

DISSERTATION

LIVE-CELL IMAGING UNCOVERS THE RELATIONSHIP BETWEEN HISTONE
ACETYLATION, RNA POLYMERASE II PHOSPHORYLATION, TRANSCRIPTION, AND
CHROMATIN DYNAMICS

Submitted by

Matthew Neeley Saxton

Department of Biochemistry and Molecular Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2023

Doctoral Committee:

Advisor: Timothy J. Stasevich

Erin O. Nishimura

Jeffrey Hansen

Diego Krapf

Copyright by Matthew N. Saxton 2023

All Rights Reserved

ABSTRACT

LIVE-CELL IMAGING UNCOVERS THE RELATIONSHIP BETWEEN HISTONE ACETYLATION, RNA POLYMERASE II PHOSPHORYLATION, TRANSCRIPTION, AND CHROMATIN DYNAMICS

Living cells are capable of turning a one dimensional strand of nucleic acids into a functional polypeptide. A host of steps and factors are involved in the process of transcription and translation, and understanding each of them is necessary for comprehending and characterizing life. While new technologies and assays have expanded our understanding of eukaryotic transcription, there is still much to be learned. In particular, single-molecule microscopy provides a powerful and versatile platform for studying the genesis of RNA with unparalleled spatiotemporal resolution (Chapter 1).

First, we characterize the timing, kinetics, and occupancy of phosphorylated RNA polymerase II (RNAP2) using a single-copy HIV-1 reporter system. This work provides strong evidence for clusters of phosphorylated, initiating RNAP2 which is spatially separated from bursty, downstream RNA synthesis. It is found that RNAP2-Ser5-phosphorylation (Ser5ph) precedes RNA output by 1 minute, and RNAP2 arrives at the locus in a phosphorylated state (Chapter 2). Then, we examine the spatial correlation between H3K27 acetylation and Ser5ph in living cells on the course of minutes to hours. Contrary to expectations based upon ChIP data, we find that the two signals are in fact spatially separated. This argues for a functional separation between transcriptional poising and initiation, likely aiding bursty behavior. Next, the dynamics of single chromatin-incorporated nucleosomes in the context of H3K27 acetylation and transcription initiation is determined with super-resolution single-molecule imaging. The physical movement of chromatin inside of H3K27ac and RNAP2-Ser5ph enriched regions is found to be significantly different, despite both marks being traditionally associated with transcriptionally active chromatin. (Chapter 3).

Much of this work utilizes bead-loading in order to introduce proteins and DNA into living cells. A simple, effective, and cheap procedure, bead-loading is a highly effective and versatile technique that is generally underutilized. To facilitate communication of this process, a detailed protocol is included (Chapter 4). While this culmination of work furthers our understanding of cellular genetic expression and eukaryotic transcription, it also introduces many new questions that are promising areas of study. Fortunately, the combination of imaging technology and knowledge developed here provides promising new fronts for studying transcription in living cells (Chapter 5).

ACKNOWLEDGEMENTS

First and foremost, I need to thank my mentor and PI, Dr. Timothy Stasevich. Throughout this long process we have always had a functional and productive working relationship. Your mentorship helped me grow immensely as a scientist, and I always felt heard and respected when we were debating projects, data, and interpretations. Furthermore, I always enjoyed the time we spent in a non-professional context. In particular, I will always cherish the time we spent together in Japan. I want to thank the members of my committee, Dr. Erin Nishimura, Dr. Jeff Hansen, and Dr. Diego Krapf. You all managed to be consistently supportive of my work while providing critical feedback. I enjoyed each and every one of our meetings, which is a rare occurrence for most graduate students. A special thanks to committee member Diego Krapf, whose willingness to lend his support and knowledge led to a fruitful collaboration which greatly increased the rigour of our work.

