ABSTRACT

Title of dissertation	DISSECTING THE GENE REGULATORY FUNCTION OF THE MYC ONCOGENE WITH SINGLE-MOLECULE IMAGING
	Simona Patange Doctor of Philosophy, 2020
Dissertation directed by	Dr. Michelle Girvan, Department of Physics, UMD
	Dr. Daniel R. Larson National Cancer Institute, NIH

The MYC oncogene contributes to an estimated 100,000 cancer-related deaths annually in the United States and is associated with aggressive tumor progression and poor clinical outcome. MYC is a nuclear transcription factor that regulates a myriad of cellular activities and has direct interactions with hundreds of proteins, which has made a unified understanding of its function historically difficult.

In recent years, several groups have put forth a new hypothesis that questions the prevailing view of MYC as a gene-specific transcription factor and instead envision it as a global amplifier of gene expression. Instead of being an on/off switch for transcription, MYC is proposed to act as a 'volume knob' to amplify and sustain the active gene expression program in a cell. The scope of the amplifier model remains controversial in part because studies of MYC largely consist of cell population-based measurements obtained at fixed timepoints, which makes distinguishing direct from indirect consequences on gene

expression difficult. A high-temporal, high-spatial precision viewpoint of how MYC acts in single living cells does not exist.

To evaluate the competing hypotheses of MYC function, we developed a single-cell assay for precisely controlling MYC and interrogating the effects on transcription in living cells. We engineered 'Pi-MYC,' an optogenetic variant of MYC that is biologically active, can be visualized under the microscope, and can be controlled with light. We combined Pi-MYC with single-molecule imaging methods to obtain the first real-time observations of how MYC affects RNA production and transcription factor mobility in single cells. We show that MYC increases the duration of active periods of genes population-wide, and globally affects the binding dynamics of core transcription factors involved in RNA polymerase II transcription complex assembly and productive elongation. These findings provide living, single-cell evidence of MYC as a global amplifier of gene expression, and suggests the mechanism is by stabilizing the active period of a gene through interactions with core transcription machinery.

DISSECTING THE GENE REGULATORY FUNCTION OF THE MYC ONCOGENE WITH SINGLE-MOLECULE IMAGING

by

Simona Patange

Dissertation submitted to Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2020

Dissertation Committee:
Dr. Michelle Girvan, Chair
Dr. Daniel Larson
Dr. Sergei Sukharev
Dr. Arpita Upadhyaya
Dr. Theodore Kwaku Dayie, Dean's Representative

Dedication

To my grandparents,

Balkrishna Patange	*Vijaya Patange*
1925-1999	1930-2017
Harish Thakkar	*Nirmala Thakkar*
1928-2019	1933-2019

Acknowledgements

I would like to thank the following people for the support of this work. First, I wish to express my sincere gratitude to my advisor Daniel Larson for the privilege of being his first graduate student, and for the opportunity to work on such a fascinating and significant research question. I am truly grateful for his unwavering enthusiasm for my project and the development of my scientific career. Dan sets the bar high for gracious professionalism and scientific rigor, and I shall forever strive to emulate these qualities as I progress in my career. I also wish to thank my co-advisors Michelle Girvan and David Levens for their support. Our joint meetings were always the highlight of my week and I enjoyed learning how to think through different perspectives on the research problems I needed to tackle. I always left our meetings with a greater clarity of mind and having renewed strength to keep pressing forward. Thank you Dan, Michelle, and Dave for your optimism and confidence in me.

Second, I wish to thank my colleagues. The members of the Larson lab have been my strongest supporters throughout my Ph.D. journey and I am indebted to them for helping me grow as a scientist and critical thinker. I especially want to thank the members who helped me get started in my first years with the lab- Murali Palangat, Yihan Wan, Joseph Rodriguez, Huimin Chen, Tineke Lenstra, Antoine Coulon, and Matthew Ferguson. I will never forget the good times we had and I will think of you every time I have a taco salad or basket of curry fries. I further wish to thank my collaborators Tatiana Karpova and David Ball of the NCI Optical Microscopy Core for their guidance with the SMT experiments, as well as Gianluca Pegoraro and Prabhakar Gudla of the NCI HiTIF Core for use of the equipment and assistance with the high-throughput analysis pipelines. I am also grateful to everyone in the LRBGE branch of NCI for making the lab a fun and intellectually stimulating place to work. Thank

you for me helping with everything from tracking down reagents and understanding protocols, to lending career advice and moral support.

Third, I wish to thank the programs that made my partnership between UMD and NIH possible: The UMD Biophysics Graduate Program, the Institute of Physical Science & Technology, the NCI-UMD Partnership Program, and the COMBINE Fellowship Program. I especially wish to thank my committee members Sergei Sukharev, Arpita Upadhyaya, and Kwaku Dayie for their constructive assessments of my dissertation project. My sincere thanks also to Souad Nejjar, Stephanie Noel, and Lucia Croce, who provided the majority of my administrative support during my time in graduate school. I am grateful to have been a part of these excellent interdisciplinary programs and wish them continued success.

I would not be here without the mentors who nurtured my love of science well before I made the decision to pursue a graduate degree. I wish to thank my seventh-grade science teacher Laurie Denio; my undergraduate mentors Tania Vu and Katye Fichter at the Oregon Health & Science University; and my postbaccalaureate mentor Rosa Puertollano at the National Heart Lung and Blood Institute, for investing in my early education and inspiring me to continue onwards.

Lastly, I wish to thank my family for their love and support— foremost my parents Kalpana and Prasad for nurturing my curiosity and providing every resource for my education when I was growing up; my brother Ronak for being the best sidekick a sister could ask for; and my partner Robert for being my steadfast anchor during my Ph.D. journey. Thank you to all my family and friends for cheering me on every step of the way— I truly could not have done it without you.

Table of Contents

List of Figures

List of Abbreviations

E2	Beta-estradiol, estrogen
FOV	Field of view
GFP	Green fluorescent protein
GR	Glucocorticoid receptor
H2B	Histone 2B
KPNB1	Karyopherin Subunit Beta 1
MED1	Mediator complex subunit 1
MYH9	Myosin heavy chain 9
Pi-mCherry	Photo-inducible mCherry
Pi-MYC	Photo-inducible MYC-mCherry
RAB7A	Ras-associated binding protein 7A
RNAPII	Pol II, RNA polymerase II
RPAP3	RNA polymerase II-associated protein 3
smFISH	Single-molecule fluoresence in-situ hybridization
SMT	Single-molecule tracking
SPT5	Transcription elongation factor SPT5
TBP	TATA binding protein
TF	Transcription factor
TS	Transcription site

Chapter 1

Introduction

1.1 Overview of Thesis

Heterogeneous gene expression has been proposed to play a fundamental role in cancer progression [59]. The causes and consequences of this variability remain largely unknown however, and represent a major challenge in cell biology. Within recent years it has become clear that gene expression heterogeneity is not static but dynamic: expression of a gene of interest can fluctuate on timescales of minutes to days [64]. Over the last few decades, parallel advances in fluorescence imaging and computational modeling have made it possible to visualize and interpret the dynamic nature of gene expression. Two experimental approaches based on direct observation of RNA in single cells—smFISH and live-cell imaging with the MS2 stem loop system—led to the current view that most eukaryotic genes are infrequently transcribed in stochastic "bursts" of RNA synthesis interspersed with long periods of inactivity [106]. These approaches of imaging transcription by directly observing RNA is a true measure of gene regulation, uncoupled from downstream steps such as mRNA export, translation, and degradation.

In this thesis, I utilized single-molecule imaging techniques to elucidate the mechanism of action of the oncogenic transcription factor c-MYC (MYC), which is overexpressed in the majority of cancers. For the last 30 years, MYC was believed to execute a gene specific program which propelled cells away from apoptosis and towards growth, cell cycle progression, and de-differentiation— eventually resulting in oncogenic transformation. Recently, several models have emerged that question the prevailing view of MYC as a 'gene-specific' transcription factor and instead envision it as a 'global amplifier' of the existing gene expression program in the cell. Studies on the effects of MYC on gene expression largely consist of cell population measurements at fixed timepoints, or "snapshot views of the average"—A real-time view of MYC activity in living cells does not exist. We sought to critically evaluate these paradigms of MYC function—the gene-specific and global amplifier models—using single-molecule imaging techniques in human cell lines. This project had the following specific aims:

1. Control the onset of MYC overexpression in living cells with high spatio-temporal precision.
2. Visualize and measure the effects of MYC on transcription kinetics with single-molecule imaging methods.

To achieve these aims, we engineered an optogenetic variant of MYC ('Pi-MYC') that could be controlled with blue-green wavelengths of light and visualized with a fluorescence microscope. We combined Pi-MYC with single-molecule imaging of RNA and protein to obtain the first real-time observations of how MYC affects gene expression events and transcription factor binding dynamics in living human cells. In Chapter 1, I provide an introduction to MYC biology and present the motivating perspective of my dissertation with a recent literature review on single-cell gene expression methods. In Chapter 2, I provide detailed protocols on smFISH and the MS2 system, the RNA imaging methods that are used in this thesis for single-cell gene expression analysis. In Chapter 3, I describe the engineering and validation of Pi-MYC. In Chapter 4, I present the results of my work using single-cell imaging to

probe the effects of the MYC oncogene on gene expression and transcription factor binding dynamics. Chapter 5 concludes with a discussion of the major findings and future outlook.

1.2 MYC Biology

c-MYC (MYC) belongs to a class of oncogenic transcription factors that are over-expressed in the majority of malignancies [135]. MYC is associated with aggressive tumor progression and poor patient survival [124]. MYC is widely present in normal proliferating somatic cells and responds to mitogens and growth factors to regulate growth and general cell proliferation. MYC expression is deregulated and elevated in cancer cells and has been shown to be a driver of cancer in mouse models [1]. MYC deregulation occurs through various means such as retroviral transduction, chromosomal translocation, gene amplification, and alterations in signaling pathways that converge on MYC expression.

The primary mode of MYC deregulation in cancer is overexpression of the normal protein product (**Figure 1.1**), and multiple reports indicate that high MYC levels cause a global increase in total RNA levels [43, 52, 71, 92]. As such, MYC has direct interactions with hundreds of proteins involved in a diverse range of functional classes such as RNA processing, ribosome biogenesis, transcription regulation, mitosis, DNA damage and replication, histone modification, and mRNA post-transcriptional modifications [56].

Given its myriad activities, a unified understanding of MYC function has been elusive. In recent years several groups have put forth a new hypothesis that describes MYC as a global amplifier of gene expression. Genome-wide ChIP-seq analysis from two groups suggests that MYC is capable of binding all promoters but does so according to their expression output [71, 92]. Instead of being an on/off switch for transcription, MYC is proposed to act as a volume knob to amplify and sustain the active gene expression program in a cell. The main findings of the Lin

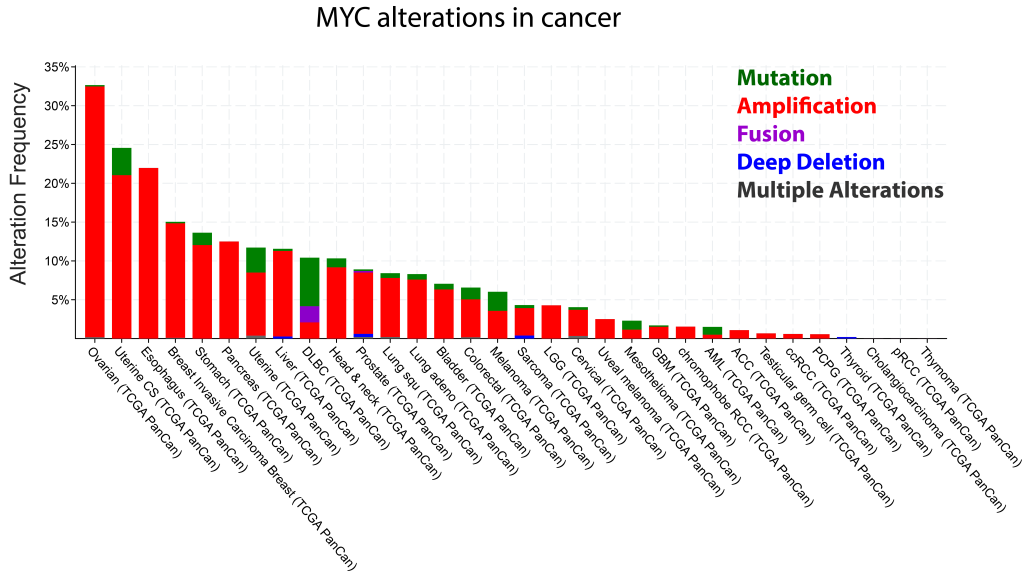


Figure 1.1: TCGA plot depicting the frequency and type of MYC alterations in cancer. *Accession date: April 11th, 2020 from cbioportal.org.*

study have since been replicated [70]. A further study on MYC occupancy at genes found the effects were limited according to the basal expression level of the gene: that is, promoters already saturated with MYC did not increase expression even if there was greater availability of MYC protein when it was induced to express at higher levels [75]. A recent study points to the promotion of RNAPII pause release through direct interactions with SPT5 as the mechanism of MYC amplification. [8]. A contrasting view is that MYC functions as a gene-specific transcription factor, with ubiquitous binding due to indirect effects of MYC binding enhancer regions of non-target genes when it is expressed at high levels [60, 115]. Thus, the scope of the amplifier model remains controversial in part because of direct and indirect consequences on gene expression. Moreover, most studies of MYC consist of cell population measurements obtained at single timepoints. A high-temporal, high-spatial precision viewpoint of how MYC acts in single living cells does not exist.

Here, we developed a single-cell assay for precisely controlling MYC and interrogating the effects on transcription in living cells. We engineered 'Pi-MYC', an optogenetic variant of MYC that is biologically active, can be visualized under

the microscope, and can be controlled with light. We used Pi-MYC along with RNA visualization in fixed and living cells to undertake a broad survey of genes that differed with respect to function, DNA features, basal expression level, and cellular context. All the genes we labeled showed transcriptional bursting, a consistent feature of human gene expression [111]. We found that MYC overexpression consistently increased the duration of active periods of genes population-wide and exerted its greatest effects on genes that did not already have high basal expression. These findings provide living, single-cell evidence of MYC as an amplifier of gene expression. Based on the measured import and export kinetics of Pi-MYC in non-transformed cells, we hypothesized that the mechanism of amplification could be through modulation of transcription factor kinetics. Using single-molecule tracking (SMT) experiments to measure the kinetics of glucocorticoid receptor (GR), TATA-box binding protein (TBP), a subunit of the Mediator complex (MED1), and a component of the DSIF complex (SPT5), we found that MYC globally affects the binding dynamics of core transcription factors involved in RNAPII complex assembly and productive elongation. Our data suggests that MYC increases the period over which productive initiation occurs through multi-faceted interactions with the core transcription machinery. In summary, our single-cell, real-time approach reveals that MYC amplifies gene expression across genes and cell types through direct modulation of the transcriptional active period.

1.3 Review on Single-Cell Gene Expression

Adapted from Simona Patange, Michelle Girvan, Daniel R. Larson. "Single-Cell Systems Biology: Probing the Basic Unit of Information Flow." Current Opinion in Systems Biology (2018);8:7-15.

1.3.1 Introduction

Gene expression is the link that connects environmental stimuli to the phenotypic responses of an organism. Early experiments aimed at identifying the key cellular

factors and genetic elements that regulated expression were in vitro and population based. Recent advances in sequencing and fluorescence microscopy now allow scientists to probe gene expression at the basic unit of information flow — the single cell. From a systems biology viewpoint, the methods of single-cell imaging and single-cell RNA sequencing (scRNA-seq) hold tremendous promise for providing an essential link between stimulus and response with the ability to directly visualize and quantify the production of RNA transcripts.

However, gene expression at the single-cell level is heterogeneous and stochastic — it varies across cells in a population and within a given cell over time. This phenomenon adds a perplexing challenge not only in being able to predict the expression behavior of a gene given known environmental inputs, but also the converse: to infer the state of the environment from a given gene's behavior. The field of single-cell gene expression is not a straightforward survey of how information gets transmitted from environment to gene product but rather grapples with a fundamentally philosophical question that often goes unappreciated: how do we get from randomness to order? At this point in time, the field appears directed towards the following questions: **1)** How does an organism coordinate a response within its body when the expression behavior of its individual cells is inherently stochastic? **2)** How much of gene expression heterogeneity is stable and represents a true biological subpopulation with different phenotypic properties? **3)** How does nuclear architecture contribute to gene regulation and variability? **4)** How do we generate a quantitative understanding of transcription and gene networks with computational modeling? In this chapter, I highlight recent literature which is at the forefront of addressing these questions.

1.3.2 Current developments in single-cell imaging and sequencing

The first methods to probe single-cell gene expression at the mRNA level were based on imaging, by labeling RNA transcripts in fixed cells via single-molecule

		Methods		
		scRNA-seq	smFISH	Live cell imaging
Features	Screen large N cells	✓	✓	✓ Recent (Wan et al. 2019)
	High throughput detection of multiple RNA species per sample	✓	✓ Recent (Moffit et al. 2016)	X
	Follow gene in real time	X	X	✓
	Observe endogenous human genes	✓	✓	✓ Recent (Wilson et al. 2015)
	Spatial information on expression	✓ Recent (Lee et al. 2015)	✓	✓

Figure 1.2: An evaluation of current features and limitations of single cell gene expression methods, with recent advances noted.

RNA FISH [35] (smFISH) and in living cells using the MS2-PP7 system [12, 55, 64]. Within a decade, imaging was followed by RNA sequencing (scRNA-seq), which allowed for in-depth gene expression profiling of individual cells [127, 138]. For the sake of brevity we refer readers to recent reviews [58,133] for an up-to-date history of the methodologies. In this section we outline the current advantages and limitations of live-cell imaging, smFISH, and scRNA-seq, and the recent work done to extend the capabilities of each technique (**Figure 1.2**).

Live-cell imaging with the MS2-PP7 stem loop system remains the most direct method for truly capturing the history of a gene's expression behavior, as it follows transcription activity of a living cell in real time. One primary limitation of the system is that it is low-throughput: of the three techniques, it is the most laborious and takes months to design and integrate constructs into the desired model system [21]. This time-intensive aspect has hindered the ability to make an extensive survey and classification of real-time transcription across genes. Our group has recently succeeded in the achievement of being able to visualize the transcription activity of thousands of individual genes across a cell population using a 'gene trap' method [136]. The development of this cell line is described in **Chapter 2**.

An inherent limitation of live-cell imaging is the existence of only two reliable orthogonal stem loop sequences (PP7 and MS2), which allows for at most two elements to be labeled within a single cell. Recent studies have used dual color labeling to their advantage to explore different phenomenon, such as splicing kinetics [20, 82], sense and antisense transcription from a single promoter on a yeast gene [67], two genes regulated by a common enhancer [40], and translation of individual RNAs [45]. Due to the difficulty of genetic manipulation in higher order eukaryotes, most studies on transcription kinetics are conducted with exogenous genes. However, homologous recombination has been successfully used to integrate stem loops into the endogenous loci of bacteria and yeast, and was recently used to integrate MS2 loops in mouse embryonic stem cells [72], and to visualize endogenous transcription dynamics of the pluripotency factors Nanog and Oct4 [95]. Within the next few years we expect advances in CRISPR/Cas9 gene editing to greatly aid in making endogenous integrations possible.

The second imaging-based technique, smFISH, has advantages to complement the limitations of live-cell imaging. The fluorescent oligo probes can be designed and commercially synthesized for virtually any endogenous gene. Because the sample is fixed, high throughput imaging can be used to routinely collect data on tens of thousands of cells for a given time point. Initially, smFISH was limited by the number of spectrally separable colors that could be used within a single cell. Earlier methods were able to visualize 10-30 RNA within a single cell through spectral separation [77], sequential labeling [78] or combinatorial labeling [69]. A recent technique (MERFISH) that uses sequential labeling increased the number of mRNA to 100-1000 [16], and a high-throughput version has also been recently published [88]. Progress has also been made in detecting small variations in nucleic acid sequence [68, 87]. Although smFISH is only a single time point measurement, advances in labeling and imaging are making the method increasingly high-throughput, allowing a significant number of genes and cells to be analyzed in a single experiment.

The main advantage of the final technique, scRNA-seq, is its breadth and depth: it produces extensive gene expression profiles that quantify the variation in abundance and sequence of all the transcripts in a cell. scRNA-seq has historically dealt with problems of bias in the type and quantity of transcripts it identifies, and several recent protocols (CEL-Seq [49], MARS-Seq [54], Cyto-Seq [34], and Drop-Seq [80]), have sought to overcome these issues. Solutions include in vitro transcription, which uses linear instead of exponential amplification, the use of barcodes to relate each transcript sequence to a unique molecular identifier, and spike-in RNA to normalize the output signal to the relative transcript abundance in each cell, to reduce technical variation. The advantage is that one can quantify the transcriptome in depth, and quantify the variation or heterogeneity of all the expressed genes of a given cell.

The primary limitation of traditional scRNA-seq is that it does not inherently contain spatial information, nor does it allow one to follow the transcriptome over time in a single cell. Recently, the protocol 'FISSEQ' [65] combined spatial information of transcripts from smFISH with sequencing of the individual transcriptomes of those cells. Sequencing methods that incorporate spatial information of the transcript, or imaging methods that incorporate sequencing, will be essential contributions to ultimately achieving a "4D" transcriptome atlas of gene regulation.

1.3.3 Single-cell heterogeneity in gene expression

Variation can be due to genetic or non-genetic causes. It can be fixed or time-dependent, programmed or random. A significant component of non-genetic heterogeneity is due to the discontinuous nature of transcription. Genes are transcribed in "bursts" and this phenomenon has recently been observed in mouse liver tissue with smFISH [6], Drosophila with live-cell imaging [30], mouse embryonic stem cells with scRNA-seq [81], zebrafish embryos with smFISH [96], and human brain tumors [81]

and melanoma [32] with single-cell qPCR.

The prevailing view of considering heterogeneity as 'extrinsic' or 'intrinsic', was first developed from experimental data using two-color reporter assays in bacteria [30], and later in yeast [108]. Extrinsic factors are those that influence many genes, for example the concentration of RNA polymerase in a cell. Intrinsic noise arises from stochastic fluctuations inherent in biochemical reactions between molecules at low copy number. This dichotomy of heterogeneity continues to be examined. Recently, Fu and Pachter [39] revisited previous data from Elowitz and colleagues [30] highlighted the importance of experimentally determining whether the two fluorescent reporters have the same distribution of mean and variance in fluorescence intensity, and to normalize them if not, as this is a major assumption of the model. Also recently, Sherman and colleagues [121] proposed that extrinsic and intrinsic variability are not exclusively orthogonal to each other. The authors examined extrinsic variability with the yeast heat shock protein SSA1, and with modeling showed how intrinsic variability can be dependent on external factors. The terms 'extrinsic' and 'intrinsic' are subjectively defined. Considering that upstream extrinsic factors may also have a timescale of fluctuation (for example due to bursting), one person's extrinsic noise could be another person's intrinsic noise.

How stable is heterogeneity, and what are its functional consequences? Whether the 'noisiness' of gene expression has a functional purpose or evolutionary advantage is still an open question. Examinations of 'bet-hedging' have largely been confined to bacteria (for a recent review see [27]), but could such transient heterogeneity confer any advantages in eukaryotes? A recent paper by Shaffer and colleagues [119] provides a compelling single-cell viewpoint of how transient switching of phenotype profiles of patient-derived melanoma cells leads to stable populations resistant to the drug vemurafenib. Using a variation on the classic Luria-Delbruck experiment [79], they observed that before application of the drug, cells transition between 'non-resistant' and 'pre-resistant states', as observed by their transient ex-

pression of resistance markers. The pre-resistant state was not heritable, and it was only with addition of vemurafenib that cellular reprogramming and a stable resistance phenotype emerged.

This intriguing example of transient heterogeneity in mammalian cells may be seen as a manifestation of "dynamical instability" [102], a model from the field of network theory based on Boolean logic to explain the mechanisms underlying gene expression variability observed in some cancer types. Supporting evidence for the dynamical instability hypothesis has been observed at the population level in 'antiprofile' studies, which shows that many cancer subtypes exhibit a high degree of gene expression variability across individuals [14, 24]. Hypervariability of gene expression is a reproducible signature of cancer tissue types. It would be interesting to determine whether at the single cell level, transient gene expression heterogeneity is an illustration of dynamical instability of gene networks, and whether this is related to hypervariability in cancer or observed more generally.

Another open question about gene expression heterogeneity is whether stochastic transcription ultimately gets transmitted through the nucleus and affects the level of translated protein product. A few recent studies propose that stochastic expression is in fact buffered, and limits the variation in cytoplasmic mRNA that is ultimately available for translation. One study attributes this buffering to microRNAs [117], and two other studies provide examples of buffering by the nuclear pore complex [5, 10]. Halpern and colleagues used whole genome RNA-seq and smFISH in various mouse cell types and found that mRNA was retained in the nuclear pore. The authors found a difference in retention times: immediate early genes tended to have the shortest retention time, and protein coding genes the longest. Their interpretation was that mature processed RNA was retained in the nuclear pore and that fast induction time was due to their release from the pore, not the mRNA synthesis rate itself. The transient nature of heterogeneity and its potential buffering are interesting observations, and it remains to be seen whether they are observed

more generally in higher eukaryotes.

1.3.4 Nuclear architecture and the role of *cis* elements in gene regulation

Our understanding of nuclear architecture and the role of enhancers has increased profoundly over the past few years. We refer readers to recent reviews on promoter-enhancer interaction [134] and the role of nuclear architecture on gene expression [42], and here focus on new research examining the role of nuclear topology in gene regulation. Hi-C, a population based assay to determine long range chromosome interactions, led to the identification of topologically associated domains (TADs) [25,94] and associations with the proteins CTCF and cohesin to act as insulators of chromosome 'neighborhoods', where enhancers interact with the promoters of genes within a neighborhood.

Within TADs, transcriptionally active genes are shown to share spatial co-regulation [118], and disruption of these topological boundaries have consequences for disease. For example, recent studies looking at the role of nuclear topology and cancer show that gene duplication (a common feature of cancer) is mis-regulated if it occurs at the boundary of a neighborhood rather than within it. A model of 'enhancer hijacking' has been proposed [140], which occurs when a boundary is disrupted and an enhancer is able to interact with the promoters of oncogenes and promote their expression. Manipulation of TAD boundaries with CRISPR was recently shown to cause oncogene activation of gliomas [50] and leukemia [36].

Recently, Bartman and colleagues [9] manipulated enhancer-promoter contacts at the locus control region in mouse erythrocytes and human primary erythroid cells. Their observations used smFISH to evaluate how transcription burst features of the beta- and gamma- globin genes were affected when they minimized contacts (via

deletion) or increased contacts with forced looping. They saw that enhancer contacts increase burst frequency, supporting the idea that enhancers increase the probability of transcription, similar to what was shown for reporter genes [63]. Importantly, they also found that active transcription of one allele lowered the probability of activity of the other allele, giving evidence to support a model of promoter-enhancer interaction where the enhancer alternates between contacts of the promoters it regulates.

Enhancer manipulation has also been carried out with live-cell imaging in Drosophila [40]. In this study, enhancers and insulators were placed between two reporter genes in developing embryos, resulting in modulations of burst frequency. Here, the authors concluded that one enhancer could activate two genes at once, in contrast to the model from the globin locus. More generally, disruption of boundary elements or mutation of CTCF results in increases in gene expression noise [109]. Thus, the interaction between enhancer and promoter as reflected in chromosome topology is a prime determinant of metazoan expression heterogeneity. Single-cell imaging coupled with manipulation of nuclear architecture will continue to provide rich insight into the physical factors governing gene regulation.

1.3.5 Modeling the transcription process and gene networks

Finally, one of the goals of studying single-cell gene expression is to reconcile the complexity of biology with the desire to find universal principles that govern living behavior. In pursuit of that understanding, researchers have drawn upon methodology from information theory [120,130]— the study of how information is received, processed, and transmitted within a system. In gene regulation, it is increasingly clear that the gene receives information not just in *trans* (i.e. chemical modifications, binding of activating and inhibitory transcription factors) but in *cis* (DNA elements in the promoter and enhancer), as noted in the previous section. Thus, the question we now ask from a theoretical standpoint is, how can a cell decode the complex collection of incoming signals to produce an effective response?

The earliest 'Telegraph' model for describing how information is processed through gene expression dynamics [100] was based on a single active and inactive state. The model proved to fit expression data in some instances [105, 108, 122], but there are increasing examples which illustrate that two states are insufficient to represent the data [13, 48, 116, 125]. Recently, Rieckh and colleagues [110] identified instances in which a multi-state promoter model performs better than a simple two-state model; however, they advocate the two-state model as the simplest theoretical baseline to start from, as it is possible to overfit the data with too many states. In **Figure 1.3** we illustrate how various cis and trans factors have the potential to affect promoter states on different timescales . Factors affecting the OFF time could include chemical modifications of the promoter and enhancer, or the displacement of the +1 nucleosome from the transcription start site. Transcription factors may prolong the ON time of a gene by stabilizing basal transcription machinery at the promoter, or enhancer-promoter contacts, for more successful re-initiation. On the other hand, a repressive transcription factor might recruit factors such as chromatin remodelers that could result in a shortening of the active period of the gene. We foresee theoretical models coupled with experimental data to continue elucidating these key factors.

Information theory has also recently been used in studies that manipulate trans factors of gene regulation. Hansen and O'Shea [46] controlled the frequency and duration of the nuclear localization of the yeast transcription factor Msn. The authors observed the effect of modulating Msn2 nuclear localization on the burst frequency and amplitude of two target genes and determined that transcriptional bursts from natural Msn2 target promoters encode 1.0-1.3 bits of information about the signal identity and intensity. Another group recently used opto-genetic stimulation to manipulate Ras in NIH3T3 cells and determined its effect on the transcriptome profile of immediate early genes [141]. Along with manipulating *cis* elements in the nucleus, as described in the previous section, experimental systems that allow re-

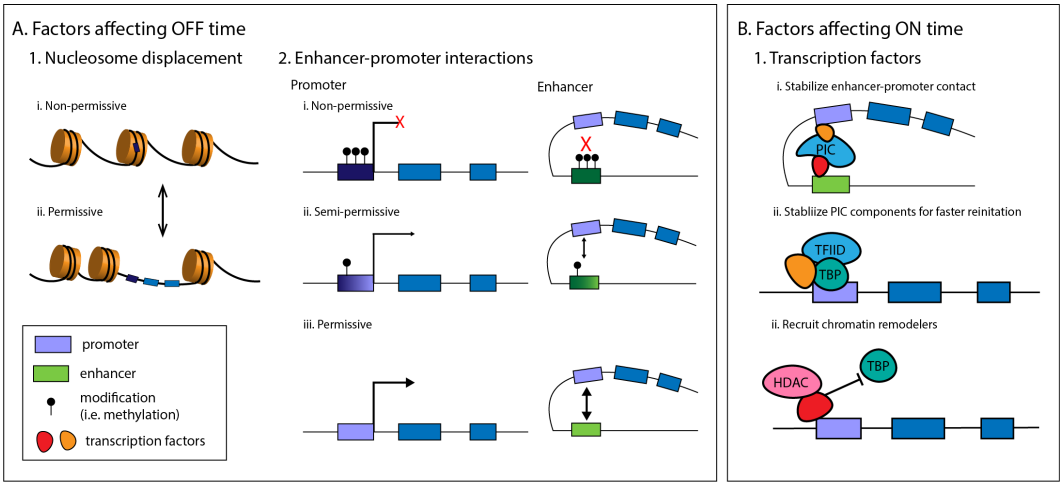


Figure 1.3: Possible factors contributing to multiple promoter states. Abbreviations: PIC, pre-initiation complex; TFIID, general transcription factor II-D; TBP, TATA binding protein; HDAC, histone deacetylase.

searchers to directly control and manipulate the localization of trans elements, such as transcription factors, will be important for achieving a systems biology viewpoint of precisely how information is transmitted to a gene.

A new development in scRNA-seq methodology is the use of principal component analysis (PCA) for stem cell lineage tracing. Several groups [22,37,132] use scRNA-seq expression data from a population of stem cells that have been induced to differentiate, and use PCA for 'pseudo-temporal ordering'- a timeline of gene expression changes gathered from the single-timepoint gene expression profiles of many cells. In sequencing, an assumption of pseudo-temporal ordering is that every cell represents a timepoint along the same continuum of differentiation. We can see a similar assumption in imaging, where it is assumed that the observed distribution of bursting comes from the same underlying kinetics present in every cell. Both methods assume that modeling will obtain parameters that are reflective of a 'mean' process. But what if it is not? Llamosi and colleagues [74] propose that the idea of fitting parameters to a 'mean cell' is faulty, and instead suggest that the goal should be to arrive at a distribution of models. This study highlights a major assumption of all three single-cell methods that is gene expression is ergodic— observing a single

cell over many timepoints yields the same information as measuring the population of cells at a single timepoint. Recent work has shown the fallacies of the ergodic assumption for dynamical models in which time averages are commonly replaced by population averages [101]. As a result, we believe the assumption of ergodicity in gene expression should be studied carefully.

