

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Using Optogenetics to Study the Effect of Transcription Factor Dynamics on Gene Activation

Permalink

<https://escholarship.org/uc/item/76m772sd>

Author

Osimiri, Lindsey

Publication Date

2021

Peer reviewed|Thesis/dissertation

Using Optogenetics to Study the Effect of Transcription Factor Dynamics on Gene Activation

by
Lindsey Osimiri

DISSERTATION

Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Bioengineering

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

AND

UNIVERSITY OF CALIFORNIA, BERKELEY

Approved:

DocuSigned by:

Hana El-Samad

Hana El-Samad

52F8E320792C4C4...

Chair

DocuSigned by:

Liana F Lareau

Liana F Lareau

DocuSigned by:

LUKE GILBERT

LUKE GILBERT

71F73C69F83C48B...

Committee Members

Copyright 2021

by

Lindsey Osimiri

Acknowledgments

My time in graduate school has been made for the better by so many around me. This is a non-exhaustive list; so many people have helped and supported me throughout the past six years that I could not possibly thank them all in this document.

Firstly, I would like to thank Prof. Doug Lauffenburger and Prof. David Clarke, two of my most impactful pre-graduate school scientific mentors. After I learned about systems biology in my junior year of college, I was so excited to do research in this subject. I worked with David as a UROP, yet he treated me as an independent scientist even though I still had so much to learn. I can't forget any of his advice; I still randomize every experiment, just like he recommended. Prof. Lauffenburger also gave me sage advice on topics ranging from siRNA to graduate school decisions to startups. I really appreciate how willing Prof. Lauffenburger was to meet with and prioritize undergraduate students like me. Another mentor from college who I would like to thank is Prof. Bevin Engelward. Though I did not work with Prof. Engelward directly, I truly appreciate how supportive she was of me in my applications to graduate school. At times when I was very unsure about myself and my future, Prof. Engelward was always there to give me the pep talk I needed to refocus and push through.

I would also like to thank the many professors at UCSF who have helped me grow as a scientist throughout graduate school. The members of my qualifying committee, Profs. Wendell Lim, Orion Weiner, and John Dueber, all gave me critical advice in improving my first project, which is described in Chapter 2 of this document. Additionally, my thesis

committee members, Profs. Liana Lareau and Luke Gilbert have always asked excellent questions, given constructive suggestions, and supported my goals as a scientist.

I also thank all those who I have collaborated with in this work. I am grateful to have been able to work with Susan, Mike, Seesha, Nina, Taylor, Alain, Rachel, Lauren, Andrew, Stefanie, Helen, and Lukasz to complete the studies described in this dissertation. I am especially thankful to Susan, Mike, Seesha, Alain, and Nina, who I have worked closely with on both yeast and mammalian CLASP. These projects would certainly not have been possible without you. Additionally, my collaborators at UCLA, Profs. Roy Wollman and Alexander Hoffmann, have been critical in helping me focus and direct my second project, which is described in Chapter 3 of this dissertation.

I would like to thank my thesis advisor, Prof. Hana El-Samad. Your dedication to doing the best science, collaborating as widely as possible, and asking thoughtful questions about synthetic and systems biology has truly shaped who I am and will be as a scientist.

Many at UCSF and UC Berkeley have supported me as scientists and as friends. In the El-Samad Lab, I would especially like to thank Joao, Elham, Alain, Andrew, Taylor, Susan, Zara, Seesha, Jesslyn, Matt, Galen, and Alex. Outside of the El-Samad lab, I'd like to thank Nairi, as well as many members of the Altschuler Wu Lab who I have laughed with over lunch for many of the days over the past 6 years. Each of you has given me scientific wisdom as well as laughter and joy. I'd like to thank all the students of the UC Berkeley - UCSF Bioengineering 2015 cohort for all the fun and commiseration that we shared over the course of our Ph.Ds. I am especially grateful to Nick, Anjali, Yiqi, Katrina, Will, and Andrew for being great friends over the past 6 years.

I also thank my college friends, many of whom live far away but still were my most important support system. Ruby, Sarah, Christian, Angel, and Amanda, thank you for listening to every rant and sending me every meme. Jon, Isaac, Steph, Pri, and Tara, thank you for always finding something new to do in the Bay Area so I had a reason to leave lab.

Finally, I would like to thank my family, without whom none of this would be possible.

Thank you to Stephanie, Derek, and Brian for being my literal Day Ones, and for being there every Sunday throughout a long pandemic. Thank you to my mom, who worked nights and weekends so that I could go to the college of my dreams, and wants only the best for me. Lastly, thank you to my dad, who was proud of me for everything that I did and bragged to every single person he knew about me. I miss you every day.

Contributions

Chapter 2 of this dissertation is a reprint of the following published paper:

S.Y. Chen*, L.C. Osimiri*, M. Chevalier, L.J. Bugaj, T.H. Nguyen, R.A. Greenstein, A.H. Ng, J. Stewart-Ornstein, L.T. Neves, H. El-Samad. Optogenetic Control Reveals Differential Promoter Interpretation of Transcription Factor Nuclear Translocation Dynamics. *Cell Syst.* 11(4):336-353 2020

Using Optogenetics to Study the Effect of Transcription Factor Dynamics on Gene Activation

Lindsey Osimiri

Abstract

Transcription factors (TFs) are important messengers in the information cascade necessary for cells to respond to their environment. In both *Saccharomyces cerevisiae* and mammalian cells, the limited set of TFs must use a variety of strategies to convey complex information about the cellular environment to the many genes that they regulate downstream. These strategies include TF identity, concentration, and combinations of TFs. A subset of TFs also modulate their subcellular localization with distinct dynamic patterns in response to environmental stimuli; this dynamic information may also be used to regulate downstream genes.

Previous studies have attempted to elucidate the effect of spatiotemporal dynamics on downstream genes by using environmental inputs to induce localization changes. Using these stimuli, these studies have shown that TF spatiotemporal dynamics likely transmit information which is then decoded by downstream genes. An important limitation of these

works is that the inputs used modulate not only TF spatiotemporal dynamics, but also other factors, like TF concentration and activation of other TFs and transcriptional regulators. In this study, we develop an optogenetic tool, termed CLASP, which sidesteps these limitations to precisely quantify the effect of TF spatiotemporal dynamics on downstream genes.

Specifically, we used CLASP in *S. cerevisiae* to control the localization of multiple TFs, and find that canonical downstream genes activate more efficiently in response to constant TF nuclear localization than to short pulses of TF localization. In contrast to these data, we focus on Crz1, a pulsatile TF, and find that many of its downstream genes activate more efficiently in response to short, pulsed TF inputs. We then use computational modeling to interrogate the promoter architecture which could yield either outcome, and find that simple promoters respond similarly to many of the downstream genes of Crz1, with a higher output in response to pulsed TF inputs. Surprisingly, a more complex model is needed to explain the phenotype seen for one Crz1 gene and the genes profiled for other TFs, where the gene is more efficiently induced by constant inputs.

In the second study, we extend CLASP to mammalian cells, and show that it is also capable of regulating localization of multiple TFs across multiple cell lines. We focus on a single TF, RelA, in mouse fibroblasts, and perform RNA-seq to measure the effect of its spatiotemporal dynamics on all genes. Additionally, we measure the effect of $\text{TNF}\alpha$, a common environmental stimulus used to modulate RelA spatiotemporal dynamics. First, we find that $\text{TNF}\alpha$ regulates many genes, even when RelA is not expressed in cells, and that this regulation is extremely different from that caused by RelA-CLASP. Secondly, we find that RelA-CLASP activates many downstream genes, and that these genes can respond differently to pulsed and constant inputs. Critically, this is a novel demonstration of the ability of RelA to regulate downstream genes using only translocation, without post-translational modifications or activation of other transcriptional regulators. Using a

simple model of gene expression to simulate a subset of genes in response to constant inputs, we qualitatively predict the response to pulsed inputs for some genes. However, more complex models are needed to explain others.

Together, these studies provide direct demonstration across eukaryotes of the importance of spatiotemporal TF dynamics in regulating gene expression. Additionally, these studies demonstrate an integrated approach of engineering novel tools, measuring dynamic gene expression, and modeling of promoter activation to elucidate how genes respond to complex inputs.

Table of Contents

1	Introduction	1
2	Optogenetic control reveals differential promoter interpretation of transcription factor nuclear translocation dynamics	4
2.1	Summary	4
2.2	Introduction	5
2.3	Results	8
2.3.1	Construction and optimization of CLASP, a dual-LOV2 optogenetic strategy for control of nuclear shuttling	8
2.3.2	CLASP achieves precise, modular control of TF nuclear translocation and activation of target genes	10
2.3.3	CLASP control of the Crz1 TF reveals that its target genes differ in their response efficiency to short pulses	12
2.3.4	A simple two-state computational model of the promoter explains pYPS1-YFP and pCMK2-YFP slope ratio data	15

2.3.5	For a two-state computational model of the promoter, slope ratio increases with decreasing mRNA half-lives but differences in half-lives cannot fully explain the measured slope ratio	16
2.3.6	For a two-state computational model of the promoter, slope ratio increases with decreasing k_d	17
2.3.7	For a two-state computational model of the promoter, slope ratio increases with slow k_{on} and k_{off}	18
2.3.8	pYPS1 promoter dynamics are constrained to be faster than its mRNA decay	20
2.3.9	Higher gene expression in response to continuous inputs by promoters can be explained by a model with a thresholded transition between non-transcribing promoter states	23
2.4	Discussion	27
2.5	Figures	29
2.6	Materials and Methods	42
2.6.1	Resource Availability	42
2.6.2	Experimental model and subject details	43
2.7	Supplemental Text	48
2.7.1	Mathematical Analysis	52
2.8	Supplemental Figures	86

3	Optogenetic control of RelA underscores the importance of transcription factor dynamics in downstream gene expression	108
3.1	Summary	108
3.2	Introduction	109
3.3	Results	111
3.3.1	Designing Mammalian CLASP as a modular optogenetic tool	111
3.3.2	Quantification of RelA-CLASP response to light reveals reversible and dose-responsive dynamics	113
3.3.3	RelA-CLASP is not modulated by environmental stress inputs	114
3.3.4	RelA-CLASP activates canonical downstream genes given a constant light input	117
3.3.5	A simple model of gene expression predicts fundamental differences in genes which respond similarly to constant light inputs	121
3.4	Discussion	126
3.5	Figures	128
3.6	Materials and Methods	138
3.6.1	Experimental details	138
3.7	Supplemental Figures	145
	References	153

List of Figures

2.1	Design, Optimization, and Characterization of CLASP.	31
2.2	CLASP can be used to control localization of many transcription factor cargos.	33
2.3	Crz1 target genes show differing interpretation to Crz1*-CLASP short nucleoplasmic pulses.	35
2.4	Higher gene expression in response to short pulses by promoters occurs when the dose response is saturated at low TF concentration.	37
2.5	Time course measurements of protein output constrain parameter relationships.	38
2.6	Higher gene expression in response to continuous inputs by promoters can be explained by a model with two transition states and with a thresholded transition between non-transcribing promoter states.	41
2.7	Approximately one-third of TFs are basally cytoplasmic in log phase and a subset are shown to exhibit transient nuclear localization, related to Figure 2.1.	88
2.8	Optimization of LANS and LOVTRAP and CLASP characterization, related to Figure 2.1.	90

2.9	Characterization of TF-CLASP strains, related to Figure 2.2.	92
2.10	Characterization of Crz1, Crz1-CLASP, and Crz1*-CLASP nuclear translocation and gene expression with CaCl ₂ or blue light input, related to Figure 2.3.	94
2.11	Higher gene expression of promoters in response to short pulses occurs when the dose response is saturated at low TF concentration, related to Figure 2.4.	96
2.12	Exploration of various models for pGYP7-YFP data, related to Figure 2.6. .	98
2.13	Exploration of three-state models for pGYP7-YFP data, related to Figure 2.6.	101
2.14	Slope ratio N_p/N_c of number of transcripts N is strongly correlated with protein slope ratio, related to Figures 2.4-2.5 and Mathematical Analysis	103
2.15	Simulations of input sequences that accompany theoretical analysis, related to Figure 2.4 and Mathematical Analysis	105
3.1	RelA-CLASP reversibly translocates to the nucleus in a dose-responsive fashion	130
3.2	RelA-CLASP does not translocate to the nucleus in response to environmental stress inputs	132
3.3	RelA-CLASP activates downstream genes in response to a constant light input	135
3.4	A simple model of gene expression predicts response of some genes to pulsed RelA-CLASP inputs	137
3.5	Improving plasma membrane localization for mammalian CLASP	145
3.6	Improving delivery of mammalian CLASP and measuring dynamics	148

3.7	Amplitude and time response of RelA-CLASP	150
3.8	Parameter values determine response to constant and pulsed inputs	152

List of Tables

2.1	NLS Optimization	106
2.2	OptoPlate Transfer Function	107

Chapter 1

Introduction

Cells receive a nearly infinite number of inputs from their environment, and must use a limited set of proteins to interpret and respond to these signals. Transcription factors (TFs) are one important class of proteins utilized in this process. Given the limited number of TFs available to encode the identity of an unlimited number of inputs, the cell must use many layers of complexity in order to correctly respond to the environmental signal at hand. These layers of complexity include TF identity, concentration, post-translational modifications, and even combinations of TFs¹⁻⁵.

An emerging area of interest is the dynamic signals that cells might also use to encode information about environmental stimuli. Critically, dynamic signals offer even more complexity than static signals such as TF identity or combinations of TFs. In response to environmental inputs, a subset of TFs in both *Saccharomyces Cerevisiae* and in mammalian cells translocate from the cytoplasm to the nucleus with distinct dynamic patterns^{6,7}. Many studies have used environmental stimuli to determine the effect of the dynamic patterns of TFs on downstream gene expression in both *S. Cerevisiae* and mammalian cells⁸⁻²¹. In yeast, researchers used an analog-sensitive protein kinase A (PKA)

allele to modulate Msn2 nuclear translocation using a chemical input. Using this approach, these researchers were able to demonstrate the ability of genes downstream of Msn2 to decode parameters such as amplitude and frequency of Msn2 nuclear translocation^{16–18,20,21}. Similarly, in mammalian cells, a recent study used a chemical input to control the feedback loop which determines p53 concentration in the nucleus. With this method, researchers were able to show that genes could decode p53 dynamic patterns, and also that p53 dynamics affected phenotypes like cell cycle arrest¹⁴. Though these studies have expanded our understanding of how TF dynamics regulate downstream gene expression, they have all been subject to the same limitation. Environmental stimuli and drug inputs regulate many proteins within the cell, and these pleiotropic effects are difficult to detangle from the effect of the dynamics of a single TF. As such, a more precise method is needed to study the effects of TF spatiotemporal dynamics on downstream gene expression.

In this work, optogenetics is used to control the spatiotemporal dynamics of a subset of transcription factors in both *S. cerevisiae* and mammalian cells. By engineering a novel tool called CLASP (Controllable Light Activated Shuttling and Plasma membrane sequestration), we precisely and reversibly control the translocation of any cargo from the plasma membrane to the nucleus with blue light. Using this tool, we can mimic the dynamic patterns of TF localization that are observed in response to stimulus, and then measure downstream gene expression to quantify the effect of these spatiotemporal dynamics.

In yeast, we focus our study on Crz1, a TF which translocates to the nucleus in a pulsatile manner following its cognate stimulus, CaCl₂¹². We induce Crz1-CLASP with both pulsed and constant light inputs and find that, unlike other TFs controlled via CLASP, many genes downstream of Crz1 induce more efficiently in response to pulsed inputs than to constant inputs. Using these data, we build computational models to elucidate how promoters would generate either response.

In mammalian cells, we study RelA, a TF which regulates a variety of pathways in the cell, including inflammation, survival, and apoptosis²². Using a mouse fibroblast cell line with endogenous I κ B, the upstream regulators of RelA, and RelA knocked out, we demonstrate that RelA-CLASP can induce downstream genes in response to pulsed and constant light inputs. This novel observation confirms that RelA translocation alone, without post-translational modifications or activation of other transcriptional regulators, can modulate downstream gene expression. Importantly, we also find that a commonly used environmental stimulus for regulating RelA spatiotemporal dynamics, TNF α , regulates many genes, even when RelA is not expressed in cells. Finally, we use transcriptomic data in conjunction with computational modeling to demonstrate how RelA-CLASP can be used to reveal qualitative differences in promoter activation across genes.

In summary, both studies in this work use precise measurements of TF inputs and gene expression outputs, to elucidate the parameters that allow genes to decode spatiotemporal TF dynamics. These models are important for understanding cell biology, but also could be used in future studies to build novel promoters or genes which respond in defined ways to a user input. As a result, these studies represent a stepping stone to further engineering of biological systems.

Chapter 2

Optogenetic control reveals differential promoter interpretation of transcription factor nuclear translocation dynamics

2.1 Summary

This study demonstrates the utility of a novel optogenetic tool, CLASP, for control of the subcellular localization of protein cargos. Furthermore, it also directly demonstrates the importance of the dynamics of subcellular localization and nuclear translocation for downstream gene expression.

2.2 Introduction

Transcription factors (TFs) are key mediators in the transmission of information from the internal and external environment of the cell to its genome. Understanding how TFs encode information about the environment in order to coordinate transcriptional programs remains one of the most pressing problems in molecular and systems biology. Many studies have explored how modulation of TF concentration, TF post-translational modifications, and combinatorial TF control can yield differential gene regulation¹⁻³, therefore explaining many important aspects of TF function and their information encoding capacity. These mechanisms, however, may not fully account for the complexity of signal multiplexing that is carried out by TFs. As a result, it has been proposed that TFs might also encode information in their spatio-temporal dynamics.

A number of studies have attempted to elucidate this TF dynamic encoding hypothesis by eliciting different TF dynamic patterns using various environmental inputs and assessing the consequences^{7,9,23-32}. For example, it was shown that p53 exhibits fixed concentration pulses in response to gamma radiation, but implements only one amplitude- and duration-dependent continuous pulse in response to UV²³. These two pulsing regimes have different physiological outcomes, with the former leading to cell cycle arrest and the latter leading to cell death²⁹. Other studies programmed different TF nuclear translocation patterns by gaining control of a signaling node upstream of the TF. A prominent example of this approach is the modulation of Msn2 dynamics using an analog-sensitive protein kinase A (PKA) allele^{16-18,20,21}. With this method, it was shown that genes in the Msn2 regulon can be differentially modulated by the amplitude, duration, and frequency of Msn2 nuclear translocation pulses. In the budding yeast *Saccharomyces cerevisiae*, there are approximately 200 known TFs, two-thirds of which are constitutively localized to the nucleus; the remaining one-third are located in the cytoplasm during exponential growth in complete media³³. At least nine of these basally cytoplasmic TFs transiently localize into

the nucleus in response to various stress conditions⁶. Furthermore, different environmental conditions elicit a range of pulsing characteristics for these TFs that differ in their duration, amplitude, and frequency⁶ (Figure 2.7), suggesting that reversible TF nuclear localization may encode regulatory information. This information may then be decoded by downstream target genes in order to produce an appropriate response³⁴. Control of TF localization through modulation of upstream regulators with small molecules or chemicals has been an essential method to put forward such a hypothesis of TF dynamic encoding^{12,16–18,20,29,35–37}. However, this method produces pleiotropic effects that can be hard to untangle. For example, PKA controls many transcriptional regulators in addition to Msn2. As a result, modulating its activity with a small molecule may yield gene expression changes that are not solely caused by Msn2 translocation dynamics, but are instead the result of combinatorial gene regulation by other PKA-responsive TFs such as Msn4^{35,38} and Dot6³⁹. Therefore, to causally and quantitatively probe the relationship between TF nuclear localization dynamics and transcriptional activity, a method by which TFs can be specifically, quickly, and reversibly localized to the nucleus is needed. Specificity is necessary to allow direct regulation of TF nuclear localization without pleiotropic effects, while speed and reversibility are necessary to recapitulate the minutes-level resolution with which TFs translocate into and out of the nucleus in response to environmental inputs. Ideally, this method would also work modularly with many TF cargos, including TFs that are basally nuclear. Optogenetic strategies are ideally suited for this purpose. There are many general optogenetic tools to control activity of molecules^{40–43}, as well as published optogenetic strategies to translocate protein cargos to the nucleus^{44–47}. A number of these tools utilized LOV2, a light-responsive protein often isolated from *A.sativa*, to uncage a nuclear localization sequence (NLS) in response to blue light and translocate the optogenetic molecule to the nucleus along with any appended protein cargo. Light Activated Nuclear Shuttle (LANS) is an example of this strategy⁴⁶ (Figure 2.1A). The architecture of this class of optogenetic tools may cause leaky nuclear

localization based on the protein cargo. An example is the TF Msn2, which when fused to LANS, in many cells exhibited constitutive nuclear localization in the absence of light stimulation (Figure 2.8A). Moreover, tools such as LANS cannot be used to regulate localization of basally nuclear TFs. A different optogenetic tool, LOVTRAP, a LOV2-based tool for protein sequestration, could be used for rapid translocation of cargo with less leaky basal localization. LOVTRAP is composed of a LOV2 fused to the mitochondria and Zdk1, a small peptide that is fused to the protein cargo. The interaction of LOV2 and Zdk1 in the dark sequesters the cargo to the surface of the mitochondria⁴⁸(Figure 2.1A). However, LOVTRAP alone does not contain targeting information, and hence cannot direct the cargo to the nucleus on demand. Therefore, to enable both robust and targeted optogenetic control of many different cargos, we sought to use LOVTRAP in concert with LANS. The idea of combining optogenetic sequestration and nuclear localization was previously investigated^{45,49}. However, the resulting tools either required complex dual color stimulation⁴⁵, thereby limiting the number of fluorescent proteins that could be used in a cell, or did not demonstrate modularity for different cargos⁴⁹. These tools also lacked optimization for use in yeast. Here, we present CLASP, an optimized optogenetic tool that can exert precise, modular, and reversible control of transcription factor localization. CLASP uses two LOV2 light-responsive domains derived from *Avena sativa* to sequester a cargo at the plasma membrane in the dark and target it to the nucleus in response to blue light. We demonstrate how CLASP can be used as a general strategy to control many TF cargos without any further optimization. With CLASP and the use of computational modeling, we investigate the consequences of transcription factor translocation dynamics and delineate the quantitative principles by which these dynamics are interpreted by different promoters. This paper therefore contributes an integrated approach through which optogenetic technology development and refinement is coupled tightly to computational modeling in order to answer fundamental questions about promoter principles that allow decoding of dynamic TF inputs. Due to the modularity of CLASP, our integrated approach

provides a general strategy for investigating these principles in many systems.

2.3 Results

2.3.1 Construction and optimization of CLASP, a dual-LOV2 optogenetic strategy for control of nuclear shuttling

To construct a modular and specific tool for yeast protein nuclear translocation, we first tackled optimization of the published LANS and LOVTRAP constructs.

Fluorescently-tagged (mCherry) LANS⁴⁶ displayed only a moderate increase (3.4%) in nuclear/cytoplasmic enrichment in response to blue light (Figure 2.1B, upper left panel).

This increase was much weaker than that seen for transcription factors in response to stress inputs (Figure 2.7B, 20-50% increase). Additionally, the published LOVTRAP tool used a TOM20 mitochondrial targeting tag that caused a strong growth defect in yeast at high expression levels (Figure 2.1B, lower panel). LOVTRAP sequestration had previously been shown to perform best when the mitochondria-bound LOV2 trap was expressed in excess of the Zdk1; as a result, these high expression levels were necessary for trapping many protein cargos and made the growth defect a concrete concern⁴⁸. To improve LANS localization properties, we replaced the published LANS NLS with a small library of yeast NLS peptides⁵⁰ (Table S1). We then screened blue light induced nuclear localization of mCherry-LANS constructs that had any one of these different NLS sequences. We identified a number of NLS sequences that showed an improvement in nuclear/cytoplasmic enrichment in response to blue light (Figure 2.8B), including an NLS that increased the fold change by eight-fold. We chose this NLS sequence to move forward as a yeast enhanced LANS (yeLANS) (Figure 2.1B). Next, to rectify the growth defect associated with LOVTRAP sequestration to the mitochondria, we swapped the mitochondrial TOM20

tag with a plasma membrane Hs-RGS2 tag⁵¹ to create pm-LOVTRAP. This modification rescued the growth defect of LOVTRAP even at high expression levels (Figure 2.8C). Finally, we combined yeLANS and pm-LOVTRAP to form CLASP (Controllable Light Activated Shuttling and Plasma membrane sequestration), a construct composed of two AsLOV2 domains. The first AsLOV2 domain is fused to the plasma membrane and sequesters a Zdk1 fused to the N-terminus of the cargo (for example, a TF). The second AsLOV2 domain is fused to the C-terminus of the cargo. This AsLOV2 domain is preceded by a nuclear export sequence (NES) and has a nuclear localization sequence (NLS) embedded in the $J\alpha$ helix. Blue light causes a conformational change in both AsLOV2 domains, yielding the simultaneous unlocking of cargo and its targeting to the nucleus (Figure 2.1A). Strains harboring CLASP did not experience any measurable growth defect (Figure 2.8D). We first tested CLASP with a red fluorescent protein (mScarlet) as a cargo. Confocal microscopy showed that mScarlet-CLASP was successfully sequestered at the membrane in the dark and translocates to the nucleus in response to blue light. Furthermore, widefield microscopy showed that nuclear localization could be maintained stably for at least 80 minutes (Figure 2.1C). Varying the duration of the light input demonstrated that CLASP could also track shorter light inputs (Figure 2.8E-G). On average, mScarlet-CLASP nuclear localization extended four minutes longer than the duration of the input light pulse, illustrating its rapid shut-off time (Figure 2.1D, Figure 2.8E). The maximum nuclear/cytoplasmic enrichment achieved by mScarlet-CLASP was also graded as a function of light amplitude; when subjected to one minute pulses of increasing amplitude (64-1024 a.u.), enrichment increased commensurately for a wide range and saturated after 256 a.u. of light (Figure 2.1D, Table 2.2). Finally, to test the ability of CLASP to respond to repeated light pulses and probe its dependence on their period, we subjected the cells to one minute pulses of blue light repeated every 2-9 minutes (Figure 2.1E, left 3 panels show one minute pulses every 9, 5, or 2 minutes). These experiments revealed that mScarlet-CLASP followed these pulses faithfully until the pulses became too

rapid, that is, when the next light pulse occurred during the time required for nuclear exit (≈ 4 minutes). This effect occurred when pulses were repeated every 2 minutes, at which point nuclear localization became almost continuous at a high level. The mean peak-to-trough difference in the amplitude of nuclear localization of single cell traces for different periodic light inputs showed a clear dependence on the period of the light pulse (Figure 2.1E). Overall, our data indicate that mScarlet-CLASP could be rapidly, reversibly, and repeatedly localized to the nucleus as frequently as every five minutes and that the duration and the magnitude of this translocation could be robustly controlled.

2.3.2 CLASP achieves precise, modular control of TF nuclear translocation and activation of target genes

The usefulness of CLASP depends on its ability to successfully control translocation of TF cargos while maintaining their function. Our next step was therefore to test the ability of CLASP to quickly and reversibly control the translocation of three basally cytoplasmic transcription factors to the nucleus. We chose a synthetic transcription factor, SynTF, constructed from Cys2-His2 zinc finger domains and a VP16 activation domain⁵², as well as Msn2, the principal transcription factor in the environmental stress response⁵³, and Pho4, the principal transcription factor in the phosphate starvation response⁵⁴. Both Msn2 and Pho4 have been known to translocate to the nucleus in response to stress^{6,54}. The three TF cargos were also tagged with a C-terminal RFP (mScarlet) for visualization. For all three TFs, TF-CLASP achieved its maximal nuclear localization in response to light within one minute of blue light exposure. Like the mScarlet cargo, the TF cargos reversibly translocated to the nucleus as frequently as every five minutes when induced with a one minute pulse of light. Furthermore, a sustained light input produced continuous nuclear localization of the TFs, indicating that CLASP was capable of maintaining robust nuclear localization of associated TF cargos for an extended period of time (Figure 2.2A). The

maximum nuclear/cytoplasmic enrichment fold change achieved with CLASP for Msn2 as a cargo was similar to that of Msn2 with a strong osmotic shock using 0.95M Sorbitol²⁶ (Figure 2.7B). To test whether nuclear localization of the TFs led to concomitant gene expression, we built yeast strains in which YFP was expressed from promoters that were responsive to SynTF (pSYNTF-YFP), Msn2 (pHSP12-YFP), and Pho4 (pPHO84-YFP). We exposed these strains to fixed-amplitude light inputs (Figure 2.9A) of increasing duration (0.5-2 hours) and measured YFP fluorescence via flow cytometry. For all three TFs, increasing the duration of the light input led to increased downstream reporter gene expression, illustrating that the TF was still functional despite its fusion to CLASP. Notably, SynTF-CLASP yielded more than 20-fold activation of pSYNTF-YFP with only 2 hours of light activation (Figure 2.2B). Gene expression in the dark downstream of the three TF-CLASP constructs was similar to basal expression, and was also commensurate after light induction to gene expression generated by a constitutively nuclear TF (Figure 2.9B-D, 2.7.1 “Measuring the basal and constitutively nuclear gene expression of TFs”). Next, we explored whether CLASP could control localization of transcription factors such as Gal4, which was basally nuclear. Gal4-CLASP was successfully sequestered to the plasma membrane in the dark and reversibly translocated to the nucleus in response to light. Nuclear translocation of Gal4-CLASP also activated expression from pGAL1, a Gal4-responsive promoter (Figure 2.9E-G), indicating that CLASP was able to control TFs irrespective of their endogenous nuclear localization. Finally, we sought to demonstrate that different TF dynamic translocation patterns generated with CLASP could yield different gene expression outputs. Several transcription factors, such as Pho4 following phosphate starvation, translocate into the nucleus in response to a stress input and reside there continuously until the response is completed⁵⁴. Others, including Msn2 following a 0.4% glucose input, have been known to translocate into the nucleus with episodic and repeated pulses in response to an activating input⁶. Moreover, Msn2 has also been known to translocate with sustained pulses in response to osmotic shock (Figure 2.7B). As a

result, we sought to explore the gene expression consequences of pulsing relative to continuous localization of the three CLASP-fused TFs (SynTF, Msn2 and Pho4). We delivered two light inputs that had different dynamic patterns but the same cumulative light duration of 40 minutes. In the first case, light was switched ON for 40 minutes, and in the second, light was given in 20 episodic pulses (2 minutes ON/10 minutes OFF) (Figure 2.2C). Delivery of the same cumulative light input and measurement at the end of the time course were necessary controls to compare the response efficiency of the promoters for pulsed input relative to continuous inputs. YFP fluorescence was measured for both inputs after 5 hours using flow cytometry. These data showed unambiguously that continuous nuclear input of SynTF-CLASP, Msn2-CLASP, and Pho4-CLASP produced higher gene expression than pulsed inputs. This directly demonstrates that TF nuclear translocation dynamics could affect downstream reporter gene expression, an idea that we wanted to explore in more depth.

2.3.3 CLASP control of the Crz1 TF reveals that its target genes differ in their response efficiency to short pulses

To further explore the modes of decoding of TF dynamics by promoters in a biologically meaningful setting, we chose to focus on Crz1, the main TF in the calcineurin-Crz1 signaling pathway that responds to calcium stress. Crz1 has been shown to exhibit two modes of pulsatile nuclear translocation in response to calcium chloride (CaCl_2) stress – a single long initial pulse (40-60 min) and subsequent episodic repeated pulsing (1-4 min) (Figure 2.10A). We reasoned that continuous nuclear localization and pulsing of Crz1 could be interpreted differently by different target genes, a behavior that could be revealed and studied by controlling its localization using CLASP. Crz1 has been shown to undergo phosphorylation on multiple residues to activate gene expression in calcium stress⁵⁵ (Figure 2.10B). Therefore, to survey the response of Crz1 target genes to dynamic inputs using

CLASP, we needed to adopt a variant of Crz1 that bypassed this regulation, an endeavor that could be necessary for studying the effects of many TFs with CLASP. We therefore built a strain expressing Crz1*, an alanine mutant with 19 S/T to A substitutions of Crz1, that was basally nuclear (data not shown) and circumvented the post-translational modification requirements for nuclear localization⁵⁵ (Figure 2.10B). To verify that Crz1* preserved the transcriptional profile of wild type Crz1, we carried out mRNA sequencing of cell populations in which the wild type allele of Crz1 was knocked out and Crz1* was expressed from a constitutive pADH1 promoter. We compared the up-regulated genes of the Crz1* strain (where Crz1* is basally nuclear) with genes upregulated by Crz1-yeLANS under CaCl₂ stress. We found similar gene expression patterns between these two gene sets as shown in the heatmap of genes that cluster together (Figure 2.10C). By probing individual Crz1 target genes with fluorescent reporters, we also found that light-induced Crz1*-CLASP, but not light-induced Crz1-CLASP, was able to elicit appreciable gene expression (Figure 2.10D). For example, Crz1*-CLASP driving pPUN1-YFP, a canonical Crz1-responsive promoter, achieved similar gene expression fold change as pPUN1-YFP in calcium stress (fold change of 1.8 versus 1.7) (Figure 2.10E). Importantly, Crz1*-CLASP did not cause increased gene expression in the absence of light, indicating that CLASP was able to successfully sequester the nuclearly localized Crz1* outside of the nucleus in the dark (Figure 2.10F). We next identified six Crz1 gene targets (Yps1, Ena1, Mep1, Put1, Cmk2, Gyp7) for follow up studies. We used the promoters of these genes, which have also been used in previous studies^{56,57}, to build YFP-expressing promoter fusions, each in a strain with Crz1*-CLASP tagged with mCherry for visualization (Figure 2.3A). We subjected these cells to two distinct types of inputs that mimic natural Crz1 translocation: 2 minute short repeated pulses with different periods or one continuous pulse of varying duration (Figure 2.3A). We confirmed that extended light exposure did not cause a growth defect in the Crz1 overexpression strain (Figure 2.10G). We then measured the nuclear fluorescence of mCherry-tagged Crz1*-CLASP continuously at 30 second intervals. We also

measured gene expression from all six YFP promoter fusions at 5 hours for all inputs given (Figure 2.3A). Every input (pulsatile or continuous) has a given nuclear fluorescence AUC, which we calculated as the integral of the measured mCherry-tagged Crz1*-CLASP nuclear fluorescence time traces and is a proxy for nuclear concentration. A given nuclear fluorescence AUC was associated with a commensurate gene expression value (measured at five hours), and these values were plotted against each other for the two input regimes for each of the six promoters. The resulting plot for all nuclear fluorescence AUC values are referred to as the Gene Output - Nuclear Fluorescence plot (Output-Fluorescence plot for short). Exploration of gene expression as a function of nuclear fluorescence AUC allowed a comparison on equal footing of the overall integrated responses to pulsed and continuous inputs. The Crz1-responsive promoters showed a spectrum of qualitative and quantitative behaviors in the Output-Fluorescence plots (Figures 2.3A-C, 2.10H-J). For pGYP7-YFP, like the promoters shown in Figure 2.2, a pulsed input generated lower gene expression output than a continuous input of the same nuclear fluorescence AUC for all values tested (Figure 2.3B). For pCMK2-YFP, pulsed and continuous inputs generated almost identical gene expression output. However, for pYPS1-YFP, pulsed inputs produced higher gene expression output at all Crz1*-CLASP nuclear fluorescence AUC values tested. These phenotypes were qualitatively reproducible despite slight quantitative day to day variability in gene expression between experiments (Figure 2.10H-J). The difference in output between pulsed and continuous inputs as a function of nuclear fluorescence AUC was quantified as the ratio of the slopes of the two lines in the Output-Fluorescence plot (termed the slope ratio) (Figure 2.3A). This metric showed that the six Crz1-responsive promoters spanned a range that is bracketed by pYPS1-YFP (slope ratio >1) and pGYP7-YFP (slope ratio <1), going from higher gene expression for pulsed than continuous inputs to the opposite phenotype (Figure 2.3C). Since all promoter fusions generated the same YFP as the protein output, these phenotypes must reflect different promoter properties as well as any differential influences of the promoters on mRNA stability. We next turned to data-backed

computational modeling to systematically explore and interpret these behaviors.

2.3.4 A simple two-state computational model of the promoter explains pYPS1-YFP and pCMK2-YFP slope ratio data

To better understand the difference in slope ratios for pYPS1-YFP and pCMK2-YFP, we represented each gene with a simple two-state promoter computational model. The promoter model (Figure 2.4A) has an ON state p_{on} and an OFF state p_{off} , with $p_{on} + p_{off} = 1$. The time-dependent nuclear concentration of the transcription factor Crz1 is given by the function $TF(t)$. The time-dependent equation for the promoter activity p_{on} in response to nuclear localization of Crz1 can be written as

$$\begin{aligned} \frac{dp_{on}}{dt} &= k_{on}p_{off}TF(t) - k_{off}p_{on} \\ &= k_{on}(1 - p_{on})TF(t) - k_{off}p_{on} \end{aligned} \quad (2.1)$$

The rate constants k_{on} and k_{off} are used to describe the transition between the two promoter states. Here, $k_{on}(1 - p_{on})TF(t)$ is the ON rate and is nonlinear due to the input $TF(t)$, while the OFF rate, $k_{off}p_{on}$, is linear. The time-dependent equations for mRNA and Protein are then given by

$$\frac{dmRNA}{dt} = \beta_0 + \beta_1 p_{on} - \gamma_1 mRNA \quad (2.2)$$

$$\frac{dProtein}{dt} = \beta_2 mRNA - \gamma_2 Protein \quad (2.3)$$

From a wide parameter search across k_{on} and k_{off} , we were able to find parameter regimes that qualitatively captured the phenotypes (slope ratio >1 and slope ratio close to one)

observed in the Output-Fluorescence plots for pYPS1-YFP and pCMK2-YFP in Figure 2.3 (Figure 2.11A; example solutions plotted in Figures 2.4E, 2.11D). We also found that the quantitative value of the slope ratio obtained from the Output-Fluorescence plot is dependent on three model parameters – mRNA and protein half-lives, the ratio k_d ($k_d = k_{off}/k_{on}$), and the absolute values of k_{on} and k_{off} . We explore these relationships below.

2.3.5 For a two-state computational model of the promoter, slope ratio increases with decreasing mRNA half-lives but differences in half-lives cannot fully explain the measured slope ratio

Our numerical results indicate that both mRNA and protein degradation values affect slope ratio. Protein degradation of YFP in yeast has been shown to be slow with a value around $\gamma_2 = 0.0083 \text{ min}^{-1}$ ⁵⁸. This value cannot be different among different promoters since they all produce the same protein (YFP). Therefore, the degradation parameter for YFP cannot explain the effects of slope ratio we are studying, and we will not explore its effects any further.

The mRNA degradation rate, on the other hand, can depend on the identity of the promoter^{59,60}. The physiological range of mRNA degradation rates in yeast has been experimentally determined to vary between $\gamma_1 = 0.025 - 10 \text{ min}^{-1}$ ⁶¹. For values of k_{on} , k_{off} , and k_d that span a wide range (k_{on} from 0.005 to $9.2 \text{ min}^{-1} \text{ a.u.}^{-1}$, k_{off} from 0.23 to 4.6 min^{-1} , k_d from 0.5 - 46 a.u.), changing γ_1 alone cannot span the range of slope ratios we observe for pCMK2-YFP and pYPS1-YFP (Figure 2.11B). We therefore conclude that while the mRNA degradation rate affects the slope ratio, with increasing slope ratio as the mRNA degradation rate increases, it alone cannot explain the difference between the slope ratios

of pCMK2-YFP and pYPS1-YFP without additional differences in the promoter characteristics. We therefore turn to investigating the role of promoter dynamics in the slope ratio phenotype.

2.3.6 For a two-state computational model of the promoter, slope ratio increases with decreasing k_d

Model simulations for different values of k_d between 0.5 and 46 a.u. revealed that decreasing k_d increases slope ratio (Figure 2.4B). To build intuition about this relationship, we first consider an asymptotic regime of extremely fast promoter dynamics, relative to transcription factor dynamics, in which k_{on} and k_{off} are large. This results in fast p_{on} and p_{off} that equilibrate to a quasi-steady state on the timescale of the TF dynamics. In this regime, $p_{on} \approx \frac{TF(t)}{(TF(t)+k_d)}$, where $k_d = \frac{k_{off}}{k_{on}}$. This expression for p_{on} can then be used in the equation that describes the dynamics of mRNA to determine the number of transcripts made. If k_d is very large compared to the maximum TF concentration (e.g. $k_d = 46$ a.u. in our system where $TF_{max} = 2.6$ a.u.), then p_{on} is approximately linear as a function of TF and p_{on} changes proportionally to the TF input (Figure 2.4C). For a smaller k_d (e.g. $k_d = 2.3$ a.u.), p_{on} rises more rapidly as a function of TF because the smaller k_d dictates that this Michaelis-Menten function should saturate faster to 1 as a function of TF.

The fact that p_{on} grows non-linearly with TF concentration means that there is excess promoter activity derived from the repeated turn-on and shut-off of the pulsed TF input (Figure 2.4D, excess activity denoted by light red shading). This allows the promoter to activate more over time than for the continuous input where there is only one activation and shut off. Therefore, for the continuous input, p_{on} cannot fully benefit non-linearly from the TF concentration. As a result, for a small k_d , the integral of p_{on} (its accumulated area as a function of time) is larger for the pulsed input than for the continuous input (Figure

2.4D, equivalent area of p_{on} is denoted by gray shading), therefore resulting in production of more mRNA. On the other hand, for large k_d ($k_d = 46$ a.u.), p_{on} follows TF in a linear way, and hence the difference between its integral for the pulsed and continuous inputs is minimal (Figure 2.11C). We provide an analytical exposition of the relationship between slope ratio and k_d for a k_{on} and k_{off} of any value in 2.7.1 (sections “Derivation of expressions of total transcripts from Eq.(1) in the main text”, “Derivation of $\frac{N_p}{N_c} > 1$ in the regime of a fast promoter relative to TF(t)”, and “Derivation of $\frac{N_p}{N_c} > 1$ for general values of k_{on} and k_{off} ” with specific examples presented in Figures 2.14-2.15). We also explore these relationships for smaller k_{on} and k_{off} numerically below.

2.3.7 For a two-state computational model of the promoter, slope ratio increases with slow k_{on} and k_{off}

The results in Figure 2.4C-D represent a promoter that is fast relative to $TF(t)$ (e.g. large k_{on} and k_{off}). We next explored the slope ratio for small values of k_{on} and k_{off} that span different k_d values (Figures 2.4B, 2.11B). For a small k_d ($k_d = 2.3$ a.u.), as k_{on} and k_{off} decreased simultaneously (e.g. $k_{\text{on}} = 2$, $k_{\text{off}} = 4.6$ versus $k_{\text{on}} = .2$, $k_{\text{off}} = .46$), the slope ratio increased (Figure 2.4E). This is because for a pulsed input, a slower k_{off} implies that the promoter stays ON for a period of time beyond the duration of the input. When repeated for every pulse, this residual activity could counteract the amplitude deficiency caused by the decrease in k_{on} and k_{off} (Figure 2.4F). On the other hand, as k_{on} and k_{off} increase, p_{on} follows the input $TF(t)$ pulses more faithfully, minimizing the gains from a pulsed input compared to a continuous one even in the regime of small k_d (Figure 2.4G). Finally, and as discussed above, for a large value of k_d (e.g. $k_d = 46$ a.u.), the slope ratio is necessarily close to 1 and is therefore relatively insensitive to the promoter dynamics (see 2.7.1, “Derivation of $\frac{N_p}{N_c} > 1$ in the regime of a fast promoter relative to TF(t)” and Figure 2.14). As a result, in this regime, slowing k_{on} and k_{off} will have minimal effects on slope

ratio for a value of k_d that is large relative to TF_{\max} (Figure 2.11D).

Taken together, these analyses indicate that both k_d and the absolute values of k_{on} and k_{off} modulate slope ratio. A small k_d is necessary for higher gene expression in response to short pulses, and a slow k_{on} and k_{off} further increase the slope ratio. These conclusions therefore point to the necessity of measuring the k_d of these promoters. To do so, we constructed strains that expressed Crz1*-CLASP from constitutive promoters of varying strengths. In each of these strains, we localized Crz1*-CLASP to the nucleus with a continuous light input for 4 hours and measured pYPS1-YFP and pCMK2-YFP (Figure 2.4H), therefore compiling a dose response. Fitting this data to a Michaelis-Menten function, we found that the experimental dose response of pYPS1-YFP had a k_d of 2.3 a.u. (Figure 2.4I) while the dose response of pCMK2-YFP had a k_d of 12.8 a.u. (Figure 2.4J).

The small k_d value for pYPS1-YFP and its large slope ratio is in agreement with our analysis showing that small k_d can allow the promoter to differentiate between short pulses and continuous inputs. It also positions the promoter in a regime where the individual values of k_{on} and k_{off} might have an important influence on its slope ratio. Moreover, the relatively large k_d for pCMK2-YFP and its slope ratio near 1 is in agreement with our analysis showing that larger k_d restricts the ability of the promoter to differentiate between short pulses and continuous input. Additionally, this relatively large k_d positions the promoter in a regime where the individual values of k_{on} and k_{off} have little bearing on its slope ratio. To explore these hypotheses further, we turned to measurement of protein dynamics in order to further constrain the values of k_{on} and k_{off} .

2.3.8 pYPS1 promoter dynamics are constrained to be faster than its mRNA decay

To constrain individual parameter values of k_{on} , k_{off} , and relate them to the mRNA degradation rate γ_1 , we measured protein dynamics for pYPS1-YFP and pCMK2-YFP in response to 2 hours of continuous light illumination (Figure 2.5A). We then used the model to fit these dynamic data by sampling within a wide range of k_{on} , k_{off} , and γ_1 values (k_{on} from 0.001-10 ($\text{min}^*\text{a.u.})^{-1}$, k_{off} from 0.000007-100 min^{-1} , and γ_1 from 0.01-10 min^{-1} ; Figure 2.5A, left and right panels, fits sought to maximize fit through the data points within the error bars; model fitting discussed in detail in 2.7.1 “Model equations and sampling details of the pYPS1-YFP and pCMK2-YFP phenotypes”). The protein dynamic data revealed a relationship that must exist between $k_{on}TF + k_{off}$ and γ_1 for the data to be explained by the model (Figure 2.5B). An analysis of mRNA dynamics in the two-state promoter model revealed that this relationship captures an important timescale T_s of the system, which we term the “settling time”. Assuming a very small degradation rate of the protein, T_s is the approximate time at which the mRNA level reaches steady state and the protein production rate becomes constant (see 2.7.1, “Asymptotic analysis of mRNA dynamics for the simple promoter to a step function input” for analytical derivations of T_s). The T_s relationship captures two characteristic timescales of the system, that of the promoter given by $\frac{1}{k_{on}TF + k_{off}}$ and that of the mRNA, given by $\frac{1}{\gamma_1}$. Their sum determines the timescale of the system (assuming that protein degradation is slow and hence has negligible contribution). Two asymptotic regimes occur if either the promoter or mRNA dynamics dominate the timescale of the system. These asymptotes serve to put a lower bound on the values of $k_{on}TF + k_{off}$ and γ_1 . At one extreme where k_{on} and k_{off} are large and γ_1 is small, $T_s \approx \frac{1}{\gamma_1}$ (the asymptote as $k_{on}TF + k_{off}$ goes to infinity in Figure 2.5B lower right data points in each panel). Therefore, the protein dynamics data would strongly constrain the values of γ_1 but not k_{on} and k_{off} . In the other extreme, for small k_{on} and k_{off} and large γ_1 ,

$T_s \approx \frac{1}{k_{on}TF + k_{off}}$ (the asymptote as γ_1 goes to infinity in Figure 2.5B, upper left data points in each panel) . In this regime, $k_{on}TF + k_{off}$ is well constrained by the protein dynamics data but γ_1 is not constrained. These asymptotes are useful for putting bounds on the parameters.

T_s can be determined from the protein dynamics data using the mean of the parameter fits. Since T_s has a given value for each gene that can be computed from the protein timecourse, the relationships between k_{on} , k_{off} , and γ_1 are correspondingly constrained for each gene. We found that $T_s = 23.8$ min (with a range of 18 to 32.8 min) for pYPS1-YFP and $T_s = 7.7$ min for pCMK2-YFP (with a range of 6.9 to 9 min) (Figure 2.5B), indicating that the pYPS1 mRNA reaches steady state approximately 3 times slower than the pCMK2 mRNA. These T_s values and their corresponding asymptotes put a lower bound for k_{on} to be $0.001 \text{ min}^{-1}\text{a.u.}^{-1}$, for k_{off} to be 0.006 a.u.^{-1} and for γ_1 to be 0.03 min^{-1} for pYPS1-YFP. These lower bound values for pCMK2-YFP are $0.001 \text{ min}^{-1}\text{a.u.}^{-1}$ (k_{on}), 0.05 a.u.^{-1} (k_{off}) and 0.11 min^{-1} (γ_1). Therefore, these constraints predict that all three parameter values might differ between the two promoters, including the mRNA decay rate. However, as we have shown above (Figure 2.11B), the potential difference in γ_1 values between pYPS1-YFP and pCMK2-YFP contributes to but cannot fully explain the difference in slope ratio between the two genes.

To further constrain the parameter values for pYPS1-YFP and pCMK2-YFP, we subjected the parameter sets constrained by the protein time course in response to a continuous input to the additional constraint of fitting the Output-Fluorescence data (Figure 2.5C-D). For pYPS1-YFP, these data constrained $k_{on}TF + k_{off}$ to be larger than 1 and less than 56, and constrained k_{on} and k_{off} values to be greater than $0.16 \text{ min}^{-1}\text{a.u.}^{-1}$ and 0.6 a.u.^{-1} , respectively. As a result, the γ_1 values were constrained to be between 0.03 and 0.05 min^{-1} , which is approximately an order of magnitude smaller than k_{on} and k_{off} (Figure 2.5D, left panel). For pCMK2-YFP, however, the Output-Fluorescence data did not further constrain

the range of parameters beyond the protein time course data (continuous input) (Figure 2.5D, right panel). Importantly, the parameters fit to the protein time course (continuous input) and Output-Fluorescence data for each gene had k_d values comparable to those measured experimentally, which provides a measure of validation for these fits. For pYPS1-YFP, the median k_d for the parameter fits was 1.6 a.u., and for pCMK2-YFP the median k_d value was 14 a.u.