I next need to thank all members of the Stasevich Lab, both past and present. Whether it was discussing data, experiments, or asking questions about protocols, I could not have done this without everyone's support. Kenneth Lyon, you were always made me laugh, whether through witty humor or self-deprecation, and I enjoyed all of our conversations. Amanda Koch, thanks for always being there when I needed an ear (or a beer), and for being a good friend. Charlotte Cialek, thanks for being a great desk neighbor and always up for a chat, I always appreciated our ability to have frank and productive conversations, and look forward to our continued friendship. Tatsuya Morisaki, thanks for always providing sage counsel and listening when I needed advice. I always greatly valued your critical mindset and input on experimental, technical, and scientific fronts. Linda Forero-Quintero, I'm glad that we were able to work together on such a nice manuscript, and I'm happy to have been a part of your career. I want to thank O'Neil Wiggan for his technical advice and for always being an engaged lab member who provided valuable input and asked probing questions. I also want to thank lab members Gabe Galindo, Adam Koch, Ning Zhao, Phil Fox, Lara Perinet, Sam Klink and Gretchen Fixen for their friendship and support. I also want to thank

the entirety of the BMB department at CSU. You all were wonderful, and the sense of community made a hard process more bearable.

I also want to thank my mother, Gaile, for her constant encouragement. You always saw the best in me, and you were never shy to tell me. Sometimes that was just what I needed. To my brother Jon, you were always there for me growing up, and were always a wonderful, kind brother. I want to thank my lovely partner, Katie Bonack, for her unwavering support. You saw me through the entirety of graduate school, and I can't have asked for more a loving, kind individual to have spent the time with. You always provided the best escape when the going was tough, and always made me feel heard. I can't thank you enough for everything you've done, and I look forward to our future together.

DEDICATION

This thesis is dedicated to my father, Steven L Saxton. You never even saw me finish my undergraduate degree, but you taught me to believe in myself, and I'll carry that with me forever.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGEMENTS	iv
DEDICATION	vi
Chapter 1 Background	1
1.1 Cellular genetic expression	1
1.2 Regulation of transcription	1
1.3 Chromatin and transcription	3
1.4 Post-translational protein modifications (PTMs)	4
1.4.1 Histone PTMs	4
1.4.2 RNAP2 c-terminal domain PTMs	5
1.5 Single-molecule microscopy as a tool to study gene regulation	6
Bibliography	8
Chapter 2 Live-cell imaging reveals the spatiotemporal organization of endogenous RNA polymerase II phosphorylation at a single gene	11
2.1 Summary	11
2.2 Introduction	12
2.3 Results	14
2.3.1 Technology to visualize endogenous RNAP2 transcription cycle dy- namics at a single gene.	14
2.3.2 Long-term imaging of fluctuations at the reporter gene reveals temporal ordering of RNAP2 phosphorylation.	19
2.3.3 Spatial organization of CTD phosphorylation at the reporter gene.	22
2.3.4 Fluctuation dynamics and statistics are captured by a simple model of transcription bursting.	24
2.3.5 Inhibiting distinct steps of the transcription cycle provides further evi- dence for spatiotemporal organization of RNAP2 phosphorylation.	32
2.4 Discussion	39
2.5 Methods	43
Bibliography	59
Chapter 3 The relationship between histone acetylation, transcription initiation, and chromatin dynamics, imaged in living-cells	68
3.1 Summary	68
3.2 Introduction	69
3.3 Results	72
3.3.1 Monitoring global chromatin dynamics in the context of histone acety- lation and RNAP2 phosphorylation	72

3.3.2	Tracking histone acetylation and RNAP2 phosphorylation dynamics at thousands of endogenous chromatin sites reveals two distinct dynamics .	77
3.3.3	Single Particle Tracking of H2B at Transcriptionally Enriched Sites Reveals Chromatin Slowdown	82
3.3.4	Enrichment of H3K27ac Predicts an Increased Nucleosome Diffusion .	89
3.4	Discussion	91
3.5	Materials and Methods	95
Bibliography		104
Chapter 4 Bead loading proteins and nucleic acids into adherent human cells 113		
4.1	Summary	113
4.2	How to introduce plasmids and proteins into cells	113
4.3	Bead loading proteins and plasmids into adherent human cells	115
4.4	Bead loading and cell health	119
4.5	Loading different cell types and materials via bead loading	119
4.6	Bead loading of live, adherent cells is inexpensive and versatile	122
4.7	Protocol	124
4.7.1	Clean, sterilize and dry glass beads so that they do not clump and can spread evenly atop cells.	124
4.7.2	Assemble the bead loader apparatus.	124
4.7.3	Prepare glass-bottom chambers of adherent cells	125
4.7.4	Bead loading cells	125
4.7.5	Imaging the bead loaded cells	127
Bibliography		128
Chapter 5 Final Conclusions and Future Directions 132		
5.1	The masking process: applications and considerations	132
5.1.1	Masking as a methodology	132
5.1.2	Histone modifications and masking	134
5.1.3	Other applications and considerations	135