Another aspect of gene expression to which mathematical modeling has begun to contribute is the elucidation of gene networks from single-cell expression data. Datasets that obtain measurements of gene expression profiles from single cells are becoming increasingly prevalent, and there is potentially much information to be gained from pairwise correlations of genes. But what can co-expression tell us about connectivity? Simply looking for correlation in mRNA as a sign of connectivity proves challenging at the single-cell level because transcription is stochastic and dynamic— timescales of the birth and decay rates of mRNA affects how much will be present at any given point in time. In **Figure 1.4** we illustrate this problem with a hypothetical gene regulatory interaction Gene A has a causal relationship with Gene B, where the protein product of Gene A promotes the expression of Gene B (i.e. the interaction between a transcription factor and its gene target). This interaction may be present in every cell of a population; however, each cell may be at different stages of the regulatory process at any given point in time. As such, fixing the population at a single time point (as is done for smFISH and scRNA-seq) may yield different combinations of Gene A and Gene B mRNA with no apparent correlation at the single cell level. As illustrated in the figure, the half-lives of both mRNA species may affect the degree to which they overlap in time (highlighted in yellow) within a single cell.

The difficulty in interpreting temporal directionality in a gene network has been observed both with single-cell imaging and scRNA-seq, where a clear functional response at the population level yields poor or no correlation when examining the RNA of pairs of involved genes within single cells [19, 51]. These observations ne-

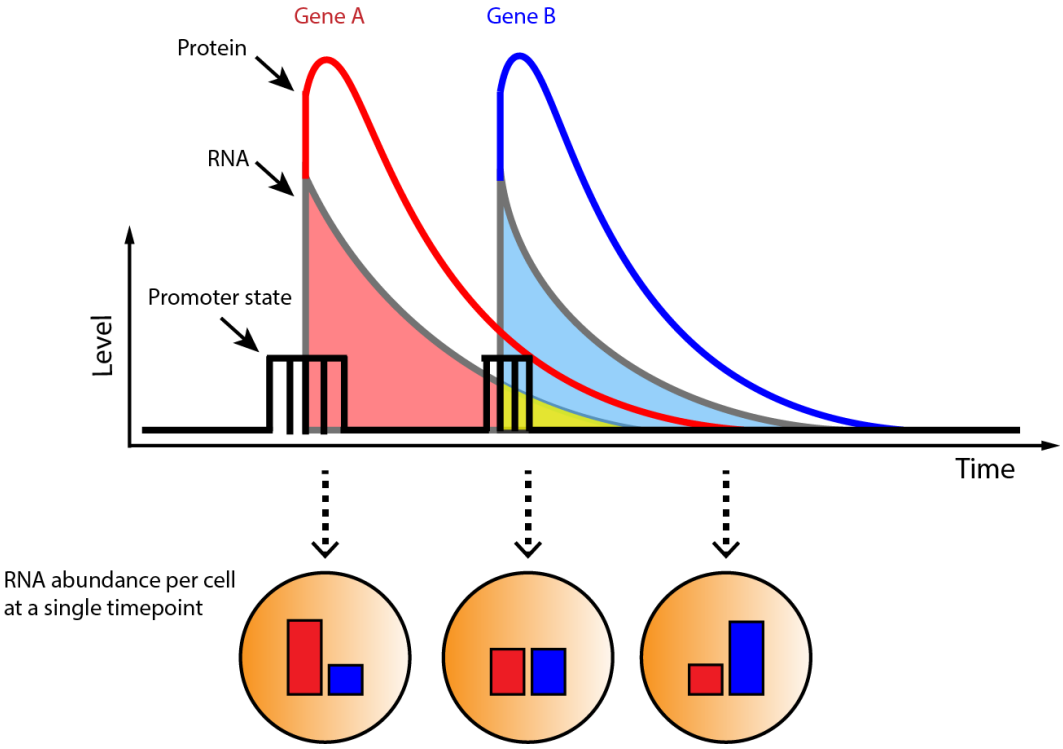


Figure 1.4: Capturing dynamic transcription regulation at a single timepoint.

cessitate the following questions for the field: Under what conditions can we expect to see correlations in expression between two interacting genes? And when is co-expression more reflective of a direct interaction rather than an indirect one?

These questions have previously been examined using principles from information theory. Ku and colleagues [61] used microarray datasets to determine whether co-expression of gene pairs was indicative of transcriptional regulatory interactions (TRIs, when Gene A codes for a transcription factor that regulates Gene B) or co-regulation (when two genes A and B are regulated by a common upstream factor). They found that co-expression was indicative of co-regulation in the bacteria *E. coli*, but more indicative of TRIs in *S. cerevisiae*. Additional such studies would be useful to the field of single-cell gene expression, as they provide a framework for us to ask what co-expression indicates in higher eukaryotes, and whether these conclusions are observable in single cells. Will population-based network models need to be modified to describe the observed stochastic nature of expression at the single-cell

level? Do gene networks change with transient heterogeneity, or when cis nuclear architecture is altered? Development of single cell assays and their applications have outpaced theoretical work to examine the data for its underlying principles. We see a need for more studies to be done in this area in the future.

1.3.6 Conclusions

The field of single-cell gene expression has the potential to generate a comprehensive and quantitative view of gene regulation. Developments in gene editing, advances in high-throughput and multiplexed assays for single cells, and increased understanding of nuclear architecture, are making significant contributions towards our ability to manipulate and understand the dynamic aspects of gene regulation. In this review, we have highlighted recent literature that advances the capabilities of three frequently-used techniques: scRNA-seq, smFISH, and live-cell imaging. Heterogeneity is still a phenomenon to grapple with – both experimentally and theoretically – but as it continues to be identified in more systems, observations of its dynamic properties are leading to hypotheses about its functionality and consequences. Models of heterogeneity continue to be revisited, and studies that manipulate nuclear architecture reveal the magnitude of its role in regulating gene expression. Population-based assays hold great utility for elucidating gene networks, and we hope single-cell data will begin to provide useful insight in this area as well. Current work and future developments in the field will contribute to a more thorough temporal and spatial understanding of gene regulation at the single-cell level, and ultimately to our ability to relate an environmental stimulus to the response of an organism.

Chapter 2

Methods in Single-Cell Gene Expression Analysis

2.1 Single-Molecule FISH

2.1.1 Method overview

Single-molecule fluorescence in-situ hybridization (smFISH) is a method to visualize RNA abundance levels in a population of fixed cells. This experimental technique is based on hybridizing a series of fluorescently-labeled short DNA oligo probes to an RNA transcript of interest. When visualized with a widefield fluorescence microscope, the tiled probes appear as a single diffraction-limited spot that corresponds to the location of the labeled RNA (**Figure 2.1**). The smFISH probes can be designed to label various features of an RNA transcript, such as the introns, exons, and UTR regions.

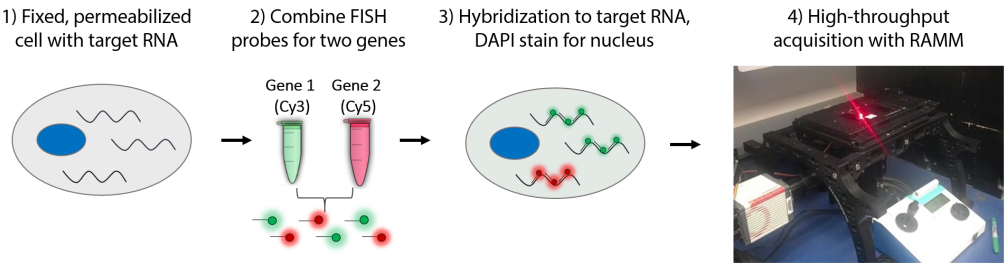


Figure 2.1: Schematic of smFISH protocol.

	Equivalent Dye	Absorption Max λ	Emission Max λ
Quasar-570	Cy3	548nm	566nm
Quasar-670	Cy5	647nm	670nm

Figure 2.2: Spectral properties of Cy3- and Cy5- equivalent Quasar dyes.

2.1.2 smFISH probe design

Probe sets are ordered from Biosearch Technologies, and each oligo sequence is conjugated to a single Cy3 derivative (Quasar570) or Cy5 derivative (Quasar670) with spectral properties outlined in **Figure 2.2**. Probes sets are designed with the Stellaris Probe Designer (biosearchtech.com) using the following parameters:

- Masking level = 5
- Max number of probes = 48
- Oligo length = 20
- Minimum spacing = 2

In order to minimize off-target labeling, probe sequences are screened by BLAT (https://blast.ncbi.nlm.nih.gov/Blast.cgi) to confirm they are unique to the target gene. Any sequences that have homology to other sites in the genome are removed from the final probe set.

2.1.3 Reagents

1. **Stellaris Probe Reconstitution**

   Dissolve in 1x TE to 25µM stock concentration and store at -20C. Working aliquots can be made of the stock concentration or a 1:10 dilution to 2.5µM. The final concentration of probe solution should be 100nM without diluting the hybridization buffer excessively.

2. **Fixation Solution (4% PFA/PBS)**

   For 12mL:

- 1.5mL 32% Paraformaldehyde
- 10.5mL 1x PBS

3. **Permeabilization Solution (70% ethanol/H20)**

   For 24mL:

   - 16.8mL 100% ethanol
   - 7.2mL H2O

4. **Wash Buffer (10% formamide/2x SSC/H20)**

   For 50mL:

   - 5.0mL 20x SSC
   - 5.0mL Formamide
   - 40.0mL H20

5. **Hybridization Buffer (make 1mL aliquots stored at -20C)**

   For 10mL:

   - 1g dextran sulfate
   - 7.0mL RNAse-free Water

   Dissolve dextran sulfate until clear. This may take up to 1 hour. Then add:

   - 1.0mL 20x SSC
   - 1.0mL Formamide
   - Add water up to 10mL final volume.

   Aliquot into Eppendorf tubes and store at -20C.

6. **smFISH Probe Solution (50µL per coverslip, calculate for 55µL)**

   - 2µL of 2.5µL diluted probe OR 0.2µL of 25µM stock
   - 48-50µL Hybridization Buffer

7. **Mounting Medium**
    - ProLong™ Gold Antifade Mountant with DAPI (ThermoFisherSci P36935)
8. **Coverslips**
    - 18mm diameter #1.5 thickness (Electron Microscopy Sciences 72222-01)

2.1.4 Labeling protocol

All volumes are 1mL unless noted.

Step 1: Fix and Permeabilization

1. Wash cells 3x with HBSS.
2. Fix cells with 4% PFA in PBS for 10 minutes.
3. Wash cells 2x 10 minutes with PBS.
4. To permeabilize, store in 2mL 70% ethanol 1-2 days at 4C.

Step 2: Prep for smFISH

1. Prepare Wash and Hybridization Buffers.

   Note: Allow formamide stored at 4C to reach room temperature before opening in a chemical fume hood.
2. Remove 70% ethanol from wells OR transfer coverslips to a new 12-well plate.
3. Incubate coverslips >5 minutes with Wash Buffer.
4. Set out smFISH probe working stocks stored at -20C to thaw in dark.
5. Prepare incubation chamber: With a lab marker, number coverslip positions on the bottom of a 10cm plate (usually 4-6 coverslips per plate).
6. Cut a piece of parafilm to size and place inside plate, wrapper side up. Press down firmly so the parafilm sits flat before peeling the wrapper off. Marked numbers will be visible underneath.

7. Fold a Kimwipe 5-6 times and tape to underside of the lid. Moisten with 1mL H20; this provides humidity for the chamber.

Step 3: smFISH Probe Hybridization

1. Prepare smFISH probe in Hybridization Buffer for 55µL per coverslip (to compensate for pipetting error). Vortex gently to mix. Store probe solution in dark while coverslips are being dried in the next step.

2. Pick up coverslips with forceps and wick off excess liquid with a Kimwipe. Rest against an Eppendorf tray on a paper towel to air dry. Set out all 4-6 coverslips for a given plate.

3. Pipette 50µl probe solution onto parafilm at numbered locations.

4. Place coverslips cell-side down onto probe solution. It helps to place the coverslip down onto the probe solution at an angle to prevent air bubbles. If needed, use forceps to tap the top of the coverslip gently to make the probe solution spread across the whole area.

5. Close the plate with the wet Kimwipe taped to the lid, making sure it does not touch the coverslips. Seal the plate with parafilm and incubate at 37C for a duration of 4 hours to overnight. **Note: Keep light exposure to the samples minimal after this point.**

Step 4: Wash and mount coverslips

1. Replace the Wash Buffer in the 12-well plate. Unseal incubation plate and place coverslips into the Wash Buffer cell-side up. Discard the parafilm and probe solution; the numbered plates can be washed and reused.

2. Incubate coverslips 2x 30 minutes in Wash Buffer at 37C. Thaw Prolong Gold mounting media after first wash. Subsequent washes are at room temperature.

3. Bring Wash 1x with 2x SSC/H20.

4. Wash 1x for 5 minutes with PBS.

5. Set out glass slides and label. 1-2 coverslips will be mounted per slide.
6. Wick coverslips and air dry as before in Step 3.2.
7. Add a small drop of mounting media to the glass slide. Gently place coverslip cell-side down onto the mounting media, making sure not to introduce bubbles.
8. Let the glass slides dry overnight in dark at room temperature.
9. Before imaging, gently wipe the coverslip surface with a moist kimwipe to remove any residue (i.e. salts from PBS) before placing on the microscope. Store slides at -20C after imaging.

2.1.5 Microscope instrumentation

We image smFISH samples on a custom-built widefield fluorescence microscope with the following components:

- Chassis: Rapid Automated Modular Microscope (RAMM, ASI Imaging, OR, USA)
- Light source: SpectraX Light Engine (Lumencor, OR, USA). The light engine contains seven color bands (max λ/FWHM, power): Violet (395/25, 295mW), Blue (440/20, 256mW), Cyan (470/24, 196mW), Teal (510/25, 62mW), Green (550/15, 260mW), Yellow (575/25, 310mW), Red (640/30, 231 mW)
- Filter set: Quad-bandpass filter for imaging DAPI/GFP/Cy3/Cy5 (VCGR-SPX-P01-PC, Chroma Technology Corp., VT, USA)
- Additional emission filters: 605/70 (Cy3) and 700/75 (Cy5) controlled by a FW-1000 High Speed Emission Filter Wheel (ASI Imaging) and TG-1000 Tiger serial controller (ASI Imaging)
- Camera: ORCA-Flash4 V2 CMOS camera (Hamamatsu Photonics K.K., Japan)
- Stage: MS-2000 motorized XY stage (ASI Imaging)
- Objective lens: 40x/1.4 N.A. oil immersion objective (Carl Zeiss, Germany)

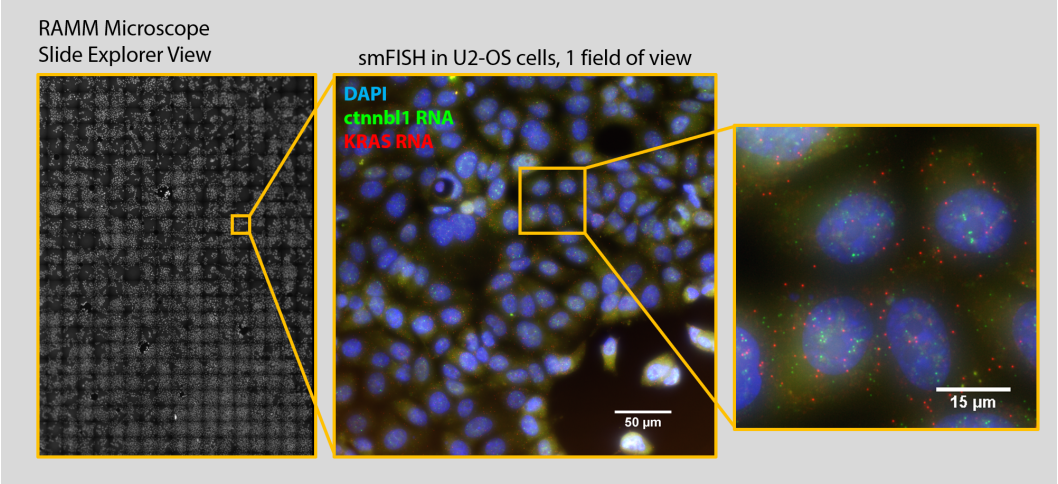


Figure 2.3: SlideExplorer view in MicroManager.

- Software: MicroManager (Version 1.4.18, https://micro-manager.org) [28]

2.1.6 Acquisition parameters

To select the regions of a coverslip to image, a large-scale tiled view of the coverslip is first obtained using the MicroManager plugin 'Slide Explorer' (**Figure 2.3**).

Areas of the coverslip with noticeable debris, air bubbles, or irregular cell densities (i.e. <10% or >90%, will depend on the normal culture conditions for a given cell type) are generally to be avoided, as these can affect the biology being measured and yield inaccurate cell segmentation and spot identification. A custom MicroManager script performs batch acquisition in the subsequent steps. Acquisition grids are created around chosen areas (usually 5x5 to 10x10 grids, 2-3 grids per coverslip). The focus across the z-plane of the grid is set by manually correcting the focus on the central frames in the upper, lower, right, and left edges of the grid. The following Lumencor excitation light and camera exposures are used for each fluorophore:

- DAPI: Violet, 25ms, no emission filter
- Cy3: Green, 500ms, 605/70 emission filter
- Cy5: Red, 500ms, 700/75 emission filter

Z-stacks with a 0.5µm step size are used to image the whole volume of the cell

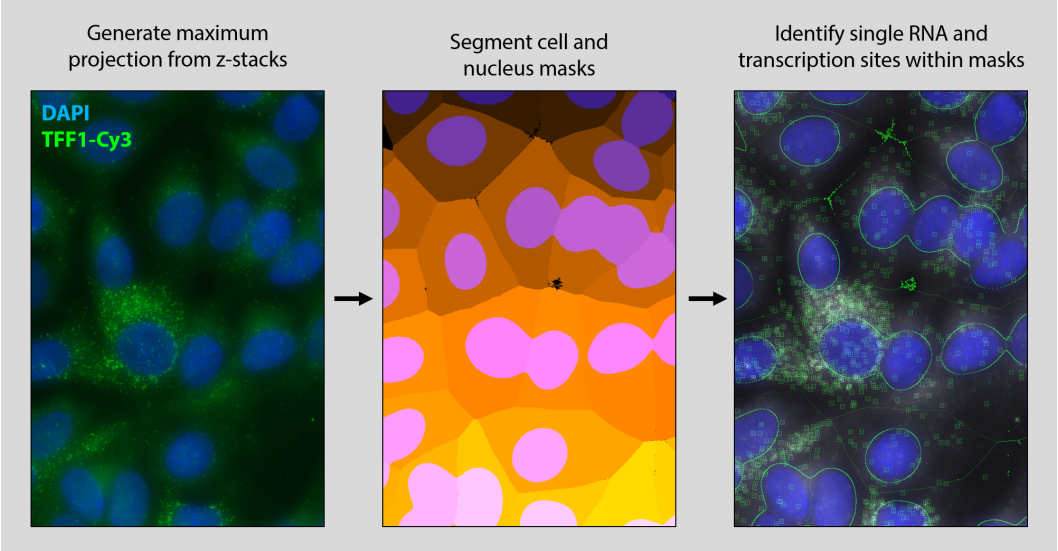


Figure 2.4: smFISH image processing pipeline.

(U2-OS and HBEC cells = approx. 11 slices, MCF7 cells= approx. 14 slices). The number of slices should be chosen such that the entire volume of the cells, across all areas of the grid, are fully captured in the stack.

2.1.7 Image processing

There are five main steps for smFISH image processing: **1)** Maximum intensity projection and **2)** Illumination correction with MicroManager, **3)** Segmentation of cell and nucleus masks with CellProfiler (version r.11710, https://cellprofiler.org), **4)** RNA spot identification with a custom IDL program 'Localize' [129, 143], and **5)** Merging of image, mask, and RNA spot information with a second custom IDL program 'FISHAuxiliary.' The steps are illustrated in **Figure 2.4** and described below:

1. **Maximum Projection:** Z-stacks are converted into a 2-D maximum intensity projection to produce a single image containing the entire volume information for a given field of view. The resulting fi

   le is a multi-channel, multi-position TIF hyperstack.

2. **Illumination Correction:** Flat-field correction is performed on images to

correct for the uneven illumination by the excitation light upon a field of view. If not compensated for, this heterogeneous illumination will produce errors in the fluorescence intensity readout of the RNA. Within the MicroManager batch acquisition script, we consider a maximum-projection image as a 2D matrix of pixel intensities and apply the following equation:

$$C = \frac{R - D}{F - D} \tag{2.1}$$

Where

- C = Illumination corrected image
- R = Raw image
- F = Flat-field image
- D = Dark image

The flat-field image is generated from an out-of-focus frame of a solution of free dye (DAPI, Cy3, or Cy5) in water excited with the appropriate wavelength. The dark image is a frame acquired without a sample or excitation light.

3. **Nucleus and Cell Segmentation:** Nucleus masks are segmented as 'primary objects' by applying the Otsu method of thresholding with two-classes to the DAPI channel (**Figure 2.5**). Cell masks are similarly generated from the smFISH channel, seeded from the nucleus primary objects.

4. **Spot Detection:** Intensity and area thresholds are applied to the smFISH channel to identify spots. The IDL program 'Localize' fits a 2D Gaussian mask to each spot and the integrated intensity of the Gaussian fit is recorded as the spot intensity. A single channel TIF stack is given as the input, and the output file contains the frame number, x-y coordinates, and integrated pixel intensity of each spot.

5. **Merge of Segmentation and Spot Identification Data:** The image, mask, and RNA spot data is merged in the IDL program 'FISHAuxiliary' to identify RNA spots specifically within the nucleus and cytosol of cells. The output file is a table with each row containing the nucleus and cytosol RNA

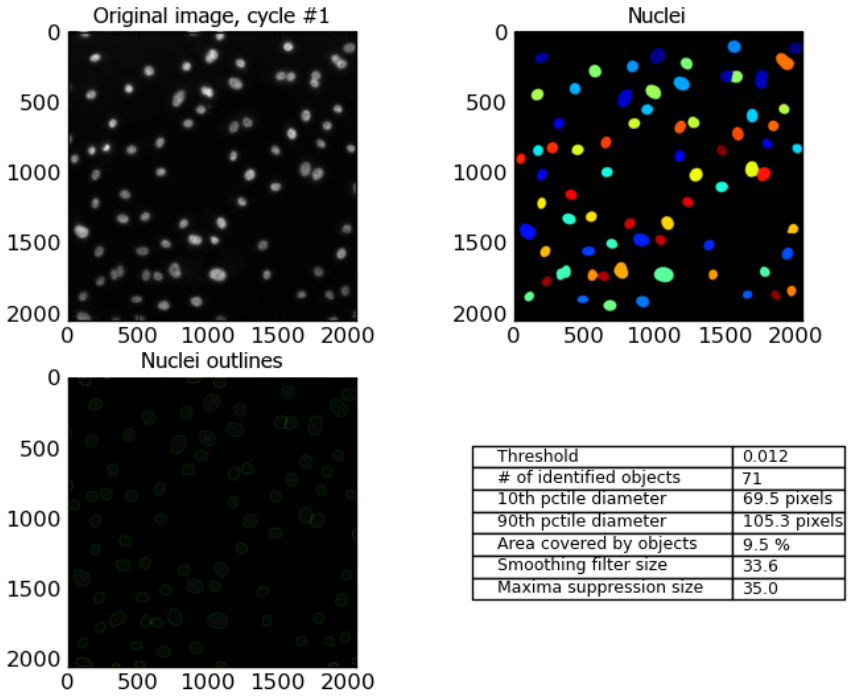


Figure 2.5: Nucleus segmentation in CellProfiler.

counts for a given cell ID. The two RNA counts can then be summed to yield the total RNA abundance for a given cell. A threshold can be further applied to identify spots of brighter intensity within a nucleus mask as transcription sites (TS). The number and intensity of these spots is recorded in a separate file.

2.1.8 Data analysis

The basic forms of output data are:

- Number of cytosolic, nuclear, and total RNA per cell.
- Number of TS per cell.
- The fluorescence intensity of each identified spot. In the case where we used an exon probe to obtain TS intensities, we can divide the TS intensities by the median intensity of a single RNA (the majority being single mRNA) to

yield the discrete number of nascent RNA (or transcribing RNAPII) at the TS.

The statistics calculated from the data are: 1) Mean with standard error for TS per cell, and 2) Two-sample Kolmogorov-Smirnov test on RNA distributions and TS intensity to determine whether two conditions (i.e. low MYC and high MYC) significantly differ from each other. The Kolmogorov-Smirnov test statistic D is calculated from the largest vertical difference between two distributions with the equation given by:

$$D = \max_x |F_{1,n}(x) - F_{2,m}(x)| \tag{2.2}$$

Where F_1 and F_2 are the empirical distribution functions of the first and second samples, and n and m are the respective sample sizes.

2.2 Live-Cell Imaging with MS2

2.2.1 Method overview

The MS2 live-cell imaging method allows for the real-time visualization of transcription events in single cells with fluorescence microscopy. This method is based on integrating a MS2 bacteriophage sequence into a gene of interest within a cell line or organism. This sequence forms secondary structures (stem loops) when the gene is transcribed into RNA. These stem loops are bound by cognate MS2 bacteriophage coat protein fused to GFP, the sequence of which is also integrated into the cells of interest and constitutively expressed. The result is fluorescently labeled RNA, where the fluctuating fluorescence signal over time is a readout of the RNA production events of a given gene (**Figure 2.6**).

2.2.2 Engineering of MS2-labeled cell lines

- **TFF1 endogenous gene in the MCF7 cell line:** The *TFF1-MS2* cell line was created with CRISPR targeting the *TFF1* 3'UTR in MCF7 cells

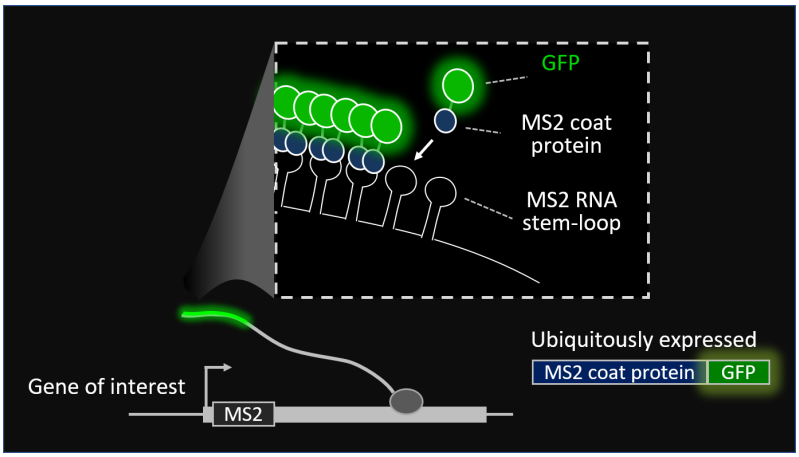


Figure 2.6: Schematic of MS2 integration and fluorescent labeling of RNA.

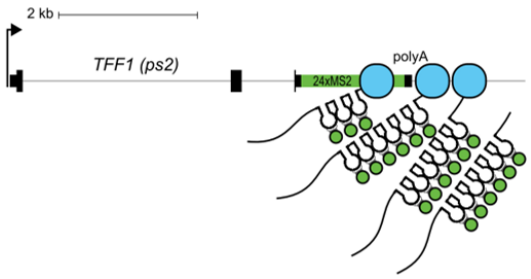


Figure 2.7: TFF1 MS2-GFP construct. Figure reproduced with permission from *Rodriguez et al., 2019.*

(**Figure 2.7**, developed in [112]). The CRISPR reagents consist of two plasmids: 1) a pX330 plasmid containing Cas9-mCherry with guide RNA sequence, and 2) a donor plasmid containing left and right homology arms to the *TFF1* locus flanking the target site with a 24x MS2 stem loop array and puromycin resistance cassette. The pX330 and donor plasmids were transfected into low passage MCF7 cells. After 7 days, antibiotic selection was started to enrich for positive cells. Surviving single clones were screened by PCR to check for on target integration. The MCF7 cell line is polyploid and three of the five *TFF1* alleles received the MS2 integration.

- **Endogenous gene panel in the HBEC cell line:** This HBEC cell line was engineered to introduce the 24x MS2 array globally into introns of genes (**Figure 2.8**, developed in [136]). A lentiviral 'gene trap' vector was engineered for this purpose. The vector consists of lentiviral backbone with

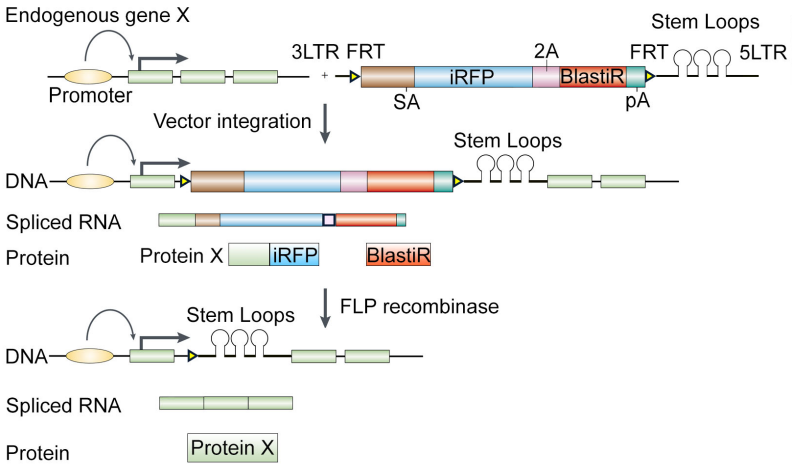


Figure 2.8: MS2 gene trap construct. Figure reproduced with permission from *Wan et al., 2019.*

FRT sequences flanking a splicing acceptor sequence, polyA (polyadenylation sequence), and a selection cassette consisting of blasticidin resistance and iRFP far-red fluorescent protein. The vector was introduced to HBEC cells by lentivirus transduction. Positive cells containing the integration were enriched by applying blasticidin antibiotic selection. MS2-GFP coat protein was then introduced with lentivirus transduction and cells were FACS sorted to enrich for GFP+/iRFP+ cells. The blasticidin-iRFP selection cassette was removed with FLP recombinase so that only the 24x MS2 array remained. Cells with GFP+ /iRFP- fluorescence were single-cell sorted with FACS and individual clones were imaged and sequenced for further analysis.

2.2.3 Aquisition parameters

MCF7 cells were imaged on a Zeiss LSM780 confocal microscope with a 37C, 5% CO2/humidity environment. Imaging was performed with 488nm and 594nm excitation at 2% power, 63x/1.4 N.A. oil immersion objective, 2.5µm pinhole size, 1.5x zoom, and 1024x1024 frame size acquired at 16-bit depth. Z-stacks were acquired at 14 slices/0.5µm step size. The imaging duration was 512 frames/100 sec intervals (approx. 14.2 hours), allowing for 1 well with 4 FOV per well to be imaged per acquisition.

For high-throughput imaging, HBEC cells were cultured in 96-well plates (Brooks Life Science Systems MGB096-1-2-LG-L) and microscopy was performed on an automated Yokogawa Cell Voyager 7000S dual spinning disk microscope with a 37C, 5% CO2 and 80% humidity environment. The microscope was equipped with a quad-bandpass filter (405/488/561/604 nm) and an additional GFP emission filter (525/50). Imaging was performed with 488nm excitation at 30% power and a 60x/1.2 N.A. water immersion objective. Fluorescence was detected with an Andor Neo 5.5 sCMOS camera with 250ms exposure time, 1278x1078 frame size, and 2x2 binning acquired at 16-bit depth. Z-stacks were acquired with 14 slices/0.5µm step size. The imaging duration was 6 or 15 hours with 100-second frame intervals, allowing for 8 wells with 5 FOV per well to be imaged during a single time course. Flat-field correction and maximum intensity projections were processed on the fly by Yokogawa acquisition software.

2.2.4 Image processing

Automated image analysis of live-cell time series data is carried out in the KNIME open-source workflow environment (64-bit Version 3.5.1, [11]) containing the following image processing and scripting nodes: KNIP (Version 1.5.3.201611190650 [23]); R (Version 3.3.1, 64-bit, https://r-project.org); and Python (Version 2.7.12, 64-bit, https://python.org). Spot tracking in KNIME incorporates algorithms from Localize. The analysis consists of five main steps: 1) Nucleus segmentation and tracking, 2) TS tracking within registered nuclei, 4) Filtering correct TS tracks. The steps are illustrated in **Figure 2.9** and described below:

1. **Nucleus segmentation:** This step is performed on the MS2-GFP channel using a Random Forest Classifier. A training set with supervised learning of a given cell type (MCF7 or HBEC) is conducted in which correctly and incorrectly segmented nuclei are manually classified in single frames of a representative time series movie. The RF model is tested on unseen data and

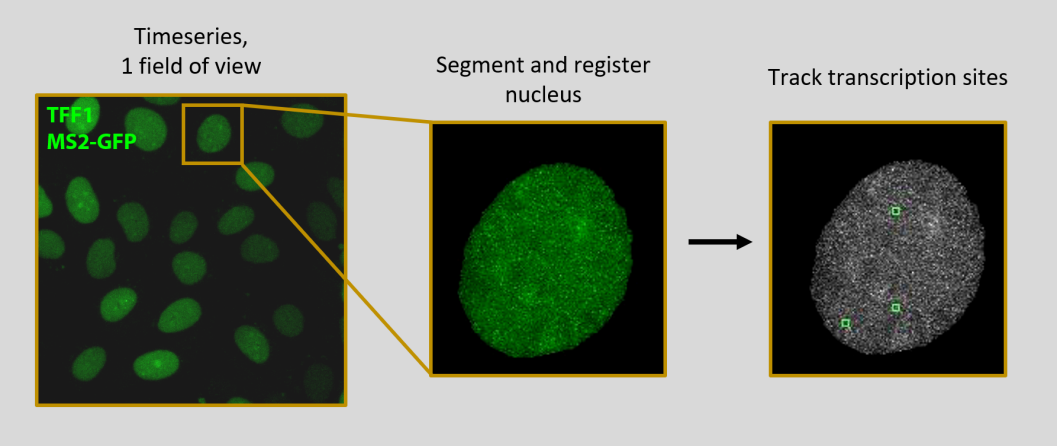


Figure 2.9: Live-cell image processing pipeline.

then applied across all subsequent time series of a given cell type. Registration of tracked nuclei in X-Y is automatically applied after the segmentation step. The reason we limit the movement of the nuclei by registration is that cell movement and long transcription OFF periods that can occur over a 6-15 hour range. This makes it difficult to set an adequate search radius that would allow bursts of the same allele to be successfully linked together.