Therefore, for pYPS1-YFP, promoter kinetics (k_{on} and k_{off}) are fast relative to mRNA decay (γ_1), indicating that mRNA decay dominates protein dynamics. Taken together with the analyses of the effect of k_d , k_{on} , and k_{off} on slope ratio, the small k_d value for this promoter indicates that its slope ratio metric can be strongly affected by its k_{on} and k_{off} values, and that the large slope ratio is likely the result of the small k_d and large k_{on} and k_{off} relative to γ_1 . On the other hand, the dynamics and small slope ratio (close to 1) of pCMK2-YFP could be generated by many combinations of parameters γ_1 , k_{on} and k_{off} that satisfy the T_s relationship (Figure 2.5D). This finding agrees with the fact that slope ratio of pCMK2-YFP is minimally affected by the values of k_{on} and k_{off} because of its measured large k_d .

Finally, to further cross-validate these insights, we asked whether the parameters identified above and the accompanying model could predict the outcome of an additional time course experiment on which the model was not trained. In this experiment, protein time course data is collected for cells induced with a pulsed (2 min ON/4 min OFF) light input for 100 minutes (Figure 2.5E, data plotted in red as circles and error bars, model predictions plotted in red as lines). As a comparison, the data for the continuous input, which these parameters were fit to previously, are also plotted (data plotted in blue as circles and error bars, model predictions plotted in blue as lines). The Output-Fluorescence plots show that pYPS1-YFP displays higher gene expression in response to pulsed TF inputs than pCMK2-YFP, given that both inputs have the same area. In this time course, the

continuous input has a much greater area than the pulsed input. It is expected that the difference between the protein outputs for the continuous and pulsed inputs would be smaller for pYPS1-YFP than for pCMK2-YFP. Furthermore, it is expected that the parameters identified will be able to predict the gene expression dynamics for the pulsed input as a result of being fit to the continuous input protein time course data. For both pYPS1-YFP and pCMK2-YFP, the parameter fits produced computational predictions that qualitatively recapitulated the protein outputs in response to pulsed TF inputs for both pYPS1-YFP and pCMK2-YFP (Figure 2.5E).

2.3.9 Higher gene expression in response to continuous inputs by promoters can be explained by a model with a thresholded transition between non-transcribing promoter states

The simple model from the previous analysis could not produce the pGYP7-YFP phenotype (Figure 2.12A). In the simple model, even as the output difference between the pulsed and continuous inputs decreased (which occurred when $k_{\text{on}} \ll k_{\text{off}}$), the output of the pulsed input was always higher than the continuous input. This is because while decreasing k_{on} reduced the output of the pulsed input, it also reduced the dynamic range of the output in response to a continuous input. This continued until to a point where k_{on} was so small that the promoter was barely activated and the much faster k_{off} quickly shut off promoter activity, resulting in a promoter that was essentially unresponsive to both continuous and pulsed inputs (Figure 2.11E).

In order to identify a minimal model that explains the pGYP7-YFP phenotype, we explored eight elaborations of the simple promoter switching model from Figure 2.4 using a sequence of fitting and cross-validation (Detailed descriptions of all models and their exploration can be found in 2.7.1 “Model exploration and sampling details for the

pGYP7-YFP phenotype: List of models” and Figure 2.12). In this process, each model was first fit to the Output-Fluorescence data in Figure 2.3; one of the eight models failed to fit. Models that fit the Output-Fluorescence data were further fit to the dose response of pGYP7-YFP, which was collected in the same way as for pCMK2-YFP and pYPS1-YFP. The pGYP7-YFP dose response was remarkably linear, and four models failed to fit it (Figure 2.12B-E). For the 3 remaining models, the dose response data served to further constrain parameter sets. For those refined parameters, we cross-validated the models on the data from an additional experiment in which we expressed Crz1*-CLASP from a stronger promoter (pTEF1 versus pADH1), and measured gene expression following a cumulative light induction of 40 minutes administered either as pulsed or continuous input. Following these rounds of fitting and cross-validation (Figure 2.12A-I), only two of the models surveyed were able to explain all the data we collected (Figures 2.6A-E, 2.12H-I).

The two models were structurally similar— they both extended the simple two-state model to contain another promoter state, thereby requiring transition through an unproductive promoter state (p_{off}) before the promoter can be fully activated. Therefore, in these models, the first transition occurred reversibly between promoter state p_0 and a non-transcribing state p_{off} with rate constants r_{on} and r_{off} , while a second transition stage occurred between p_{off} and p_{on} with rate constants k_{on} and k_{off} . Both models also involved a linear dependence on TF in the second transition stage, whose effect was to prevent the dose response from exhibiting a thresholded behavior. Finally, the two models necessitated a thresholded interaction in the first promoter transition stage, but differed in where it was applied – in one model, r_{on} was a thresholded function of TF, while in the other model, it was r_{off} that was thresholded by TF (Figure 2.6A, 2.12H-I). The threshold on either r_{on} or r_{off} acted as a “reset” for short pulses such that the system quickly returns to the starting p_0 state. Since these two models were able to recapitulate the data gathered for pGYP7-YFP in slightly different parameter regimes, we analyzed the features common to both models rather than focusing on individual values of the model parameters (k_{on} , k_{off} ,

$r_{\text{on}}, r_{\text{off}}, \gamma_1$) for pGYP7-YFP.

First, we examined one of these two models, the 3-state r_{off} -threshold model, in more depth (Figure 2.6A). To do so, we fixed r_{on} and k_{on} to values that fit the data from Figure 2.6B-D and varied r_{off} and k_{off} within a range of four logs. We then generated Output-Fluorescence plots for every parameter set and computed its corresponding slope ratio metric, which we plotted in the $\log_{10}(\frac{k_{\text{on}}}{k_{\text{off}}}) - \log_{10}(\frac{r_{\text{on}}}{r_{\text{off}}})$ plane (Figure 2.6E). Overall, we found that this model can generate both higher expression with a continuous input (slope ratio <1 , black region in Figure 2.6E, left panel) and higher expression with short pulses (slope ratio >1 , colored region on Figure 2.6E, left and right panels).

Quantitatively, there seemed to be three parameter constraints for this promoter model to elicit higher gene expression in response to a continuous input than a pulsed one. First, the rate of transition from p_0 to p_{off} should be slow; second, r_{off} should be fast relative to r_{on} ; third, k_{off} should be fast relative to k_{on} . An analysis of the 3-state r_{on} threshold model demonstrated similar requirements (Figure 2.13A-B). When r_{on} and r_{off} were increased tenfold, there were no parameter combinations that generated higher expression for continuous inputs than short pulses (Figures 2.6E, right panel, 2.6F-G, top panel). The difference in the protein outputs between the pulsed and continuous inputs was determined by the amplitude differences of promoter activity p_{on} (Figure 2.6F-G, bottom panel), which was in turn dictated by the amplitudes of depletion from p_0 for the short pulsed and continuous inputs (Figure 2.6F-G, middle panel). A slow transition from p_0 prevented the quick and full depletion of this state before a short pulse ended, while p_0 was fully depleted for the continuous input (Figure 2.6F, middle panel). By contrast, when r_{on} and r_{off} were fast, this difference disappeared as the transition from p_0 was now able to reach the same maximal amplitude in the duration of the short input (Figure 2.6G, middle panel). Hence, the incomplete depletion of the p_0 state in the duration of the short pulsed input accounted for the difference in protein outputs between the short pulsed and continuous inputs.

The requirement that the value of r_{off} be large relative to r_{on} was motivated by the fact that r_{off} dictated how quickly the promoter state transitioned back to the initial OFF state p_0 after the end of a short pulse. When the value of r_{off} decreased relative to r_{on} (Figure 2.13C), the depletion of p_0 could proceed to completion during a short pulse (Figure 2.13C, middle panel), and the resulting maximum amplitudes of the active promoter state p_{on} were more comparable for a pulsed or continuous input (Figure 2.13C, bottom panel). Lastly, as k_{off} was decreased while keeping all other parameters constant, the p_{on} to p_{off} switching also slowed, and promoter activity continued unabated between two pulses, hence maximizing the gain of promoter activity from every input pulse and causing stronger gene expression from pulses than from a continuous input (Figure 2.13D). This was in essence the same mechanism as described in Figure 2.4E-G. In summary, slow transition from the initial OFF state (p_0) to the intermediate OFF state (p_{off}) prevented the short pulsed input from achieving a quick depletion of the initial OFF state (p_0), essentially creating a filter for short inputs.

Finally, in addition to the constraints above, we found that a threshold of $\log_{10}(\frac{k_{\text{on}}}{k_{\text{off}}}) \rightarrow -1.5$ seemed to demarcate the transition between a linear and nonlinear promoter dose response in the parameter regime probed (light gray points, Figure 2.6E, left panel), therefore imposing quantitative bounds on this promoter model to exhibit a graded dose response as seen in the data.

Mechanistically, we hypothesized that the additional promoter state (p_0) and the transition through a non-transcribing promoter state (p_{off}) of the multi-state models could represent transitions induced by chromatin remodeling at the promoter. This hypothesis was further supported by analysis of previously published nucleosome occupancy data for Crz1 target genes, which showed a negative correlation between nucleosome occupancy and slope ratio (Figure 2.13E-F). To test this hypothesis more specifically, we measured H3 nucleosome occupancy of the pGYP7 and pYPS1 promoter fusions (Figure 2.6H). H3 occupancy was

much greater in pGYP7-YFP than pYPS1-YFP, which recapitulates the nucleosome occupancy profile at the native promoter/ORF junctions previously cited in literature. This result lends credence to the proposed multi-state model of pGYP7 activation by Crz1.

2.4 Discussion

In this work, we devised an integrated approach that combines engineering CLASP, a precise and modular optogenetic tool, with computational modeling and experimentation to investigate the transcriptional consequences of dynamic TF nuclear shuttling. Through precise perturbations afforded by CLASP, we were able to collect data of sufficient resolution to formulate quantitative hypotheses that can explain promoter behaviors. This productive dialogue between the technology, the modeling, and the biological findings allowed us to constrain both underlying mechanisms and quantitative parameter relationships in decoding of dynamic TF signals. Specifically, we showed that some target promoters of Crz1, a naturally pulsatile TF, had higher gene expression in response to short pulsed inputs compared to a continuous input of the same area. We demonstrated that this behavior could be explained by a two-state promoter model, and delineated its quantitative requirements of an easily saturable dose response, fast activation, and slow inactivation. By contrast, pGYP7, which had higher gene expression in response to continuous inputs than pulsed inputs of the same area and also had a linear dose response, required a more elaborate multi-state model with thresholded activation steps and a dependence on the TF at each step. These insights constitute general principles that wouldn't have been possible without the combined development of CLASP and organization of the resulting data into computational models.

The quantitative principles delineated by our experiments have clear biological implications. Recent studies have used a 3-state promoter model similar to that used for

pGYP7 to recapitulate decoding of NFkB dynamics³⁷. The multi-state promoter model we built therefore generated a robust hypothesis that the chromatin structure of pGYP7 was different from the other Crz1 target genes we profiled. To test this hypothesis, we measured nucleosome occupancy for pYPS1-YFP and pGYP7-YFP, and analyzed available occupancy data for Crz1 target genes. We found that genes that respond with higher expression to short pulses exhibited lower nucleosome occupancy (Figures 2.6H, 2.13E-F). This correlative data suggests the presence of additional promoter regulation, such as a TF-gated promoter transition between non-transcribing promoter states, for promoters that have lower gene expression for short inputs. Phenomenologically, the TF-thresholded transitions between non-transcribing states can represent transcription factor interactions with chromatin remodelers or nucleosomes⁶²⁻⁶⁸. However, mechanistic studies, such as direct observation of promoter dynamics, are needed to pinpoint the exact biochemical mechanisms⁶⁹. Additionally, eukaryotic endogenous gene regulation can be controlled by multiple TFs, phosphorylation, and other factors. Therefore, study of multiple TF-promoter relationships is needed to reflect the full complexity of gene regulation.

What biological function may differential interpretation of TF dynamic inputs carry for the Crz1 stress response? Under stress, Crz1 undergoes an initial 40-60 minute nuclear localization, followed by pulsing in the “maintenance” phase of the calcium response. Cohorts of genes could activate strongly only in the first long pulse, while other genes maintain high expression even in the short pulsing phase of Crz1 response to calcium, therefore programming a staggered response. Moreover, since Crz1 pulses exhibit different amplitudes in the “maintenance” phase (Figure 2.10A), a 2-state promoter with a dose response that saturates quickly as a function of TF, like pYPS1, would turn on strongly for all pulse amplitudes. By contrast, promoters with a linear dose response like pCMK2 would activate in a graded fashion, presenting yet another mechanism by which dynamic inputs could be differentially interpreted.

Finally, while our studies focused on decoding of TF inputs at the promoter level, other modes of differential decoding can be implemented. For example, modulation of translation and degradation of mRNA and protein are alternative strategies for differential decoding. As we discovered for pYPS1-YFP and pCMK2-YFP, different mRNA degradation rates additionally contribute to differential decoding. It will be fascinating to study the bounds of dynamic decoding explored by endogenous genes.

2.5 Figures

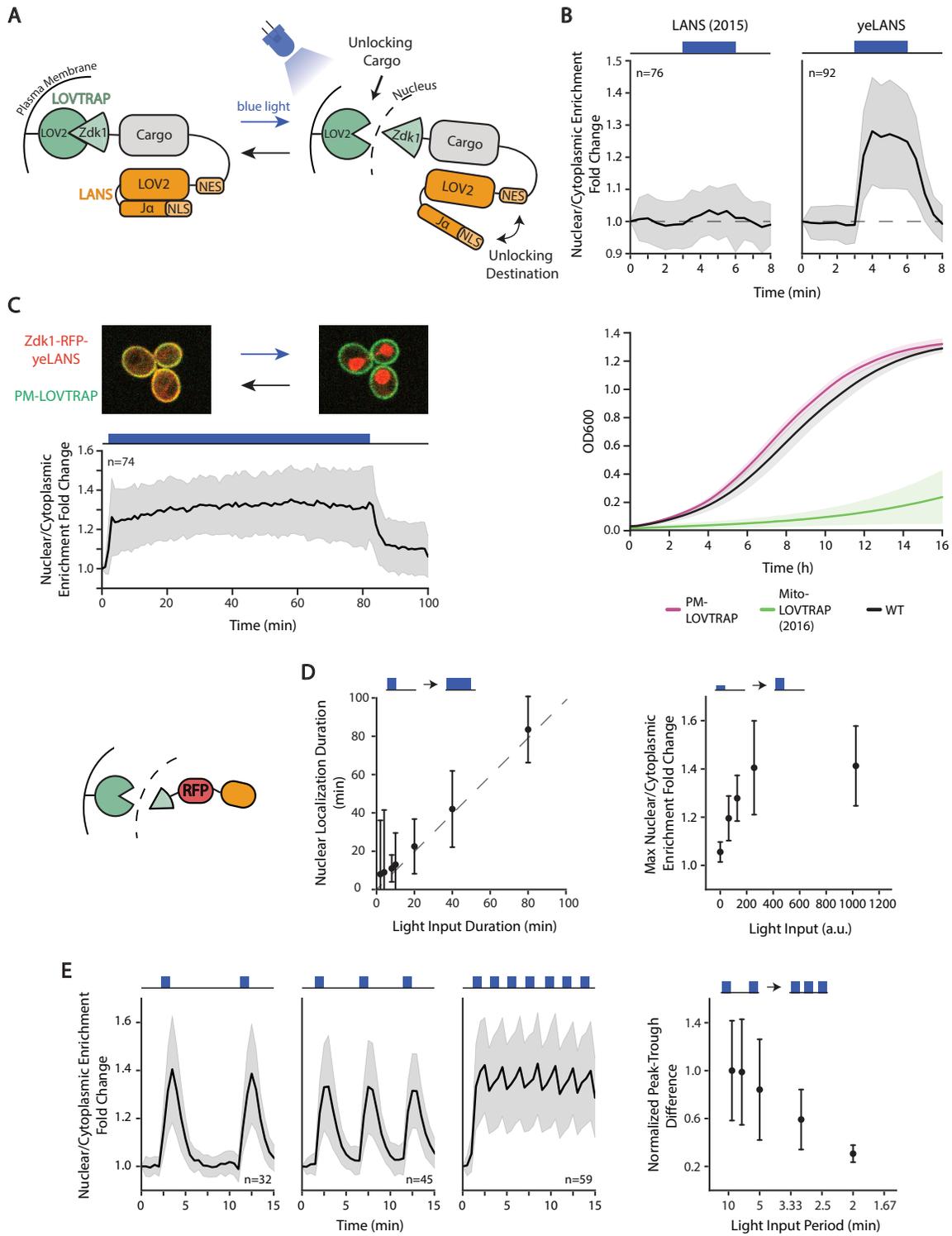


Figure 2.1: Design, Optimization, and Characterization of CLASP. A) Schematic illustrating CLASP mechanism. B) Optimization of LANS NLS (top panels) and LOVTRAP localization (bottom panel). Top panels show mean value of nuclear/cytoplasmic enrichment fold change for original NLS and optimized NLS (yeLANS) as a function of time when given a pulse of blue light. Nuclear/Cytoplasmic enrichment fold change is calculated relative to the nuclear/cytoplasmic enrichment at $t=0$. Bottom panel shows mean of OD600 in 3 independent growth experiments for original LOVTRAP targeted to mitochondria and the optimized plasma membrane targeted LOVTRAP. C) (top panel) Confocal microscopy image showing mScarlet-CLASP localization at the plasma membrane in the dark (left) and in the nucleus (right) after 3 minutes of light exposure. Images are an overlay of the mCherry and Cy7 channels. (bottom panel) Quantification of mean nuclear/cytoplasmic enrichment fold change of mScarlet-CLASP as a function of time in response to a prolonged light input (80 minutes, 1024 a.u. light input amplitude). Black line represents the mean of 74 cells. D) Quantification of the response of mScarlet-CLASP to light inputs with different dynamic characteristics. Left plot shows median time to return within 25% of basal nuclear/cytoplasmic enrichment for light pulses of different durations and constant 1024 a.u. amplitude. Median is used to minimize the effect of outliers. The dotted line is $Y=X$ line. Right plot shows the mean response to one minute light pulses of different amplitudes. Points in both plots represent at least 21 cells. E) Nuclear/cytoplasmic enrichment fold change of mScarlet-CLASP in response to light pulsing with different periods. Left three graphs show mean enrichment fold change as a function of time in response to pulsed light inputs (1 minute light given in a 9, 5, or 2 minute period, respectively) with 1024 a.u. amplitude. Right plot quantifies median peak-to-trough difference (normalized to the median peak-to-trough difference generated by the longest period). Median is used to minimize the effect of outliers. Each point in right plot represents at least 32 cells. Error bars and shaded area, except where noted, represent standard deviation to show the spread of the data. For all panels, n represents the number of cells tracked and light input regimes are depicted on top of panels. Cartoon (left of D) represents mScarlet-CLASP. yeLANS – yeast enhanced LANS, PM-LOVTRAP – Plasma Membrane LOVTRAP, Mito-LOVTRAP – Mitochondrial LOVTRAP. See also Figures S1-2.

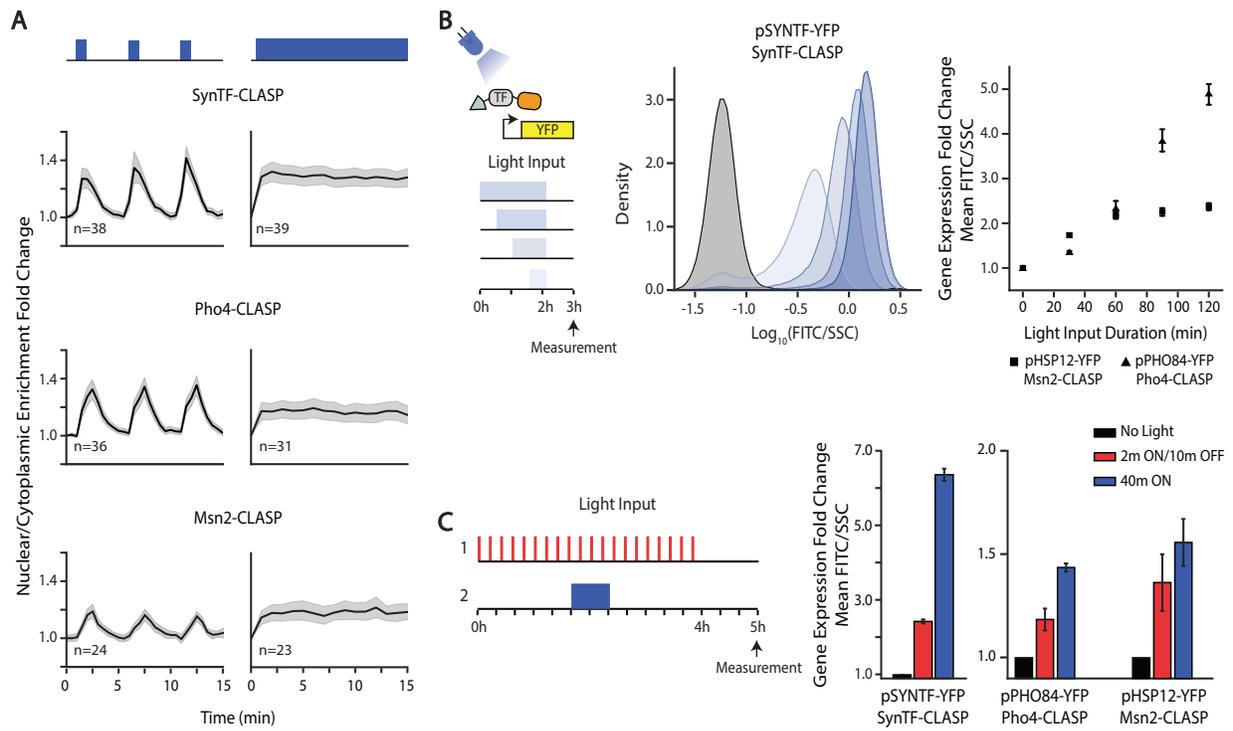


Figure 2.2: CLASP can be used to control localization of many transcription factor cargos. A) Nuclear/cytoplasmic enrichment fold change in response to pulsed (left panels) and continuous light (right panels) for several TF-CLASP cargos. Graph shows mean of single-cell traces for transcription factors tagged with CLASP. Light is delivered for one minute at the start of each five-minute period or continuously. Shaded gray area represents 95% confidence interval and light inputs are represented in blue above graphs. n represents number of cells tracked. B) Fluorescent reporter expression due to TF-CLASP localization. Left panel shows a schematic of the experiment – the TF is localized to the nucleus for 0.5, 1, 1.5 or 2hrs. A fluorescent reporter is measured via flow cytometry one hour after light shut-off. Center panel shows the population response of pSYNTF-YFP (promoter downstream of SynTF-CLASP) for inputs shown on the left. Darker blue shades correspond to longer light duration. Black histogram corresponds to no light. Right panel shows quantification of the YFP fold change as a function of light duration for promoters responsive to other TF-CLASP constructs following the same experimental protocol. Fluorescence readings are normalized by side scatter and then normalized to the 0min dose for each strain to show fold change. Error bars represent standard error of the mean for 9 biologically independent replicates. C) Fluorescent reporter response to pulsatile versus continuous localization of different TF-CLASP constructs. TF-CLASP constructs are given either 20 two-minute pulses of light or 1 forty-minute pulse of light, as depicted in the schematic on the left. Reporter expression is measured via flow cytometry one hour after light shut-off. Right panels show quantification of YFP fold change in response to pulsed light input, continuous light input, or no input. Error bars represent standard error of the mean for 9 biologically independent replicates. In all panels, strains are induced with a given amplitude of light (SynTF-CLASP – 1024 a.u.; Msn2-CLASP – 2048 a.u., Pho4-CLASP – 4095 a.u.). See also Figure 2.9.

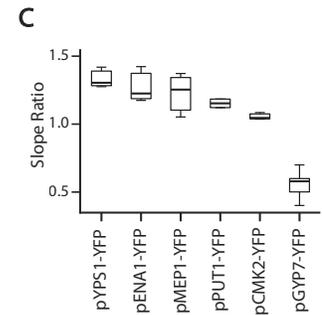
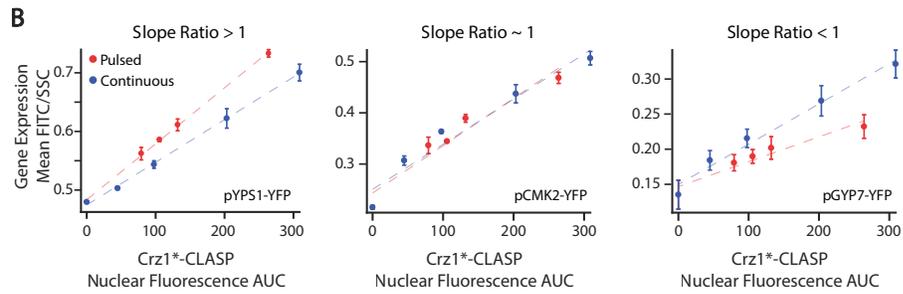
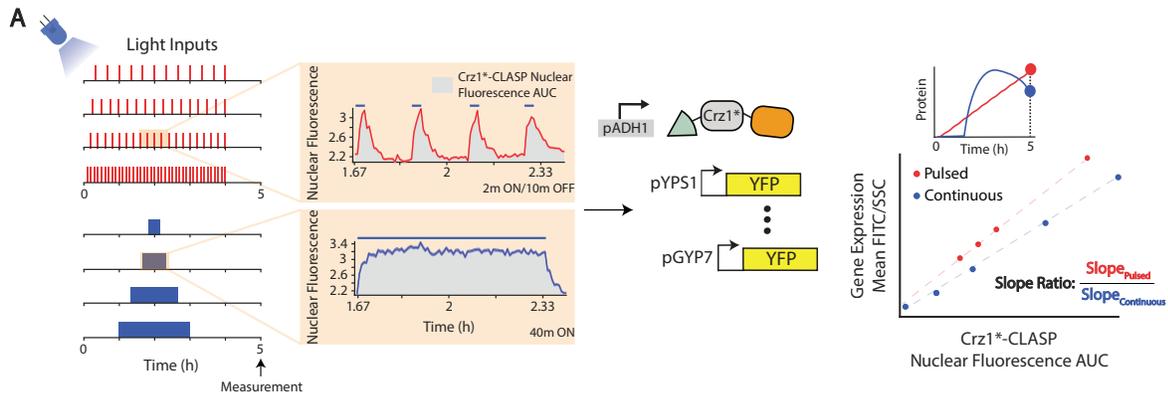


Figure 2.3: Crz1 target genes show differing interpretation to Crz1*-CLASP short nucleo-cytoplasmic pulses.A) Schematic of experimental setup used. Two types of light inputs are given to cells expressing Crz1*-CLASP: 2 minute pulses with decreasing period (20, 15, 12, and 6 minute periods) and single pulses with increasing duration (20, 40, 80, 120 minutes). Light-induced Crz1*-CLASP nuclear localization is measured with fluorescence microscopy. The mean of single cell fluorescence values is plotted (solid red for pulsed input or blue line for continuous inputs), with the shaded area representing 95% confidence interval (red or blue shading). Cells tracked for the pulsed and continuous inputs are 187 and 91 cells, respectively. Crz1*-CLASP nuclear fluorescence AUC (x-axis in rightmost panel) is quantified as the area under the nuclear fluorescence traces (gray shading in middle panel). Gene expression (mean FITC/SSC) is measured for 6 promoter fusions of target gene driving a fluorescent protein (YFP) at 5 hours after light input. A schematic shows gene expression values for different light input regimes are plotted as a function of nuclear fluorescence AUC, generating the Output-Fluorescence plot referred to in the text. Each point in the plot is an endpoint measurement of gene expression, as highlighted by the YFP time course schematic above. Red circles represent output fluorescence for short 2 minute pulses with increasing period, and blue circles represent that for continuous single pulse with increasing durations. A best fit line (red for pulsed inputs and blue for continuous inputs) is fit through the data points for the pulsed and continuous inputs. For each Output-Fluorescence plot we define the slope ratio as the ratio of the slope of the pulsed to continuous best fit lines. B) Output-Fluorescence plot for three representative Crz1 target promoters pYPS1-YFP, pCMK2-YFP, and pGYP7-YFP. The error bars are standard deviation of at least 3 biological replicates. C) Slope ratios for 6 Crz1 target genes plotted in order of highest to lowest slope ratio. Data for 3 biological replicates is plotted. In all panels, Crz1*-CLASP is induced with a 512 a.u. light input. See also Figure 2.10.

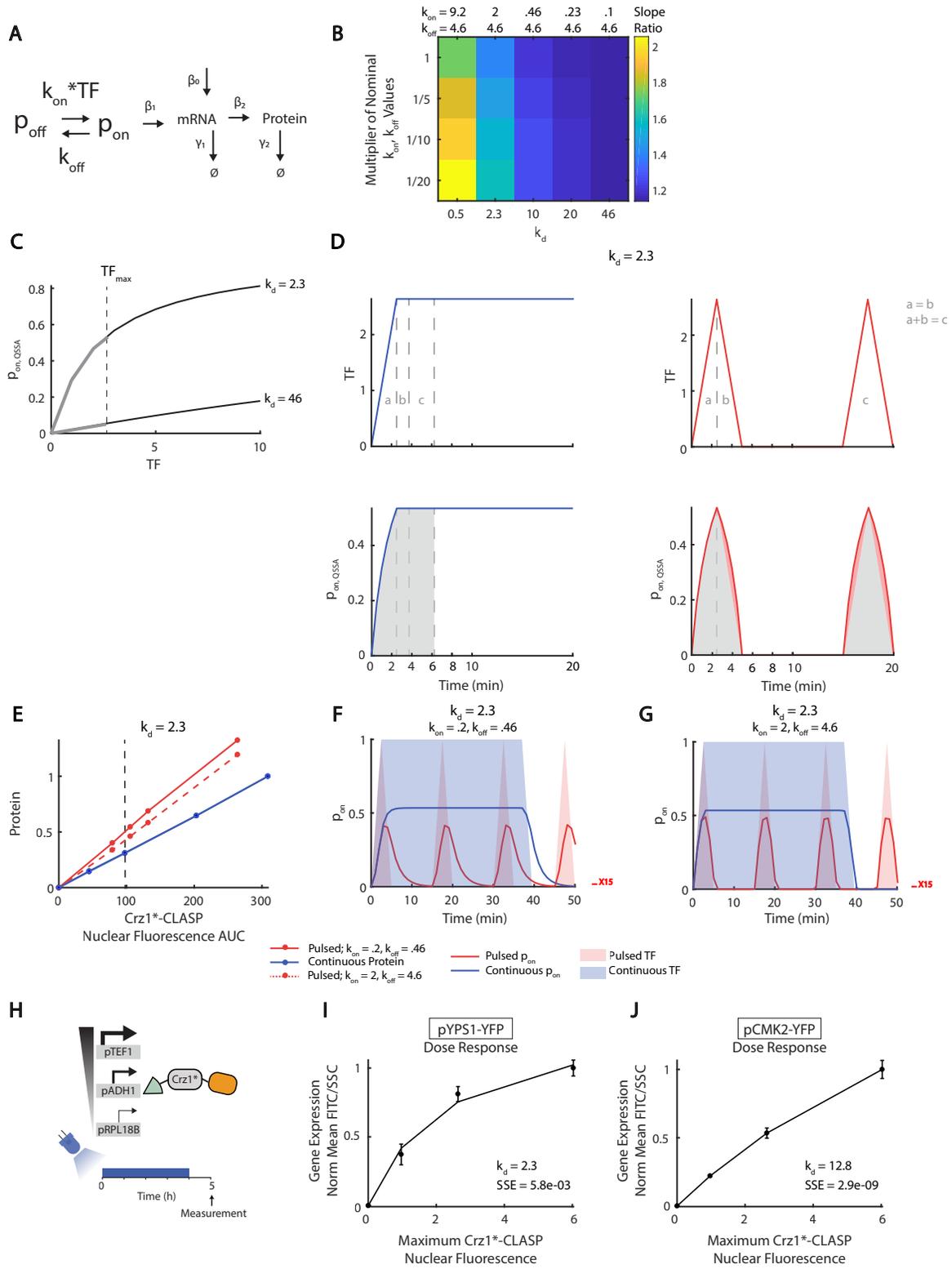


Figure 2.4: Higher gene expression in response to short pulses by promoters occurs when the dose response is saturated at low TF concentration. A) Schematic of a two-state promoter model, where the input is Crz1*-CLASP nuclear localization (TF) and the output is fluorescent protein level (Protein). The promoter turns ON with rate constant k_{on} and turns OFF with rate constant k_{off} . B) Heatmap of slope ratio for increasing k_d and different values of k_{on} and k_{off} . Each column has a given value of k_d and each row has different values for k_{on} and k_{off} that produce the same k_d . The nominal k_{on} and k_{off} values used in the first row are noted at the top of each column, and every subsequent row uses a fraction of these values (1/5, 1/10, and 1/20). The values of β_1 , β_2 , and β_0 are 2.01, 4.92, and 0.0032, respectively. C) Plot of p_{on} as a function of TF for $k_d = 2.3$ and 46, assuming a fast promoter. This quantity is denoted as p_{on} , QSSA and calculated as p_{on} , QSSA = $\frac{TF}{TF+k_d}$. The dotted line represents max TF input, TF_{max} , which is 2.6. D) Plot of p_{on} , QSSA as a function of time assuming quasi-steady state of promoter dynamics as in panel (C). In these panels, $k_d = 2.3$. (Top panels) Red and blue lines represent pulsed and continuous TF inputs, respectively. Gray lines and text denote equivalent area of TF input. The area labeled “a” represents the rise for both pulsed and continuous inputs. The area labeled “b” represents the fall of the pulsed input, and the equivalent area for the continuous input. The area labeled “c” represents a single pulse of the pulsed input, and the equivalent area of the continuous input. The area labeled “c” is equivalent to the sum of the areas labeled “a” and “b”. The areas labeled “a” and “b” are equivalent to each other. (Bottom panels) Red and blue lines represent p_{on} , QSSA in response to pulsed and continuous TF inputs, respectively. Gray shading denotes equivalent area of p_{on} , QSSA for continuous and pulsed inputs. Light red shading denotes excess p_{on} , QSSA area resulting from the rise and fall of the pulsed input. E) Output-Fluorescence plots generated by the model for two parameter sets that qualitatively represent pYPS1-YFP. The solid lines represent $k_{on} = 0.2$ and $k_{off} = 0.46$. The dashed lines represent $k_{on} = 2$ and $k_{off} = 4.6$, and $k_d = \frac{k_{off}}{k_{on}} = 2.3$ for both parameter sets. The red lines represent the output of the pulsed input. The blue lines represent the output of the continuous input; for both parameter sets, this output is the same. F) Plot of p_{on} as a function of time for continuous and pulsed inputs for $k_d = 2.3$ with $k_{on} = 0.2$, $k_{off} = 0.46$. Red and blue solid lines represent the p_{on} resulting from pulsed and continuous inputs, respectively. The red and blue shading represent pulsed and continuous TF inputs, respectively. G) Same as F for $k_{on} = 2$, $k_{off} = 4.6$. H) Schematic of dose response experiment. Cells with different expression levels of Crz1*-CLASP are induced with light for 4 hours and YFP expression is measured after 5 hours. I) The experimental dose response for pYPS1-YFP was fit to the equation: normalized protein output = $\frac{C \cdot TF}{TF+k_d}$ where TF = maximum Crz1*-CLASP nuclear fluorescence, C = scaling factor, and $k_d = \frac{k_{off}}{k_{on}}$. k_d and squared error of prediction (SSE) of the fit for each gene is noted in the bottom right corner of the plot. J) Same as (I) for pCMK2-YFP. For I-J, error bars represent standard deviation of 3 biologically independent replicates. See also Figures S5, S8-9.

Figure 2.5: Time course measurements of protein output constrain parameter relationships. A) (left panel) Schematic of the input used for protein time course experiment. Cells are induced with constant blue light and pYPS1-YFP or pCMK2-YFP expression is measured continuously (every 8 minutes) throughout induction for two hours. (middle panel) Plot of normalized protein expression (FITC/SSC) as a function of time for pYPS1-YFP. The magenta lines represent fits through the data (plotted as black dots with error bars) for model in Figure 2.4A. The model was simulated using 33,000 parameter sets varying k_{on} , k_{off} , and γ_1 , and fit to the dynamic gene expression data was assessed. k_{on} was varied from 0.001-10, k_{off} from 0.000007-100, and γ_1 from 0.01-10. (right panel) Same as middle panel for pCMK2-YFP. For both panels, β_1 was set to 0.1, β_2 set to 0.06, γ_2 set to 0.0083, and β_0 set to 0.001. For middle and right panels, error bars represent standard deviation of 3 biologically independent replicates. B) (left panel) Plot of $\log_{10}(\gamma_1)$ as a function of $\log_{10}(k_{on}TF + k_{off})$ for pYPS1-YFP. Magenta dots represent 2355 parameter fits to the dynamic protein time course data (continuous input) as discussed in Figure 2.5A middle panel. Gray dots represent parameters that were tested but did not fit to the data. (right panel) Same as left panel for pCMK2-YFP. Blue dots represent 807 parameter fits to the dynamic protein time course data (continuous input) as discussed in Figure 2.5A right panel. Gray dots represent parameters that were tested but did not fit to the data. C) (left panel) Schematic of input for Output-Fluorescence experiment. Experiment is as described in Figure 2.3A. (middle panel) Output-Fluorescence plot of simulated outputs and data for pYPS1-YFP. Parameters determined to fit the dynamic protein time course with a continuous input are used to predict the Output-Fluorescence data. Red and blue lines represent the model outputs for all parameters that fit the Output-Fluorescence and protein time course data from panel A. Red and blue circles and error bars represent experimentally measured means and standard deviations for pulsed and continuous inputs, respectively, for 3 biologically independent replicates. (right panel) Same as middle panel for pCMK2-YFP. D) (left panel) Plot of $\log_{10}(\gamma_1)$ as a function of $\log_{10}(k_{on}TF + k_{off})$ for pYPS1-YFP, where the magenta dots represent 300 parameters that fit both the Output-Fluorescence and dynamic protein time course (continuous input). Gray dots represent parameters that were tested but did not fit to the data. (right panel) Same as left panel for pCMK2-YFP. The blue dots represent 321 parameters that fit both the Output-Fluorescence and dynamic protein time course data (continuous input). E) (left panel) Schematic of input for protein time course experiment for pulsed and continuous inputs. Cells were induced with either constant light or pulsed light (2m ON/4m OFF) and gene expression was measured every 8 minutes. (middle panel) Plot of normalized protein expression (FITC/SSC) as a function of time for pYPS1-YFP for the experiment denoted in the left panel. Red and blue circles and error bars denote experimentally measured means and standard deviations for pulsed and continuous inputs, respectively, for 3 biologically independent replicates.

Figure 2.5: Red and blue lines represent the model outputs for the experiment denoted in the left panel, using parameters that fit the dynamic protein time course (continuous input) and Output-Fluorescence data. (right panel) Same as left panel for pCMK2-YFP. See also Figure 2.14.

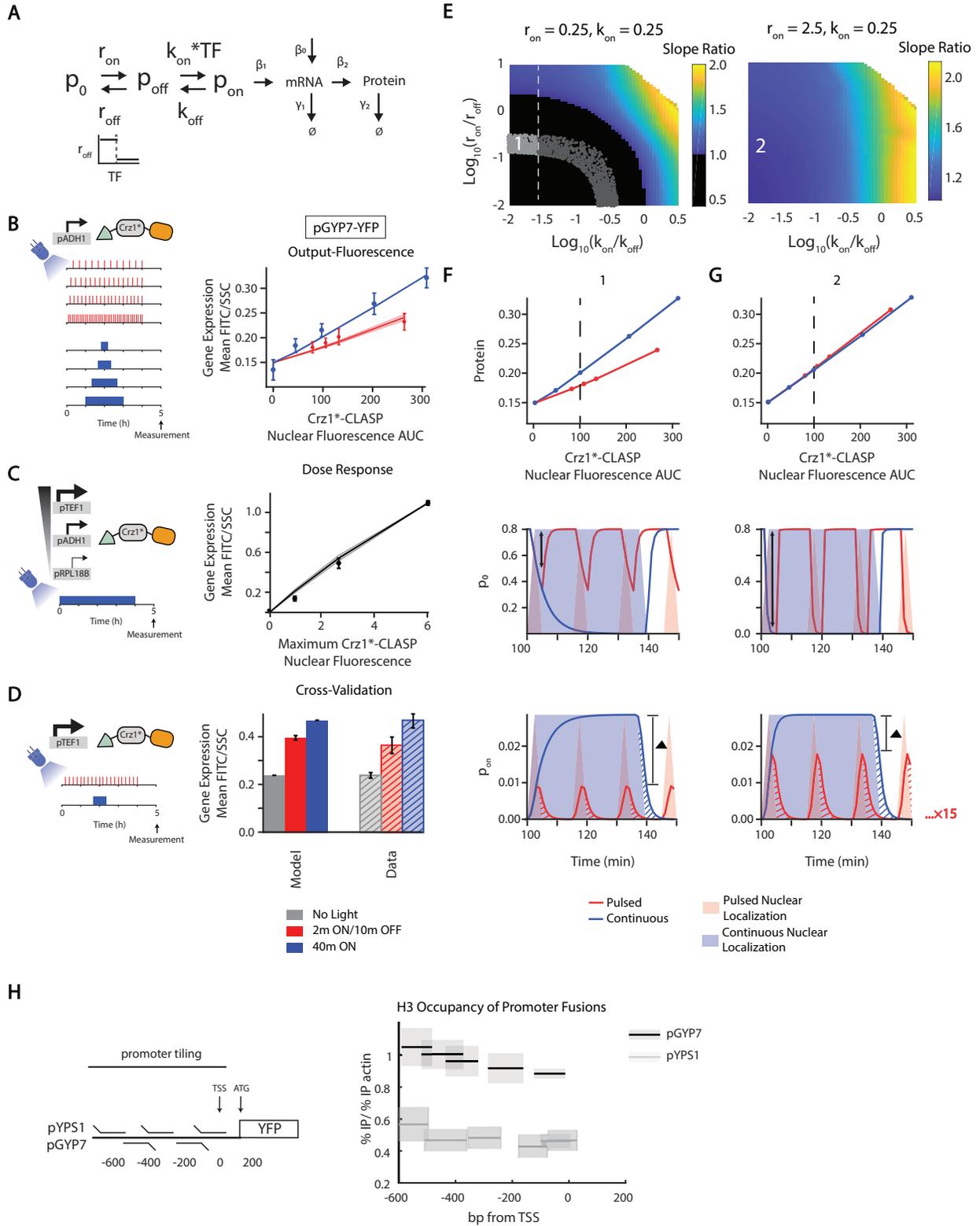


Figure 2.6: Higher gene expression in response to continuous inputs by promoters can be explained by a model with two transition states and with a thresholded transition between non-transcribing promoter states. A) Schematic of the three-state model where r_{off} , the inactivation rate constant from p_0 to p_{off} , is thresholded by TF concentration and where the activation from p_{off} to p_{on} is linearly dependent on TF. B) (left panel) Schematic of experimental setup. (right panel) Output-Fluorescence plot for pGYP7-YFP. Circles are experimentally measured values while lines denote the mean model output for 96 parameter sets that fit the data points within the error bars, the same metric as used in Figure 2.5. The solid line denotes the mean and shaded areas denote the standard deviation of the model outputs for these parameter sets. Parameters were sampled (r_{on} from 0.1-100, r_{off} from 0.1-100, k_{on} from 0.0001-1, k_{off} from 0.0001-1, β_1 from 0.0001-10, β_0 from 0.000001-0.01, threshold from 0-0.5) or set ($\beta_2 = 0.06$, $\gamma_1 = 0.05$, $\gamma_2 = 0.0083$). Red circles, error bars, and lines relate to the pulsed input, while blue circles, error bars, and lines relate to the continuous input. Error bars show standard deviation from 3 biologically independent replicates. C) (left panel) Schematic of experimental setup. (right panel) Dose response plot for pGYP7-YFP. The parameters that fit the Output-Fluorescence data were used to further fit the dose response of pGYP7-YFP using a least squared error criterion (25 parameter sets). Solid black line is the mean generated by the model. The black circles are the mean of the experimentally measured dose response and error bars are the standard deviation of 3 biologically independent replicates. D) (left panel) Schematic of experimental setup. (right panel) The parameters that fit the Output-Fluorescence are subjected to cross-validation using an experiment where Crz1*-CLASP expression is increased (construct expressed from a pTEF1 promoter), and cells are exposed to either short pulsed (2 minutes ON/10 minutes OFF) or continuous input (40 minutes of light). The model generated outputs (solid gray, red, and blue bars) are plotted with the experimental data (hashed gray, red, and blue bars). The gray bars correspond to no light input. E) (left panel) Heatmap shown in the $\log_{10}(\frac{k_{\text{on}}}{k_{\text{off}}}) - \log_{10}(\frac{r_{\text{on}}}{r_{\text{off}}})$ plane of slope ratio of Output-Fluorescence relationship resulting from the model in (A). Parameters are sampled (r_{off} from 0.0025-25, k_{off} from 0.0025-25) or set ($r_{\text{on}} = 0.25$, $k_{\text{on}} = 0.25$, $\beta_1 = 0.0001$, $\beta_2 = 0.06$, $\gamma_1 = 0.05$, $\gamma_2 = 0.0083$, threshold = 0.5, $\beta_0 = 0.000001$). Point 1 highlights a parameter set that fits the output-fluorescence, dose response, and cross-validation datasets for pGYP7-YFP. Black region is where slope ratio < 1 . Gray dotted line indicates when $\log_{10}(\frac{k_{\text{on}}}{k_{\text{off}}}) \approx -1.5$, at which point the dose response changes from linear to nonlinear with increase in the $\log_{10}(\frac{k_{\text{on}}}{k_{\text{off}}})$ value. All parameters that show a qualitative fit to Output-Fluorescence data are displayed as light and dark gray dots. The light gray dots represent parameter sets where all pGYP7-YFP data are quantitatively fit. (right panel) Heatmap of slope ratio as in (E, left panel) with a $r_{\text{on}} = 2.5$, 10 times larger than that in (E, left panel). k_{on} is also set to 0.25. Parameters are sampled (r_{off} from 0.025-250, k_{off} from 0.0025-25) or set ($\beta_2 = 0.0001$, $\beta_2 = 0.06$, $\gamma_1 = 0.05$, $\gamma_2 = 0.0083$, threshold = 0.5, $\beta_0 = 0.000001$). Point 2 highlights the effect of increasing both r_{on} and r_{off} while maintaining the ratio $\log_{10}(\frac{r_{\text{on}}}{r_{\text{off}}})$.

Figure 2.6: F-G) (upper panels) Output-Fluorescence plots generated by the model for different parameter sets that correspond to points 1 and 2 in the heatmaps of panel E. The slope ratio for point 1 is 0.51 with $\log_{10}(\frac{k_{on}}{k_{off}}) = -1.58$ and $\log_{10}(\frac{r_{on}}{r_{off}}) = -0.89$. The slope ratio for point 2 is 1.04 with $\log_{10}(\frac{k_{on}}{k_{off}}) = -1.58$ and $\log_{10}(\frac{r_{on}}{r_{off}}) = -0.89$. (middle panels) Example of a time course of promoter state p_0 for a light input that produces the equivalent of 40 minutes (dotted line in upper panel) in nuclear localization either continuously or in short pulses. Solid lines are the p_0 pulses while shading denotes TF nuclear localization. The black double arrows denote the maximum depletion of the p_0 state for the pulsed input. (lower panels) Example of a time course of promoter activity p_{on} for a light input that produces the equivalent of 40 minutes (dotted line in upper panel) in nuclear localization either continuously or in short pulses, similar to middle panels. The red and blue hashes represent residual promoter activity beyond the TF nuclear localization input. The red residual promoter activity is repeated 15 times while the blue residual activity is repeated one time. The Δ bar denotes the difference between the amplitudes generated by the 2 minute pulsed and 40 minute continuous input. H) (left panel) Schematic of chromatin immunoprecipitation experiment. (right panel) H3 histone occupancy is plotted for regions of the promoter fusions pYPS1-YFP and pGYP7-YFP. H3 histone occupancy is calculated as the ratio of % immunoprecipitation (% IP) of the promoter fusion target to % immunoprecipitation of an actin control. % Immunoprecipitation is calculated relative to the input DNA. Black lines show the mean measured value and gray shading shows the standard deviation of 3 biologically independent replicates. See also Figures S6-7.

2.6 Materials and Methods

2.6.1 Resource Availability

Lead Contact Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Hana El-Samad (hana.el-samad@ucsf.edu).

Materials Availability

To request reagents, please submit a form to UCSF at <https://ita.ucsf.edu/researchers/mta>. Key plasmids have been deposited on Addgene and can be requested from there. For other plasmids, please contact the Lead Contact.

Data and code availability

All microscopy, flow cytometry, RNA-seq, ChIP-qPCR data, and modeling results (except for those for Fig S6A-I) can be accessed on Mendeley Data at

<http://dx.doi.org/10.17632/jxjnjmmj83.1>. Raw microscopy images supporting the current study and modeling results for Fig S6A-I have not been deposited in a public repository due to file size constraints, but are available from the Lead Contact on request. All code necessary to generate the figures is available via Github at <https://github.com/lindseyo/CLASP-Crz1>.

2.6.2 Experimental model and subject details

Saccharomyces Cerevisiae

Plasmid and strain construction

Hierarchical golden gate assembly was used to assemble plasmids for yeast strain construction using the method in Lee et al⁷⁰. BsaI, BsmBI, and NotI cut sites were removed from individual parts to facilitate downstream assembly and linearization. Parts were either generated via PCR or purchased as gBlocks from IDT. For promoters, these parts incorporate the 1000 bp upstream of the open reading frame. These parts were then assembled into transcriptional units (promoter-gene-terminator) on cassette plasmids. These cassettes were assembled together to form multi-gene plasmids for insertion into the yeast genome at the TRP, URA, or LEU locus. Cassette plasmids were grown and prepared from either DH5alpha or Mach1 competent cells (Macrolab, Berkeley, CA). Cassettes were digested with NotI and then transformed into yeast as described in Lee S et al., 2013⁷¹ or Lee ME et al., 2015⁷⁰.

