w_1y_1)^{n_1} + a_2(w_2y_2)^{n_2} + a_{12}\rho(w_1y_1)^{n_1}(w_2y_2)^{n_2}}{1 + (w_1y_1)^{n_1} + (w_2y_2)^{n_2} + \rho(w_1y_1)^{n_1}(w_2y_2)^{n_2}} \quad (\text{A5.13})$$

If in this scenario both TFs are the same (one TF species can bind up to two sites), and additionally if maximal activation is 100% ($\alpha = 1$), this simplifies to:

$$f = \frac{a_0 + 2(wy)^n + \rho(wy)^{2n}}{1 + 2(wy)^n + \rho(wy)^{2n}} \quad (\text{A5.14})$$

In a scenario without Hill cooperativity for TF-DNA binding ($n = 1$) and without TF cooperativity ($\rho = 1$), this further simplifies to:

$$f = \frac{a_0 + 2wy + (wy)^2}{1 + 2wy + (wy)^2} \quad (\text{A5.15})$$

We extend the above case to any number of binding sites. Adding sites could affect ρ for each term in the numerator and denominator, but for simplicity we constrain the possible values by assuming all $\rho =$

1. This assumption is applied in the lower plots of the first and second landscapes in **Figure 3.2c**. Examples are shown below for three, four, five, and six binding sites. Coefficients are derived using Pascal's triangle:

$$f_3 = \frac{a_0 + 3(\mathbf{w}\mathbf{y})^n + 3(\mathbf{w}\mathbf{y})^{2n} + (\mathbf{w}\mathbf{y})^{3n}}{1 + 3(\mathbf{w}\mathbf{y})^n + 3(\mathbf{w}\mathbf{y})^{2n} + (\mathbf{w}\mathbf{y})^{3n}} \quad (\text{A5.16})$$

$$f_4 = \frac{a_0 + 4(\mathbf{w}\mathbf{y})^n + 6(\mathbf{w}\mathbf{y})^{2n} + 4(\mathbf{w}\mathbf{y})^{3n} + (\mathbf{w}\mathbf{y})^{4n}}{1 + 4(\mathbf{w}\mathbf{y})^n + 6(\mathbf{w}\mathbf{y})^{2n} + 4(\mathbf{w}\mathbf{y})^{3n} + (\mathbf{w}\mathbf{y})^{4n}} \quad (\text{A5.17})$$

$$f_5 = \frac{a_0 + 5(\mathbf{w}\mathbf{y})^n + 10(\mathbf{w}\mathbf{y})^{2n} + 10(\mathbf{w}\mathbf{y})^{3n} + 5(\mathbf{w}\mathbf{y})^{4n} + (\mathbf{w}\mathbf{y})^{5n}}{1 + 5(\mathbf{w}\mathbf{y})^n + 10(\mathbf{w}\mathbf{y})^{2n} + 10(\mathbf{w}\mathbf{y})^{3n} + 5(\mathbf{w}\mathbf{y})^{4n} + (\mathbf{w}\mathbf{y})^{5n}} \quad (\text{A5.18})$$

$$f_6 = \frac{a_0 + 6(\mathbf{w}\mathbf{y})^n + 15(\mathbf{w}\mathbf{y})^{2n} + 20(\mathbf{w}\mathbf{y})^{3n} + 15(\mathbf{w}\mathbf{y})^{4n} + 6(\mathbf{w}\mathbf{y})^{5n} + (\mathbf{w}\mathbf{y})^{6n}}{1 + 6(\mathbf{w}\mathbf{y})^n + 15(\mathbf{w}\mathbf{y})^{2n} + 20(\mathbf{w}\mathbf{y})^{3n} + 15(\mathbf{w}\mathbf{y})^{4n} + 6(\mathbf{w}\mathbf{y})^{5n} + (\mathbf{w}\mathbf{y})^{6n}} \quad (\text{A5.19})$$

For iii–iv in **Figure 3.2c**, m values for spaced and compact promoters were substituted for a in each term of the numerator and denominator. As an example, the equation for three sites is:

$$f_3 = \frac{a_0 + 3m_1(\mathbf{w}\mathbf{y})^n + 3m_2(\mathbf{w}\mathbf{y})^{2n} + m_3(\mathbf{w}\mathbf{y})^{3n}}{1 + 3m_1(\mathbf{w}\mathbf{y})^n + 3m_2(\mathbf{w}\mathbf{y})^{2n} + m_3(\mathbf{w}\mathbf{y})^{3n}} \quad (\text{A5.20})$$

Since m values can exceed 1, f no longer represents *fractional* activation defined with the range of zero to one. This interpretational note also applies to f in the COMET model.

To investigate modes of transcriptional regulation independent of the effects of cell heterogeneity, the plots in **Figure 3.2c,d** depict homogeneous (one-cell) expression (whereas the fits shown as lines in **Figure 3.2a** depict heterogeneous population means). In **Figure 3.2c**, outcomes were scaled for a maximum attainable value of 1 within each model.

Transcriptional inhibition: The model used to generate predictions presented in **Figure 3.4c,d** was developed as follows. Within the COMET framework, a competitive inhibitor is represented as:

$$f = \frac{b + m \cdot w_A \cdot \mathbf{ZFa}_{\text{Protein}}}{1 + w_A \cdot \mathbf{ZFa}_{\text{Protein}} + w_I \cdot \mathbf{ZFi}_{\text{Protein}}} \quad (\text{A5.21})$$

where m and w_A correspond to the ZFa, and w_I corresponds to the inhibitor. However, the observed effect of the inhibitors (**Figure 3.4**) was greater than that predicted by competitive inhibition alone. We found that outcomes with ZFi-DsRed or with a spaced promoter could be explained by also accounting for a decrease in *effective cooperativity* at the promoter. Removal of cooperativity from a multi-site promoter is a complex process involving an ensemble of promoter states within and between cells. For simplicity, we represent this as a non-mechanistic heuristic function that depends upon the amounts and properties of both the ZFa

and the ZFi. The value m is replaced by a ramp down function from baseline cooperativity without inhibitor to no cooperativity at a high amount of inhibitor:

$$f = \frac{b + \max\left(\min\left(\frac{\left(\frac{w_I \cdot \mathbf{ZFi}_{\text{Protein}}}{w_A \cdot \mathbf{ZFa}_{\text{Protein}}} - l\right)(1 - m)}{u - l} + m, m\right), 1\right) \cdot w_A \cdot \mathbf{ZFa}_{\text{Protein}}}{1 + w_A \cdot \mathbf{ZFa}_{\text{Protein}} + w_I \cdot \mathbf{ZFi}_{\text{Protein}}} \quad (\text{A5.22})$$

where l and u are empirically determined values for the weight-normalized ratio of inhibitor to activator at which the ramp down from m to 1 begins and ends, respectively.

We found that compared to ZFi, ZFi-DsRed was a more potent inhibitor. Multiplying its weight in the equation by a factor of four improved the fit to data, and ramp down parameters were adjusted accordingly to maintain the shape profile:

$$f = \frac{b + \max\left(\min\left(\frac{\left(\frac{4w_I \cdot \mathbf{ZFiDsRed}_{\text{Protein}}}{w_A \cdot \mathbf{ZFa}_{\text{Protein}}} - 4l\right)(1 - m)}{4u - 4l} + m, m\right), 1\right) \cdot w_A \cdot \mathbf{ZFa}_{\text{Protein}}}{1 + w_A \cdot \mathbf{ZFa}_{\text{Protein}} + 4w_I \cdot \mathbf{ZFiDsRed}_{\text{Protein}}} \quad (\text{A5.23})$$

For inhibitor dose responses in **Figure 3.4c**, cooperativity was more readily removed with ZFi-DsRed than with ZFi, and with a spaced promoter than with a compact one. However, cooperativity was maintained with ZFi and a compact promoter, and this held across ZF1i mutants and doses in **Figure 3.4d**.

Transcriptional logic gates: In **Figure 3.5c**, we used the standard model from **Figure 3.2** to investigate properties of AND gates. For simplicity, leaky transcription (a_0) is set to zero and Hill coefficients (n_1 and n_2) are set to one. **Figure 3.5c** shows four variations that differ in whether each TF's maximal activation (a_1 and a_2) is less than or equal to the maximum activation with both present ($a_{12} = 1$), and synergy (ρ) is present or absent.

$$f = \frac{a_1 w_1 y_1 + a_2 w_2 y_2 + a_{12} \rho w_1 w_2 y_1 y_2}{1 + w_1 y_1 + w_2 y_2 + \rho w_1 w_2 y_1 y_2} \quad (\text{A5.24})$$

TFs were assigned identical properties such that landscapes were symmetric about the dose response diagonal. Simulations used the homogeneous model.

In **Figure 3.5c**, TF dose responses span 0 to 200 ng of plasmid, and target gene expression is linearly scaled to a maximum attainable value of 1. Comparison between experiments and simulations

shows that the hybrid COMET promoter exhibits hybrid cooperative activity: it resembles x3-S with either ZFa individually, and it resembles x6-C if both ZFa are present in sufficient amounts.