2. **Spot identification and tracking:** Spot identification is performed on registered nuclei based on 2D Gaussian masking methods developed in Localize. A square border around the detected spot is used to track bursting events across frames with the following KNIME parameters used for MCF7 and HBEC cells:

   - 'LAP Tracker Maximum Distance' = 15 pixels: Max distance a visible spot can travel between frames.
   - 'LAP Tracker GAP-Closing Maximum Distance' = 15 pixels: Max distance a spot can travel when it reappears after disappearing.

   The pixel intensity values of filtered traces are then normalized to the highest and lowest value on a per track basis.

3. **Classifying gene ON and OFF periods using a 2-state Hidden Markov Model:** The batch of normalized traces for a given condition are fit to a 2-

state HMM as a way to classify the active and inactive states of each gene in an unsupervised manner. The model we use is based on an algorithm for characterizing single molecule FRET transitions [66] that factors in the environmental heterogeneity in the measurements such as variation in background fluorescence intensity (as is the case with nuclear MS2-GFP coat protein levels varying across cells) and variation in peak heights. The HMM is based on two states—a gene being 'on' or 'off'—and the observables being the absence or presence of the fluorescence signal of RNA over time. In this classification, the emission probability b of a given observable fluorescence intensity O given a state I is defined as:

$$b_i(O) = \frac{1}{\sqrt{2\pi}\sigma_j} exp[-\frac{(O-\mu_j)^2}{2{\sigma_j}^2}] \tag{2.3}$$

2.2.5 Data analysis

The basic forms of output data for each track are:

- Raw intensity values
- Normalized intensity values
- HMM fit of each track
- List of the calculated ON and OFF periods within each track, in units of frame number.

The distributions are plotted as normalized cumulative distribution functions and two statistics are calculated from the data: 1) The mean for the ON and OFF state with 95% confidence intervals calculated by bootstrap, and 2) A two-sample Kolmogorov-Smirnov test to compare the ON and OFF time distributions between two conditions.

Chapter 3

Optogenetic Engineering of the MYC Oncogene

3.1 Controlling MYC translocation with the *Avena sativa* LOV2 domain

Real-time changes in the ON and OFF time of a gene's transcription periods can be reflected in smFISH measurements in different ways depending on downstream factors like mRNA processing and decay. The longer the window of time observed, the more likely secondary effects come into play that affect the population-level snapshot of gene expression. To achieve acute control of MYC overexpression on fast timescales in living cells, we engineered photo-inducible MYC (Pi-MYC') whereby nuclear translocation of MYC is controlled by light. This capability is achieved using the light oxygen voltage 2 (LOV2) phototrophic domain from the *Avena sativa* plant. The Jα helix at the C-terminus of the LOV2 domain is coiled in the dark (**Figure 3.1**), but when irradiated with blue-green wavelengths of light (450-500nm), excites a flavin cofactor to elicit a conformational change in the domain. This process is reversible when the activating light is removed. We sought to use the caging mechanism of the LOV2 domain to control the nuclear translocation of MYC.

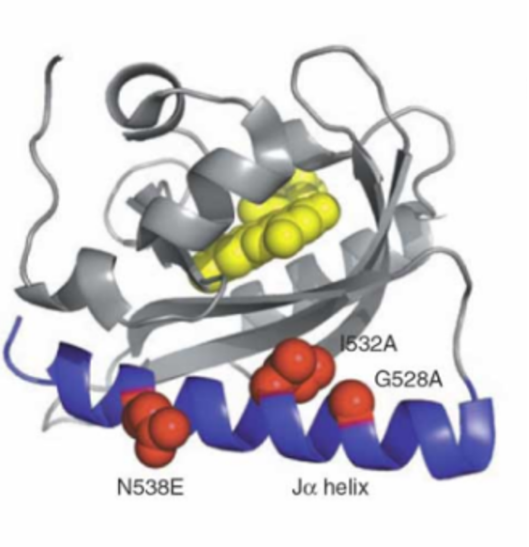


Figure 3.1: Crystal structure of the LOV2 domain (PDB 2V1A).

3.2 Overview of the Pi-MYC Transgene

The Pi-MYC transgene consists of MYC-mCherry fused to a nuclear export signal (NES) followed by the LOV2 domain enclosing the c-MYC nuclear localization signal (NLS) (**Figure 3.2**). This transgene is driven by the human Ubiquitin C promoter, allowing MYC-mCherry to be constitutively expressed yet retained in the cytoplasm due to the combination of an exposed NES and LOV2-caged NLS. Upon induction with light, the LOV2 domain unhinges and exposes the NLS, allowing MYC to be imported into the nucleus. We found it was necessary to inactivate the endogenous MYC NLS located in exon 3 of the gene in order for MYC to be under control of the LOV2-caged NLS. To do this we replaced the basic lysine and arginine residues of the 'PAAKRVKLD' sequence with alanine (**Figure 3.2**). As a control to make sure any effects we saw in our experiments were specifically due to MYC, we created "Pi-mCherry"— a transgene that contains all the domains except MYC.

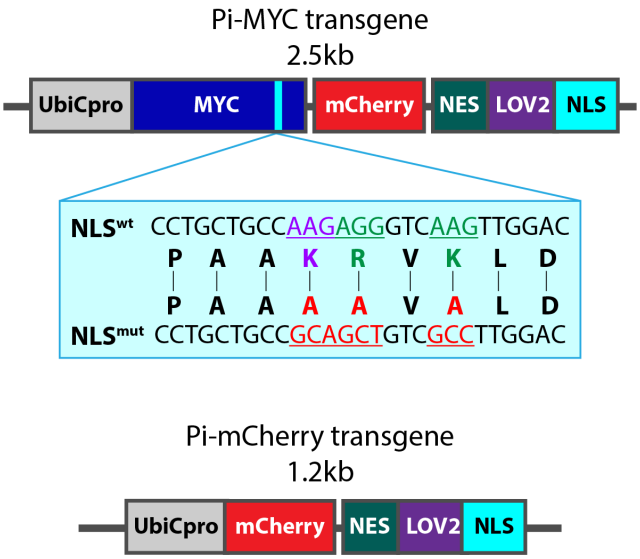


Figure 3.2: Schematic of Pi-MYC. The transgene consists of the human Ubiquitin C promoter (UbiCpro), *c-myc* exons 2 and 3 (MYC), alanine mutations in the native NLS (blue inset shows changed residues), mCherry, a nuclear export signal (NES), and the LOV2 domain followed by the wildtype *c-myc* NLS sequence. The Pi-mCherry control transgene contains all domains except MYC.

3.3 Translocation Kinetics

To test the import and export kinetics, we introduced Pi-MYC via lentiviral integration into U2-OS cells. We found MYC exhibited reversible nuclear translocation within minutes of addition and removal of activating 488nm light (**Figure 3.4**). We found the translocation was reversible over repeated induction periods of 1 hour with alternating 10-minute periods of induction light, with an entry rate of $k_{1/2} = 0.380 \pm 0.018$ min^{-1} and an exit rate of $k_{1/2} = 0.261 \pm 0.022$ min^{-1}(**Figure 3.3**). Analysis of Pi-MYC nuclear fluorescence intensity in cells before and after induction shows a 1.6-fold increase in nuclear MYC levels under these illumination conditions.

3.4 Biological Validation

We next determined whether Pi-MYC was biologically functional using a classic transformation assay in NIH3T3 mouse fibroblast cells [62]. In this assay, introduction of two oncogenes—RAS containing a G12V mutation and MYC overex-

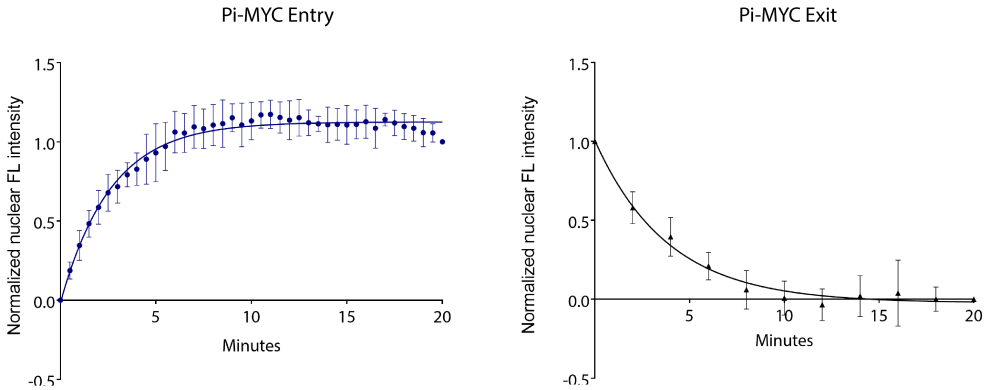


Figure 3.3: Pi-MYC entry and exit rates in U2-OS cells fit to a single exponential model. Data points are mean with 95% confidence interval of n=5 cells.

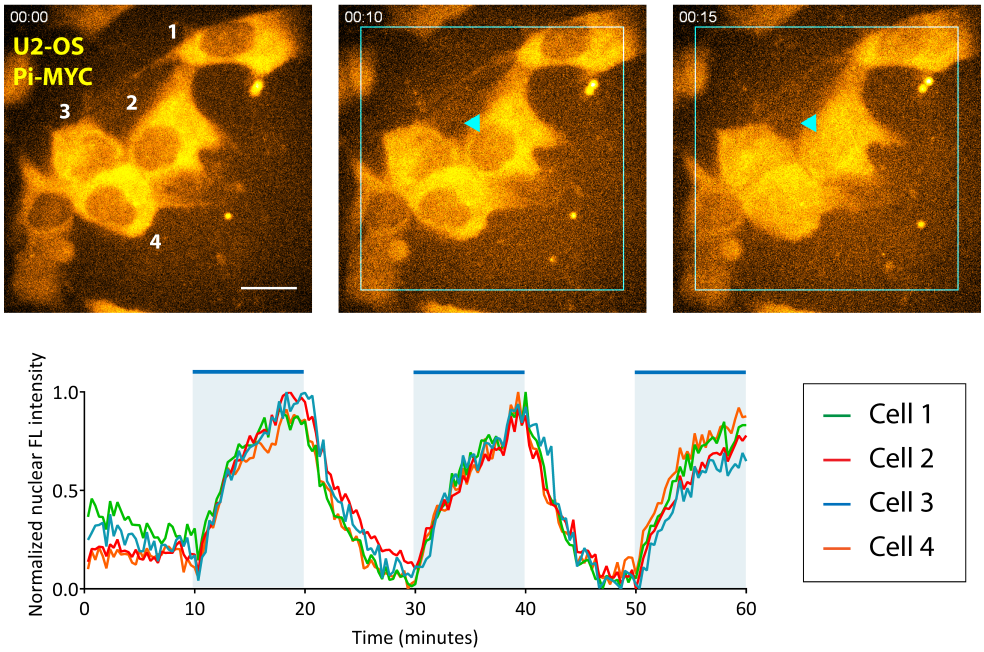


Figure 3.4: Top: Pi-MYC stable expression in U2-OS cells. Scale bar = 15µm. In the absence of irradiating light, the NES allows Pi-MYC retention in the cytoplasm. Upon irradiation with blue-green wavelengths of light (450-500nm), the LOV2 domain exposes the enclosed NLS and allows Pi-MYC to be imported into the nucleus (indicated by blue triangles). Pi-mCherry operates via the same mechanism. **Bottom:** Quantification of nuclear fluorescence intensity from four cells numbered in (B) over a 1-hour time series. The field of view was subjected to alternating 10-minute periods of activating light indicated by blue rectangles on plot.

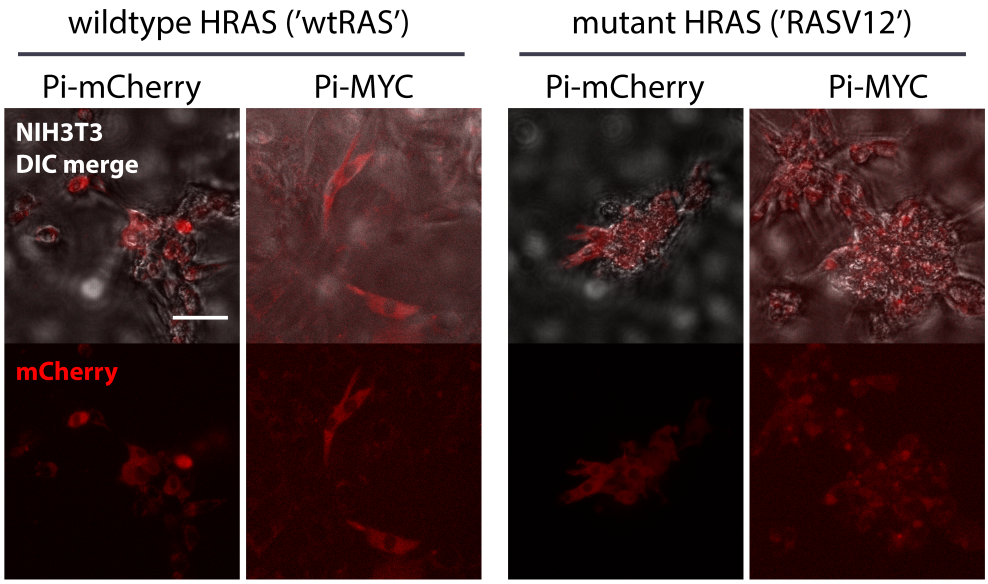


Figure 3.5: DIC merge and fluorescence images of NIH3T3 fibroblasts stably expressing Pi-mCherry or Pi-MYC (visualized with mCherry) in a background of wildtype HRAS or V12 mutant. Scale bar = 50µm.

pression—causes transformation of mouse fibroblasts, resulting in several hallmark phenotypes of oncogenesis: focus formation, faster growth in culture, and anchorage-independent growth in soft agar. We generated NIH3T3 cell lines that expressed either a background of wildtype RAS or mutant RASV12, combined with either Pi-mCherry or Pi-MYC stable expression. Pi-MYC and Pi-mCherry expression is visible by fluorescence microscopy (**Figure 3.5**), and both MYC and RAS protein levels can be detected by Western blot (**Figure 3.6**). We found that Pi-MYC co-expressed with mutant RAS transformed NIH3T3 cells, resulting in focus formation, faster growth in culture (**Figure 3.7**), and anchorage independent growth (**Figure 3.8**). We note that the NIH3T3 cells were sensitive enough that the transformation occurred without needing to specifically incubate them with blue light. We reason this 'leakiness' could be due to the LOV2 domain being activated by trace amounts of room light, or some LOV2 domains being improperly folded and not able to completely cage the NLS. Nevertheless, these data indicate that Pi-MYC is biologically functional, allowing us to probe for transcription changes when perturbing

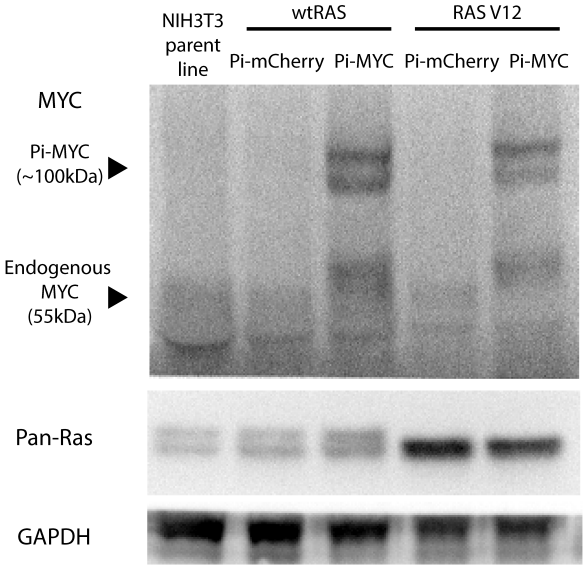


Figure 3.6: Western blot of MYC and RAS expression in the NIH3T3 stable lines.

cells on short timescales.

3.5 Stable Expression of Pi-MYC in HBEC Cells

Having confirmed that Pi-MYC is both controllable and functional, we introduced Pi-mCherry or Pi-MYC into HBEC cells by lentiviral integration. We created a total of 10 stable cell lines where either Pi-mCherry or Pi-MYC was expressed in five clonal backgrounds, the MS2 polyclonal line and four single-cell clones *RPAP3, RAB7A, KPNB1, MYH9*, via lentiviral transduction (**Figure 3.9**). The MS2 polyclonal cell line consists of introns of approximately 900 unique genes labeled with MS2 and so we used an smFISH probe set to MS2 to detect the activity of nascent RNA at transcription sites (**Figure 3.10**). We found that even under conditions of no activating light, the Pi-MYC cell line exhibited brighter TS compared to the Pi-mCherry line, indicating the production of more nascent RNA at active TS (**Figure 3.11**). This result may follow from slight leakiness of the LOV2 cage, consistent with what we saw in NIH3T3 cells, even though by fluorescence microscopy MYC appears to be excluded from the nucleus. Similarly, Pi-MYC induced a growth

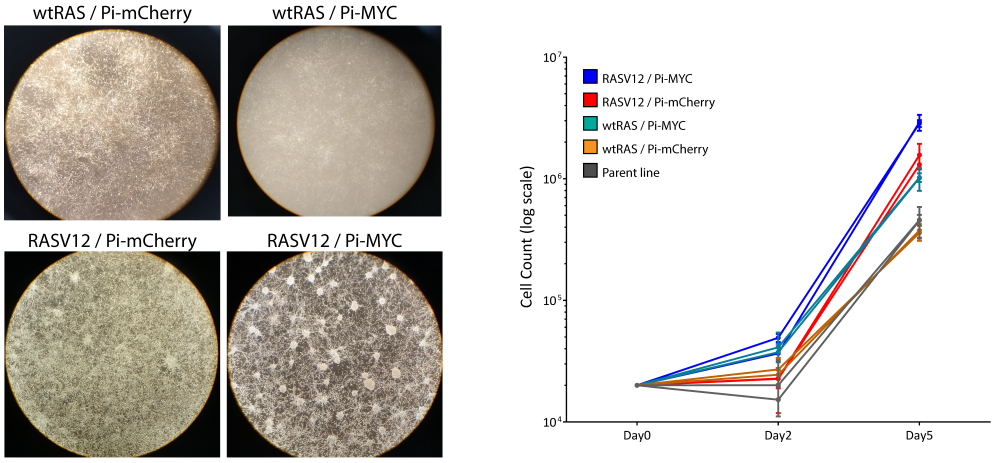


Figure 3.7: Left: Growth and focus formation in monolayer culture of four NIH3T3 lines. **Right:** Quantification of mean growth rates, two biological replicates per cell line. Error bars are SD of three technical replicates.

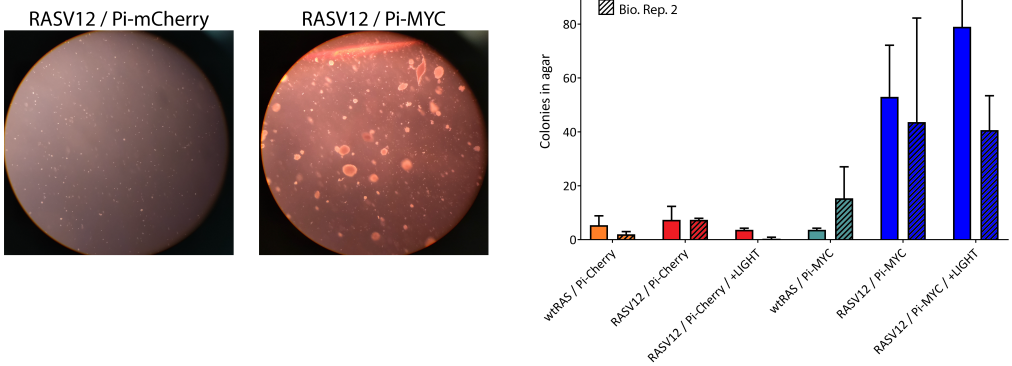


Figure 3.8: Left: Colony formation in soft agar of Pi-mCherry and Pi-MYC stable lines in the RAS mutant background. **Right:** Quantification of colonies visible after 2 weeks growth in soft agar of the four stable NIH3T3 lines and the two RASV12 lines cultured in 24 hours light before embedding in agar (+LIGHT). Bars are mean with SD of two biological replicates.

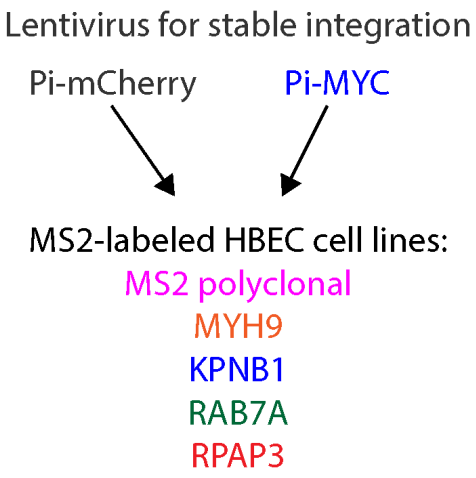


Figure 3.9: Diagram of HBEC cell lines generated by stably integrating Pi-mCherry or Pi-MYC into the MS2 polyclonal cell line and four single cell clones (10 cell lines total).

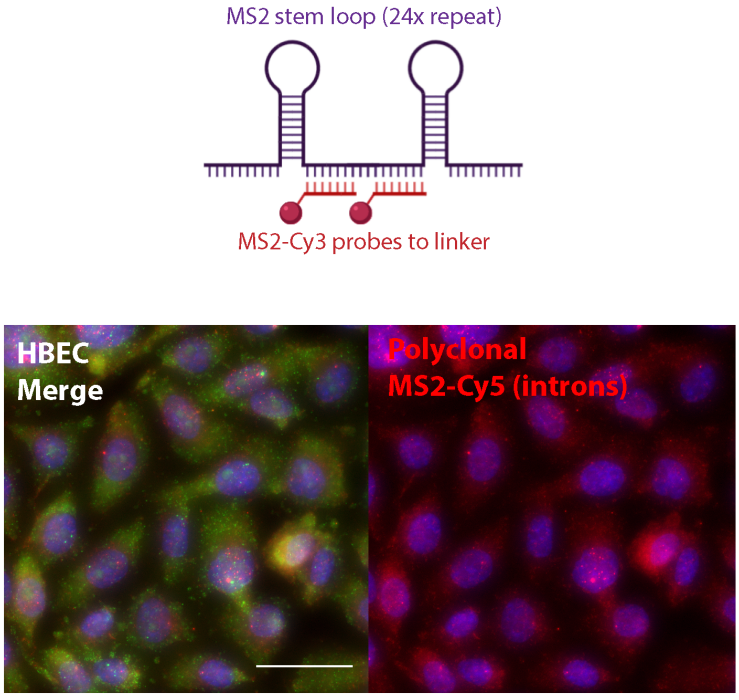


Figure 3.10: **Top:** MS2-Cy3 smFISH probes are localized to the linker region between a single MS2 stem loop repeat. **Bottom:** smFISH images of the MS2 polyclonal line (blue=DAPI) with MS2 probes to highlight the nascent RNA (introns) of genes. Scale bar = 30µm.

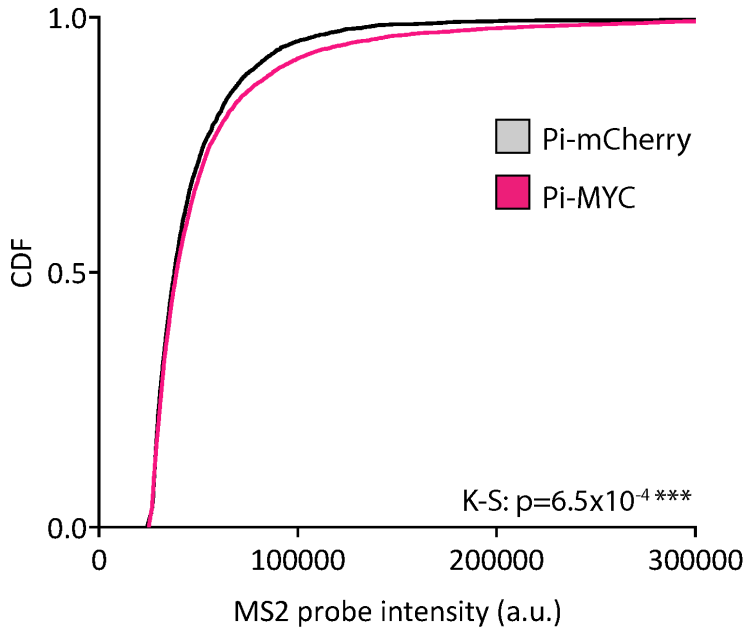


Figure 3.11: Normalized CDF of MS2 probe intensity in the HBEC polyclonal line, indicating the number of nascent RNA present globally at transcription sites with stable expression of Pi-mCherry (n=3364 cells) or Pi-MYC (n=6670 cells). Shown are two biological replicates combined per condition.

advantage in HBEC cells over a period of two days in culture (**Figure 3.12**). However, in contrast to NIH3T3 cells, this growth difference was more pronounced when the cell line was cultured under blue light. Interestingly, the Pi-mCherry line also displayed increased growth when cultured under blue light. It has been previously reported that blue and/or green light may have antioxidant effects [18], have protective effects on protein subjected to UV irradiation [33], and initiate biophysical processes in cells such as increased wound healing and cell proliferation [2, 113]. These data speak to the importance of using the Pi-mCherry control for our experiments where the cell is subjected to all factors involved in the light induction except translocation of MYC.

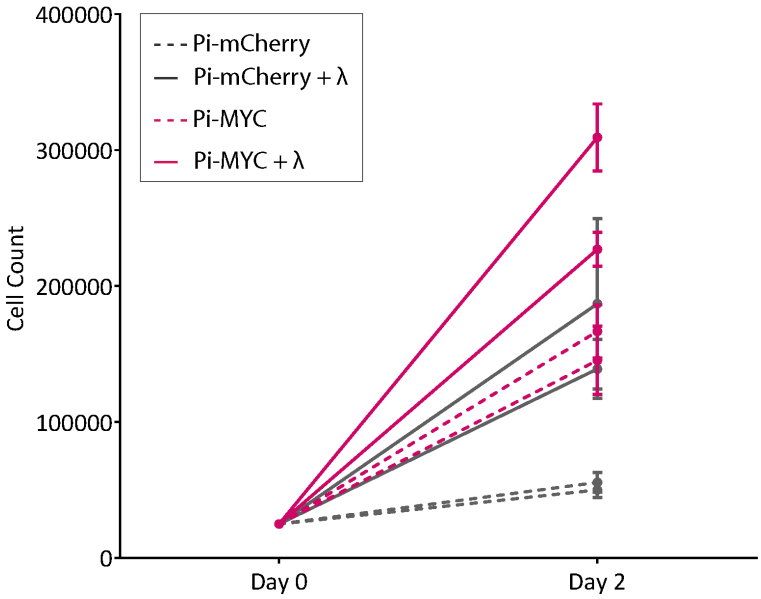


Figure 3.12: Two-day growth curve of HBEC cells with Pi-mCherry or Pi-MYC stable integration, cultured in the dark or under 455nm light (+λ).

Chapter 4

Investigating the Effect of MYC on Gene Expression with Single-Molecule Imaging

4.1 MYC Increases the Active Transcription Period of the Estrogen-Responsive Gene *TFF1* in MCF7 Breast Cancer Cells

Our strategy to dissect the direct effect of MYC on human genes was to use multiple methods for changing MYC levels (including transient transfection, induction of stably-integrated genes, siRNA knockdown, and the photo-inducible MYC described later) followed by multiple methods of transcriptional readout (including RT-qPCR, nascent RNA counting by smFISH, live-cell-imaging of RNA synthesis by MS2). The first candidate gene we chose was *TFF1*, an estrogen-responsive gene that is upregulated in breast cancer [4, 123]. *TFF1* has a canonical E-box motif (a DNA response element that MYC binds) in its promoter as well as its enhancer, *TMPRSS3* (**Figure 4.1**). Studies with luciferase reporters show MYC interacts with the *TFF1* gene although its influence on other members of the TFF family are stronger [3]. *TFF1* RNA is transcribed in bursts, with an average ON time of 16.0 ± 0.5 minutes and an average OFF time of 66 ± 7 minutes when fully induced with estradiol (E2) [112]. The EC50 of *TFF1* when induced with E2 is 50pM [83] and increasing E2 increases the frequency of *TFF1* bursts, but has no effect on the

Figure 4.1: Genome browser screenshot of MYC ChIP-seq peaks (gray) and the E-box consensus motif (blue) at the loci of *TFF1* and its enhancer *TMPRSS3*.

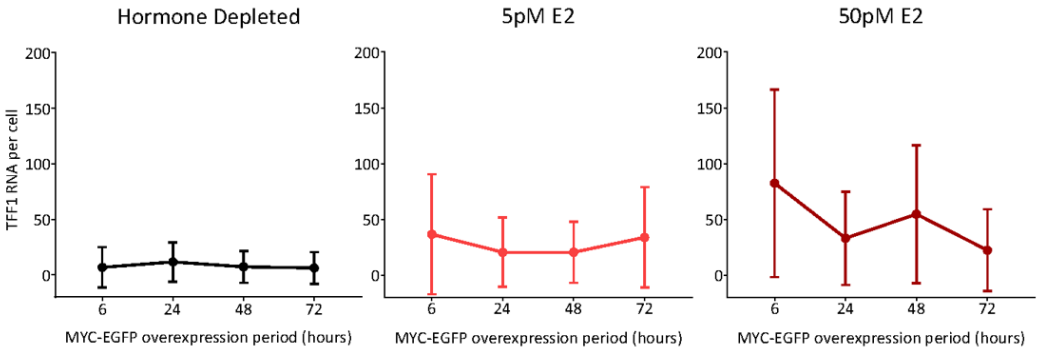


Figure 4.2: smFISH of TFF1 RNA abundance levels at varying E2 concentrations over a 72-hour time course of MYC overexpression

duration of bursts.

We used a previously-created MCF7 cell line where *TFF1* contains 24x repeat of MS2 loops in the 3'UTR, thus enabling live-cell imaging of transcription (**Figure 4.3**) [112]. First, we overexpressed MYC with transient transfection, and observed no detectable changes in *TFF1* mRNA abundance as measured by smFISH over a 72-hour MYC overexpression time course (**Figure 4.2**), similar to previous results obtained by a luciferase assay [3]. We sought to evaluate this finding by directly measuring pre-mRNA synthesis of *TFF1* at the transcription site (TS). We subjected MCF7 *TFF1*-MS2 cells to either 1 week of growth in saturating E2 or hormone depletion followed by induction with 50pM E2, followed by MYC plasmid overexpression or MYC siRNA knockdown (**Figure 4.4**). Cells in each condition were imaged for a timeseries of 14.2 hours and the *TFF1* TS were tracked over that duration (**Figure 4.4**). The TS intensity is a direct readout of the nascent RNA produced at the *TFF1* gene locus, and we fit the normalized time traces of *TFF1* fluores-

Figure 4.3: Schematic of the MS2 insertion in the *TFF1* endogenous locus.

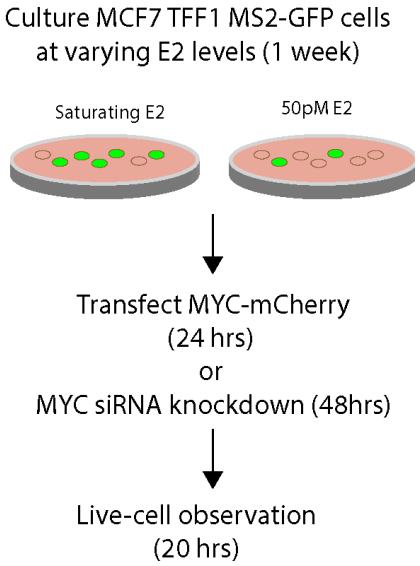


Figure 4.4: Schematic of live-cell experiment with MYC perturbation.