Yeast strains, media, and growth conditions

The base *S. cerevisiae* strain used for experimentation was W303 α or BY4741. Base strain for each engineered strain is noted in the strain list. From these base strains, knockout of endogenous transcription factors was done with a one-step replacement using a plasmid that contains 40 base pair overlaps in the 5' and 3' UTR of the transcription factor (Gardner and Jaspersen, 2014)⁷². The 40 base pair overhangs flank the Candida Albicans HIS selectable marker.

Single colonies were picked from auxotrophic SD (6.7 g/L Bacto-yeast nitrogen base without amino acids, BD Difco, Franklin Lakes, NJ; 2 g/L supplement amino acid mix minus necessary amino acids, MP Biomedical, Irvine, CA; 20 g/L dextrose, Sigma-Aldrich, St Louis, MO) agar plates. For microscopy and growth measurement studies, colonies were picked into 1 ml SDC media. For flow cytometry studies, colonies were picked into 1 ml YPD (yeast extract, Alfa Aesar, Haverhill, MA; peptone, BD Biosciences, Franklin Lakes, NJ; 2% glucose, Sigma-Aldrich, St Louis, MO) or SDC (6.7 g/L Bacto-yeast nitrogen base without amino acids, BD Difco, Franklin Lakes, NJ; 2 g/L complete supplement amino acid mix, MP Biomedical, Irvine, CA; 20 g/L dextrose, Sigma-Aldrich, St Louis, MO) media. Colonies were grown overnight from 30°C to saturation. Prior to the start of an experiment, cells were diluted into 1-3 ml of SDC and grown for 4 hours to an OD of 0.05-0.1 prior to the start of an experiment. A TECAN Spark 10M plate reader (TECAN, Mannedorf, Switzerland) was used for growth measurements.

Flow Cytometry

Blue light optogenetic stimulation of samples was done using a custom built “optoPlate” as described in Bugaj et al⁷³. Analysis of fluorescent protein reporter expression was performed with a BD LSRII flow cytometer (BD Biosciences, Franklin Lakes, NJ) equipped

with a high-throughput sampler. For steady-state measurements, cultures were diluted in TE before running through the instrument. Cultures were run on the instrument 1 hour (+/- 20 min) after optical stimulation using the optoPlate, to allow for YFP maturation. YFP (Venus) fluorescence was measured using the FITC channel and RFP (mCherry/mScarlet) was measured using the PE-Texas Red channel. For steady-state measurements, a maximum of 10,000 events were collected per sample.

Growth Assays

Growth was measured using a TECAN Spark 10M plate reader (TECAN, Mannedorf, Switzerland) using 600nm excitation. Cultures were plated into Corning 3904 96-well assay plates (Corning, Corning, NY) and grown at 30°C while shaking until saturation.

Treatment with CaCl₂ stress

Cells were grown at 30°C in YPD medium to saturation overnight. Cells were then diluted prior to the start of an experiment and grown for 4 hours to an OD of 0.05-0.1. For microscopy experiments, cells were plated in SDC with concanavalin A (conA; Sigma-Aldrich, St Louis, MO) for 15 minutes to adhere them to the bottom of the glass imaging plate. Prior to imaging, the SDC was removed and replaced with a solution of SDC with 0.2M CaCl₂ (Fisher Scientific, Waltham, MA). For flow cytometry experiments, cells in SDC were diluted to OD 0.1 in a media of SDC with 0.2M CaCl₂ and grown in the media for the duration of the experiment. Prior to measurement, the 0.2M CaCl₂ media was removed by centrifugation with 3 washes in 1X TE (Fisher Scientific, Waltham, MA).

Automated Flow Cytometry

Cells were grown at 30°C in YPD medium to saturation overnight. Cells were then diluted prior to the start of an experiment to 0.15 OD and grown for 4 hours in SDC media. Cells were then back diluted to 0.2 OD into SDC media with 0.5X Penicillin-Streptomycin to prevent culture contamination (Thermo Fisher Scientific, Waltham, MA ; 10,000 U/mL). Cells were then outgrown for 1 hour at 30°C while shaking on the automated flow cytometer setup (described in Harrigan et al, 2018⁷⁴). Following the outgrowth, cells were illuminated with 40 mA pulsed or continuous blue light (455 nm) and sampled every 8 minutes for 6 hours using custom LabView scripts.

Chromatin Immunoprecipitation

Chromatin Immunoprecipitation (ChIP) followed by qPCR was performed as described in Greenstein et al, 2018⁷⁵ with the following modifications. *S.cerevisiae* cells were grown at 30°C in YPD overnight to saturation. Cells were then diluted prior to experiment to 0.2 OD and then grown for 4 hours in SDC media. To fix cells, 1% formaldehyde (Thermo Fisher Scientific, Waltham, MA) was added directly to the media and cultures were incubated with shaking for 15 minutes at 30°C. Fixation was quenched with 0.25M glycine (Fisher Scientific, Waltham, MA) for 5 minutes at 30°C. Cell pellets were washed twice with cold 1xTBS (Teknova, Hollister, CA) and flash frozen prior to lysis. Cells were lysed using a Mini Bead-Beater (Biospec, Bartlesville, OK) with 7 rounds of 1 minute ON followed by 2 minute incubations on ice. Sonication was performed using a Diagenode BioRuptor Standard machine (Diagenode, Liege, Belgium) for 30 rounds of 30s ON/30s rest at 4°C. The insoluble fraction was removed by centrifugation and then pre-cleared with Protein A Dynabeads (Thermo Fisher Scientific, Waltham, MA) for 3 hours with rotation. Beads were then removed with a magnetic stand (Dyna/Thermo Fisher

Scientific, Waltham, MA). 2.1µg of H3 antibody (Active Motif, Carlsbad, CA; 39064) was added per ChIP sample after a 25% was set aside as Input/WCE. Following overnight incubation with antibody, immune complexes were collected with Protein A Dynabeads (Invitrogen) and washed as described⁷⁵ with the exception that the wash buffer step was performed twice. DNA was quantified by RT-qPCR and %IP (ChIP DNA / Input DNA) was calculated as described⁷⁵. %IP values for each qPCR target were normalized to the %IP value for ACT1, an internal positive control.

RNAseq of Crz1 19A and 5A mutant

Single colonies were picked and grown to saturation in YPD at 30°C overnight. Cells were then diluted in SDC and grown for 4 hours to an OD of 0.3. Cells were harvested by centrifugation and frozen with liquid nitrogen. RNA was extracted using phenol chloroform (Sambrook and Russell, 2006; Thermo Fisher Scientific, Waltham, MA). RNA quality was assessed using the Agilent RNA Pico kit (Agilent, Santa Clara, CA). The Lexogen Quantseq 3' mRNA-Seq Library Prep Kit (Lexogen, Vienna, Austria) was used for RNA preparation. mRNA libraries were quantified using Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA) and subject to single-end sequencing on an Illumina HiSeq 4000.

Delivery of stress inputs for microscopy

For each environmental perturbation, cells were grown overnight to saturation in YPD, diluted in SDC prior to the experiment, and grown to an OD of 0.1. 200ul cells were plated with conA. Just before imaging, the SDC media was removed from the microscopy well and the appropriate environmental stress media was applied to the cells. The media for glucose depletion consisted of 0.67% YNB w/o AA w/ ammonium sulfate, 0.79% CSM, 0.05%

glucose. The media for osmotic shock was composed of 0.67% YNB w/o AA w/ ammonium sulfate, 0.79% CSM, 2% glucose, and 0.95M sorbitol⁵³ (Sigma-Aldrich, St Louis, MO).

2.7 Supplemental Text

Measuring basal and constitutively nuclear gene expression of TFs

To further assess how TF-CLASP-induced expression compares to endogenous gene expression, we measured the level of reporter gene expression when the TFs were constitutively localized to the nucleus by C-terminally tagging them with the same NLS used in yeLANS (TF-NLS), or in their basal localization by C-terminally tagging them with only mScarlet. All TF-NLS, TF-mScarlet, and TF-CLASP constructs were expressed from pRPL18b. We compared this value to expression achieved when TF-CLASP was induced with 2 hours of blue light. SynTF-CLASP achieved 52% of pSYNTF-YFP expression produced through constitutive nuclear localization of SynTF (Figure 2.9B). Furthermore, the mean SynTF-CLASP-induced gene expression in the dark (.09) was similar to the mean basal gene expression in a strain in which the SynTF was only tagged with mScarlet (.07) (Figure 2.9B). Pho4-CLASP activated pPHO84-YFP to 14% of the gene expression achieved with constitutive nuclear localization (Figure 2.9C) while Msn2-CLASP was more efficient at inducing pHSP12-YFP gene expression than constitutive Msn2 nuclear localization (23% greater expression, Figure 2.9D). Since Msn2 is subject to faster degradation in the nucleus^{76,77}, transient localization with CLASP may be more efficient at inducing gene expression. For both pHSP12-YFP and pPHO84-YFP, reporter expression in the dark was lower in a strain that had either Msn2-CLASP or Pho4-CLASP than in their respective controls with either Msn2 or Pho4 when only tagged with mScarlet (28% and 88% lower, respectively). In fact, pPHO84-YFP showed basal bimodal expression in the constitutively expressed Pho4 strain, but not in the Pho4-CLASP strain (Figure 2.9C).

These data suggest that CLASP can potentially sequester TFs in the dark.

Effect of protein and mRNA half-lives on slope ratio

Large degradation rates (hence short half-lives) of mRNA and protein result in large slope ratio, which decreases as these rates decrease (Figure 2.11B). This is not unexpected given that the continuous input ends before the pulsed input. Since the slope ratio is a comparison of the continuous and pulsed inputs measured one hour after the experiment ends, with larger degradation rates, the protein output of the continuous input is reduced more than the output of the pulsed input.

Computational modeling

Ordinary differential equation (ODE) models of gene expression focusing on promoter kinetics were constructed. For the simple kinetic model that described higher gene expression in response to short pulses, a model was constructed with three state variables and seven parameters. Nine models were constructed and tested for higher gene expression in response to continuous pulses. These models either contained three or five state variables with up to ten parameters. Latin hypercube sampling was done to randomly sample parameters. ODE solvers 45 and 113 in Matlab were used. Least squared error and fit within the error bars of the data were metrics used to obtain model fits. More details of the modeling methods are below.

Model equations and sampling details of the pYPS1-YFP and pCMK2-YFP phenotypes

Simple Two-State Promoter Model - This model described the higher gene expression in response to short pulses for pYPS1-YFP. The model described a two-state promoter that activates mRNA production which then activates protein production and is depicted in Figure 2.4A. We modeled these interactions using equations (1.4)-(1.6), which are detailed in the main text as (1.1)-(1.3):

$$\begin{aligned}\frac{dp_{on}}{dt} &= k_{on}p_{off}TF(t) - k_{off}p_{on} \\ &= k_{on}(1 - p_{on})TF(t) - k_{off}p_{on}\end{aligned}\tag{2.4}$$

$$\frac{dmRNA}{dt} = \beta_0 + \beta_1 p_{on} - \gamma_1 mRNA\tag{2.5}$$

$$\frac{dProtein}{dt} = \beta_2 mRNA - \gamma_2 Protein\tag{2.6}$$

In these equations, p_{on} represented promoter activity due to increased nuclear localization of Crz1*, while mRNA and Protein represented concentration of mRNA and protein, respectively. Here we assumed that promoter activity was conserved such that $1 = p_{on} + p_{off}$. TF represented the concentration, or nuclear fluorescence, of nuclear transcription factor. The model was characterized by 7 parameters. Most of the activation/inactivation and production/degradation terms were modeled by first-order mass action kinetics. The parameter β_0 was zeroth-order, to reflect basal promoter activity. We chose this simple model form because we were interested in a parsimonious model that could explain the experimental phenotype of pYPS1-YFP. Note that the rate of promoter activation was dependent on TF concentration because Crz1 has been shown to activate

genes through binding of a known promoter element, the calcineurin-dependent response element (CDRE), through its zinc finger domain⁵⁵. The input to the model was the concentration, or nuclear fluorescence, of nuclear transcription factor (TF), while the output represented protein concentration (Protein).

Dose Response Fitting - The experimental dose response was fit to the equation protein output = $\frac{C \cdot TF}{(TF + k_d)}$ where TF = amplitude of transcription factor input, C = scaling factor, and $k_d = k_{off}/k_{on}$. The model fit showed a k_d of 2.3 for pYPS1-YFP (with a squared error of prediction (SSE) of 5.8e-03) and a k_d of 12.8 for pCMK2-YFP (with an SSE of 2.9e-09).

Parameter Search for Figure 2.11 - The equations were numerically solved by the ODE solver ode113 for nonstiff differential equations via MATLAB. Parameters k_{on} , k_{off} , β_0 , and β_1 were sampled over 4-5 orders of magnitude systematically and randomly using Latin Hypercube Sampling (LHS). k_{on} and k_{off} varied from 1e-1 to 1e2. β_0 was varied from 1e-6 to 1e-2. β_1 was varied from 1e-4 to 1e1. Parameters $\beta_2 = 0.06$, $\gamma_1 = 0.06$, and $\gamma_2 = 0.0083$ were fixed to values according to literature^{16,61}. From the parameter sets sampled, the slope ratio (defined in Figure 2.3), a summary metric for the degree of efficiency in response to short pulses, was calculated for each parameter set. The parameter search yielded model outputs that qualitatively recapitulate the pYPS1-YFP and pCMK2-YFP slope ratios.

Model Fitting - The model was used to fit the experimental data. Fits of the experimental data to the simple two-promoter state model (Figure 2.4) were obtained by the following procedure: 1. 10000 parameters were randomly sampled using LHS with the aforementioned parameter ranges (Figure 2.11A) or more comprehensively sampled using 33,000 parameter sets varying k_{on} , k_{off} , and γ_1 , (k_{on} was varied from 0.001-10, k_{off} from 0.000007-100, and γ_1 from 0.01-10, Figure 2.5). 2. Fits were determined to be model outputs that fit through ten or more data points within error of the dynamic protein time course data (Figure 2.5B) or seven or more data points within error of the Output-Fluorescence data (Figure 2.5D). Note that the ability of the model to fit the data

was the same regardless of the criteria of fit used – whether the criteria was the model output fit within the error bars of the data or least squared error of the model to the best fit line to the data. 3. Cross-validation of parameter fits to the dynamic protein time course for light input of 2 minute ON/4 minutes OFF was done. From this procedure, we identified parameter sets that recapitulated all of the experimental data for both pYPS1-YFP and pCMK2-YFP.

2.7.1 Mathematical Analysis

Derivation of expressions of total transcripts from Eq(1.1)

In this section, we derive expressions for the total mRNA produced by the promoter. Using these expressions, we then study properties of total transcripts produced from pulsed and continuous inputs. Starting with the mRNA equation

$$\frac{dmRNA}{dt} = \beta_1 p_{on} - \gamma_1 mRNA \quad (2.7)$$

where there is an instantaneous transcript creation rate $\beta_1 p_{on}$, with β_1 being the maximum transcription rate. The instantaneous mRNA loss rate is $\gamma_1 mRNA$. We will calculate the total integrated output of the promoter, that is the total number of transcripts produced over the entire experiment. By comparing this output for pulsed and continuous inputs that have the same Nuclear Fluorescence AUC, we will be able to link the promoter properties to the ability of producing more transcripts from either pulsed or continuous inputs.

The expression for the total number of transcripts, N , produced over the entire experiment for a given finite input function $TF(t)$ is described by the equation

$$N = \int_0^5 \beta_1 p_{on} dt \quad (2.8)$$

In the particular setting of our experiments, light input of constant amplitude is administered either continuously (with TF input denoted by $TF_c(t)$, and maximum TF input reaching TF_{max}) or pulsed (with TF input denoted by $TF_p(t)$) for a period of time (a maximum of 4 hours), after which the light is shut-off and the output is measured at 5 hours. For the pulsed experiment, light is given in pulses, which yields a TF input of approximately triangular pulses. Each triangular pulse reaches a maximum amplitude of TF_{max} after which the light is shut-off. By design, for a given Nuclear Fluorescence AUC value, the integrals of $TF_c(t)$ and $TF_p(t)$ are the same, i.e. $\int_0^5 TF_p(t)dt = \int_0^5 TF_c(t)dt$. Dynamic protein measurements (Figure 2.5B) indicate that the combined timescales of promoter and mRNA decay are faster than 25 minutes (for pYPS1-YFP and pCMK2-YFP). Therefore, p_{on} and mRNA levels will have decayed to zero at the time of measurement. In addition, prior to any inputs at $t = 0$, we assume that p_{on} is zero. As a result, $p_{on}(5) = 0$ and $p_{on}(0) = 0$.

Eq. (1.1) indicates that:

$$\frac{dp_{on}}{dt} = k_{on}(1 - p_{on})TF(t) - k_{off}p_{on} \quad (2.9)$$

Integrating this equation from $t = 0$ to $t = 5$ hours:

$$\int_0^5 \frac{dp_{on}}{dt} dt = \int_0^5 k_{on}(1 - p_{on})TF(t)dt - \int_0^5 k_{off}p_{on}dt \quad (2.10)$$

Now, $\int_0^5 \frac{dp_{on}}{dt} dt = p_{on}(5) - p_{on}(0) = 0$. Thus, setting the two terms on the right side equal to each other and multiplying by β_1/k_{off} we get

$$N = \int_0^5 \beta_1 p_{on} dt = \int_0^5 \frac{\beta_1}{k_d} (1 - p_{on})TF(t)dt \quad (2.11)$$

where $k_d = k_{off}/k_{on}$. The left two expressions are Eq. (??) while the rightmost expression describes how p_{on} negatively impacts N , for a given input $TF(t)$, through the term $(1 - p_{on})$.

To assess the value of N , we use the rightmost expression of Eq. (2.11). We further denote the total number of transcripts generated by $TF_p(t)$ to be N_p and that generated by $TF_c(t)$ to be N_c . Before we delve into the thorough mathematical treatment, we make two arguments. The first is an intuitive illustration of why the ratio N_p/N_c is larger than 1 for the two-state promoter model and the second with respect to how the ratio of total transcripts can be extrapolated to the analysis of the slope ratio we measure, which is that of the protein.

Intuitive Explanation of $N_p/N_c > 1$ for two-state promoter model

Irrespective of the input, p_{on} starts at zero and cannot exceed the steady-state solution of p_{on} from Eq. (2.1) evaluated at TF_{max} . This steady-state for p_{on} is given by $TF_{max}/(TF_{max} + k_d)$. For a non-zero $TF_c(t)$, after the initial rise, $TF_c(t)$ is primarily equal to TF_{max} (Figure 2.4G (blue shading)) and p_{on} is primarily equal to $TF_{max}/(TF_{max} + k_d)$ (Figure 2.4G (blue curve)). This is of course assuming that the promoter dynamics allow for reaching steady-state well before the input shuts off. On the other hand, for the pulsed input, the transcription factor input will spend more of its time in the state where $TF_p(t) < TF_{max}$ (Figure 2.4G (red shading)) and therefore, p_{on} will satisfy $p_{on} < TF_{max}/(TF_{max} + k_d)$ (Figure 2.4G (red curve)) for more of the time course. Accordingly, the $(1 - p_{on})$ term on the right hand side of Eq. (2.11) will predominantly be greater for the pulsed input than the continuous input when $TF(t)$ is non-zero. Taken together with the fact that $\int TF_p(t)dt = \int TF_c(t)dt$, then, according to Eq. (2.11), one might postulate that N_p is larger than N_c . We also prove this to be mathematically true in 2.7.1 Sections "Derivation of $N_p/N_c > 1$ in the regime of a fast promoter relative to $TF(t)$ "

(fast promoter) and "Derivation of $N_p/N_c > 1$ for general values of k_{on} and k_{off} " (general case). A limiting scenario occurs when $TF_{max} \ll k_d$, where $TF_{max}/(TF_{max} + k_d)$, the upper bound of p_{on} , remains much smaller than one (heat map of N_p/N_c in Figure 2.14A). In this case, the right hand side of Eq. (2.11) dictates that N_p/N_c will be close to one independent of the magnitudes of the rate constants k_{on} and k_{off} (see Figure 2.14B, $k_d = 46$).

N_p/N_c can be used to understand slope ratio

We can use the ratio of N for the pulsed input to N for the continuous input, i.e. N_p/N_c , to infer the characteristics of the protein slope ratio, which is the quantity experimentally measured (compare Figures 2.4B and Figure 2.14A). This is not surprising since transcripts are causal to proteins. This correlation between N_p/N_c and protein slope ratio can also be gleaned from the plot of N as a function of nuclear fluorescence AUC by comparing Figure 2.14B with the plot of protein as a function of nuclear fluorescence AUC data (Figures 2.4E ($k_d = 2.3$) and Figure 2.11D ($k_d = 46$)). Furthermore, for all parameters of the model that fit pYPS1-YFP and pCMK2-YFP experimental data, we computed N_p/N_c and the slope ratio of the protein output to compare their relationships. The results in Figure 2.14A-B show that much like for proteins, pulsed inputs yield more transcripts per nuclear fluorescence AUC than continuous inputs. Furthermore N_p/N_c linearly correlates with slope ratio for both pYPS1-YFP and pCMK2-YFP (Figure 2.14C). Fitting both plots to a line shows that the value of N_p/N_c is generally smaller than slope ratio. Thus, protein and mRNA degradation increase the slope ratio for the protein output.

Derivation of $N_p/N_c > 1$ in the regime of a fast promoter relative to $TF(t)$

In the case that the promoter dynamics are much faster than those of the transcription factor nuclear dynamics, p_{on} reaches pseudo-steady state on the timescale of $TF(t)$ such that $p_{on} \approx \frac{TF(t)}{TF(t)+k_d}$. Using this equation, we can write the middle expression in Eq. (2.11) as

$$\beta_1 \int_0^5 p_{on} dt = \beta_1 \int_0^5 \frac{TF(t)}{TF(t) + k_d} dt \quad (2.12)$$

And the right side of Eq. (2.11) is

$$\begin{aligned} \frac{\beta_1}{k_d} \int_0^5 (1 - p_{on}) TF(t) dt &= \frac{\beta_1}{k_d} \int_0^5 \left(1 - \frac{TF(t)}{TF(t) + k_d} \right) TF(t) dt \\ &= \frac{\beta_1}{k_d} \int_0^5 \left(\frac{TF(t) + k_d}{TF(t) + k_d} - \frac{TF(t)}{TF(t) + k_d} \right) TF(t) dt \\ &= \frac{\beta_1}{k_d} \int_0^5 \frac{k_d}{TF(t) + k_d} TF(t) dt \\ &= \beta_1 \int_0^5 \frac{TF(t)}{TF(t) + k_d} dt \end{aligned} \quad (2.13)$$

Accordingly, Eq. (2.12) is identical to Eq. (2.13) as is required by Eq. (2.11).

To begin, for the continuous TF input, the initial rise of $TF_c(t)$ ends at $t = \tau_{r,c}$ ('r' corresponds to rise, and 'c' corresponds to continuous). After this rise, $TF_c(t) = TF_{max}$ until the input starts shutting off at $t = \tau_{f,c}$ (Figure 2.14D) (where 'f' corresponds to fall).

We will divide N_c into three parts as follows

$$\begin{aligned} N_c &= \beta_1 \int_0^{\tau_{r,c}} \frac{TF_c(t)}{TF_c(t) + k_d} dt + \beta_1 \int_{\tau_{r,c}}^{\tau_{f,c}} \frac{TF_c(t)}{TF_c(t) + k_d} dt + \beta_1 \int_{\tau_{f,c}}^5 \frac{TF_c(t)}{TF_c(t) + k_d} dt \\ &= N_r + N_{c,m} + N_f \end{aligned} \quad (2.14)$$

where N_r is the transcriptional contribution due to the rise of the TF pulse, N_f is the

transcriptional contribution during the fall of this pulse, and $N_{c,m}$ is the contribution in between where TF assumes its maximum value TF_{max} (where 'm' corresponds to middle).

The maximum amplitude for the pulsed input $TF_p(t)$ is also TF_{max} . The initial rise of $TF_p(t)$ is the same as that of $TF_c(t)$, and the final decay of the continuous and pulsed inputs are also the same (Figure 2.14D). If $\tau_{r,p}$ is the initial rise time of $TF_p(t)$ and $\tau_{f,p}$ is the time at which the final pulse starts decaying, then $\int_0^{\tau_{r,c}} TF_c(t)dt = \int_0^{\tau_{r,p}} TF_p(t)dt$ and $\int_{\tau_{f,c}}^5 TF_c(t)dt = \int_{\tau_{f,p}}^5 TF_p(t)dt$.

We again divide N_p into three parts:

$$\begin{aligned} N_p &= \beta_1 \int_0^{\tau_{r,p}} \frac{TF_p(t)}{TF_p(t) + k_d} dt + \beta_1 \int_{\tau_{r,p}}^{\tau_{f,p}} \frac{TF_p(t)}{TF_p(t) + k_d} dt + \beta_1 \int_{\tau_{f,p}}^5 \frac{TF_p(t)}{TF_p(t) + k_d} dt \\ &= N_r + N_{p,m} + N_f \end{aligned} \quad (2.15)$$

Where $N_{p,m}$ is the contribution of the series of pulses occurring between $\tau_{r,p}$ and $\tau_{f,p}$.

Therefore, in order to show that $N_p/N_c > 1$, we just have to show that $N_{p,m}/N_{c,m} > 1$.

To begin, it is important to note that

$$\int_{\tau_{r,c}}^{\tau_{f,c}} TF_c(t)dt = \int_{\tau_{r,p}}^{\tau_{f,p}} TF_p(t)dt \quad (2.16)$$

For the continuous input, $TF_c(t) = TF_{max}$ between $\tau_{r,c} \leq t \leq \tau_{f,c}$. Accordingly,

$$\begin{aligned} N_{c,m} &= \beta_1 \int_{\tau_{r,c}}^{\tau_{f,c}} \frac{TF_c(t)}{TF_c(t) + k_d} dt \\ &= \beta_1 \int_{\tau_{r,c}}^{\tau_{f,c}} \frac{TF_{max}}{TF_{max} + k_d} dt \\ &= \beta_1 \frac{TF_{max}[\tau_{f,c} - \tau_{r,c}]}{TF_{max} + k_d} \end{aligned} \quad (2.17)$$

In the steps above we left $TF_c(t)$ in the numerator of the integrand term without replacing

it with TF_{max} in the middle step from Eq. (2.17) in order to use the equality in Eq. (2.16) such that $N_{c,m}$ becomes:

$$\begin{aligned} N_{c,m} &= \beta_1 \int_{\tau_{r,c}}^{\tau_{f,c}} \frac{TF_c(t)}{TF_{max} + k_d} dt \\ &= \beta_1 \int_{\tau_{r,p}}^{\tau_{f,p}} \frac{TF_p(t)}{TF_{max} + k_d} dt \end{aligned} \quad (2.18)$$

On the other hand, the expression for $N_{p,m}$ is given by:

$$N_{p,m} = \beta_1 \int_{\tau_{r,p}}^{\tau_{f,p}} \frac{TF_p(t)}{TF_p(t) + k_d} dt \quad (2.19)$$

Since $TF_p(t) + k_d \leq TF_{max} + k_d$ for every value of t , then $\frac{TF_p(t)}{TF_p(t) + k_d} \geq \frac{TF_p(t)}{TF_{max} + k_d}$ for every value of t . In particular, since $TF_p(t)$ is pulsing, then $TF_p(t) + k_d < TF_{max} + k_d$ for most of $\tau_{r,p} \leq t \leq \tau_{f,p}$. We can therefore write down the inequality

$$N_{p,m} = \beta_1 \int_{\tau_{r,p}}^{\tau_{f,p}} \frac{TF_p(t)}{TF_p(t) + k_d} dt > \beta_1 \int_{\tau_{r,p}}^{\tau_{f,p}} \frac{TF_p(t)}{TF_{max} + k_d} dt = N_{c,m} \quad (2.20)$$

As a result, $\frac{N_{p,m}}{N_{c,m}} > 1$ and thus $\frac{N_p}{N_c} > 1$. It should be noted that an important requirement for Eq. (2.20) and all following expressions is that the pulses in $TF_p(t)$ transition from zero to TF_{max} at a finite speed (e.g, less than infinitely fast). This allows finite integral contributions to the left side of Eq. (2.20) during this transition when $0 < TF_p(t) < TF_{max}$. This ensures the inequality. However for rectangular pulses with no transition time, i.e $TF_p(t)$ jumps infinitely fast from zero to TF_{max} , the left side of Eq. (2.20) will equal the right side and thus $N_p = N_c$, for this special case.

The ratio $\frac{N_p}{N_c}$ is a monotonically decreasing function of k_d and approaches 1 as k_d becomes very large compared to TF_{max}

One can see on the left side of Eq.(2.20) that the $\frac{TF_p(t)}{TF_p(t)+k_d}$ term becomes larger in magnitude as k_d decreases for all t when $TF_p(t) > 0$. This implies that $N_{p,m}$ increases for decreasing k_d for a given $TF_p(t)$. Likewise, $N_{c,m}$ increases for decreasing k_d . We will next show however that $\frac{N_p}{N_c}$ itself decreases and approaches the value of 1 as k_d increases and becomes large relative to TF_{max} . Under this constraint Eq (2.20) becomes

$$N_{p,m} = \beta_1 \int_{\tau_{r,p}}^{\tau_{f,p}} \frac{TF_p(t)}{TF_p(t) + k_d} dt \approx \beta_1 \int_{\tau_{r,p}}^{\tau_{f,p}} \frac{TF_p(t)}{k_d} dt \approx \beta_1 \int_{\tau_{r,p}}^{\tau_{f,p}} \frac{TF_p(t)}{TF_{max} + k_d} dt = N_{c,m} \quad (2.21)$$

Therefore $N_{p,m} \approx N_{c,m}$ and hence $N_p \approx N_c$

As mentioned above, the contribution to N_c from the initial rise and final fall of $TF_c(t)$ corresponds to one full pulse of a pulsed input $TF_p(t)$. Therefore, since

$N_p = \beta_1 \int_0^5 \frac{TF_p(t)}{TF_p(t)+k_d} dt$, then if $TF_p(t)$ consists of M pulses, then one pulse would be the equivalent of $\frac{1}{M}\beta_1 \int_0^5 \frac{TF_p(t)}{TF_p(t)+k_d} dt$. Furthermore, as also explained above: $N_{r,c} = N_{r,p}$ and $N_{f,c} = N_{f,p}$ and we will refer to both as N_r and N_f , respectively. We will therefore represent N_c as:

$$\begin{aligned} N_c &= N_r + N_{c,m} + N_f \\ &= \frac{1}{M}\beta_1 \int_0^5 \frac{TF_p(t)}{TF_p(t) + k_d} dt + \beta_1 \int_{\tau_{r,c}}^{\tau_{f,c}} \frac{TF_c(t)}{TF_{max} + k_d} dt \end{aligned} \quad (2.22)$$

where we have used Eq (2.17). Similarly, we represent N_p as

$$\begin{aligned} N_p &= N_r + N_{p,m} + N_f \\ &= \frac{1}{M}\beta_1 \int_0^5 \frac{TF_p(t)}{TF_p(t) + k_d} dt + \beta_1 \int_{\tau_{r,p}}^{\tau_{f,p}} \frac{TF_p(t)}{TF_p(t) + k_d} dt \\ &= \frac{1}{M}\beta_1 \int_0^5 \frac{TF_p(t)}{TF_p(t) + k_d} dt + \frac{M-1}{M}\beta_1 \int_0^5 \frac{TF_p(t)}{TF_p(t) + k_d} dt \end{aligned} \quad (2.23)$$

The derivative of N_c with respect to k_d would be

$$\begin{aligned}\frac{dN_c}{dk_d} &= \frac{dN_r}{dk_d} + \frac{dN_{c,m}}{dk_d} + \frac{dN_f}{dk_d} \\ &= -\frac{1}{M}\beta_1 \int_0^5 \frac{TF_p(t)}{(TF_p(t) + k_d)^2} dt - \beta_1 \int_{\tau_{r,c}}^{\tau_{f,c}} \frac{TF_c(t)}{(TF_{max} + k_d)^2} dt\end{aligned}\quad (2.24)$$

Therefore,

$$\frac{dN_r}{dk_d} + \frac{dN_f}{dk_d} = -\frac{1}{M}\beta_1 \int_0^5 \frac{TF_p(t)}{(TF_p(t) + k_d)^2} dt \quad (2.25)$$

and

$$\begin{aligned}\frac{dN_{c,m}}{dk_d} &= -\beta_1 \int_{\tau_{r,c}}^{\tau_{f,c}} \frac{TF_c(t)}{(TF_{max} + k_d)^2} dt \\ &= -\frac{N_{c,m}}{(TF_{max} + k_d)}\end{aligned}\quad (2.26)$$

Likewise, the derivative of N_p with respect to k_d would be

$$\begin{aligned}\frac{dN_p}{dk_d} &= \frac{dN_r}{dk_d} + \frac{dN_{p,m}}{dk_d} + \frac{dN_f}{dk_d} \\ &= -\frac{1}{M}\beta_1 \int_0^5 \frac{TF_p(t)}{(TF_p(t) + k_d)^2} dt - \beta_1 \int_{\tau_{r,p}}^{\tau_{f,p}} \frac{TF_p(t)}{(TF_p(t) + k_d)^2} dt \\ &= -\frac{1}{M}\beta_1 \int_0^5 \frac{TF_p(t)}{(TF_p(t) + k_d)^2} dt - \frac{M-1}{M}\beta_1 \int_0^5 \frac{TF_p(t)}{(TF_p(t) + k_d)^2} dt\end{aligned}\quad (2.27)$$

where

$$\begin{aligned}\frac{dN_{p,m}}{dk_d} &= -\beta_1 \int_{\tau_{r,p}}^{\tau_{f,p}} \frac{TF_p(t)}{(TF_p(t) + k_d)^2} dt \\ &= -\frac{M-1}{M}\beta_1 \int_0^5 \frac{TF_p(t)}{(TF_p(t) + k_d)^2} dt\end{aligned}\quad (2.28)$$

We will next compute the derivative of N_p/N_c with respect to k_d , and use the expressions above to assess its sign, demonstrating that it is negative. The derivative of N_p/N_c with respect to k_d is given by:

$$\begin{aligned}
\frac{d\frac{N_p}{N_c}}{dk_d} &= \frac{\frac{dN_p}{dk_d} N_c}{N_c^2} - \frac{\frac{dN_c}{dk_d} N_p}{N_c^2} \\
&= \frac{1}{N_c^2} \left[\left(\frac{dN_{p,m}}{dk_d} + \frac{d[N_r + N_f]}{dk_d} \right) \left(N_{c,m} + [N_r + N_f] \right) \right. \\
&\quad \left. - \left(\frac{dN_{c,m}}{dk_d} + \frac{d[N_r + N_f]}{dk_d} \right) \left(N_p + [N_r + N_f] \right) \right] \\
&= \frac{1}{N_c^2} \left[\frac{dN_{p,m}}{dk_d} N_{c,m} + \frac{d[N_r + N_f]}{dk_d} N_{c,m} + \frac{dN_{p,m}}{dk_d} [N_r + N_f] + \frac{d[N_r + N_f]}{dk_d} [N_r + N_f] \right. \\
&\quad \left. - \frac{dN_{c,m}}{dk_d} N_{p,m} - \frac{d[N_r + N_f]}{dk_d} N_{p,m} - \frac{dN_{c,m}}{dk_d} [N_r + N_f] - \frac{d[N_r + N_f]}{dk_d} [N_r + N_f] \right] \\
&= \frac{1}{N_c^2} \left[\frac{dN_{p,m}}{dk_d} N_{c,m} + \frac{d[N_r + N_f]}{dk_d} N_{c,m} + \frac{dN_{p,m}}{dk_d} [N_r + N_f] \right. \\
&\quad \left. - \frac{dN_{c,m}}{dk_d} N_{p,m} - \frac{d[N_r + N_f]}{dk_d} N_{p,m} - \frac{dN_{c,m}}{dk_d} [N_r + N_f] \right] \tag{2.29}
\end{aligned}$$

We will now analyze different pairs of terms from the right hand side of Eq. (2.29). We will

start with

$$\begin{aligned}
\frac{dN_{p,m}}{dk_d} N_{c,m} - \frac{dN_{c,m}}{dk_d} N_{p,m} &= -N_{c,m} \beta_1 \int_{\tau_{r,p}}^{\tau_{f,p}} \frac{TF_p(t)}{(TF_p(t) + k_d)^2} dt \\
&\quad + N_{p,m} \beta_1 \int_{\tau_{r,c}}^{\tau_{f,c}} \frac{TF_c(t)}{(TF_{max} + k_d)^2} dt \\
&= -N_{c,m} \beta_1 \int_{\tau_{r,p}}^{\tau_{f,p}} \frac{TF_p(t)}{(TF_p(t) + k_d)^2} dt \\
&\quad + \frac{N_{p,m}}{TF_{max} + k_d} \beta_1 \int_{\tau_{r,c}}^{\tau_{f,c}} \frac{TF_c(t)}{TF_{max} + k_d} dt \\
&= -N_{c,m} \beta_1 \int_{\tau_{r,p}}^{\tau_{f,p}} \frac{TF_p(t)}{(TF_p(t) + k_d)^2} dt + \frac{N_{p,m}}{TF_{max} + k_d} N_{c,m} \\
&= -N_{c,m} \beta_1 \int_{\tau_{r,p}}^{\tau_{f,p}} \frac{TF_p(t)}{(TF_p(t) + k_d)^2} dt + N_{c,m} \frac{N_{p,m}}{TF_{max} + k_d} \\
&= -N_{c,m} \beta_1 \int_{\tau_{r,p}}^{\tau_{f,p}} \frac{TF_p(t)}{(TF_p(t) + k_d)^2} dt \\
&\quad + N_{c,m} \beta_1 \int_{\tau_{r,p}}^{\tau_{f,p}} \frac{TF_p(t)}{(TF_p(t) + k_d)(TF_{max} + k_d)} dt \tag{2.30}
\end{aligned}$$

where in the third line we have used the relationship $N_{c,m} = \beta_1 \int_{\tau_{r,c}}^{\tau_{f,c}} \frac{TF_c(t)}{TF_{max} + k_d} dt$ from Eq. (2.17). In the last line we used the relationship $N_{p,m} = \beta_1 \int_{\tau_{r,p}}^{\tau_{f,p}} \frac{TF_p(t)}{TF_p(t) + k_d} dt$ from Eq. (2.23). When $TF_p(t)$ is non-zero, $\frac{TF_p(t)}{(TF_p(t) + k_d)^2} > \frac{TF_p(t)}{(TF_p(t) + k_d)(TF_{max} + k_d)}$ for all t except when $TF_p(t) = TF_{max}$ which occurs only at the peak of the pulse. Therefore,

$\frac{dN_{p,m}}{dk_d} N_{c,m} - \frac{dN_{c,m}}{dk_d} N_{p,m} < 0$. Next, the second pair of terms is

$$\begin{aligned}
\frac{d[N_r + N_f]}{dk_d} N_{c,m} - \frac{d[N_r + N_f]}{dk_d} N_{p,m} &= [-N_{c,m} + N_{p,m}] \frac{1}{M} \beta_1 \cdot \\
&\quad \int_0^5 \frac{TF_p(t)}{(TF_p(t) + k_d)^2} dt \tag{2.31}
\end{aligned}$$

Where we have used the expression for $\frac{dN_r+dN_f}{dk_d}$ from above. Finally, the third pair is

$$\begin{aligned}
\frac{dN_{p,m}}{dk_d}[N_r + N_f] - \frac{dN_{c,m}}{dk_d}[N_r + N_f] &= [N_r + N_f] \left[-\beta_1 \int_{\tau_{r,p}}^{\tau_{f,p}} \frac{TF_p(t)}{(TF_p(t) + k_d)^2} dt \right. \\
&\quad \left. + \beta_1 \int_{\tau_{r,c}}^{\tau_{f,c}} \frac{TF_c(t)}{(TF_{max} + k_d)^2} dt \right] \\
&= [N_r + N_f] \left[-\beta_1 \int_{\tau_{r,p}}^{\tau_{f,p}} \frac{TF_p(t)}{(TF_p(t) + k_d)^2} dt \right. \\
&\quad \left. + \frac{N_{c,m}}{(TF_{max} + k_d)} \right] \\
&= \frac{1}{M} N_p \left[-\beta_1 \int_{\tau_{r,p}}^{\tau_{f,p}} \frac{TF_p(t)}{(TF_p(t) + k_d)^2} dt \right. \\
&\quad \left. + \frac{N_{c,m}}{(TF_{max} + k_d)} \right] \\
&= \frac{1}{M} N_p \left[-\frac{M-1}{M} \beta_1 \int_0^5 \frac{TF_p(t)}{(TF_p(t) + k_d)^2} dt \right. \\
&\quad \left. + \frac{N_{c,m}}{(TF_{max} + k_d)} \right] \\
&= -\frac{M-1}{M} N_p \frac{1}{M} \beta_1 \int_0^5 \frac{TF_p(t)}{(TF_p(t) + k_d)^2} dt \\
&\quad + \frac{1}{M} N_p \frac{N_{c,m}}{(TF_{max} + k_d)} \\
&= -N_{p,m} \frac{1}{M} \beta_1 \int_0^5 \frac{TF_p(t)}{(TF_p(t) + k_d)^2} dt + \frac{1}{M} N_p \frac{N_{c,m}}{(TF_{max} + k_d)} \\
&= -N_{p,m} \frac{1}{M} \beta_1 \int_0^5 \frac{TF_p(t)}{(TF_p(t) + k_d)^2} dt + N_{c,m} \frac{1}{M} \frac{N_p}{(TF_{max} + k_d)} \\
&= -N_{p,m} \frac{1}{M} \beta_1 \int_0^5 \frac{TF_p(t)}{(TF_p(t) + k_d)^2} dt \\
&\quad + N_{c,m} \frac{1}{M} \beta_1 \int_0^5 \frac{TF_p(t)}{(TF_p(t) + k_d)(TF_{max} + k_d)} dt \quad (2.32)
\end{aligned}$$

where in the second line we have used the relationship $\beta_1 \int_{\tau_{r,c}}^{\tau_{f,c}} \frac{TF_c(t)}{(TF_{max} + k_d)^2} dt = \frac{N_{c,m}}{(TF_{max} + k_d)}$

from Eq.(2.26). In the fourth line we have used the relationship

$-\beta_1 \int_{\tau_{r,p}}^{\tau_{f,p}} \frac{TF_p(t)}{(TF_p(t) + k_d)^2} dt = -\frac{M-1}{M} \beta_1 \int_0^5 \frac{TF_p(t)}{(TF_p(t) + k_d)^2} dt$ from Eq. (2.28). In the last line we used

the relationship $N_p = \beta_1 \int_0^5 \frac{TF_p(t)}{TF_p(t)+k_d} dt$. We can now add Eq (2.31) and Eq (2.32) to get

$$\begin{aligned}
& \frac{d[N_r + N_f]}{dk_d} N_{c,m} - \frac{d[N_r + N_f]}{dk_d} N_{p,m} + \frac{dN_{p,m}}{dk_d} [N_r + N_f] - \frac{dN_{c,m}}{dk_d} [N_r + N_f] \\
& = [-N_{c,m} + N_{p,m}] \frac{1}{M} \beta_1 \int_0^5 \frac{TF_p(t)}{(TF_p(t) + k_d)^2} dt \\
& - N_{p,m} \frac{1}{M} \beta_1 \int_0^5 \frac{TF_p(t)}{(TF_p(t) + k_d)^2} dt + N_{c,m} \frac{1}{M} \beta_1 \int_0^5 \frac{TF_p(t)}{(TF_p(t) + k_d)(TF_{max} + k_d)} dt \\
& = -N_{c,m} \frac{1}{M} \beta_1 \int_0^5 \frac{TF_p(t)}{(TF_p(t) + k_d)^2} dt + N_{c,m} \frac{1}{M} \beta_1 \int_0^5 \frac{TF_p(t)}{(TF_p(t) + k_d)(TF_{max} + k_d)} dt \\
& < 0 \quad (2.33)
\end{aligned}$$

since $\frac{TF_p(t)}{(TF_p(t)+k_d)^2} > \frac{TF_p(t)}{(TF_p(t)+k_d)(TF_{max}+k_d)}$ for all t and when $TF_p(t)$ is non-zero, except when $TF_p(t) = TF_{max}$ which occurs only at the peak of the pulse. Thus, from the results in Eq. (2.30) and Eq. (2.33), we can conclude that $\frac{dN_p}{dk_d} < 0$, Therefore, N_p/N_c is a monotonically decreasing function of k_d as is demonstrated in Figure 2.4 and Figure 2.10 in the main text.

Derivation of $N_p/N_c > 1$ for general values of k_{on} and k_{off}

Below we derive the result that N_p , the total transcripts for the pulsed input $TF_p(t)$, is always greater than N_c , the total transcripts for the continuous input $TF_c(t)$. The first section below (2.7.1 Section "Treatment of a class of inputs whose decreasing edge proceeds through steps") introduces a class of inputs and subsequent analysis and simulation results that form the foundation of this proof. This section continues at a high level to describe the construction of more complex classes of inputs, related analysis, and simulation results. The extremes of these classes represent the continuous and pulsed inputs. We leave the detailed mathematical derivations, analysis, and technical formulation of the proof for later sections, but each derivation is referenced in the appropriate area in 2.7.1 Section "Treatment of a class of inputs whose decreasing edge proceeds through steps". This serves

as a roadmap to understand the complete proof for those interested. Otherwise, 2.7.1 Section "Treatment of a class of inputs whose decreasing edge proceeds through steps" by itself serves as a high level intuitive summary of the proof that is aimed to be accessible to a more general audience.

Treatment of a class of inputs whose decreasing edge proceeds through steps

We first consider an ensemble of inputs $TF(t)$ that will become instrumental in the general demonstration, and derive properties generated by the two-state promoter model when stimulated with these inputs. The first class of inputs is shown in the top panel of Figure 2.15A. After an initial rise of the pulse, all the inputs decrease by a step except one which remains constant (Figure 2.15A, top plot, blue), reminiscent of the continuous input. The other inputs then remain constant for a duration and then decrease by a step except for one which remains constant (Figure 2.15A, top plot, red). This behavior repeats itself until the last remaining input step decreases to zero and remains at zero (Figure 2.15A, top plot, light blue), representing a single pulse. This produces a series of inputs, where each plateaus (indefinitely) at a different value. Simulation results of the model with this input ensemble show that each input's corresponding p_{on} trajectory maintains the same relative order to the other inputs' p_{on} trajectories as the inputs do with each other (Figure 2.15A, second panel from the top, compare for example light blue and green traces). This is a general property of the system regardless of the values of k_{on} and k_{off} . We provide a proof in 2.7.1 Section "Proof showing that $p_{on}(t, \alpha_1) > p_{on}(t, \alpha_2)$ for $t > t_0$ and $\alpha_1 > \alpha_2$ " using the analytical results derived in 2.7.1 Section "Derivation of an expression for $p_{on}(t)$ in response to a step input starting from an initial condition $p_{on}(t_0)$ ".

Then we plot p_{on} versus accumulated TF area (defined as $\int_0^t TF(v)dv$) for each input. This plot (Figure 2.15A, second panel from the bottom) shows that the relative order is the same as it was for p_{on} versus accumulated time. We derive these results in 2.7.1 Section

”Proof showing that $p_{on}(t_0 + \sigma, \alpha_1) > p_{on}(t_0 + \frac{\alpha_1}{\alpha_2}\sigma, \alpha_2)$ for $\sigma > 0$, where σ is proportional to accumulated TF area”. These results are then used to prove, as we observe in our simulations, that for accumulated transcripts versus accumulated TF area, the relative order is reversed from the first three panels (Figure 2.15A, bottom plot). Intuitively this makes sense since total transcripts from the right hand side of Eq. (2.11) is a function of $1 - p_{on}$ (multiplied by $TF(t)$) and not p_{on} . **Overall, these results show that the lower the plateau value is for this class of inputs, the higher the accumulated transcripts as a function of accumulated area (see 2.7.1 Section ”Deriving relationships for the number of transcripts, specifically $\Delta N_{\alpha_2}(\sigma) > \Delta N_{\alpha_1}(\sigma)$ for $\sigma > 0$ ” for derivation).**

Towards the construction of multiple pulses, in 2.7.1 Section ”Analysis of a class of inputs that relate the number of transcripts of the continuous input $TF_c(t)$ to that of pulsed input $TF_p(t)$ for the two pulse case” we extend our analysis to a class of inputs that initially follow the same trajectories as Figure 2.15A (top plot), where each plateaus at a different value, but then rise again to TF_{max} before shutting off (Figure 2.15B, top plot). Here the input with the highest valued plateau represents the continuous input $TF_c(t)$ while the input with the lowest valued plateau represents the pulsed input $TF_p(t)$ (two pulses).

Importantly, all inputs share the same total TF area. Through a similar analysis as above, we show that the lower the plateau value of an input, the more total transcripts that are produced (Figure 2.15B, inputs (top panel), transcripts (bottom panel)). Finally, in 2.7.1 Section ”Extending the approach to a higher number of pulses” we extend this approach first to three pulses and then to arbitrary numbers of pulses to demonstrate the generality of the two pulse result (examples in Figure 2.15C, inputs (top panel), accumulated transcripts (bottom panel), and Figure 2.15E, inputs (left panel), accumulated transcripts (right panel)).

Finally, in our examples we used $I = 6$ inputs to visually demonstrate our approach. As I

gets larger, each step change is smaller and better represents a smooth input, but the conclusions of our proof do not change. Thus, we can take the limit of steps becoming infinitesimally small to represent smooth pulses with large I to prove that for smooth inputs, $N_p > N_c$.