Chapter 1

Background

1.1 Cellular genetic expression

The oft repeated mantra of the biological central dogma - DNA makes RNA makes Protein presents a simplistic view of how life functions. In reality, gene expression is a complex process with many unique environmental and interactional influences on final product and function. Macroscopic cellular decisions are thereby determined by small, molecular events. If regulation is disrupted in excess, deleterious effects lead to disease states. The stochastic and variable nature of genetic expression means that each protein experiences a unique genesis and understanding this process from start to finish would allow us to pinpoint the moment an error occurs.

In eukaryotic cells, the pathway from DNA to protein involves many steps, each providing the opportunity for regulatory effects and genetic control. A cell's archival copy of DNA is first transcribed by RNA polymerase into various forms of RNA, termed 'transcription'. Relatively short-lived copies, termed messenger RNA or mRNA, are then exported from the nucleus. In the cytoplasmic space mRNAs are engaged by ribosomes, massive ribonucleoprotein complexes which are responsible for translating the linear mRNA message into protein. Conjugated amino acids then fold and form a final protein, which diffuses into the cellular space to perform one (or many) cellular functions. Although each cell carries the same copy of genetic material, the protein content of each cell will vary considerably. Therefore, in the most explicit and direct way, regulating and controlling when, what, and how many RNA or proteins are expressed in a cellular environment determines behavior, phenotype, metabolism, or even cell death.

1.2 Regulation of transcription

Before a protein can be made, RNA must be transcribed. Transcription begins with the recognition of a specific DNA sequence by a host of transcription factors (TFs) [1]. These factors bind

to specific regions of DNA, known as promoters, which are located near the transcription start site (TSS) of a gene. Promoters play a crucial role in regulating gene expression, and their sequence and location can greatly impact the level of transcription [2]. Once TFs have bound to the promoter, they recruit RNA polymerase II (RNAP2), which is responsible for the synthesis of mRNA. RNAP2 binds to the promoter, and a series of events occur that lead to the formation of a transcription initiation complex. This complex includes RNAP2, TFs, and other regulatory proteins, and is responsible for the initial steps of mRNA synthesis [3]. After initiation, RNAP2 moves along the DNA template strand, synthesizing an RNA molecule that is complementary to the template strand. RNAP2 procedural synthesis is aided by elongation factors and chromatin remodeling complexes [4]. These factors help to ensure that the DNA template strand is properly unwound and that the RNA molecule is properly synthesized. Once RNAP2 has reached the end of the gene, it must be released from the DNA template strand. This process is known as termination and is controlled by specific sequences within the gene itself. These sequences signal to RNAP2 that it has reached the end of the gene, and the enzyme is released from the DNA template [5].

Additionally, transcription is coordinated by enhancers. Enhancers are regulatory DNA sequences that are often far away from a promoter on the linear genome, but interact with promoters in 3D space. Enhancers function by recruiting transcriptional activators and TFs. These TFs are often sequence specific and aid with catalyzing transcriptional activation and processivity. However, the exact mechanism and function of enhancer elements is still poorly understood. Some hypotheses posit that enhancers loop to interact with promoters, and the proximity of recruited proteins activates transcription. Similarly, enhancers may function by altering the local chromatin environment by recruiting chromatin remodeling complexes, increasing accessibility to TFs. Conversely, repressor proteins and insulating elements can serve to block enhancer interaction and thereby block gene activity. Enhancers are therefore another critical target for transcription regulation, and understanding their function and mechanism of action is a core component of characterizing gene control [6].

1.3 Chromatin and transcription

As hinted earlier, another critical element of gene expression is the structure and organization of genetic material. Linear DNA is wrapped around interspersed core histones to form a 10nm fiber. Early *in-vitro* studies proposed a model of hierarchical folding in which chromatin is progressively compacted into dense, rope-like structures. Active chromatin was unfolded to become loose and accessible. Inactive chromatin was hypothesized to be condensed into higher-order structures inaccessible to transcription machinery. This hypothesis has been contradicted as electron microscopy [7,8] and small-angle X-ray scattering [9,10] have failed to show a structure greater than 11nm in both mitotic and interphase chromosomes. These data have thereby shifted the nuclear chromosome model to 10nm fibers interdigitated into a polymer-melt [11]. Single nucleosome tracking experiments have provided further evidence that this melt is organized into 100-500nm chromatin domains [12]. This structure is more dynamic than classic helical folding and has important implications for cellular regulation of genetic expression.