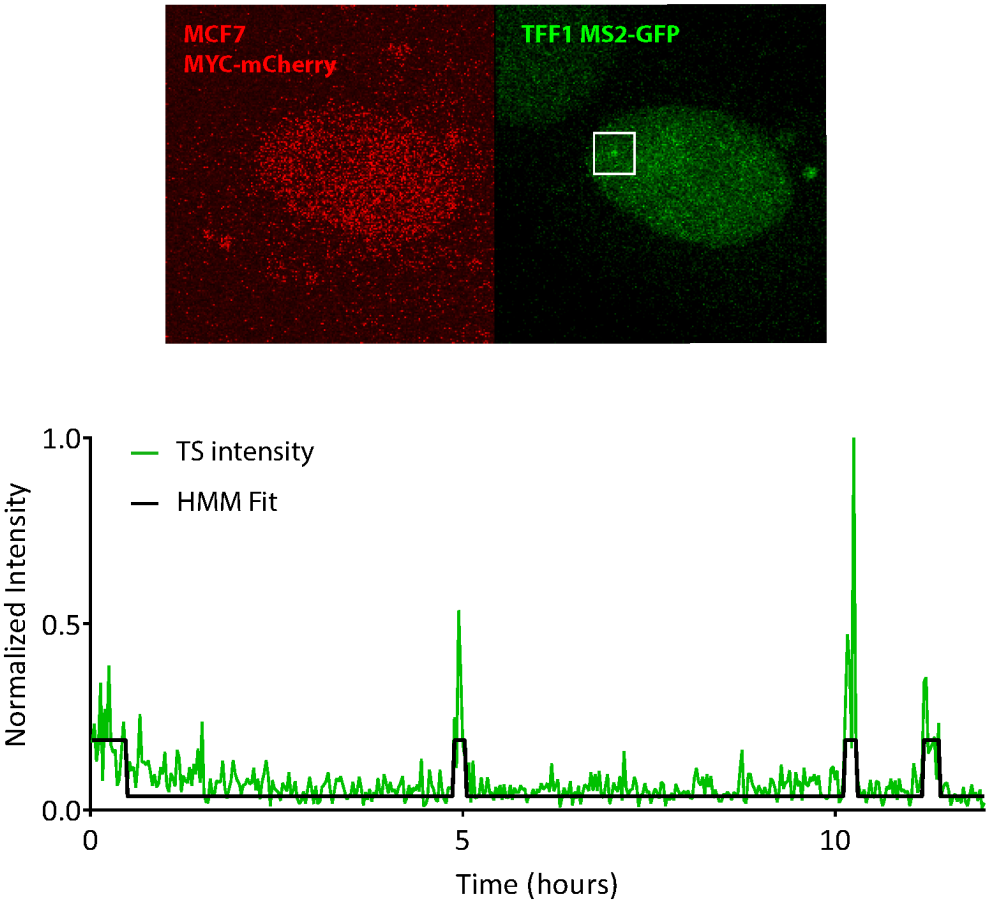


Figure 4.5: Top: Example screenshot of a live-cell movie in a MCF7 cell with MYC-mCherry overexpression. Frames are a maximum projection of a z-stack. *TFF1* transcription site indicated by white square. **Bottom:** Example trace of *TFF1* transcription site activity (green line) fit to a 2-state HMM (black line) to categorize gene activity into 'ON' and 'OFF' periods.

cence intensity to a 2-state HMM to threshold the ON and OFF periods of the gene (**Figure 4.5**). We then plotted the distributions of ON and OFF periods as a normalized cumulative distribution function (CDF) to determine whether MYC modulates the conditions. We found no effects on transcription ON and OFF time at saturating E2 conditions, but we did find that MYC affected *TFF1* transcription events at 50pM E2, the EC50 for the gene (**Figure 4.6**). We observed a significant increase in the ON time distribution with the mean increasing from 5.6 ± 0.4 to 7.1 ± 0.7 minutes, as well as an increase in the OFF time from 44.2 ± 4.3 to 70.2 ± 7.0 minutes. In summary, despite not seeing changes in total cellular *TFF1* mRNA, we

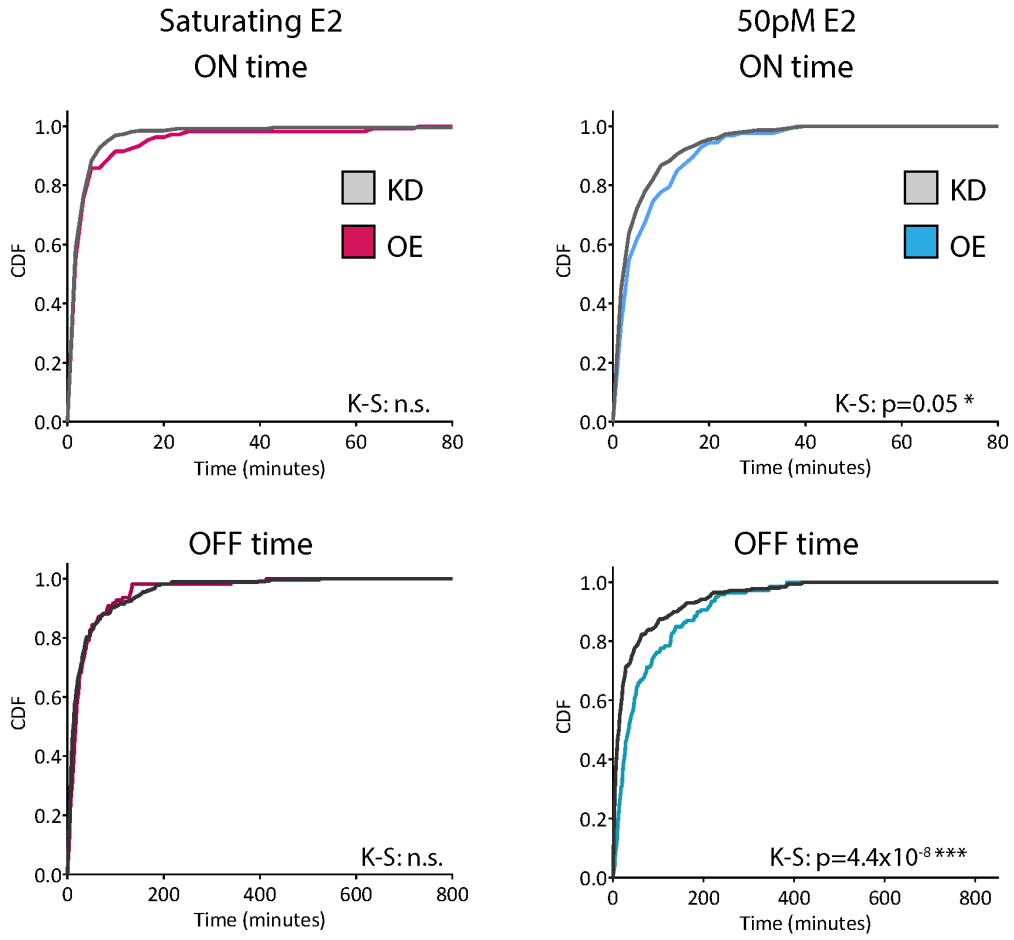


Figure 4.6: Normalized CDF of *TFF1* ON and OFF times with MYC knockdown (KD) or overexpression (OE). Cells were cultured at saturating E2 (KD=15 cells, OE=8 cells) or 50pM E2 (KD=22 cells, OE=13 cells).

observed that MYC increased the duration of *TFF1* transcription events, but also decreased the frequency of the events occurring. Yet, this effect was only observed under non-saturating levels of expression.

We also tested our findings for the *TFF1* gene transcribing under non-saturating conditions (50pM E2) with an orthogonal method of MYC overexpression that has been previously characterized [103]. In this MCF7 cell line, a stably-integrated TetON MYC transgene induces MYC-EGFP overexpression upon induction with doxycycline (**Figure 4.7**). Under the microscope, this difference in MYC level

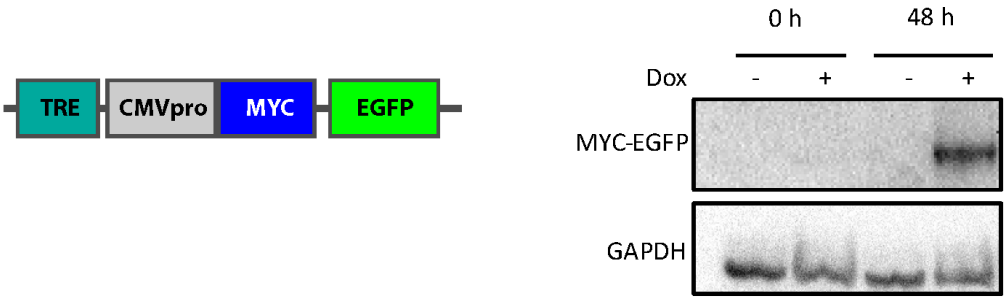


Figure 4.7: Left: Schematic of TetON MYC-EGFP transgene, containing a doxycycline inducible Tet-responsive element, and a CMV promoter driving expression of MYC-EGFP. **Right:** MYC western blot of MCF7 TetON MYC-EGFP expression at 0 and 48 hours of vehicle or doxycycline induction.

upon induction can be detected both at the protein level with MYC-EGFP fluorescence and at the RNA level with smFISH (**Figure 4.8**) Induction with doxycycline leads to a 1.7-fold increase in MYC RNA levels over 48 hours, from 40.2 $\pm$ 0.4 to 67.7 $\pm$ 0.7 RNA per cell (**Figure 4.9**). Using an smFISH probe set to the intron of *TFF1* to identify nascent unspliced pre-mRNA at the TS, we found that TS intensity increased slightly from 0-24 hours of a 48-hour time course, whereas the fraction of active TS per cell peaked at 6 hours and then decreased below initial levels (**Figure 4.10**). Taken together, two different methods of MYC over-expression and two different methods of transcriptional readout indicate MYC-dependent changes in *TFF1* transcription that were not visible at the total mRNA level.

We verified these findings for another gene in a different transformed cell line. We used a previously-created TetON beta-globin reporter gene stably integrated into the genome of U2-OS osteosarcoma cells [20]. The expression level of the beta-globin reporter is tunable by doxycycline and driven by a bacterial CMV promoter. Like *TFF1*, this gene also contains 24x MS2 stem loops in the 3' UTR, but unlike *TFF1*, the beta-globin reporter is exogenously expressed, present as a multi-copy array, and does not contain any cis-acting elements designed to be MYC-responsive. Interestingly, we saw that MYC overexpression over a 24-hour period increased the RNA abundance above the levels achieved by doxycycline and increased the mean

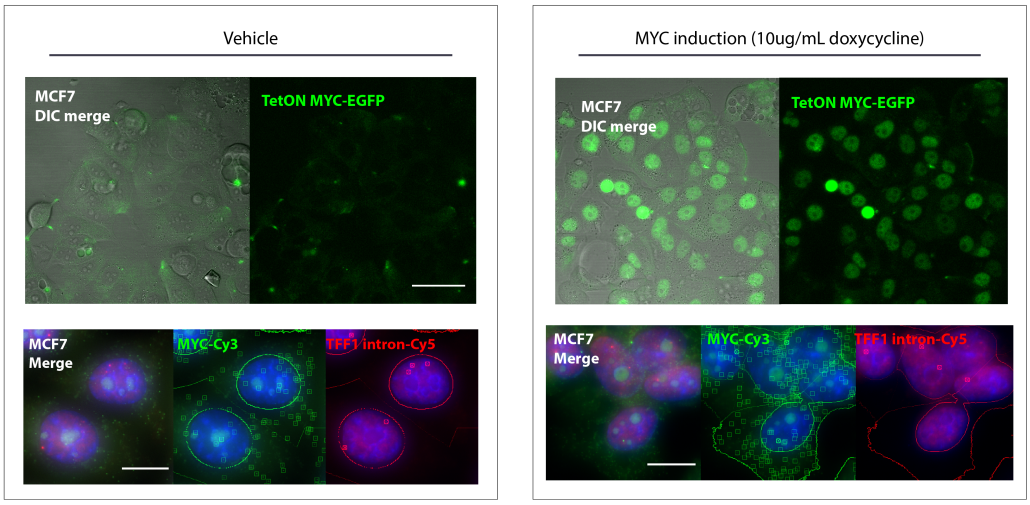


Figure 4.8: Left: TetON MYC-EGFP expression induced with vehicle. The top panel shows DIC-merged and fluorescence images of the TetON MYC-EGFP protein expression (scale bar=50µm).The bottom panel shows smFISH images of merged, MYC, and TFF1 intron channels (blue=DAPI, scale bar=12µm.). RNA transcripts and transcription sites are identified with squares and circles, respectively. **Right:** TetON MYC-EGFP induced with doxycycline for 24 hours.

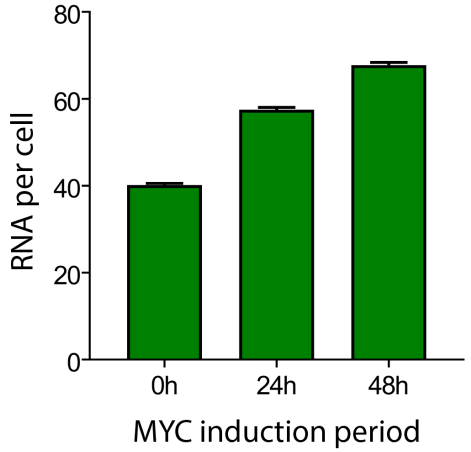


Figure 4.9: smFISH quantification of MYC RNA per cell at 0 (n=3448 cells), 24 (n=7890 cells), and 48 hours (n=6599 cells) of TetON MYC-EGFP induction with 10µg/mL doxycycline. Bars represent mean with SEM.

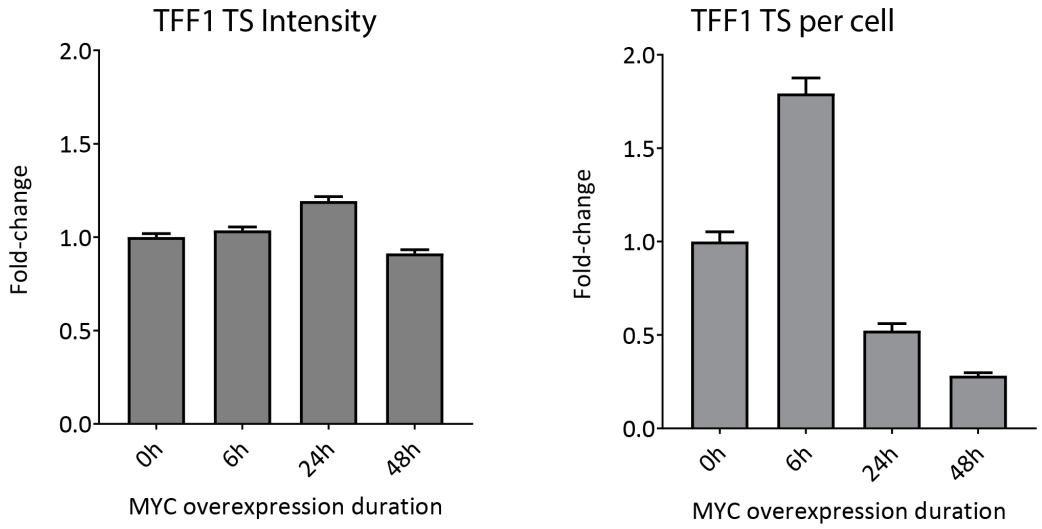


Figure 4.10: Left: smFISH quantification of fold-change in fluorescence intensity (dox/veh, normalized to 0h) of *TFF1* TS over 48-hour time course of MYC overexpression. **Right:** smFISH quantification of fold-change in *TFF1* TS per cell (dox/veh, normalized to 0h) over 48-hour time course of MYC overexpression. Approximately 2000-6000 cells were imaged per condition.

ON time from 32.1 ± 2.0 to 41.4 ± 5.3 minutes (**Figure 4.11**). We observed a slight increase in the mean OFF time by a similar magnitude (70.8 ± 5.5 to 79.3 ± 9.9 minutes), but this effect was not statistically significant. Thus, even a transgene with no canonical E-box sequences in the promoter proximal region responds to MYC over-expression over long timescales. In summary, single-cell imaging of two genes (*TFF1* and a beta-globin reporter) in two transformed cell lines (MCF7 and U2-OS) show that MYC increases burst duration.

4.2 MYC Exerts Changes in Gene Expression in the Non-Transformed HBEC Cell Line

In order to expand our single-cell, single-gene survey of MYC's effects on transcription to genes that exhibit a range of functions and expression levels and also to examine MYC overexpression in a non-cancer context, i.e. an immortalized human cell line that was not transformed), we studied the role of MYC in a human bronchial epithelial cell line (HBEC) [107], immortalized through hTERT and Cdk4 expres-

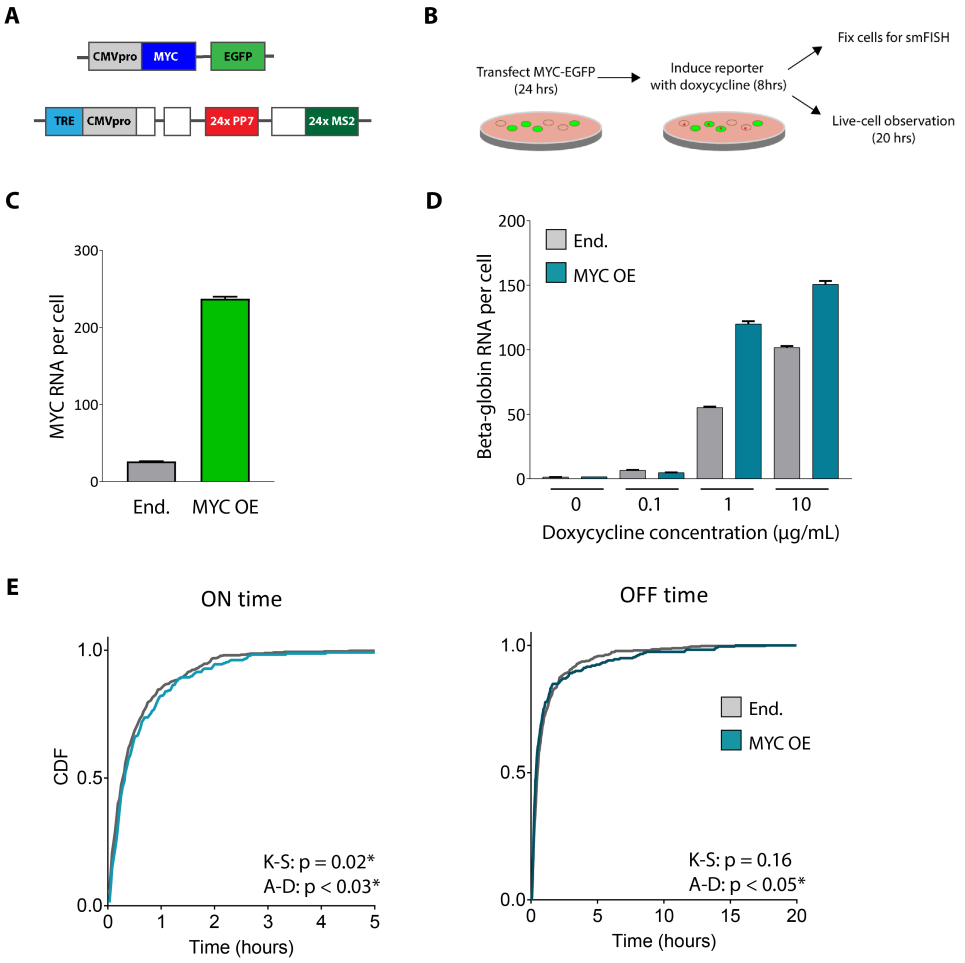


Figure 4.11: **(A)** Schematic of the TetON beta-globin reporter gene stably integrated in U2-OS cells, and the MYC-EGFP transgene added as a plasmid transfection. For live-cell imaging, transcription activity was measured from the PP7 stem loop signal in intron 2 observed by the fluorescence of the interacting PP7-mCherry coat protein. **(B)** Diagram of experimental setup. **(C)** smFISH quantification of MYC RNA per cell with endogenous levels of MYC (End.) and 24 hours MYC-EGFP overexpression (OE). **(D)** smFISH quantification of induced beta-globin reporter expression with endogenous MYC levels or MYC overexpression. **(E)** Normalized CDF of beta-globin reporter ON and OFF times with endogenous MYC levels or MYC overexpression. Statistical significance from the Kolmogorov-Smirnov and Anderson-Darling tests are both reported.

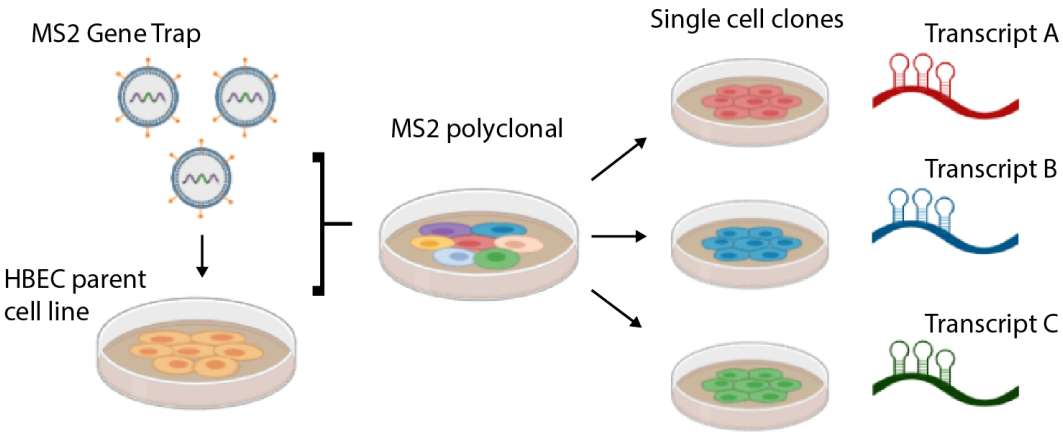


Figure 4.12: Schematic of the gene trap system to insert MS2 stem loops globally into the introns of endogenous genes in the HBEC cell line.

sion. We previously used a 'gene trap' to introduce MS2 loops globally into the introns of genes in this HBEC cell line [136]. Briefly, the gene trap works through random insertion, followed by drug selection and sequencing to identify genes containing MS2 stem loops in introns. We derived two types of cell lines from this method: An 'MS2 polyclonal' cell line in which every cell in the population contains a single unique gene tagged with MS2, and 'single-cell clones' that were derived from the MS2 polyclonal line by single-cell sorting, followed by sequencing to determine the identity of the gene that received the MS2 tag (**Figure 4.12**). We established 10 single-cell clones that each contain an MS2-tagged gene, and from bulk RNA sequencing of the HBEC parental cell line [99] we determined these genes represent the upper 50-percent range of expression of the entire transcriptome, varying from 6 to 230 RPKM (Reads Per Kilobase of transcript, per Million mapped reads) (**Figure 4.13**).

We first assayed whether any of the 10 genes for which we had MS2 live-cell clones responded to MYC perturbations. We performed 24 hours MYC overexpression (MYC-mCherry plasmid) or 48 hours MYC siRNA knockdown and determined the fold-change in expression compared to controls (mCherry plasmid or scrambled siRNA) with RT-qPCR to total mRNA (**Figure 4.14**).

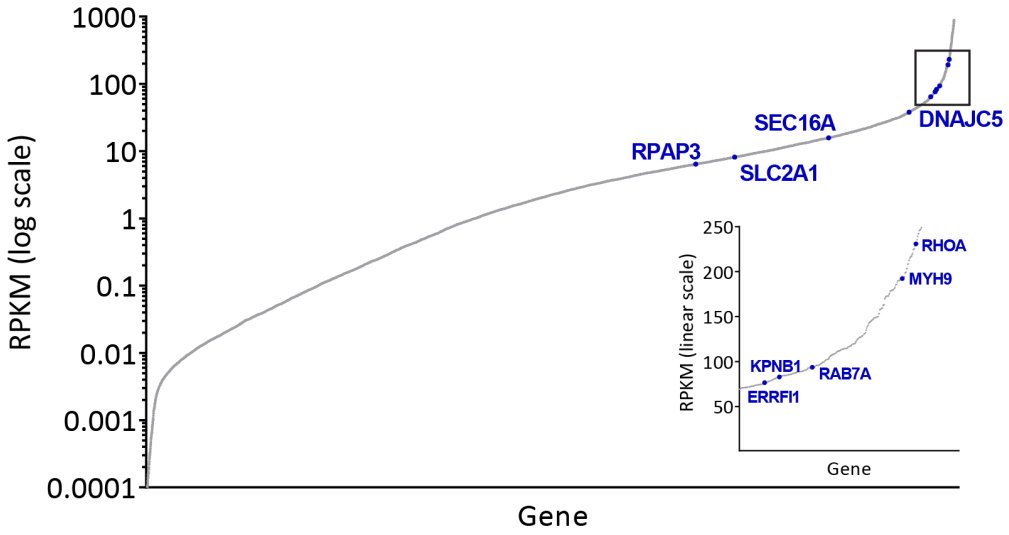


Figure 4.13: RNA expression profile of the HBEC transcriptome. Genes for which MS2 single-cell clones were generated are labeled in blue.

We found that increasing MYC levels increased the expression of most genes, and correspondingly MYC knockdown generally decreased expression. We picked four genes— *RPAP3*, *RAB7A*, *KPNB1*, and *MYH9*— for more detailed analysis because they represented a range of features. First, these genes encode for proteins with diverse cellular functions: RPAP3 is an RNAPII associated protein, RAB7A is a GTP-binding protein that is a structural component of lysosomes, KPNB1 is a member of the Importin-beta family of nuclear chaperones, and MYH9 is a subunit of the non-muscle Myosin IIA protein. Second, all these genes except *KPNB1* showed a response to MYC overexpression detectable by RT-qPCR. Third, these genes represent a range of RNA abundance levels as quantified by smFISH, from *RPAP3* with an average of 8.2 ± 0.1 RNA per cell, to *MYH9* with 140 ± 0.6 RNA per cell (**Figure 4.15**). Fourth, they represent a range of DNA features: All have MYC binding detectable by MYC ChIP in various cell lines (**Figure 4.16**). None of the genes have a canonical E-box sequence in their promoter-proximal regions (within 200 bp of the transcription start site, yellow highlight), although there are canonical E-boxes within 760bp of *RAB7A* and 900bp of *RPAP3*.

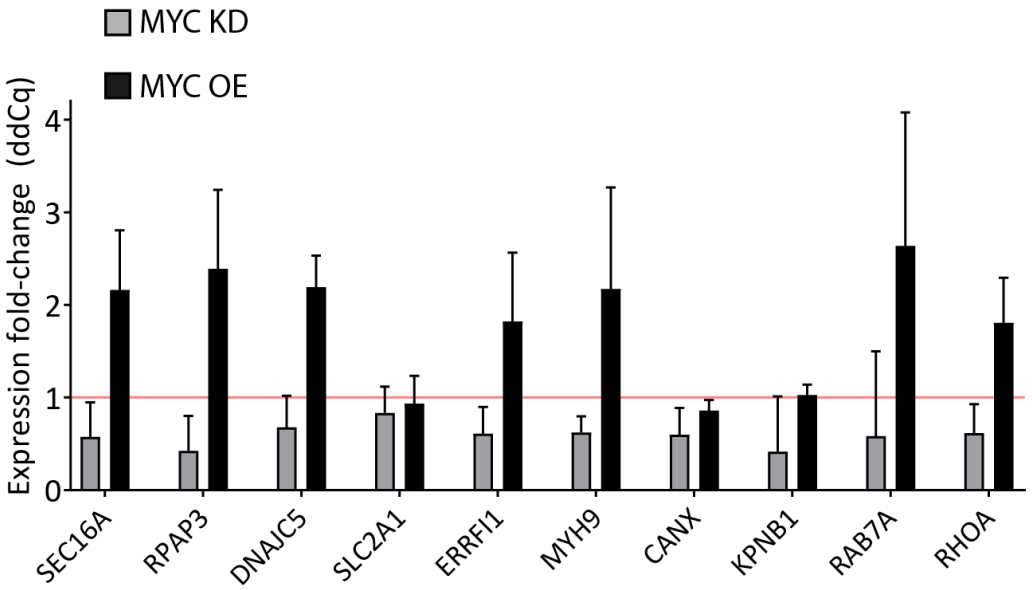


Figure 4.14: qPCR analysis of the effects of transient MYC-mCherry overexpression (MYC OE) or siRNA knockdown (MYC KD) on a panel of genes for which MS2 single-cell clones are available. Expression is normalized to the control perturbation (mCherry plasmid or scrambled siRNA).

From the unmodified HBEC parental cell line, we generated a TetON MYC-EGFP stable cell line to look at the effects of MYC overexpression on the four genes. MYC-EGFP overexpression is visible in the nuclei of HBEC cells after a few hours of doxycycline induction, and stable expression in the population can be observed by microscopy or western blot after 24 hours of induction (**Figure 4.17**). We conducted smFISH experiments in which we overexpressed MYC for 24 hours, then fixed and labeled the cells with smFISH probes to the 3'UTR of the four genes *RPAP3*, *RAB7A*, *KPNB1*, and *MYH9* (**Figure 4.18**). We found that MYC increased the RNA abundance of the lower expressed genes, *RPAP3* and *RAB7A*. Curiously, *KPNB1* showed a decrease in RNA abundance. The highest expressing gene, *MYH9*, showed no change (**Figure 4.20**). For all four genes, the number of nascent RNA measured by TS intensity did not show significant changes with MYC overexpression: all had changes of less than 1 RNA (**Figure 4.19**). The two genes that showed increased RNA abundance, *RPAP3* and *RAB7A*, also showed a 1.3-

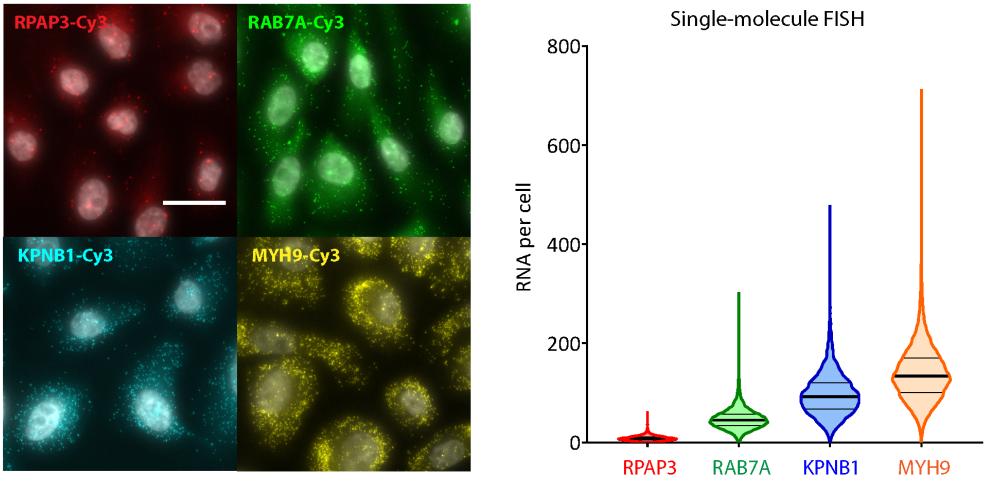


Figure 4.15: Left: Representative smFISH images of *RPAP3*, *RAB7A*, *KPNB1*, and *MYH9* RNA (gray = DAPI). Scale bar = 20µm. **Right:** Violin plot of distribution of RNA per cell of *RPAP3* (n=5955 cells), *RAB7A* (n=4911 cells), *KPNB1* (n=3511 cells), and *MYH9* (n=7999 cells). Each point represents a cell with a given number of RNA. Black bar=median; gray bars=25th and 75th quartiles.

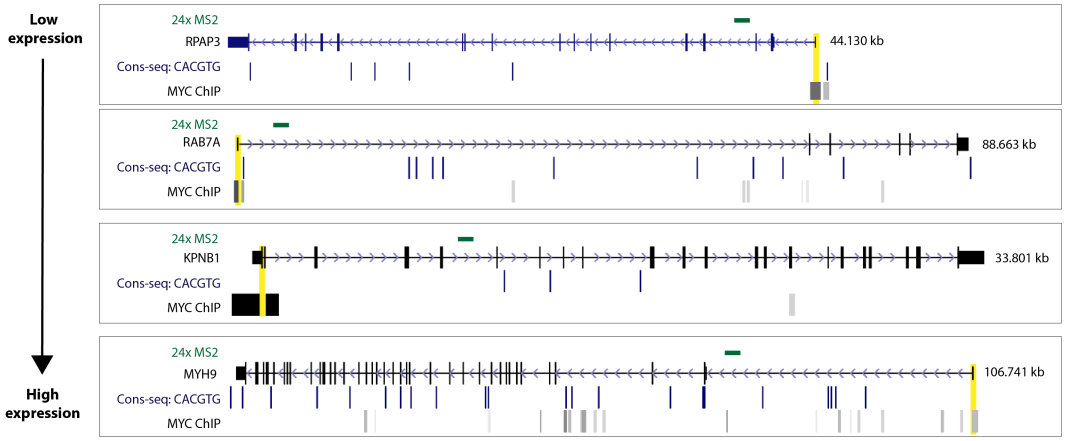


Figure 4.16: Genome browser shots of the presence of E-box consensus motifs and MYC ChIP-seq binding at four selected genes. Cell lines for which MYC binding at promoters was detected were GM12878, H1-hESC, HeLa, HUVEC, HepG2, K562, MCF7, MCF10A, and NB4. MS2 insertion location indicated in green. Promoter-proximal region indicated in yellow.

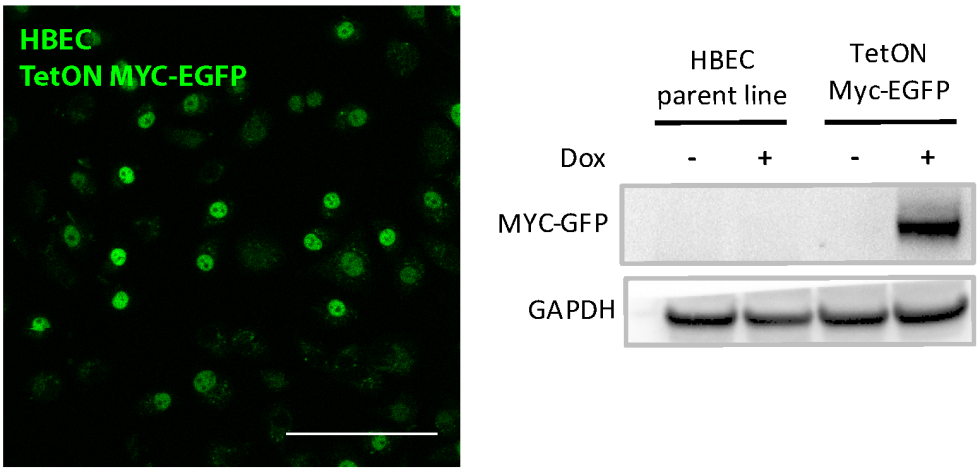


Figure 4.17: Left: Fluorescence image of TetON MYC-EGFP transgene stable expression in HBEC cells induced with 10µg/mL doxycycline for 24 hours. Scale bar = 70µm. **Right:** MYC western blot of HBEC parent line and TetON MYC-EGFP stable line with 24 hours vehicle or doxycycline induction.