To explain this effect, we consider a scenario of a ZFa inducing transcription at a x6-C promoter:

$$f = \frac{m_{6xCompact} \cdot w \cdot \mathbf{ZFa}}{1 + w \cdot \mathbf{ZFa}} \quad (\text{A5.25})$$

Hypothetically, if the pool of ZFa protein in a cell could be partitioned into two sub-pools of equal concentration, each with access to a distinct set of three alternating sites on the reporter promoter, then if only one sub-pool were active the promoter activity would decrease to:

$$f = \frac{m_{3xSpaced} \cdot w \cdot \frac{1}{2} \mathbf{ZFa}}{1 + w \cdot \frac{1}{2} \mathbf{ZFa}} \quad (\text{A5.26})$$

If sub-pools differed in properties that affected m and w , then they could be treated as distinct TFs:

$$f = \frac{m_{3xSpacedZFa1} \cdot w_1 \cdot \mathbf{ZFa}_1}{1 + w_1 \cdot \mathbf{ZFa}_1} \quad (\text{A5.27})$$

$$f = \frac{m_{3xSpacedZFa2} \cdot w_2 \cdot \mathbf{ZFa}_2}{1 + w_2 \cdot \mathbf{ZFa}_2} \quad (\text{A5.28})$$

An inhibitor for either ZFa would act specifically on the corresponding binding sites, such that maximal inhibition would require inhibitor species that tile both sets of sites.

In the limit of high doses of both ZFa, the contribution of each individually to total activation is:

$$f = \frac{m_{6xCompactZFa1ZFa2} \cdot w_1 \cdot \mathbf{ZFa}_1}{1 + w_1 \cdot \mathbf{ZFa}_1 + w_2 \cdot \mathbf{ZFa}_2} \quad (\text{A5.29})$$

$$f = \frac{m_{6xCompactZFa1ZFa2} \cdot w_2 \cdot \mathbf{ZFa}_2}{1 + w_1 \cdot \mathbf{ZFa}_1 + w_2 \cdot \mathbf{ZFa}_2} \quad (\text{A5.30})$$

Together, these contributions sum to:

$$f = \frac{m_{6xCompactZFa1ZFa2} \cdot (w_1 \cdot \mathbf{ZFa}_1 + w_2 \cdot \mathbf{ZFa}_2)}{1 + w_1 \cdot \mathbf{ZFa}_1 + w_2 \cdot \mathbf{ZFa}_2} \quad (\text{A5.31})$$

If both ZFa are identical, this expression becomes identical to the original expression.

Table A5.1. ZFa and fitted parameters for x6-C promoters. These ZFa contain VP16. The reference number is the nomenclature used by Khalil, et al.¹⁴³ *Certain ZFa exhibited squelching at high ZFa plasmid doses. n.d. indicates no data.

Zinc finger	Reference number	<i>b</i>	<i>m</i>	<i>w</i>
ZF1	43-8	0.08	33	0.036
ZF2	37-12	0.25	54	0.018
ZF3	158-2	n.d.		
ZF4	97-4			
ZF5	92-1			
ZF6	150-4	0.02	58	0.043
ZF7	172-5	0.11	46	0.025
ZF8*	173-3	0.07	43	0.041
ZF9*	42-10	0.46	33	0.096
ZF10	13-6	0.01	31	0.037
ZF11	36-4	0.08	32	0.025
ZF12*	62-1	0.15	33	0.065
ZF13	21-16	0.04	41	0.012
ZF14*	14-3	0.20	30	0.069
ZF15*	129-3	0.18	33	0.007
ZF16	54-8	n.d.		
ZF17	55-1			
ZF18	93-10			
ZF19	151-1			

Table A5.2. Fitted parameters for modifications to promoter architecture and ZFa domain.

ZF	AD	Promoter	<i>b</i>	<i>m</i>	<i>w</i>
ZF1	VP16	ZF1x1	0.08	1.0	0.036
		ZF1x3-S		1.3	
		ZF1x6-S		3	
		ZF1x12-S		8.5	
		ZF1x3-C		7.1	
		ZF1x6-C v1		33	
		ZF1x12-C		41	
ZF1	VP16	ZF1x6-C CMV_Min	0.26	33	0.058
		ZF1x6-C SV40_Min	0.43	7.5	0.046
ZF1(RARR)	VP16	ZF1x6-C v1	0.08	26	0.018
ZF1(ARRR)				19	0.010
ZF1(AARR)				15	0.011
ZF1(RAAR)				13	0.0043
ZF1(RAAA)				13	0.0023
ZF1(AAAR)				7	0.0040
ZF1(AAAA)				7	0.0017
ZF1(AAAA)	VP64	ZF1x6-C v1	0.08	24	0.012
	VPR			78	0.020

APPENDIX 6. Model equations for genetic components

This appendix corresponds to **Chapter 4**, and a version of this appendix is in preparation as:

Muldoon J.J., Kandula V., Hong M., Donahue P.S., Boucher J.D., Bagheri N., Leonard J.N. Design-driven engineering of mammalian genetic programs.¹⁶¹ *In preparation.*

Dynamical models: Genetic programs are represented by systems of ODEs. State variables include RNA and protein species in arbitrary concentration units. Processes include transcription (constitutive, inducible, inhibitable), RNA degradation, protein translation, split intein-mediated splicing, small molecule-based reconstitution, and protein degradation. Some parameter values are from the COMET study¹²⁹ and others are newly estimated or fitted (**Table A6.1**).

Constitutive transcription from an EF1 α or CMV promoter is treated as proportional to plasmid dose (ng). Functions for regulated transcription are broadly represented by f . The dose term d for a regulated gene is empirically defined and calculated by dividing the plasmid dose (ng) by 200 ng; then, the square root of this fraction is used. E.g., for 200 ng, $d^{1/2} = 1$, and for 50 ng, $d^{1/2} = 0.5$.

Inducible transcription uses the COMET model with b , m , and w . We model the activation mediated by AD-ZF-containing proteins that also contain intC, intN, or additional ZF domains like that by a ZFa. Transcription can be inhibited by a ZF, which sterically blocks the activator from binding sites in a promoter. We model the inhibition mediated by ZF proteins that also contain intC, intN, FKBP, or additional ZF domains like that by a ZF. Transcription can also be inhibited by a DsDed-ZF, which uses the dual mechanism. The effect of the latter mechanism is that at increasing strength or dose of inhibitor compared to activator, the effective cooperativity ramps down to $m = 1$. Inhibition mediated by DsDed-ZF-containing proteins that also contain intC, intN, or additional ZF domains is modeled like that by a DsDed-ZF.

Split intein-mediated splicing is a second-order reaction with a fitted rate constant between intN-containing and intC-containing proteins.

$$k_{\text{rec}} \cdot \text{Species1}_{\text{Protein}} \cdot \text{Species2}_{\text{Protein}} \quad (\text{A6.1})$$

Small molecule-based reconstitution to form a RaZFa uses the Heaviside function H with ligand treatment at time τ (hours) post-transfection.

$$k_{\text{rec}} \cdot \text{AD-FRB}_{\text{Protein}} \cdot \text{FKBP-ZF}_{\text{Protein}} \cdot H(t - \tau) \quad (\text{A6.2})$$

Prior to reconstitution, FKBP-ZF can act as a ZF-like inhibitor against RaZFa or ZFa at a promoter.

$$k_{\text{txZF}} \cdot \frac{b + m \cdot w \cdot \text{RaZFa}_{\text{Protein}}}{1 + w \cdot \text{RaZFa}_{\text{Protein}} + w \cdot \text{FKBP-ZF}_{\text{Protein}}} \quad (\text{A6.3})$$

The following system of equations represents the reconstitution of a ZFa and induction of a reporter. This system produces an AND gate for: if AD-intN and intC-ZF are present, then induce reporter.

$$\frac{d\text{AD-intN}_{\text{RNA}}}{dt} = z_{i,1} \cdot k_{\text{txEF1a}} \cdot \text{dose}_{\text{AD-intN}} - k_{\text{degRNA}} \cdot \text{AD-intN}_{\text{RNA}} \quad (\text{A6.4})$$

$$\frac{d\text{AD-intN}_{\text{Protein}}}{dt} = k_{\text{tl}} \cdot \text{AD-intN}_{\text{RNA}} - k_{\text{rec}} \cdot \text{AD-intN}_{\text{Protein}} \cdot \text{intC-ZF}_{\text{Protein}} - k_{\text{degZFP}} \cdot \text{AD-intN}_{\text{Protein}} \quad (\text{A6.5})$$

$$\frac{d\text{intC-ZF}_{\text{RNA}}}{dt} = z_{i,2} \cdot k_{\text{txEF1a}} \cdot \text{dose}_{\text{intC-ZF}} - k_{\text{degRNA}} \cdot \text{intC-ZF}_{\text{RNA}} \quad (\text{A6.6})$$

$$\frac{d\text{intC-ZF}_{\text{Protein}}}{dt} = k_{\text{tl}} \cdot \text{intC-ZF}_{\text{RNA}} - k_{\text{rec}} \cdot \text{AD-intN}_{\text{Protein}} \cdot \text{intC-ZF}_{\text{Protein}} - k_{\text{degintC}} \cdot \text{intC-ZF}_{\text{Protein}} \quad (\text{A6.7})$$

$$\frac{d\text{AD-ZF}_{\text{Protein}}}{dt} = k_{\text{rec}} \cdot \text{AD-intN}_{\text{Protein}} \cdot \text{intC-ZF}_{\text{Protein}} - k_{\text{degZFP}} \cdot \text{AD-ZF}_{\text{Protein}} \quad (\text{A6.8})$$

$$\frac{d\text{intC/intN}_{\text{Protein}}}{dt} = k_{\text{rec}} \cdot \text{AD-intN}_{\text{Protein}} \cdot \text{intC-ZF}_{\text{Protein}} - k_{\text{degintC}} \cdot \text{intC/intN}_{\text{Protein}} \quad (\text{A6.9})$$

$$\frac{d\text{Reporter}_{\text{RNA}}}{dt} = z_{i,3} \cdot k_{\text{txZF}} \cdot d_{\text{Reporter}}^{1/2} \cdot \frac{b + m \cdot w \cdot \text{AD-ZF}_{\text{Protein}}}{1 + w \cdot \text{AD-ZF}_{\text{Protein}} + w \cdot \text{intC-ZF}_{\text{Protein}}} - k_{\text{degRNA}} \cdot \text{Reporter}_{\text{RNA}} \quad (\text{A6.10})$$

$$\frac{d\text{Reporter}_{\text{Protein}}}{dt} = k_{\text{tl}} \cdot \text{Reporter}_{\text{RNA}} - k_{\text{degRep}} \cdot \text{Reporter}_{\text{Protein}} \quad (\text{A6.11})$$