In a simple way, the physical nature of chromatin structure may regulate transcription by modulating TF access. TFs have been found to have relatively short binding times *in-vivo*, and chromatin domains silence certain genes by physically restricting access to TFs (i.e. inactive X-locus) [13]. Given that TF binding is dynamic, it is important that TFs are able to find their target in three dimensional space. Highly static chromatin would not facilitate this target-searching function, and thus dynamism is typically a property associated with active euchromatin. Similarly, to facilitate transcription, nucleosomes must be removed from the genetic template by ATP-dependent remodeling factors, which again may further reinforce TF searching [14].

Unsurprisingly, chromatin domains exhibit varying levels of nucleosome density and mobility [12], yet the underlying mechanism defining these physical dynamics is unexplained. Chromatin behavior has yet to be foundationally associated with specific cellular processes, whether is be replication, transcription, or completely silent genes. While this is primarily due to experimental limitations, this thesis lays a groundwork for studying chromatin behavior in context that will be applicable to many facets of transcription.

1.4 Post-translational protein modifications (PTMs)

1.4.1 Histone PTMs

Histone proteins, which form the core nucleosome, have N-terminal "tails" which protrude from the core globule [15]. After being incorporated into chromatin, histone tails are targets for post-translational modifications (PTMs) such as acetylation, methylation or ubiquitination. In general, acetylation has been associated with transcription activation [16]. Conversely, methylation has been generally associated with transcription repression. However, these are not immutable rules. There is an entire suite of histone PTMs, each with its own host of associated proteins and factors. This has led to a "histone code" hypothesis. Generally, these modifications work in tandem to create a code for expression or repression, meaning that precise combinations of modifications allow for fine-tuning of transcription initiation and thus full genetic control [17]. Critics, however, go as far as to believe that histone PTMs serve relatively little function, and are simply the enzymatic byproduct of surrounding protein activity. The exact nature and mechanism of these functions and interactions is thus unclear, and an ongoing topic of intense study.

Chromatin structure is also sensitive to tail-modification changes. Histones are positively charged proteins, while DNA has a negatively charged phosphate backbone. While this helps chromatin form, this interaction can also be exploited [18]. In general, acetylation of histone tails has been shown to weaken the interaction between histones and DNA, leading to a more open chromatin structure. This theoretically increases the accessibility of genetic elements to transcriptional machinery and reinforces gene activation. In contrast, methylation of histones is associated with transcriptional repression and the formation of heterochromatic regions [19]. As previously mentioned, however, this is not a hard rule. Methylation of histone 3 at lysine 4 (H3K4me1) is associated with active promoters and enhancers [20], while H3K9me and H3K27me3 are associated with silent genes and the inactive-X locus [21].

Given the general intellectual quagmire created by having so many modifications, it is useful to first focus on them individually. One well studied and important modification is acetylation of histone 3 at lysine 27 (H3K27ac). H3K27ac is catalyzed by histone acetyltransferases (HATs)

and can be removed by histone deacetylases (HDACs). This modification is predominantly associated with active gene transcription and is found in promoter and enhancer regions of the genome. H3K27acetylation is also often used as a marker of active regulatory regions, and its levels are correlated with the level of gene expression. H3K27ac is also highly correlated with enhancer and super-enhancer regions. Thus, H3K27ac is thought to help facilitate the recruitment of transcriptional coactivators to enhancer regions and activate nearby genes [22]. Due to its role in transcription regulation, H3K27ac is also tightly regulated. Aberrant H3K27ac levels has been implicated in a variety of diseases, including cancer [23]. However, H3K27ac activity can be depleted in mouse embryonic stem cells with little loss in expression [24], and is not present at every active gene, so the exact nature of H3K27ac involvement in transcription activation is unclear.