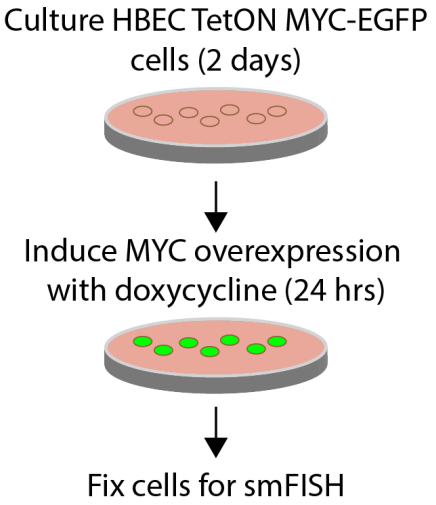


Figure 4.18: Schematic of smFISH experiment with MYC overexpression.

and 1.8- fold increase in the number of active TS.

Taken together, these data agree with previous observations of MYC effects on gene expression. Both population and single-cell methods show that the lower expressed genes, *RPAP3* and *RAB7A*, increase expression in response to MYC overexpression. The more highly expressed genes, *MYH9* and *KPNB1*, responded to MYC knockdown but show discordant results with MYC overexpression between the two assays. Previous studies indicated that highly expressed genes are MYC-saturated and fail to respond to MYC overexpression [75]. However, these assays provide a mixed view as to whether the effects are direct. They are performed at >24 hours of MYC overexpression, potentially resulting in homeostatic compensation which would obscure direct effects of MYC on transcription. We thus desired a more acute perturbation of MYC activity.

A method to acutely perturb MYC in living cells. To address the need for a more acute control of MYC activity in living cells, we engineered a photoinducible variant of MYC (Pi-MYC, described in **Chapter 3**) which allows us to control the nuclear translocation (and hence nuclear overexpression) of MYC with 450-500nm wavelengths of light. To make sure any effects on gene expression we observed were specifically due to MYC and not indirect effects of the light induction method, we created, 'Pi-mCherry' which contains all the domains of the Pi-MYC transgene except the MYC sequence (**Figure 3.2**).

4.3 Pi-MYC Overexpression on Short Timescales Reveals Genome-Wide Increases in Transcription ON Time

We proceeded to use Pi-MYC to conduct simultaneous perturbation and real-time imaging of RNA synthesis. These experiments were carried out with high-throughput

MYH9							
Function	Conventional non-muscle myosin; myosin IIA heavy chain						
Transcription Sites							
	veh1	veh2	dox1	dox2	veh (ave)	dox (ave)	%Δ TS per cell
n Cells	8695	9866	7999	6531	9280.5	7265	
n TS	5798	6414	3884	2901	6106	3392.5	
TS per cell	0.667	0.650	0.486	0.444	**0.658**	**0.465**	**0.706**
Nascent RNA							
	veh1	veh2	dox1	dox2	veh (ave)	dox (ave)	Δ nascent RNA
Mean	8.3	8.1	7.4	7.4	**8.2**	**7.4**	**-0.8**
Median	7.3	7.2	6.6	6.7	7.2	6.6	
75% percentile	9.6	9.3	8.5	8.6	9.4	8.5	
SD	3.4	3.3	3.0	2.6	3.4	2.8	

KPNB1							
Function	Importin beta family member						
Transcription Sites							
	veh1	veh2	dox1	dox2	veh (ave)	dox (ave)	%Δ TS per cell
n Cells	4876	3965	3436	3511	4420.5	3473.5	
n TS	2398	2133	1457	1258	2265.5	1357.5	
TS per cell	0.492	0.538	0.424	0.358	**0.515**	**0.391**	**0.760**
Nascent RNA							
	veh1	veh2	dox1	dox2	veh (ave)	dox (ave)	Δ nascent RNA
Mean	6.0	5.9	5.6	6.1	**5.9**	**5.8**	**-0.1**
Median	5.3	5.1	5.0	5.4	5.2	5.2	
75% percentile	6.5	6.2	6.1	6.6	6.4	6.4	
SD	3.4	4.1	2.0	2.3	3.8	2.2	

RAB7A							
Function	RAB family member; regulates vesicle traffic in the late endosomes and lysosomes						
Transcription Sites							
	veh1	veh2	dox1	dox2	veh (ave)	dox (ave)	%Δ TS per cell
n Cells	11562	9540	3546	4911	10551	4228.5	
n TS	4007	2318	1804	2897	3162.5	2350.5	
TS per cell	0.347	0.243	0.509	0.590	**0.295**	**0.549**	**1.864**
Nascent RNA							
	veh1	veh2	dox1	dox2	veh (ave)	dox (ave)	Δ nascent RNA
Mean	5.5	4.8	5.9	5.7	**5.2**	**5.8**	**0.6**
Median	4.2	3.9	4.2	4.3	4.0	4.3	
75% percentile	5.8	5.0	6.6	6.1	5.4	6.3	
SD	3.7	3.4	4.6	3.9	3.6	4.3	

RPAP3							
Function	RNA polymerase II-associated protein						
Transcription Sites							
	veh1	veh2	dox1	dox2	veh (ave)	dox (ave)	%Δ TS per cell
n Cells	9459	6881	5852	5955	8170	5903.5	
n TS	2583	2110	2410	1941	2346.5	2175.5	
TS per cell	0.273	0.307	0.412	0.326	**0.290**	**0.369**	**1.273**
Nascent RNA							
	veh1	veh2	dox1	dox2	veh (ave)	dox (ave)	Δ nascent RNA
Mean	3.9	4.7	4.4	4.6	**4.3**	**4.5**	**0.2**
Median	3.5	3.7	3.9	4.0	3.6	3.9	
75% percentile	4.3	4.7	4.8	5.0	4.5	4.9	
SD	1.8	4.0	2.2	2.3	2.9	2.2	

Figure 4.19: Fraction of active TS and number of nascent RNA at TS from two biological replicates per condition for each gene.

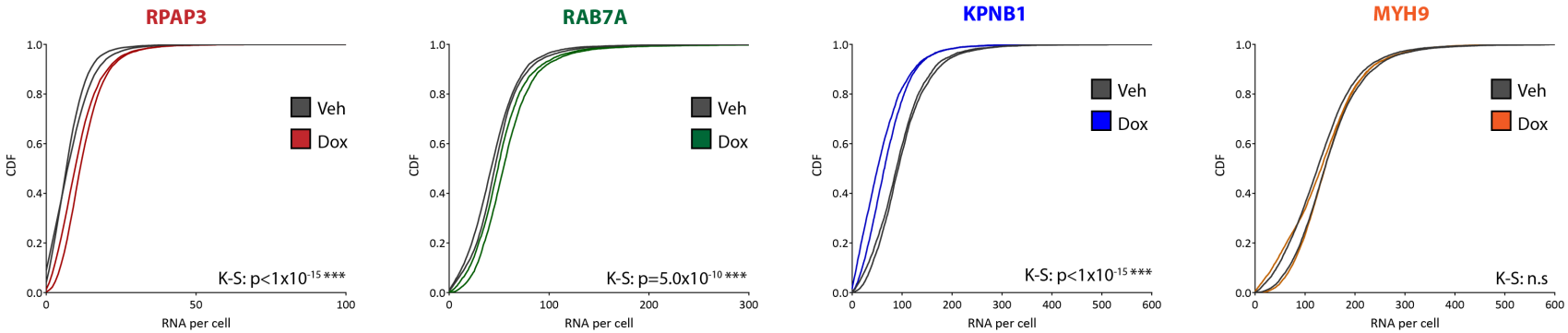


Figure 4.20: Normalized CDF of *MYH9*, *KPNB1*, *RAB7A*, and *RPAP3* RNA levels with vehicle and doxycycline induction of TetON MYC-EGFP. Two biological replicates per condition shown. 6000 cells were imaged on average per replicate (see table in **Figure 4.19**). K-S test performed on the replicates combined.

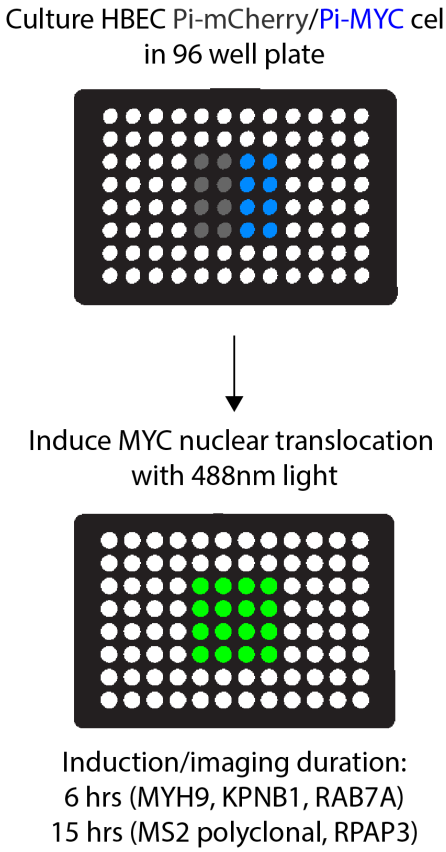


Figure 4.21: Schematic of HBEC cells grown in 96-well plates for high-throughput live-cell imaging of transcription site activity.

live-cell imaging immediately after Pi-MYC translocation to determine how MYC affected the ON and OFF periods of transcriptional bursts. For a given MS2 cell line, we cultured the Pi-mCherry and Pi-MYC variants in parallel for high-throughput imaging (**Figure 4.21**). We chose an induction/imaging duration of 6 hours in order to allow us to measure short timescales of effects while still capturing the OFF time distribution. By RNA sequencing we found that *RPAP3* is expressed 30-fold lower than *MYH9* (**Figure 4.13**), so we ultimately extended our perturbation/imaging window to 15 hours in order to more fully capture the OFF time distribution for this gene.

We induced Pi-MYC and Pi-mCherry nuclear translocation using the 488nm laser we used to image the MS2-GFP channel; in this way we could use a single wavelength of light to couple the induction/constitutive nuclear expression of the Pi-

transgene with imaging of the downstream gene. We segmented nuclei and tracked transcription site intensities of each MS2 labeled gene with an automated analysis pipeline previously developed to characterize HBEC live-cell imaging data [136]. We normalized the fluorescence intensity traces and fit each dataset to a 2-state HMM (adapted from [66]) to threshold the ON and OFF times of the gene (**Figure 4.22**). The final output of this analysis is a cumulative distribution of ON and OFF times reflecting transcription dynamics immediately (within 100 seconds) after Pi-MYC translocation.

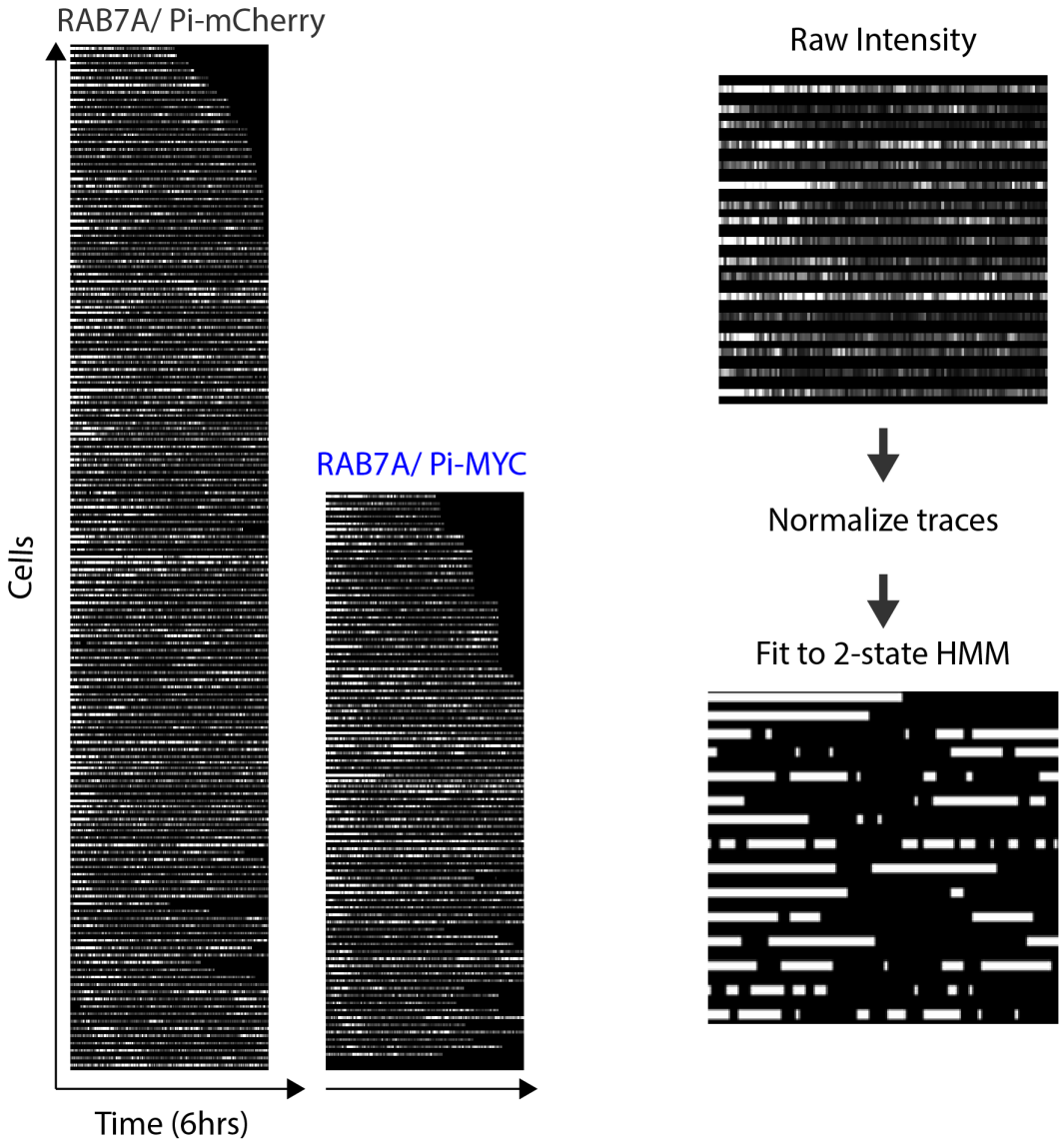


Figure 4.22: Left: Example raw data of *RAB7A* transcription site activity with 6 hours of Pi-mCherry or Pi-MYC overexpression. **Right:** Raw fluorescence intensity traces are normalized and fit to a 2-state HMM to yield 'ON' and 'OFF' periods of each gene.

We found that three of four genes showed an increase in burst duration ranging from 2.4 to 21.6 minutes. For the single-cell clones, the extent of the ON time increase was related to the expression level (**Figure 4.23**). The burst duration of the lowest expressed gene *RPAP3* increased from 11.5 ± 0.8 to 33.1 ± 2.3 minutes upon Pi-MYC translocation. The highest expressed gene, *MYH9*, showed no detectable change in burst duration (8.9 ± 1.1 to 9.5 ± 1.3 minutes). These results are in agreement with the fixed-cell data measured earlier, in that the genes showing the greatest increases in expression by RT-qPCR and smFISH at 24 hours were *RPAP3* and *RAB7A*. Interestingly, although *KPNB1* expression showed no change by RT-qPCR and a decrease in expression by smFISH, we saw a statistically significant increase in *KPNB1* burst duration by the live-cell assay. With the exception of *RPAP3*, we observed no change in the OFF periods (i.e. burst frequency) under Pi-MYC translocation conditions. Thus, live-cell imaging with Pi-MYC shows that the duration of transcription active periods is longer immediately after translocation of Pi- MYC, but that this effect can be masked by downstream effects on the overall abundance of the cellular mRNA.

Our studies on single-cell clones demonstrated the feasibility of an optogenetic real-time approach to dissecting MYC function. We next sought to extend these experiments to the MS2 polyclonal cell line which contains approximately 900 labeled genes. Although we cannot ascertain which particular gene is being recorded in each cell, the advantage of the polyclonal population is the ability to obtain a global-level measurement of gene expression. When averaged together, these data represent a unique live-cell 'metagene' analysis. For the polyclonal analysis, we used a 15-hour experimental window since this cell line represents a distribution of transcription dynamics that also includes infrequently transcribed genes. We observed a substantial global increase in ON time of 20 minutes, but the OFF time was unchanged (**Figure 4.24**). Thus, the average response to acute Pi-MYC translocation is an increase in burst duration.

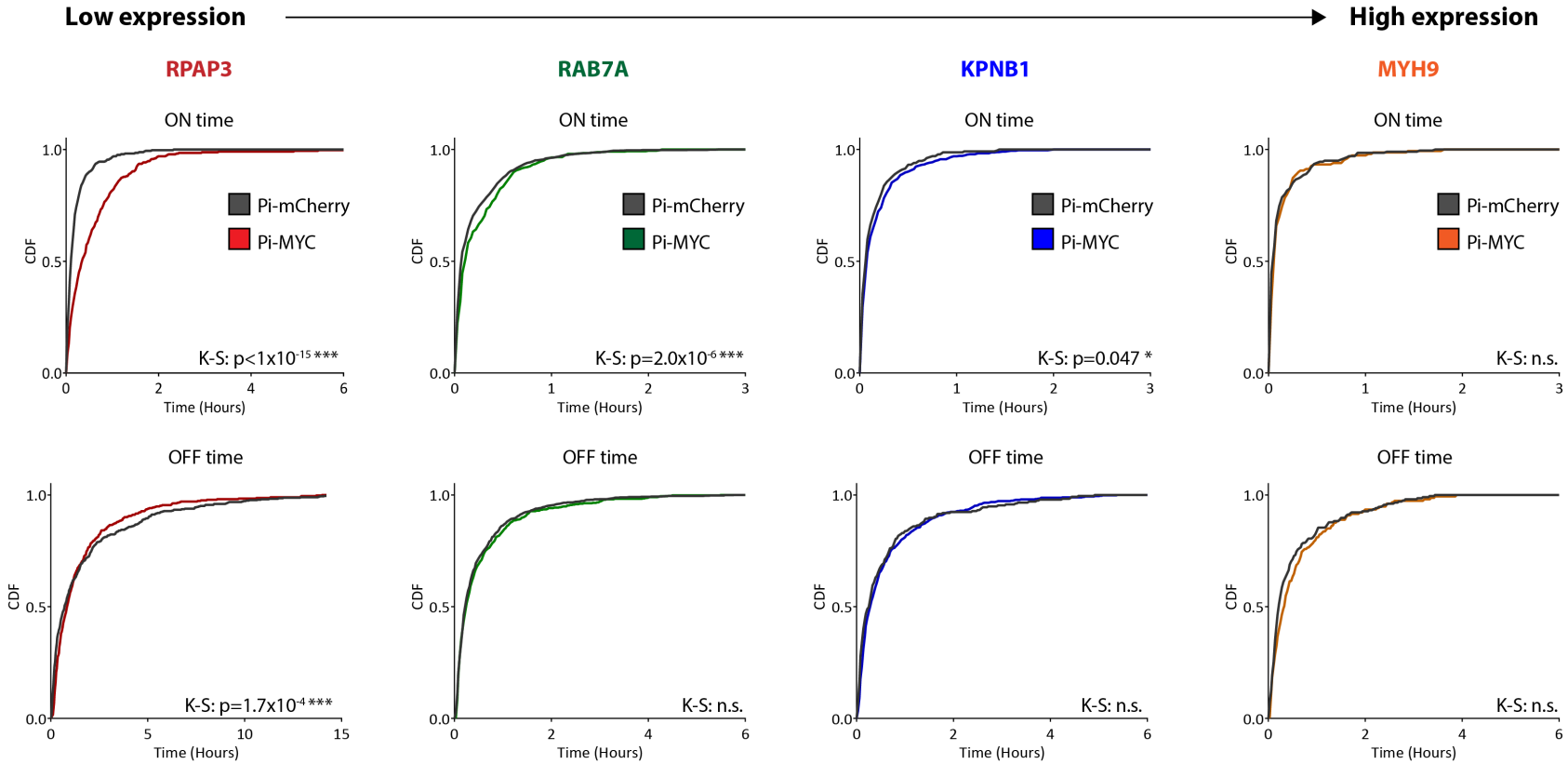


Figure 4.23: Normalized CDF of transcription ON and OFF times in HBEC single-cell clones with Pi-mCherry overexpression (RPAP3=96 cells, RAB7A=255 cells, KPNB1=40 cells, MYH9=31 cells) or Pi-MYC overexpression (RPAP3=126 cells, RAB7A=117 cells, KPNB1=92 cells, MYH9=31 cells).

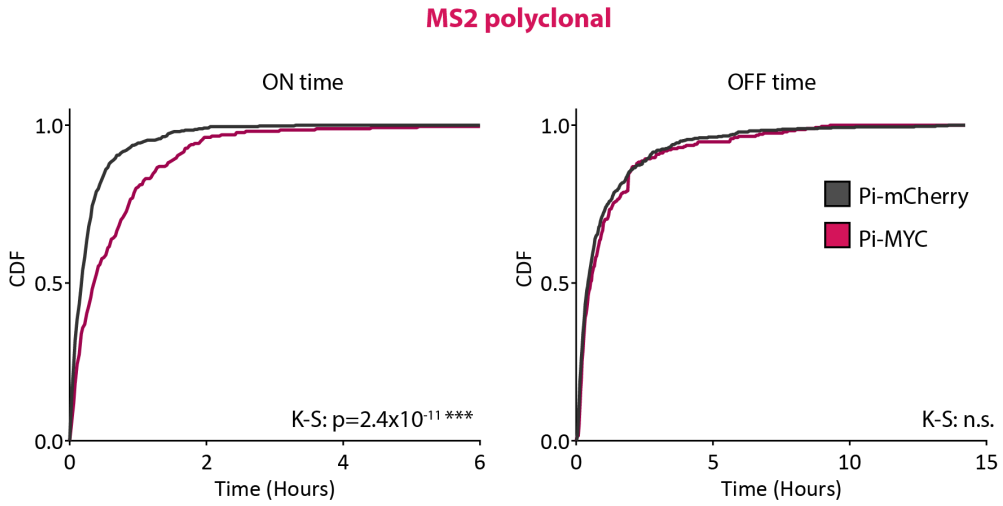


Figure 4.24: Normalized CDF of transcription ON and OFF times in the MS2 polyclonal cell line (Pi-mCherry n=83 cells, Pi-MYC n=52 cells).

Finally, we summarize the live-cell imaging data of all our genes across cell type and measurement approach as a 'delta plot' that indicates the change MYC overexpression had on the OFF and ON time of a gene compared to the control (**Figure 4.16**). Regardless of the cell type, mode of overexpression, or the gene studied, we see that at no point did MYC ever reduce the ON time of a gene. Intriguingly, MYC also appears to generally increase the time between bursts (reduced burst frequency), but this change was not statistically significant under all experimental conditions. The relative balance of these countervailing effects – increased burst duration and reduced burst frequency – determines the net RNA synthesis rate.

4.4 MYC Globally Affects Residence Times of Transcription Factors involved in Pre-Initiation Complex Formation and Pol II Pause Release

During the course of our experiments, we observed that Pi-MYC translocation kinetics varied across cell types. Specifically, fluorescence time-lapse imaging indicated that nuclear export was slower in the non-transformed HBEC cells than either MCF7 or U2-OS, both of which are cancer cell lines. We found that HBEC Pi-MYC cells displayed nuclear retention even when the induction light was removed

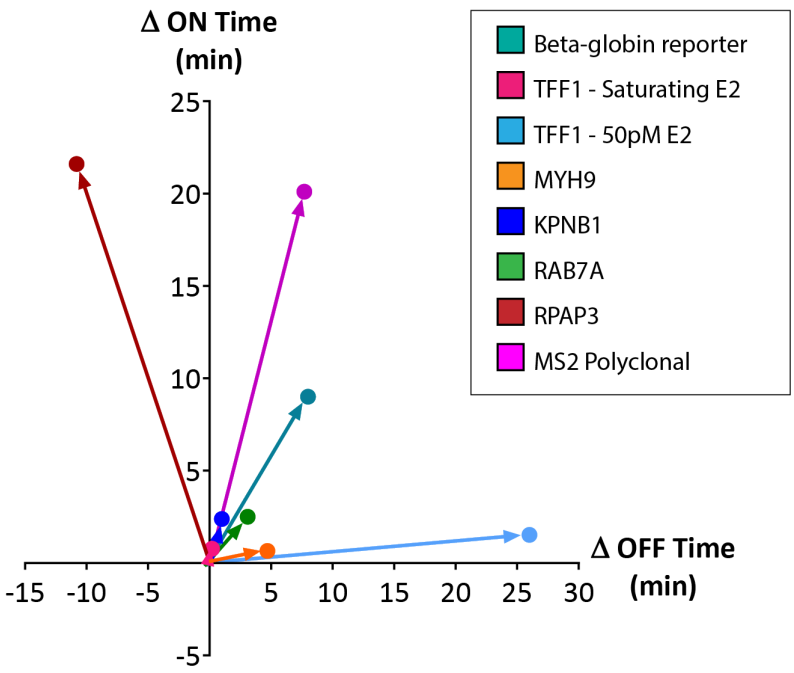


Figure 4.25: Summary delta plot showing the change in ON and OFF times for all genes studied.

(**Figure 4.26**), unlike the U2-OS Pi-MYC cells that displayed reversible translocation (**Figure 4.26**). We replicated the induction experiment of **Figure 3.4** with the HBEC cell lines and found that Pi-MYC displayed different exit and entry kinetics compared to the Pi-mCherry control and compared to Pi-MYC expressed in MCF7 cells (**Figure 4.27**). Notably, there was a decreased export rate of Pi-MYC (0.257 ± 0.037 min^{-1}) compared to Pi-mCherry (0.347 ± 0.015 min^{-1}) in HBEC cells.

Because of the non-transformed nature of the HBEC cells and the concomitant lower endogenous MYC levels compared to U2-OS and MCF7 cancer cells (**Figure 4.28**), we hypothesized that the slower export kinetics could be due to binding of MYC to available sites throughout the genome, possibly modulating transcription activity across the nucleus. Consequently, if we were indeed observing Pi-MYC acting on the genome in a global manner, this activity might be reflected in the global residence times of core transcription factors on DNA. Such a phenomenon would also be consistent with the widespread changes we observed in burst duration in the MS2

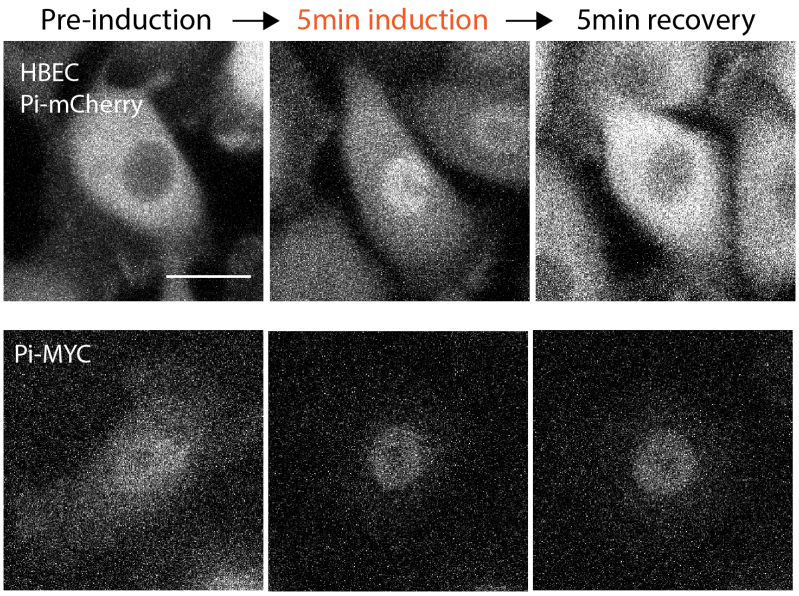
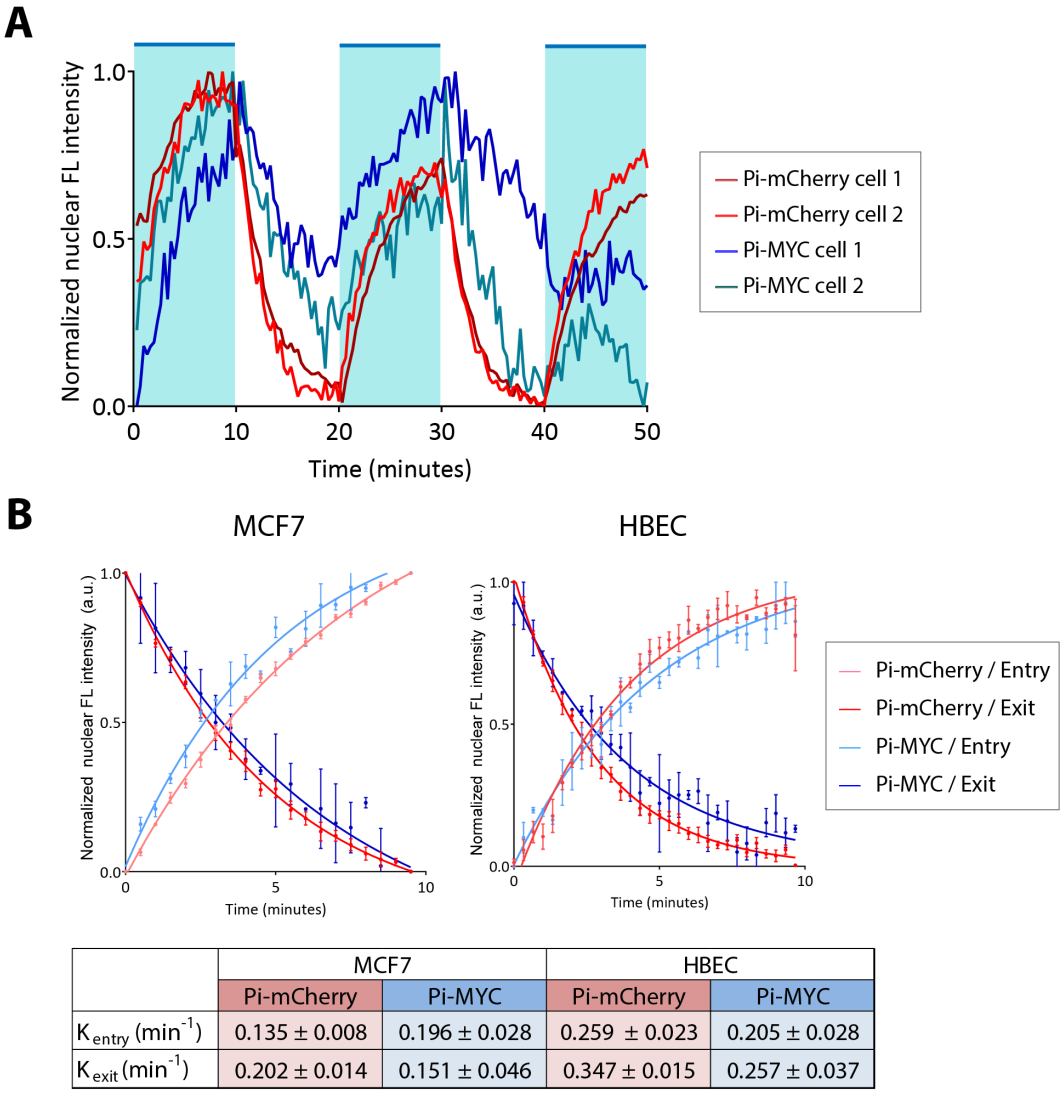


Figure 4.26: Example fluorescence images comparing nuclear entry and exit between Pi-mCherry and Pi-MYC stably expressed in HBEC cells. Scale bar = 15µm.

polyclonal population (**Figure 4.24**).

We tested this hypothesis using single-molecule tracking of transcription factors (TF) in the nucleus using HaloTag [76], an approach which has been previously used to detect residence times on DNA [17, 47, 57, 98, 126]. We chose four proteins (**Figure 4.29**): glucocorticoid receptor (GR), a gene-specific transcription factor; TATA binding protein (TBP), a core component of the pre-initiation complex (PIC) formation shown to have direct interactions with MYC [139]; mediator complex subunit 1 (MED1); and elongation factor SPT5 that is recruited by MYC to initiate RNAPII pause release and productive elongation [8].