Derivation of an expression for $p_{on}(t)$ in response to a step input starting from an initial condition $p_{on}(t_0)$

For the two-state promoter model that experiences a step input $TF(t) = \alpha_1 TF_{max}$, starting from an initial condition $p_{on}(t_0)$, the dynamic equations are given by:

$$\begin{aligned}
\frac{dp_{on}}{dt} &= k_{on}TF(t)(1 - p_{on}) - k_{off}p_{on} \\
&= k_{on}\alpha_1 TF_{max}(1 - p_{on}) - k_{off}p_{on} \\
&= -[k_{on}\alpha_1 TF_{max} + k_{off}]p_{on} + k_{on}\alpha_1 TF_{max}
\end{aligned} \tag{2.34}$$

We seek the time dependent solution of this equation, which we denote by $p_{on}(t, \alpha_1)$. This is a linear first order ordinary differential equation, whose solution takes the form $p_{on} = a_0 + a_1 \exp(-[k_{on}\alpha_1 TF_{max} + k_{off}][t - t_0])$ for $t \geq t_0$. Following standard procedure, we solve for the constants a_0 and a_1 by evaluating the system at $t = t_0$ and $t = \infty$. At $t = \infty$, $\frac{dp_{on}}{dt} = 0$ therefore dictating that $a_0 = \frac{k_{on}\alpha_1 TF_{max}}{k_{on}\alpha_1 TF_{max} + k_{off}}$. At $t = t_0$ the solution must equal the initial condition, that is $a_0 + a_1 = p_{on}(t_0)$. Thus, $a_1 = p_{on}(t_0) - \frac{k_{on}\alpha_1 TF_{max}}{k_{on}\alpha_1 TF_{max} + k_{off}}$

and the full solution becomes:

$$\begin{aligned}
p_{on}(t, \alpha_1) &= \frac{k_{on}\alpha_1 TF_{max}}{k_{on}\alpha_1 TF_{max} + k_{off}} \\
&+ \left[p_{on}(t_0) - \frac{k_{on}\alpha_1 TF_{max}}{k_{on}\alpha_1 TF_{max} + k_{off}} \right] \exp(-[k_{on}\alpha_1 TF_{max} + k_{off}][t - t_0]) \\
&= \frac{k_{on}\alpha_1 TF_{max}}{k_{on}\alpha_1 TF_{max} + k_{off}} \\
&+ \left[p_{on}(t_0) - \frac{k_{on}\alpha_1 TF_{max}}{k_{on}\alpha_1 TF_{max} + k_{off}} \right] \exp(-[k_{on}\alpha_1 TF_{max} + k_{off}][t - t_0]) \\
&+ p_{on}(t_0) - p_{on}(t_0) \\
&= p_{on}(t_0) + \left(\frac{k_{on}\alpha_1 TF_{max}}{k_{on}\alpha_1 TF_{max} + k_{off}} - p_{on}(t_0) \right) \times \\
&\quad \left(1 - \exp(-[k_{on}\alpha_1 TF_{max} + k_{off}][t - t_0]) \right) \tag{2.35}
\end{aligned}$$

Proof showing that $p_{on}(t, \alpha_1) > p_{on}(t, \alpha_2)$ for $t > t_0$ and $\alpha_1 > \alpha_2$

Simulation results of the model with the input ensemble in Figure 2.15A (top panel) show that each input's corresponding p_{on} time trajectory maintains the same relative order to the other inputs' p_{on} trajectories as the inputs do with each other (Figure 2.15A, second panel from the top, compare for example light blue and green traces). This is a general property of the system regardless of the values of k_{on} and k_{off} . We now provide a proof of this observation. First, for any pair of inputs with adjacent plateaus (Figure 2.15A (top plot), blue/red, red/orange, orange/purple, purple/green, green/light blue), the point in time where they diverge we'll denote as $t = t_0$. Likewise, since the pair have identical input behavior up until $t = t_0$, the value of $p_{on}(t)$ at $t = t_0$ will be the same for both. We'll denote this shared value as $p_{on}(t_0)$. For $t > t_0$, the first (upper) input has $TF(t) = \alpha_1 TF_{max}$ while the second (step down, lower) input has $TF(t) = \alpha_2 TF_{max}$. Accordingly, $\alpha_1 > \alpha_2$. We'll apply the analytical solution from 2.7.1 Section "Derivation of an expression for $p_{on}(t)$ in response to a step input starting from an initial condition $p_{on}(t_0)$ " to compare the solutions

of $p_{on}(t)$ for the upper and lower inputs, which we'll refer to as $p_{on}(t, \alpha_1)$ and $p_{on}(t, \alpha_2)$, respectively. We'll then prove that $p_{on}(t, \alpha_1) > p_{on}(t, \alpha_2)$ for $t > t_0$. We begin by subtracting the analytical solution of $p_{on}(t, \alpha_2)$ from that of $p_{on}(t, \alpha_1)$ to get

$$\begin{aligned}
p_{on}(t, \alpha_1) - p_{on}(t, \alpha_2) &= p_{on}(t_0) + \left(\frac{k_{on}\alpha_1 TF_{max}}{k_{on}\alpha_1 TF_{max} + k_{off}} - p_{on}(t_0) \right) \\
&\quad \times \left(1 - \exp(-[k_{on}\alpha_1 TF_{max} + k_{off}][t - t_0]) \right) \\
&\quad - p_{on}(t_0) - \left(\frac{k_{on}\alpha_2 TF_{max}}{k_{on}\alpha_2 TF_{max} + k_{off}} - p_{on}(t_0) \right) \\
&\quad \times \left(1 - \exp(-[k_{on}\alpha_2 TF_{max} + k_{off}][t - t_0]) \right) \quad (2.36) \\
&= \left(\frac{k_{on}\alpha_1 TF_{max}}{k_{on}\alpha_1 TF_{max} + k_{off}} - p_{on}(t_0) \right) \\
&\quad \times \left(1 - \exp(-[k_{on}\alpha_1 TF_{max} + k_{off}][t - t_0]) \right) \\
&\quad - \left(\frac{k_{on}\alpha_2 TF_{max}}{k_{on}\alpha_2 TF_{max} + k_{off}} - p_{on}(t_0) \right) \\
&\quad \times \left(1 - \exp(-[k_{on}\alpha_2 TF_{max} + k_{off}][t - t_0]) \right)
\end{aligned}$$

There are three cases to consider depending on the value of the initial condition $p_{on}(t_0)$.

1. $\frac{k_{on}\alpha_1 TF_{max}}{k_{on}\alpha_1 TF_{max} + k_{off}} - p_{on}(t_0) > \frac{k_{on}\alpha_2 TF_{max}}{k_{on}\alpha_2 TF_{max} + k_{off}} - p_{on}(t_0) > 0$: In this case, $p_{on}(t, \alpha_1) - p_{on}(t, \alpha_2) > 0$ for $t > t_0$. This is because $1 - \exp(-[k_{on}\alpha_1 TF_{max} + k_{off}][t - t_0]) > 1 - \exp(-[k_{on}\alpha_2 TF_{max} + k_{off}][t - t_0])$, since by definition $\alpha_1 > \alpha_2$.
2. $\frac{k_{on}\alpha_1 TF_{max}}{k_{on}\alpha_1 TF_{max} + k_{off}} - p_{on}(t_0) > 0$ and $\frac{k_{on}\alpha_2 TF_{max}}{k_{on}\alpha_2 TF_{max} + k_{off}} - p_{on}(t_0) < 0$: In this case, all terms are positive, thus, $p_{on}(t, \alpha_1) - p_{on}(t, \alpha_2) > 0$ for $t > t_0$.
3. $0 > \frac{k_{on}\alpha_1 TF_{max}}{k_{on}\alpha_1 TF_{max} + k_{off}} - p_{on}(t_0) > \frac{k_{on}\alpha_2 TF_{max}}{k_{on}\alpha_2 TF_{max} + k_{off}} - p_{on}(t_0)$. For this case, we will present a more detailed analysis.

To simplify notation, We will rewrite Eq. (2.36) as:

$$p_{on}(t, \alpha_1) - p_{on}(t, \alpha_2) = b_2 \left(1 - \exp(-c_2[t - t_0]) \right) - b_1 \left(1 - \exp(-c_1[t - t_0]) \right) \quad (2.37)$$

where $b_1 = p_{on}(t_0) - \frac{k_{on}\alpha_1 TF_{max}}{k_{on}\alpha_1 TF_{max} + k_{off}}$ and $b_2 = p_{on}(t_0) - \frac{k_{on}\alpha_2 TF_{max}}{k_{on}\alpha_2 TF_{max} + k_{off}}$, with $b_2 > b_1 > 0$.

Also, $c_1 = [k_{on}\alpha_1 TF_{max} + k_{off}]$ and $c_2 = [k_{on}\alpha_2 TF_{max} + k_{off}]$, with $c_1 > c_2 > 0$.

Since the initial conditions for both inputs is the same, the initial condition of their difference in Eq. (2.37) is zero at $t = t_0$. We therefore need to show that the slope of Eq. (2.37) evaluated for $t \geq t_0$ is positive.

Taking the derivative of Eq. (2.37) we get:

$$\frac{d[p_{on}(t, \alpha_1) - p_{on}(t, \alpha_2)]}{dt} = b_2 c_2 \exp(-c_2[t - t_0]) - b_1 c_1 \exp(-c_1[t - t_0]) \quad (2.38)$$

We evaluate this expression at $t = t_0$ to get:

$$\left. \frac{d[p_{on}(t, \alpha_1) - p_{on}(t, \alpha_2)]}{dt} \right|_{t=t_0} = b_2 c_2 - b_1 c_1 \quad (2.39)$$

We now use the expressions for b_1, c_1, b_2, c_2 in the above equation, to get that:

$$\left. \frac{d[p_{on}(t, \alpha_1) - p_{on}(t, \alpha_2)]}{dt} \right|_{t=t_0} = k_{on} TF_{max} (1 - p_{on}(t_0)) (\alpha_1 - \alpha_2) \quad (2.40)$$

Which is positive since $\alpha_1 > \alpha_2$. Furthermore, $\exp(-c_2 t)$ decays slower than $\exp(-c_1 t)$, and thus $\exp(-c_2 t) > \exp(-c_1 t)$ for all t . Therefore, it follows that $b_2 c_2 > b_1 c_1$, and that $b_2 c_2 \exp(-c_2 t) > b_1 c_1 \exp(-c_1 t)$. This concludes our proof that $p_{on}(t, \alpha_1) - p_{on}(t, \alpha_2) > 0$ for $t > t_0$. As discussed above, in Figure 2.15A (top plot), this result applies to any pair of inputs with adjacent plateaus (blue/red, red/orange, orange/purple, purple/green, green/light blue). That is, for a given pair, $p_{on}(t)$ due to the upper input (α_1) remains

higher over time relative to $p_{on}(t)$ due to the lower input (α_2). Now for the sequence of plateaued input pairs, the lower input in one pair is the upper input for the next pair (e.g., blue/red then red/orange in Figure 2.15A). **Thus, this observation combined with the analytical results implies that the observed order of inputs (Figure 2.15A, top plot) must be shared by their corresponding p_{on} trajectories (Figure 2.15A, second panel from the top). Furthermore, this means that the input with the highest valued plateau (Figure 2.15A, top plot, blue, continuous-like input) produces the highest $p_{on}(t)$, and the input with the lowest-valued plateau (zero) input (Figure 2.15A, top plot, light blue, single pulse input) produces the lowest.**

Furthermore, since all time t_0 and values α_i are arbitrary in the analytical treatment, these conclusions hold for input sequences that step down with any resolution, and therefore for any input that has a decreasing edge. We use $I = 6$ inputs to visually demonstrate our approach (Figure 2.15A). As I gets larger, where each step change is smaller and better represents a smooth input, the conclusions of our proof do not change. Our results are invariant to I . Thus, we can represent smooth pulses with large I .

Proof showing that $p_{on}(t_0 + \sigma, \alpha_1) > p_{on}(t_0 + \frac{\alpha_1}{\alpha_2}\sigma, \alpha_2)$ for $\sigma > 0$, where σ is proportional to accumulated TF area

In Figure 2.15A (second panel from the bottom), we plot p_{on} versus accumulated TF area for each input. This plot shows that the relative order is the same as it was for p_{on} versus accumulated time. To understand this, we will prove that this observed result must hold for any pair of inputs with adjacent plateaus (Figure 2.15A (top plot), blue/red, red/orange, orange/purple, purple/green, green/light blue) regardless of the values of k_{on} and k_{off} . With the same definitions for $p_{on}(t_0)$, t_0 , α_1 , and α_2 as in 2.7.1 Section "Proof showing that $p_{on}(t, \alpha_1) > p_{on}(t, \alpha_2)$ for $t > t_0$ and $\alpha_1 > \alpha_2$ " we now proceed.

For the $\alpha_1 TF_{max}$ input, over the duration $t_0 < t < t_0 + \sigma$, the accumulated $TF(t)$ area, denoted as $A_{\alpha_1}(\sigma)$, can formally be written in integral form as

$$\begin{aligned} A_{\alpha_1}(\sigma) &= \int_{t_0}^{t_0+\sigma} \alpha_1 TF_{max} dt \\ &= \alpha_1 TF_{max} \sigma \end{aligned} \tag{2.41}$$

Similarly, for the $\alpha_2 TF_{max}$ input, over the duration $t_0 < t < t_0 + \sigma'$, the accumulated $TF(t)$ area is $\alpha_2 TF_{max} \sigma'$. If we required that the two inputs have equal accumulated area then $\alpha_1 TF_{max} \sigma = \alpha_2 TF_{max} \sigma'$. Accordingly, $\sigma' = \frac{\alpha_1}{\alpha_2} \sigma$, i.e. σ' must be larger than σ in order to have the same area since the transcription factor amplitude is lower.

Here one can see that σ is a variable that can be used to transform both inputs (α_1 and α_2) from time to equal accumulated $TF(t)$ area. This allows us to analyze quantities such as $p_{on}(t)$ and accumulated transcripts as a function of σ . Now, we derive a relationship between the value of p_{on} evaluated at $t_0 + \sigma$ for an input $\alpha_1 TF_{max}$ input (which we denote $p_{on}(t_0 + \sigma, \alpha_1)$) and its value evaluated at $t_0 + \frac{\alpha_1}{\alpha_2} \sigma$ for an input $\alpha_2 TF_{max}$ (which we denote by $p_{on}(t_0 + \frac{\alpha_1}{\alpha_2} \sigma, \alpha_2)$). These two p_{on} values are therefore generated by two inputs of identical area.

Using the analytical solutions derived above, we can write:

$$\begin{aligned}
p_{on}(t_0 + \sigma, \alpha_1) - p_{on}(t_0 + \frac{\alpha_1}{\alpha_2}\sigma, \alpha_2) &= p_{on}(t_0) + \left(\frac{k_{on}\alpha_1 TF_{max}}{k_{on}\alpha_1 TF_{max} + k_{off}} - p_{on}(t_0) \right) \\
&\quad \times \left(1 - \exp(-[k_{on}\alpha_1 TF_{max} + k_{off}]\sigma) \right) \\
&\quad - p_{on}(t_0) - \left(\frac{k_{on}\alpha_2 TF_{max}}{k_{on}\alpha_2 TF_{max} + k_{off}} - p_{on}(t_0) \right) \\
&\quad \times \left(1 - \exp(-[k_{on}\alpha_2 TF_{max} + k_{off}]\frac{\alpha_1}{\alpha_2}\sigma) \right) \quad (2.42) \\
&= \left(\frac{k_{on}\alpha_1 TF_{max}}{k_{on}\alpha_1 TF_{max} + k_{off}} - p_{on}(t_0) \right) \\
&\quad \times \left(1 - \exp(-[k_{on}\alpha_1 TF_{max} + k_{off}]\sigma) \right) \\
&\quad - \left(\frac{k_{on}\alpha_2 TF_{max}}{k_{on}\alpha_2 TF_{max} + k_{off}} - p_{on}(t_0) \right) \\
&\quad \times \left(1 - \exp(-[k_{on}\alpha_2 TF_{max} + k_{off}]\frac{\alpha_1}{\alpha_2}\sigma) \right)
\end{aligned}$$

There are 3 possibilities for the relationship between $p_{on}(t_0 + \sigma, \alpha_1)$ and $p_{on}(t_0 + \frac{\alpha_1}{\alpha_2}\sigma, \alpha_2)$.

1. $\frac{k_{on}\alpha_2 TF_{max}}{k_{on}\alpha_2 TF_{max} + k_{off}} - p_{on}(t_0) < \frac{k_{on}\alpha_1 TF_{max}}{k_{on}\alpha_1 TF_{max} + k_{off}} - p_{on}(t_0) < 0$: Here, the expression in Eq. (2.42) is greater than zero. This is because $1 - \exp(-[k_{on}\alpha_2 TF_{max} + k_{off}]\frac{\alpha_1}{\alpha_2}\sigma) > 1 - \exp(-[k_{on}\alpha_1 TF_{max} + k_{off}]\sigma)$ for $t > t_0$.
2. $\frac{k_{on}\alpha_1 TF_{max}}{k_{on}\alpha_1 TF_{max} + k_{off}} - p_{on}(t_0) > 0$ and $\frac{k_{on}\alpha_2 TF_{max}}{k_{on}\alpha_2 TF_{max} + k_{off}} - p_{on}(t_0) < 0$: Here, all terms in Eq. (2.42) are positive and thus greater than zero.
3. $\frac{k_{on}\alpha_1 TF_{max}}{k_{on}\alpha_1 TF_{max} + k_{off}} - p_{on}(t_0) > \frac{k_{on}\alpha_2 TF_{max}}{k_{on}\alpha_2 TF_{max} + k_{off}} - p_{on}(t_0) > 0$: For this case, we will present a more detailed analysis.

To begin we'll differentiate Eq. (2.42) with respect to σ . We get

$$\begin{aligned}
& \frac{d \left[p_{on}(t_0 + \sigma, \alpha_1) - p_{on}(t_0 + \frac{\alpha_1}{\alpha_2} \sigma, \alpha_2) \right]}{d\sigma} \\
&= \frac{d \left[\left(\frac{k_{on}\alpha_1 TF_{max}}{k_{on}\alpha_1 TF_{max} + k_{off}} - p_{on}(t_0) \right) \times \left(1 - \exp(-[k_{on}\alpha_1 TF_{max} + k_{off}]\sigma) \right) \right]}{d\sigma} \\
&\quad - \frac{d \left[\left(\frac{k_{on}\alpha_2 TF_{max}}{k_{on}\alpha_2 TF_{max} + k_{off}} - p_{on}(t_0) \right) \times \left(1 - \exp(-[k_{on}\alpha_2 TF_{max} + k_{off}]\frac{\alpha_1}{\alpha_2}\sigma) \right) \right]}{d\sigma} \\
&= [k_{on}\alpha_1 TF_{max} + k_{off}] \left(\frac{k_{on}\alpha_1 TF_{max}}{k_{on}\alpha_1 TF_{max} + k_{off}} - p_{on}(t_0) \right) \times \exp(-[k_{on}\alpha_1 TF_{max} + k_{off}]\sigma) \\
&\quad - [k_{on}\alpha_2 TF_{max} + k_{off}] \frac{\alpha_1}{\alpha_2} \left(\frac{k_{on}\alpha_2 TF_{max}}{k_{on}\alpha_2 TF_{max} + k_{off}} - p_{on}(t_0) \right) \\
&\quad \times \exp(-[k_{on}\alpha_2 TF_{max} + k_{off}]\frac{\alpha_1}{\alpha_2}\sigma) \\
&= \left(k_{on}\alpha_1 TF_{max} - [k_{on}\alpha_1 TF_{max} + k_{off}]p_{on}(t_0) \right) \times \exp(-[k_{on}\alpha_1 TF_{max} + k_{off}]\sigma) \\
&\quad - \left(k_{on}\alpha_1 TF_{max} - [k_{on}\alpha_1 TF_{max} + \frac{\alpha_1}{\alpha_2}k_{off}]p_{on}(t_0) \right) \times \exp(-[k_{on}\alpha_2 TF_{max} + k_{off}]\frac{\alpha_1}{\alpha_2}\sigma)
\end{aligned} \tag{2.43}$$

Because $\frac{\alpha_1}{\alpha_2} > 1$, then

$$k_{on}\alpha_1 TF_{max} - [k_{on}\alpha_1 TF_{max} + k_{off}]p_{on}(t_0) > k_{on}\alpha_1 TF_{max} - [k_{on}\alpha_1 TF_{max} + \frac{\alpha_1}{\alpha_2}k_{off}]p_{on}(t_0) > 0$$

and also $\exp(-[k_{on}\alpha_2 TF_{max} + k_{off}]\frac{\alpha_1}{\alpha_2}\sigma)$ is smaller than $\exp(-[k_{on}\alpha_1 TF_{max} + k_{off}]\sigma)$ for

all σ . Therefore, the expression in Eq. (2.43) is greater than zero. This taken together with

the fact that the initial condition in Eq. (2.42) at $\sigma = 0$ is equal to zero dictates that

Eq. (2.42) is greater than zero for $\sigma > 0$. Thus, $p_{on}(t_0 + \sigma, \alpha_1) > p_{on}(t_0 + \frac{\alpha_1}{\alpha_2}\sigma, \alpha_2)$ for $\sigma > 0$.

Thus, this analytical result implies that the observed order of inputs (Figure 2.15A, top plot) which are shared by their corresponding time-dependent p_{on} trajectories (Figure 2.15A, second panel from the top) must also be shared by the corresponding σ -dependent p_{on} trajectories. Indeed, when we plot p_{on} versus accumulated TF area for each input (Figure 2.15A, second panel from the bottom) the relative order is the same as it was for p_{on} versus accumulated

time.

Deriving relationships for the number of transcripts, specifically

$\Delta N_{\alpha_2}(\sigma) > \Delta N_{\alpha_1}(\sigma)$ for $\sigma > 0$

With the same definitions for σ , $p_{on}(t_0)$, t_0 , α_1 and α_2 , we now calculate the number of transcripts produced for a given value of σ . We will denote $\Delta N_{\alpha_1}(\sigma)$ as accumulated transcripts for the $\alpha_1 TF_{max}$ input. We will also denote by $\Delta N_{\alpha_2}(\sigma)$ the number of accumulated transcripts for the $\alpha_2 TF_{max}$ input over the same input area as for the $\alpha_1 TF_{max}$ input. Applying the right hand side of Eq. (2.11) to the $\alpha_1 TF_{max}$ input we get

$$\begin{aligned}
\Delta N_{\alpha_1}(\sigma) &= \int_{t_0}^{t_0+\sigma} \frac{\beta_1}{k_d} \left(1 - p_{on}(t, \alpha_1) \right) \alpha_1 TF_{max} dt \\
&= \int_{t_0}^{t_0+\sigma} \frac{\beta_1}{k_d} \left(1 - p_{on}(t_0) - \left(\frac{k_{on}\alpha_1 TF_{max}}{k_{on}\alpha_1 TF_{max} + k_{off}} - p_{on}(t_0) \right) \right. \\
&\quad \left. \times \left(1 - \exp(-[k_{on}\alpha_1 TF_{max} + k_{off}](t - t_0)) \right) \right) \alpha_1 TF_{max} dt \\
&= \frac{\beta_1}{k_d} \left[\alpha_1 TF_{max} \sigma - p_{on}(t_0) \alpha_1 TF_{max} \sigma \right. \\
&\quad \left. - \left(\frac{k_{on}\alpha_1 TF_{max}}{k_{on}\alpha_1 TF_{max} + k_{off}} - p_{on}(t_0) \right) \right. \\
&\quad \left. \times \left(\alpha_1 TF_{max} \sigma - \frac{1 - \exp(-[k_{on}\alpha_1 TF_{max} + k_{off}]\sigma)}{k_{on}\alpha_1 TF_{max} + k_{off}} \alpha_1 TF_{max} \right) \right] \quad (2.44)
\end{aligned}$$

Likewise, $\Delta N_{\alpha_2}(\sigma)$ is given by:

$$\begin{aligned}
\Delta N_{\alpha_2}(\sigma) &= \int_{t_0}^{t_0 + \frac{\alpha_1}{\alpha_2} \sigma} \frac{\beta_1}{k_d} \left(1 - p_{on}(t, \alpha_2)\right) \alpha_2 T F_{max} dt \\
&= \int_{t_0}^{t_0 + \frac{\alpha_1}{\alpha_2} \sigma} \frac{\beta_1}{k_d} \left(1 - p_{on}(t_0) \right. \\
&\quad \left. - \left(\frac{k_{on} \alpha_2 T F_{max}}{k_{on} \alpha_2 T F_{max} + k_{off}} - p_{on}(t_0) \right) \right. \\
&\quad \left. \times \left(1 - \exp(-[k_{on} \alpha_2 T F_{max} + k_{off}](t - t_0))\right) \right) \alpha_2 T F_{max} dt \\
&= \frac{\beta_1}{k_d} \left[\alpha_1 T F_{max} \sigma - p_{on}(t_0) \alpha_1 T F_{max} \sigma \right. \\
&\quad \left. - \left(\frac{k_{on} \alpha_2 T F_{max}}{k_{on} \alpha_2 T F_{max} + k_{off}} - p_{on}(t_0) \right) \right. \\
&\quad \left. \times \left(\alpha_1 T F_{max} \sigma - \frac{1 - \exp(-[k_{on} \alpha_2 T F_{max} + k_{off}] \frac{\alpha_1}{\alpha_2} \sigma)}{k_{on} \alpha_2 T F_{max} + k_{off}} \alpha_2 T F_{max} \right) \right] \quad (2.45)
\end{aligned}$$

Next we must show that $\Delta N_{\alpha_2} - \Delta N_{\alpha_1} > 0$. Explicitly this expression (divided by $\frac{\beta_1}{k_d}$) is

$$\begin{aligned}
\frac{k_d}{\beta_1} [\Delta N_{\alpha_2}(\sigma) - \Delta N_{\alpha_1}(\sigma)] &= \left(\frac{k_{on} \alpha_1 T F_{max}}{k_{on} \alpha_1 T F_{max} + k_{off}} - p_{on}(t_0) \right) \\
&\quad \times \left(\alpha_1 T F_{max} \sigma - \frac{1 - \exp(-[k_{on} \alpha_1 T F_{max} + k_{off}]\sigma)}{k_{on} \alpha_1 T F_{max} + k_{off}} \alpha_1 T F_{max} \right) \\
&\quad - \left(\frac{k_{on} \alpha_2 T F_{max}}{k_{on} \alpha_2 T F_{max} + k_{off}} - p_{on}(t_0) \right) \\
&\quad \times \left(\alpha_1 T F_{max} \sigma - \frac{1 - \exp(-[k_{on} \alpha_2 T F_{max} + k_{off}] \frac{\alpha_1}{\alpha_2} \sigma)}{k_{on} \alpha_2 T F_{max} + k_{off}} \alpha_2 T F_{max} \right) \\
&= \alpha_1 T F_{max} \left[\left(\frac{k_{on} \alpha_1 T F_{max}}{k_{on} \alpha_1 T F_{max} + k_{off}} - p_{on}(t_0) \right) \right. \\
&\quad \times \left(\sigma - \frac{1 - \exp(-[k_{on} \alpha_1 T F_{max} + k_{off}]\sigma)}{k_{on} \alpha_1 T F_{max} + k_{off}} \right) \\
&\quad - \left(\frac{k_{on} \alpha_2 T F_{max}}{k_{on} \alpha_2 T F_{max} + k_{off}} - p_{on}(t_0) \right) \\
&\quad \times \left(\sigma - \frac{1 - \exp(-[k_{on} \alpha_2 T F_{max} + k_{off}] \frac{\alpha_1}{\alpha_2} \sigma)}{[k_{on} \alpha_2 T F_{max} + k_{off}] \frac{\alpha_1}{\alpha_2}} \right) \left. \right] \quad (2.46)
\end{aligned}$$

At $\sigma = 0$, Eq. (2.46) is zero as it should be. We just need to show that the slope of Eq. (2.46) is positive with respect to σ . The expression for the slope is

$$\begin{aligned}
\frac{k_d}{\beta_1} \frac{d[\Delta N_{\alpha_2}(\sigma) - \Delta N_{\alpha_1}(\sigma)]}{d\sigma} &= \alpha_1 TF_{max} \left[\left(\frac{k_{on}\alpha_1 TF_{max}}{k_{on}\alpha_1 TF_{max} + k_{off}} - p_{on}(t_0) \right) \right. \\
&\quad \times \left(1 - \exp(-[k_{on}\alpha_1 TF_{max} + k_{off}]\sigma) \right) \\
&\quad - \left(\frac{k_{on}\alpha_2 TF_{max}}{k_{on}\alpha_2 TF_{max} + k_{off}} - p_{on}(t_0) \right) \\
&\quad \left. \times \left(1 - \exp(-[k_{on}\alpha_2 TF_{max} + k_{off}]\frac{\alpha_1}{\alpha_2}\sigma) \right) \right] \\
&= \alpha_1 TF_{max} \left[p_{on}(t_0) + \left(\frac{k_{on}\alpha_1 TF_{max}}{k_{on}\alpha_1 TF_{max} + k_{off}} - p_{on}(t_0) \right) \right. \\
&\quad \times \left(1 - \exp(-[k_{on}\alpha_1 TF_{max} + k_{off}]\sigma) \right) \\
&\quad - p_{on}(t_0) - \left(\frac{k_{on}\alpha_2 TF_{max}}{k_{on}\alpha_2 TF_{max} + k_{off}} - p_{on}(t_0) \right) \\
&\quad \left. \times \left(1 - \exp(-[k_{on}\alpha_2 TF_{max} + k_{off}]\frac{\alpha_1}{\alpha_2}\sigma) \right) \right] \\
&= \alpha_1 TF_{max} \left[p_{on}(t_0 + \sigma, \alpha_1) - p_{on}(t_0 + \frac{\alpha_1}{\alpha_2}\sigma, \alpha_2) \right] \\
&> 0
\end{aligned} \tag{2.47}$$

This is because for $\sigma > 0$, $p_{on}(t_0 + \sigma, \alpha_1) - p_{on}(t_0 + \frac{\alpha_1}{\alpha_2}\sigma, \alpha_2) > 0$ as demonstrated above. Thus, $\Delta N_{\alpha_2}(\sigma) - \Delta N_{\alpha_1}(\sigma) > 0$ for $\sigma > 0$. **This inequality verifies that more transcripts are produced from an input TF whose amplitude is reduced ($\alpha_2 TF_{max}$ input) and which extends more in time relative to a shorter but higher TF amplitude input ($\alpha_1 TF_{max}$ input) when the TF input area is equal between the two inputs. These results imply that given the relative order of the inputs (Figure 2.15A, top plot) the associated order of accumulated transcripts versus accumulated TF area must be reversed (compare the top plot in Figure 2.15A to the bottom plot). Overall, these results show that the lower the plateau value is for this class of inputs, the higher the accumulated**

transcripts as a function of accumulated area.

Analysis of a class of inputs that relate the number of transcripts of the continuous input $TF_c(t)$ to that of pulsed input $TF_p(t)$ for the two pulse case

Here, we will start using all the properties derived above to establish relationships between the number of transcripts generated by a continuous input $TF_c(t)$ and that of a pulsed input $TF_p(t)$ for $M = 2$ pulses. Figure 2.15B (top plot) shows the classes of inputs we will consider here. These are similar to those in Figure 2.15A (top plot), in that the inputs descend to their respective plateaus in exactly the same manner, but later rise to TF_{max} and then shut off. Here the input with the highest valued plateau represents the continuous input $TF_c(t)$ (Figure 2.15B (top plot, dark blue)) while the input with the lowest valued plateau represents the pulsed input $TF_p(t)$ (Figure 2.15B (top plot, light blue, two pulses)).

Importantly, the inputs are constructed such that the total TF area is the same

for all inputs. As we did for the analysis above, we analyze each pair of inputs with adjacent plateau values (Figure 2.15B (top plot, blue/red, red/orange, orange/purple, purple/green, green/light blue)) where the input with the higher plateau value will be called $TF_{\alpha_1}(t)$ and the input with the lower plateau value is $TF_{\alpha_2}(t)$. Similar to the above analysis, we'll denote the corresponding p_{on} for each these inputs as $p_{on}(t, \alpha_1)$ and $p_{on}(t, \alpha_2)$, respectively. To begin, we will first consider a continuous input that rises, plateaus to a value $\alpha_1 TF_{max}$, ($\alpha_1 = 1$ in this case for illustration purposes) and then shuts off (Figure 2.15B (top middle plot), dark blue, $TF_{\alpha_1}(t)$). The input with a plateau value adjacent to the continuous input (Figure 2.15B (top middle plot), red, $TF_{\alpha_2}(t)$) rises in the same fashion, plateaus, and then drops at time t_0 to a lower plateau value of $\alpha_2 TF_{max}$, staying there until time $t_0 + \frac{\alpha_1}{\alpha_2} \sigma^*$. The input then jumps up to TF_{max} and then shuts off. The total accumulated TF area for $TF_{\alpha_1}(t)$ at $t = t_0 + \sigma^*$ is identical to that for $TF_{\alpha_2}(t)$ at $t = t_0 + \frac{\alpha_1}{\alpha_2} \sigma^*$. The shut-off for the two TF inputs are identical, albeit shifted in time with

respect to each other. That is, $TF_{\alpha_1}(t_0 + \sigma^* + \tau) = TF_{\alpha_2}(t_0 + \frac{\alpha_1}{\alpha_2}\sigma^* + \tau) = TF_{off}(\tau)$ for $\tau > 0$.

We know from our discussion and derivations above that the total transcripts produced from $TF_{\alpha_1}(t)$ at $t = t_0 + \sigma^*$ is less than that of $TF_{\alpha_2}(t)$ at $t = t_0 + \frac{\alpha_1}{\alpha_2}\sigma^*$. To analyze the transcripts produced from the shut-off of the input for each case, we will begin by first showing that for the identical shut-off of the TF inputs, $p_{on}(t_0 + \frac{\alpha_1}{\alpha_2}\sigma^* + \tau, \alpha_2)$ for $TF_{\alpha_2}(t_0 + \frac{\alpha_1}{\alpha_2}\sigma^* + \tau)$ is always less than the $p_{on}(t_0 + \sigma^* + \tau, \alpha_1)$ for $TF_{\alpha_1}(t_0 + \sigma^* + \tau)$ at every τ for $\tau > 0$. Let us first model $TF_{off}(\tau)$ for $\tau > 0$ as a series of step changes with arbitrary resolution. This allows us to apply our analytical solutions over the duration of each step for two different initial conditions where we will show that the lower initial condition will remain lower over the duration of the step. To derive this result, consider $p_{on}(t, \alpha)$ which has the initial condition $p_{on}(t_0)$ at $t = t_0$, and which has the solution for the input $TF(t) = \alpha TF_{max}$ for $0 \leq \alpha \leq 1$ and for $t > t_0$ is

$$\begin{aligned} p_{on}(t, \alpha) &= p_{on}(t_0) + \left(\frac{k_{on}\alpha TF_{max}}{k_{on}\alpha TF_{max} + k_{off}} - p_{on}(t_0) \right) \times \left(1 - \exp(-[k_{on}\alpha TF_{max} + k_{off}][t - t_0]) \right) \end{aligned} \quad (2.48)$$

Likewise consider $\tilde{p}_{on}(t, \alpha)$ which has initial condition $\tilde{p}_{on}(t_0)$ at $t = t_0$ and where $\tilde{p}_{on}(t_0) < p_{on}(t_0)$. The solution for $\tilde{p}_{on}(t, \alpha)$ would be

$$\begin{aligned} \tilde{p}_{on}(t, \alpha) &= \tilde{p}_{on}(t_0) + \left(\frac{k_{on}\alpha TF_{max}}{k_{on}\alpha TF_{max} + k_{off}} - \tilde{p}_{on}(t_0) \right) \times \left(1 - \exp(-[k_{on}\alpha TF_{max} + k_{off}][t - t_0]) \right) \end{aligned} \quad (2.49)$$

Now, $p_{on}(t, \alpha) - \tilde{p}_{on}(t, \alpha)$ is given by:

$$\begin{aligned}
p_{on}(t, \alpha) - \tilde{p}_{on}(t, \alpha) &= p_{on}(t_0) + \left(\frac{k_{on}\alpha TF_{max}}{k_{on}\alpha TF_{max} + k_{off}} - p_{on}(t_0) \right) \\
&\quad \times \left(1 - \exp(-[k_{on}\alpha TF_{max} + k_{off}][t - t_0]) \right) \\
&\quad - \tilde{p}_{on}(t_0) - \left(\frac{k_{on}\alpha TF_{max}}{k_{on}\alpha TF_{max} + k_{off}} - \tilde{p}_{on}(t_0) \right) \\
&\quad \times \left(1 - \exp(-[k_{on}\alpha TF_{max} + k_{off}][t - t_0]) \right) \\
&= \left[p_{on}(t_0) - \tilde{p}_{on}(t_0) \right] \left[1 - \left(1 - \exp(-[k_{on}\alpha TF_{max} + k_{off}][t - t_0]) \right) \right] \\
&> 0
\end{aligned} \tag{2.50}$$

Now recall that we are modeling $TF_{off}(\tau)$ for $\tau > 0$ as a series of step changes with arbitrary resolution. This analytical result implies that the function with the lower initial condition at $\tau = 0$, $p_{on}(t_0 + \frac{\alpha_1}{\alpha_2}\sigma^* + \tau, \alpha_2)$, will remain lower over the series of steps in $TF_{off}(\tau)$ for $\tau > 0$. This is simply because at the end of a given step, the lower valued function at this point in time represents the lower initial condition for the next step and will remain lower over that step. Thus, the lower valued function must remain lower through the whole series of steps until the input shuts off. To reiterate, since $p_{on}(t_0 + \sigma^* + \tau, \alpha_1) > p_{on}(t_0 + \frac{\alpha_1}{\alpha_2}\sigma^* + \tau, \alpha_2)$ at $\tau = 0$, and since both are experiencing $TF_{off}(\tau)$, our results above imply that $p_{on}(t_0 + \sigma^* + \tau, \alpha_1) > p_{on}(t_0 + \frac{\alpha_1}{\alpha_2}\sigma^* + \tau, \alpha_2)$ for $\tau > 0$. Importantly, this is independent of the shape of $TF_{off}(\tau)$.

We are now poised to determine which input, $TF_{\alpha_1}(t)$ or $TF_{\alpha_2}(t)$, produces more total transcripts. For the $TF_{\alpha_1}(t)$ input, applying the right hand side of Eq. (2.11), we have

$$\begin{aligned}
N_{\alpha_1} &= \int_0^5 \frac{\beta_1}{k_d} (1 - p_{on}(t, \alpha_1)) TF_{\alpha_1}(t) dt \\
&= \int_0^{t_0 + \sigma^*} \frac{\beta_1}{k_d} (1 - p_{on}(t, \alpha_1)) TF_{\alpha_1}(t) dt + \int_{t_0 + \sigma^*}^5 \frac{\beta_1}{k_d} (1 - p_{on}(t, \alpha_1)) TF_{\alpha_1}(t) dt \\
&= N_{\alpha_1}(t_0 + \sigma^*) + \int_0^{5 - [t_0 + \sigma^*]} \frac{\beta_1}{k_d} (1 - p_{on}(t_0 + \sigma^* + \tau, \alpha_1)) TF_{off}(\tau) d\tau
\end{aligned} \tag{2.51}$$

And for the $TF_{\alpha_2}(t)$ input we have

$$\begin{aligned}
N_{\alpha_2} &= \int_0^5 \frac{\beta_1}{k_d} (1 - p_{on}(t, \alpha_2)) TF_{\alpha_2}(t) dt \\
&= \int_0^{t_0 + \frac{\alpha_1}{\alpha_2} \sigma^*} \frac{\beta_1}{k_d} (1 - p_{on}(t, \alpha_2)) TF_{\alpha_2}(t) dt + \int_{t_0 + \frac{\alpha_1}{\alpha_2} \sigma^*}^5 \frac{\beta_1}{k_d} (1 - p_{on}(t, \alpha_2)) TF_{\alpha_2}(t) dt \\
&= N_{\alpha_2}(t_0 + \frac{\alpha_1}{\alpha_2} \sigma^*) + \int_0^{5 - [t_0 + \frac{\alpha_1}{\alpha_2} \sigma^*]} \frac{\beta_1}{k_d} (1 - p_{on}(t_0 + \frac{\alpha_1}{\alpha_2} \sigma^* + \tau, \alpha_2)) TF_{off}(\tau) d\tau
\end{aligned} \tag{2.52}$$

We have previously shown that $N_{\alpha_2}(t_0 + \frac{\alpha_1}{\alpha_2} \sigma^*) > N_{\alpha_1}(t_0 + \sigma^*)$ (Eq. (2.47)). Since $p_{on}(t_0 + \sigma^* + \tau, \alpha_1) > p_{on}(t_0 + \frac{\alpha_1}{\alpha_2} \sigma^* + \tau, \alpha_2)$ for $\tau > 0$, this means that $1 - p_{on}(t_0 + \sigma^* + \tau, \alpha_1) < 1 - p_{on}(t_0 + \frac{\alpha_1}{\alpha_2} \sigma^* + \tau, \alpha_2)$ for $\tau > 0$. Thus, the integral term on the last line of the N_{α_2} equation, Eq. (2.52), will be greater than that of the N_{α_1} equation, Eq. (2.51). We can therefore conclude that $N_{\alpha_2} > N_{\alpha_1}$ for $\alpha_1 > \alpha_2$. Additionally, analysis of the next pair of inputs with adjacent plateau values (Figure 2.15B (bottom middle plot), red ($TF_{\alpha_1}(t)$) and orange ($TF_{\alpha_2}(t)$)) follows exactly the same logic and will yield the same result. This remains true for every subsequent pair of inputs with adjacent plateau values from Figure 2.15B (top plot) where the $TF_{\alpha_2}(t)$ input always produces more total transcripts than the $TF_{\alpha_1}(t)$ input. **Thus, for this class of inputs, the lower an input's plateau value, the more total transcripts that are produced (Figure 2.15B (bottom plot)), with the continuous input $TF_c(t)$ producing the least and the pulsed input $TF_p(t)$ producing the most.**

Extending the approach to a higher number of pulses

This analysis also applies to inputs with larger numbers of pulses. We'll first start with the three pulse input (Figure 2.15C). We extend the analysis from the two pulse input for the first two adjacent input pairs (Figure 2.15B (middle plots)) to accommodate the three pulse input (Figure 2.15C (middle plots)). As above with the two pulse input, $t = t_0$

represents the time at which divergence of the inputs TF_{α_1} and TF_{α_2} occurs, and where TF_{α_2} drops to its plateau value. Additionally, at $t = t_0$, both $TF_{\alpha_1}(t)$ and $TF_{\alpha_2}(t)$ have the same TF area. As above with the two pulse input, at $t = t_0 + \frac{\alpha_1}{\alpha_2}\sigma^*$, $TF_{\alpha_2}(t)$ jumps back up to $TF_{\alpha_1}(t)$, and where $TF_{\alpha_1}(t)$ at $t = t_0 + \sigma^*$ has the same corresponding accumulated TF area as $TF_{\alpha_2}(t)$ does at $t = t_0 + \frac{\alpha_1}{\alpha_2}\sigma^*$. For the three pulse input, unlike the two pulse input, $TF_{\alpha_2}(t)$ then drops a second time to its plateau value at $t = t_1(\alpha_2)$. Correspondingly, the TF area at $t = t_1(\alpha_1)$ for $TF_{\alpha_1}(t)$ is the same as the TF area at $t = t_1(\alpha_2)$ for $TF_{\alpha_2}(t)$. For the first adjacent input pair, $t_0 + \frac{\alpha_1}{\alpha_2}\sigma^*$ happens to share the same value as $t_1(\alpha_1)$ (Figure 2.15C (top middle plot)). This is the result of the particular choice we make—i.e. triangular pulses, that we are modeling with steps. However, the second adjacent input pair does not exhibit this equality (Figure 2.15C (bottom middle plot)) nor would any succeeding pair.

Given the analysis above for the two pulse input, we already know that for the three pulse input, the total transcripts produced from $TF_{\alpha_1}(t)$ at $t = t_0 + \sigma^*$ is less than that of $TF_{\alpha_2}(t)$ at $t = t_0 + \frac{\alpha_1}{\alpha_2}\sigma^*$ (Figure 2.15C (middle plots)). Then, similar to the two pulse input, the inputs are identical but shifted, i.e. $TF_{\alpha_1}(t_0 + \sigma^* + \tau) = TF_{\alpha_2}(t_0 + \frac{\alpha_1}{\alpha_2}\sigma^* + \tau)$, but only over the duration $0 < \tau \leq t_1(\alpha_1) - [t_0 + \sigma^*] = t_1(\alpha_2) - [t_0 + \frac{\alpha_1}{\alpha_2}\sigma^*]$ (Figure 2.15C (middle plots)). From the general implications of Eq. (2.50), we know that since $p_{on}(t_0 + \sigma^*, \alpha_1 + \tau) > p_{on}(t_0 + \frac{\alpha_1}{\alpha_2}\sigma^* + \tau, \alpha_2)$ at $\tau = 0$, this relationship must hold over the duration of τ . $TF_{\alpha_2}(t)$ will therefore yield more transcripts through the right-hand side of Eq (2.11) over this duration. Thus, for the same accumulated TF area, i.e. $TF_{\alpha_1}(t)$ up until $t = t_1(\alpha_1)$ and $TF_{\alpha_2}(t)$ up until $t = t_1(\alpha_2)$, $TF_{\alpha_2}(t)$ produces more transcripts.

Now for the three pulse input, $TF_{\alpha_1}(t_1(\alpha_1) + \tau)$ is identical to $TF_{\alpha_1}(t_0 + \tau)$ for the two pulse input for $\tau > 0$. Additionally, for the three pulse input, $TF_{\alpha_2}(t_1(\alpha_2) + \tau)$ is identical to $TF_{\alpha_2}(t_0 + \tau)$ for the two pulse input for $\tau > 0$. For the two pulse input, $p_{on}(t_0, \alpha_1) = p_{on}(t_0, \alpha_2)$ and $TF_{\alpha_2}(t)$ produces more transcripts. However, $p_{on}(t_1(\alpha_1), \alpha_1) > p_{on}(t_1(\alpha_2), \alpha_2)$ for the three pulse input. Thus, $TF_{\alpha_2}(t)$ must produce

more transcripts for $\tau > 0$. We can therefore conclude that $TF_{\alpha_2}(t)$ produces more transcripts than $TF_{\alpha_1}(t)$ for the three pulse input, i.e. $N_{\alpha_2} > N_{\alpha_1}$. This same analysis can then be applied to each successive pair of inputs with adjacent plateau values to ultimately show that the pulsed input $TF_p(t)$ yields the most transcripts for this class of inputs while the continuous input $TF_c(t)$ yields the least (Figure 2.15C (bottom plot)).

The logic and analysis developed for the three pulse input can be applied to the same class of inputs with an arbitrary number of pulses to show that $TF_p(t)$ yields the most transcripts for this class of inputs while the continuous input $TF_c(t)$ yields the least, regardless of the number of pulses. To begin, we discuss the first two adjacent input pairs (Figure 2.15D) where we plot just past the beginning of the third drop to its plateau value for $TF_{\alpha_2}(t)$. As above, $t = t_0$ represents the time where $TF_{\alpha_2}(t)$ first drops to its plateau value, and at which both $TF_{\alpha_1}(t)$ and $TF_{\alpha_2}(t)$ have the same TF area. In general, all successive drops to the plateau value of $TF_{\alpha_2}(t)$ start at $t = t_i(\alpha_2)$, where $1 \leq i \leq M - 2$, where M is the total number of pulses for this class of inputs. Given $t = t_i(\alpha_2)$ for $TF_{\alpha_2}(t)$, the time that corresponds to the same TF area for $TF_{\alpha_1}(t)$ is $t = t_i(\alpha_1)$.

For the three pulse input, we used the right-hand side of Eq (2.11) to show that the transcripts produced from $t = t_0$ to $t = t_1(\alpha_1)$ for $TF_{\alpha_1}(t)$ is less than that for $TF_{\alpha_2}(t)$ from $t = t_0$ to $t = t_1(\alpha_2)$. This comparison is over equivalent TF area. Now for $TF_{\alpha_1}(t)$ from $t = t_1(\alpha_1)$ to $t = t_2(\alpha_1)$, the input is the same as it was from $t = t_0$ to $t = t_1(\alpha_1)$. Similarly, for $TF_{\alpha_2}(t)$ from $t = t_1(\alpha_2)$ to $t = t_2(\alpha_2)$, the input is the same as it was from $t = t_0$ to $t = t_1(\alpha_2)$. This repeats at every $t = t_i(\alpha_1)$ for $TF_{\alpha_1}(t)$ and every $t = t_i(\alpha_2)$ for $TF_{\alpha_2}(t)$ until $i = M - 2$ just prior to shutoff. Importantly, at $t = t_0$, $p_{on}(t_0, \alpha_1) = p_{on}(t_0, \alpha_2)$, but at the beginning of every succeeding repeated sequence pair, $p_{on}(t_i(\alpha_1), \alpha_1) > p_{on}(t_i(\alpha_2), \alpha_2)$ for all $1 \leq i \leq M - 2$. Hypothetically, if $p_{on}(t_i(\alpha_1), \alpha_1) = p_{on}(t_i(\alpha_2), \alpha_2)$, we know that $TF_{\alpha_2}(t)$ will yield more transcripts through Eq (2.11) from $t = t_i(\alpha_2)$ to $t = t_{i+1}(\alpha_2)$ than $TF_{\alpha_1}(t)$ from $t = t_i(\alpha_1)$ to $t = t_{i+1}(\alpha_1)$. Thus, since $p_{on}(t_i(\alpha_1), \alpha_1) > p_{on}(t_i(\alpha_2), \alpha_2)$, this

must still hold. We can therefore conclude that over every succeeding repeated sequence pair, $TF_{\alpha_2}(t)$ will always yield more transcripts. Furthermore, we can apply the shutoff results from the three pulse input above to conclude that $TF_{\alpha_2}(t)$ will yield more transcripts for $t > t_{M-2}(\alpha_2)$ (through shutoff) than $TF_{\alpha_1}(t)$ for $t > t_{M-2}(\alpha_1)$ (through shutoff). Thus, in general, for any number of pulses M , $N_{\alpha_2} > N_{\alpha_1}$. This same analysis can then be applied to each successive pair of inputs with adjacent plateau values to ultimately show that the pulsed input $TF_p(t)$ yields the most transcripts for this class of inputs while the continuous input $TF_c(t)$ yields the least (see twenty pulse example in Figure 2.15E).