1.4.2 RNAP2 c-terminal domain PTMs

RNAP2 is another target for post-translational modification. RNAP2 has a C-terminal domain (CTD) that contains a repetitive heptapeptide sequence (Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7), which serves as a platform for the recruitment of various TFs or RNA processing machinery²⁵. Two of the most important modifications occur at Serine 5 (RNAP2-Ser5ph) and Serine 2 (RNAP2-Ser2ph). Ser5ph is primarily added by CDK7, a subunit of general transcription factor TFIIF. This serves to disrupt the interactions between the CTD and mediator complex, helping assist in promoter escape [25]. Thus Ser5ph is an excellent general marker of transcription initiation, and is typically found near promoter and enhancer elements via ChIP [26].

Ser2ph, on the other hand, is primarily carried out by the cyclin-dependent kinase 9 (CDK9) of the transcription elongation factor b complex (P-TEFb). Ser2ph has also been shown to recruit RNA processing enzymes, and is thought to be important in releasing RNAP2 from promoter-proximal pausing. Ser2ph thus accumulates as transcription progresses, and is a marker for productive RNAP2 elongation [25]. However, is it still unclear the degree to which Ser5ph and Ser2ph are separated in time and space during actual transcription at a single-gene level.

1.5 Single-molecule microscopy as a tool to study gene regulation

Given the complex and multifaceted nature of genetic expression, it can be difficult to effectively study *in-vivo*. Often data is gathered from experiments relying upon the aggregation and averaging of many events, from thousands or even millions of cells. This means that the inherent dynamics of individual events are lost, and these averages may not reflect what is occurring in any one cell or at any one locus. This principle is also true in traditional fluorescence microscopy, in which dyes and fluorescent proteins are used to gather spatial information within a cell or cellular population. Fortunately, however, advancements in dye technology, optical engineering, and sensors have paved the way for single-molecule microscopy.

In single-molecule microscopy, individual molecules can be observed and tracked with incredible spatiotemporal resolution, which has numerous advantages. First is high spatial resolution. While conventional microscopy's resolution is limited by the diffraction of light, single molecule microscopy detects individual fluorophores. This means that these point sources can be localized with high precision, with resolution in the 10-20nm range [27]. Single-molecule microscopy also allows for real-time observation of dynamic processes. Many biological processes happen on very fast time scales, and by tracking individual molecules, it is possible to capture the breadth and heterogeneity of individual events. Lastly, the combination of spectral channels and fluorescent labelling allows for multiplexed detection. This means that multiple events and molecules can be tracked simultaneously, and their behaviors related to one another. Developments such as MS2-MCP [28], PP7 coat protein [29], and HaloTag [30] combined with JaneliaFluor [31] and AlexaFluor dyes have allowed researchers to effectively tag many different types of RNA and protein with single-molecule resolution inside of living cells, leading to a breadth of new knowledge.

While tagging and tracking proteins with single-molecule resolution is fairly straightforward, the same is not true of PTMs. PTMs are often dynamic, and cannot be tagged by conjugating them to a fluorescent protein. Ultimately, PTMs have been challenging to study at single-molecule

resolution due to the difficulty of labeling small, chemical modifications. Thus, the spatiotemporal dynamics of transcription activity and chromatin behavior associated with histone PTMs and RNAP2 PTMs remains uncharacterized.

However, this limitation can be circumvented by using fragmented antigen binding regions of antibodies (Fab). While antibodies are unsuitable for most live-cell imaging applications, the Fc regions can be digested and separated, leaving only the antigen binding component, eliminating many of the issues that traditional antibodies present. This small polypeptide can be conjugated to a fluorescent dye, and then introduced into living cells where it will freely diffuse and bind transiently but specifically to various endogenous post-translational modifications [32]. The following thesis work thus leverages the power of single-molecule microscopy and Fab-based labeling to characterize the dynamics of transcription and chromatin with unprecedented spatiotemporal resolution, further uncovering how living cells regulate and control selective genetic expression.