We introduced the HaloTag constructs to the HBEC cells with electroporation and labeled the Halo-TFs with Halo ligand conjugated to JaneliaFluor-646 dye (**Figure 4.30**). We initiated quick timescale SMT measurements consisting of a 6-minute induction period of Pi-mCherry or Pi-MYC translocation with a Halo-SMT measurement in the last 2 minutes of the induction after translocation had visibly occurred (GR was stimulated with dexamethasone prior to imaging (full methods



	MCF7		HBEC	
	Pi-mCherry	Pi-MYC	Pi-mCherry	Pi-MYC
K_{entry} (min^{-1})	0.135 ± 0.008	0.196 ± 0.028	0.259 ± 0.023	0.205 ± 0.028
K_{exit} (min^{-1})	0.202 ± 0.014	0.151 ± 0.046	0.347 ± 0.015	0.257 ± 0.037

Figure 4.27: **(A)** Quantification of nuclear fluorescence intensity over a one-hour time course of HBEC Pi-mCherry or Pi-MYC cells with repeated 10-minute periods of activating light (indicated by blue rectangles). **(B)** Nuclear entry and exit rates of Pi-mCherry and Pi-MYC in the HBEC and MCF7 cell lines (3-5 cells imaged per condition). The data is fit to a single-exponential model, with the rate K and standard error reported in the table.

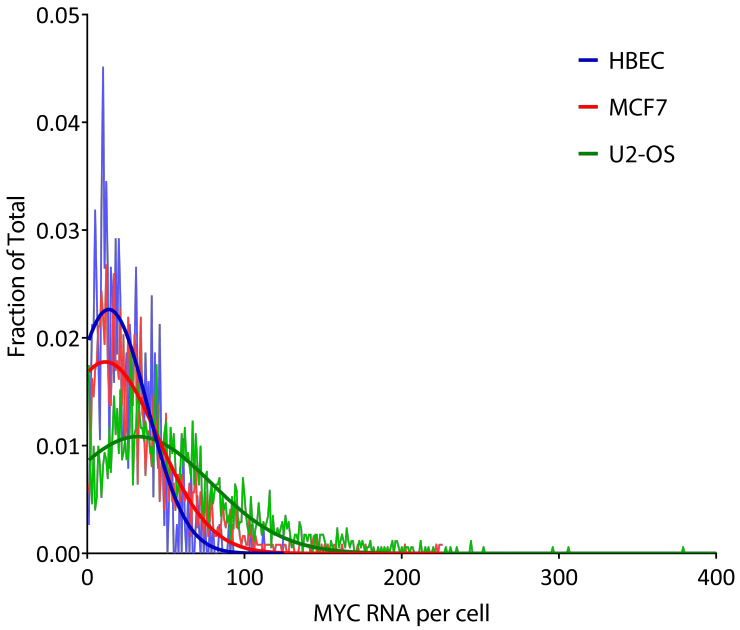


Figure 4.28: smFISH distribution of endogenous MYC RNA per cell in HBEC (n=125 cells), MCF7 (n=226 cells) and U2-OS (n=479 cells).

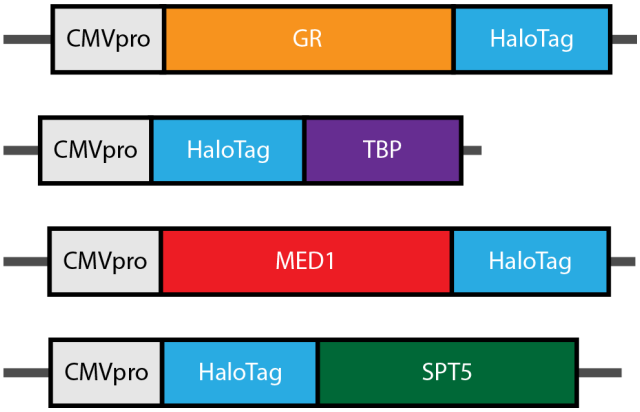


Figure 4.29: HaloTag constructs used for SMT experiments. The HaloTag is fused to the C-terminus of GR and MED1, and the N-terminus of TBP and SPT5.

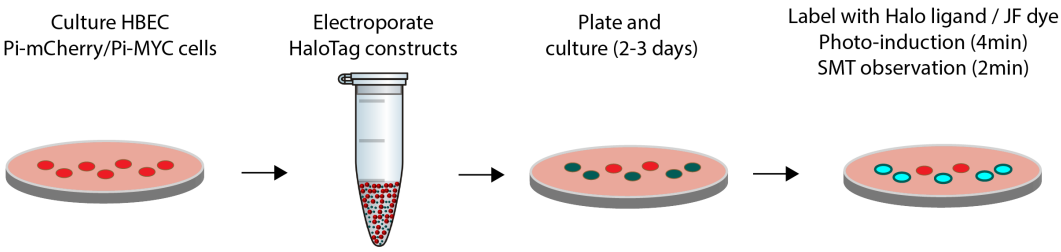


Figure 4.30: Schematic representation of SMT experiments.

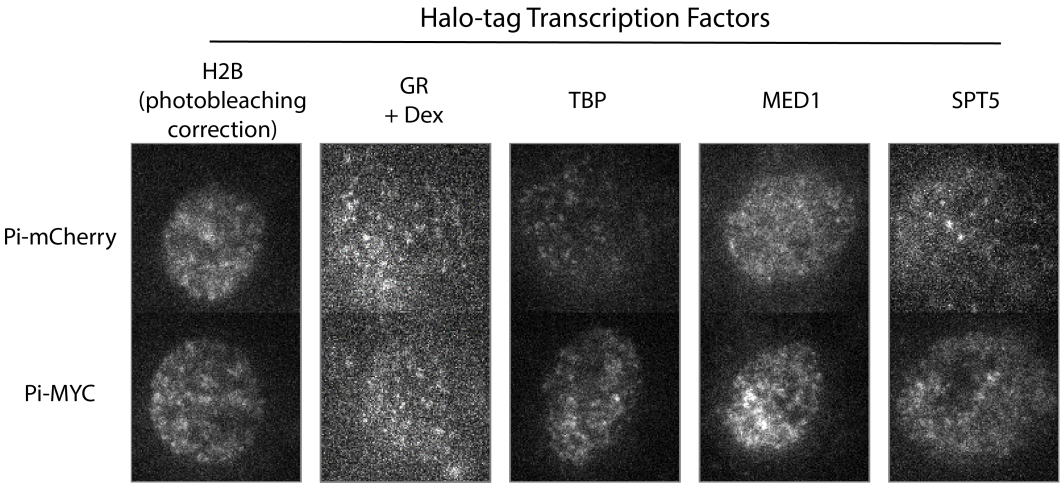


Figure 4.31: Representative images of the nuclear area of HBEC cells that stably express Pi-mCherry or Pi-MYC (fluorescence channel not shown) and transiently express the Halo-tag transcription factors measured with SMT. Individual proteins are visible as puncta if they are slow diffusing/immobile, and appear as blurry regions of the nucleus in areas where they are fast diffusing.

in section **A.6**). In this manner, single TFs could be observed exhibiting various states of mobility and were tracked to determine the distribution of residence times (**Figure 4.31**).

We found that Pi-MYC translocation changed the residence times of transcription factors related to PIC assembly and pause release. We show this data as a survival probability, which is computed from the cumulative distribution function and reflects the frequency of dwell time events in the time-series analysis (**Figure 4.32**). We observed increased residence time of TBP and SPT5 and decreased residence time of MED1 with Pi-MYC overexpression. We found that residence time of the sequence-specific factor GR was not affected.

There are multiple approaches for fitting residence time data, but we relied on two models which have been widely used. The first is a power law model [41] which assumes a continuous distribution of residence times, where the slope of the line in log-log space αindicates differences in the distributions (**Equation A.4**). The second method is to fit the data to a bi-exponential model [7,84,86,104], which assumes two populations that are conventionally understood to be a non-specifically bound population with a faster off rate k_1, and a specifically-bound fraction with a slower off rate k_2 (**Equation A.3, Figure 4.33**). Using the Baysian Information Criterion (BIC) calculated for each of the fits, we found that the power law model consistently had the lowest BIC and was the better fit in all conditions (**Figure 4.34**).

We found that the fitting parameters reflect the statistical significance of the differences in the raw data (**Figure 4.35**). Strikingly, the TBP residence time of the specifically-bound population, interpreted as the more stable component of the bi-exponential fit, nearly doubles after Pi-MYC translocation, increasing from 9.6 ± 1.1 to 18.6 ± 2.1s (the off rate k_1 decreasing from 0.104 ± 0.011 to 0.054 ± 0.009s^{-1}) with only a small change in the non-specific binding. In contrast, MED1 residence time decreases from 18.1 ± 3.3 to 6.9 ± 0.6s (k_1 increasing from 0.055 ± 0.009 to 0.146 ± 0.012s^{-1}). However, SPT5 appears to show comparable increases in the residence times of both the specific and non-specific bound fraction after Pi-MYC translocation. GR, which is involved in the early steps of transcription initiation for its target genes, was not affected by Pi-MYC overexpression. Taken together, these data suggest that MYC primarily serves to affect the mobility of transcription factors involved later in the transcription cycle, and that global changes in the dynamics of the core transcription machinery is observable seconds after Pi-MYC translocation.

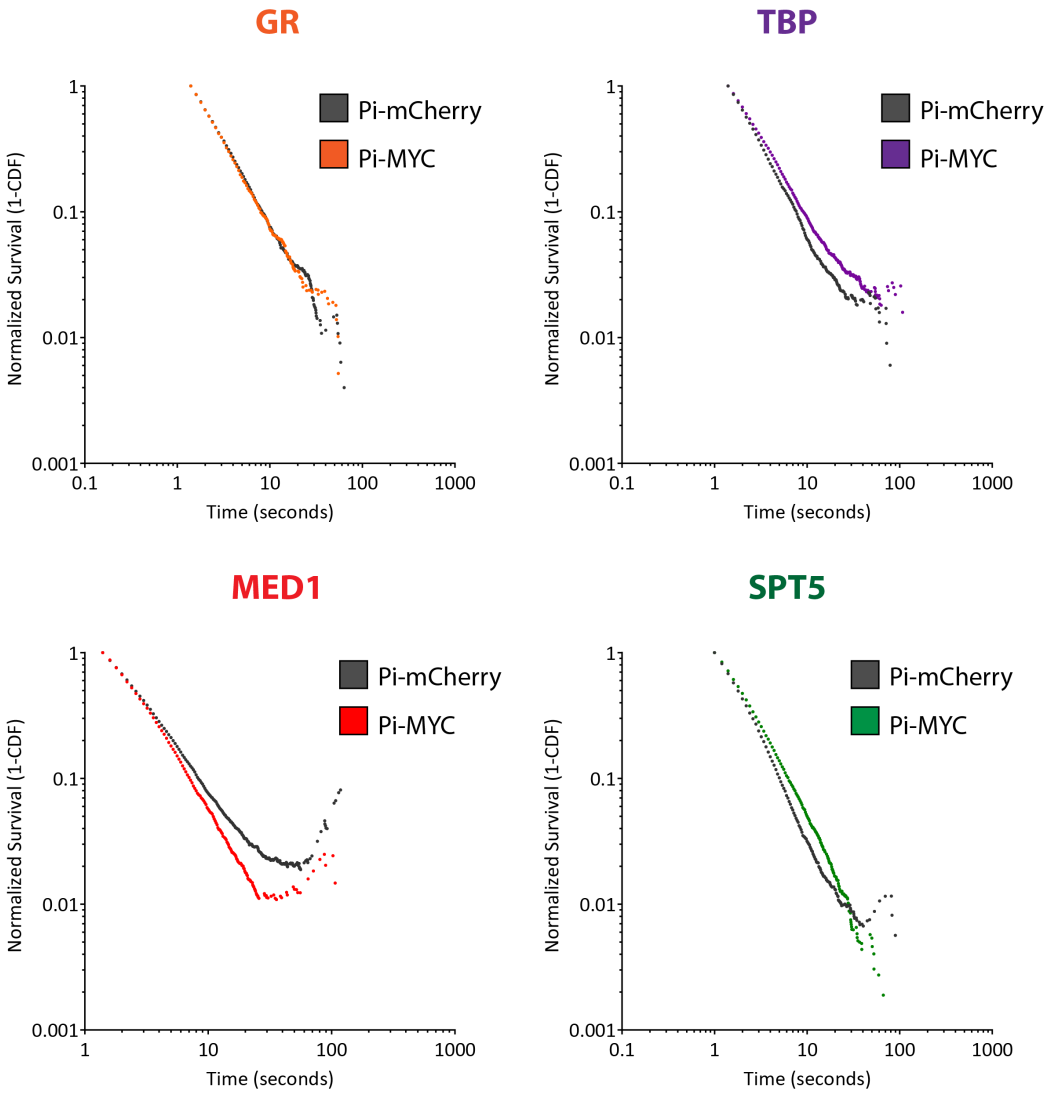


Figure 4.32: Log-log plot of transcription factor residence time distributions with Pi-mCherry overexpression (GR=19 cells, TBP=28 cells, MED1=32 cells, SPT5=20 cells) or Pi-MYC overexpression (GR=13 cells, TBP=20 cells, MED1=22 cells, SPT5=20 cells).

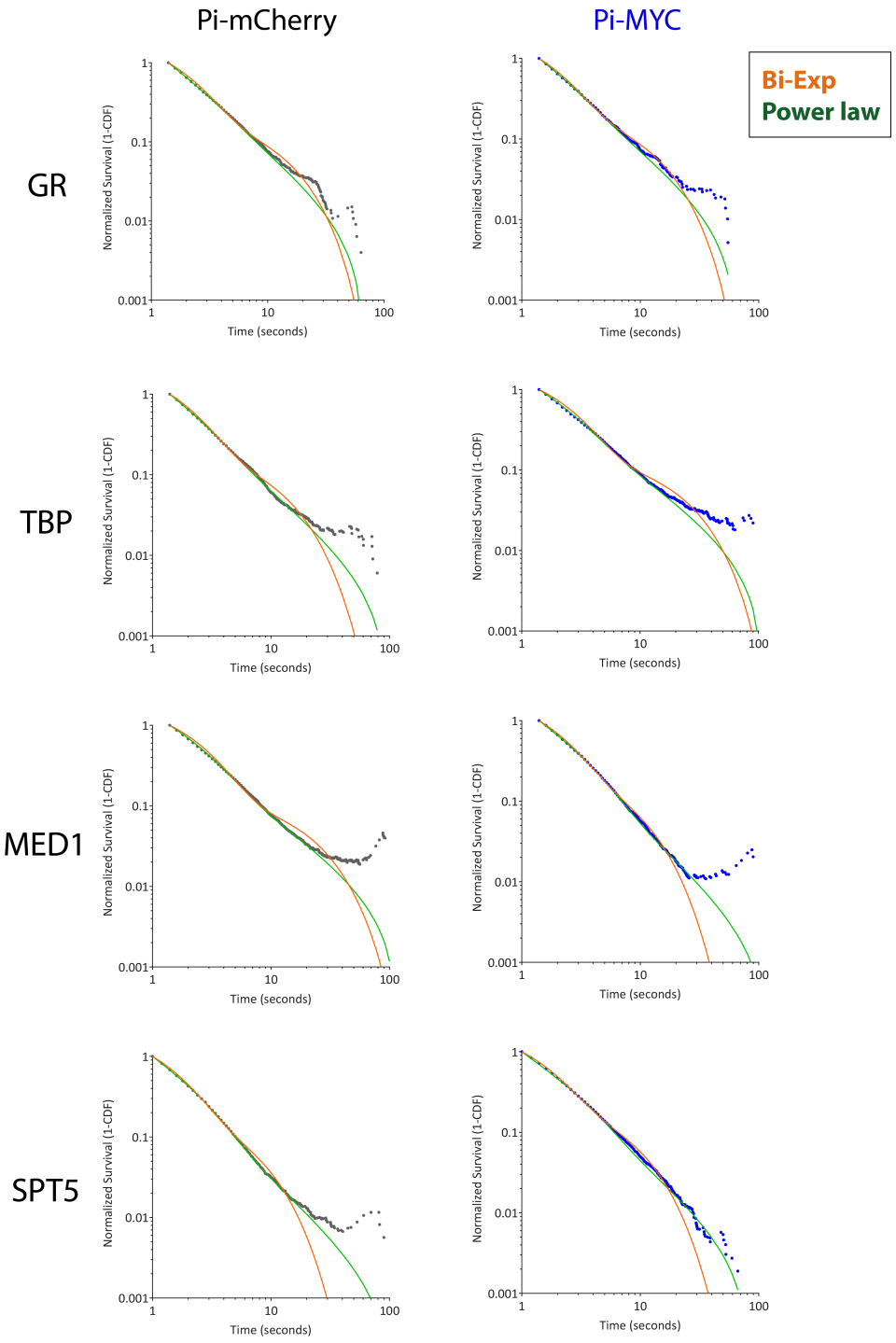


Figure 4.33: Fit of residence time distributions (gray and blue scatter plots) to the bi-exponential (orange) or power law (green) model.

	BIC VALUES	
	Bi-exponential	Power Law
GR_Pi-mCherry	1.523E+06	1.099E+05
GR_Pi-MYC	1.239E+06	4.859E+04
TBP_Pi-mCherry	1.409E+07	7.036E+05
TBP_Pi-MYC	5.200E+06	5.193E+05
MED1_Pi-mCherry	3.435E+07	6.032E+06
MED1_Pi-MYC	5.208E+07	1.897E+06
SPT5_Pi-mCherry	6.679E+07	1.133E+06
SPT5_Pi-MYC	1.017E+07	5.595E+05

Figure 4.34: BIC values for each of the conditions fitted to a bi-exponential or power law model.

		Bi-exponential Fit		Power Law Fit
		k1 (sec^{-1})	k2 (sec^{-1})	Exp. α
GR	Pi-mCherry	0.732 ± 0.031	0.090 ± 0.008	1.219± 0.018
	Pi-MYC	0.795 ± 0.034	0.101 ± 0.009	1.180 ± 0.024
TBP	Pi-mCherry	0.787 ± 0.038	0.104 ± 0.011	1.243 ± 0.035
	Pi-MYC	0.590 ± 0.023	0.054± 0.006	1.031 ± 0.027
MED1	Pi-mCherry	0.584 ± 0.031	0.055 ± 0.009	1.032 ± 0.070
	Pi-MYC	0.769 ± 0.036	0.146 ± 0.012	1.345 ± 0.030
SPT5	Pi-mCherry	1.022 ± 0.043	0.182 ± 0.016	1.341 ± 0.015
	Pi-MYC	0.941 ± 0.028	0.147 ± 0.007	1.234 ± 0.015

Figure 4.35: Table of values from fitting residence time distributions to a bi-exponential or power law model. For the bi-exponential fit, *k1* corresponds to the faster 'non-specific' bound fraction, and *k2* corresponds to the the slower 'specific' bound fraction.

Chapter 5

Conclusions

5.1 Discussion of Results

Single-cell imaging provides a window into the molecular mechanisms of gene expression. Although the MYC oncogene has been studied for many decades, an outstanding question remains as to how MYC acts directly on transcriptional bursts — the discrete events of gene expression — in single living cells. We view the question of primary effects as a question of *timing:* that is, when MYC levels are perturbed in the nucleus of a living cell, what happens to gene expression events *immediately* afterwards?

To answer this question, we undertook a systematic survey of MYC using single-cell imaging approaches to quantify RNA production, gene active and inactive periods, and transcription factor residence times. Our study was based on four essential components:

1. Live-cell imaging to obtain real-time readout of gene expression activity.
2. Optogenetic engineering to acutely control and visualize MYC overexpression in the nucleus.
3. Surveying a broad panel of genes that represented a range of cellular function,

basal expression level, DNA features, and cell type.

4. Measurements made with a large range of timescales (minutes to hours) to assess the most immediate effects of MYC perturbation.

This approach allowed us to build an empirical model that relates the effects of MYC nuclear localization to discrete quantities and rates of transcription factor binding, gene state, and RNA production (**Figure 5.1**).

We find that MYC increases the active period of transcription and exerts global changes in the dynamics of the basal transcription machinery. In breast carcinoma cells (MCF7), MYC extended the transcriptional bursts of the estrogen responsive gene *TFF1* but only when estradiol levels were non-saturating. In immortalized lung epithelial cells (HBEC), we again observed that MYC increased gene active periods for a panel of genes, and that the significance of the increase corresponded to the basal expression level of the gene: the higher the gene was expressed, the less the active period changed. Finally, by using Pi-MYC, a photo-inducible MYC in an MS2 polyclonal cell line, we observed that the increase in gene active period was a global effect. This global increase in the transcription burst duration was accompanied by an increase in the dwell time for TBP and SPT5 but a decrease in dwell time for MED1. We thus achieved the first real-time readout of how MYC affects the discrete events of gene expression in living human cells.

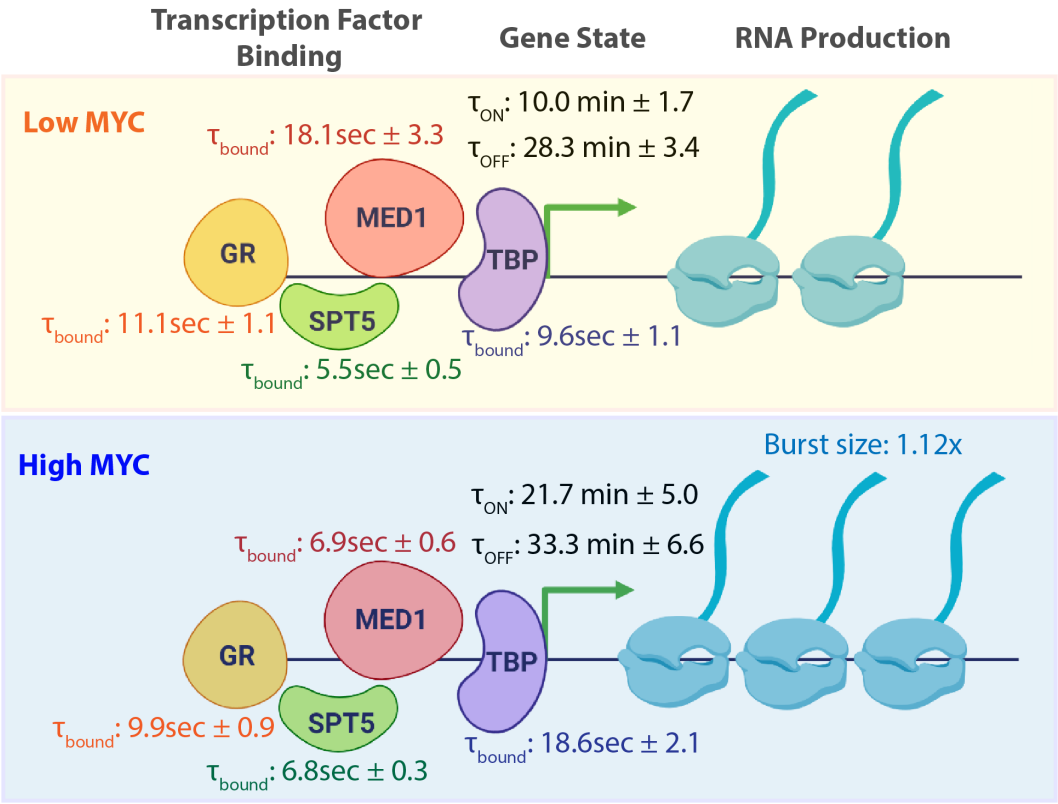


Figure 5.1: Model of MYC mechanism of action. Rates and characteristic timescales of transcription factor binding, gene ON/OFF state, and RNA production are from measurements taken in the HBEC MS2 polyclonal cell line, at low and high MYC levels.

5.1.1 Single cell perturbation and readout with Pi-MYC

We sought a form of MYC overexpression that was compatible with imaging MS2-GFP labeled RNA, could be controlled with high spatio-temporal position, and was potentially reversible on fast timescales. We met these design criteria with a photo-inducible variant of MYC possessing a LOV2-caged NLS that we engineered for acute spatial-temporal control of MYC nuclear localization. The caging mechanism of LOV2 has previously been used to control proteins such as cyclins, kinases, and motor proteins [90,93,137,142]. Pi-MYC exhibited nuclear translocation within minutes of light induction and further exhibited hallmark phenotypes of an oncogene in growth and transformation assays. While TetON MYC-EGFP enabled stable MYC overexpression and is visible under the microscope, the onset of overexpression with this method takes several hours and occurs at different rates in different cells. An

earlier form of MYC overexpression is c-MycER, where MYC is fused to the ligand binding domain of the mouse estrogen receptor and is induced to translocate to the nucleus with the addition of 4-hydroxytamoxifen [29, 114]. The advantage of this method is that the MYC transgene is constitutively expressed and already present in the cytosol. Induction stimulates MYC nuclear translocation, and it occurs within a matter of minutes. However, c-MycER likewise relies on a chemical inducer, and reversing MYC nuclear overexpression relies on replacement of media to washout the chemical inducer. In contrast, Pi-MYC immediately begins to exit the nucleus after turning off blue light. To our knowledge, our study is the first report of using LOV2 to spatially control a human oncogenic transcription factor.

We note several caveats when working with a phototrophic domain. First, it can be sensitive to trace amounts of light found in a normal laboratory setting. Room light itself may cause the LOV2 domain to react and uncage the NLS, and cause trace amounts of nuclear translocation during passaging, culture, and transport of the cell lines. We made every effort to reduce leaky activation by covering cell culture plates, working with the cell lines in minimal light, and keeping the samples in the dark for 1 hour prior to beginning an imaging acquisition. Second, under certain conditions that we document here, blue light can change growth properties in tissue culture. Although we were able to control for these effects in our biological studies, future technical efforts will be directed toward developing tighter cages, possibly through simultaneous caging of the NLS and re-directing to other cellular compartments as demonstrated previously [137].

Surprisingly, we observed slower export and sustained nuclear retention of Pi-MYC in human bronchial epithelial cells. This phenomenon was not seen for Pi-MYC in the U2-OS and MCF7 cell lines, and mCherry export kinetics were the same across cell lines. Given that HBEC cells express MYC at lower levels than the cancer-derived cell lines, the nuclear retention of Pi-MYC in HBEC cells may be due to greater avidity of MYC for chromatin: there may be more unoccupied MYC bind-

ing sites. Thus, even though the LOV2 domain cages the NLS when the activating light is removed, Pi-MYC may be hindered from exiting the nucleus due to multiple interactions with DNA and nuclear proteins at sites of active transcription. This observation is reminiscent of a previous in vivo single-molecule tracking study in which MYC exhibited non-compact 'global' exploration of nuclear space [53]. Regardless of the underlying mechanism of Pi-MYC retention, these observations motivated us to further probe MYC activity through two assays which average the behavior of many individual genes: 1) meta-gene bursting analysis of a polyclonal population, and 2) examination of transcription factor residence times across the nucleus.

5.1.2 MYC regulates transcriptional bursts and core transcription machinery dynamics

Our findings from live-cell imaging across genes and cell lines show that MYC increases the active period of a gene, making it unique among transcription factors and providing clues to the mechanism of activity. Several experimental and computational studies now show that gene regulation occurs predominantly through burst frequency modulation [9, 15, 38, 40, 46, 112]. Thus, MYC plays an orthogonal role, modulating burst duration rather than burst frequency. The robust increase in the duration of transcriptional bursts suggested to us that MYC is acting on transcription at the early steps of initiation and may be exerting effects through multiple interactions with core transcription machinery. Our survey of the effects of acute MYC translocation on transcription factor residence times with SMT supports this hypothesis, as we saw changes in the basal machinery such as TBP, MED1, and SPT5, but not the sequence specific activator GR. MYC overexpression caused increased residence times for TBP and decreased residence times for MED1, suggesting stabilization of the PIC and recruitment of RNAPII while facilitating dissociation of the Mediator complex to allow for promoter clearance. Increased residence time of SPT5 may coincide with a greater proportion of productive elongation events. There is evidence for physical interaction between MYC and each

of these factors [8, 73, 85], but our assay reports on functional outcomes and does not distinguish between direct and indirect mechanisms. Based on this evidence, we predict that MYC increases in gene expression through factors that affect the ON time (i.e. core transcription machinery) rather than factors that affect the OFF time (i.e. pioneer transcription factors, chromatin remodelers)

An outstanding question to emerge from these studies is: How do changes in dwell time on the order of seconds lead to changes in burst duration on the order of minutes? The average dwell time of GR (11.1 seconds) is similar to previous studies, but the measured TBP dwell time (9.6 seconds) is different [104, 128]. Yet despite this difference, the bursting timescale is still much longer. One possibility which is hard to exclude is that SMT misses a population of extremely long-lived events. Our SMT acquisition parameters introduce a bias in the sampling the full distribution of residence times; our short measurement window of two minutes causes an underestimation of the dwell times of the highly stable population of transcription factors. A second technical limitation is that the position of the stem loops in the target gene can determine the resolution of bursting analysis [111]. One biological possibility is that cooperative effects on many transcription factors leads to assembly of a more stable permissive environment for sustained transcription. In some cases, such as the GAL locus in budding yeast, there is a clear relationship between activator dwell time and burst duration [26]. However, a general model in metazoans has not emerged.

In conclusion, we envision MYC as the mortar holding the bricks of active transcription assembly together to more easily facilitate product transcription while the gene is permissive for transcription. Although burst frequency is the dominant mode of gene regulation, MYC exerts a minimal role in this function. One consequence of MYC increasing the active period of a gene with little to no change in the long inactive periods is an increase in gene expression heterogeneity. We speculate that the resulting increase in gene expression heterogeneity could drive cells to populate

pathological states and phenotypes that may ultimately make them susceptible to cancer.

5.2 Future work

5.2.1 Pi-MYC

Applying the Pi-system to study other transcription factors. Photo-inducible nuclear translocation with the LOV2 domain offers unprecedented spatio-temporal control of a transcription factor within living cells. This system could readily be applied to other transcription factors to observe their direct and immediate effects on downstream gene expression after nuclear import. We expect that each transcription factor would be need to be evaluated and possibly modified on a case-by-case basis. MYC has a strong NLS in exon 2, which we had to mutate in order for it to be under control of the LOV2-caged NLS. Such modifications may be necessary for other factors that contain nuclear import and export signals within their coding region.

Determining the MYC functional domains responsible for gene amplification. We have achieved an observable readout of how MYC effects gene expression on short timescales, namely that it increases gene ON time. A follow-up study would be to probe the domains in the MYC protein responsible for this function. Future experiments would be to create Pi-MYC variants that are mutated in the various functional domains, such as the DNA binding domain, the transactivation domain, and the conserved sequences in Myc boxes I-IV. We would expect that the domains responsible for gene expression amplification, when mutated, would result in a loss of gene ON time increases as observed by live-cell imaging, and loss of mobility changes in core transcription factors as observed by SMT.

5.2.2 MS2-MYC

There is currently no MS2 live-cell imaging system to visualize the transcriptional bursting of the endogenous *c-myc* gene directly. A current follow-up study is to generate an 'MS2-MYC' live-cell imaging system in HBEC and U2-OS cells so that we may directly observe the RNA synthesis of *c-myc* itself. Some possible questions we seek to ask with this system are:

1. How does *c-myc* expression correlate with changes in nuclear architecture, specifically with the compaction of its topologically associated domain and association with the nuclear periphery?

2. How do transcriptional burst profiles of *c-myc* differ in a non-transformed cell line (HBEC) vs. cancer derived cell line (U2-OS)?

3. Can we perturb the *c-myc* locus to induce the alterations normally seen in cancer, such as the (8;14) chromosomal translocation to put MYC expression under control of the Ig locus, or force a gene duplication, to see how the transcriptional burst profile of MYC changes as a result?

To answer these questions, we are currently engineering the MS2 system to be targeted to the 3' end of the *c-myc* coding region using CRISPR. The integration is delivered by a donor plasmid described in **Figure 5.2**, and consists of the following domains:

- **miRFP670nano [97]**: A far-red fluorescent protein to serve as a proxy for MYC protein expression levels; can be used as a selection marker in fluorescence-activated cell sorting.

- **Blasticidin**: Antibiotic resistance marker.

- **24x MS2 loops**: Visualization of RNA synthesis.

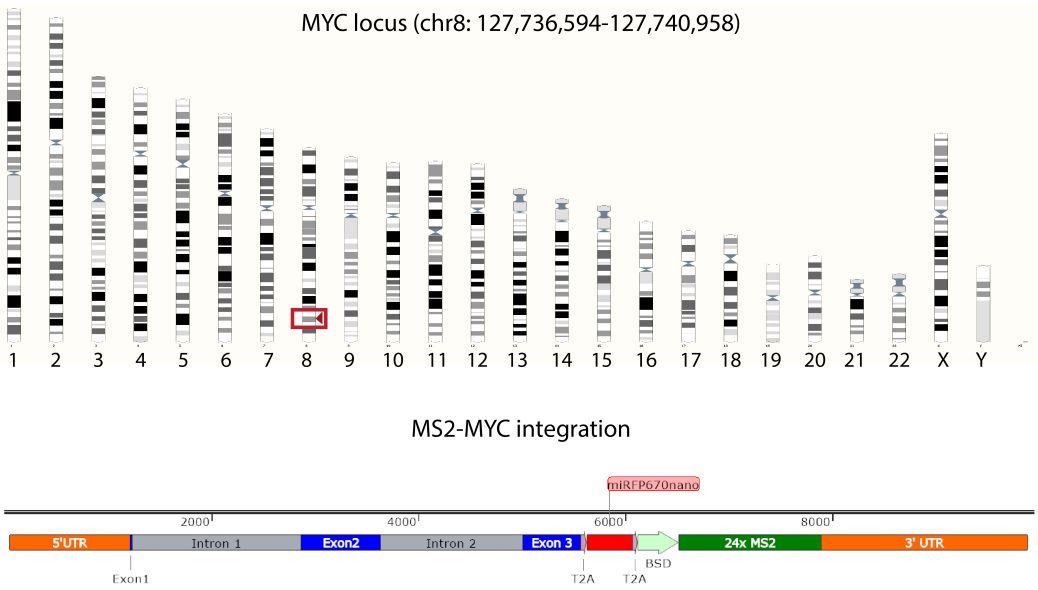


Figure 5.2: Top: MYC locus on chromosome 8 indicated by the red rectangle. **Bottom:** MS2-MYC integration design. The predicted integration contains the fluorescent marker *miRFP670nano* and antibiotic selection marker *blasticidin* (BSD) after the MYC exon 3 coding seqence. The markers are separated by T2A sequences in order to be translated as separate peptides from MYC, so as not to interfere with the function of the MYC protein. A stop codon follows the antibiotic marker, after which follows a 24x MS2 stem loop sequence at the begnning of the 3'UTR. The total size of the integration sequence is 2.3kb.