Ultrasensitivity: Ultrasensitivity is a type of nonlinear signal processing in which a small change in an input produces a large change in an output. We demonstrate how this property can be achieved with motifs such as a double inhibition cascade (**Figure 4.1k**), activation thresholded by an inhibitor (**Figure 4.3b**), and reconstitutable activation (**Figure 4.3c**). The ultrasensitivity of experimental and simulated dose responses is quantified using the Hill coefficient n from a modified Hill equation, in which x is input plasmid dose (ng), y is reporter signal (MEPTRs or MEFLs), y_0 is reporter signal for zero input, and a and b are other fitted parameters. Standard ZFa dose responses have $n \sim 1$. Ultrasensitive responses have $n > 1$.

$$y = y_0 + \frac{a \cdot x^n}{\left(\frac{1}{b}\right)^n + x^n} \quad (\text{A6.12})$$

Table A6.1. Model parameters for genetic programs.

Symbol	Description	Value	Source
b_1	Basal transcription at ZF1x6-C promoter	0.08	COMET
m_1	Max. induction for CMV-driven VP16-ZF1 at ZF1x6-C promoter	33	COMET
w_1	Steepness for CMV-driven VP16-ZF1 at ZF1x6-C promoter	0.036	Fitted here
m_{1E64}	Max. induction by EF1 α -driven VP64-ZF1 at ZF1x6-C promoter	52	Fitted here
w_{1E64}	Steepness for EF1 α -driven VP16-ZF1 at ZF1x6-C promoter	0.192	Fitted here
b_2	Basal transcription at ZF2x6-C promoter	0.25	COMET
m_2	Max. induction by CMV-driven VP16-ZF2 at ZF2x6-C promoter	54	COMET
w_2	Steepness for CMV-driven VP16-ZF2 at ZF2x6-C promoter	0.082	Fitted here
b_H	Basal transcription at ZF1/2x6-C promoter	b_1	Assumed
m_{1H}	Max. induction by CMV-driven VP16-ZF1 at ZF1/2x6-C promoter	m_1	Assumed
m_{2H}	Max. induction by CMV-driven VP16-ZF2 at ZF1/2x6-C promoter	m_2	Assumed
w_{1H}	Steepness for CMV-driven VP16-ZF1 at ZF1/2x6-C promoter	0.072	Fitted here
w_{2H}	Steepness for CMV-driven VP16-ZF2 at ZF1/2x6-C promoter	0.170	Fitted here
b_{10}	Basal transcription at ZF10x6-C promoter	0.01	COMET
m_{10E64}	Max. induction by EF1 α -driven VP64-ZF10 at ZF10x6-C promoter	m_{1E64}	Assumed
w_{10E64}	Steepness for EF1 α -driven VP16-ZF10 at ZF10x6-C promoter	w_{1E64}	Assumed
l	Start of loss of cooperativity (0.5 in COMET)	0	Adjusted here
u	End of loss of cooperativity (2 in COMET)	1.5	Adjusted here
w_{r1E64}	Steepness for DsDed-ZF1 inhibition of EF1 α -driven VP64-ZF1 at ZF1x6-C	$4 * w_{r1E64}$	Definition
w_{r10E64}	Steepness for DsDed-ZF10 inhibition of EF1 α -driven VP64-ZF10 at ZF10x6-C	$4 * w_{r10E64}$	Definition
w_{r1H}	Steepness for DsDed-ZF1 inhibition of CMV-driven VP64-ZF1 at ZF1/2x6-C	$4 * w_{r1H}$	Definition
rec	Reconstitution of split TFs (fitted based on split inteins; also applied to RaZFa)	$0.34 \text{ U}^{-1} \text{ h}^{-1}$	Fitted here
k_{txCMV}	Transcription at CMV promoter	1	Default
k_{txEF1a}	Transcription at EF1 α promoter	1	Default
k_{txZF}	Transcription multiplier for COMET promoters	1	Assumed
k_{tl}	Translation	1	Default
k_{degR}	Degradation of RNA	2.7 h^{-1}	COMET
k_{degZFP}	Degradation of TF protein (default)	0.35 h^{-1}	COMET
k_{degZFP_PEST}	Degradation of PEST-tagged TF protein	0.7 h^{-1}	Assumed
$k_{degintC}$	Degradation of intC-containing TF protein	1.3 h^{-1}	Fitted here
k_{degRep}	Degradation of reporter protein	0.029 h^{-1}	COMET

APPENDIX 7. Identification of determinants of network inference algorithm performance

A version of this appendix was previously published as:

Muldoon J.J.*, Yu J.S.*, Fassia M.-K., Bagheri N. Network inference performance complexity: a consequence of topological, experimental and algorithmic determinants. *Bioinformatics* **35**, 3421–3432. (2019).²⁸¹ *Equal contributions

Summary: Network inference algorithms aim to uncover regulatory interactions governing cellular decision-making, disease progression and therapeutic interventions. Having a blueprint of this regulation is essential for understanding and controlling cell behavior. However, the utility and impact of these approaches are limited because the ways in which various factors shape inference outcomes remain largely unknown. We identify and systematically evaluate determinants of performance—including network properties, experimental design choices and data processing—by developing new metrics that quantify confidence across algorithms in comparable terms. We conducted a multifactorial analysis that demonstrates how stimulus target, regulatory kinetics, induction and resolution dynamics, and noise differentially impact widely used algorithms in significant and previously unrecognized ways. The results show how even if high-quality data are paired with high-performing algorithms, inferred models are sometimes susceptible to giving misleading conclusions. This new characterization approach provides a way to more rigorously interpret how algorithms infer regulation from biological datasets.

Background: The advent of genome-scale and high-throughput experiments demands network inference algorithms that accurately uncover regulation of gene expression and protein activity²⁸²⁻²⁸⁶. These computational tools have been invaluable for studying cell differentiation²⁸⁷, identifying genetic regulators and their targets in disease²⁸⁸⁻²⁹², classifying diseases into subtypes^{293,294}, and predicting mechanisms of drug responses²⁹⁵⁻³⁰⁰. Having a blueprint of the underlying network comprising genetic components and their regulation is key to understanding and controlling cellular processes. Elucidating these blueprints directly from experimental data has proven challenging. Each algorithm offers advantages and limitations, and its reliability is shaped by biological context and experimental design. For instance, algorithms infer certain motifs with different accuracy, and so their performance depends on the presence of these motifs²⁸⁵. These observations have helped spur efforts to benchmark algorithm performance on experimental or in silico datasets with varying properties³⁰¹⁻³⁰⁶, and some of these studies have yielded tools for further exploration of algorithm-dataset pairings³⁰⁷⁻³⁰⁹. Throughout, the most widely used metrics are predominantly AUROC and AUPR: the area under the receiver operator characteristic and precision-recall curves, respectively. This approach treats the inference as a binary classification, which is possible only if a gold standard network is known. However, applications with experimental data rarely have a gold standard network, making it infeasible to use AUROC or AUPR. We postulate that factors relating to network

properties, experimental design and data processing affect algorithm performance, but that the type and extent of these effects remain challenging to discern, in part, because of how they typically might be assessed.

Here, we develop an *in silico* framework and new confidence metrics [edge score (ES), edge rank score (ERS)], and evaluate effects of kinetic parameters, network motifs, logic gates, stimulus target, stimulus temporal profile, noise, and data sampling on algorithms spanning widely used statistical learning methods. The analysis distinguishes between inference accuracy and confidence, quantifies how well algorithms utilize the input data, and enables comparisons in a manner that was not previously possible. The guiding principle is that outcomes across algorithms can now be assessed in like terms through normalization to null models, circumventing the need for a gold standard network. The results show that several factors—some within and others outside one’s direct control—exert significant and previously unrecognized effects, raising questions on how datasets and algorithms ought to be effectively paired.

Methods: Five-node networks were formulated in which the nodes represent genes or proteins and the signed directed edges are regulation. Each network has two nodes (A, B) that fan in to a target node (C), and two nodes (D, E) that fan out from the target node. The fan-in is assigned a logic gate. Various gates have been described for cellular mechanisms³¹⁰⁻³¹⁴. For AND, OR, and SUM gates, both inputs are activators. For AND, inputs act with multiplicative synergy, and for OR and SUM they act independently. In OR, either input can yield maximum activation, and in SUM both are required. For NAND, NOR, and SUB, both inputs are inhibitory. For NAND, both are required for inhibition, and for NOR and SUB either is sufficient. While SUB has not yet been described in a cellular context, it is included for completeness w.r.t. SUM. Six motifs and six logic gates are considered, for 36 motif-gate combinations. Target node activation is defined as a function of the concentrations of the inputs and their affinities for the target node³¹⁵⁻³¹⁷.