Bibliography

- [1] G. Orphanides, T. Lagrange, and D. Reinberg. The general transcription factors of rna polymerase ii. *Genes Dev*, 10:26572683.
- [2] V. Haberle and A. Stark. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat. Rev. Mol. Cell Biol*, 19:621637.
- [3] Y. He, J. Fang, D.J. Taatjes, and E. Nogales. Structural visualization of key steps in human transcription initiation. *Nature*, 495:481486.
- [4] Y. Yamaguchi. Structure and function of the human transcription elongation factor dsif *. *J. Biol. Chem*, 274:80858092.
- [5] J.B. Rodríguez-Molina, S. West, and L.A. Passmore. Knowing when to stop: Transcription termination on protein-coding genes by eukaryotic rnapii. *Mol. Cell*, 83:404415.
- [6] C.-T. Ong and V.G. Corces. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat. Rev. Genet*, 12:283293.
- [7] C. Bouchet-Marquis, J. Dubochet, and S. Fakan. Cryoelectron microscopy of vitrified sections: a new challenge for the analysis of functional nuclear architecture. *Histochem. Cell Biol*, 125:4351.
- [8] M. Eltsov, K.M. MacLellan, K. Maeshima, A.S. Frangakis, and J. Dubochet. Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ. *Proc. Natl. Acad. Sci*, 105:1973219737.
- [9] Y. Joti. Chromosomes without a 30-nm chromatin fiber. *Nucleus*, 3:404410.
- [10] Y. Nishino. Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. *EMBO J*, 31:16441653.

- [11] K. Maeshima, R. Imai, S. Tamura, and T. Nozaki. Chromatin as dynamic 10-nm fibers. *Chromosoma*, 123:225237.
- [12] T. Nozaki. Dynamic organization of chromatin domains revealed by super-resolution live-cell imaging. *Mol. Cell*, 67:282293 7.
- [13] T. Morisaki, W.G. Müller, N. Golob, D. Mazza, and J.G. McNally. Single-molecule analysis of transcription factor binding at transcription sites in live cells. *Nat. Commun*, 5:4456.
- [14] C.R. Clapier and B.R. Cairns. The biology of chromatin remodeling complexes. *Annu. Rev. Biochem*, 78:273304.
- [15] K. Luger, A.W. Mäder, R.K. Richmond, D.F. Sargent, and T.J. Richmond. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389:251260.
- [16] Y. Zhao and B.A. Garcia. Comprehensive catalog of currently documented histone modifications. *Cold Spring Harb. Perspect. Biol*, 7:025064.
- [17] T. Jenuwein and C.D. Allis. Translating the histone code. *Science*, 293:10741080.
- [18] R. Zhang, J. Erler, and J. Langowski. Histone acetylation regulates chromatin accessibility: Role of h4k16 in inter-nucleosome interaction. *Biophys. J*, 112:450459.
- [19] S.I. Grewal and J.C. Rice. Regulation of heterochromatin by histone methylation and small rnas. *Curr. Opin. Cell Biol*, 16:230238.
- [20] A. Local. Identification of h3k4me1-associated proteins at mammalian enhancers. *Nat. Genet*, 50:7382.
- [21] A.-V. Gendrel. Smchd1-dependent and -independent pathways determine developmental dynamics of cpg island methylation on the inactive x chromosome. *Dev. Cell*, 23:265279.
- [22] Y. Kang, Y.W. Kim, J. Kang, and A. Kim. Histone h3k4me1 and h3k27ac play roles in nucleosome eviction and ena transcription, respectively, at enhancers. *FASEB J*, 35:21781.

- [23] D. Sengupta. Disruption of brd4 at h3k27ac-enriched enhancer region correlates with decreased c-myc expression in merkel cell carcinoma. *Epigenetics*, 10:460466.
- [24] T. Zhang, Z. Zhang, Q. Dong, J. Xiong, and B. Zhu. Histone h3k27 acetylation is dispensable for enhancer activity in mouse embryonic stem cells. *Genome Biol*, 21:45.
- [25] M. Heidemann, C. Hintermair, K. VoSS, and D. Eick. Dynamic phosphorylation patterns of rna polymerase ii ctd during transcription. *Biochim. Biophys. Acta BBA - Gene Regul. Mech*, 1829:5562.
- [26] T.J. Stasevich. Regulation of rna polymerase ii activation by histone acetylation in single living cells. *Nature*, 516:272275.
- [27] M. Lelek. Single-molecule localization microscopy. *Nat. Rev. Methods Primer*, 1:127.
- [28] E. Bertrand. Localization of ash1 mrna particles in living yeast. *Mol. Cell*, 2:437445.
- [29] N. Katz. An in vivo binding assay for rna-binding proteins based on repression of a reporter gene. *ACS Synth. Biol*, 7:27652774.
- [30] G.V. Los. Halotag: A novel protein labeling technology for cell imaging and protein analysis. *ACS Chem. Biol*, 3:373382.
- [31] J.B. Grimm, T.A. Brown, B.P. English, T. Lionnet, and L.D. Lavis. Synthesis of janelia fluor halotag and snap-tag ligands and their use in cellular imaging experiments. In *Super-Resolution Microscopy: Methods and Protocols*. ed. Erfle, H.) 179188 (Springer, 2017).
- [32] Y. Hayashi-Takanaka. Tracking epigenetic histone modifications in single cells using fab-based live endogenous modification labeling. *Nucleic Acids Res*, 39:64756488.