Finally, this analysis proves that the continuous input $TF_c(t)$ will always produce less total transcripts than the pulsed input $TF_p(t)$, given the constraint of equal total accumulated TF. Thus, the slope ratio of the total transcripts, N_p/N_c , will remain greater than one. Importantly, while we use triangle-like pulses as a visual example, the analysis is general for any pulse type that has a single rise (step-ups allowed as in the examples presented) followed by a fall (step-downs allowed as in the examples).

Asymptotic analysis of mRNA dynamics for the simple promoter to a step function input

We can evaluate the promoter model (Figure 2.4A and Eq.(1) in the main text) and the resulting mRNA dynamics for a step input using linear systems analysis. For a four hour constant input experiment, $TF(t) \approx TF_{max}u(t)$, a step function, since $TF(t)$ equilibrates to TF_{max} within two minutes (Figure 2.3A). The solution for p_{on} would be

$$p_{on}(t) = \frac{TF_{max}}{TF_{max} + \frac{k_{off}}{k_{on}}} u(t) [1 - \exp(-[k_{on}TF_{max} + k_{off}]t)] \quad (2.53)$$

The mRNA equation would then be

$$\frac{dmRNA}{dt} = \beta_1 p_{on}(t) - \gamma_1 mRNA \quad (2.54)$$

where β_1 is the production rate of the promoter and γ_1 is the mRNA decay rate. The solution for the mRNA equation would then be

$$\begin{aligned} mRNA(t) &= \beta_1 \frac{TF_{max}}{TF_{max} + \frac{k_{off}}{k_{on}}} \left[u(t) - u(t) \exp(-[k_{on}TF_{max} + k_{off}]t) \right] * u(t) \exp(-\gamma_1 t) \\ &= \beta_1 \frac{TF_{max}}{TF_{max} + \frac{k_{off}}{k_{on}}} u(t) \left[\frac{1}{\gamma_1} - \frac{1}{\gamma_1} \exp(-\gamma_1 t) \right. \\ &\quad - \frac{1}{\gamma_1 - [k_{on}TF_{max} + k_{off}]} \exp(-[k_{on}TF_{max} + k_{off}]t) \\ &\quad \left. + \frac{1}{\gamma_1 - [k_{on}TF_{max} + k_{off}]} \exp(-\gamma_1 t) \right] \end{aligned} \quad (2.55)$$

where '*' denotes convolution. Here we have used the formula

$$[u(t) \exp(-\lambda_1 t)] * [u(t) \exp(-\lambda_2 t)] = \frac{1}{\lambda_2 - \lambda_1} \left[u(t) \exp(-\lambda_1 t) - u(t) \exp(-\lambda_2 t) \right] \quad (2.56)$$

for $\lambda_1 \neq \lambda_2$. Now let's look at two extreme regimes. The first being $k_{on}TF_{max} + k_{off} \gg \gamma_1$ (where $k_{on}TF_{max} + k_{off} \gg 1$ is also true). For this case

$$mRNA(t) \approx \beta_1 \frac{TF_{max}}{TF_{max} + \frac{k_{off}}{k_{on}}} u(t) \left[\frac{1}{\gamma_1} - \frac{1}{\gamma_1} \exp(-\gamma_1 t) \right] \quad (2.57)$$

Here is dominant effect in the time signal $mRNA(t)$ is the slow γ_1 with exponential timescale $T_s = 1/\gamma_1$. The other extreme regime is $k_{on}TF_{max} + k_{off} \ll \gamma_1$ (where $\gamma_1 \gg 1$ is also true). For this case

$$mRNA(t) \approx \beta_1 \frac{TF_{max}}{TF_{max} + \frac{k_{off}}{k_{on}}} u(t) \left[\frac{1}{\gamma_1} - \frac{1}{\gamma_1} \exp(-[k_{on}TF_{max} + k_{off}]t) \right] \quad (2.58)$$

Here the dominant effect in the time signal $mRNA(t)$ is the slow $k_{on}TF_{max} + k_{off}$ with exponential timescale $T_s = 1/(k_{on}TF_{max} + k_{off})$. We denote T_s as the settling time for the dynamic data presented in Figure 2.5 in the main text.

2.8 Supplemental Figures

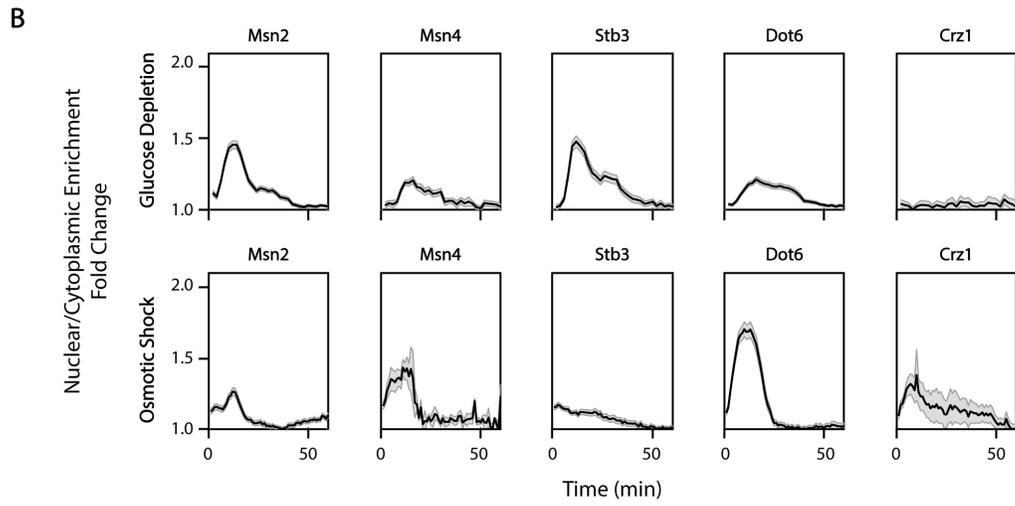
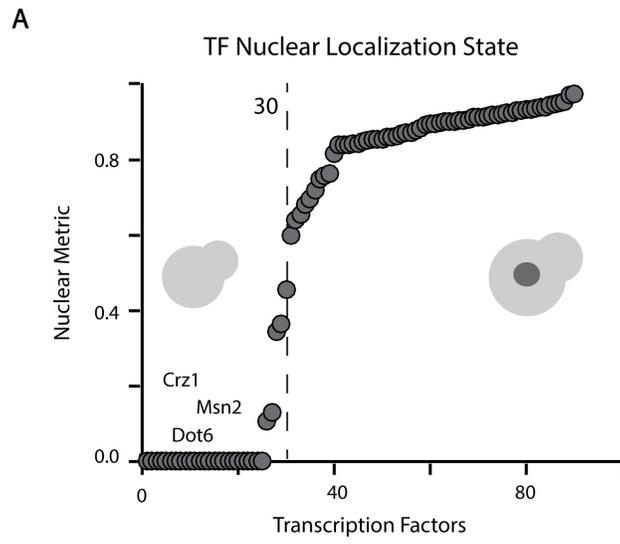


Figure 2.7: Approximately one-third of TFs are basally cytoplasmic in log phase and a subset are shown to exhibit transient nuclear localization, related to Figure 2.1. A) LOC scores of available transcription factors from the CYCLOPs database are plotted (Chong et al 2015). The LOC score³³ is the number of cells assigned to a specific location (nucleus in this instance) over the total number of cells in any subcellular location. Increasing LOC score denotes increasing nuclear enrichment. B) Fold change of nuclear enrichment for a panel of stress-responsive transcription factors (Msn2, Msn4, Stb3, Dot6, and Crz1) are plotted as a function of time in response to environmental inputs (Glucose depletion and osmotic shock). For glucose depletion, SDC media (2% glucose) is replaced with SD media with 0.05% glucose. For osmotic shock, SDC media is replaced with SDC media with 0.95M sorbitol. Imaging begins at $t = 0$ after addition of environmental perturbation and samples are imaged every 30 seconds. The solid black lines represent the mean of single cell traces and the shading represents the standard error of the mean.

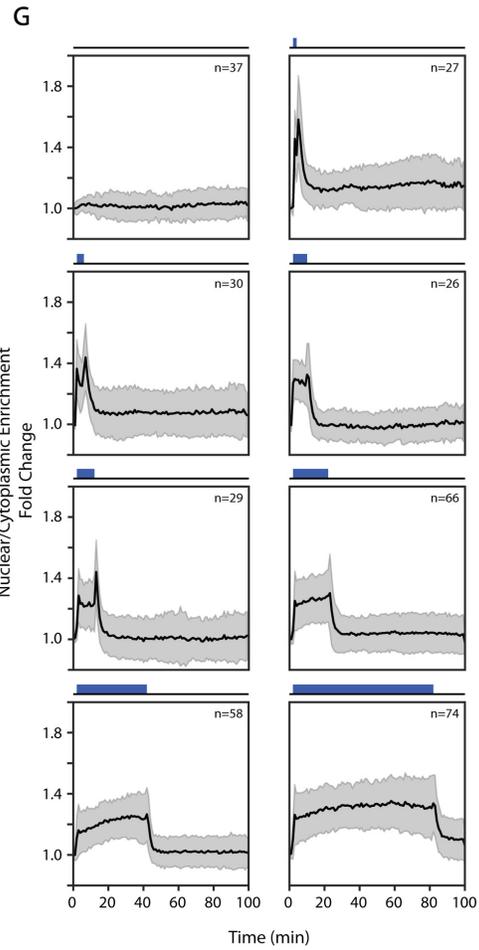
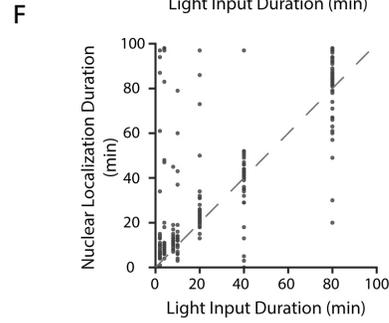
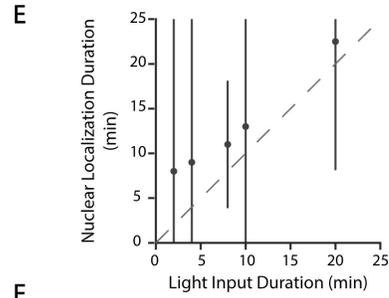
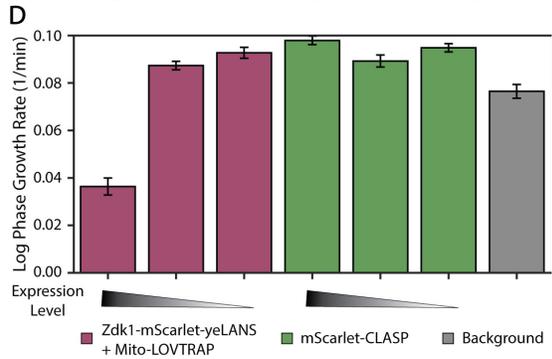
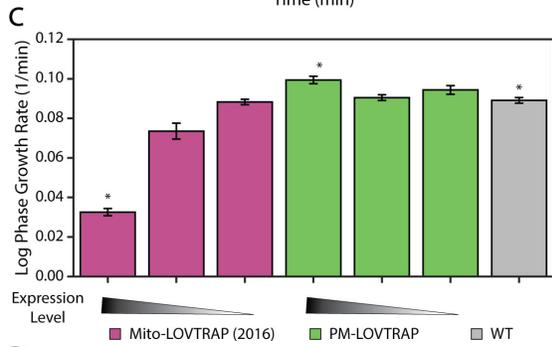
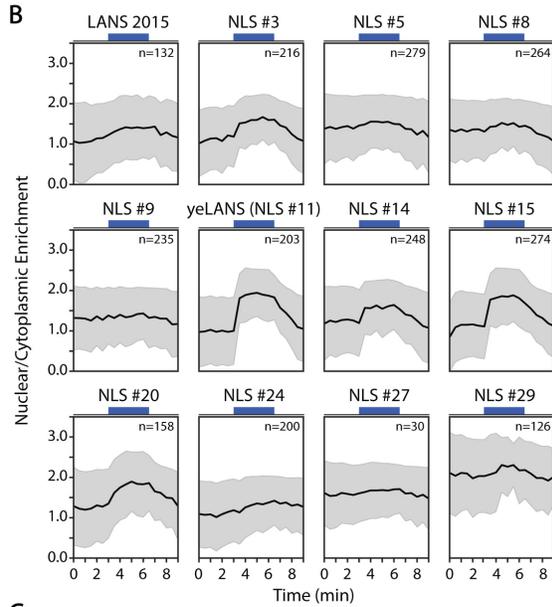
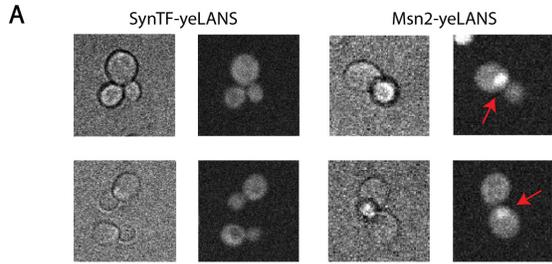


Figure 2.8: Optimization of LANS and LOVTRAP and CLASP characterization, related to Figure 2.1. A) Confocal microscopy images of yeast expressing SynTF-yeLANS and Msn2-yeLANS in the absence of blue light. Red arrows (inset) denote examples of cells that exhibit nuclear/cytoplasmic localization of Msn2-yeLANS. B) Mean nuclear/cytoplasmic enrichment (nuclear intensity divided by cytoplasm intensity) is plotted as a function of time. Shaded error represents standard deviation and light input regimes are illustrated above graphs. n refers to number of cells tracked and subplot headings (e.g., NLS 3) correspond to NLS peptides listed in Table 2.1. C) Comparison of Mito-LOVTRAP and PM-LOVTRAP strains. Mito-LOVTRAP and PM-LOVTRAP are expressed from pTDH3 (highest), pRPL18B (medium), and pREV1 (lowest) promoters. Strains marked with an asterisk denote those for which growth curves are plotted in Figure 2.1B of main text. Error bars represent standard error of the linear regression for data from 3 growth experiments. D) Comparison of Zdk1-mScarlet-yeLANS + Mito-LOVTrap and CLASP. Both components (e.g. Mito-LOVTRAP and Zdk1-mScarlet-yeLANS) are expressed at the same level, using either pTDH3, pRPL18B, or pREV1 promoters. Background (control strain) denotes the WT strain with pSYN1-YFP integrated in the LEU locus. Zdk1-mScarlet-yeLANS + Mito-LOVTRAP and mScarlet-CLASP strains also have this integration. Error bars represent standard error of the linear regression for data from 3 growth experiments. E) A zoom-in of Figure 2.1D in the main text that shows median duration of nuclear localization as a function of light input duration; the line $X = Y$ is denoted by the dashed line. The zoomed graph illustrates that for short pulse durations, the OFF time – time that nuclear localization extends past the pulse – is not linearly related to light input duration. F) A scatterplot that shows duration of nuclear localization as a function of light input duration. Each point represents a single cell. G) Mean nuclear/cytoplasmic enrichment fold change as a function of time for mScarlet-CLASP induced with blue light. Light input regimes are illustrated above graphs (indicating 0, 2, 4, 8, 10, 20, 40, or 80 minute light input). In all plots, except where noted, error (bars or shading) represents standard deviation.

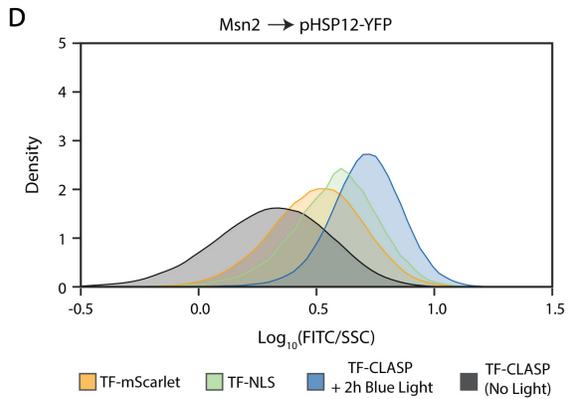
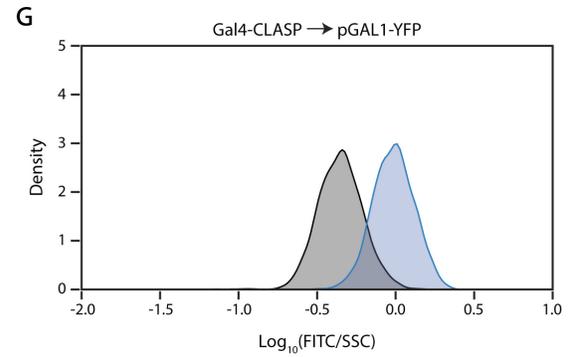
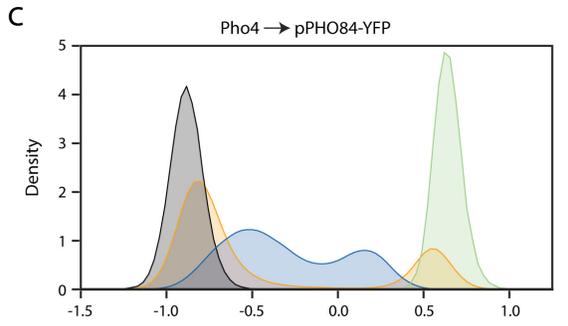
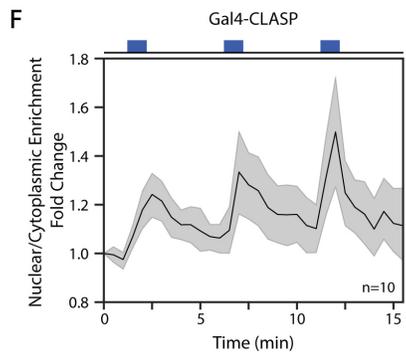
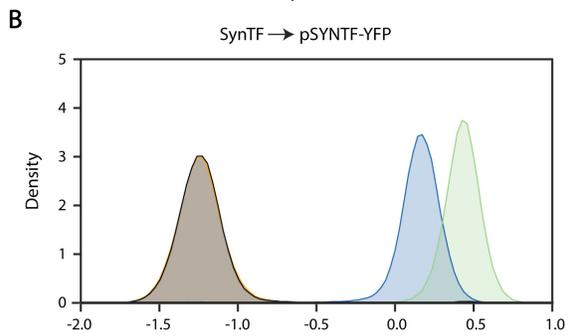
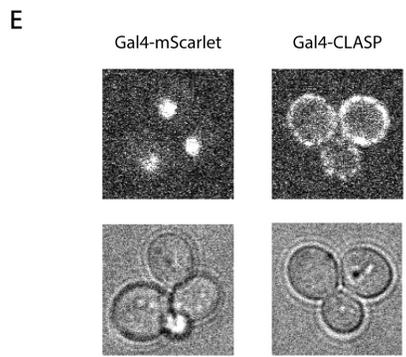
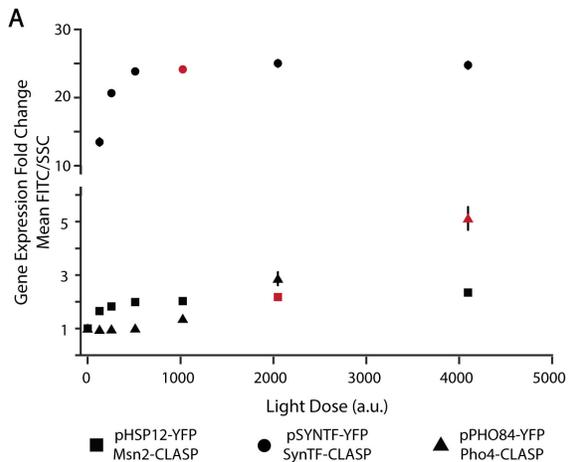


Figure 2.9: Characterization of TF-CLASP strains, related to Figure 2.2. A) Mean FITC/SSC is plotted as a function of light intensity (a.u.) for strains that are exposed to different amplitudes of light for two hours (continuous input). Marked in red is the lowest light dose which yielded near-maximal expression for each strain (>90%); this dose is used in all microscopy and flow cytometry experiments for each strain. Error bars represent standard error of the mean for 9 biological replicates. Light doses are quantified in mW in Table 2.2. B-D) Each subplot shows the probability density functions of $\log_{10}(\text{FITC}/\text{SSC})$ of gene expression of corresponding fluorescent promoter fusion for TF-CLASP, TF-NLS (constitutive nuclear localization), and TF-mScarlet (basal localization) strains. Distributions display expression from 9 biological replicates (data from replicates are pooled). TF-mScarlet strains are not exposed to light, for facile comparison to TF-CLASP (No Light) expression. TF-NLS strains are exposed to two hours of blue light (continuous input) to control for the effect of blue light on YFP fluorescence when comparing to TF-CLASP (Light) expression. For all panels, TF cargos are expressed from pRPL18B. E) RFP (top panels) and brightfield (bottom panels) images of mScarlet-tagged Gal4 (left panels) and mScarlet-tagged Gal4-CLASP (right panels). F) Gal4 nuclear/cytoplasmic enrichment is plotted as a function of time. Light input regime is illustrated above graph. Shaded gray area represents 95% confidence interval. n refers to the number of cells tracked. G) Gene expression of pGal1-YFP resulting from Gal4-CLASP nuclear localization following two hours of blue light input at 25% of the maximal Optoplate intensity. Distributions display expression from 3 biological replicates (data from replicates is pooled). Shown is the probability density function of $\log_{10}(\text{FITC}/\text{SSC})$ of pGAL1-YFP in the dark (gray) or after light exposure (blue).

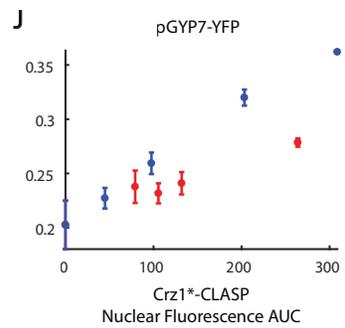
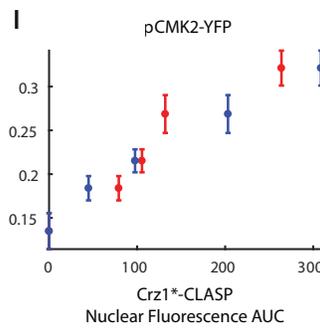
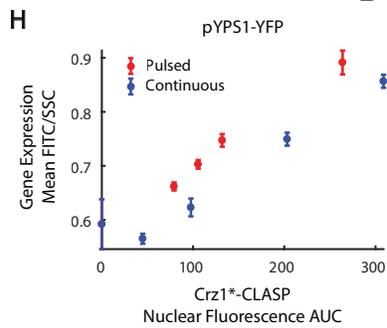
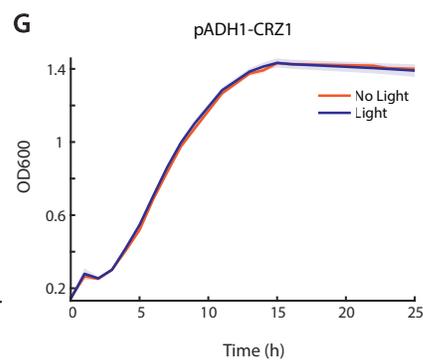
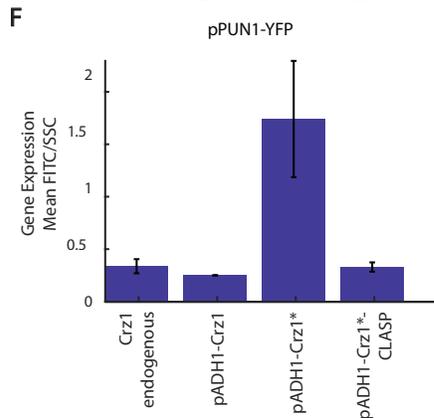
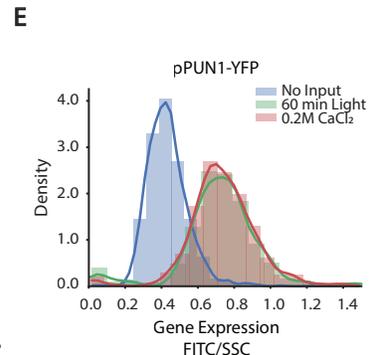
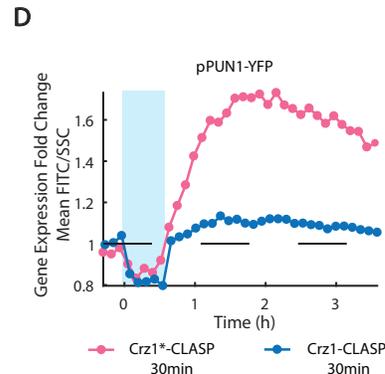
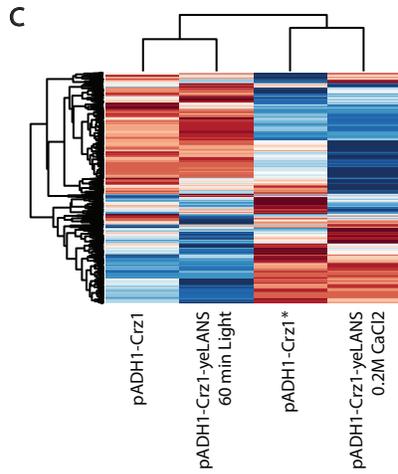
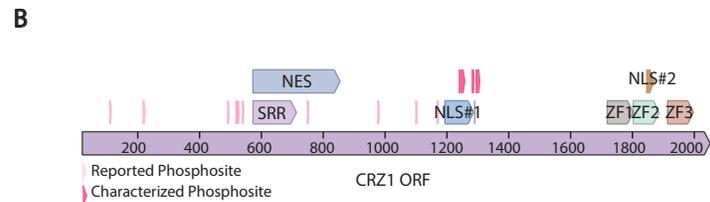
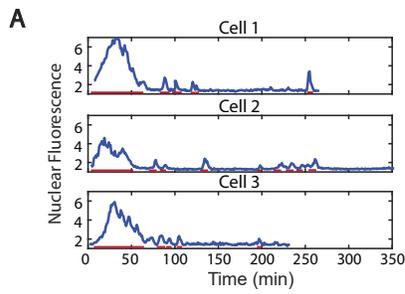


Figure 2.10: Characterization of Crz1, Crz1-CLASP, and Crz1*-CLASP nuclear translocation and gene expression with CaCl₂ or blue light input, related to Figure 2.3. A) Single cell traces of Crz1 nuclear fluorescence over time for 3 representative cells following 0.2M CaCl₂. The red lines indicate nuclear localization events. B) Schematic of the CRZ1 Open Reading Frame (ORF). Labeled are the Nuclear Localization Sequences (NLS1 and NLS2) and the Nuclear Exit Sequence (NES), as well as the Serine-Rich Region (SRR), which is calcium responsive. The light pink triangles denote reported S/T phosphosites, while the dark pink triangles denote reported and characterized S/T phosphosites. The 19 dark and light pink phosphosites are mutated from S/T -i A to construct Crz1*. Phosphosites were identified using PhosphoGRID. C) Heatmap of clustered gene expression for 5657 genes. Samples in each column of the heatmap are pADH1-Crz1 with no input, pADH1-Crz1-yeLANS with 60 minutes of light, pADH1-Crz1* with no input, and pADH1-Crz1-yeLANS with 0.2M CaCl₂ delivered at the start of the experiment. All samples are in log phase and all measurements are taken 60 minutes after delivery of input. D) Gene expression (mean FITC/SSC) of the Crz1 reporter gene pPUN1-YFP driven by either Crz1-CLASP (blue) or Crz1*-CLASP (pink) when given 30 minutes of blue light. Data plotted is for 1 biological replicate. E) Probability density functions of gene expression of pPUN1-YFP, measured by FITC/SSC, in response to 0.2M CaCl₂ (which causes an initial Crz1 nuclear localization pulse of 40-60 minutes) in a pADH1-Crz1 strain, 60 minutes of blue light exposure in Crz1*-CLASP, and no input in a pADH1-Crz1 strain. Measurements are taken at 4 hours after delivery of input. Data plotted is for 1 biological replicate. F) Basal gene expression of pPUN1-YFP for different Crz1 strains: endogenous Crz1, pAdh1-Crz1 in a Crz1 KO background, pAdh1-Crz1* in a Crz1 KO background, and pAdh1-Crz1*-CLASP (without light input) in a Crz1 KO background. Error bars show standard deviation of 3 biological replicates. G) OD600, a measurement for growth, plotted as a function of time, for pAdh1-Crz1 with (blue) and without (red) light input (intensity 512 a.u.) over a period of 24 hours, indicating that light exposure does not affect population growth. Measurements are taken every hour. Data plotted is for 3 technical replicates. H) Characterization of additional Crz1*-CLASP gene expression in response to blue light, as in Figure 2.3. Output-Fluorescence plot for pYPS1-YFP. I) Output-Fluorescence plot for pCMK2-YFP. J) Output-Fluorescence plot for pGYP7-YFP. These data are for 3 biological replicates taken from different days in addition to the data shown in Figure 2.3. For H-J, error bars are standard deviation of 3 biological replicates.

Figure S5

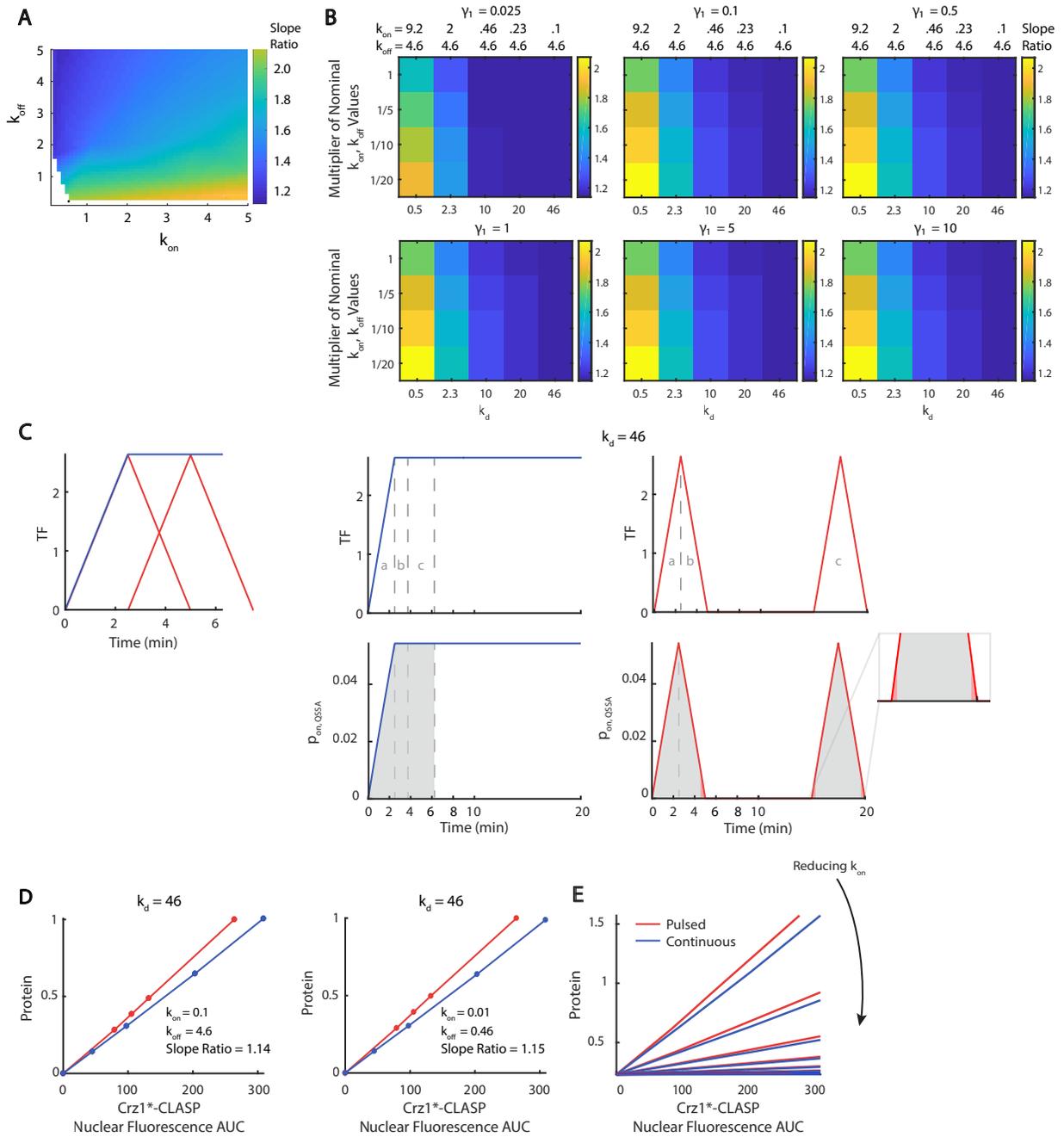


Figure 2.11: Higher gene expression of promoters in response to short pulses occurs when the dose response is saturated at low TF concentration, related to Figure 2.4. A) Heatmap of slope ratio resulting from the model in (Figure 2.4A) as a function of k_{on} and k_{off} , which both vary from 0.1-5. The ratio k_{off}/k_{on} decreases in a counterclockwise direction on the heatmap. β_1 varies from 0.0001-10, β_0 from 0.000001-0.01 and β_2 from 0.0001-10. The parameter γ_1 is set to 0.06 and γ_2 to 0.0083. B) Heatmap of slope ratio for increasing k_d and different values of k_{on} and k_{off} . Each column has a given value of k_d and each row has different values for k_{on} and k_{off} that produce the same k_d . The nominal k_{on} and k_{off} values used in the first row are noted at the top of each column, and every subsequent row uses a fraction of these values (1/5, 1/10, and 1/20). Each heatmap is compiled for a different value of γ_1 shown at the top of each panel. The values of β_1 , β_2 , and β_0 are 2.01, 4.92, and 0.0032, respectively. C) (Left panel) Pulsed and continuous TF inputs, of an equivalent area, used in Figure 2.4D and S5C are superimposed for comparison of their area. Red and blue lines represent pulsed and continuous inputs, respectively. (Middle and right panels) Plot of TF and p_{on} as a function of time for $k_d = 46$, assuming a fast promoter, therefore generating p_{on} , QSSA calculated as p_{on} , $QSSA = \frac{TF}{TF+k_d}$. (Top panels) Red and blue lines represent pulsed and continuous TF inputs, respectively. Gray lines and text denote equivalent area of TF input. The area labeled “a” represents the rise for both pulsed and continuous inputs. The area labeled “b” represents the fall of the pulsed input, and the equivalent area for the continuous input. The area labeled “c” represents a single pulse of the pulsed input, and the equivalent area of the continuous input. The area labeled “c” is equivalent to the sum of the areas labeled “a” and “b”. The areas labeled “a” and “b” are equivalent to each other. (Bottom panels) Red and blue lines represent p_{on} , QSSA in response to pulsed and continuous TF inputs, respectively. Gray shading denotes equivalent area of p_{on} , QSSA for continuous and pulsed inputs. Light red shading denotes excess p_{on} , QSSA area resulting from the rise and fall of the pulsed input. Inset shows the second p_{on} pulse at 200% resolution. D) (left panel) Output-Fluorescence plot as in Figure 2.11D, $k_{on} = 0.1$ and $k_{off} = 4.6$. (right panel) Output-Fluorescence plot where $k_{on} = 0.01$ and $k_{off} = 0.46$. E) Output-Fluorescence plot for a parameter set with higher gene expression in response to short pulses. Parameter values are: $k_{on} = 1$, $k_{off} = 0.8$, $\beta_1 = 0.0001$, $\beta_2 = 0.1$, $\gamma_1 = 0.05$, $\gamma_2 = 0.0083$, and $\beta_0 = 0.000001$. k_{on} is multiplied by 1/2, 1/4, 1/8, 1/16, and 1/32. The red line represents protein resulting from pulsed inputs and the blue line from continuous inputs.

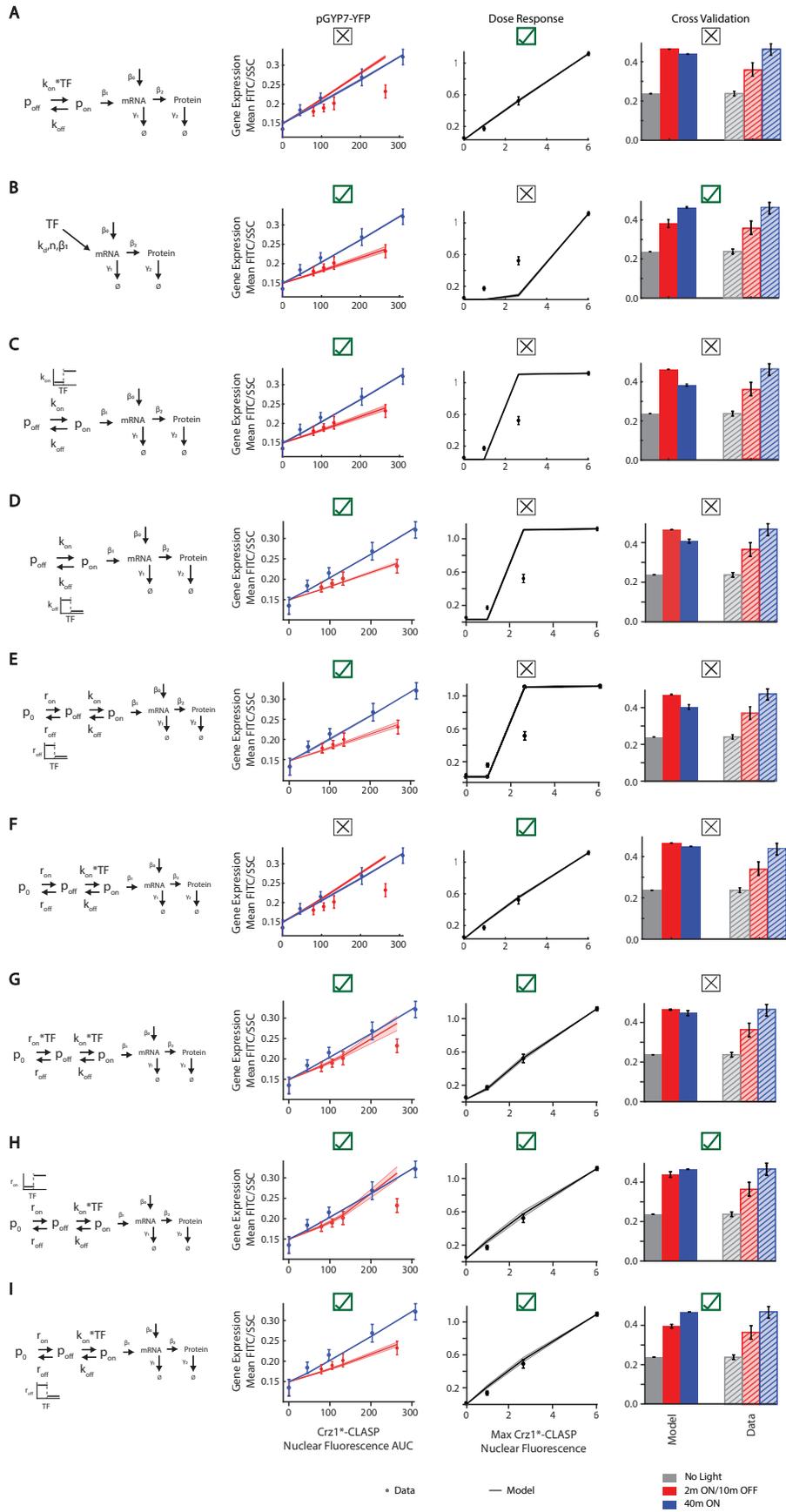


Figure 2.12: Exploration of various models for pGYP7-YFP data, related to Figure 2.6. A) (left panel) Schematic of the kinetic model, where the input is Crz1*-CLASP nuclear localization (TF) and the output is fluorescent protein level (Protein). (middle left panel) Output-Fluorescence plot for pGYP7-YFP. Circles are experimentally measured values for at least 3 biological replicates, error bars are standard deviation of those values, while lines denote the output of the model for 200 parameter sets out of 10000 that maximize fits through data points. The solid line denotes the mean and shaded areas the standard deviation of the model outputs. Parameters are sampled (k_{on} from 0.0001-1, k_{off} from 0.0001-1, β_1 from 0.0001-10, β_0 from 0.000001-0.01) or set ($\beta_2 = 0.06$, $\gamma_1 = 0.05$, $\gamma_2 = 0.0083$). (middle right panel) Dose response plot for pGYP7-YFP. The parameters that fit the Output-Fluorescence data are used to further fit the dose response of pGYP7-YFP using a best fit to least squared error criterion. Parameter sets below the mean of the least squared error distribution are plotted (solid black line is the mean generated by the model). The black dots are the experimentally measured dose response, and error bars represent standard deviation of at least 3 biological replicates. (right panel) The parameters that fit the Output-Fluorescence are then subject to cross-validation using an experiment where Crz1*-CLASP expression is increased (expressed from a pTEF1 promoter), and cells are exposed to either short pulsed (2 minutes ON/10 minutes OFF) or continuous input (40 minutes of light). The model generated outputs (solid red and blue bar) are plotted with the experimental data (hashed red and blue bar). The gray bars are samples not exposed to light. The error bars are the standard deviation of 3 biological replicates. B) (left panel) Schematic of a model with cooperativity. (middle left panel) Same plots as in (A, middle left panel), with 2481 parameter sets for this model. Parameters are sampled (k_d from 0.01-100, n from 0.5-4, β_1 from 0.0001-10, β_0 from 0.000001-0.01) or set ($\beta_2 = 0.06$, $\gamma_1 = 0.05$, $\gamma_2 = 0.0083$). (middle right panel, right panel) Plotted in the same manner as in (A, middle right panel, right panel) with 35 parameter sets. C) (left panel) Schematic of a 2-state model with thresholding on the activation constant, k_{on} . (middle left panel) Same plots as in (A, middle left panel), with 148 parameter sets for this model. Parameters are sampled (r_{on} from 0.1-100, r_{off} from 0.1-100, β_1 from 0.0001-10, β_0 from 0.000001-0.01) or set ($\beta_2 = 0.06$, $\gamma_1 = 0.05$, $\gamma_2 = 0.0083$, threshold = 0.5). (middle right panel, right panel) Plotted in the same manner as in (A, middle right panel, right panel) with 16 parameter sets. D) (left panel) Schematic of a two-state promoter model with a thresholded promoter inactivation constant, k_{off} . (middle left panel) Same plots as in (A, middle left panel), with 380 parameter sets for this model. Parameters are sampled (r_{on} from 0.0001-1, r_{off} from 0.0001-1, β_1 from 0.0001-10, β_0 from 0.000001-0.01, threshold from 0-2.7) or set ($\beta_2 = 0.06$, $\gamma_1 = 0.05$, $\gamma_2 = 0.0083$). (middle right panel, right panel) Plotted in the same manner as in (A, middle right panel, right panel) with 52 parameter sets.

Figure 2.12: E) (left panel) Schematic of a 3-state model with thresholding in the inactivation constant, r_{off} , between the promoter off-states, p_0 and p_{off} , and no TF dependence in the step before promoter activation. (middle left panel) Same plots as in (A, middle left panel), with 423 parameter sets for this model. Parameters are sampled (r_{on} from 0.1-100, r_{off} from 0.1-100, k_{on} from 0.0001-1, k_{off} from 0.0001-1, β_1 from 0.0001-10, β_0 from 0.000001-0.01, threshold from 0-2.7) or set ($\beta_2 = 0.06$, $\gamma_1 = 0.05$, $\gamma_2 = 0.0083$). (middle right panel, right panel) Plotted in the same manner as in (A, middle right panel, right panel) with 84 parameter sets. F) (left panel) Schematic of a 3-state model with constant rate of transition from p_0 to p_{off} . (middle left panel) Same plots as in (A, middle left panel), with 1288 parameter sets for this model. Parameters are sampled (r_{on} from 0.1-100, r_{off} from 0.1-100, k_{on} from 0.0001-1, k_{off} from 0.0001-1, β_1 from 0.0001-10, β_0 from 0.000001-0.01) or set ($\beta_2 = 0.06$, $\gamma_1 = 0.05$, $\gamma_2 = 0.0083$). (middle right panel, right panel) Plotted in the same manner as in (A, middle right panel, right panel) with 16 parameter sets. G) (left panel) Schematic of 3-state model with linear dependence on TF in both transitions from p_0 to p_{off} and p_{off} to p_{on} . (middle left panel) Same plots as in (A, middle left panel), with 1638 parameter sets for this model. Parameters are sampled (r_{on} from 0.1-100, r_{off} from 0.1-100, k_{on} from 0.0001-1, k_{off} from 0.0001-1, β_1 from 0.0001-10, β_0 from 0.000001-0.01) or set ($\beta_2 = 0.06$, $\gamma_1 = 0.05$, $\gamma_2 = 0.0083$). (middle right panel, right panel) Plotted, in the same manner as in (A, middle right panel, right panel) with 228 parameter sets. H) (left panel) Schematic of the 3-state model with thresholding in the activation constant, r_{on} , between promoter off-states, p_0 and p_{off} , and linear dependence on TF in transition from p_{off} to p_{on} . (middle left panel) Same plots as in (A, middle left panel), with 1649 parameter sets for this model. Parameters are sampled (r_{on} from 0.1-100, r_{off} from 0.1-100, k_{on} from 0.0001-1, k_{off} from 0.0001-1, β_1 from 0.0001-10, β_0 from 0.000001-0.01, threshold from 0-0.5) or set ($\beta_2 = 0.06$, $\gamma_1 = 0.05$, $\gamma_2 = 0.0083$). (middle right panel, right panel) Plotted in the same manner as in (A, middle right panel, right panel) with 455 parameter sets. I) (left panel) Schematic of the 3-state model with thresholding in the inactivation constant, r_{off} , between promoter off-states, p_0 and p_{off} and linear dependence on TF in transition from p_{off} to p_{on} . (middle left panel) Same plots as in (A, middle left panel), with 96 parameter sets for this model. Parameters are sampled (r_{on} from 0.1-100, r_{off} from 0.1-100, k_{on} from 0.0001-1, k_{off} from 0.0001-1, β_1 from 0.0001-10, β_0 from 0.000001-0.01, threshold from 0-0.5) or set ($\beta_2 = 0.06$, $\gamma_1 = 0.05$, $\gamma_2 = 0.0083$). (middle right panel, right panel) Plotted in the same manner as in (A, middle right panel, right panel) with 25 parameter sets.