5.2.3 Halo-MYC

We seek to track MYC protein mobility with single-molecule imaging in living cells. To achieve this, we engineered a 'Halo-MYC' plasmid that can be delivered into cells by transient transfection or electroporation. Halo-MYC can be used to probe our 'mortar and brick' hypothesis of how MYC binds both protein and DNA at active sites of transcription to retain the open state of the gene and the nuclear factors involved. We can use Halo-MYC with the super-resolution imaging technique PALM (photoactivated localization microscopy) to determine if multiple MYC proteins assemble at a DNA locus and whether they appear in direct complex with core transcription factors such as TBP, MED1, and SPT5 that have their mobility affected by MYC overexpression.

In addition to PALM, we can perform SMT to observe Halo-MYC mobility in real time, and use MYC mutants in various functional domains to assess what contributions DNA binding or protein-protein interactions make to the MYC dwell time distribution. We succesfully cloned Halo- wildtype MYC and a Halo-MYC NLS deletion mutant and conducted preliminary observations with SMT on its mobility in HBEC cells (**Figure 5.3**). Interestingly, preliminary movies show aggregates of molecules with long-lived residence times in the wildtype MYC condition. As expected, the NLS deletion mutant appears not to enter the nucleus, and no puncta are observed. Future study will be needed to follow up on these initial observations. While much can be learned by delivering Halo-MYC plasmid transiently, the goal in coming years would be ultimately to integrate the Halo-tag into the endogenous locus of *c-myc* with CRISPR, similar to MS2-MYC, so that we can measure the mobility of endogenous MYC protein with single-molecule imaging.

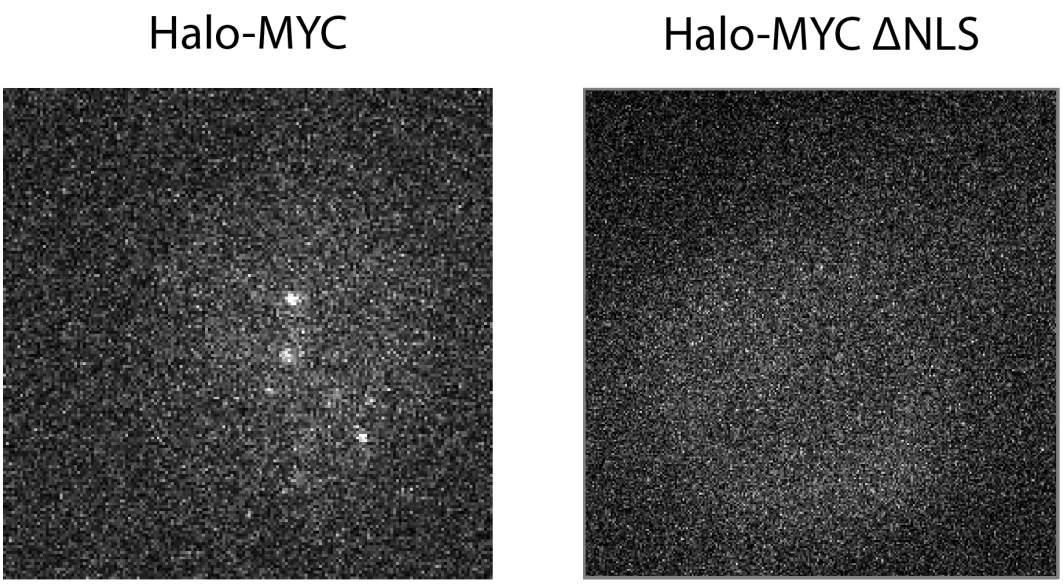


Figure 5.3: Halo-MYC expressed in HBEC cells. Shown are the nuclear area of cells with Halo-tagged wildtype MYC (left) and an NLS deletion mutant (right). Puncta of slow diffusing Halo-MYC protein is visible in the wildtype condition but not in the case where the NLS is deleted and MYC is unable to be imported into the nucleus.

Appendices

Appendix A

Additional Methods

A.1 Mammalian cell culture conditions

A.1.1 U2-OS human osteosarcoma cell line

U2-OS cells were cultured in DMEM with low glucose, GlutaMAX, and pyruvate (Gibco 10567) supplemented with 1% Penicillin-Streptomycin (ThermoFisher) and 10% FBS (Sigma).

A.1.2 MCF7 human breast cancer cell line

MCF7 TFF1-MS2 GFP cells were cultured in MEM (Corning 15-010) supplemented with 1% Pen-Strep, 10% FBS, and 2mM L-glutamine (HyClone SH30034). Pi-MYC and Pi-mCherry stable lines were cultured the same media.
MCF7 TetON MYC-EGFP cells were cultured in MEM supplemented with 1% Pen-Strep, 2mM L-glutamine, 10% Tetracycline-free FBS (Clontech) and 0.5 micrograms/mL Puromycin to maintain selection.

Saturating E2 imaging media is phenol-free MEM (Corning 17-305) supplemented with 1x Pen-Strep, 2mM L-Glutamine, and 10% FBS. Hormone depletion media is phenol-free MEM supplemented with 1% Pen-Strep, 2mM L-Glutamine, and char-

coal/dextran treated FBS (HyClone). E2 (β-estradiol, Sigma) is added to this media to achieve the desired concentration. This E2/hormone depletion media is also used as imaging media.

A.1.3 HBEC human bronchial epithelial cell line

HBEC cells were cultured and imaged in Keratinocyte Serum-Free Media supplemented with bovine pituitary extract and EGF (Gibco 17005) and 1% Pen-Strep.

A.1.4 NIH3T3 mouse fibroblasts

NIH3T3 fibroblasts (ATCC CRL-1658) were cultured in DMEM (Gibco 10564) supplemented with 1% pen-strep and 10% FBS.

A.1.5 Harvest cells for virus generation

HEK293T and Phoenix-AMPHO cells were cultured in DMEM with high glucose, GlutaMAX and pyruvate (Gibco 10569), supplemented with 1% pen-strep and 10% FBS.

A.1.6 Treatment vehicles

E2 (β-estradiol) was dissolved in 100% ethanol at a stock concentration of 1mM and stored in aliquots at -20C. Doxycycline was dissolved in water and at a stock concentration of 9µg/mL and stored in aliquots at -20C. HBEC and MCF7 TetON MYC-EGFP stable lines were induced with 1µg/mL doxycycline to induce MYC overexpression.

A.2 Transient MYC overexpression and knockdown in MCF7 cells at saturating and 50pM estradiol

MCF7 TFF1-MS2 cells were plated in a 2-well #1.5 glass bottom chamber (Nunc Lab-Tek 155379) in culture media and allowed to recover. The saturating E2 condition had a media change on Day 2 and changed into Saturating E2 imaging media

along transfections of 1.5µg MYC-mCherry plasmid per well or 40nM MYC siRNA (ThermoFisher) on Day 5. Cells were imaged at Day 6 (24 hours MYC overexpression) and Day 7 (48 hours MYC knockdown). For the 50pM E2 condition, cells were hormone depleted with 2 washes and a replacement into hormone depletion media on Day 2. On Day 5, cells were changed into hormone depletion media containing 50pM E2. On Day 7 the 50pM E2/hormone depletion media was replaced and cells were transfected with MYC-mCherry plasmid or MYC siRNA and imaged after 24 hours (Day 8) or 48 hours (Day 9), respectively.

A.3 TetON MYC-EGFP Stable Expression in HBEC cells

Frozen stock of purified pTRIPZ TetON MYC-EGFP virus was obtained from the Levens lab and transduced with the Lenti-X system (Takara Bio USA, Inc.) as per their protocol [91]. HBEC cells were plated in T-75 flasks in antibiotic free media and grown to 80% confluency. A viral transduction mixture was prepared with 100µL virus, 40µL Lenti-X beads, and 260µL 1x PBS. The viral mixture was incubated for 30min at RT, tapping briefly every 5min to mix. The mixture was then added dropwise to the flask of cells and placed on a Lenti-X magnetic plate for 5min and a media exchange afterwards. The flask was then incubated on the magnetic plate overnight at 37C. After 1 week, positive cells were enriched by induction 5µg/mL doxycycline for 24 hours to stimulate MYC-EGFP expression, and then FACS sorted for fluorescence. Two rounds of induction and sorting were required to fully enrich for the transduced cells. Although the pTRIPZ construct contains a puromycin antibiotic resistance marker, we had to rely on fluorescence expression of induced MYC-EGFP HBEC cells already contain puromycin resistance as part of their immortalization process.

A.4 Biochemistry

A.4.1 qRT-PCR

qRT-PCR was performed in HBEC cells to measure effects of MYC overexpression and knockdown for 11 genes identified in [136]. RNA was purified from cell pellets using a Quick-RNA Microprep kit (Zymo). cDNA was generated using Protoscript II (NEB) with Random Hexamers (IDT). qPCR was performed with iQ SYBR Green Supermix (BioRad) using primers for 10 genes *CANX, DNAJC5, ERRFI1, KPNB1, MYH9, PFN1, RAB7A, RHOA, RPAP3, SEC16A, SLC2A1*, and primers to beta-actin as a control.

A.4.2 Western blotting

Cells were resuspended in 20µL RIPA buffer per 350,000 cells (50mM TrisHCl pH 7.5, 150mM NaCl, 0.05% Tween-20, 1% sodium deoxycholate, 1 tablet Protease Inhibitor Cocktail in H20). Lysates were incubated in RIPA on ice for 30min, with a brief vortex every 10min. Lysates were spun at 14,000rpm at 4C for 10min and supernatant was collected into new tubes. 4x LDS Sample Buffer (Invitrogen) was added to lysates and boiled for 10min at 70C. Samples were loaded onto 10-lane or 15-lane NuPAGE 4-12% Bis-Tris gel (Invitrogen) and run with Spectra BR protein ladder (ThermoFisher Scientific). Samples were transferred to a PDVF membrane using a Trans-Blot Turbo Transfer System (Bio-Rad). Transfer of protein bands were verified with Ponceau stain, and membrane was then blocked in 5% milk /TBST (19.98mM Tris, 136mM NaCl, pH 7.4) for 1 hour at RT. Membrane was incubated with primary antibodies at 4C overnight at 1:1000 in milk block. Primary antibodies used were Rabbit monoclonal anti-Myc Y69 (Abcam ab32072), Rabbit polyclonal anti-mCherry (Abcam ab167453), Rabbit polyclonal anti-GFP (Abcam ab290), Mouse monoclonal Pan-Ras (Sigma-Aldrich OP40), and Mouse monoclonal anti-GAPDH (Ambion AM4300). The next day, membranes were washed 3X in TBST and incubated with secondary antibodies 1:10,000 in milk block using anti-rabbit HRP (GE Healthcare Life Sciences NA934) or anti-mouse

HRP (GE Healthcare Life Sciences NA931V) for 1 hour at RT. The membranes were then washed 3X with TBST and developed with ECL reagent (SuperSignal West Dura Extended Duration Substrate, ThermoFisher Scientific).

A.5 Pi-MYC

A.5.1 Plasmid design and construction

The mCherry-NES-LOV2-NLS sequence from pTRIEX-mCherry-LOV2 plasmid (gift from Klaus Hahn laboratory, UNC Chapel Hill, Addgene plasmid 81041) was PCR amplified with NotI-5' and AgeI-3' ends. The fragment was cloned into a pd4-MYC-EGFP plasmid where the GFP had been removed via restriction digest with AgeI-NotI. The resulting plasmid (pd4-MYC-LOV2) was digested with SacII-AgeI to remove the c-myc wildtype NLS and was replaced with a GeneBlock (IDT) in which the lysine and arginine residues of the c-MYC NLS sequence were replaced with alanine. The entire sequence was then PCR amplified with NotI-5' and NdeI-3' ends, with the 5' primer starting at either MYC exon 2 (for Pi-MYC) or mCherry (for Pi-mCherry), and the 3' primer at the end of the NLS that follows the LOV2 sequence. The PCR fragments were ligated into lentiviral vectors containing ampicillin and bleomycin resistance cassettes, resulting in the plenti-Pi-MYC and pLenti-Pi-mCherry plasmids. All plasmids were purified with Qiagen maxiprep kit and verified with Sanger sequencing.

Additional control plasmids were made without photo-inducible capability. pLenti-wtMYC-mCherry (with the wildtype NLS sequence in c-myc exon 3) was cloned by PCR amplifying and inserting the MYC-mCherry sequence from pd4-MYC-LOV2 into a lentiviral vector. The resulting plasmid was digested to remove the MYC sequence and re-ligated with a short linker to create the pLenti-mCherry plasmid.

A.5.2 Lentivirus generation

HEK 293T cells were plated at 162cm flasks and grown to 80% confluency. The cells were exchanged into media with no antibiotics for 4.5 hours prior to transfection. Cells were transfected with Fugene HD (Promega) containing a mixture of 1.5µg each of Tat, Rev, Gag/Pol, and VsvG plasmids, and 30µg of the pLenti-Pi-Myc or pLenti-Pi-mCherry plasmid. Cells were incubated overnight at 37C and moved to a 32C incubator the next day to facilitate more efficient virus production. 25mL supernatant was harvested every 24 hours for two days and purified with the Virabind Lentivirus Purification Kit (Cell BioLabs) in 100µL aliquots of glycerol and stored at -80C.

A.5.3 NIH3T3 and HBEC stable lines

NIH3T3 cells with pBABE-HRasV12 retrovirus and Pi-mCherry or Pi-MYC lentivirus to create the following cell lines for stable expression: 1) wtRAS/Pi-mCherry, 2) wtRAS/Pi-MYC, 3) RASV12/Pi-mCherry, 4) RASV12/Pi-MYC. We generated the RASV12 lines first; briefly, Phoenix-AMPHO harvest cells were plated at a density of 5 million cells in 10cm plates. Cells were exchanged into media without serum or antibiotics and transfected with 10µg DNA per plate. Plates were moved to a 32C incubator to improve virus stability. The next day, the target cells (NIH3T3) were plated at 350,000 cells/well in a 6-well plate to prepare for transduction. Phoenix-AMPHO cell media containing viral supernatant was collected at 48 hours post transfection and applied directly to target cells at a ratio of 1:1 with NIH3T3 media, and polybrene at a concentration of 5.4µg/mL. The target cell plate was spun at 2900rpm for 2 hours at RT to facilitate efficient viral transduction and incubated at 32C overnight. The next day, cells were expanded into a T-75 flasks and returned to incubation at 37C. The following day, G418 antibiotic selection was applied for 7 days. After NIH3T3 wtRAS / RASV12 backgrounds were established, the cells were plated in 6-well plates at 200,000 cells per well in 2mL media. The next day, cells were transduced with virus solution containing 100µL glycerol aliquot of Pi-Myc or

Pi-mCherry virus and 10µg polybrene in 2mL media per well. The plates were spun at 2900rpm at 32C for 2 hours to facilitate efficient transduction, and then placed into a 32C incubator overnight. Media was replaced the next day and the plates were moved to a 37C incubator. After reaching confluency, cells were transferred to T-75 flasks and subjected to antibiotic selection for 7 days with zeocin (Invivogen). The same lentivirus transduction method was used for generating stable expression of Pi-mCherry and Pi-MYC in HBEC cells.

A.5.4 NIH3T3 MYC/RAS growth assay

NIH3T3 stable lines and the parental line were plated at 20,000 cells per well for two timepoints with biological replicates for. Cells were trypsinized and resuspended in ¡2mL media, and the total count per well was calculated with an automated cell counter at Days 2 and 5.

A.5.5 Colony formation in soft agar

NIH3T3 cells were seeded in soft agar (Sigma A1296) at a concentration of 40,000 cells/well in 6-well plates, with three biological replicates per condition, and NIH3T3 culture media containing 20% FBS. Each well consisted of a bottom layer of 1.5mL 1% agar/media, a middle layer containing cells suspended in 1mL 0.35% agar/media, and a top feeder layer of 1mL media. Media was added to the feeder layer twice a week for two weeks. After two weeks, cells were stained with crystal violet and imaged with Epson scanner (RBG color, 1200dpi) and scored for colonies. For the light induction condition, cells were incubated in pulses of blue light (10s on/10s off) with a 455nm collimated LED (Thorlabs M455L4-C1) for 24 hours prior seeding in agar.

A.5.6 Photoinduction conditions

Nuclear translocation of Pi-mCherry and Pi-MYC was achieved with either a 488nm laser to illuminate a field of view in the microscope or a 455nm collimated blue LED to illuminate a 35mm area of a culture dish. Translocation of Pi-mCherry

and Pi-MYC (**Figures 3.3, 4.26, 4.27**) were induced using a 488nm laser at 10% power with a 150ms pulse at 10 second intervals. For the HBEC growth assay in **Figure 3.12**, cells were seeded at 20,000 cells per well in two 12-well plates separated by a barrier in a 37C 5% CO2 incubator. One plate was cultured under constant 455nm light at 50mW for two days (+λ condition). For the HBEC live-cell imaging experiments, cells were induced with a 488nm laser at 30% power for approximately 15 seconds at 100 second intervals.

A.6 SMT

A.6.1 HaloTag labeling protocol

HBEC cells (MS2-polyclonal background) expressing Pi-mCherry or Pi-MYC were electroporated with HaloTag plasmids at a ratio of 50,000 cells:500ng DNA:10µL total volume with the Neon Transfection System (Invitrogen). The following electroporation pulse settings were used: Voltage = 1400V, Width = 20ms, Pulses = 2. After electroporation, cells were plated into 8-well #1.5 glass bottom chamber at a density of 50,000 cells/well in HBEC media and were cultured for 2-3 days to allow for recovery and expression of HaloTag plasmid.

On the day of imaging, cells were labeled with Halo ligand conjugated to JaneliaFluor-646 (JF646) dye [44]. Cells were labeled and incubated for 15min then rinsed with HBEC media with a 3x wash, 10-minute incubation, and a final 3x wash. For the Halo-GR condition, cells were induced with 100nM dexamethasone to stimulate GR nuclear translocation immediately prior to imaging.

A.6.2 Microscope setup

For SMT, we used a custom wide-field microscope capable of simultaneous imaging in three channels that is based on designs previously described [31,89]. Briefly, three solid-state excitation lasers at wavelengths 488, 561, and 647 nm (OBIS, Coherent Inc., CA, USA) are combined, expanded to provide more even illumination at the

sample, and focused at the back focal plane of a 100x, 1.49 N.A., oil immersion objective (Olympus Scientific Solutions, MA, USA). To reduce background, HILO illumination [131] is achieved by moving the radial position of the beam in the objective back aperture with a movable mirror, and the thickness of the excitation is adjusted by use of a manual diaphragm. Fluorescence emission is separated from scattered laser light by use of a quad-band dichroic (ZT405/488/561/647, Chroma Technology Corp., VT, USA), and the emission bands are separated by two long-pass filters (T588lpxr, T660lpxr), and emission filters (525/50, 609/58 and 736/128, Semrock, Inc., NY, USA) before being focused on separate EMCCD cameras (Evolve 512 Delta, Teledyne Photometrics, Tucson, AZ), with 200mm tube lenses. The combination of objective lens and tube lens results in a total magnification of 111x, corresponding to a XY pixel size of 144 nm. The sample is held on a motorized XY translation stage with piezo Z (PZ-2000 XYZ, ASI Imaging, OR, USA), which is mounted on a Rapid Automated Modular Microscope System (ASI Imaging). All lasers and cameras are synchronized using a microcontroller board (Arduino UNO), and images are collected using the open source microscope control software, Micro-Manager [28].

A.6.3 Imaging

Time-lapse images were collected at intervals of 200ms, for a total of 2min (600 images). Prior to imaging, the field-of-view was exposed to 100µW of 488 nm laser light for 4min to induce translocation of either Pi-mCherry or Pi-Myc into the nucleus. The 488nm laser was left on during imaging of the JF646 to retain Pi-mCherry or Pi-Myc in the nucleus (supplement movie). The JF646 dye attached to the protein of interest was excited with 647nm with a laser power of 1 mW measured at the back aperture of the objective. Tracking was performed automatically with the MATLAB-based Track Record software (freely available at https://sourceforge.net/projects/single-molecule-tracking). Track Record links molecule positions in two consecutive planes using a nearest-neighbor approach [84]. Molecules are considered bound if they move less than $r_{max} = 530$nm from one frame

to the next (Ball et al., 2016) for at least $N_{min} = 8$ frames. The value for r_{max} is the distance that 99% of tagged histone H2B molecules moved in the same imaging conditions. Assuming a moderate diffusion coefficient, D_f, of 0.5 $\mu m^2/s$, with our frame interval, δt, of 200ms, The minimum number of frames, N_{min}, is calculated such that the probability, P, that a freely diffusing molecule will be classified as bound is below 1% [84]:

$$P = [1 - e^{-\frac{r_{max}^2}{4D_f \Delta t}}]^{N_{min}} \tag{A.1}$$

A.6.4 Analysis of Survival Curves

We used two different methods to analyze the residence times of the various target proteins, both of which attempt to correct for photobleaching. The first method relies on correcting for photobleaching separately in each experiment by fitting the number of particles over time to a bi-exponential. The survival curve is then divided by this photobleaching bi-exponential before being fit to both mono- and bi-exponentials [84]. An F-test is performed to determine if the added parameters in the bi-exponential provide a statistically significant improvement in the fit.

The second method requires a separate dataset where a very stable molecule, such as the histone H2 is measured [41]. The survival curve for H2 is then fit with a tri-exponential distribution:

$$Y = [A_1 e^{-k_1 t} + A_2 e^{-k_2 t} + (1 - A_1 - A_2) e^{-k_3 t}]C \tag{A.2}$$

Where the slowest decay rate k_3 is assumed to be loss of fluorescence due to photobleaching (as opposed to the molecule physically leaving the location), and is used as the correction in the samples of interest. This latter method fits the data to alternate models to the mono- and bi-exponential including a power law, and a combination of a power law and an exponential. The Bayesian Information Criterion (BIC) is calculated for each of the four fits, and the one with the lowest BIC is

selected. The bi-exponential model is given as:

$$Y = [Ae^{-k_1 t} + (1 - A)e^{-k_2 t}]C \tag{A.3}$$

And the power-law model is given as:

$$Y = [At^{-\alpha} + (1 - A)e^{-\beta t}]C \tag{A.4}$$

Appendix B

Calculating confidence intervals of distribution means with bootstrapping

The 95% confidence intervals of the distribution means were calculated in R with 10,000 bootstrap replicates.

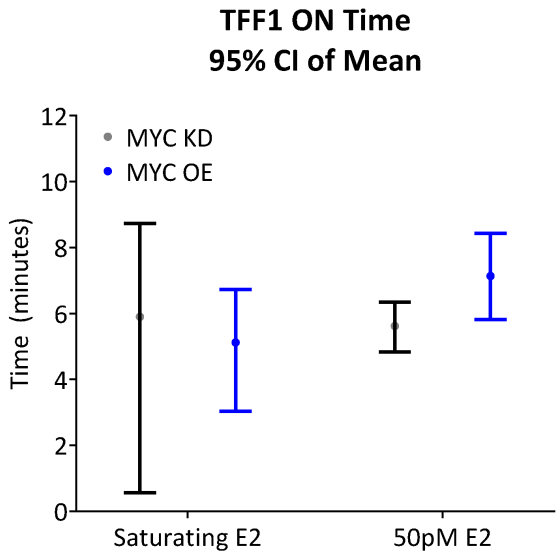


Figure B.1: TFF1 transcription ON time (Related to **Figure 4.6**).

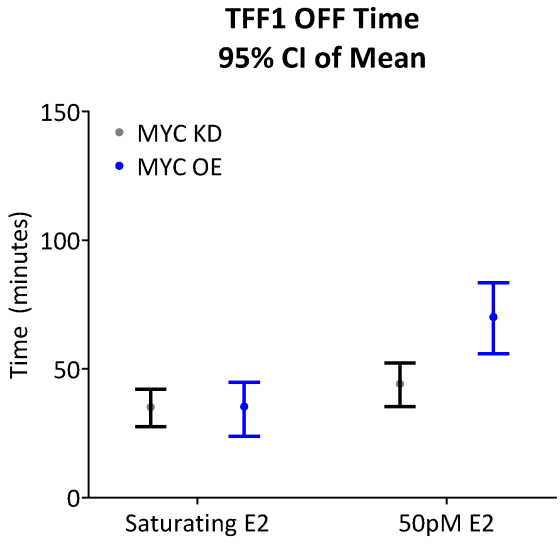


Figure B.2: TFF1 transcription OFF time (Related to **Figure 4.6**).

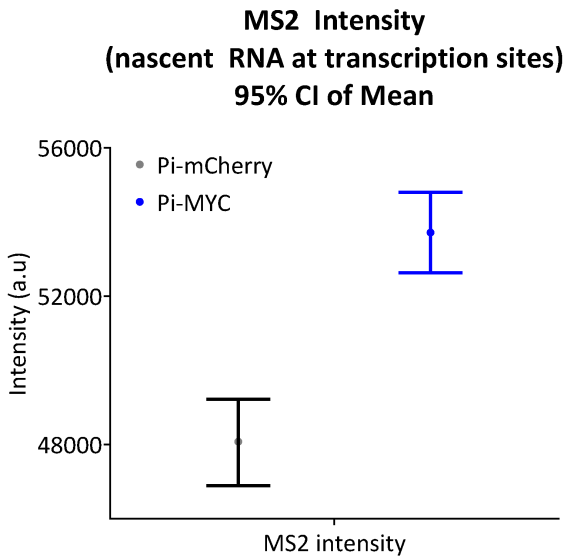


Figure B.3: TS intensity in MS2 polyclonal line (Related to (**Figure 3.11**).

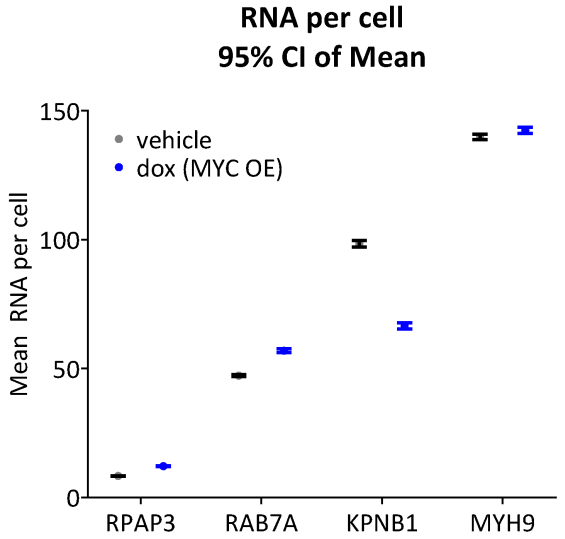


Figure B.4: HBEC gene RNA distribution (Related to **Figure 4.20**).

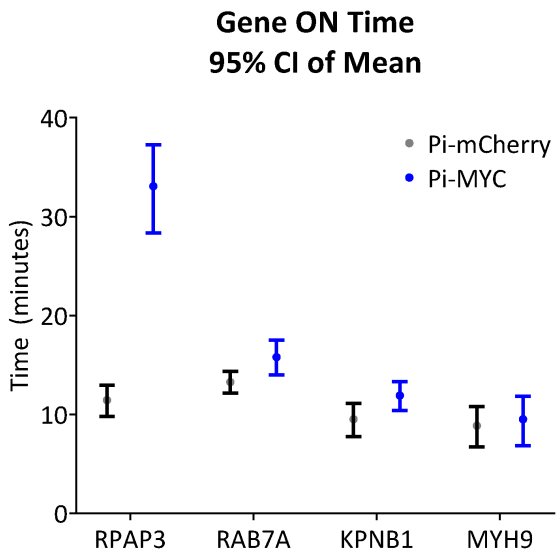


Figure B.5: HBEC gene ON Time (Related to **Figure 4.23**).

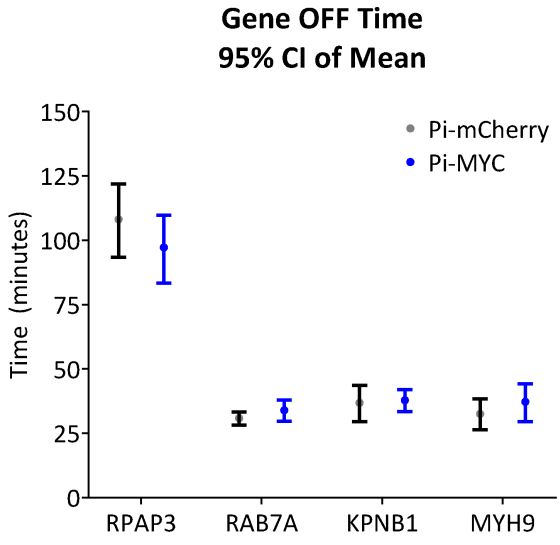


Figure B.6: HBEC gene OFF time (Related to **Figure 4.23**).

Bibliography

[1] Adams, J. M., Harris, A. W., Pinkert, C. A., Corcoran, L. M., Alexander, W. S., Cory, S., Palmiter, R. D., and Brinster, R. L. The c-myc oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice. *Nature 318*, 6046 (dec 1985), 533–538.

[2] Adamskaya, N., Dungel, P., Mittermayr, R., Hartinger, J., Feichtinger, G., Wassermann, K., Redl, H., and Van Griensven, M. Light therapy by blue LED improves wound healing in an excision model in rats. *Injury 42*, 9 (sep 2011), 917–921.

[3] Al-azzeh, E., Dittrich, O., Vervoorts, J., Blin, N., Gött, P., and Lüscher, B. Gastroprotective peptide trefoil factor family 2 gene is activated by upstream stimulating factor but not by c-Myc in gastrointestinal cancer cells. *Gut 51*, 5 (nov 2002), 685–690.

[4] Amiry, N., Kong, X., Muniraj, N., Kannan, N., Grandison, P. M., Lin, J., Yang, Y., Vouyovitch, C. M., Borges, S., Perry, J. K., Mertani, H. C., Zhu, T., Liu, D., and Lobie, P. E. Trefoil Factor-1 (TFF1) Enhances Oncogenicity of Mammary Carcinoma Cells. *Endocrinology 150*, 10 (oct 2009), 4473–4483.

[5] Bahar Halpern, K., Caspi, I., Lemze, D., Levy, M., Landen, S., Elinav, E., Ulitsky, I., and Itzkovitz, S. Nuclear Retention of mRNA in Mammalian Tissues. *Cell reports 13*, 12 (dec 2015), 2653–62.

[6] Bahar Halpern, K., and Itzkovitz, S. Single molecule approaches for quantifying transcription and degradation rates in intact mammalian tissues. *Methods 98* (apr 2016), 134–142.

[7] Ball, D. A., Mehta, G. D., Salomon-Kent, R., Mazza, D., Morisaki, T., Mueller, F., Mcnally, J. G., and Karpova, T. S. Single molecule tracking of Ace1p in Saccharomyces cerevisiae defines a characteristic residence time for non-specific interactions of transcription factors with chromatin. *Nucleic Acids Research 44* (2016).

[8] Baluapuri, A., Hofstetter, J., Dudvarski Stankovic, N., Endres, T., Bhandare, P., Vos, S. M., Adhikari, B., Schwarz, J. D., Narain, A., Vogt, M., Wang, S. Y., Düster, R., Jung, L. A., Vanselow, J. T., Wiegering, A., Geyer, M., Maric, H. M., Gallant, P., Walz, S., Schlosser, A., Cramer, P., Eilers, M., and Wolf, E. MYC Recruits

SPT5 to RNA Polymerase II to Promote Processive Transcription Elongation. *Molecular Cell 74*, 4 (may 2019), 674–687.e11.

[9] Bartman, C., Hsu, S., Hsiung, C.-S., Raj, A., and Blobel, G. Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. *Molecular Cell 62*, 2 (apr 2016), 237–247.

[10] Battich, N., Stoeger, T., and Pelkmans, L. Control of Transcript Variability in Single Mammalian Cells. *Cell 163*, 7 (dec 2015), 1596–1610.

[11] Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., Ohl, P., Sieb, C., Thiel, K., and Wiswedel, B. KNIME: The Konstanz Information Miner. Springer, Berlin, Heidelberg, 2008, pp. 319–326.

[12] Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S. M., Singer, R. H., and Long, R. M. Localization of ASH1 mRNA particles in living yeast. *Molecular cell 2*, 4 (oct 1998), 437–45.

[13] Bothma, J. P., Garcia, H. G., Esposito, E., Schlissel, G., Gregor, T., and Levine, M. Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living Drosophila embryos. *Proceedings of the National Academy of Sciences of the United States of America 111*, 29 (jul 2014), 10598–603.

[14] Bravo, H. C., Pihur, V., McCall, M., Irizarry, R. A., and Leek, J. T. Gene expression anti-profiles as a basis for accurate universal cancer signatures. *BMC bioinformatics 13* (oct 2012), 272.

[15] Cai, L., Dalal, C. K., and Elowitz, M. B. Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature 455*, 7212 (sep 2008), 485–490.

[16] Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S., and Zhuang, X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science (New York, N.Y.) 348*, 6233 (apr 2015), aaa6090.