Chapter 2

Live-cell imaging reveals the spatiotemporal organization of endogenous RNA polymerase II phosphorylation at a single gene

2.1 Summary

The carboxyl-terminal domain of RNA polymerase II (RNAP2) is phosphorylated during transcription in eukaryotic cells. While residue-specific phosphorylation has been mapped with exquisite spatial resolution along the 1D genome in a population of fixed cells using immunoprecipitation-based assays, the timing, kinetics, and spatial organization of phosphorylation along a single-copy gene have not yet been measured in living cells. Here, we achieve this by combining multi-color, single-molecule microscopy with fluorescent antibody-based probes that specifically bind to different phosphorylated forms of endogenous RNAP2 in living cells. Applying this methodology to a single-copy HIV-1 reporter gene provides live-cell evidence for heterogeneity in the distribution of RNAP2 along the length of the gene as well as Serine 5 phosphorylated RNAP2 clusters that remain separated in both space and time from nascent mRNA synthesis. Computational models determine that 5 to 40 RNAP2 cluster around the promoter during a typical transcriptional burst, with most phosphorylated at Serine 5 within 6 seconds of arrival and roughly half escaping the promoter in ~ 1.5 minutes. Taken together, our data provide live-cell support for the notion of effi-

cient transcription clusters that transiently form around promoters and contain high concentrations of RNAP2 phosphorylated at Serine 5.¹

2.2 Introduction

In eukaryotic cells, the catalytic RPB1 subunit of RNA polymerase II (RNAP2) possesses an extended carboxy terminal domain (CTD) that consists of heptapeptide repeats (52 in humans) with a consensus sequence (Tyr₁-Ser₂-Pro₃-Thr₄-Ser₅-Pro₆-Ser₇). The CTD region is dynamically phosphorylated as RNAP2 progresses through the transcription cycle, regulating each step of transcription, from initiation to termination. In some models, RNAP2 is recruited to promoters in an unphosphorylated form (CTD-RNAP2), but is later phosphorylated at Serine 5 (Ser5ph-RNAP2) upon initiation and at Serine 2 (Ser2ph-RNAP2) during active elongation [1–4]. Interest in the CTD has recently increased due to observations of highly dynamic RNAP2 clustering [4–6] that correlates with the phosphorylation status of the CTD [7, 8]. In particular, recent data suggest that a transcriptional cluster forms around gene promoters early in the transcription cycle. The cluster is thought to be enriched in unphosphorylated- and Ser5ph-RNAP2 that appear to constrain chromatin movement near the transcription start site [9]. However, upon transcriptional activation, hyperphosphorylation of RNAP2 at Ser2 allows the enzyme to escape the cluster and begin active elongation [7, 9]. The dynamic clustering of RNAP2 involves many steps and a complex orchestration of multiple factors and could therefore represent a global form of transcriptional regulation [10].

RNAP2 phosphorylation throughout the transcription cycle has traditionally been studied in fixed cells using immunoprecipitation-based assays [1, 3, 11]. These studies provide precise spatial maps of the average positions of RNAP2 along the 1D genome. Unfortunately, the inherent av-

¹This chapter is adapted from published work with the same title in which I am a co-author. Forero-Quintero et al. *Nat Commun* 12, 3158 (2021). <https://doi.org/10.1038/s41467-021-23417-0>. Author Contributions: Conceptualization, LSFQ, BM, and TJS; Performed experiments/collected data, LSFQ, TH, and MS; antibodies, HK and TH; Fab preparation, TJS, LSFQ, and HK; H-128 cells, EB; Software implementation and development, LSFQ, WR, TM, BM, and TJS; Formal analysis, LSFQ; Computational modeling, WR and BM; Wrote the original draft, LSFQ, WR, BM, and TJS; Review and edit drafts, LSFQ, WR, TH, MS, TM, EB, HK, BM, and TJS; Resources, Supervision and Funding acquisition, BM and TJS.