In silico data were generated from simulations. Each network is specified by a system of ODEs. The change over time in the concentration X of a node i regulated by nodes $j = 1:J$ is given by:

$$\frac{dX_i}{dt} = \frac{s(i) + \sum_{j=1}^J k \cdot M_{ji} \cdot X_j}{1 + s(i) + \sum_{j=1}^J k \cdot M_{ji} \cdot X_j} - k_d \cdot X_i \quad (\text{A7.1})$$

where $\mathbf{s}(i)$ is a time-dependent stimulus to node i , M_{ji} is an adjacency matrix entry indicating the presence or absence of a directed edge between nodes j and i (**Table A7.1**), and k_d is a degradation constant (0.5 inverse time units). In the five-node networks, non-logic gate nodes have at most one incoming edge. For the gate node c :

$$\frac{dX_c}{dt} = \mathbf{g}(a,b) - k_d X_c \quad (\text{A7.2})$$

where $\mathbf{g}(a,b)$ is the gate function indicating how nodes a and b activate node c (**Table A7.2**).

Efficiencies (k_{cat}/K_M) for enzyme activity and gene regulation span $[10^0, 10^{-9}] \text{ M}^{-1} \text{ s}^{-1}$ ³¹⁸⁻³²⁰. To capture wide kinetic variation, different values for gate edge parameters k were evaluated. Values between $[10^{-2}, 10^2]$ were observed to produce a variety of dynamical profiles given the network formulation and range of node concentrations, and values outside of this kinetic range tended not to provide further variety to the profiles. Non-gate edges were set to 0.5 inverse concentration units, and the two gate edges were varied across 17 log-spaced values in the range $[10^{-2}, 10^2]$.

Stimulus was applied to either or both gate parent nodes. The base case stimulus was applied for the first half of the timecourse and removed for the second half to produce activation and relaxation dynamics. Initial values are steady-state concentrations in the absence of stimulus. The initial value of each node is set to the steady-state value for the given combination of motif, gate, and stimulus. For most cases, the initial value is zero. Simulations were run from time = 0 to 10 a.u. Trajectories were sampled at intervals of 0.5 a.u. to yield a 21 data points per node. For each data point x , relative noise was added such that:

$$x = x_0 \left(1 + \frac{\rho \mathcal{N}(0,1)}{3} \right) \quad (\text{A7.3})$$

where x_0 is the original simulated value, ρ is the percent noise (0, 5, 10, 20, or 50%), and $\mathcal{N}(0,1)$ is a random number drawn from a Gaussian distribution with zero mean and unit variance. With a Gaussian, 99.7% of values are within 3 S.D. of the mean. Division by 3 therefore ensures the noise is essentially bounded by $[-\rho, \rho]$. We note that data with a value of zero remain zero, x values are non-negative, and in principle other distributions could also be used to introduce noise.

Algorithms assign a weight to each edge describing the regulation of one node by another. For each true dataset and each of $N = 100$ null datasets, inferred weights (IW) and null weights (NW) were

inferred for each edge, respectively, in triplicate; replicates were averaged before calculating ES and ERS. The panel of algorithms includes GENIE3³²¹ (which uses Random Forests³²²), TIGRESS³²³, BANJO^{324,325}, MIDER³²⁶, and correlation (abbreviated here as CORR). Our focus was not to span a large number of algorithms or to determine a most effective one, but rather to evaluate determinants of performance using a concise set of established algorithms spanning different statistical methods. Therefore, and per convention, we do not include an exhaustive analysis with other algorithms, but note that the presented analysis is extensible. Null datasets for five-node networks were generated by shuffling data across gate/motif dimensions. Null datasets for GNW networks were generated by shuffling data across nodes and stimulus conditions. To calculate ES and ERS, the necessary outcome from any method of generating the nulls is that the inferred weights (IW) and null weights (NW) are uncorrelated.

Simulated data have the dimensions: (6 motifs) x (6 gates) x (3 stimulus conditions) x (5 noise levels) x (17 values for k_a) x (17 values for k_b) x (5 nodes) x (21 time points). Inference outcomes have the dimensions: (6 motifs) x (6 gates) x (3 stimulus conditions) x (5 noise levels) x (3 time intervals) x (17 values for k_a) x (17 values for k_b) x (5 algorithms). Inference outcomes include four metrics for each possible edge: IW, NW (averaged across the 100 nulls), ES, and ERS. Instances of non-inferable edges are removed from true data and null data before calculating ES and ERS. For each algorithm, depending on whether it infers edges in a manner that is affected by other edges, nodes with identical trajectories can be assigned edge weights that are necessarily identical or potentially different. If the data for nodes A and B are identical, this affects whether the ES and ERS landscapes are diagonally symmetric.

A methodology to assess and compare algorithm performance: To identify how different factors affect inference outcomes in a controlled manner, we started by formulating *in silico* networks representing scenarios for cellular regulation. Given the large combinatorial space, and the potential for a large network to complicate interpretation, we used a concise testbed (a strategy that has also been used in other studies^{305,306,327}). Each network has five nodes: three (A, B, C) comprise a fan-in and the other two (D, E) are downstream of the fan-in target (C). Regulation among A, B and C is specified by a motif, and C is activated via a logic gate (**Figure A7.1a**). We considered 36 gate-motif combinations and four orders of magnitude of kinetic variation in gate edges. For the network inference, we chose algorithms representative of widely used statistical methods, including top performers in DREAM challenges^{321,323} (**Figure A7.1b**).

We take a multifactorial approach to evaluate performance. Parameter values for gate edges are varied to reflect different strengths of regulation. Nodes A and/or B receive a stimulus representing the start of an experiment, such as ligand-induced pathway activation. At the halfway point, the stimulus is discontinued, representing its removal (or treatment with an inhibitor) as in the DREAM challenge³²⁸. Timecourse data from simulations are sampled at regular intervals, and varying levels of noise are added. Lastly, algorithms are provided for different time intervals of the data.

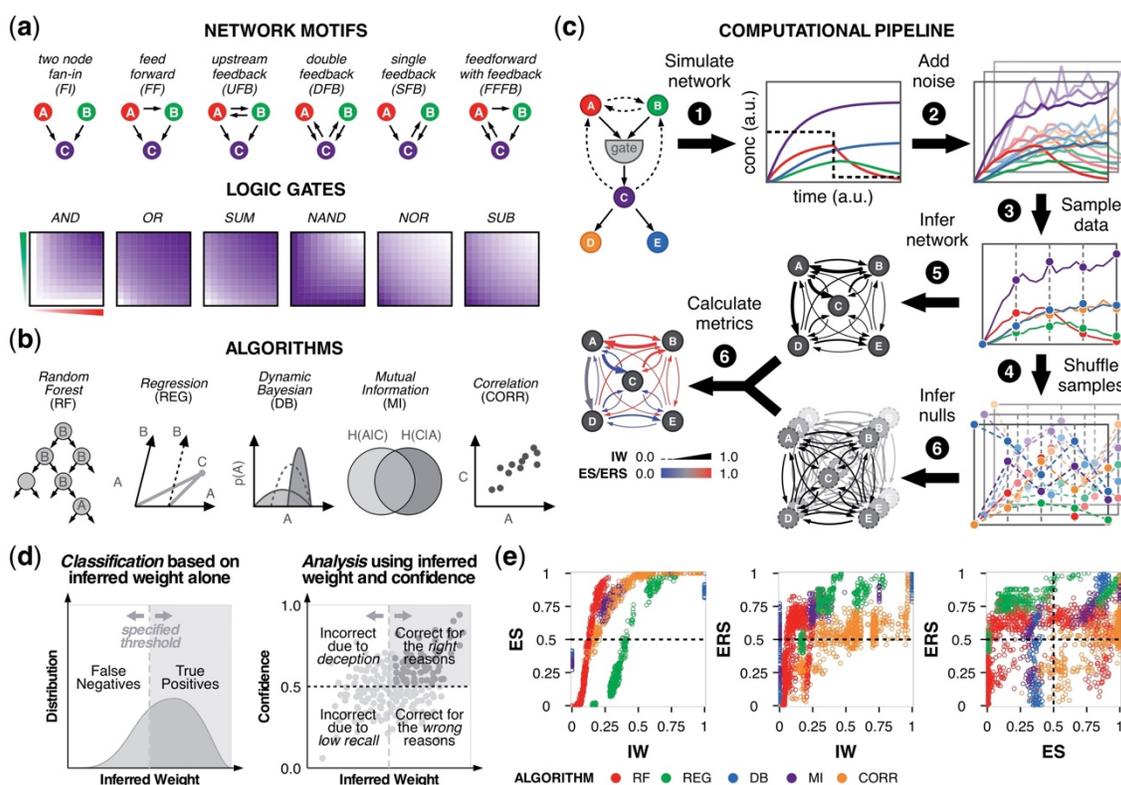


Figure A7.1. Evaluating performance of network inference. Evaluating performance of network inference. **a** Networks differ in features such as motifs and gates. Gates differentially regulate node C based on the activity of nodes A and B. Color-coding (white to purple for low to high activity) characterizes node C in the fan-in motif. **b** Panel of algorithms that use distinct statistical learning methods. **c** Networks were simulated under different conditions to produce timecourse data. Noise was added before data samples were obtained, and true data were permuted to produce null data. Regulation was inferred by each algorithm, and inferred weights (IW) and null weights (NW) were compared to determine the confidence metrics ES and ERS. **d** Left: for a true edge, the two possible outcomes from a binary classification are true positive and false negative. The IW classification threshold depends on algorithm and context. Right: four-quadrant analysis of confidence and IW suggests reasons for algorithm performance. Confidence values above 0.5 indicate that a predicted model tends to outperform null models. Ideal outcomes are in the upper-right quadrant. **e** Left and middle: analysis with IW and confidence; right: comparison of confidence metrics. Results are color-coded by algorithm. For the 36 gate-motif combinations, inference outcomes are shown that are specific to edge A→C, using: nine representative kinetic parameters ($k_A, k_B \in [10^{-2}, 10^0, 10^2]$), stimulus to nodes A and B, no added noise, and data sampled from the full timecourse.