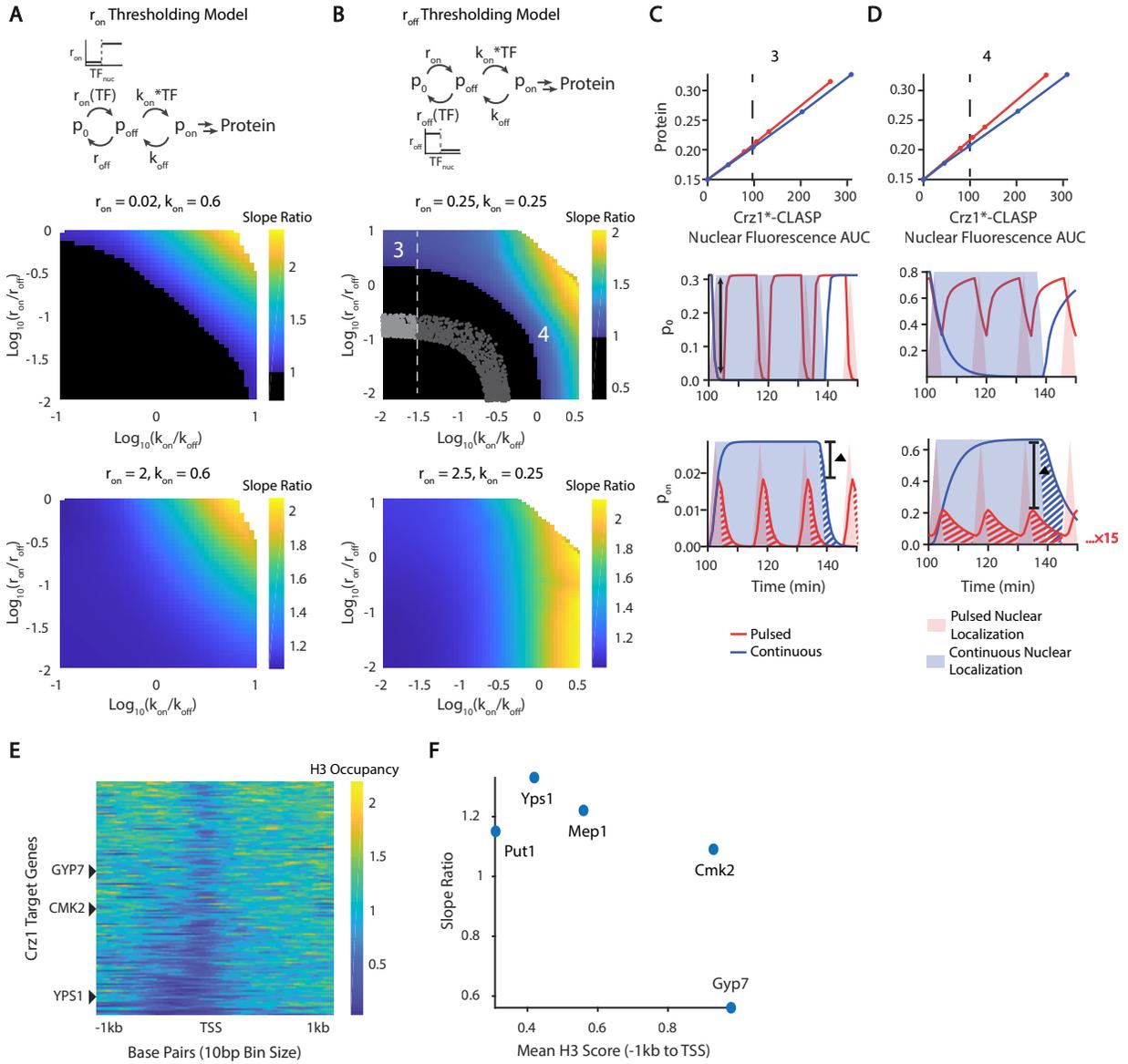


Figure 2.13: Exploration of three-state models for pGYP7-YFP data, related to Figure 2.6. A) Comparison of the 3-state models with either r_{on} or r_{on} thresholding in the transition from p_0 to p_{off} (upper panel) Schematic of the 3-state model with thresholding in the activation rate constant, r_{on} , between promoter off-states p_0 and p_{off} . (middle panel) Heatmap of slope ratio in the $\log_{10}(\frac{k_{on}}{k_{off}}) - \log_{10}(\frac{r_{on}}{r_{off}})$ plane. r_{on} is set to 0.02 and $k_{on} = 0.6$. Parameters are sampled (r_{on} from 0.0002-0.02, k_{off} from 0.002-0.2) or set ($\beta_1 = 0.0001$, $\beta_2 = 0.06$, $\gamma_1 = 0.05$, $\gamma_2 = 0.0083$, threshold = 0.5, $\beta_0 = 0.000001$). (lower panel) Same heatmap as in (A, middle panel) except with r_{on} set to 2, and r_{on} ranges from 0.02-2. B) (upper panel) Schematic of the 3-state model with thresholding in the inactivation constant, r_{on} , between promoter OFF-states, p_0 and p_{off} . (middle panel) Same heatmap as in (A, middle panel) with r_{on} is set to 0.25 and $k_{on} = 0.25$, that is previously described in Figure 2.6E. Parameters are sampled (r_{on} from 0.0025-2.5, k_{off} from 0.0025-0.25) or set ($\beta_1 = 0.0001$, $\beta_2 = 0.06$, $\gamma_1 = 0.05$, $\gamma_2 = 0.0083$, threshold = 0.5, $\beta_0 = 0.000001$). (lower panel) Same heatmap as in (B, middle panel) except with r_{on} set to 2.5, and r_{on} ranges from 0.025-25. C-D) Additional parameter requirements of the 3-state r_{on} threshold model for fitting pGYP7-YFP. (upper panels) Output-Fluorescence plots are generated by the model for different parameter sets that correspond to points 3 and 4 in the heatmap in B. The slope ratio for point 3 is 1.05 with $\log_{10}(\frac{k_{on}}{k_{off}}) = -1.58$ and $\log_{10}(\frac{r_{on}}{r_{off}}) = 0.6$. The slope ratio for point 4 is 1.25 with $\log_{10}(\frac{k_{on}}{k_{off}}) = 0.1$ and $\log_{10}(\frac{r_{on}}{r_{off}}) = -0.89$. Point 3 is chosen to highlight the effect of decreasing r_{on} , while Point 4 is chosen to highlight the effect of decreasing k_{off} . (middle panels) Example of a time course of promoter state p_0 for a light input that produces the equivalent of 40 minutes (dotted line in upper panel) in nuclear localization either continuously or in short pulses. Solid lines are the p_0 activity while shading denotes TF nuclear localization. The black double arrow denotes the maximum depletion of the p_0 state for the pulsed input. (lower panels) Example of a time course of promoter activity p_{on} for a light input that produces the equivalent of 40 minutes (dotted line in upper panel) in nuclear localization either continuously or in short pulses, similar to the (middle panels). The red and blue hashing represents residual promoter activity beyond the nuclear localization input. The red residual promoter activity is repeated 15 times while the blue residual activity is repeated one time. The Δ bar denotes the difference between the amplitudes generated by the 2 minute pulsed and 40 minute continuous input. E) Correlation of nucleosome occupancy and sensitivity to pulsing. Heatmap of H3 occupancy for the Crz1 target genes as specified by Yoshimoto 2002. H3 occupancy is defined as counts of H3 enrichment over the IgG antibody, which is a control for no pull down of histones. The dataset and determination of start sites are obtained from Sen et al., 2015 and Malabat et al., 2015, respectively^{78,79}. The software deepTools 2.0 is used to compute the H3 occupancy values. -1 and +1 kb from the transcription start site (TSS) is used. The positions of YPS1, CMK2, and GYP7 in the heatmap are denoted with black triangles. F) Slope ratios of Crz1 target genes as a function of their mean H3 nucleosome occupancy scores averaged from -1kb to the Transcription Start Site (TSS). The correlation coefficient is $r^2 = 0.26$.

Figure S8

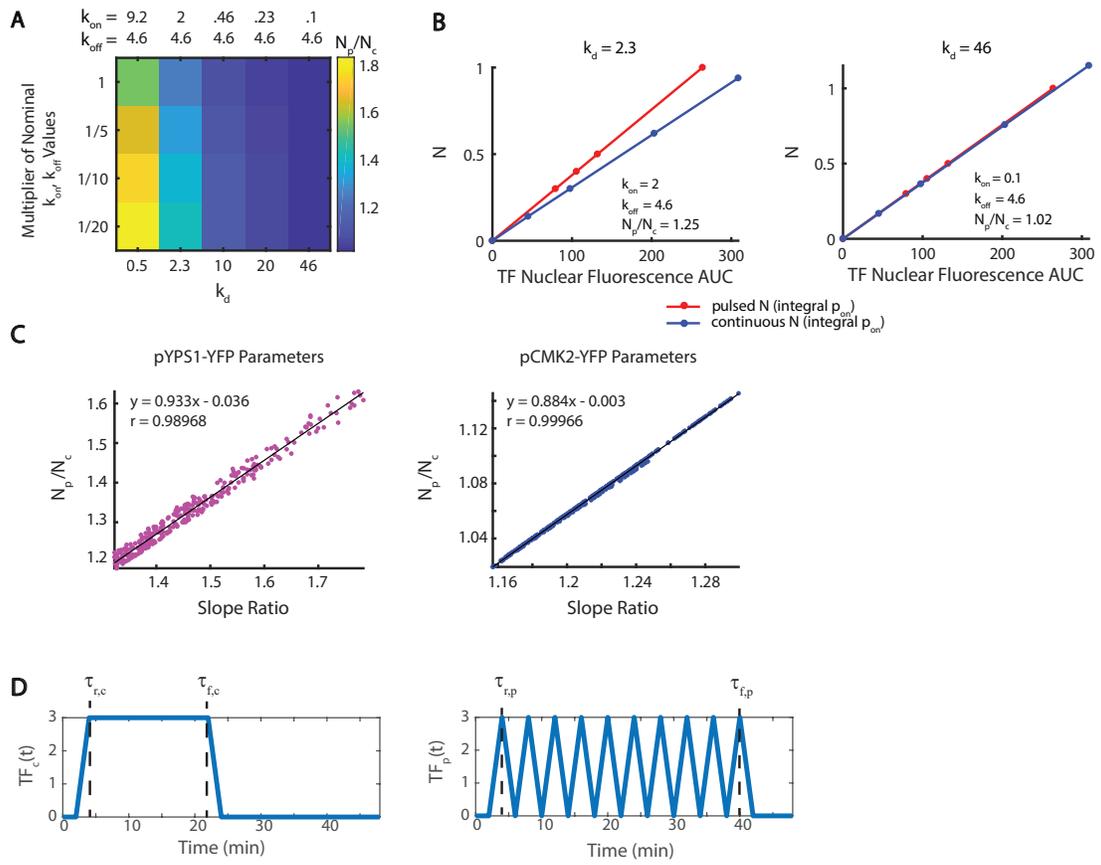


Figure 2.14: Slope ratio N_p/N_c of number of transcripts N is strongly correlated with protein slope ratio, related to Figures 4-5 and 2.7.1 Mathematical Analysis.

A) Heatmap of N_p/N_c for increasing k_d and different values of k_{on} and k_{off} . Each column has a given value of k_d and each row has different values for k_{on} and k_{off} that produce the same k_d . The nominal k_{on} and k_{off} values used in the first row are noted at the top, and every subsequent row uses a fraction of these values (1/5, 1/10, and 1/20). B) Plot of N , total transcripts produced, as a function of TF nuclear fluorescence AUC. Left Panel: $k_{on} = 2$, $k_{off} = 4.6$, $k_d = 2.3$ and $\frac{N_p}{N_c} = 1.25$. Right Panel: $k_{on} = 0.1$, $k_{off} = 4.6$, $k_d = 46$ and $\frac{N_p}{N_c} = 1.02$. For both panels, $\beta_1 = 2$, $\gamma_1 = 0.06$, $\beta_2 = 4.92$, $\gamma_2 = 0.0083$, $\beta_0 = 0.0032$. C) Plot of slope ratio $\frac{N_p}{N_c}$ of transcripts versus protein slope ratio. N_p are the transcripts produced by $TF_p(t)$ while N_c are those produced by $TF_c(t)$. Each dot represents values computed using the model for a parameter set that fit the protein timecourse and Output-Fluorescence data. Left Panel: data for pYPS1-YFP (also plotted in Figure 2.5C). Right Panel: Same plot as left panel for pCMK2-YFP (also plotted in Figure 2.5C). D) Time-dependent transcription factor concentration $TF(t)$ for the continuous ($TF_c(t)$, left plot) and pulsed ($TF_p(t)$, right plot) cases. Here $\int TF_c(t) dt = \int TF_p(t) dt$ and both have the same initial rise (shown between 0 and $\tau_{r,c}$ (left graph) and 0 and $\tau_{r,p}$ (right graph)) and final shutoff behavior (shown between $\tau_{f,c}$ and end of input (left graph) and $\tau_{f,p}$ and end of input (right graph)).

Figure S9

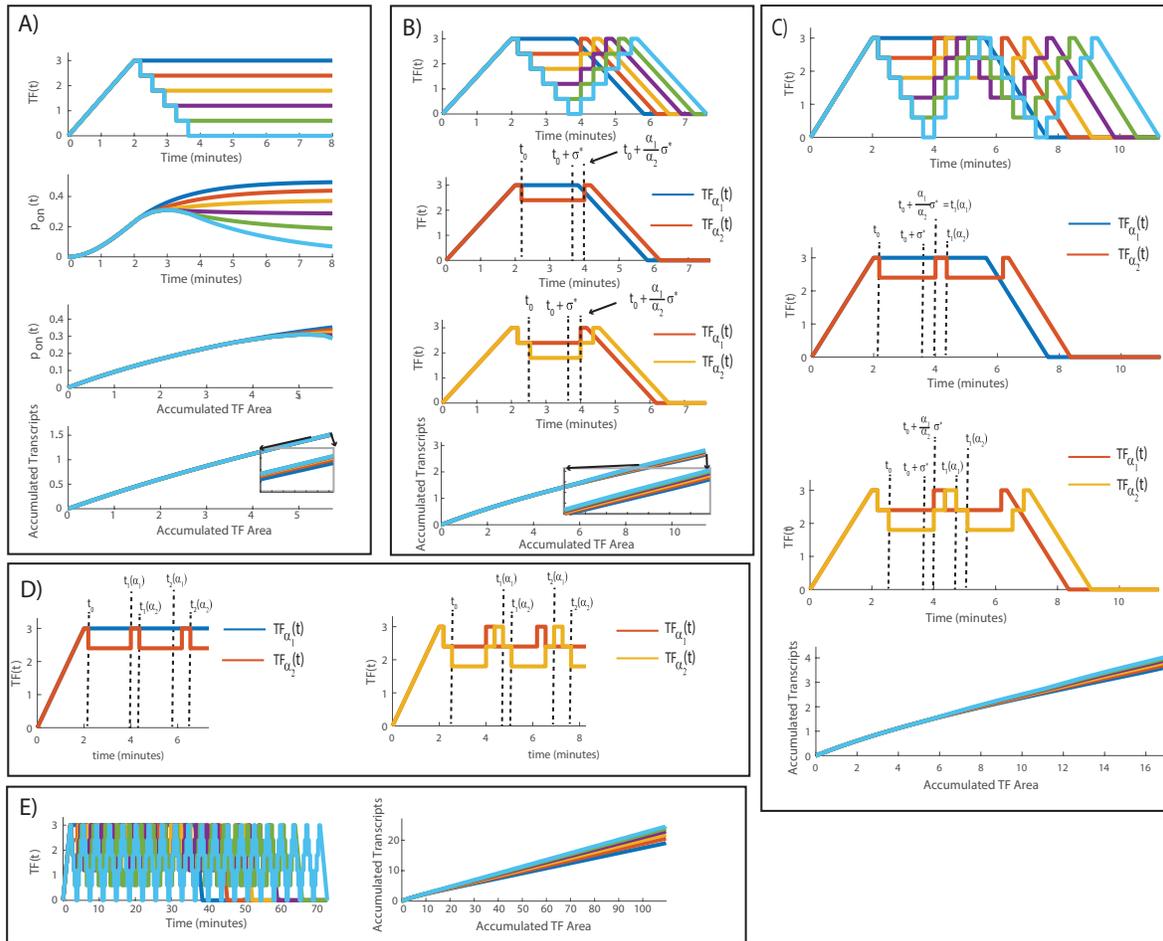


Figure 2.15: Simulations of input sequences that accompany theoretical analysis, related to Figure 2.4 and 2.7.1 Mathematical Analysis. For all simulations $I=6$ (I is the number of inputs), $k_d=3$, $k_{\text{off}}=.33$, $k_{\text{on}}=.11$, and $\beta_1=1$. A) Sequence of inputs used and simulation results for them. (Top panel) Sequence of TF inputs with progressively lower plateaus at successive times. (Top Middle panel) Plots of $p_{\text{on}}(t)$ for the different inputs. (Bottom Middle panel) Plots of p_{on} versus cumulative TF area $\int_0^t TF(v) dv$ for the different inputs. (Bottom plot) Cumulative transcripts ($\frac{\beta_1}{k_d} \int_0^t (1-p_{\text{on}}(v))TF(v) dv$) versus cumulative TF area. B) Simulation results of cumulative transcripts for a class of inputs that have the same total accumulated TF area. (Top panel) Sequence of TF inputs with progressively lower plateaus at successive times that later rise (step up) to the level of $TF_c(t)$ shown in dark blue, and then shut off. Light blue line is $TF_p(t)$ with 2 pulses. (Top middle panel) First adjacent input pairs with dark blue being $TF_{\alpha_1}(t)$ and red being $TF_{\alpha_2}(t)$. The plot shows the time of divergence $t=t_0$ for the first two inputs (where $TF_{\alpha_2}(t)$ first drops to its plateau), the time $t = t_0 + \frac{\alpha_1}{\alpha_2}\sigma^*$ when $TF_{\alpha_2}(t)$ jumps back up to $TF_{\alpha_1}(t)$, and the time $t = t_0 + \sigma^*$ where $TF_{\alpha_1}(t)$ has the same corresponding accumulated TF area as $TF_{\alpha_2}(t)$ does at $t = t_0 + \frac{\alpha_1}{\alpha_2}\sigma^*$. (Bottom middle panel) Second adjacent input pairs (red is $TF_{\alpha_1}(t)$ and orange is $TF_{\alpha_2}(t)$) with similarly marked time points as in panel above it. (Bottom panel) Cumulative transcripts versus cumulative TF area for the inputs in the top panel. C) Extending the class of inputs with equivalent total accumulated TF area from two (panel B) to 3 pulses. (Top panel): Sequence of TF inputs with progressively lower plateaus at successive times that later rise (step up) to the level of $TF_c(t)$ shown in dark blue. Same as in panel B but for 3 pulses. (Top Middle Panel) First adjacent input pairs with dark blue being $TF_{\alpha_1}(t)$ and red being $TF_{\alpha_2}(t)$. There are two additional time labels relative to those in (B). For the $TF_{\alpha_2}(t)$ input, $t = t_1(\alpha_2)$ is the time at which $TF_{\alpha_2}(t)$ drops a second time to its plateau. For the $TF_{\alpha_1}(t)$ input, $t = t_1(\alpha_1)$ is the time at which $TF_{\alpha_1}(t)$ has the same TF area as $TF_{\alpha_2}(t)$ does at $t = t_1(\alpha_2)$. (Bottom middle panel) Second adjacent input pairs (red is $TF_{\alpha_1}(t)$ and orange is $TF_{\alpha_2}(t)$) with similarly marked time points as in panel above it. (Bottom panel): Cumulative transcripts versus cumulative TF(t) area for 3 pulse sequence of inputs shown in top panel. D) Extending the class of inputs to M pulses. Plots showing time locations when $TF_{\alpha_2}(t)$ drops to its first, second, and third plateaus for the first adjacent input pair (left plot) and second adjacent input pair (right plot). The times $t = t_0$, $t = t_1(\alpha_2)$ and $t = t_2(\alpha_2)$ are the ordered locations when $TF_{\alpha_2}(t)$ drops to the corresponding plateau. For the $TF_{\alpha_1}(t)$ input, the times $t = t_1(\alpha_1)$ and $t = t_2(\alpha_1)$ map to the same TF area for $TF_{\alpha_1}(t)$ as $t = t_1(\alpha_2)$ and $t = t_2(\alpha_2)$ does for $TF_{\alpha_2}(t)$. E) Simulation results for the 20 pulse case. Left plot: All inputs. Dark blue corresponds to the continuous input $TF_c(t)$ of equal area to the other inputs. Light blue corresponds to the pulsed input $TF_p(t)$ (20 pulses). Right plot: Cumulative transcripts versus cumulative TF(t) area.

Table 2.1: NLS Optimization This table contains the amino acid and peptide sequences used in the screen for NLSs in the optimization of CLASP for maximal nuclear localization dynamic range.

Number	NLS	Score (Kosugi 2009)	AA Length	Class
0	paaKRvKld	na	9	original LANS, class2
3	raaKRpRtt	10	9	class2
5	paaKRpRtt	9	9	class2
8	apaKRaRtt	8	9	class2
9	paaKRlCtt	9	9	class2
11	aaaKRswsmaf	10	11	class3
14	aaaKRswvmf	9	11	class3
15	aaaKRswsaaf	10	11	class3
20	KRpatlandspaaKRR	9	16	bipartite
24	KRKRwendip	na	10	class1
27	psRKRKRdhyav	na	12	class1
29	tspsRKRKwdqv	na	12	class1

Table 2.2: OptoPlate Transfer Function The transfer function of the OptoPlate. Conversion from arbitrary units (a.u.) to intensity (milliWatts).

Light Dose	milliWattage	Std Dev
0	-0.0001499	0.0002004
0	-0.0002513	0.000252
0	-0.001819	0.000978
64	0.1011	0.008522
64	0.07921	0.007053
64	0.1279	0.01123
128	0.1944	0.02126
128	0.1962	0.01824
128	0.2633	0.02653
256	0.3961	0.0393
256	0.4012	0.0362
256	0.5107	0.04832
512	0.7849	0.07396
512	0.7776	0.09084
512	1.041	0.09508
1024	1.53	0.1346
1024	1.477	0.1491
1024	1.986	0.2349
2048	3.088	0.3326
2048	3.026	0.2563
2048	3.45	0.9
3072	4.302	0.4414
3072	4.235	0.4559
4095	5.407	0.5494
4095	5.486	0.4617
4095	6.56	0.587

Chapter 3

Optogenetic control of RelA underscores the importance of transcription factor dynamics in downstream gene expression

3.1 Summary

This study extends the results of the previous study by using CLASP in mammalian cells. In this study, optogenetics is used to control the translocation of RelA, as well as other transcription factors, into the nucleus, and transcriptomics is used to assess the genome-wide effects of optogenetically-controlled translocation.

3.2 Introduction

Transcription factors are critical intracellular messengers that receive information from upstream nodes, like kinases and receptors, and then pass that information to genes by causing their activation and repression. As discussed in Chapter 2, TFs transmit information to downstream genes in many ways, including post-translational modifications and changes in concentration^{80–82}. Much like in *S. cerevisiae*, a subset of TFs, including NFAT and NF- κ B, additionally regulate their spatiotemporal dynamics in response to environmental inputs^{26,83}. Additionally, TFs like p53 modulate their nuclear concentration with different dynamics²³.

Previous studies have used a variety of inputs, including chemical and optogenetic inputs, to regulate temporal dynamics of transcriptional regulators in order to elucidate the effects of these dynamics on downstream gene expression, which they have measured using techniques such as RNA-seq and reporter genes^{9,13–15,19,84–86}. For example, a study used gamma irradiation to induce pulses of p53, and then measured expression changes across the genome over a period of 12 hours using RNA-seq. By combining these data with ChIP-seq data, the authors were able to show that using a single dynamic input, p53 can activate downstream genes with different temporal patterns due to differences in binding kinetics¹⁹. Another study regulated NFAT dynamics through optogenetic control of calcium concentration. With this method, researchers found that a synthetic reporter did not differentiate between NFAT nuclear translocation dynamics, but instead activated in proportion to the integral of the NFAT nuclear localization¹³. Finally, a recent study used microfluidics to precisely control the concentration of TNF α delivered to cells, which then regulated the amplitude of NF- κ B pulses of nuclear localization. As a result, the authors were able to demonstrate that downstream genes had differential responses to pulsed inputs of different amplitudes⁸⁴.

A commonality across these studies, though, is that they all modulate upstream regulators, which, in turn, modulate many downstream effectors, to study dynamics of a single TF. As a result, many other transcriptional regulators are activated in addition to the TF of interest, and these pleiotropic effects confound conclusions about the effect of dynamics of the TF of interest. As an example, a recent study elucidated the effects of two stimuli, LPS and TNF, on NF- κ B downstream signaling by knocking out *Ifnar*, an upstream regulator induced by LPS, and *Nfkb1a*, a protein which regulates NF- κ B activation through a feedback loop. Using these genetic modifications, researchers were able to create similar dynamics for the two inputs, thereby allowing them to model stimulus-specific effects. In this study, researchers found hundreds of genes downstream of NF- κ B which were differentially regulated by these stimuli as opposed to the dynamics of NF- κ B itself³⁷. It is important to note that many genes in mammalian cells are regulated by multiple TFs which can bind each gene separately or as a complex. As a result, modulation of multiple transcriptional regulators at the same time may be necessary to induce downstream gene expression changes. Still, fine-tuned control over a single TF can allow researchers to begin to untangle the effects of dynamic perturbations on gene activation.

To elucidate the effects of TF spatiotemporal dynamics on downstream gene expression in mammalian cells, we have modified the yeast CLASP system for use across multiple cultured cell lines. With mammalian CLASP, we demonstrate optogenetic control of nuclear translocation for several TFs, including NFAT1 and RelA. We focus on RelA-CLASP, and modulate its dynamics by inducing cells with pulsed and constant light inputs, thereby mimicking the effects of different stimuli. Following induction, we measure expression of downstream genes using RNA-seq. Using a model of promoter activation and gene expression, we glean high-level information about the complexity of gene activation downstream of RelA. In summary, this study demonstrates the effect of differential dynamics of a single TF on downstream genes, and provides a framework for future studies on the same topic.

3.3 Results

3.3.1 Designing Mammalian CLASP as a modular optogenetic tool

In Chapter 2, we showed the utility of CLASP, a blue-light-responsive optogenetic tool, for regulating nuclear translocation of various transcription factors in yeast. To extend these results to mammalian cells, we first needed to engineer a CLASP DNA construct that would express readily in cultured cells, and would work across cell lines. As discussed in Chapter 2, optimization of LANS was necessary to achieve nuclear translocation to the scale seen with transcription factors activated by environmental stress inputs⁴⁶. This optimization yielded yeLANS, a LANS with a stronger nuclear localization sequence (NLS) than that which was originally published. Furthermore, the NLS in yeLANS was a sequence that was predicted to work in mammalian cells in addition to yeast⁵⁰.

Accordingly, we found that nuclear translocation was robust in HEK293T cells using the yeLANS sequence (Figure 3.5B). However, the pm-LOVTRAP was poorly targeted to the plasma membrane in HEK293T cells, despite the plasma membrane targeting sequence being derived from RGS2, a protein expressed in *Homo sapiens*⁵¹. Accordingly, we swapped the Hs_RGS2 plasma membrane tag for a tag derived from Lyn kinase, which led to improved plasma membrane targeting for the LOVTRAP and sequestration for the Zdk1-cargo-yeLANS construct⁸⁷ (Figure 3.5A).

We next proceeded to optimize the method of delivery of CLASP DNA constructs. In our first iteration, we delivered the entire construct on one plasmid, using PiggyBac transposase to integrate into the genome and selected cells using an antibiotic selectable marker⁸⁸. This method had several disadvantages: firstly, changing the cargo of interest required completely rebuilding the plasmid construct, and secondly, such a large

integration was prone to silencing and loss of expression of parts of the circuit. As a result, many cells would only express part of the construct that was delivered (Figure 3.6A). To address both issues, we created a three plasmid system to deliver the CLASP system to mammalian cells. The CLASP system is divided into the three plasmids as follows: a plasmid expressing a fluorescent nuclear marker, a plasmid expressing the fluorescently-tagged pm-LOVTRAP system, and a plasmid expressing the fluorescently-tagged Zdk1-cargo-yeLANS protein (Figure 3.6B). These plasmids are delivered through lentiviral integration and can be transduced simultaneously or in series, depending on the goal of the user. Lentiviral transduction was preferred to PiggyBac transposase for its superior efficiency; multiple plasmids were used because this would reduce the size of each integration, thereby increasing the likelihood of continued expression over time, and also assisting with efficient lentiviral packaging.

A critical benefit of separating the plasmids is that it allows us to create a system which was more amenable to rapid screening of different cargo constructs across multiple cell lines. Using only the nuclear marker and pm-LOVTRAP plasmids, we were able to transduce and create "chassis" cell lines in HEK293T, 3T3, and MCF10A cells which could be used to theoretically screen any Zdk1-cargo-yeLANS construct of interest (Figure 3.6C). These chassis cell lines can be used with transient transfection to test whether a given cargo can be effectively sequestered and translocated by CLASP. With this method, we show that a variety of cargos, including mScarlet, p53, NFAT1, and RelA could be translocated to and from the nucleus in various cell lines with CLASP (Figures 3.5B, 3.6D-E, 3.1B). Overall, these data indicate that CLASP can be used in mammalian cells in a modular fashion to regulate nuclear translocation of multiple cargo proteins across multiple cell lines.

3.3.2 Quantification of RelA-CLASP response to light reveals reversible and dose-responsive dynamics

For the remaining studies in this chapter, we will focus on one specific optogenetic construct, RelA-CLASP. RelA-CLASP is expressed in 3T34KO cells. These cells are mouse fibroblasts which do not express RelA and three Inhibitor of nuclear kappa light polypeptide gene enhancer in B-cells (I κ B) proteins: I κ B α , I κ B β , and I κ B ϵ . In wildtype cells, the I κ B proteins sequester RelA in the cytoplasm in the absence of stimulus. When an environmental stimulus, such as Tumor Necrosis Factor α (TNF α), is added to the cells, an upstream regulator of the I κ B proteins, I κ B kinase (IKK) phosphorylates the I κ B proteins, which leads to their dissociation from RelA and subsequent degradation (Figure 3.1A)²⁶. This cell line creates an ideal chassis for testing the ability of CLASP to regulate nuclear translocation of RelA, as the primary regulators of RelA translocation are non-functional. Additionally, several mutations were made to the sequence of RelA used in CLASP to improve likelihood of inducing downstream gene expression⁴.

Using confocal microscopy, we show that translocation from the plasma membrane to nucleus occurs rapidly; within 5 minutes of stimulation with blue light, nuclear translocation of Zdk1-RelA-mScarlet-yeLANS is evident (Figure 3.1B). Quantification of the response to blue light stimulation shows that nuclear translocation reaches its maximum rapidly, in less than 10 minutes. Additionally, nuclear translocation is reversible, and exit occurs within 10-20 minutes after cessation of blue light (Figure 3.1C). Furthermore, RelA-CLASP can be pulsed into the nucleus repeatedly, with similar increase in nuclear/cytoplasmic enrichment in each pulse, when given light inputs separated by at least 20 minutes (Figure 3.1D).

We undertook additional experiments to more specifically quantify how responsive RelA-CLASP is to inputs of different lengths and intensities. First, we vary the time of

blue light induction from 1-10 minutes, to measure exactly how sensitive RelA-CLASP translocation is to the width of the light pulse. We find that the nuclear/cytoplasmic enrichment of the cell population increases with as little as 1 minute of blue light. Furthermore, the population reaches its maximum nuclear/cytoplasmic enrichment within 8 minutes (Figures 3.1E, 3.7A). Next, we vary the intensity of blue light given to the cells over 15 minutes of induction and measure the maximum nuclear/cytoplasmic enrichment achieved across the population of cells. Importantly, we find that RelA-CLASP responds almost linearly to different intensities of light within the range sampled (0-3000 a.u., induced with the Optoplate-96) (Figures 3.1F, 3.7B)^{73,89}. Taken together, these data show that RelA-CLASP can quickly (≤ 8 minutes) and reversibly translocate to the nucleus, and that the magnitude of translocation can be tuned with light intensity.

3.3.3 RelA-CLASP is not modulated by environmental stress inputs

We have demonstrated remarkably fine control over RelA-CLASP using blue light inputs. However, the endogenous regulation of RelA leads to its activation in response to many environmental inputs. $\text{TNF}\alpha$, Interleukin-1 β (IL-1 β), and lipopolysaccharide (LPS) are three environmental stimuli known to activate RelA. Each of these inputs is recognized by a different extracellular receptor: $\text{TNF}\alpha$ is recognized by the Tumor Necrosis Factor Receptor (TNFR); IL-1 β activates the Interleukin-1 Receptor (IL1R); LPS binds the Toll Like Receptor 4 (TLR4)²² (Figure 3.2A, top panel). RelA regulates many cellular pathways, such as survival and inflammation, and is canonically studied for its response to immune system stimuli. $\text{TNF}\alpha$ and IL-1 β are cytokines produced in the immune response, and LPS is an endotoxin produced by gram-negative bacteria which causes an immune response. All three inputs converge on RelA through similar, yet distinct pathways.

As a control for RelA-CLASP, we used lentiviral transduction to express a TagBFP nuclear marker and a wildtype *Mus musculus* RelA fused to mScarlet in NIH3T3 background cells. Unlike the RelA-CLASP cells, the NIH3T3 cells expressing the RelA-mScarlet construct had no mutations to their endogenous RelA or I κ B sequences, and therefore had full endogenous control over nuclear translocation of RelA.

Prior work has shown that TNF α binding of the TNFR yields formation of signaling complexes that activate IKK, as well as other kinases like c-Jun N-terminal Kinase (JNK)^{22,24}. Given the intact IKK-I κ B pathway in RelA-mScarlet cells, RelA is rapidly activated and translocated to the nucleus upon stimulation with TNF α in this cell line (Figure 3.2B, green trace, top panel). This stimulation is specific to TNF α ; addition of PBS + .1% BSA vehicle to RelA-mScarlet cells yields no such response (Figure 3.2B, black trace, top panel). Approximately 10 minutes after induction with 1 ng/mL TNF α , RelA translocates into the nucleus in all cells. Nuclear/cytoplasmic enrichment across the population peaks approximately 30 minutes after induction and resolves within 100-120 minutes after induction. Previous studies have shown that TNF α induction leads to a coordinated first wave of translocation into the nucleus, followed by additional damped pulses into the nucleus^{8,30,84,90}. Individual cell traces show a similar finding, with some cells undergoing a short, single pulse within the first hour after induction, while other cells display two or more pulses of RelA translocation after induction (Figure 3.2B, inset, top panel). By contrast, no nuclear translocation is seen for RelA-CLASP in response to induction with 1 ng/mL TNF α (Figure 3.2B, green trace, bottom panel). At a population level, the nuclear/cytoplasmic enrichment observed is similar to that seen with vehicle induction (Figure 3.2B, black trace, bottom panel). Individual cell traces confirm that within the population, RelA-CLASP is not induced by this concentration of TNF α (Figure 3.2B, inset, bottom panel).

Il-1 β activates the Il1R, which then signals through MyD88, a pathway that is also

activated in TLR4 signaling. MyD88 activation leads to activation of IKK, as well as other regulators like IL-1 receptor-associated kinase 1 (IRAK1) and JNK²². Accordingly, RelA-mScarlet cells induced with 1 ng/mL of Il-1 β respond with a coordinated translocation to the nucleus (Figure 3.2C, orange trace, top panel). As seen in prior work, the translocation induced by Il-1 β resolves more rapidly than that induced by TNF α ; within 80 minutes after induction, the median change in nuclear/cytoplasmic enrichment has decreased to 0³⁷. Additionally, the RelA translocation response is fully concluded after this single pulse –for the remaining 100 minutes, no additional change in nuclear/cytoplasmic enrichment is seen for RelA-mScarlet cells induced with 1 ng/mL Il-1 β . Individual cell traces also confirm that the response to Il-1 β is highly uniform across cells (Figure 3.2C, inset, top panel). Importantly, no such translocation response is seen in RelA-CLASP cells treated with 1 ng/mL Il-1 β (Figure 3.2C, orange trace, bottom panel). The observed change in nuclear/cytoplasmic enrichment for RelA-CLASP cells induced with 1 ng/mL is similar to that seen in RelA-CLASP cells induced with a vehicle control (Figure 3.2C, black trace, bottom panel). This lack of response to Il-1 β is also evident in individual cell traces (Figure 3.2C, inset, bottom panel).

LPS is one of many peptides which bind the TLR4. After binding, TLR4 activates the MyD88 pathway as well as the TRIF pathway, which leads to activation of interferon signaling^{22,24}. Unlike the response to TNF α and Il-1 β , RelA translocation in RelA-mScarlet cells in response to LPS is more delayed. Translocation does not begin until approximately 30 minutes after induction with 100 ng/mL LPS, and does not peak until 60 minutes after induction (Figure 3.2D, pink trace, top panel). The population response to LPS also extends for nearly two hours, which is similar to the response to TNF α and much longer than the response to Il-1 β . However, individual cell traces demonstrate that this prolonged translocation is uniform across cells, a key difference between the response to LPS and TNF α (Figure 3.2D, inset, top panel). Remarkably, despite the strong, coordinated, and prolonged response seen in RelA-mScarlet cells, translocation of RelA-CLASP in response

to LPS is not distinct from translocation of RelA-CLASP after addition of vehicle (Figure 3.2D, bottom panel). The lack of RelA-CLASP response is also shown in individual traces from RelA-CLASP cells induced with 100 ng/mL LPS (Figure 3.2D, inset, bottom panel).

As described in Chapter 2, blue light stimulation activates RelA-CLASP by opening the LOV2 domain which sequesters the Zdk1-RelA-yeLANS at the plasma membrane, and also by opening the yeLANS to reveal the NLS. The activation of both LOV2 domains leads to nuclear translocation of Zdk1-RelA-yeLANS (Figure 3.2E, bottom panel). As shown in Figure 3.1, RelA-CLASP translocates to the nucleus rapidly upon induction with blue light. When given a 2 hour light input, RelA-CLASP maintains nuclear localization throughout the duration of the input, and then exits the nucleus shortly after the light input is turned off (Figure 3.2F, blue trace, bottom panel). Individual cell traces show that within the population of RelA-CLASP cells, nuclear translocation and nuclear exit are closely timed with the blue light input. For RelA-mScarlet cells, blue light does not lead to appreciable nuclear translocation over the no light control. This is confirmed by the individual cell traces, which show uniform lack of induction of RelA-mScarlet in response to blue light input of 1350 a.u.

These data quantify both the canonical RelA response to various environmental stress inputs and the unique responsiveness of RelA-CLASP to blue light as an input. The lack of RelA-CLASP response to LPS, $\text{TNF}\alpha$, and $\text{Il-1}\beta$ inputs reflects the robust knockout of the I κ B proteins, as well as the sequestration ability of CLASP.

3.3.4 RelA-CLASP activates canonical downstream genes given a constant light input

While important, nuclear translocation of RelA does not guarantee activation of downstream genes in the known RelA regulon. Though nuclear translocation of RelA differs

in response to different environmental inputs (Figure 3.2), so too do the pathways and regulators activated by these inputs. Previous studies have shown that post-translational modifications of RelA, such as phosphorylation, are critical for downstream gene expression⁹¹. Additionally, prior work has shown that activation of co-regulators, like JNK, CBP:p300, I κ B β , and many others, boosts the ability of RelA to activate genes^{4,92,93}.

Prior to this work, no publication has demonstrated precise, reversible control over RelA nuclear translocation without environmental stress inputs. As a result, it was previously impossible to directly demonstrate whether nuclear translocation of RelA alone was sufficient to induce downstream gene activation, without concurrent phosphorylation of critical residues, such as S536, or activation of important co-regulators. RelA-CLASP represents a novel opportunity to interrogate solely the effect of nuclear translocation of RelA on its downstream genes.

To probe this question, we performed RNA-seq on RelA-CLASP cells induced with constant blue light at 460 a.u. (Figure 3.3A). This intensity of light was chosen due to its ability to activate RelA-CLASP nuclear translocation while potentially minimizing the effect of blue light toxicity. Samples were measured at 0h, 1h, and 2h of blue light induction. Additionally, as a control, pm-LOVTRAP cells were induced with the same light input. The gene expression changes seen in pm-LOVTRAP were then used to benchmark the changes seen in RelA-CLASP cells at the same time points. By subtracting the log₂ fold change seen in pm-LOVTRAP from that seen in RelA-CLASP cells, we were able to estimate the induction caused by RelA-CLASP alone as opposed to that caused by blue light. We termed this metric log₂FC RelA-CLASP induction.

Using this method, we first found that several canonical genes downstream of RelA were upregulated in RelA-CLASP cells in response to blue light (Figure 3.3B). Trim30 α is upregulated in response to LPS and other TLR agonists and its activation is NF- κ B-dependent⁹⁴. Nfkbiz produces I κ B ζ , another inhibitor of RelA nuclear

translocation; activation of I κ B proteins, including I κ B ζ , acts as a feedback loop on RelA activation⁹⁵. Ccl2 is a chemokine induced in response to many inflammatory inputs, and Tnfrsf11b is a secreted protein which binds TNF-related apoptosis inducing ligand (TRAIL) and Receptor Activator of NF- κ B Ligand (RANKL)^{96,97}. These data represent the first direct demonstration that nuclear translocation of RelA alone can induce downstream genes.

Many previous studies have used environmental inputs to regulate the nuclear translocation of RelA and quantify the effect of differential translocation on downstream genes^{8,9,84,85}. However, these inputs can have pleiotropic effects, activating many other pathways and regulators in addition to modulating RelA translocation. To quantify these pleiotropic effects, we delivered 1 ng/mL TNF α to pm-LOVTRAP cells, which do not express RelA, I κ B α , I κ B β , and I κ B ϵ , and measured gene expression through RNA-seq at 0h, 1h, and 2h of TNF α induction (Figure 3.3C). Interestingly, we found 1188 genes significantly regulated (FDR $p < .05$ as compared to 0h) by 1h or 2h TNF α induction even in the absence of RelA expression. 650 genes were significantly (FDR $p < .05$ as compared to no input for the same cell line) regulated both by constant light induction in RelA-CLASP and TNF α induction in pm-LOVTRAP. A subset of these genes are plotted in Figure 3.3D, showing the induction caused by TNF α compared to that caused by RelA-CLASP translocation. Phlda1, a pro-survival factor, and Atf3, a transcriptional repressor, reach similar maximal induction whether induced by TNF α alone or RelA-CLASP (given constant light)^{98,99}. Despite the similar range of induction, the dynamics of their induction are different depending which induction is delivered to the cells. By contrast, Cmtm6, a regulator of PD-L1 that is broadly expressed across cell types, and Rnaset2b, a ribonuclease, are both induced by RelA-CLASP given constant light and repressed by TNF α alone. Furthermore, many genes which are strongly (FDR $p < .05$, \log_2 FC RelA-CLASP induction > 85 th percentile) induced by constant light induction in RelA-CLASP and regulated by TNF α induction in pm-LOVTRAP, are in fact repressed by TNF α alone, but induced by

RelA-CLASP. These data illustrate the wide ranging effects of $\text{TNF}\alpha$ induction and the potential confounding effect environmental inputs can have when used to study the effect of nuclear translocation dynamics of a single transcriptional regulator.

To further probe the genes induced by RelA-CLASP induction with constant light, we calculated $\log_2\text{FC}$ RelA-CLASP induction for all genes significantly (FDR $p < .05$, 5034 genes) regulated by RelA-CLASP in response to constant light. We focused on those genes which had the top 15% of $\log_2\text{FC}$ RelA-CLASP induction at either the 1h or 2h time points, thereby reducing the set of genes of interest to 896 genes. By normalizing the $\log_2\text{FC}$ RelA-CLASP induction to the maximum for each gene, we can use longitudinal k-means clustering to reveal 8 clusters of dynamic gene expression trajectories¹⁰⁰. Cluster A, which represents 230 genes, peaks in induction at 1h of constant light, and maintains or slightly decreases induction at 2h of constant light. Clusters B (167 genes) and C (156 genes) both peak at 2h of constant light induction; the difference between these clusters is that cluster C has a nearly-maximal induction at 1h of constant light. Similar to cluster A, cluster D (140 genes) genes peak at 1h of constant light, yet these genes decrease on average 40% in induction at 2h of constant light. Cluster E, on the other hand, represents 83 genes which induce minimally at 1h of constant light induction, and reach their peak at 2h of constant light induction. Clusters F (80 genes) and G (27 genes), similarly to clusters A and D, peak at 1h of constant light induction, after which their $\log_2\text{FC}$ RelA-CLASP induction decreases strongly to near 0 (cluster F) or even below 0 (Cluster G). Finally, cluster H represents 13 genes which are not activated or even repressed at 1h of constant light induction, but are then strongly induced at 2h of constant light induction (Figure 3.3E).

These 8 clusters can be qualitatively reduced to 4 clusters: early genes, proportional genes, late genes, and feedback genes, a gene activation structure that has been seen in previous studies^{19,53,54}. Early genes, like those in clusters A and C, are those which peak in $\log_2\text{FC}$

RelA-CLASP induction at 1h of constant light and stay activated through 2h of constant light. Proportional genes, like those in cluster B, are those which increase proportionally in \log_2 FC RelA-CLASP induction at both 1h and 2h. Late genes are not induced or even repressed at 1h of constant light induction, and are proceedingly induced at 2h of constant light induction; genes in clusters E and H qualitatively fall into this category. Finally, feedback genes are those which reach peak RelA-CLASP induction after 1h of constant light, and then decrease induction by 2h of constant light, which is seen in clusters D, F, and G. These data show not only the breadth of activation across the genome that is possible with RelA-CLASP, but also the variety of gene expression dynamics that a constant pulse of RelA can induce.

3.3.5 A simple model of gene expression predicts fundamental differences in genes which respond similarly to constant light inputs

Given the differences in response to constant RelA-CLASP nuclear translocation across the four qualitative clusters, it is clear that fundamental differences between genes determine whether they will respond with a given type of trajectory, whether early, proportional, late, or feedback. These differences can be caused by promoter architecture or activation, mRNA decay, or opposing effects of other regulators within the gene regulatory network. Still, though genes within each of the four qualitative clusters have similar responses to a constant input of RelA-CLASP, it remains possible that genes within these clusters have very different underlying parameters.

To further interrogate this question, we generated a simple model of gene expression, capable of recapitulating a subset of the responses to constant light input. This model consists of a promoter in the off state (p_{off}) which transitions to a promoter in the on state

(p_{on}) with rate that is dependent on TF nuclear concentration and an on rate, k_{on} . The promoter also transitions back to an off state using a constant off rate, k_{off} . p_{on} can then produce mRNA with a rate of β_1 . mRNA is also constitutively produced with a rate of β_0 . Finally, the mRNA decays with a rate of γ_1 (Figure 3.4A). This simple model cannot reproduce gene expression dynamics which respond slowly to induction, such as late genes, as proved in 2.7.1 and shown in other works³⁷. Additionally, it cannot reproduce dynamics which display a feedback response because no feedback loop is present in the model. Still, this simple model is useful to explore the parameters which underlie early and proportional genes.

In addition to inducing RelA-CLASP with constant light input, we induced the cells with a pulsed light input of 1h light ON/1h light OFF/1h light ON (Figure 3.4B). We theorized that the gene expression dynamics in response to pulsed light input could be used to differentiate the parameters of genes with qualitatively similar responses to constant RelA-CLASP induction. Accordingly, we used our model to probe whether genes with an early response to constant TF input would display varied responses to a pulsed TF input.

To do this, we generated 8,000 parameter sets varying k_{on} , k_{off} , and γ_1 , and then simulated the mRNA dynamics for each combination of parameters in response to a constant TF input. Each of these simulations was normalized to the maximum mRNA value simulated at 1h or 2h of induction. Then, to find those parameter sets which generated gene expression dynamics similar to those seen in the early genes identified by the RNA-seq data, we filtered the simulated trajectories to those with 1h and 2h normalized induction values within the bound of those seen in clusters A and C in the RNA-seq data. These trajectories are plotted in Figure 3.4C. In general, the parameters which recapitulated the early gene response seen in our RNA seq dataset had a higher k_{on} and higher γ_1 than seen across all parameter sets tested (Figure 3.8A). Given a higher k_{on} and a higher γ_1 value, the gene expression dynamics will reach steady state expression more quickly (2.7.1, Figure

3.8C). These parameter differences explain why genes in clusters A and C might achieve their peak induction, or close to peak induction, after 1h of constant TF input.

Using only this filtered group of parameter sets, we then simulated their response to pulsed TF dynamics generated by the light input illustrated in Figure 3.4B. Remarkably, despite the uniform response to constant light inputs, the simulated responses to pulsed light inputs are varied. They can be grouped into 5 clusters using longitudinal k-means clustering, which are further condensed into 3 qualitatively different clusters (Figure 3.4D)¹⁰⁰. The first cluster, pictured in purple, responds in a similarly early fashion to pulsed inputs as constant inputs, and can actually turn on more in response to a pulsed TF input than a constant input. The second cluster (orange) responds proportionally to pulsed TF inputs, increasing expression at both the 1h and 2h timepoints. Finally, the third cluster (green) responds in a late fashion, turning on only at the 2h timepoint.

The different dynamic trajectories generated by these clusters are due to their relative differences in k_{on} , k_{off} , and γ_1 with respect to all parameters that can generate early gene dynamics in response to a constant light input. The early response to pulsed inputs, pictured in purple, is generated by higher k_{on} and lower k_{off} values relative to those seen across all early gene parameter sets (Figures 3.4E, 3.8B). Higher k_{on} values help the gene to reach the p_{on} state more quickly while the light is on during the first pulse, and lower k_{off} values keep the gene from turning off completely during the 1 hour light OFF period which follows (Figure 3.8C). Together, these opposing forces generate the quick on response to pulsed TF inputs. On average, the proportional response to pulsed inputs (orange, Figure 3.4D) is generated through lower k_{on} , k_{off} , and γ_1 ; however, these parameters can vary widely across simulations with the same qualitative response (Figures 3.4E, 3.8B). Lower k_{on} and k_{off} cause the gene to reach steady state gene expression more slowly during the first hour of TF input, and lower γ_1 prevents mRNA decay during the time that the light is off. This leads to additional mRNA accumulation after the light turns on a second

time (2.7.1, Figure 3.8C). For the late response to pulsed inputs, a higher k_{off} and γ_1 are needed (Figures 3.4E, 3.8B). With a fast k_{on} and k_{off} , these genes will reach steady state quickly, within the first hour that the light is on. However, the high γ_1 also leads to complete decay of the mRNA produced during the 1 hour light OFF period. Still, due to the fast kinetics of these parameter sets, the gene will again reach maximum induction when the light turns on for a second time (Figure 3.8C).

To determine the utility of these predictions, we ran a similar analysis using the RNA-seq data to see if the same responses to pulsed TF inputs were seen amongst the genes with a qualitative early response to constant TF inputs. The genes which make up clusters A and C represent the qualitative early response (Figures 3.3E, 3.4F). Clustering the response of these genes to pulsed light inputs yields 7 clusters of gene expression dynamics (Figure 3.4G). Three of these clusters closely resemble the early, proportional, and late responses predicted by the simple model (27.7% of early genes), which implies that these genes might be well-modeled by a simple promoter. However, there are still 4 clusters of genes which are not predicted by the simple model. The first two clusters of genes both peak early in response to pulsed inputs, and then decrease induction during the the second hour of light input (fourth, fifth graphs, Figure 3.4G; 8.3% of early genes). This decrease in induction in response to TF input cannot be predicted by the simple model. The third cluster consists of genes which have a maximal induction that is much higher for pulsed light inputs than for constant light inputs, which is also not predicted by any parameter set tested with the simple model (sixth graph, Figure 3.4G; 19.6% of early genes). The final cluster, which is not plotted, consists of all those genes which are not significantly induced (FDR $p < .05$) by pulsed light inputs (44.3% of early genes). This lack of induction in response to pulsed inputs additionally cannot be predicted by a simple model.

As discussed previously, the simple model can also be used to recapitulate the proportional response to constant TF inputs. 8,000 parameter sets varying k_{on} , k_{off} , and γ_1 were

generated, and then used to simulate the mRNA dynamics for each combination of parameters in response to a constant TF input. Each of these simulations was normalized to the maximum mRNA value simulated, and then all parameter sets were filtered for those which generated gene expression dynamics similar to those seen in the proportional genes (cluster B, Figure 3.3E) identified by the RNA-seq data (Figure 3.4H). Overall, these parameter sets had lower γ_1 and lower k_{off} values than the full set of parameters tested. Lower k_{off} values cause the gene to reach steady state more slowly. Furthermore, lower γ_1 means that mRNA decay occurs slowly, allowing the mRNA to accumulate over the length of the constant TF input (Figure 3.8A, E).

Next, we used these parameter sets to simulate their response to pulsed TF inputs. This generated two qualitatively different responses to pulsed TF inputs: a proportional response and a quick on response (Figure 3.4I). For both trajectories, a wide range of simulated induction is seen; some parameter sets cause up to 40% more induction in response to pulsed TF inputs than constant TF inputs, while others cause slightly lower induction in response to pulsed TF inputs. The parameter sets which generate a proportional response to pulsed TF inputs have higher k_{on} and k_{off} values than those parameter sets which generate a proportional response to constant TF inputs (Figures 3.4J, 3.8D). As a result, these genes increase expression more quickly during the first TF pulse, and also rapidly shut off after the TF exits the nucleus. Due to the low γ_1 values across all proportional gene sets, the mRNA does not completely decay during the time that the light is off. Finally, during the second pulse of light, the gene reaches its maximal activation (Figure 3.8E). The parameter sets which respond with a quick on dynamic to pulsed TF inputs have lower k_{off} values (Figure 3.4J, Figure 3.8C). A low k_{off} value causes the gene to turn off slowly during the 1h light OFF period, which gives it a slightly higher mRNA value after 1h light ON/1h light OFF than the other parameter sets (Figure 3.8E).