[17] Cho, W. K., Jayanth, N., Mullen, S., Tan, T. H., Jung, Y. J., and Cissé, I. I. Super-resolution imaging of fluorescently labeled, endogenous RNA Polymerase II in living cells with CRISPR/Cas9-mediated gene editing. *Scientific Reports 6*, 1 (oct 2016), 1–8.

[18] Comorosan, S., Kappel, W., Constantinescu, I., Gheorghe, M., Ionescu, E., Pîrvu, C., Cinca, S., and Cristache, L. Green light effects on biological systems: a new biophysical phenomenon. *Journal of Biological Physics 35*, 3 (aug 2009), 265–277.

[19] Cote, A. J., McLeod, C. M., Farrell, M. J., McClanahan, P. D., Dunagin, M. C., Raj, A., and Mauck, R. L. Single-cell differences in matrix gene expression do not predict matrix deposition. *Nature Communications 7* (mar 2016), 10865.

[20] Coulon, A., Ferguson, M. L., de Turris, V., Palangat, M., Chow, C. C., and Larson, D. R. Kinetic competition during the transcription cycle results in stochastic RNA processing. *eLife 3* (oct 2014).

[21] Coulon, A., and Larson, D. Fluctuation Analysis. In *Methods in enzymology*. 2016, pp. 159–191.

[22] DeLaughter, D. M., Bick, A. G., Wakimoto, H., McKean, D., Gorham, J. M., Kathiriya, I. S., Hinson, J. T., Homsy, J., Gray, J., Pu, W., Bruneau, B. G., Seidman, J., and Seidman, C. E. Single-Cell Resolution of Temporal Gene Expression during Heart Development. *Developmental Cell 39*, 4 (2016), 480–490.

[23] Dietz, C., and Berthold, M. R. KNIME for open-source bioimage analysis: A tutorial. *Advances in Anatomy Embryology and Cell Biology 219* (may 2016), 179–197.

[24] Dinalankara, W., and Bravo, H. C. Gene Expression Signatures Based on Variability can Robustly Predict Tumor Progression and Prognosis. *Cancer informatics 14* (2015), 71–81.

[25] Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., and Ren, B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature 485*, 7398 (apr 2012), 376–80.

[26] Donovan, B. T., Huynh, A., Ball, D. A., Patel, H. P., Poirier, M. G., Larson, D. R., Ferguson, M. L., and Lenstra, T. L. Live-cell imaging reveals the interplay between transcription factors, nucleosomes, and bursting. *The EMBO Journal 38*, 12 (jun 2019).

[27] Dueck, H., Eberwine, J., and Kim, J. Variation is function: Are single cell differences functionally important? *BioEssays 38*, 2 (feb 2016), 172–180.

[28] Edelstein, A. D., Tsuchida, M. A., Amodaj, N., Pinkard, H., Vale, R. D., and Stuurman, N. Advanced methods of microscope control using μManager software. *Journal of Biological Methods 1*, 2 (nov 2014), 10.

[29] Eilers, M., Schirm, S., and Bishop, J. M. The MYC protein activates transcription of the alpha-prothymosin gene. *The EMBO Journal 10*, 1 (jan 1991), 133–141.

[30] Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S. Stochastic Gene Expression in a Single Cell. *Science 297*, 5584 (2002).

[31] English, B. P., and Singer, R. H. A three-camera imaging microscope for high-speed single-molecule tracking and super-resolution imaging in living cells. In *Biosensing and Nanomedicine VIII* (aug 2015), vol. 9550, SPIE, p. 955008.

[32] Ennen, M., Keime, C., Kobi, D., Mengus, G., Lipsker, D., Thibault-Carpentier, C., and Davidson, I. Single-cell gene expression signatures reveal melanoma cell heterogeneity. *Oncogene 34*, 25 (jun 2015), 3251–3263.

[33] Espinoza, J. H., and Mercado-Uribe, H. Visible light neutralizes the effect produced by ultraviolet radiation in proteins. *Journal of Photochemistry and Photobiology B: Biology 167* (feb 2017), 15–19.

[34] Fan, H. C., Fu, G. K., and Fodor, S. P. A. Combinatorial labeling of single cells for gene expression cytometry. *Science 347*, 6222 (feb 2015), 1258367–1258367.

[35] Femino, A. M. Visualization of Single RNA Transcripts in Situ. *Science 280*, 5363 (1998), 585–590.

[36] Flavahan, W. A., Drier, Y., Liau, B. B., Gillespie, S. M., Venteicher, A. S., Stemmer-Rachamimov, A. O., Suvà, M. L., and Bernstein, B. E. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature 529*, 7584 (jan 2016), 110–4.

[37] Fletcher, R. B., Das, D., Gadye, L., Street, K. N., Baudhuin, A., Wagner, A., Cole, M. B., Flores, Q., Choi, Y. G., Yosef, N., Purdom, E., Dudoit, S., Risso, D., and Ngai, J. Deconstructing Olfactory Stem Cell Trajectories at Single-Cell Resolution. *Cell Stem Cell 20*, 6 (2017), 817–830.e8.

[38] Fritzsch, C., Baumgärtner, S., Kuban, M., Steinshorn, D., Reid, G., and Legewie, S. Estrogen-dependent control and cell-to-cell variability of transcriptional bursting. *Molecular Systems Biology 14*, 2 (feb 2018).

[39] Fu, A. Q., and Pachter, L. Estimating intrinsic and extrinsic noise from single-cell gene expression measurements. *Statistical applications in genetics and molecular biology 15*, 6 (dec 2016), 447–471.

[40] Fukaya, T., Lim, B., and Levine, M. Enhancer Control of Transcriptional Bursting. *Cell 166*, 2 (jul 2016), 358–368.

[41] Garcia, D. A., Fettweis, G., Presman, D. M., Paakinaho, V., Jarzynski, C., Upadhyaya, A., and Hager, G. L. A New Model for Single-Molecule Tracking Analysis of Transcription Factor Dynamics. *bioRxiv* (may 2019), 637355.

[42] Gibcus, J. H., and Dekker, J. The hierarchy of the 3D genome. *Molecular cell 49*, 5 (mar 2013), 773–82.

[43] Grandori, C., Gomez-Roman, N., Felton-Edkins, Z. A., Ngouenet, C., Galloway, D. A., Eisenman, R. N., and White, R. J. c-Myc binds to human ribosomal DNA and stimulates transcription of rRNA genes by RNA polymerase I. *Nature Cell Biology 7*, 3 (mar 2005), 311–318.

[44] Grimm, J. B., Muthusamy, A. K., Liang, Y., Brown, T. A., Lemon, W. C., Patel, R., Lu, R., Macklin, J. J., Keller, P. J., Ji, N., and Lavis, L. D. A general method to fine-tune fluorophores for live-cell and in vivo imaging. *Nature Methods 14*, 10 (oct 2017), 987–994.

[45] Halstead, J. M., Lionnet, T., Wilbertz, J. H., Wippich, F., Ephrussi, A., Singer, R. H., and Chao, J. A. An RNA biosensor for

imaging the first round of translation from single cells to living animals. *Science (New York, N.Y.) 347*, 6228 (mar 2015), 1367–671.

[46] Hansen, A. S., and O'Shea, E. K. Limits on information transduction through amplitude and frequency regulation of transcription factor activity. *eLife 4* (may 2015).

[47] Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R., and Darzacq, X. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *eLife 6* (may 2017).

[48] Harper, C. V., Finkenstädt, B., Woodcock, D. J., Friedrichsen, S., Semprini, S., Ashall, L., Spiller, D. G., Mullins, J. J., Rand, D. A., Davis, J. R. E., and White, M. R. H. Dynamic Analysis of Stochastic Transcription Cycles. *PLoS Biology 9*, 4 (apr 2011), e1000607.

[49] Hashimshony, T., Senderovich, N., Avital, G., Klochendler, A., de Leeuw, Y., Anavy, L., Gennert, D., Li, S., Livak, K. J., Rozenblatt-Rosen, O., Dor, Y., Regev, A., and Yanai, I. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome biology 17* (apr 2016), 77.

[50] Hnisz, D., Weintraub, A. S., Day, D. S., Valton, A.-L., Bak, R. O., Li, C. H., Goldmann, J., Lajoie, B. R., Fan, Z. P., Sigova, A. A., Reddy, J., Borges-Rivera, D., Lee, T. I., Jaenisch, R., Porteus, M. H., Dekker, J., and Young, R. A. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science (New York, N.Y.) 351*, 6280 (mar 2016), 1454–1458.

[51] Hsu, S. C., Gilgenast, T. G., Bartman, C. R., Edwards, C. R., Stonestrom, A. J., Huang, P., Emerson, D. J., Evans, P., Werner, M. T., Keller, C. A., Giardine, B., Hardison, R. C., Raj, A., Phillips-Cremins, J. E., and Blobel, G. A. The BET Protein BRD2 Cooperates with CTCF to Enforce Transcriptional and Architectural Boundaries. *Molecular Cell 66*, 1 (2017), 102–116.e7.

[52] Hsu, T. Y.-T., Simon, L. M., Neill, N. J., Marcotte, R., Sayad, A., Bland, C. S., Echeverria, G. V., Sun, T., Kurley, S. J., Tyagi, S., Karlin, K. L., Dominguez-Vidaña, R., Hartman, J. D., Renwick, A., Scorsone, K., Bernardi, R. J., Skinner, S. O., Jain, A., Orellana, M., Lagisetti, C., Golding, I., Jung, S. Y., Neilson, J. R., Zhang, X. H.-F., Cooper, T. A., Webb, T. R., Neel, B. G., Shaw, C. A., and Westbrook, T. F. The spliceosome is a therapeutic vulnerability in MYC-driven cancer. *Nature 525*, 7569 (sep 2015), 384–388.

[53] Izeddin, I., Récamier, V., Bosanac, L., Cissé, I. I., Boudarene, L., Dugast-Darzacq, C., Proux, F., Bénichou, O., Voituriez, R., Bensaude, O., Dahan, M., and Darzacq, X. Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus. *eLife 3* (jun 2014).

[54] Jaitin, D. A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A., and Amit, I. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science (New York, N.Y.) 343*, 6172 (feb 2014), 776–9.

[55] Janicki, S. M., Tsukamoto, T., Salghetti, S. E., Tansey, W. P., Sachidanandam, R., Prasanth, K. V., Ried, T., Shav-Tal, Y., Bertrand, E., Singer, R. H., and Spector, D. L. From silencing to gene expression: real-time analysis in single cells. *Cell 116*, 5 (mar 2004), 683–98.

[56] Kalkat, M., Resetca, D., Lourenco, C., Chan, P. K., Wei, Y., Shiah, Y. J., Vitkin, N., Tong, Y., Sunnerhagen, M., Done, S. J., Boutros, P. C., Raught, B., and Penn, L. Z. MYC Protein Interactome Profiling Reveals Functionally Distinct Regions that Cooperate to Drive Tumorigenesis. *Molecular Cell 72*, 5 (dec 2018), 836–848.e7.

[57] Khan, A. O., White, C. W., Pike, J. A., Yule, J., Slater, A., Hill, S. J., Poulter, N. S., Thomas, S. G., and Morgan, N. V. Optimised insert design for improved single-molecule imaging and quantification through CRISPR-Cas9 mediated knock-in. *Scientific Reports 9*, 1 (dec 2019), 1–13.

[58] Kolodziejczyk, A., Kim, J. K., Svensson, V., Marioni, J., Teichmann, S., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., and et Al. The Technology and Biology of Single-Cell RNA Sequencing. *Molecular Cell 58*, 4 (may 2015), 610–620.

[59] Kreso, A., O'Brien, C. A., Van Galen, P., Gan, O. I., Notta, F., Brown, A. M., Ng, K., Jing, M., Wienholds, E., Dunant, C., Pollett, A., Gallinger, S., McPherson, J., Mullighan, C. G., Shibata, D., and Dick, J. E. Variable clonal repopulation dynamics influence chemotherapy response in colorectal cancer. *Science 339*, 6119 (feb 2013), 543–548.

[60] Kress, T. R., Sabò, A., and Amati, B. MYC: Connecting selective transcriptional control to global RNA production, sep 2015.

[61] Ku, W. L., Duggal, G., Li, Y., Girvan, M., and Ott, E. Interpreting Patterns of Gene Expression: Signatures of Coregulation, the Data Processing Inequality, and Triplet Motifs. *PLoS ONE 7*, 2 (feb 2012), e31969.

[62] Land, H., Chen, A. C., Morgenstern, J. P., Parada, L. F., and Weinberg, R. A. Behavior of myc and ras oncogenes in transformation of rat embryo fibroblasts. *Molecular and Cellular Biology 6*, 6 (jun 1986), 1917–1925.

[63] Larson, D. R., Fritzsch, C., Sun, L., Meng, X., Lawrence, D. S., and Singer, R. H. Direct observation of frequency modulated transcription in single cells using light activation. *eLife 2013*, 2 (sep 2013).

[64] Larson, D. R., Zenklusen, D., Wu, B., Chao, J. A., and Singer, R. H. Real-Time Observation of Transcription Initiation and Elongation on an Endogenous Yeast Gene. *Science 332*, 6028 (apr 2011), 475–478.

[65] Lee, J. H., Daugharthy, E. R., Scheiman, J., Kalhor, R., Ferrante, T. C., Terry, R., Turczyk, B. M., Yang, J. L., Lee, H. S., Aach, J., Zhang, K., and Church, G. M. Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nature protocols 10*, 3 (mar 2015), 442–58.

[66] Lee, T. H. Extracting kinetics information from single-molecule fluorescence resonance energy transfer data using hidden markov models. *Journal of Physical Chemistry B 113*, 33 (2009), 11535–11542.

[67] Lenstra, T., Coulon, A., Chow, C., and Larson, D. Single-Molecule Imaging Reveals a Switch between Spurious and Functional ncRNA Transcription. *Molecular Cell 60*, 4 (2015), 597–610.

[68] Levesque, M. J., Ginart, P., Wei, Y., and Raj, A. Visualizing SNVs to quantify allele-specific expression in single cells. *Nature methods 10*, 9 (sep 2013), 865–7.

[69] Levsky, J. M., Shenoy, S. M., Pezo, R. C., and Singer, R. H. Single-Cell Gene Expression Profiling. *Science 297*, 5582 (aug 2002), 836–840.

[70] Lewis, L. M., Edwards, M. C., Meyers, Z. R., Talbot, C. C., Hao, H., Blum, D., Iorns, E., Tsui, R., Denis, A., Perfito, N., and Errington, T. M. Replication study: Transcriptional amplification in tumor cells with elevated c-Myc. *eLife 7* (jan 2018).

[71] Lin, C. Y., Lovén, J., Rahl, P. B., Paranal, R. M., Burge, C. B., Bradner, J. E., Lee, T. I., and Young, R. A. Transcriptional amplification in tumor cells with elevated c-Myc. *Cell 151*, 1 (sep 2012), 56–67.

[72] Lionnet, T., Czaplinski, K., Darzacq, X., Shav-Tal, Y., Wells, A. L., Chao, J. A., Park, H. Y., de Turris, V., Lopez-Jones, M., and Singer, R. H. A transgenic mouse for in vivo detection of endogenous labeled mRNA. *Nature Methods 8*, 2 (feb 2011), 165–170.

[73] Liu, X., Vorontchikhina, M., Wang, Y.-L., Faiola, F., and Martinez, E. STAGA Recruits Mediator to the MYC Oncoprotein To Stimulate Transcription and Cell Proliferation. *Molecular and Cellular Biology 28*, 1 (jan 2008), 108–121.

[74] Llamosi, A., Gonzalez-Vargas, A. M., Versari, C., Cinquemani, E., Ferrari-Trecate, G., Hersen, P., and Batt, G. What Population Reveals about Individual Cell Identity: Single-Cell Parameter Estimation of Models of Gene Expression in Yeast. *PLOS Computational Biology 12*, 2 (feb 2016), e1004706.

[75] Lorenzin, F., Benary, U., Baluapuri, A., Walz, S., Jung, L. A., von Eyss, B., Kisker, C., Wolf, J., Eilers, M., and Wolf, E. Different

promoter affinities account for specificity in MYC-dependent gene regulation. *eLife 5*, JULY (jul 2016).

[76] Los, G. V., Encell, L. P., McDougall, M. G., Hartzell, D. D., Karassina, N., Zimprich, C., Wood, M. G., Learish, R., Ohana, R. F., Urh, M., Simpson, D., Mendez, J., Zimmerman, K., Otto, P., Vidugiris, G., Zhu, J., Darzins, A., Klaubert, D. H., Bulleit, R. F., and Wood, K. V. HaloTag: A novel protein labeling technology for cell imaging and protein analysis. *ACS Chemical Biology 3*, 6 (jun 2008), 373–382.

[77] Lubeck, E., and Cai, L. Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nature methods 9*, 7 (jun 2012), 743–8.

[78] Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M., and Cai, L. Single-cell in situ RNA profiling by sequential hybridization. *Nature methods 11*, 4 (apr 2014), 360–1.

[79] Luria, S. E., and Delbrück, M. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics 28*, 6 (nov 1943), 491–511.

[80] Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M., Trombetta, J. J., Weitz, D. A., Sanes, J. R., Shalek, A. K., Regev, A., and McCarroll, S. A. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell 161*, 5 (2015), 1202–1214.

[81] Mantsoki, A., Devailly, G., and Joshi, A. Gene expression variability in mammalian embryonic stem cells using single cell RNA-seq data. *Computational Biology and Chemistry 63* (2016), 52–61.

[82] Martin, R. M., Rino, J., Carvalho, C., Kirchhausen, T., and Carmo-Fonseca, M. Live-cell visualization of pre-mRNA splicing with single-molecule sensitivity. *Cell reports 4*, 6 (sep 2013), 1144–55.

[83] May, F. E., and Westley, B. R. Identification and characterization of estrogen-regulated RNAs in human breast cancer cells. *The Journal of Biological Chemistry 263*, 26 (sep 1988), 12901–8.

[84] Mazza, D., Abernathy, A., Golob, N., Morisaki, T., and McNally, J. G. A benchmark for chromatin binding measurements in live cells. *Nucleic acids research 40*, 15 (aug 2012), e119.

[85] McEwan, I. J., Dahlman-Wright, K., Ford, J., and Wright, A. P. Functional interaction of the c-Myc transactivation domain with the TATA binding protein: Evidence for an induced fit model of transactivation domain folding. *Biochemistry 35*, 29 (1996), 9584–9593.

[86] Mehta, G. D., Ball, D. A., Eriksson, P. R., Chereji, R. V., Clark, D. J., McNally, J. G., and Karpova, T. S. Single-Molecule Analysis Reveals Linked Cycles of RSC Chromatin Remodeling and Ace1p Transcription Factor Binding in Yeast. *Molecular Cell 72*, 5 (dec 2018), 875–887.e9.

[87] Mellis, I. A., Gupte, R., Raj, A., and Rouhanifard, S. H. Visualizing adenosine-to-inosine RNA editing in single mammalian cells. *Nature methods 14*, 8 (aug 2017), 801–804.

[88] Moffitt, J. R., Hao, J., Wang, G., Chen, K. H., Babcock, H. P., and Zhuang, X. High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. *Proceedings of the National Academy of Sciences of the United States of America 113*, 39 (sep 2016), 11046–51.

[89] Morisaki, T., Lyon, K., DeLuca, K. F., DeLuca, J. G., English, B. P., Zhang, Z., Lavis, L. D., Grimm, J. B., Viswanathan, S., Looger, L. L., Lionnet, T., and Stasevich, T. J. Real-time quantification of single RNA translation dynamics in living cells. *Science 352*, 6292 (jun 2016), 1425–1429.

[90] Nakamura, M., Chen, L., Howes, S. C., Schindler, T. D., Nogales, E., and Bryant, Z. Remote control of myosin and kinesin motors using light-activated gearshifting. *Nature Nanotechnology 9*, 9 (aug 2014), 693–697.

[91] Nie, Z., Guo, C., Das, S. K., Chow, C. C., Batchelor, E., Simons, S. S., and Levens, D. Dissecting transcriptional amplification by MYC. *[Manuscript submitted for publication]* (2020).

[92] Nie, Z., Hu, G., Wei, G., Cui, K., Yamane, A., Resch, W., Wang, R., Green, D. R., Tessarollo, L., Casellas, R., Zhao, K., and Levens, D. c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells. *Cell 151*, 1 (2012), 68–79.

[93] Niopek, D., Benzinger, D., Roensch, J., Draebing, T., Wehler, P., Eils, R., and Di Ventura, B. Engineering light-inducible nuclear localization signals for precise spatiotemporal control of protein dynamics in living cells. *Nature Communications 5*, 1 (jul 2014), 1–11.

[94] Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N. L., Meisig, J., Sedat, J., Gribnau, J., Barillot, E., Blüthgen, N., Dekker, J., and Heard, E. Spatial partitioning of the regulatory landscape of the X-inactivation center. *Nature 485*, 7398 (2012), 381.

[95] Ochiai, H., Sugawara, T., and Yamamoto, T. Simultaneous live imaging of the transcription and nuclear position of specific genes. *Nucleic acids research 43*, 19 (oct 2015), e127.

[96] Oka, Y., and Sato, T. N. Whole-mount single molecule FISH method for zebrafish embryo. *Scientific reports 5* (feb 2015), 8571.

[97] Oliinyk, O. S., Shemetov, A. A., Pletnev, S., Shcherbakova, D. M., and Verkhusha, V. V. Smallest near-infrared fluorescent protein evolved from cyanobacteriochrome as versatile tag for spectral multiplexing. *Nature Communications 10*, 1 (dec 2019), 1–13.

[98] Paakinaho, V., Presman, D. M., Ball, D. A., Johnson, T. A., Schiltz, R. L., Levitt, P., Mazza, D., Morisaki, T., Karpova, T. S., and Hager, G. L. Single-molecule analysis of steroid receptor and cofactor action in living cells. *Nature Communications 8*, 1 (jun 2017), 1–14.

[99] Palangat, M., Anastasakis, D. G., Fei, D. L., Lindblad, K. E., Bradley, R., Hourigan, C. S., Hafner, M., and Larson, D. R. The splicing factor U2AF1 contributes to cancer progression through a noncanonical role in translation regulation. *Genes and Development 33*, 9-10 (may 2019), 482–497.

[100] Peccoud, J., and Ycart, B. Markovian Modeling of Gene-Product Synthesis. *Theoretical Population Biology 48*, 2 (oct 1995), 222–234.

[101] Peters, O., and Klein, W. Ergodicity Breaking in Geometric Brownian Motion. *Physical Review Letters 110*, 10 (mar 2013), 100603.

[102] Pomerance, A., Ott, E., Girvan, M., and Losert, W. The effect of network topology on the stability of discrete state models of genetic control. *Proceedings of the National Academy of Sciences of the United States of America 106*, 20 (may 2009), 8209–14.

[103] Porter, J. R., Fisher, B. E., Baranello, L., Liu, J. C., Kambach, D. M., Nie, Z., Koh, W. S., Luo, J., Stommel, J. M., Levens, D., and Batchelor, E. Global Inhibition with Specific Activation: How p53 and MYC Redistribute the Transcriptome in the DNA Double-Strand Break Response. *Molecular Cell 67*, 6 (sep 2017), 1013–1025.e9.

[104] Presman, D. M., Ball, D. A., Paakinaho, V., Grimm, J. B., Lavis, L. D., Karpova, T. S., and Hager, G. L. Quantifying transcription factor binding dynamics at the single-molecule level in live cells. *Methods 123* (jul 2017), 76–88.

[105] Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y., and Tyagi, S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biology 4*, 10 (2006), 1707–1719.

[106] Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A., and Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nature methods 5*, 10 (oct 2008), 877–9.

[107] Ramirez, R. D., Sheridan, S., Girard, L., Sato, M., Kim, Y., Pollack, J., Peyton, M., Zou, Y., Kurie, J. M., DiMaio, J. M., Milchgrub, S., Smith, A. L., Souza, R. F., Gilbey, L., Zhang, X., Gandia, K., Vaughan, M. B., Wright, W. E., Gazdar, A. F., Shay, J. W., and Minna, J. D. Immortalization of human bronchial epithelial cells in the absence of viral oncoproteins. *Cancer Research 64*, 24 (dec 2004), 9027–9034.

[108] Raser, J. M., and O'Shea, E. K. Control of stochasticity in eukaryotic gene expression. *Science (New York, N.Y.) 304*, 5678 (jun 2004), 1811–4.

[109] Ren, G., Jin, W., Cui, K., Rodrigez, J., Hu, G., Zhang, Z., Larson, D. R., and Zhao, K. CTCF-Mediated Enhancer-Promoter Interaction Is

a Critical Regulator of Cell-to-Cell Variation of Gene Expression. *Molecular Cell 67*, 6 (sep 2017), 1049–1058.e6.

[110] Rieckh, G., and Tkačik, G. Noise and information transmission in promoters with multiple internal States. *Biophysical journal 106*, 5 (mar 2014), 1194–204.

[111] Rodriguez, J., and Larson, D. R. Transcription in Living Cells: Molecular Mechanisms of Bursting. *Annual Review of Biochemistry 89*, 1 (jun 2020), annurev–biochem–011520–105250.

[112] Rodriguez, J., Ren, G., Day, C. R., Zhao, K., Chow, C. C., and Larson, D. R. Intrinsic Dynamics of a Human Gene Reveal the Basis of Expression Heterogeneity. *Cell 176*, 1-2 (jan 2019), 213–226.e18.

[113] Rohringer, S., Holnthoner, W., Chaudary, S., Slezak, P., Priglinger, E., Strassl, M., Pill, K., Mühleder, S., Redl, H., and Dungel, P. The impact of wavelengths of LED light-therapy on endothelial cells. *Scientific Reports 7*, 1 (dec 2017).

[114] Rudolph, B., Hueber, A. O., and Evan, G. I. Reversible activation of c-Myc in thymocytes enhances positive selection and induces proliferation and apoptosis in vitro. *Oncogene 19*, 15 (apr 2000), 1891–1900.

[115] Sabò, A., Kress, T. R., Pelizzola, M., De Pretis, S., Gorski, M. M., Tesi, A., Morelli, M. J., Bora, P., Doni, M., Verrecchia, A., Tonelli, C., Fagà, G., Bianchi, V., Ronchi, A., Low, D., Müller, H., Guccione, E., Campaner, S., and Amati, B. Selective transcriptional regulation by Myc in cellular growth control and lymphomagenesis. *Nature 511*, 7510 (jul 2014), 488–492.

[116] Sanchez, A., Choubey, S., and Kondev, J. Stochastic models of transcription: From single molecules to single cells. *Methods 62*, 1 (jul 2013), 13–25.

[117] Schmiedel, J. M., Klemm, S. L., Zheng, Y., Sahay, A., Bluthgen, N., Marks, D. S., and van Oudenaarden, A. MicroRNA control of protein expression noise. *Science 348*, 6230 (apr 2015), 128–132.

[118] Schoenfelder, S., Sexton, T., Chakalova, L., Cope, N. F., Horton, A., Andrews, S., Kurukuti, S., Mitchell, J. A., Umlauf, D., Dimitrova, D. S., Eskiw, C. H., Luo, Y., Wei, C.-L., Ruan, Y., Bieker, J. J., and Fraser, P. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nature Genetics 42*, 1 (jan 2010), 53–61.

[119] Shaffer, S. M., Dunagin, M. C., Torborg, S. R., Torre, E. A., Emert, B., Krepler, C., Beqiri, M., Sproesser, K., Brafford, P. A., Xiao, M., Eggan, E., Anastopoulos, I. N., Vargas-Garcia, C. A., Singh, A., Nathanson, K. L., Herlyn, M., and Raj, A. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature 546*, 7658 (jun 2017), 431–435.

[120] Shannon, C. E. A Mathematical Theory of Communication. *Bell System Technical Journal 27*, 3 (jul 1948), 379–423.

[121] Sherman, M. S., Lorenz, K., Lanier, M. H., and Cohen, B. A. Cell-to-cell variability in the propensity to transcribe explains correlated fluctuations in gene expression. *Cell systems 1*, 5 (nov 2015), 315–325.

[122] So, L.-H., Ghosh, A., Zong, C., Sepúlveda, L. A., Segev, R., and Golding, I. General properties of transcriptional time series in Escherichia coli. *Nature genetics 43*, 6 (jun 2011), 554–60.

[123] Soutto, M., Peng, D. F., Katsha, A., Chen, Z., Piazuelo, M. B., Washington, M. K., Belkhiri, A., Correa, P., and El-Rifai, W. Activation of β-catenin signalling by TFF1 loss promotes cell proliferation and gastric tumorigenesis. *Gut 64*, 7 (jul 2015), 1028–1039.

[124] Spencer, C. A., and Groudine, M. Control of c-myc regulation in normal and neoplastic cells. *Advances in Cancer Research 56* (jan 1991), 1–48.

[125] Suter, D. M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., and Naef, F. Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics. *Science 332*, 6028 (apr 2011), 472–474.

[126] Swinstead, E. E., Miranda, T. B., Paakinaho, V., Baek, S., Goldstein, I., Hawkins, M., Karpova, T. S., Ball, D., Mazza, D., Lavis, L. D., Grimm, J. B., Morisaki, T., Grøntved, L., Presman, D. M., and Hager, G. L. Steroid Receptors Reprogram FoxA1 Occupancy through Dynamic Chromatin Transitions. *Cell 165*, 3 (apr 2016), 593–605.

[127] Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., Siddiqui, A., Lao, K., and Surani, M. A. mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods 6*, 5 (may 2009), 377–382.

[128] Teves, S. S., An, L., Bhargava-Shah, A., Xie, L., Darzacq, X., and Tjian, R. A stable mode of bookmarking by TBP recruits RNA polymerase II to mitotic chromosomes. *eLife 7* (jun 2018).

[129] Thompson, R. E., Larson, D. R., and Webb, W. W. Precise nanometer localization analysis for individual fluorescent probes. *Biophysical Journal 82*, 5 (2002), 2775–2783.

[130] Tkačik, G., and Bialek, W. Information processing in living systems. 89–117.

[131] Tokunaga, M., Imamoto, N., and Sakata-Sogawa, K. Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nature Methods 5*, 2 (feb 2008), 159–161.

[132] Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N. J., Livak, K. J., Mikkelsen, T. S., and Rinn, J. L. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature biotechnology 32*, 4 (apr 2014), 381–6.

[133] Vera, M., Biswas, J., Senecal, A., Singer, R. H., and Park, H. Y. Single-Cell and Single-Molecule Analysis of Gene Expression Regulation. *Annual review of genetics 50* (nov 2016), 267–291.

[134] Vernimmen, D., and Bickmore, W. A. The Hierarchy of Transcriptional Activation: From Enhancer to Promoter. *Trends in Genetics 31*, 12 (dec 2015), 696–708.

[135] Vita, M., and Henriksson, M. The Myc oncoprotein as a therapeutic target for human cancer, aug 2006.

[136] Wan, Y., Anastasakis, D. G., Rodriguez, J., Palangat, M., Gudla, P., Zaki, G., Tandon, M., Pegoraro, G., Chow, C. C., Hafner, M., and Larson, D. R. Dynamic Imaging of Nascent RNA Reveals General Principles of Transcription Dynamics And Stochastic Splice Site Selection. *SSRN Electronic Journal* (oct 2019).

[137] Wang, H., Vilela, M., Winkler, A., Tarnawski, M., Schlichting, I., Yumerefendi, H., Kuhlman, B., Liu, R., Danuser, G., and Hahn, K. M. LOVTRAP: An optogenetic system for photoinduced protein dissociation. *Nature Methods 13*, 9 (aug 2016), 755–758.

[138] Wang, Z., Gerstein, M., and Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics 10*, 1 (jan 2009), 57–63.

[139] Wei, Y., Resetca, D., Li, Z., Johansson-Åkhe, I., Ahlner, A., Helander, S., Wallenhammar, A., Morad, V., Raught, B., Wallner, B., Kokubo, T., Tong, Y., Penn, L. Z., and Sunnerhagen, M. Multiple direct interactions of TBP with the MYC oncoprotein. *Nature Structural and Molecular Biology 26*, 11 (nov 2019), 1035–1043.

[140] Weischenfeldt, J., Dubash, T., Drainas, A. P., Mardin, B. R., Chen, Y., Stütz, A. M., Waszak, S. M., Bosco, G., Halvorsen, A. R., Raeder, B., Efthymiopoulos, T., Erkek, S., Siegl, C., Brenner, H., Brustugun, O. T., Dieter, S. M., Northcott, P. A., Petersen, I., Pfister, S. M., Schneider, M., Solberg, S. K., Thunissen, E., Weichert, W., Zichner, T., Thomas, R., Peifer, M., Helland, A., Ball, C. R., Jechlinger, M., Sotillo, R., Glimm, H., and Korbel, J. O. Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. *Nature Genetics 49*, 1 (nov 2016), 65–74.

[141] Wilson, M. Z., Ravindran, P. T., Lim, W. A., and Toettcher, J. E. Tracing Information Flow from Erk to Target Gene Induction Reveals Mechanisms of Dynamic and Combinatorial Control. *Molecular cell 67*, 5 (sep 2017), 757–769.e5.

[142] Wu, Y. I., Frey, D., Lungu, O. I., Jaehrig, A., Schlichting, I., Kuhlman, B., and Hahn, K. M. A genetically encoded photoactivatable Rac controls the motility of living cells. *Nature 461*, 7260 (sep 2009), 104–108.

[143] Zenklusen, D., Larson, D. R., and Singer, R. H. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nature Structural & Molecular Biology 15*, 12 (2008), 1263–1271.