Importantly, as each algorithm uses a distinct statistical method and infers edge weights with different ranges and distributions, the output values cannot be directly compared. Additionally, if an algorithm correctly identifies an edge, it is not possible to determine if this outcome was discerned from information within the data or if it could have been recovered spuriously. These shortcomings motivated us to develop new, generalizable metrics to compare performance across algorithms and assess the confidence of true edges, which we arrive at by comparing IW from true data to NW from N permuted datasets (**Figure A7.1c**). The first metric, ES, quantifies the frequency with which the true-data model outperforms a set of permuted-data models. It represents the confidence of the IW. ES for the edge from node i to node j , across N null datasets indexed by k , is given by:

$$ES_{ij} = \frac{1}{N} \sum_{k=1}^N \begin{cases} 1.0, & \mathbf{IW}_{ij} > \mathbf{NW}_{ijk} \\ 0.5, & \mathbf{IW}_{ij} = \mathbf{NW}_{ijk} \\ 0.0, & \mathbf{IW}_{ij} < \mathbf{NW}_{ijk} \end{cases} \quad (\text{A7.4})$$

The second metric, ERS, quantifies the frequency with which an edge is more highly ranked in the true-data (predicted) model versus permuted-data models. ERS represents the confidence for if a true edge is inferred relative to other edges in a network, and is given by:

$$ERS_{ij} = \frac{1}{N} \sum_{k=1}^N \begin{cases} 1.0, & \text{rank}(\mathbf{IW}_{ij}) > \text{rank}(\mathbf{NW}_{ijk}) \\ 0.5, & \text{rank}(\mathbf{IW}_{ij}) = \text{rank}(\mathbf{NW}_{ijk}) \\ 0.0, & \text{rank}(\mathbf{IW}_{ij}) < \text{rank}(\mathbf{NW}_{ijk}) \end{cases} \quad (\text{A7.5})$$

Both metrics quantify the extent to which algorithms utilize the input data. Values between (0.5, 1] indicate that the predicted model outperforms null models; 0.5 indicates equivalent performance; and [0, 0.5) indicates that null models outperform the predicted model. The use of permuted data, as opposed to random values, ensures that the null data have a distribution consistent with that of the true data.

To situate the new metrics in an existing framework, we consider a standard binary classification. Among the four outcomes [true positive (TP), false positive (FP), true negative (TN) and false negative (FN)], a true edge can be TP or FN (**Figure A7.1d**, left). An algorithm that predicts true edges correctly has high recall (i.e., sensitivity), defined as TP divided by condition positive (TP+FN). However, the recall does not inform whether an algorithm truly discerns regulation based on the data or if the inference can be made by chance. To gain this insight, we use confidence to sub-categorize TP and FN (**Figure A7.1d**, right). If

IW is high and confidence > 0.5 , then we deduce that the algorithm is correct for the right reasons. If IW is high but confidence < 0.5 , it is correct for the wrong reasons; it guessed correctly by chance. If IW is low and confidence > 0.5 , it is incorrect due to deception; it does not uncover the edge well but still outperforms the nulls, suggesting features of the data deceive the algorithm. Lastly, if IW is low and confidence < 0.5 , it is incorrect due to a difficult inference; the outcome is incorrect and has no confidence. Among the four quadrants, ideal performance is in the upper right. We note that this analysis applies to true edges. For false edges, while IW should be low, the interpretation is not defined analogously for the four quadrants.

We observed that each algorithm has characteristic trends for its IW distribution and the relationship between IW and confidence (**Figure A7.1e**, left). For the IW values, Random Forests is low, regression is intermediate, dynamic Bayesian is binary (as expected), mutual information is clustered, and correlation is wide-ranging. Because of these differences, a low IW by one algorithm can potentially convey better edge recovery than a high IW by another, confounding direct comparisons. However, this limitation could be overcome by mapping each IW distribution onto a shared metric. To this end, we note that (i) the IW–ES relationship is monotonic for each algorithm, and (ii) for algorithms that are continuous in IW, ES surpasses 0.5 (y-axis) at a characteristic IW value (x-axis)—which in this context is 0.15 for Random Forests, 0.2 for correlation and 0.4 for regression—such that these values indicate equivalent performance compared to null models. Therefore, for a given network context, ES can be used as a common currency to directly compare IW across algorithms.

The relationship between IW and ERS is different than that with ES, because ERS also accounts for within-model rankings. ERS therefore captures the possibility that a low IW can convey better recovery than a high IW by the same algorithm (such as given a difference in motif, gate, kinetics). For example, the vertical Random Forests pattern (**Figure A7.1e**) shows that one IW value can occupy different within-model rankings relative to the null expectation, and the horizontal pattern for correlation shows that different IW values can occupy similar ones. In summary, ES and ERS provide complementary information that can be applied across algorithms to augment the standard interpretation of IW.

Performance characteristics are highly variable: Given the high dimensionality of the data, we focus on the region where combinatorial variation was introduced and for which the results might be the most informative: the fan-in edges. The kinetic landscapes for confidence show striking patterns that vary as uniform, graded, steep boundaries and speckled. The shapes of the regions for these patterns also vary as kinetically symmetric (mirror imaged across the diagonal), bounded by one or both kinetic parameters, or thin bands with linear or curved boundaries. Many landscapes have surprising combinations of features resembling phase diagrams with phase boundaries and triple points. As a representative example, we highlight a single network that produces different timecourse trajectories depending on the kinetics and stimulus (**Figure A7.2a**), and for which algorithm performance varies as a function of kinetics, stimulus, time interval of input data and gate edge (**Figure A7.2b**). The range of outcomes all from the same network underscores the fundamentally challenging task of network inference.

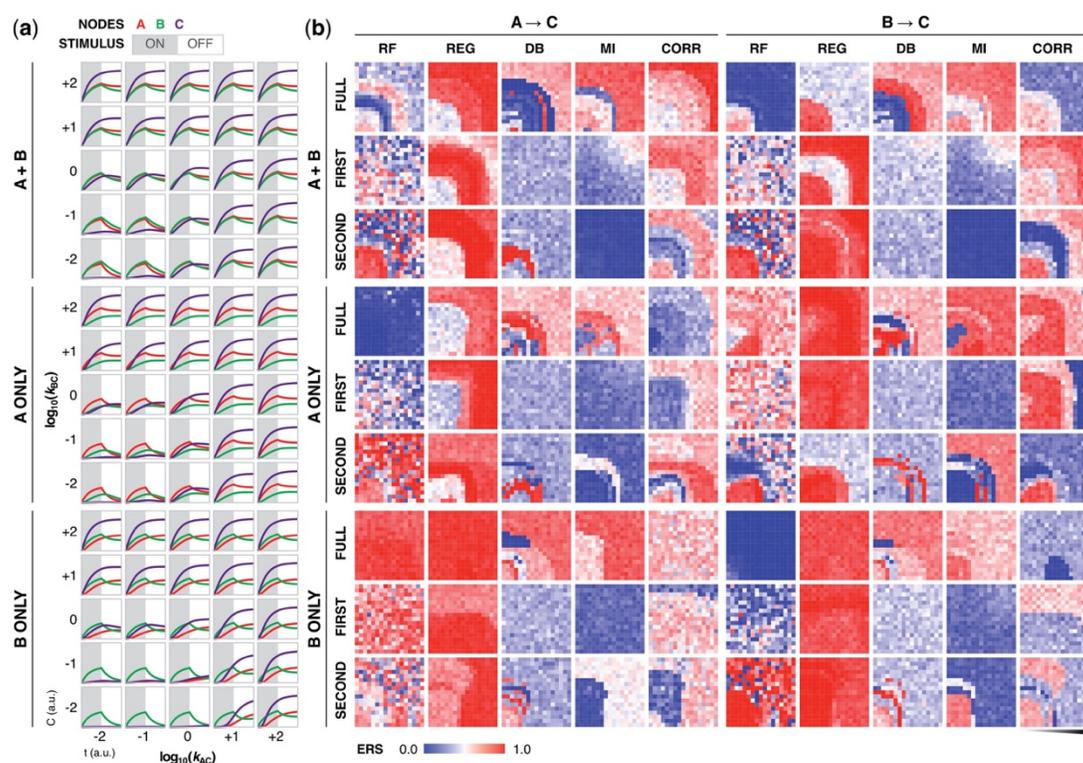


Figure A7.2. Network confidence varies across algorithms. Network confidence varies across algorithms. **a** Trajectories of nodes A, B, and C, and **b** ERS for the two gate edges ($A \rightarrow C$ and $B \rightarrow C$) for a network containing an FFFB motif and OR gate. ERS is provided as a function of stimulus condition (A only, B only or both A and B), time interval of input data (first half, second half and full timecourse), and gate kinetics (plot axes are in log space). Simulations in **a** show a subset of the kinetic landscape, and heatmaps in **b** show the full 17×17 landscape. Gate kinetics (a network property), stimulus target (an experimental choice), and time interval and algorithm (post-experimental choices) strongly affect inference outcomes.