Finally, we wanted to compare the predictions from the simple model to the data observed

with RNA-seq. We clustered the response to pulsed light inputs for the genes assigned to cluster B. This resulted in 5 clusters, two of which qualitatively recapitulated the proportional and quick on responses predicted by the simple model (Figure 3.4L; 23.3% of proportional genes). The third cluster of gene trajectories displays a feedback response, with strong induction at the 1h timepoint and decreased induction at 2h (6.6% of proportional genes). As discussed previously, the simple model has no feedback, so it cannot recapitulate a decrease in induction in response to TF input. The fourth cluster consists of two genes which respond to pulsed light inputs with late gene dynamics, which is also not predicted by the simple model. The fifth cluster displays genes which activate much more strongly in response to pulsed light inputs than constant light inputs, which also cannot be predicted by the simple model (6% of proportional genes). Lastly, many genes from cluster B (61.1% of proportional genes) are in fact not significantly (FDR $p < .05$) regulated by pulsed light inputs.

In summary, the simple model of gene expression is able to predict the response to pulsed light for approximately one-quarter of the genes with an early or proportional response to constant light inputs. The genes in this subset which cannot be predicted likely have other features involved in their activation, such as feedback or more complex gene regulatory networks. Given that RelA is known to regulate genes in conjunction with other transcriptional regulators, and activates several feedback loops, it is remarkable that a meaningful subset of its downstream genes can be modeled as simple promoters.

3.4 Discussion

In this study, we extended our previous efforts to build a modular and reversible optogenetic tool to an additional model system, mammalian cells, and then used this tool to generate a novel transcriptomic dataset that measures the effect of RelA nuclear

translocation on downstream genes. Importantly, we were able to provide the first demonstration that RelA translocation, without phosphorylation or activation of co-regulators, is able to induce downstream gene activation. Using a computational model of gene expression, we were able to further probe this dataset to categorize genes into those which may resemble a simple promoter, and those with more complex features such as feedback or multistep activation. In summary, these results present an important step forward for our understanding of TF dynamics and provide a blueprint for future studies on the same topic.

In Figure 3.3, we use a knockout cell line to measure the RelA- and I κ B-independent gene activation that occurs downstream of a commonly used stimulus, TNF α . These data show that it is critical to use direct methods of regulating TF dynamics in order to study its downstream effects. In future work, it would be useful to measure the RelA-independent gene activation in response to TNF α across additional cell lines, since cell identity can affect gene response⁸⁵. These data, combined with RelA ChIP-seq data, could be useful to better identify direct targets of RelA.

A limitation of this study is that combinatorial control of gene expression is widespread, if not ubiquitous, in mammalian cells. Given the prevalence of combinatorial regulation, it may seem ineffective to study gene activation by activating only one TF at a time. In future studies, though, CLASP could be composed with chemical inputs to precisely regulate RelA translocation during stimulus, thereby allowing other pathways to be activated at the same time. This would be possible because of the robust sequestration of pm-LOVTRAP and the knockout of I κ B proteins, as seen in Figure 3.2. A recent study used chemical inputs and optogenetic control of Erk to discern the effect of Erk translocation on cell proliferation; future studies focusing on RelA could also elucidate the effects of dynamics in a similar fashion¹⁵. Additionally, other studies have compared dynamics across stimuli to understand what features of a dynamic RelA input help the cell

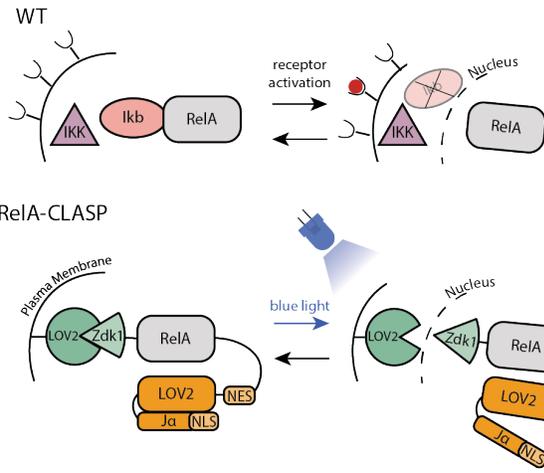
to differentiate between inputs^{101,102}. RelA-CLASP could be used to further substantiate the computational analyses that were used to draw conclusions on how genes use parameters such as amplitude, pulse width, and frequency to differentiate inputs.

Another frontier for improvement in future studies will be to see how single isogenic cells respond differently to the same input. This study is limited by the use of bulk RNA-seq data, whereas future studies could use single cell RNA-seq (scRNA-seq), single molecule RNA FISH (smFISH), or MS2 to further probe heterogeneity in response to the same dynamic TF signal. Several recent studies have used chemical inputs to modulate localization of transcriptional regulators like RelA and Erk, and then have measured the heterogenous response of single cells with these methods^{9,10,103}. Critically, the response of single cells to some chemical inputs is heterogeneous⁹⁰; as a result, studying the heterogenous response of downstream genes may be more complicated given heterogenous TF dynamics instead of uniform dynamics generated by optogenetic inputs.

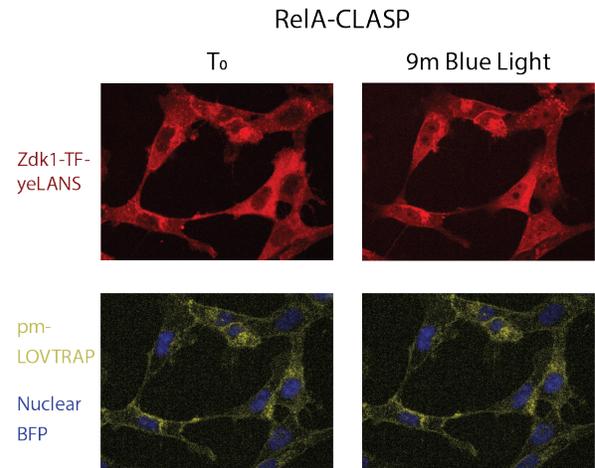
Finally, more complex models are clearly necessary to understand the activation of many genes downstream of RelA-CLASP. A simple model predicts the response to pulsed and constant inputs of only 16% of all the genes most strongly induced by RelA-CLASP. The vast majority of genes identified in this study require models with more complexity, such as feedback, regulation by other genes, or a multi-step promoter activation model³⁷. Future studies could build new models to explain novel observations, such as genes which respond with much higher mRNA induction in response to pulsed inputs than to constant inputs.

3.5 Figures

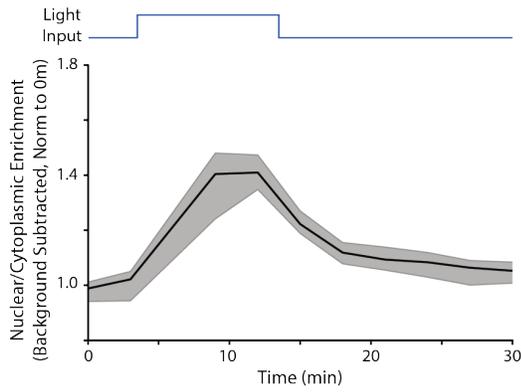
A



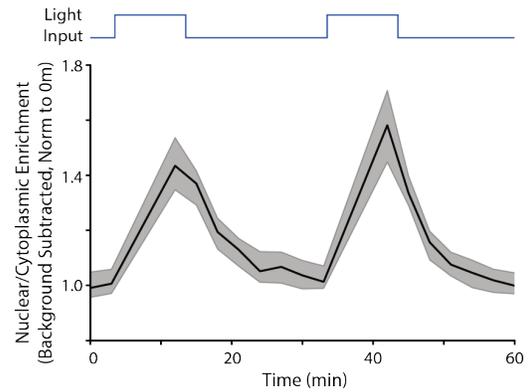
B



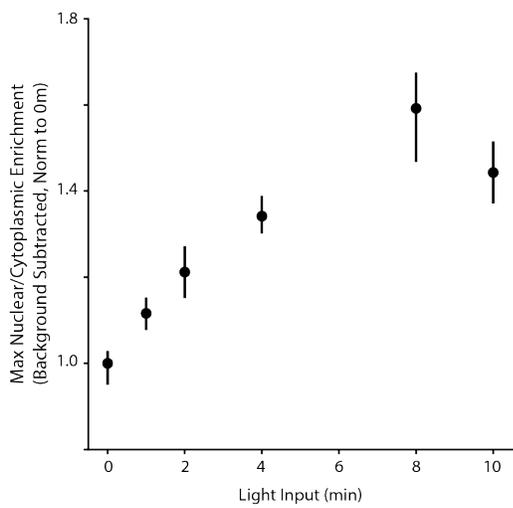
C



D



E



F

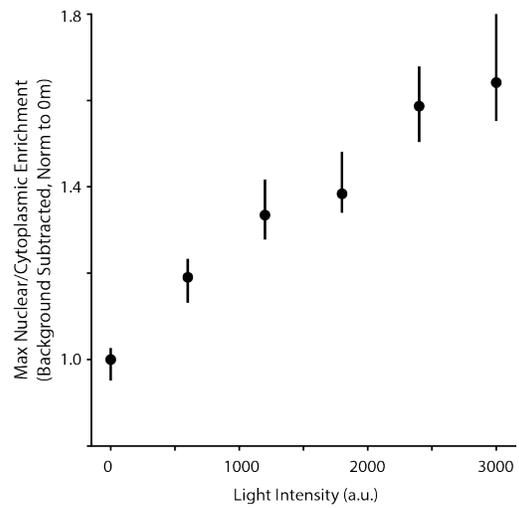


Figure 3.1: RelA-CLASP reversibly translocates to the nucleus in a dose-responsive fashion A) Cartoon illustrating wildtype (WT) activation of RelA and optogenetic activation of RelA-CLASP. (Top panel) IKK, I κ B and RelA proteins are cytoplasmic in unactivated cells. I κ B sequesters RelA in the cytoplasm, preventing nuclear entry. After an activating input, such as TNF α , binds an extracellular receptor, IKK phosphorylates I κ B proteins, causing their degradation. Degradation of I κ B proteins allows RelA to translocate to the nucleus. Through a native feedback loop, RelA activates transcription of I κ B proteins. Translation of these I κ B proteins leads to sequestration of RelA after receptor activation. (Bottom panel) pm-LOVTRAP (green) sequesters Zdk1-RelA-yeLANS at the plasma membrane in the absence of blue light stimulation. In response to blue light, pm-LOVTRAP releases Zdk1-RelA-yeLANS and yeLANS opens to reveal an NLS. Opening of yeLANS allows Zdk1-RelA-yeLANS to be imported into the nucleus. Once the blue light is shut off, Zdk1-RelA-yeLANS exits the nucleus and returns to the cytoplasm. B) Images showing RelA-CLASP translocation. (Top panels) RFP channel showing Zdk1-RelA-mScarlet-yeLANS; image has been false-colored red. (Bottom panels) Merged IRFP (false colored yellow) and BFP (false colored blue) channels showing pm-LOVTRAP and nuclear BFP marker, respectively. Cells are induced with 9 mins of 488nm GFP imaging light between images. C) Quantification of RelA-CLASP response to a single pulse of light input. RelA-CLASP is induced with 10 minutes, 1350 a.u. light. Median nuclear/cytoplasmic enrichment is normalized to its value at 0 minutes and plotted as a dark line; 95% CI is plotted as a gray band. Data is from 3 replicates and 160 cells tracked. D) Quantification of RelA-CLASP response to two pulses of blue light. RelA-CLASP is induced with 10 minutes, 1350 a.u. light, followed by 20 minutes of no light, and then 10 minutes 1350 a.u. light. Median nuclear/cytoplasmic enrichment is normalized to its value at 0 minutes and plotted as a dark line; 95% CI is plotted as a gray band. Data is from 3 replicates and 204 cells tracked. E) RelA-CLASP responsiveness to short inputs. RelA-CLASP is induced with light inputs of lengths varying from 0-10 mins at 1350 a.u. light. 553-665 cells are tracked for each input across 3 replicates. For each replicate and input, the frame with the maximum median nuclear/cytoplasmic enrichment is selected, and the nuclear/cytoplasmic enrichment for each cell present in that frame is normalized to the maximum median nuclear/cytoplasmic enrichment for the no light input in that replicate. Median is plotted for all replicates and 95% CI is plotted as the error bars. F) RelA-CLASP amplitude response. RelA-CLASP is induced with light inputs of intensities varying from 0-3000 a.u. for 15 mins. 152-209 cells are tracked for each input across 3 replicates. For each replicate and input, the frame with the maximum median nuclear/cytoplasmic enrichment is selected, and the nuclear/cytoplasmic enrichment for each cell present in that frame is normalized to the maximum median nuclear/cytoplasmic enrichment for the no light input in that replicate. Median is plotted for all replicates and 95% CI is plotted as the error bars.

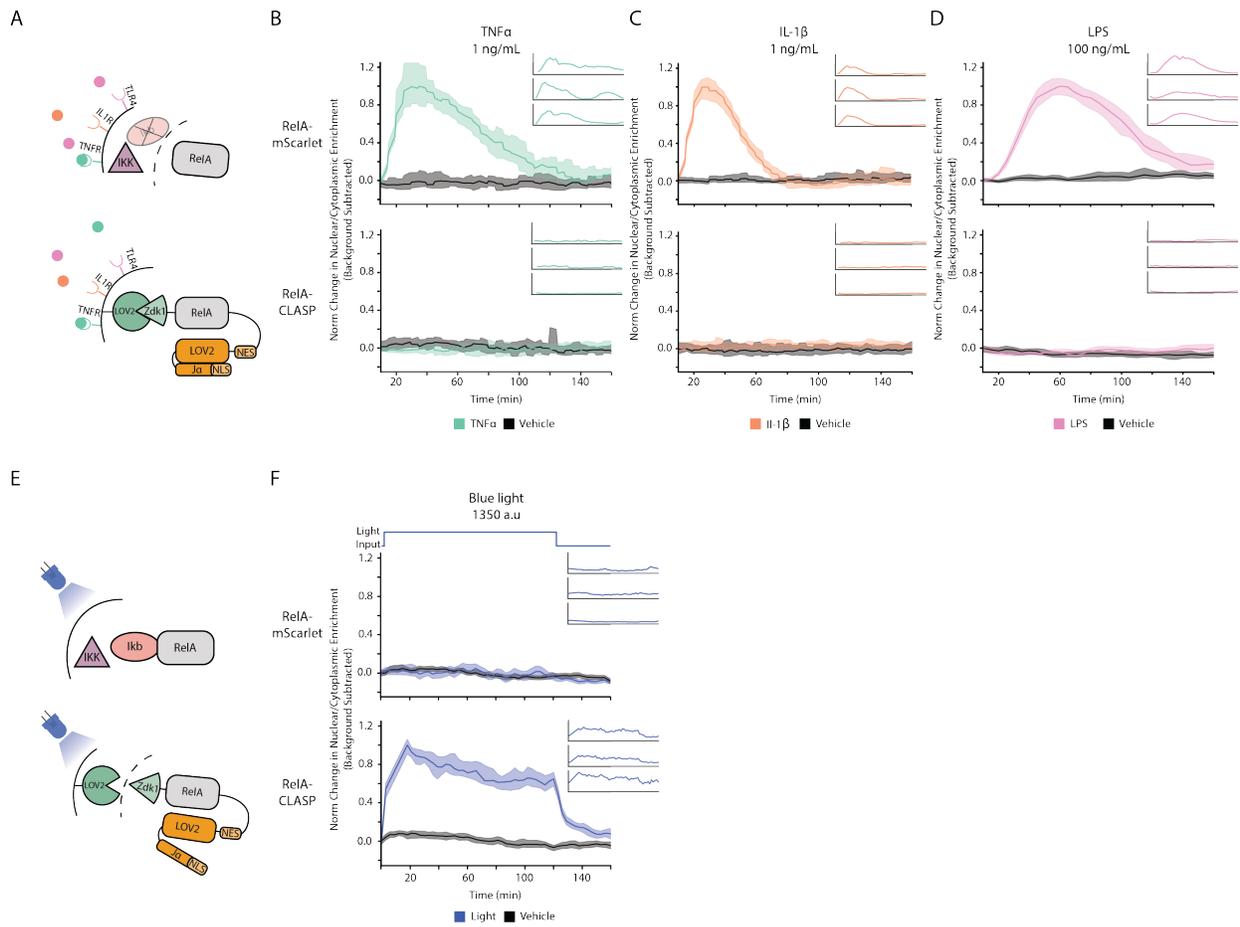


Figure 3.2: RelA-CLASP does not translocate to the nucleus in response to environmental stress inputs A) Cartoon illustrating RelA-mScarlet and RelA-CLASP responses to stress inputs. (Both panels) TNFR (green), IL1R (orange), and TLR4 (pink) are shown as extracellular receptors capable of binding their cognate input. (Top panel) Cartoon illustrating WT system. Binding of $\text{TNF}\alpha$ to TNFR, as an example, leads to activation of IKK, which then leads to degradation of I κ B. I κ B degradation allows RelA to translocate to the nucleus. (Bottom panel) Cartoon illustrating RelA-CLASP system. Binding of $\text{TNF}\alpha$ to TNFR, for example, does not affect subcellular localization of RelA-CLASP. B) Response of RelA-mScarlet and RelA-CLASP to 1 ng/mL $\text{TNF}\alpha$ or phosphate buffered saline (PBS) + .1% BSA vehicle. (Both panels) Norm change in Nuclear/Cytoplasmic Enrichment is calculated as the median change in nuclear/cytoplasmic enrichment for each replicate and cell line divided by the maximum value of the median change in nuclear/cytoplasmic enrichment across both cell lines. Green and black lines show smoothed (window = 2) median of all cells tracked across 3 replicates; green and gray confidence bands show 95% CI. (Top panel) Response of RelA-mScarlet cell line to $\text{TNF}\alpha$ input (green) or PBS + .1% BSA vehicle (black); at least 187 cells tracked across 3 replicates for each input. Inset in upper right shows traces of 3 randomly selected cells responding to $\text{TNF}\alpha$ input. (Bottom panel) Response of RelA-CLASP cell line to $\text{TNF}\alpha$ input (green) or PBS + .1% BSA vehicle (black); at least 103 cells tracked across 3 replicates for each input. Inset in upper right shows traces of 3 randomly selected RelA-CLASP cells responding to $\text{TNF}\alpha$ input. C) Response of RelA-mScarlet and RelA-CLASP to 1 ng/mL IL-1 β or H₂O + .1% BSA vehicle. (Both panels) Norm change in Nuclear/Cytoplasmic Enrichment is calculated and plotted as described in B. (Top panel) Response of RelA-mScarlet cell line to IL-1 β input (orange) or H₂O + .1% BSA vehicle (black); at least 128 cells tracked across 3 replicates for each input. Inset in upper right shows traces of 3 randomly selected cells responding to IL-1 β input. (Bottom panel) Response of RelA-CLASP cell line to IL-1 β input (orange) and H₂O + .1% BSA vehicle (black); at least 89 cells tracked across 3 replicates for each input. Inset in upper right shows traces of 3 randomly selected RelA-CLASP cells responding to IL-1 β input. D) Response of RelA-mScarlet and RelA-CLASP to 100 ng/mL LPS or PBS vehicle. (Both panels) Norm change in Nuclear/Cytoplasmic Enrichment is calculated and plotted as described in B. (Top panel) Response of RelA-mScarlet cell line to LPS input (pink) or PBS vehicle (black); at least 85 cells tracked across 3 replicates for each input. Inset in upper right shows traces of 3 randomly selected cells responding to LPS input. (Bottom panel) Response of RelA-CLASP cell line to LPS input (pink) and PBS vehicle (black); at least 70 cells tracked across 3 replicates for each input. Inset in upper right shows traces of 3 randomly selected RelA-CLASP cells responding to LPS input. E) Cartoon illustrating RelA-mScarlet and RelA-CLASP responses to blue light. (Top panel) Cartoon illustrating WT system. Blue light input does not cause any changes to IKK, I κ B, or RelA, and no nuclear translocation of RelA occurs. (Bottom panel) Cartoon illustrating RelA-CLASP system. Blue light stimulation activates both pm-LOVTRAP and yeLANS, causing release of Zdk1-RelA-yeLANS from the plasma membrane and opening of yeLANS, thereby revealing an NLS. Revealing NLS leads to nuclear import of RelA-CLASP.

Figure 3.2: F) Response of RelA-mScarlet and RelA-CLASP to 1350 a.u. light or no light. (Both panels) Norm change in Nuclear/Cytoplasmic Enrichment is calculated and plotted as described in B. (Top panel) Response of RelA-mScarlet cell line to blue light (blue) or no light (black); at least 928 cells tracked across 3 replicates for each input. Inset in upper right shows traces of 3 randomly selected cells responding to blue light input. (Bottom panel) Response of RelA-CLASP cell line to blue light input (blue) and no light (black); at least 508 cells tracked across 3 replicates for each input. Inset in upper right shows traces of 3 randomly selected RelA-CLASP cells responding to blue light input.

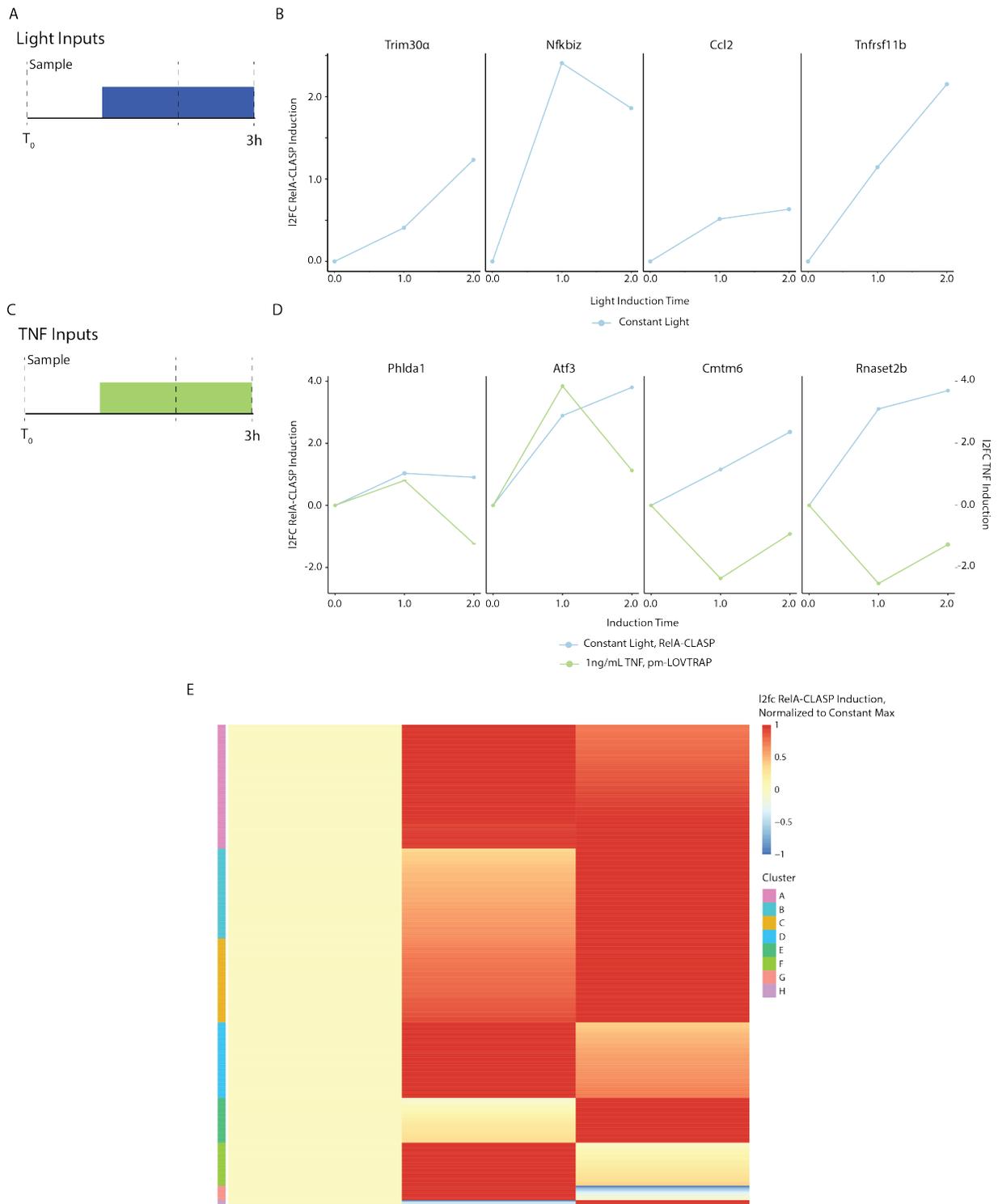


Figure 3.3: RelA-CLASP activates downstream genes in response to a constant light input A) Constant light input of 460 a.u. given to cells preceding RNA-seq. Dashed lines show when samples were taken for RNA-seq. Samples were taken at T_0 , with no light input. Samples were also taken after 1 hour of light input and 2 hours of light input. B) Induction of canonical genes downstream of RelA. \log_2 FC RelA-CLASP Induction is calculated as the \log_2 FC for RelA-CLASP at time t compared to RelA-CLASP at time 0 minus \log_2 FC for pm-LOVTRAP at time t compared to pm-LOVTRAP at time 0; this is calculated only for genes which are significantly (FDR $p < .05$) induced by RelA-CLASP. \log_2 FC RelA-CLASP Induction is plotted for Trim30a, Nfkbiz, Ccl2, Tnfrsf11b given a constant light input. C) $TNF\alpha$ input of 1 ng/mL given to cells preceding RNA-seq. Dashed lines show when samples were taken for RNA-seq. Samples were taken at T_0 , with no $TNF\alpha$ input. Samples were also taken after 1 hour of $TNF\alpha$ input and 2 hours of $TNF\alpha$ input. D) Induction and repression of canonical genes downstream of RelA. \log_2 FC $TNF\alpha$ Induction is calculated as pm-LOVTRAP $TNF\alpha$ at time t compared to pm-LOVTRAP $TNF\alpha$ at time 0. Blue lines show \log_2 FC RelA-CLASP Induction; green lines show \log_2 FC $TNF\alpha$ induction. \log_2 FC is calculated from 5 replicates for each cell line and input. E) Heatmap showing clustered, normalized \log_2 FC RelA-CLASP induction for genes most strongly induced by RelA-CLASP (FDR $p < .05$, \log_2 FC RelA-CLASP induction > 85 th percentile). Normalized \log_2 FC RelA-CLASP induction is calculated by dividing \log_2 FC RelA-CLASP induction for a given gene by the maximum \log_2 FC RelA-CLASP induction for that gene. Longitudinal gene trajectories are grouped into 8 clusters using k-means clustering.

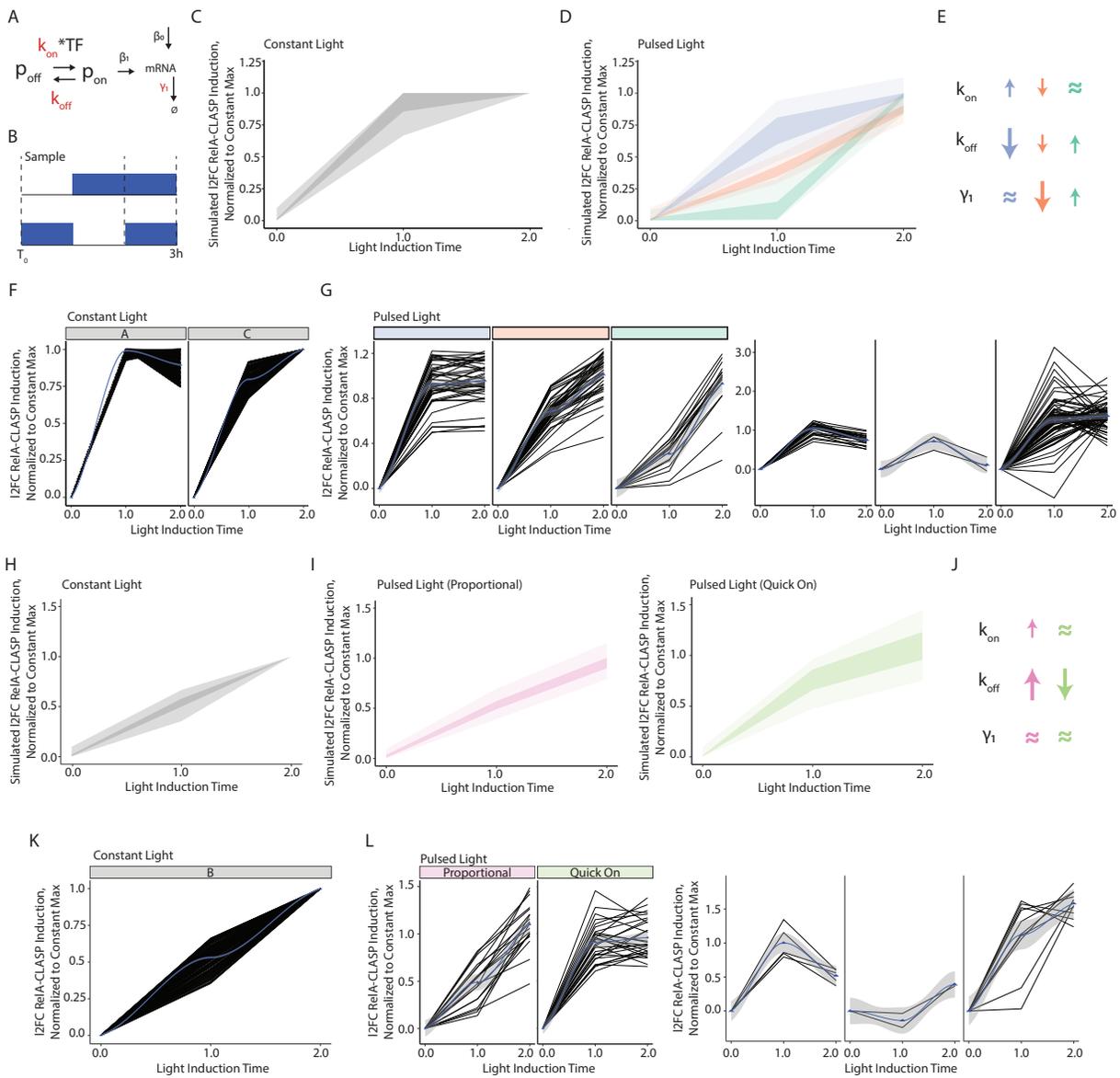


Figure 3.4: A simple model of gene expression predicts response of some genes to pulsed RelA-CLASP inputs A) Simple model of gene expression. p_{off} represents an off promoter, which transitions to an on promoter (p_{on}) dependent on TF nuclear concentration and the parameter k_{on} . p_{on} transitions back to p_{off} with the parameter k_{off} . p_{on} produces mRNA with the rate β_1 , and mRNA is also produced basally with the rate β_0 . mRNA then decays with the rate γ_1 . k_{on} is varied from .002-2, k_{off} is varied from .004-4, and γ_1 is varied from .001-1. B) Light input of 460 a.u. given to cells preceding RNA-seq. Dashed lines show when samples were taken for RNA-seq. (Both panels) Samples were taken at T_0 , with no light input. (Top panel) Samples were also taken after 1 hour of light induction and 2 hours of light induction. (Bottom panel) Samples were also taken after 1 hour of light ON/1 hour light OFF and 2 hours of pulsed light induction. C) Graph of simulated $\log_2\text{FC}$ RelA-CLASP induction for parameter sets that recapitulate early gene response to constant light observed in RNA-seq. Simulated $\log_2\text{FC}$ RelA-CLASP induction, normalized to constant max is calculated by dividing the mRNA values at 0, 1, and 2h by the maximum mRNA value across those timepoints for each parameter set. Dark gray bands show 25th-75th percentile of normalized simulated mRNA values at 1h and 2h; light gray bands show 0-100th percentile of the same quantity. D) Graph of simulated $\log_2\text{FC}$ RelA-CLASP induction in response to pulsed light for parameter sets shown in C. Parameter sets seen in panel C are used to simulate the response to pulsed light for genes which display an early response to constant light. Simulated $\log_2\text{FC}$ RelA-CLASP induction, normalized to constant max is calculated by dividing the mRNA values at 0, 1, and 2h of pulsed light input by the maximum mRNA value observed for constant light input for each parameter set. Longitudinal gene trajectories are clustered with k-means clustering and then further condensed to 3 clusters by grouping visually similar clusters. Gene trajectories in purple represent a quick on response to pulsed inputs; orange trajectories show a proportional response to pulsed inputs; green trajectories represent a late response to pulsed inputs. Dark colored bands show 25th-75th percentile of normalized simulated mRNA values at 1h and 2h; light colored bands show 0-100th percentile of the same quantity. E) Graphic showing relative differences in k_{on} , k_{off} , and γ_1 parameters across clusters in D. Up arrows indicate higher parameter values, down arrows indicate lower parameter values, and \approx indicates similar parameter values to all early gene parameter sets. Size of arrow denotes size of relative difference. Purple, or quick on, gene dynamics have on higher k_{on} , lower k_{off} , and a similar γ_1 when compared to all early response to constant light input parameter sets. Orange, or proportional, gene trajectories display a similar k_{on} , higher k_{off} , and lower γ_1 compared to all relevant parameter sets. Green, or late response, gene dynamics have a similar k_{on} , higher k_{off} , and higher γ_1 as compared to all early response to constant light input parameter sets. F) Clusters A and C, seen in Figure 3.3E, plotted as line graphs to demonstrate early gene response to constant light inputs. Blue line represents fit and blue triangles represent mean $\log_2\text{FC}$ RelA-CLASP induction at a given timepoint. G) Graphs showing pulsed light response for genes in clusters A and C. Dynamic trajectories in response to pulsed light for genes in clusters A and C are clustered using k-means clustering. Three clusters are qualitatively similar to those shown in panel D; colored bars above each graph show which simulated cluster each is most similar to.

Figure 3.4: Four clusters (three pictured) are not predicted by the model: the first two clusters display a feedback response; the next has much higher induction in response to pulsed light inputs than constant light inputs; one (not pictured) is not significantly induced by pulsed inputs. H) Graph of simulated \log_2 FC RelA-CLASP induction for parameter sets that recapitulate proportional gene response to constant light observed in RNA-seq. Simulated \log_2 FC RelA-CLASP induction, normalized to constant max is calculated and plotted as described for panel C. I) Graph of simulated \log_2 FC RelA-CLASP induction in response to pulsed light for parameter sets shown in H. Parameter sets seen in panel H are used to simulate the response to pulsed light for genes which display a proportional response to constant light. Simulated \log_2 FC RelA-CLASP induction, normalized to constant max is calculated and plotted as described in panel D. Longitudinal gene trajectories are clustered with k-means clustering and then further condensed to 2 clusters by grouping qualitatively similar clusters. J) Graphic showing relative differences in k_{on} , k_{off} , and γ_1 parameters across clusters in I. Arrows indicate relative differences as described in E. Pink, or proportional, gene dynamics have higher k_{on} , higher k_{off} , and similar γ_1 when compared to all proportional response to constant light input parameter sets. Green, or quick on, gene trajectories have lower k_{off} and similar k_{on} and k_{off} as all relevant parameter sets. K) Cluster B, seen in Figure 3.3E, plotted as a line graph to demonstrate proportional gene response to constant light inputs. Blue line and triangles are as described in panel F. L) Graphs showing pulsed light response for genes in cluster B. Dynamic trajectories in response to pulsed light for genes in cluster B are clustered using k-means clustering. Two clusters are qualitatively similar to those shown in panel D; colored bars above each graph show which simulated cluster each is most similar to. Four clusters (three pictured) are not predicted by the model: the first cluster displays a feedback response; the second cluster has a late gene response to pulsed inputs; the third cluster displays much stronger induction in response to pulsed inputs than to constant inputs; the fourth cluster (not pictured) is not significantly induced by pulsed inputs.

3.6 Materials and Methods

3.6.1 Experimental details

Mammalian cell culture

HEK293T and LX293T cells were cultured in 1g/L glucose DMEM (Life Technologies 11885076), 1% Antibiotic-Antimycotic (Thermo 15240062), and 10% Fetal Bovine Serum.

NIH3T3 and 3T34KO cells were maintained in 1g/L glucose DMEM (Life Technologies 11885076), 1% Antibiotic-Antimycotic (Thermo 15240062), and 10% heat-inactivated Bovine Calf Serum (UCSF Cell Culture Facility, HyClone, lot number AZM197696). Heat inactivation was accomplished by heating serum at 56°C for 30 minutes. After heating, serum was cooled to room temperature before media production. MCF10A cells were cultured in DMEM/F12 (Thermo 21331020), 5% Horse Serum (UCSF Cell Culture Facility), .1mg/mL EGF, 4 mg/mL Insulin (Gibco 12585014), 1 mg/mL Hydrocortisone, .1mg/mL Cholera toxin (Sigma C8052), and 1% Antibiotic-Antimycotic (Thermo 15240062). All cells were cultured at 37°C and 5% CO₂. NIH3T3 and 3T34KO cells were passaged every three days; HEK293T, LX293T, and MCF10A cells were passaged every other day. For microscopy, a 96-well glass-bottom plate (Thermo Fisher 164588) or a 24-well microscopy plate (Ibidi 82406) was incubated with .1mg/mL Poly-D-Lysine (Gibco A3890401) at room temperature for 1 hour, after which the plate was washed three times with sterile water and left to dry for 2 hours. After drying, cells were seeded at 8000 cells/well (96 well plate) or 35000 cells/well (24 well plate). 48 hours later, cells were imaged.

Plasmid and cell line construction

Hierarchical golden gate assembly was used to assemble all plasmids^{70,87}. BsaI and BsmBI sites were removed from parts to enable further assembly. Parts were generated through PCR or ordered as gBlocks from IDT. Plasmids were grown and prepared from DH5 α , Mach1, or Stbl1 competent cells (Macrolab, Berkeley, CA). For lentiviral transduction, plasmids were first transfected into LX293T cells at 80% confluency using Lipofectamine 2000 (Thermo 11668019), the plasmid of interest, and two plasmids encoding second generation lentiviral envelope and packaging vectors (MDG.2 and CMV). Transfection reagent and media were removed from LX293T cells the next day, approximately 16 hours

later, and transfected cells were refed with 1g/L glucose DMEM (Life Technologies 11885076), 1% Antibiotic-Antimycotic (Thermo 15240062), and 10% Fetal Bovine Serum. 24 hours later, the media was removed from transfected LX293T cells and filtered through a .45 micron filter to remove cell debris. For 3T34KO cells, polybrene was added to the filtered viral supernatant to achieve a final concentration of 4 $\mu\text{g}/\text{mL}$ after adding to cells. Viral supernatant was then added to cells at for transduction slowly on top of media. After addition of viral supernatant, cells were spun at 800xg for 45 minutes to increase transduction efficiency. After 16-24 hours of incubation with viral supernatant, cells were refed with fresh media. After transduction, cells were sorted to select the population of interest. For Piggybac transfection, cells were transfected with the TF-CLASP plasmid in addition to PiggyBac Transposase (pCMV-hyPBase) using Lipofectamine 2000 (Thermo 11668019) according to the manufacturer's instructions.

Cell selection via sorting

To prepare for sorting, cells were lifted using trypsin and resuspended in the corresponding media to quench trypsin activity. Afterwards, cells were spun down at 400xg for 5 minutes to form a pellet and placed on ice. This pellet was then resuspended in PBS for sorting. Sorting was performed on a BD FACSAria II. BFP was assessed using the BV405 channel (405nm excitation, 450/50nm filter), mScarlet was measured using the mCherry channel (561nm excitation, 610/20nm filter), and IRFP was assessed using the APC-Cy7 channel (633nm excitation, 780/60nm filter). Cells were sorted into fresh media and re-plated after sorting.

RelA-CLASP cell line generation

As described in Figure 3.2, RelA-CLASP was generated through lentiviral transduction of the bulk-sorted 3T34KO chassis cell line. After transduction, this cell line was further selected to generate a clonal cell line which is referred to as RelA-CLASP in this study. Single cells expressing BFP, low IRFP, and low RFP were sorted into a 96 well plate and then clonally expanded. After expansion, clonal cell lines were assessed for continued expression of fluorophores and responsiveness of CLASP construct. A single cell line, termed F8 lo, was selected for use in this study.

Microscopy

Microscopy for all figures (except for Figures 3.1B and 3.6D-E) was performed on an inverted Nikon Ti microscope equipped with a CSU-22 spinning disk confocal, EMCCD camera, and custom 4-line solid state laser launch. Imaging took place inside a cage incubator which maintained temperature, CO₂, and humidity throughout the experiment. Images were taken using a 40x/0.95 objective, and cells were illuminated with 405, 561, and 640nm lasers. For any images where cells are induced with light on this microscope, cells were covered with a BreatheEasy seal and a custom-printed Optoplate holder was mounted on top of the cells. The Optoplate was then placed on top of the holder to induce the cells. For Figures 3.1B and 3.6D-E, microscopy was performed on an inverted Nikon Ti microscope with an Andor iXon Ultra DU888 1k x 1k EMCCD and Andor 4-line laser launch. An Oko stage was used to maintain temperature and atmosphere control. For these panels, cells were induced using 488 nm light produced by imaging GFP.

Drug induction

IL-1 β (Peprotech, 211-11B) was diluted to a stock concentration of 100ng/mL in sterile water with .1% Bovine Serum Albumin (BSA). LPS (Sigma, L2880) was diluted to 10 μ g/mL in phosphate-buffered saline (PBS). TNF α (RD Systems, 410-MT) was diluted to a 100ng/mL stock solution in PBS with .1% BSA. All stocks were at 100X concentration. Prior to induction, stocks were diluted to 2X or 3X in media and then added to cells to a final 1X concentration (TNF α , IL-1 β : 1ng/mL, LPS: 100ng/mL).

Light induction using Optoplate-96

Optoplate-96 was programmed using the OptoConfig-96 program¹⁰⁴. Cells were induced with up to 12 minutes of constant light input, followed by a pulsed light input of 2 seconds ON/2 seconds OFF to reduce blue light toxicity.

RNA-seq

Cells were seeded with 40,000 cells/well in 24 well plates (Ibidi 82406) 2 days prior to experiment, so that they would be 80% confluent when induced. Five replicate wells were seeded for each input and cell line. Immediately prior to experiment, cells were induced with vehicle (PBS + .1% BSA) or TNF α diluted in media to a 2X concentration. For induction, .5 mL of media was removed from the well and .5 ml of induction media was added. After induction, a BreatheEasy seal (Sigma Z380059) was placed on cell plate. For cells induced with TNF α for 1h, cells were placed into experiment incubator for 2 hours, removed, induced, and then removed 1h hour later for harvesting. For cells induced with TNF α for 2h, cells were placed in experiment incubator 1h before induction. For cells induced with light, cells were induced with vehicle media and placed into incubator with

Optoplate-96 directly on top of BreatheEasy seal for 3 hours. All cells were harvested immediately after induction, and RNA was isolated from cell pellets using the Lexogen SPLIT RNA extraction kit. After extraction, RNA quality was assessed using the Agilent Pico RNA kit, and quantified using a Nanodrop. Following extraction, RNA samples were diluted using concentrations estimated by Nanodrop. Libraries were prepared using the Lexogen Quantseq 3' mRNA-Seq Library Prep Kit on 250 ng RNA from each sample. Library quality was assessed using the Agilent High Sensitivity DNA Kit, and quantity was measured using the Qubit dsDNA HS Assay Kit. Libraries were diluted to equimolar concentrations and pooled. Pools were subject to single-end sequencing on the Illumina HiSeq 400.

Analysis of RNA-seq data

Reads were aligned to the *Mus musculus* genome using the cloud service Bluebee designed for Lexogen Quantseq data. Briefly, the reads were trimmed using Bbduk and then aligned to the GRCm38 genome using STAR¹⁰⁵. After alignment, counts were generated using HTSeq-count¹⁰⁶. Raw counts data was then used with DESeq2 to generate \log_2FC and FDR p values¹⁰⁷. Genes with an average raw counts ≤ 2 across all samples of interest were dropped from the analysis. \log_2FC RelA-CLASP induction was calculated for genes significantly regulated (FDR $p < .05$) by RelA-CLASP as such: \log_2FC (RelA-CLASP time t vs time 0) - \log_2FC (pm-LOVTRAP time t vs time 0), where t is either 1 or 2 hours of light induction. \log_2FC TNF α induction was simply \log_2FC (pm-LOVTRAP time T vs time 0) for 1 or 2 hours of TNF α induction.

Computational modeling

Modeling is as described in Chapter 2 for the two-state promoter model. Briefly, ordinary differential equations representing promoter kinetics were constructed with three state variables and seven parameters. Parameters for k_{on} , k_{off} , and γ_1 were sampled across the log space of .002-2, .004-4, and .001-1, respectively.

Image analysis

Microscopy images are analyzed for nuclear and cytoplasmic intensity using StarDist, Scikit-image, and OpenCV^{108,109}. First, StarDist is used on nuclear BFP images to create masks of nuclei. Then, the cytoplasm is approximated by dilating the nuclear mask four times and subtracting a twice-dilated nuclear mask. Background of each image is estimated by expanding all nuclear masks in an image by 50 pixels, which approximates the cell radius, and then taking the mode of the intensity of the pixels which are not labeled by a mask. Finally, OpenCV is used to track centroids of the nuclear masks throughout the experiment. Nuclear/cytoplasmic enrichment (background subtracted) is calculated as (avg nuclear intensity - background intensity) / (avg cytoplasmic intensity - background intensity).

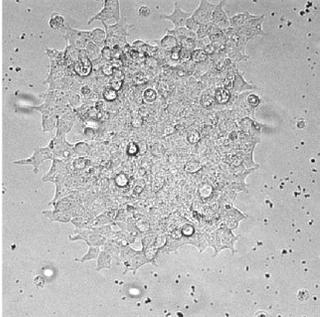
Data processing

Data processing was done with custom-written Python, R, and Matlab scripts.

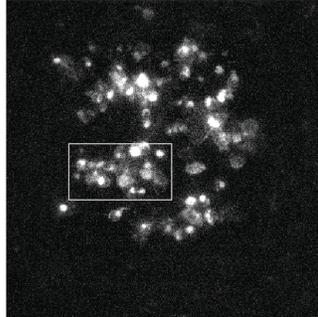
3.7 Supplemental Figures

A

Hs_RGS2
BF



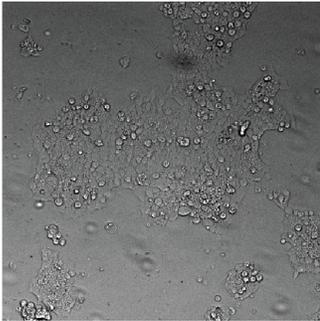
IRFP



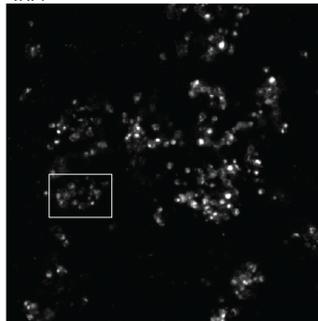
Zoom



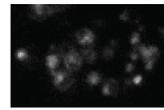
Lyn
BF



IRFP



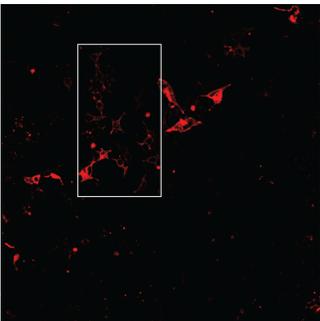
Zoom



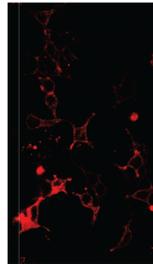
B

T_0

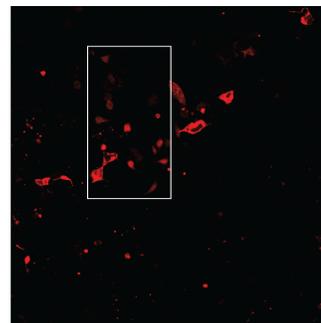
Zdk1-
mScarlet-
yeLANS



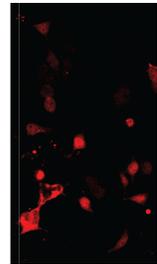
Zoom



5 min
Blue light



Zoom



pm-
LOVTRAP
(Lyn tag)

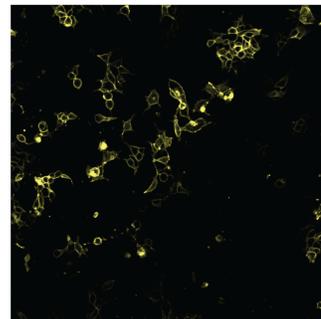
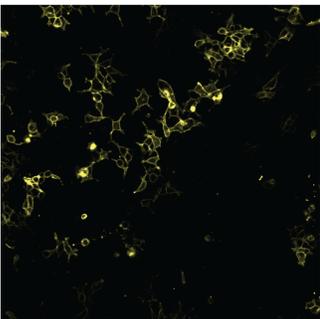


Figure 3.5: Improving plasma membrane localization for mammalian CLASP

A) Plasma membrane targeting improvements for mammalian CLASP. (Top panel) Left photo shows brightfield image of HEK293T cells expressing Hs'RGS2-tagged pm-LOVTRAP. Right image shows IRFP fluorescence for same cells. A zoomed image is shown on the far right to display lack of plasma membrane targeting. (Bottom panel) Left photo shows brightfield image of HEK293T cells expressing Lyn kinase-tagged pm-LOVTRAP. Right image shows IRFP fluorescence for same cells. A zoomed image is shown on the far right to display plasma membrane targeting. B) Improved plasma membrane targeting and robust translocation for mScarlet-CLASP. (Left panel) Top image shows RFP fluorescence seen in confocal microscopy of mScarlet-CLASP. RFP fluorescence denotes localization of Zdk1-mScarlet-yeLANS. Zoomed image on right shows localization in more detail. Bottom photo shows IRFP fluorescence, denoting localization of pm-LOVTRAP. (Right panel) Top image shows RFP fluorescence and Zdk1-mScarlet-yeLANS localization after 5 minutes of 488 nm blue light induction. Zoomed image on right shows localization in more detail. Bottom photo shows IRFP fluorescence, denoting localization of pm-LOVTRAP.