Despite wide variation, the results are still informative. First, much of the variation originates from decisions that in principle are within one's control but in practice are nonobvious. For instance, confidence varies based on the employed algorithm and the time interval of input data. Additionally, stimulus choices that increase confidence in one edge are not necessarily advantageous for recovering another edge. Second, some outcomes of low confidence are due to high NW (rather than only low IW), which, in context, suggest that an algorithm would have an elevated propensity to call FPs. For regression, NW values are relatively high, and for mutual information, NW values are low with the full timecourse dataset but high with the first and second time intervals. Lastly, each algorithm has characteristic contours in the landscapes: Random Forests, dynamic Bayesian and mutual information have defined boundaries; regression is usually highly uniform; and correlation often has several tiers. These patterns hold across networks, indicating that algorithms differ in sensitivity to kinetic variation.

Discussion: This study develops a way to evaluate the confidence and robustness of inference outcomes, which is enabled through comparisons to null models. Although the analysis utilizes *in silico* data with many timepoints, it can also be applied to experimental datasets with fewer timepoints, and in principle it is extensible to any algorithm. Overall, we find that performance is significantly shaped by previously unrecognized factors, some within and others outside of one's control. The no free lunch theorems for optimization³²⁹ state that one should expect identical average algorithm performance in the limit of all possible problems, but that some algorithms will outperform others if better aligned with the subset of problems at hand. In this light, network inference applications would benefit from a greater understanding of the types of data that algorithms interpret well and the circumstances under which experiments produce data that align with these criteria. Achieving this goal will require experimental design (e.g., stimulus target/profile and number/spacing of measurements) that is based, in part, on how well the experiments position algorithms to extract information. Reciprocally, this goal will also require more study on which algorithms effectively utilize data portraying characteristic features of gene regulation and cell signaling. Further characterization of the factors that benefit and hinder algorithms, and investigation on how data and algorithms should be paired, will enable more accurate models and their effective applications.

Acknowledgements: I thank Jessica Yu for rigorous and creative analysis in this collaboration, and Mohammad-Kasim Fassia for the opportunity to be a part of this project.

Table A7.1. Network motifs. Six three-node motifs containing a fan-in are considered. In each adjacency matrix, entries indicate the presence (1) or absence (0) of a directed edge from one node (row) to another node (column), for nodes A, B, and C. All networks have two nodes (D and E) downstream of the fan-in.

Motif	Abbreviation	Adjacency Matrix
Two-node fan-in	FI	$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$
Feedforward	FF	$\begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$
Upstream feedback	UFB	$\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$
Single feedback	SFB	$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$
Double feedback	DFB	$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$
Feedforward with feedback	FFFB	$\begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$

Table A7.2. Logic gates. Functions defining the gates for the fan-in to node C, with parameters for edges k_A and k_B from nodes A and B, respectively.

Gate	Function
AND	$\frac{k_A k_B AB}{1 + k_A A + k_B B + k_A k_B AB}$
OR	$\frac{k_A A + k_B B + k_A k_B AB}{1 + k_A A + k_B B + k_A k_B AB}$
SUM	$\frac{k_A A/2 + k_B B/2 + k_A k_B AB}{1 + k_A A + k_B B + k_A k_B AB}$
NAND	$1 - \frac{k_A k_B AB}{1 + k_A A + k_B B + k_A k_B AB}$
NOR	$1 - \frac{k_A A + k_B B + k_A k_B AB}{1 + k_A A + k_B B + k_A k_B AB}$
SUB	$1 - \frac{k_A A/2 + k_B B/2 + k_A k_B AB}{1 + k_A A + k_B B + k_A k_B AB}$

APPENDIX 8. New ligand-sensing for receptors

A version of this appendix was submitted as:

Edelstein H.I.*, Donahue P.S.*, Muldoon J.J., Kang A.K., Dolberg T.B., Battaglia L.M., Allchin E.R., Hong M., Leonard J.N. Elucidation and refinement of synthetic receptor mechanisms. *In revision*.¹⁹⁸ *Equal contributions

Extending MESA sensing to new cues: As part of an investigation into mechanisms of MESA signaling, I investigated whether the trends observed for different transmembrane domains (TMDs) on a rapamycin-sensing receptor would extend to different ectodomains (ECDs) and ligands. I built new receptors to sense small molecules—gibberellin (GA3-AM is a cell-permeable analog) and abscisic acid (ABA)—and explored considerations that are typically expected to be ECD-specific, such as how linker length affects expression and cell-surface localization^{41,174}. Functional assays showed similar TMD-associated trends across ECDs (**Figure A8.1–A8.2**). The TMD choice for each chain significantly affected background signaling and induced signaling, and the interaction between TMD choices was also significant, indicating that the choice of TC TMD or PC TMD alone does not fully explain the trends. Additionally, the TMD choices together account for most of the variance in background and induced signaling observed.

The observations show that satisfying any one design objective (e.g., maximizing F.D, minimizing background) requires choosing a pair of TMDs suited to that goal. Additionally, some general trends held across the new receptors. For example, high background and modest induced signaling were observed for pairs including FGFR1-TMD TC, resulting in generally low F.D. Conversely, FGFR4-TMD-containing pairs often exhibited low background signaling and high F.D. In summary, some effects of TMD choice extend across receptors, and a limited scan of these choices enables one to generate new functional receptors.

Acknowledgements: I thank my colleagues Hailey Edelstein, Patrick Donahue, Anthony Kang, Taylor Dolberg, Lauren Battaglia, Everett Allchin, and Amy Hong for collaboration on this study.

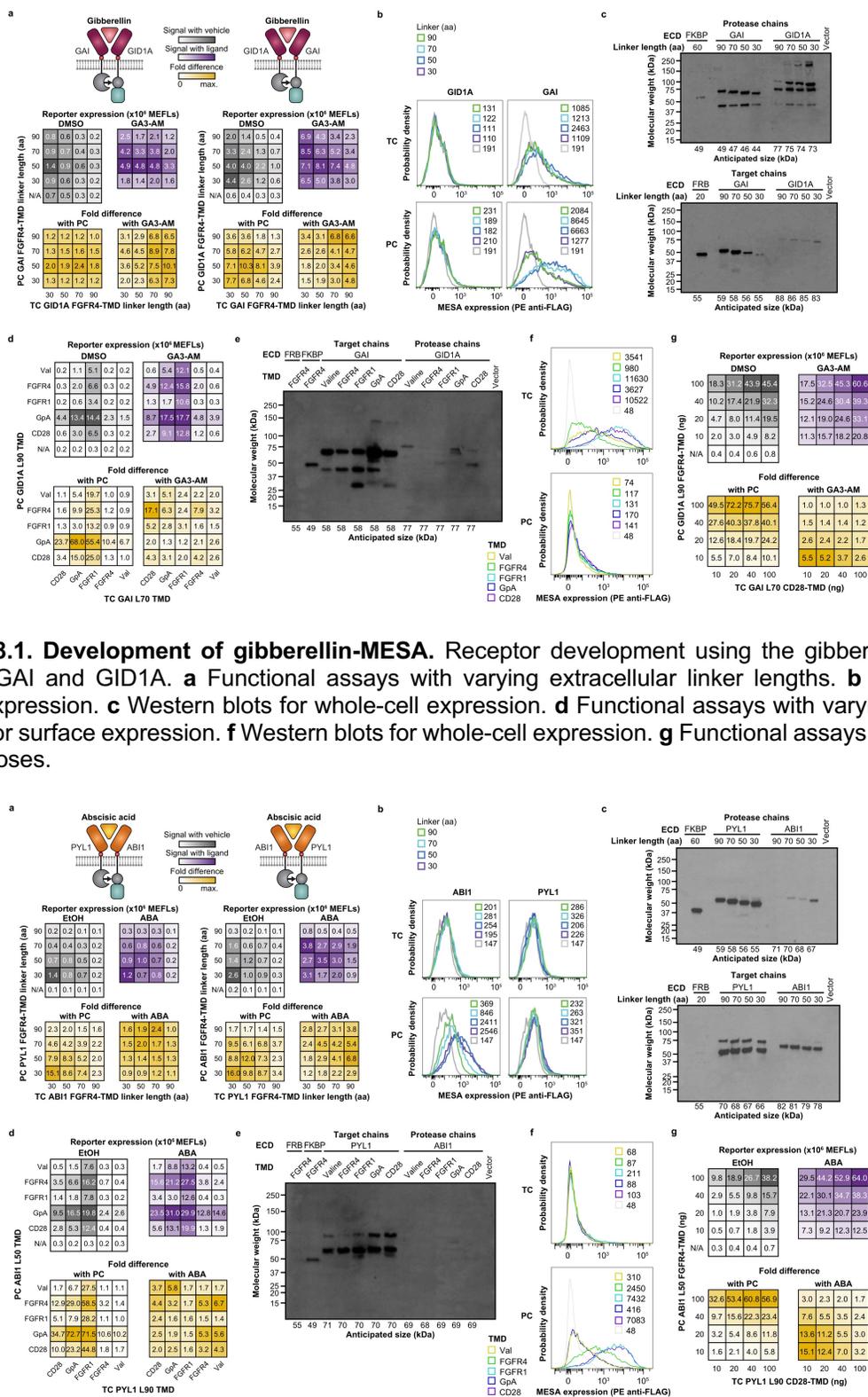


Figure A8.1. Development of gibberellin-MESA. Receptor development using the gibberellin-binding domains GAI and GID1A. **a** Functional assays with varying extracellular linker lengths. **b** Staining for surface expression. **c** Western blots for whole-cell expression. **d** Functional assays with varying TMDs. **e** Staining for surface expression. **f** Western blots for whole-cell expression. **g** Functional assays with varying plasmid doses.

Figure A8.2. Development of abscisic-acid MESA. **a–g** Analogous receptor development process as for gibberellin-MESA using the ABA-binding domains PYL1 and ABI1.

APPENDIX 9. New intracellular cargo for receptors

This appendix corresponds to **Chapter 4**, and a version of this appendix is in preparation as:

Muldoon J.J., Kandula V., Hong M., Donahue P.S., Boucher J.D., Bagheri N., Leonard J.N. Design-driven engineering of mammalian genetic programs.¹⁶¹ *In preparation.*

We expanded the COMET toolkit by incorporating gp41-1: a split intein that was identified putatively in a bioinformatic analysis³³⁰, characterized *in vitro* and in *E. coli*¹⁷⁸, and later utilized in mammalian cells¹⁶⁸. The TRSGY motif from the native sequence upstream of intN (at the end of the intN-adjointing extein) was maintained, as done by^{178,331-333} to retain high splicing activity; however, gp41-1 splicing has also been reported without this motif¹⁶⁸. An additional note is that it is important to use cysteine as the first amino acid of intN (“1” site) and serine as the first amino acid downstream of intC (“+1” site)³³⁴.

The protein sequence for intN was:

CLDLKTQVQTPQGMKEISNIQVGLVLSNTGYNEVLNVFPKSKKKSYSKITLEDGKEIICSEEHLFPTQTGE
MNISGGLKEGMCLYVKE, where the first amino acid is the “1” site, and TRSGY precedes this site.

The protein sequence for intC was:

MMLKKILKIEELDERELIDIEVSGNHLFYANDILTHNS, where the last amino acid is the “+1” site.

The mutagenesis investigation was informed by a crystal structure of the gp41-1 C1A catalytically dead mutant³³⁵. Electrostatic interactions between the $\beta 3$ strand at the end of intN and $\beta 6$ strand at the start of intC were previously identified to form a charge zipper and proposed to be important in the capture and collapse mechanism that precedes splicing. In this mechanism, was which previously elucidated using the *Npu* DnaE split intein³³⁶, capture involves electrostatic interactions between extended regions of the two fragments and collapse involves compaction and stabilization of their initially disordered regions.

Based on the investigation here, split inteins were ultimately used in non-receptor components. However, COMET activators and inhibitors were usable as receptor cargo.

Split intein mutagenesis

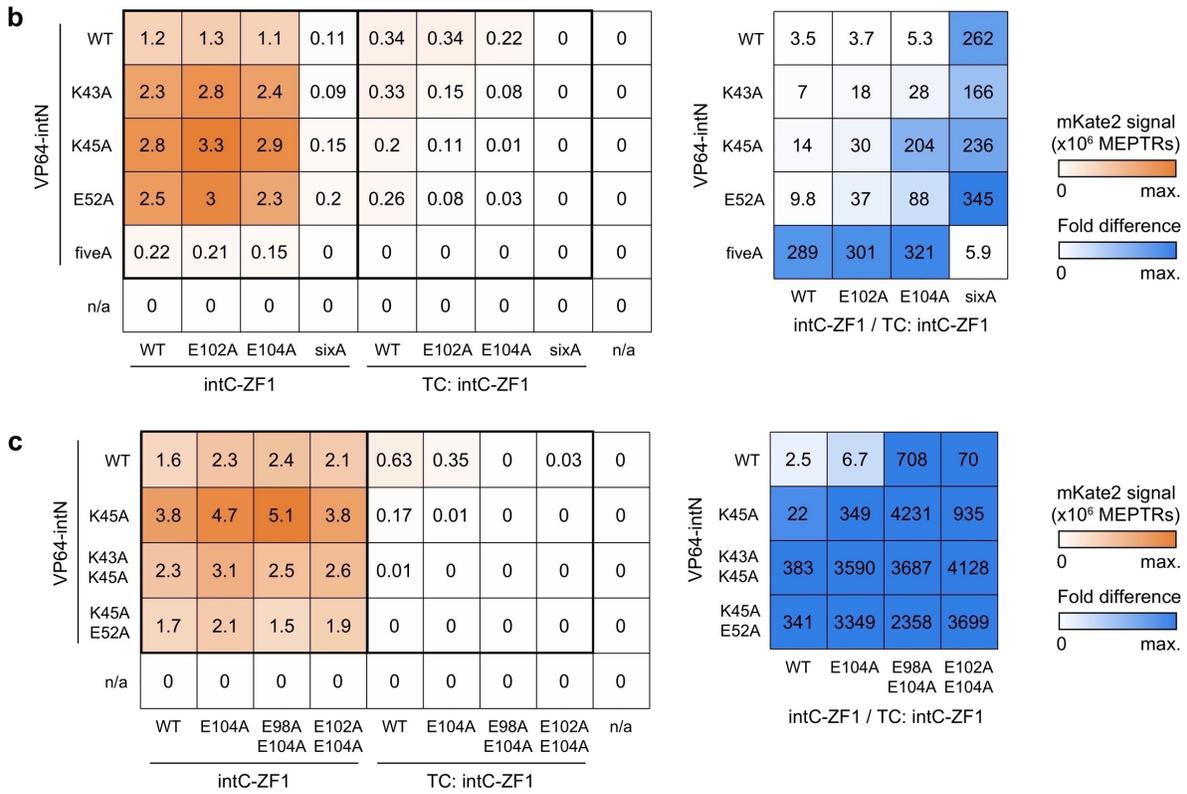
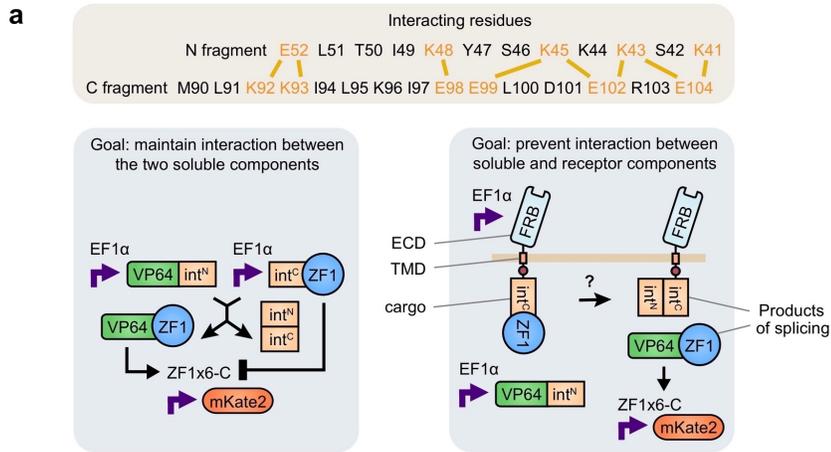
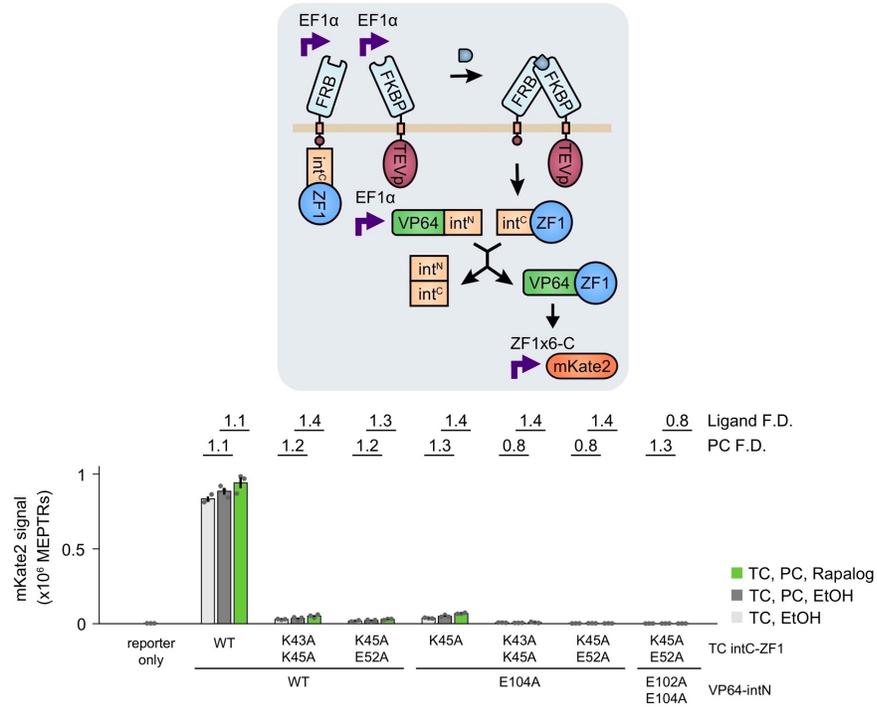