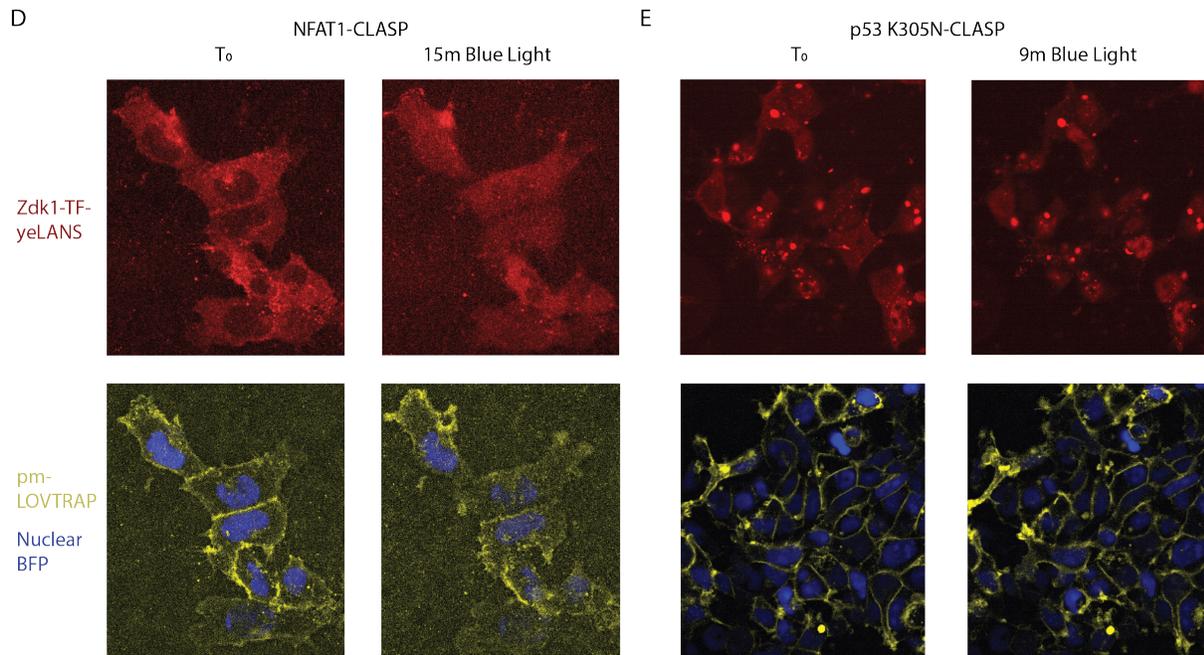
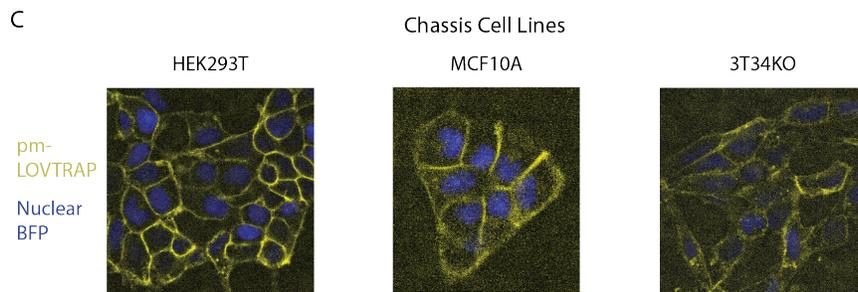
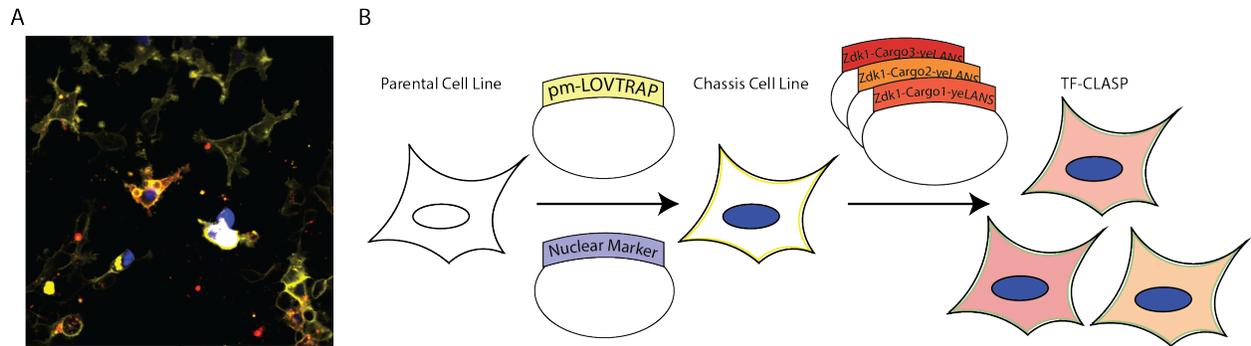
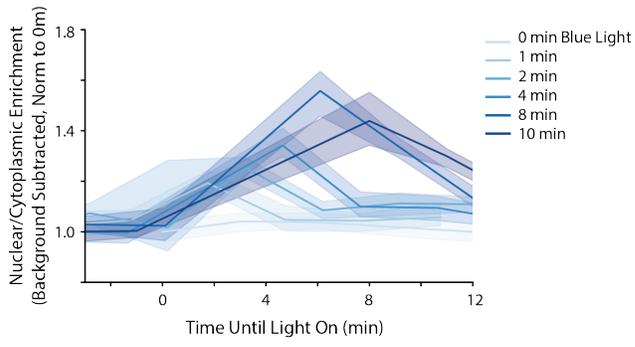


Figure 3.6: Improving delivery of mammalian CLASP and measuring dynamics A) Confocal microscopy demonstrating silencing of expression of mScarlet-CLASP in HEK293T cells. Image is false colored; yellow denotes IRFP fluorescence (pm-LOVTRAP), red denotes RFP fluorescence (Zdk1-mScarlet-yeLANS), and blue denotes BFP fluorescence (Nuclear BFP marker). All cells have been selected for uptake of an mScarlet-CLASP plasmid which should yield expression of IRFP, RFP, and BFP. B) Schematic of three-plasmid CLASP system. Parental cell lines can be lentivirally transduced with two plasmids at once, expressing pm-LOVTRAP and a nuclear marker, thereby creating a chassis cell line. This chassis cell line can then be transduced with a variety of Zdk1-cargo-yeLANS plasmids to test CLASP regulation of TF translocation. C) Images of chassis cell lines. Chassis cell lines created in HEK293T, MCF10A, and 3T34KO are shown. IRFP and BFP channels are merged; IRFP, representing pm-LOVTRAP localization, is false-colored yellow, and BFP, which is used as the nuclear marker, is false-colored blue. D) Images showing NFAT1-CLASP translocation. (Top panels) RFP channel showing Zdk1-NFAT1-mScarlet-yeLANS; image has been false-colored red. (Bottom panels) Merged IRFP (false colored yellow) and BFP (false colored blue) channels showing pm-LOVTRAP and nuclear BFP marker, respectively. Cells are induced with 15 mins of 488nm GFP imaging light between images. E) Images showing p53 K305N-CLASP translocation. (Top panels) RFP channel showing Zdk1-p53 K305N-mScarlet-yeLANS; image has been false-colored red. (Bottom panels) Merged IRFP (false colored yellow) and BFP (false colored blue) channels showing pm-LOVTRAP and nuclear BFP marker, respectively. Cells are induced with 9 mins of 488nm GFP imaging light between images.

A



B

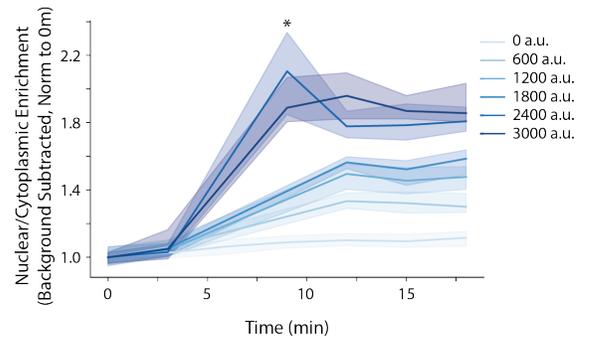


Figure 3.7: Amplitude and time response of RelA-CLASP A) Timecourse graph showing RelA-CLASP translocation in response to short inputs of blue light. Light input turns on at 0 minutes as shown on x-axis. Darker shades of blue correspond to longer inputs of light. Data is from 3 replicates with 553-665 cells tracked per input. Nuclear/Cytoplasmic Enrichment for each input is normalized to the value measured just before the light turns on. B) Timecourse graph showing RelA-CLASP translocation in response to different amplitudes of blue light. Light input turns on at 5 minutes. Darker shades of blue correspond to higher light intensity. Data is from 3 replicates with 152-209 cells tracked per input. Asterisk denotes 9-minute timepoint for 2400 a.u., which has data only from 1 replicate. Nuclear/Cytoplasmic Enrichment for each input is normalized to the value measured at 0 minutes.

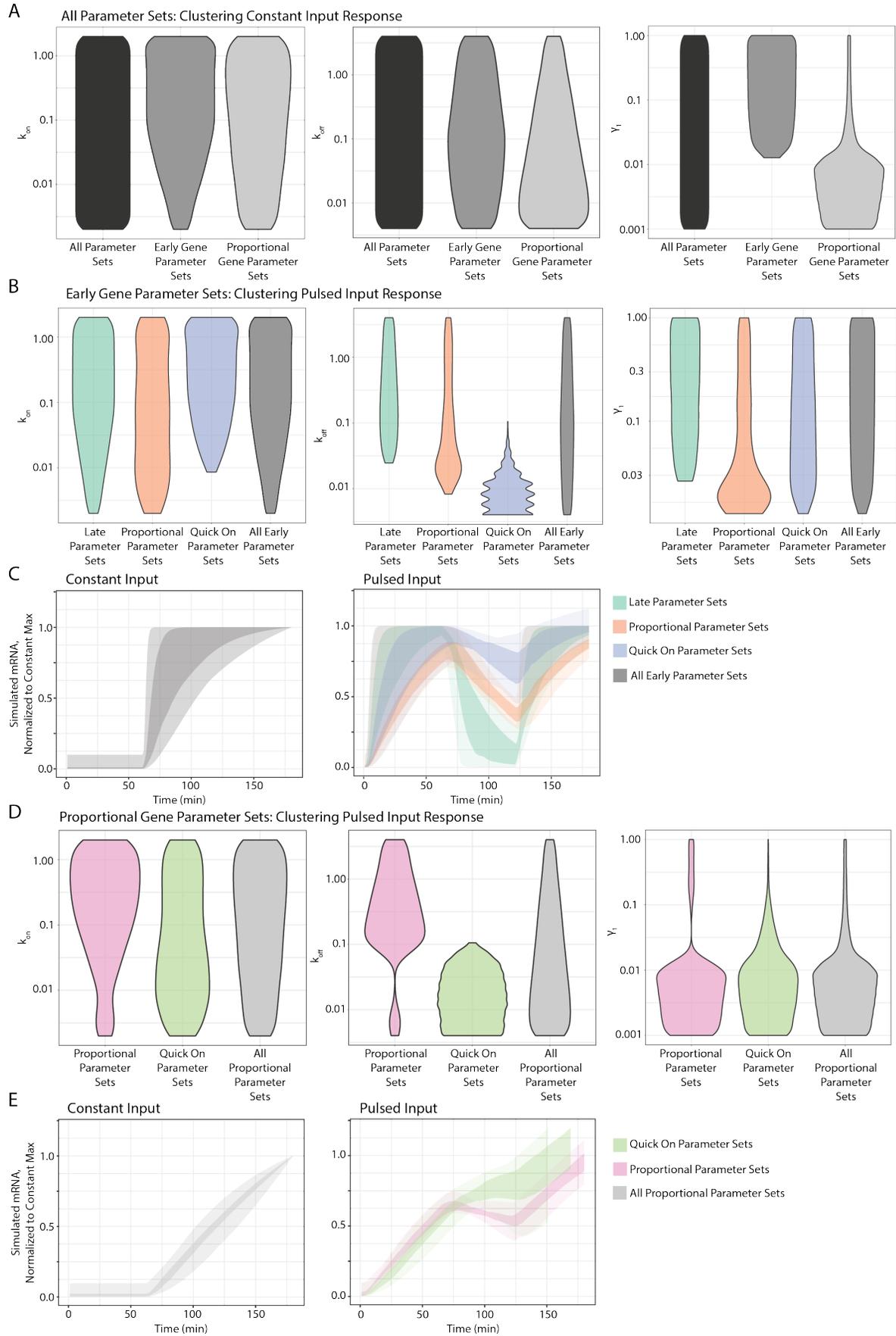


Figure 3.8: Parameter values determine response to constant and pulsed inputs

A) Plots of parameter values that yield early and proportional responses to constant inputs. Violin plots for k_{on} , k_{off} , γ_1 are shown. All parameter sets are plotted in black, parameter sets with an early gene response are plotted in dark gray, and parameter sets with a proportional response are plotted in light gray. B) Plots of early gene parameter values that lead to differential responses to pulsed light inputs. All early gene parameter sets are shown in each graph for reference. Parameter sets which yield a late gene response to pulsed TF inputs are shown in green. Parameters which lead to a proportional response after pulsed inputs are shown in orange, and those which yield a quick on response are shown in purple. C) Plots of normalized simulated mRNA vs time for all early gene parameter sets. (All panels) Simulated mRNA is normalized by dividing the simulated mRNA for each parameter set across all timepoints by the maximum simulated mRNA value measured at 0, 120, or 180 min. Dark colored bands show 25th-75th percentile of normalized simulated mRNA values; light colored bands show 0-100th percentile of the same quantity. (Left panel) A subset of early gene parameter sets in response to constant light input (n=100). (Right panel) A subset of late, proportional, and quick on parameter sets in response to pulsed light input (n=100 for each). D) Plots of proportional gene parameter values that lead to differential responses to pulsed light inputs. All proportional gene parameter sets are shown in each graph for reference. Parameter sets that yield a proportional response after pulsed TF inputs are shown in pink, and those which yield a quick on response are shown in green. E) Plots of normalized simulated mRNA vs time for all proportional gene parameter sets. (All panels) Simulated mRNA is normalized and plotted as described in C. (Left panel) A subset of proportional gene parameter sets in response to constant light input (n=100). (Right panel) A subset of proportional and quick on parameter sets in response to pulsed light input (n=100 for each).

References

- [1] A. Sadeh, N. Movshovich, M. Volokh, L. Gheber, and A. Aharoni. Fine-tuning of the Msn2/4-mediated yeast stress responses as revealed by systematic deletion of Msn2/4 partners. *Mol. Biol. Cell*, 22:3127–3138, 2011.
- [2] M. Springer, D.D. Wykoff, N. Miller, and E.K. O’Shea. Partially phosphorylated Pho4 activates transcription of a subset of phosphate-responsive genes. *PLoS Biol.*, 1:E28, 2003.
- [3] M. Czyz, M.M. Nagiec, and R.C. Dickson. Autoregulation of GAL4 transcription is essential for rapid growth of *Kluyveromyces lactis* on lactose and galactose. *Nucleic Acids Res.*, 21:4378–4382, 1993.
- [4] Sulakshana P. Mukherjee, Marcelo Behar, Harry A. Birnbaum, Alexander Hoffmann, Peter E. Wright, and Gourisankar Ghosh. Analysis of the RelA:CBP/p300 Interaction Reveals Its Involvement in NF- κ B-Driven Transcription. *PLOS Biology*, 11(9): e1001647, 9 2013. ISSN 1545-7885. doi: 10.1371/JOURNAL.PBIO.1001647. URL <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001647>.
- [5] Myong-Hee Sung, Ning Li, Qizong Lao, Rachel A. Gottschalk, Gordon L. Hager, and Iain D. C. Fraser. Switching of the Relative Dominance Between Feedback Mechanisms in Lipopolysaccharide-Induced NF- κ B Signaling. *Science Signaling*, 7(308):ra6–ra6, 1 2014. ISSN 1945-0877. doi: 10.1126/SCISIGNAL.2004764. URL

<https://stke.sciencemag.org/content/7/308/ra6>

<https://stke.sciencemag.org/content/7/308/ra6.abstract>.

- [6] C.K. Dalal, L. Cai, Y. Lin, K. Rahbar, and M.B. Elowitz. Pulsatile dynamics in the yeast proteome. *Curr. Biol.*, 24:2189–2194, 2014.
- [7] J.E. Purvis and G. Lahav. Encoding and decoding cellular information through signaling dynamics. *Cell*, 152:945–956, 2013.
- [8] L. Ashall, C.A. Horton, D.E. Nelson, P. Paszek, C.V. Harper, K. Sillitoe, S. Ryan, D.G. Spiller, J.F. Unitt, D.S. Broomhead, D.B. Kell, D.A. Rand, V. Sée, and M.R.H. White. Pulsatile stimulation determines timing and specificity of NF-kappa B-dependent transcription. *Science (New York, N.Y.)*, 324(5924):242, 4 2009. doi: 10.1126/SCIENCE.1164860. URL </pmc/articles/PMC2785900/>
</pmc/articles/PMC2785900/?report=abstract>
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2785900/>.
- [9] K. Lane, D. Van Valen, M.M. DeFelice, D.N. Macklin, T. Kudo, A. Jaimovich, A. Carr, T. Meyer, D. Pe’er, S.C. Boutet, and M.W. Covert. 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.
- [10] Samuel Zambrano, Alessia Loffreda, Elena Carelli, Giacomo Stefanelli, Federica Colombo, Edouard Bertrand, Carlo Tacchetti, Alessandra Agresti, Marco E. Bianchi, Nacho Molina, and Davide Mazza. First Responders Shape a Prompt and Sharp NF- κ B-Mediated Transcriptional Response to TNF- α . *iScience*, 23(9):101529, 9 2020. ISSN 2589-0042. doi: 10.1016/J.ISCI.2020.101529.
- [11] Alexander E Davies, Michael Pargett, Stefan Siebert, Gerald Quon, Mina J Bissell, and John G Albeck. Article Systems-Level Properties of EGFR-RAS-ERK Signaling Amplify Local Signals to Generate Dynamic Gene Expression Heterogeneity. 2020.

doi: 10.1016/j.cels.2020.07.004. URL

<https://doi.org/10.1016/j.cels.2020.07.004>.

- [12] Long Cai, Chiraj K. Dalal, and Michael B. Elowitz. Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* 2008 455:7212, 455(7212): 485–490, 9 2008. ISSN 1476-4687. doi: 10.1038/nature07292. URL <https://www.nature.com/articles/nature07292>.
- [13] Pimkhuan Hannanta-Anan and Brian Y. Chow. Optogenetic Control of Calcium Oscillation Waveform Defines NFAT as an Integrator of Calcium Load. *Cell Systems*, 2(4):283–288, 4 2016. ISSN 2405-4712. doi: 10.1016/J.CELS.2016.03.010.
- [14] Marie D Harton, Woo Seuk Koh, Amie D Bunker, Abhyudai Singh, and Eric Batchelor. p53 pulse modulation differentially regulates target gene promoters to regulate cell fate decisions. *Molecular Systems Biology*, 15(9):e8685, 9 2019. ISSN 1744-4292. doi: 10.15252/MSB.20188685. URL <https://www.embopress.org/doi/full/10.15252/msb.20188685> <https://www.embopress.org/doi/abs/10.15252/msb.20188685>.
- [15] Alexander G. Goglia, Maxwell Z. Wilson, Siddhartha G. Jena, Jillian Silbert, Lena P. Basta, Danelle Devenport, and Jared E. Toettcher. A Live-Cell Screen for Altered Erk Dynamics Reveals Principles of Proliferative Control. *Cell Systems*, 10(3): 240–253, 3 2020. ISSN 2405-4712. doi: 10.1016/J.CELS.2020.02.005.
- [16] A.S. Hansen and E.K. O’Shea. Promoter decoding of transcription factor dynamics involves a trade-off between noise and control of gene expression. *Mol. Syst. Biol.*, 9: 704, 2013.
- [17] A.S. Hansen and E.K. O’Shea. cis determinants of promoter threshold and activation timescale. *Cell Rep.*, 12:1226–1233, 2015.

- [18] A.S. Hansen and E.K. O’Shea. Limits on information transduction through amplitude and frequency regulation of transcription factor activity. *eLife*, 4:e06559, 2015.
- [19] Antonina Hafner, Jacob Stewart-Ornstein, Jeremy E Purvis, William C Forrester, Martha L Bulyk, and Galit Lahav. p53 pulses lead to distinct patterns of gene expression albeit similar DNA-binding dynamics. *Nature Structural & Molecular Biology* 2017 24:10, 24(10):840–847, 8 2017. ISSN 1545-9985. doi: 10.1038/nsmb.3452. URL <https://www.nature.com/articles/nsmb.3452>.
- [20] N. Hao and E.K. O’Shea. Signal-dependent dynamics of transcription factor translocation controls gene expression. *Nat. Struct. Mol. Biol.*, 19:31–39, 2011.
- [21] N. Hao, B.A. Budnik, J. Gunawardena, and E.K. O’Shea. Tunable signal processing through modular control of transcription factor translocation. *Science*, 339:460–464, 2013.
- [22] L. Verstrepen, T. Bekaert, T.-L. Chau, J. Tavernier, A. Chariot, and R. Beyaert. TLR-4, IL-1R and TNF-R signaling to NF- κ B: variations on a common theme. *Cellular and Molecular Life Sciences* 2008 65:19, 65(19):2964–2978, 6 2008. ISSN 1420-9071. doi: 10.1007/S00018-008-8064-8. URL <https://link.springer.com/article/10.1007/s00018-008-8064-8>.
- [23] Eric Batchelor, Alexander Loewer, Caroline Mock, and Galit Lahav. Stimulus-dependent dynamics of p53 in single cells. *Molecular Systems Biology*, 7(1), 1 2011. ISSN 1744-4292. doi: 10.1038/msb.2011.20.
- [24] M.W. Covert, T.H. Leung, J.E. Gaston, and D. Baltimore. Achieving stability of lipopolysaccharide-induced NF-kappaB activation. *Science*, 309:1854–1857, 2005.
- [25] Y. Gotoh, E. Nishida, T. Yamashita, M. Hoshi, M. Kawakami, and H. Sakai. Microtubule-associated-protein (MAP) kinase activated by nerve growth factor and

- epidermal growth factor in PC12 cells. Identity with the mitogen-activated MAP kinase of fibroblastic cells. *Eur. J. Biochem.*, 193:661–669, 1990.
- [26] A. Hoffmann, A. Levchenko, M.L. Scott, and D. Baltimore. The I κ B-NF- κ B signaling module: temporal control and selective gene activation. *Science*, 298:1241–1245, 2002.
- [27] D.E. Nelson, A.E.C. Ihekweba, M. Elliott, J.R. Johnson, C.A. Gibney, B.E. Foreman, G. Nelson, V. See, C.A. Horton, D.G. Spiller, and et al. Oscillations in NF- κ B signaling control the dynamics of gene expression. *Science*, 306:704–708, 2004.
- [28] T.T. Nguyen, J.C. Scimeca, C. Filloux, P. Peraldi, J.L. Carpentier, and E. Van Obberghen. Co-regulation of the mitogen-activated protein kinase, extracellular signal-regulated kinase 1, and the 90-kDa ribosomal S6 kinase in PC12 cells. Distinct effects of the neurotrophic factor, nerve growth factor, and the mitogenic factor, epidermal growth factor. *J. Biol. Chem.*, 268:9803–9810, 1993.
- [29] J.E. Purvis, K.W. Karhohs, C. Mock, E. Batchelor, A. Loewer, and G. Lahav. p53 dynamics control cell fate. *Science*, 336:1440–1444, 2012.
- [30] S. Tay, J.J. Hughey, T.K. Lee, T. Lipniacki, S.R. Quake, and M.W. Covert. Single-cell NF- κ B dynamics reveal digital activation and analogue information processing. *Nature*, 466:267–271, 2010.
- [31] S. Traverse, N. Gomez, H. Paterson, C. Marshall, and P. Cohen. Sustained activation of the mitogen-activated protein (MAP) kinase cascade may be required for differentiation of PC12 cells. Comparison of the effects of nerve growth factor and epidermal growth factor. *Biochem. J.*, 288:351–355, 1992.
- [32] S.L. Werner, J.D. Kearns, V. Zadorozhnaya, C. Lynch, E. O’Dea, M.P. Boldin, A. Ma, D. Baltimore, and A. Hoffmann. Encoding NF- κ B temporal control in

- response to TNF: distinct roles for the negative regulators IkappaBalpha and A20. *Genes Dev.*, 22:2093–2101, 2008.
- [33] Y.T. Chong, J.L. Koh, H. Friesen, S.K. Duffy, M.J. Cox, A. Moses, J. Moffat, C. Boone, and B.J. Andrews. Yeast proteome dynamics from single cell imaging and automated analysis. *Cell*, 161:1413–1424, 2015.
- [34] A.A. Granados, J.M.J. Pietsch, S.A. Cepeda-Humerez, I.L. Farquhar, G. Tkačik, and P.S. Swain. Distributed and dynamic intracellular organization of extracellular information. *Proc. Natl. Acad. Sci. USA*, 115:6088–6093, 2018.
- [35] Zohreh AkhavanAghdam, Joydeb Sinha, Omar P Tabbaa, and Nan Hao. Dynamic control of gene regulatory logic by seemingly redundant transcription factors. *eLife*, 5, 9 2016. ISSN 2050-084X. doi: 10.7554/eLife.18458.
- [36] Y.T. Lin and C.R. Doering. Gene expression dynamics with stochastic bursts: construction and exact results for a coarse-grained model. *Phys. Rev. E*, 93:022409, 2016.
- [37] S. Sen, Z. Cheng, K.M. Sheu, Y.H. Chen, and A. Hoffmann. Gene regulatory strategies that decode the duration of NF κ B dynamics contribute to LPS- versus TNF-specific gene expression. *Cell Syst.*, 10:169–182, 2020.
- [38] C. Garmendia-Torres, A. Goldbeter, and M. Jacquet. Nucleocytoplasmic oscillations of the yeast transcription factor Msn2: evidence for periodic PKA activation. *Curr. Biol.*, 17:1044–1049, 2007.
- [39] D. Pincus, A. Aranda-Díaz, I.A. Zuleta, P. Walter, and H. El-Samad. Delayed Ras/PKA signaling augments the unfolded protein response. *Proc. Natl. Acad. Sci. USA*, 111:14800–14805, 2014.

- [40] S. Shimizu-Sato, E. Huq, J.M. Tepperman, and P.H. Quail. A light-switchable gene promoter system. *Nat. Biotechnol.*, 20:1041–1044, 2002.
- [41] D. Strickland, Y. Lin, E. Wagner, C.M. Hope, J. Zayner, C. Antoniou, T.R. Sosnick, E.L. Weiss, and M. Glotzer. TULIPs: tunable, light-controlled interacting protein tags for cell biology. *Nat. Methods*, 9:379–384, 2012.
- [42] J.E. Toettcher, C.A. Voigt, O.D. Weiner, and W.A. Lim. The promise of optogenetics in cell biology: interrogating molecular circuits in space and time. *Nat. Methods*, 8:35–38, 2011.
- [43] J.E. Toettcher, O.D. Weiner, and W.A. Lim. Using optogenetics to interrogate the dynamic control of signal transmission by the Ras/erk module. *Cell*, 155:1422–1434, 2013.
- [44] D. Niopek, D. Benzinger, J. Roensch, T. Draebing, P. Wehler, R. Eils, and B. Di Ventura. Engineering light-inducible nuclear localization signals for precise spatiotemporal control of protein dynamics in living cells. *Nat. Commun.*, 5:4404, 2014.
- [45] T.A. Redchuk, E.S. Omelina, K.G. Chernov, and V.V. Verkhusha. Near-infrared optogenetic pair for protein regulation and spectral multiplexing. *Nat. Chem. Biol.*, 13:633–639, 2017.
- [46] H. Yumerefendi, D.J. Dickinson, H. Wang, S.P. Zimmerman, J.E. Bear, B. Goldstein, K. Hahn, and B. Kuhlman. Control of protein activity and cell fate specification via light-mediated nuclear translocation. *PLoS One*, 10:e0128443, 2015.
- [47] A. Gautier, D.P. Nguyen, H. Lusic, W. An, A. Deiters, and J.W. Chin. Genetically encoded photocontrol of protein localization in mammalian cells. *J. Am. Chem. Soc.*, 132:4086–4088, 2010.

- [48] H. Wang, M. Vilela, A. Winkler, M. Tarnawski, I. Schlichting, H. Yumerefendi, B. Kuhlman, R. Liu, G. Danuser, and K.M. Hahn. LOVTRAP: an optogenetic system for photoinduced protein dissociation. *Nat. Methods*, 13:755–758, 2016.
- [49] H. Yumerefendi, H. Wang, D.J. Dickinson, A.M. Lerner, P. Malkus, B. Goldstein, K. Hahn, and B. Kuhlman. Light-dependent cytoplasmic recruitment enhances the dynamic range of a nuclear import photoswitch. *Chembiochem*, 19:1319–1325, 2018.
- [50] S. Kosugi, M. Hasebe, N. Matsumura, H. Takashima, E. Miyamoto-Sato, M. Tomita, and H. Yanagawa. Six classes of nuclear localization signals specific to different binding grooves of importin alpha. *J. Biol. Chem.*, 284:478–485, 2009.
- [51] S.P. Heximer, H. Lim, J.L. Bernard, and K.J. Blumer. Mechanisms governing subcellular localization and function of human RGS2. *J. Biol. Chem.*, 276:14195–14203, 2001.
- [52] A.S. Khalil, T.K. Lu, C.J. Bashor, C.L. Ramirez, N.C. Pyenson, J.K. Joung, and J.J. Collins. A synthetic biology framework for programming eukaryotic transcription functions. *Cell*, 150:647–658, 2012.
- [53] A.P. Gasch, P.T. Spellman, C.M. Kao, O. Carmel-Harel, M.B. Eisen, G. Storz, D. Botstein, and P.O. Brown. Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell*, 11:4241–4257, 2000.
- [54] N. Vardi, S. Levy, Y. Gurvich, T. Polacheck, M. Carmi, D. Jaitin, I. Amit, and N. Barkai. Sequential feedback induction stabilizes the phosphate starvation response in budding yeast. *Cell Rep.*, 9:1122–1134, 2014.
- [55] A. Stathopoulos-Gerontides, J.J. Guo, and M.S. Cyert. Yeast calcineurin regulates nuclear localization of the Crz1p transcription factor through dephosphorylation. *Genes Dev*, 13:798–803, 1999.

- [56] A.M. Stathopoulos and M.S. Cyert. Calcineurin acts through the CRZ1/TCN1-encoded transcription factor to regulate gene expression in yeast. *Genes Dev.*, 11:3432–3444, 1997.
- [57] H. Yoshimoto, K. Saltsman, A.P. Gasch, H.X. Li, N. Ogawa, D. Botstein, P.O. Brown, and M.S. Cyert. Genome-wide analysis of gene expression regulated by the calcineurin/Crz1p signaling pathway in *Saccharomyces cerevisiae*. *J. Biol. Chem.*, 277:31079–31088, 2002.
- [58] R. Christiano, N. Nagaraj, F. Fröhlich, and T.C. Walther. Global proteome turnover analyses of the yeasts *S. cerevisiae* and *S. pombe*. *Cell Rep.*, 9:1959–1965, 2014.
- [59] A. Bregman, M. Avraham-Kelbert, O. Barkai, L. Duek, A. Guterman, and M. Choder. Promoter elements regulate cytoplasmic mRNA decay. *Cell*, 147:1473–1483, 2011.
- [60] T. Trcek, D.R. Larson, A. Moldón, C.C. Query, and R.H. Singer. Single-molecule mRNA decay measurements reveal promoter-regulated mRNA stability in yeast. *Cell*, 147:1484–1497, 2011.
- [61] Y. Wang, C.L. Liu, J.D. Storey, R.J. Tibshirani, D. Herschlag, and P.O. Brown. Precision and functional specificity in mRNA decay. *Proc. Natl. Acad. Sci. USA*, 99:5860–5865, 2002.
- [62] N. Dillon and R. Festenstein. Unravelling heterochromatin: competition between positive and negative factors regulates accessibility. *Trends Genet.*, 18:252–258, 2002.
- [63] F. Spitz and E.E.M. Furlong. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.*, 13:613–626, 2012.
- [64] C.R. Lickwar, F. Mueller, S.E. Hanlon, J.G. McNally, and J.D. Lieb. Genome-wide

- protein-DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature*, 484:251–255, 2012.
- [65] J.M. Platt, P. Ryvkin, J.J. Wanat, G. Donahue, M.D. Ricketts, S.P. Barrett, H.J. Waters, S. Song, A. Chavez, K.O. Abdallah, and et al. Rap1 relocalization contributes to the chromatin-mediated gene expression profile and pace of cell senescence. *Genes Dev.*, 27:1406–1420, 2013.
- [66] Chao Cheng, Chong Shou, Kevin Y Yip, and Mark B Gerstein. Genome-wide analysis of chromatin features identifies histone modification sensitive and insensitive yeast transcription factors. *Genome Biology* 2011 12:11, 12(11):1–17, 11 2011. ISSN 1474-760X. doi: 10.1186/GB-2011-12-11-R111. URL <https://genomebiology.biomedcentral.com/articles/10.1186/gb-2011-12-11-r111>.
- [67] A.C. Gaupel, T. Begley, and M. Tenniswood. High throughput screening identifies modulators of histone deacetylase inhibitors. *BMC Genomics*, 15:528, 2014.
- [68] I. Steinfeld, R. Shamir, and M. Kupiec. A genome-wide analysis in *Saccharomyces cerevisiae* demonstrates the influence of chromatin modifiers on transcription. *Nat. Genet.*, 39:303–309, 2007.
- [69] Won-Ki Cho, Jan-Hendrik Spille, Micca Hecht, Choongman Lee, Charles Li, Valentin Grube, and Ibrahim I. Cisse. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science*, 361(6400):412–415, 7 2018. ISSN 0036-8075. doi: 10.1126/SCIENCE.AAR4199. URL <https://science.sciencemag.org/content/361/6400/412>
<https://science.sciencemag.org/content/361/6400/412.abstract>.
- [70] M.E. Lee, W.C. DeLoache, B. Cervantes, and J.E. Dueber. A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synth. Biol.*, 4:975–986, 2015.

- [71] S. Lee, W.A. Lim, and K.S. Thorn. Improved blue, green, and red fluorescent protein tagging vectors for *S. cerevisiae*. *PLoS One*, 8:e67902, 2013.
- [72] J.M. Gardner and S.L. Jaspersen. Manipulating the yeast genome: deletion, mutation, and tagging by PCR. *Methods Mol. Biol.*, 1205:45–78, 2014.
- [73] L.J. Bugaj, A.J. Sabnis, A. Mitchell, J.E. Garbarino, J.E. Toettcher, T.G. Bivona, and W.A. Lim. Cancer mutations and targeted drugs can disrupt dynamic signal encoding by the Ras-Erk pathway. *Science*, 361:6405, 2018.
- [74] P. Harrigan, H.D. Madhani, and H. El-Samad. Real-time genetic compensation defines the dynamic demands of feedback control. *Cell*, 175:877–886, 2018.
- [75] R.A. Greenstein, S.K. Jones, E.C. Spivey, J.R. Rybarski, I.J. Finkelstein, and B. Al-Sady. Noncoding RNA-nucleated heterochromatin spreading is intrinsically labile and requires accessory elements for epigenetic stability. *eLife*, 7:e32948, 2018.
- [76] Yong Chi, Michael J. Huddleston, Xiaolong Zhang, Richard A. Young, Roland S. Annan, Steven A. Carr, and Raymond J. Deshaies. Negative regulation of Gcn4 and Msn2 transcription factors by Srb10 cyclin-dependent kinase. *Genes and Development*, 15(9):1078–1092, 5 2001. doi: 10.1101/GAD.867501.
- [77] E. Durchschlag, W. Reiter, G. Ammerer, and C. Schüller. Nuclear localization destabilizes the stress-regulated transcription factor Msn2. *J. Biol. Chem.*, 279: 55425–55432, 2004.
- [78] P. Sen, W. Dang, G. Donahue, J. Dai, J. Dorsey, X. Cao, W. Liu, K. Cao, R. Perry, J.Y. Lee, and et al. H3K36 methylation promotes longevity by enhancing transcriptional fidelity. *Genes Dev.*, 29:1362–1376, 2015.
- [79] C. Malabat, F. Feuerbach, L. Ma, C. Saveanu, and A. Jacquier. Quality control of

- transcription start site selection by nonsense-mediated-mRNA decay. *eLife*, 4:e06722, 2015.
- [80] Rita Moreno, Jürgen-Markus Sobotzik, Christian Schultz, and M. Lienhard Schmitz. Specification of the NF- κ B transcriptional response by p65 phosphorylation and TNF-induced nuclear translocation of IKK ϵ . *Nucleic Acids Research*, 38(18):6029, 5 2010. doi: 10.1093/NAR/GKQ439. URL [/pmc/articles/PMC2952868/](#)
[/pmc/articles/PMC2952868/?report=abstract](#)
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2952868/>.
- [81] Jan-Philipp Kruse and Wei Gu. Modes of p53 Regulation. *Cell*, 137(4):609–622, 5 2009. ISSN 0092-8674. doi: 10.1016/J.CELL.2009.04.050. URL
<http://www.cell.com/article/S009286740900511X/fulltext>
<http://www.cell.com/article/S009286740900511X/abstract>
[https://www.cell.com/cell/abstract/S0092-8674\(09\)00511-X](https://www.cell.com/cell/abstract/S0092-8674(09)00511-X).
- [82] Holger Buss, Anneke Dörrie, M. Lienhard Schmitz, Elke Hoffmann, Klaus Resch, and Michael Kracht. Constitutive and Interleukin-1-inducible Phosphorylation of p65 NF- κ B at Serine 536 Is Mediated by Multiple Protein Kinases Including I κ B Kinase (IKK)- α , IKK β , IKK, TRAF Family Member-associated (TANK)-binding Kinase 1 (TBK1), and an Unknown Kinase and Couples p65 to TATA-binding Protein-associated Factor II31-mediated Interleukin-8 Transcription. *Journal of Biological Chemistry*, 279(53):55633–55643, 12 2004. ISSN 0021-9258. doi: 10.1074/JBC.M409825200.
- [83] N. Yissachar, T.S. Sharar Fischler, A.A. Cohen, S. Reich-Zeliger, D. Russ, E. Shifrut, Z. Porat, and N. Friedman. Dynamic response diversity of NFAT isoforms in individual living cells. *Mol. Cell*, 49:322–330, 2013.
- [84] Minjun Son, Andrew G. Wang, Hsiung-Lin Tu, Marie Oliver Metzigg, Parthiv Patel,

- Kabir Husain, Jing Lin, Arvind Murugan, Alexander Hoffmann, and Savaş Tay. NF- κ B responds to absolute differences in cytokine concentrations. *Science Signaling*, 14(666):4382, 1 2021. ISSN 1945-0877. doi: 10.1126/SCISIGNAL.AAZ4382. URL <https://stke.sciencemag.org/content/14/666/eaaz4382>
<https://stke.sciencemag.org/content/14/666/eaaz4382.abstract>.
- [85] Erik W. Martin, Alicja Pacholewska, Heta Patel, Himanshu Dashora, and Myong Hee Sung. Integrative analysis suggests cell type-specific decoding of NF- κ B dynamics. *Science Signaling*, 13(620), 2 2020. doi: 10.1126/SCISIGNAL.AAX7195.
- [86] Maxwell Z. Wilson, Pavithran T. Ravindran, Wendell A. Lim, and Jared E. Toettcher. Tracing Information Flow from Erk to Target Gene Induction Reveals Mechanisms of Dynamic and Combinatorial Control. *Molecular Cell*, 67(5):757–769, 9 2017. ISSN 1097-2765. doi: 10.1016/J.MOLCEL.2017.07.016.
- [87] João Pedro Fonseca, Alain R. Bonny, G. Renuka Kumar, Andrew H. Ng, Jason Town, Qiu Chang Wu, Elham Aslankoochi, Susan Y. Chen, Galen Dods, Patrick Harrigan, Lindsey C. Osimiri, Amy L. Kistler, and Hana El-Samad. A Toolkit for Rapid Modular Construction of Biological Circuits in Mammalian Cells. *ACS Synthetic Biology*, 8(11):2593–2606, 11 2019. doi: 10.1021/ACSSYNBIO.9B00322. URL <https://pubs.acs.org/doi/abs/10.1021/acssynbio.9b00322>.
- [88] Shuang Zhao, Enze Jiang, Shuangshuang Chen, Yuan Gu, Anna Junjie Shanguan, Tangfeng Lv, Liguo Luo, and Zhenghong Yu. PiggyBac transposon vectors: the tools of the human gene encoding. *Translational Lung Cancer Research*, 5(1):120, 2016. doi: 10.3978/J.ISSN.2218-6751.2016.01.05. URL </pmc/articles/PMC4758974/>
</pmc/articles/PMC4758974/?report=abstract>
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4758974/>.
- [89] Lukasz J. Bugaj and Wendell A. Lim. High-throughput multicolor optogenetics in

- microwell plates. *Nature Protocols* 2019 14:7, 14(7):2205–2228, 6 2019. ISSN 1750-2799. doi: 10.1038/s41596-019-0178-y. URL <https://www.nature.com/articles/s41596-019-0178-y>.
- [90] Qihong Zhang, Sanjana Gupta, David L. Schipper, Gabriel J. Kowalczyk, Allison E. Mancini, James R. Faeder, and Robin E.C. Lee. NF- κ B Dynamics Discriminate between TNF Doses in Single Cells. *Cell Systems*, 5(6):638–645, 12 2017. ISSN 2405-4712. doi: 10.1016/J.CELS.2017.10.011.
- [91] Manfred Neumann and Michael Naumann. Beyond I κ Bs: alternative regulation of NF-KB activity. *The FASEB Journal*, 21(11):2642–2654, 9 2007. ISSN 1530-6860. doi: 10.1096/FJ.06-7615REV. URL <https://faseb.onlinelibrary.wiley.com/doi/full/10.1096/fj.06-7615rev>
<https://faseb.onlinelibrary.wiley.com/doi/abs/10.1096/fj.06-7615rev>
<https://faseb.onlinelibrary.wiley.com/doi/10.1096/fj.06-7615rev>.
- [92] Kim A. Ngo, Kensei Kishimoto, Jeremy Davis-Turak, Aditya Pimplaskar, Zhang Cheng, Roberto Spreafico, Emily Y. Chen, Amy Tam, Gourisankar Ghosh, Simon Mitchell, and Alexander Hoffmann. Dissecting the Regulatory Strategies of NF- κ B RelA Target Genes in the Inflammatory Response Reveals Differential Transactivation Logics. *Cell Reports*, 30(8):2758–2775, 2 2020. ISSN 2211-1247. doi: 10.1016/J.CELREP.2020.01.108.
- [93] Ping Rao, Mathew S. Hayden, Meixiao Long, Martin L. Scott, A. Philip West, Dekai Zhang, Andrea Oeckinghaus, Candace Lynch, Alexander Hoffmann, David Baltimore, and Sankar Ghosh. I κ B β acts to inhibit and activate gene expression during the inflammatory response. *Nature* 2010 466:7310, 466(7310):1115–1119, 8 2010. ISSN 1476-4687. doi: 10.1038/nature09283. URL <https://www.nature.com/articles/nature09283>.

- [94] Mude Shi, Weiwen Deng, Enguang Bi, Kairui Mao, Yongyong Ji, Guomei Lin, Xiaodong Wu, Zhiyun Tao, Zhenhu Li, Xinfen Cai, Shuhui Sun, Charlie Xiang, and Bing Sun. TRIM30 α negatively regulates TLR-mediated NF- κ B activation by targeting TAB2 and TAB3 for degradation. *Nature Immunology* 2008 9:4, 9(4): 369–377, 3 2008. ISSN 1529-2916. doi: 10.1038/ni1577. URL <https://www.nature.com/articles/ni1577>.
- [95] Anne Müller, André Hennig, Sebastian Lorscheid, Paula Grondona, Klaus Schulze-Osthoff, Stephan Hailfinger, and Daniela Kramer. I κ B ζ is a key transcriptional regulator of IL-36–driven psoriasis-related gene expression in keratinocytes. *Proceedings of the National Academy of Sciences*, 115(40): 10088–10093, 10 2018. ISSN 0027-8424. doi: 10.1073/PNAS.1801377115. URL <https://www.pnas.org/content/115/40/10088> <https://www.pnas.org/content/115/40/10088.abstract>.
- [96] Brad H. Rovin, Jennifer A. Dickerson, Laura C. Tan, and Caroline A. Hebert. Activation of nuclear factor- κ B correlates with MCP-1 expression by human mesangial cells. *Kidney International*, 48(4):1263–1271, 10 1995. ISSN 0085-2538. doi: 10.1038/KI.1995.410.
- [97] Fengming Luan, Xiaomei Li, Xiaojing Cheng, Longtao Huangfu, Jing Han, Ting Guo, Hong Du, Xianzi Wen, and Jiafu Ji. TNFRSF11B activates Wnt/ β -catenin signaling and promotes gastric cancer progression. *International Journal of Biological Sciences*, 16(11):1956, 2020. doi: 10.7150/IJBS.43630. URL [/pmc/articles/PMC7211174/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7211174/) [/pmc/articles/PMC7211174/?report=abstract](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7211174/?report=abstract) <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7211174/>.
- [98] I Kastrati, E Canestrari, and J Frasar. PHLDA1 expression is controlled by an estrogen receptor-NF κ B-miR-181 regulatory loop and is essential for formation of ER+ mammospheres. *Oncogene* 2015 34:18, 34(18):2309–2316, 6 2014. ISSN

- 1476-5594. doi: 10.1038/onc.2014.180. URL
<https://www.nature.com/articles/onc2014180>.
- [99] Ji-Woong Kwon, Hyuk-Kwon Kwon, Hyeon-Jun Shin, Yong-Min Choi, Muhammad Ayaz Anwar, and Sangdun Choi. Activating transcription factor 3 represses inflammatory responses by binding to the p65 subunit of NF- κ B. *Scientific Reports* 2015 5:1, 5(1):1–9, 9 2015. ISSN 2045-2322. doi: 10.1038/srep14470. URL
<https://www.nature.com/articles/srep14470>.
- [100] Christophe Genolini and Bruno Falissard. KmL: k-means for longitudinal data. *Computational Statistics* 2009 25:2, 25(2):317–328, 11 2009. ISSN 1613-9658. doi: 10.1007/S00180-009-0178-4. URL
<https://link.springer.com/article/10.1007/s00180-009-0178-4>.
- [101] Adewunmi Adelaja, Brooks Taylor, Katherine M Sheu, Yi Liu, Stefanie Luecke, and Alexander Hoffmann Correspondence. Six distinct NF- κ B signaling codons convey discrete information to distinguish stimuli and enable appropriate macrophage responses. *Immunity*, 54:916–930, 2021. doi: 10.1016/j.immuni.2021.04.011. URL
<https://doi.org/10.1016/j.immuni.2021.04.011>.
- [102] Miriam V. Gutschow, John C. Mason, Keara M. Lane, Inbal Maayan, Jacob J. Hughey, Bryce T. Bajar, Debha N. Amatya, Sean D. Valle, and Markus W. Covert. Combinatorial processing of bacterial and host-derived innate immune stimuli at the single-cell level. <https://doi.org/10.1091/mbc.E18-07-0423>, 30(2):282–292, 1 2019. doi: 10.1091/MBC.E18-07-0423. URL
<https://www.molbiolcell.org/doi/abs/10.1091/mbc.E18-07-0423>.
- [103] Alexander E. Davies, Michael Pargett, Stefan Siebert, Taryn E. Gillies, Yongin Choi, Savannah J. Tobin, Abhineet R. Ram, Vaibhav Murthy, Celina Juliano, Gerald Quon, Mina J. Bissell, and John G. Albeck. Systems-Level Properties of

- EGFR-RAS-ERK Signaling Amplify Local Signals to Generate Dynamic Gene Expression Heterogeneity. *Cell Systems*, 11(2):161–175, 8 2020. ISSN 2405-4712. doi: 10.1016/J.CELS.2020.07.004.
- [104] Oliver S. Thomas, Maximilian Hörner, and Wilfried Weber. A graphical user interface to design high-throughput optogenetic experiments with the optoPlate-96. *Nature Protocols* 2020 15:9, 15(9):2785–2787, 7 2020. ISSN 1750-2799. doi: 10.1038/s41596-020-0349-x. URL <https://www.nature.com/articles/s41596-020-0349-x>.
- [105] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T.R. Gingeras. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29:15–21, 2013.
- [106] Simon Anders, Paul Theodor Pyl, and Wolfgang Huber. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2):166, 1 2015. doi: 10.1093/BIOINFORMATICS/BTU638. URL [/pmc/articles/PMC4287950/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4287950/) [/pmc/articles/PMC4287950/?report=abstract](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4287950/?report=abstract) <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4287950/>.
- [107] M.I. Love, W. Huber, and S. Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, 15:550, 2014.
- [108] Martin Weigert, Uwe Schmidt, Robert Haase, Ko Sugawara, and Gene Myers. Star-convex Polyhedra for 3D Object Detection and Segmentation in Microscopy. *The IEEE Winter Conference on Applications of Computer Vision (WACV)*, 3 2020. URL <https://github.com/>.
- [109] Uwe Schmidt, Martin Weigert, Coleman Broaddus, and Gene Myers. Cell Detection with Star-convex Polygons. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11071

LNCS:265–273, 6 2018. doi: 10.1007/978-3-030-00934-2_30. URL
<https://arxiv.org/abs/1806.03535v2>.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:

3DB0A2A7FEE54AC... Author Signature

9/1/2021
Date