Figure A9.1. Split intein mutagenesis. A strategy was investigated for modulating split intein splicing efficiency. 11 glutamine and lysine residues across intN and intC were selected based on crystallographic evidence³³⁵ for electrostatic involvement in the initial capture step in the mechanism for gp41-1 folding and splicing. We hypothesized that mutating residues to alanine (either individually or in combination, to reduce the number of electrostatic interactions) would decrease the likelihood of intN and intC folding upon coming into contact with each other—and by extension decrease TF reconstitution efficiency—and that this effect would be greater for intC as membrane-proximal cargo on a MESA target chain (TC) than as an intracellular protein. We reasoned that a sufficient differential effect between these two contexts would enable effective fusion of intC-ZF1 onto a TC, such that interactions with intracellular intN-containing components would occur only after proteolytic cargo release in ligand-induced receptor signaling. Abbreviations: ectodomain (ECD), transmembrane domain (TMD). **(a)** The cartoon illustrates the evaluation of reconstitution between two intracellular components (VP64-intN and intC-ZF1) and between an intracellular component and a receptor (VP64-intN and Rapa-MESA TC:intC-ZF1). Mutations were considered ideal if reporter signal was retained in the former scenario and not produced in the latter. **(b–c)** Several mutants were generated and tested **(b)**, and based on these results, double mutants with K45A for intN and with E104A for intC were generated and tested **(c)**. Heatmaps denote the mean reporter signal from three biological replicates (left heatmaps) and the fold difference in mean signal with intC-ZF1 vs. TC:intC-ZF1 (right heatmaps). The results indicate that the new pairings disrupted interactions with intN more for TC-fused intC than for intracellular intC. In some cases, pairings also produced up to several fold greater reporter signal in the intracellular context than did the WT-WT case, and in tandem with the reduction in signal in the receptor context there was a several thousand fold context-dependent difference. High-performing variants were carried forward for further investigation. For the mutation of all 11 residues using the pairing of intN fiveA (K41A, K43A, K45A, K48A, E52A) and intC sixA (K92A, K93A, E98A, E99A, E102A, E104A) **(C)**, reporter expression was not induced in either context. Thus, mutations can be used tune reconstitution efficiency from WT level to effectively none, and there exists an intermediate regime with a differential effect based on whether intC is TC cargo or intracellular.

a Protease chain and ligand treatment



b Target chain intracellular linker

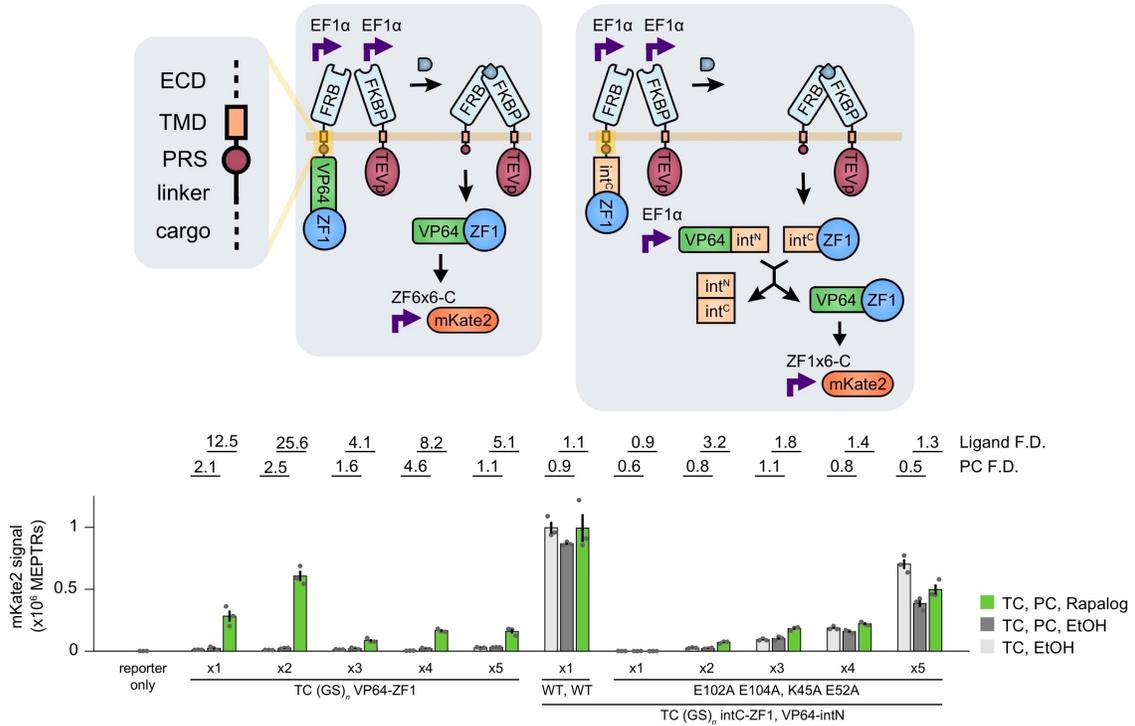


Figure A9.2. Functional test of split intein cargo signaling. a A functional assay was conducted for ligand-inducible receptor signaling incorporating split inteins. A panel of intC and intN pairings from the mutagenesis assay was evaluated for compatibility with the MESA signaling mechanism by measuring reporter signal in three scenarios: (1) VP64-intN and TC:intC-ZF1 with vehicle (EtOH), (2) VP64-intN, TC:intC-ZF1, and MESA protease chain (PC) with vehicle, and (3) VP64-intN, TC:intC-ZF1, and PC with receptor ligand (rapalog). Outcomes are considered ideal if reporter signal is low in the first two scenarios and high in the third. However, we observed that for each pairing, reporter signal was similar regardless of PC co-expression or ligand treatment. This result does not support the ability of PC to cleave intC-containing cargo from the TC, and instead indicates that the conditions in which reporter signal was observed were due to residual interactions between intracellular intN and TC-bound intC partial-mutant variants. These designs were not carried forward, but variants with different intracellular linker lengths were evaluated in the next panel. **b** Functional assay to assess the effect of TC intracellular linker length. We hypothesized that introducing more physical distance or geometric flexibility between the protease recognition sequence (PRS) and cargo would enable PC-mediated cleavage of TC:intC-ZF1. The base case TC linker had one glycine-serine (GS) repeat, and between one and five repeats were tested for mutant intC-ZF1 cargo co-expressed with mutant VP64-intN and for VP64-ZF1 cargo. For the case with intC-ZF1 cargo, reporter signal increased with increasing linker length; however, a signal increase occurred regardless of PC co-expression and ligand treatment. Therefore, modifications to the TC membrane-proximal region did not alleviate the inability of the PC to cleave these TCs. It is possible that the effect of increasing linker length is to make the membrane-proximal intC more accessible (resembling intracellular intC) to intracellular intN. For the case with VP64-ZF1 cargo, there was low reporter signal with PC and no ligand, and there was high reporter signal with PC and ligand, demonstrating that a ZFa can be effectively used on MESA. For increasingly long linkers, the reporter signal decreased. Based on these findings, we chose to use full transcription factors such as ZFa as cargo. Linkers for subsequent TCs contained one GS repeat.

Adapting MESA for inhibitory signaling

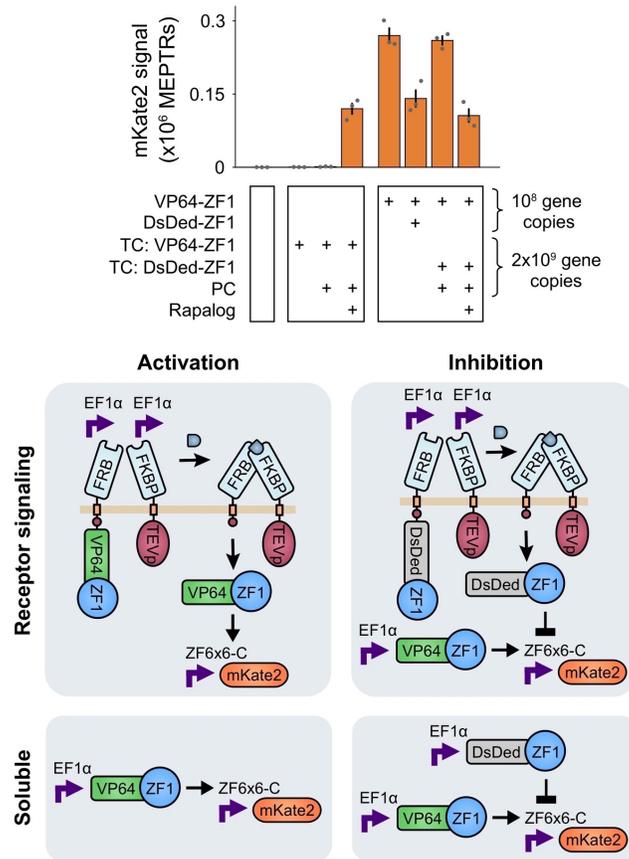


Figure A9.3. Adapting MESA for inhibitory signaling. MESA with DsDed-ZF1 cargo can ligand-inducibly signal to inhibit target gene expression. In this assay, treatment with EtOH or rapalog was applied both at the time of transfection and at the time of media change (rather than only at the latter) to promote inhibitory signaling upon expression of the receptor, analogously to inhibition that could take place upon expression of the intracellular inhibitor.