eraging masks heterogeneity and the procedure limits temporal resolution to timescales of tens of minutes or longer [12]. RNAP2 dynamics can instead be imaged and quantified in living cells using fluorescence microscopy, overcoming the limitations of traditional assays. Recent single-molecule tracking technologies [13–17] have made it possible to monitor single RNAP2 as they bind at non-specific locations throughout the genome [5, 18] as well as at specific, single-copy genes [6, 17] pre-marked with MS2 [19, 20] or PP7 [21] RNA stem loops (that are lit up co-transcriptionally when, respectively, fluorescent MS2 or PP7 coat proteins bind to them). Each of these studies used permanent fluorescent fusion tags to track RNAP2. Fusion tags are incapable of discerning post-translational modifications to RNAP2, including transcription cycle associated phosphorylation events.

One way to resolve post-translational modifications to RNAP2 is to use antibody-based probes that bind and light-up specific modifications to residues within the CTD *in vivo* [22–26]. However, the signal-to-noise is limited with this approach because of the presence of unbound and freely diffusing probes that increase the fluorescence background. Applications have therefore been restricted to large tandem gene arrays. Signal-to-noise is amplified by the multiple copies of a gene within these arrays, but heterogeneity from one gene copy to another is again masked by averaging [27]. Therefore, the spatiotemporal dynamics of RNAP2 phosphorylation at single-copy genes remain unclear.

Here we combine multi-color single-molecule microscopy, complementary fluorescent antibody-based probes, and rigorous computational modeling to visualize, quantify, and predict endogenous RNAP2 phosphorylation dynamics at a single-copy reporter gene in living cells. This unique combination of technologies allows us to directly visualize the temporal ordering and spatial organization of RNAP2 phosphorylation and mRNA synthesis throughout the transcription cycle at the reporter gene. We find evidence for relatively high concentrations of RNAP2 near the beginning versus end of the gene that are both spatially and temporally separate from elongating RNAP2 and nascent mRNA synthesis. Collectively, our data provide live-cell support for the existence of

higher-order, phosphorylation-dependent transcriptional clusters that dynamically form and surround active genes throughout the transcription cycle.

2.3 Results

2.3.1 Technology to visualize endogenous RNAP2 transcription cycle dynamics at a single gene.

To visualize the spatiotemporal dynamics of endogenous RNAP2 phosphorylation at a single gene, we used an established HeLa cell line (H-128) harboring an MS2-tagged HIV-1 reporter gene and stably expressing both GFP-tagged MS2 coat protein (MCP) and an untagged HIV-1 trans-activator of transcription (Tat) [20]. We chose HIV-1 as our reporter gene because it is a prototypical model for RNAP2 phosphorylation [28]. The HIV-1 reporter is strongly active in our cell line due to persistent stimulation by Tat, producing a bright MCP signal that pinpoints the location of the transcription site and gauges its activity in real-time [20] (Fig 2.1a). Consistent with this strong signal, immunostaining experiments in fixed cells revealed the transcription site is highly enriched in RNAP2 and relatively depleted in histones and their epigenetic modifications (Fig. 2.2a-b). Chromatin immunoprecipitation (ChIP) experiments furthermore confirmed the presence of CTD-RNAP2 and its phosphorylated forms Ser5ph- and Ser2ph-RNAP2, respectively. In particular, we detected that CTD-RNAP2 and Ser5ph-RNAP2 signals are highest at the transcription start site, whereas Ser2ph-RNAP2 is highest towards the end of the gene (Fig 2.1b). However, because these data come from a population of fixed cells, whether the various forms of RNAP2 are present at the same time and place and whether or not they appear in a preferred order is difficult to extract from this assay.

To better characterize the spatiotemporal dynamics of single-cell RNAP2 modifications during transcription, we loaded fluorescent fragmented antibodies (Fab, generated from the same antibodies used in ChIP) [22, 29] recognizing (1) the CTD of RNAP2 (anti-CTD RNAP2) without or with residue-specific phosphorylations, and (2) heptad repeats within the CTD that are phosphorylated at Serine 5 (anti-Ser5ph RNAP2). These antibodies have previously been shown to be specific for

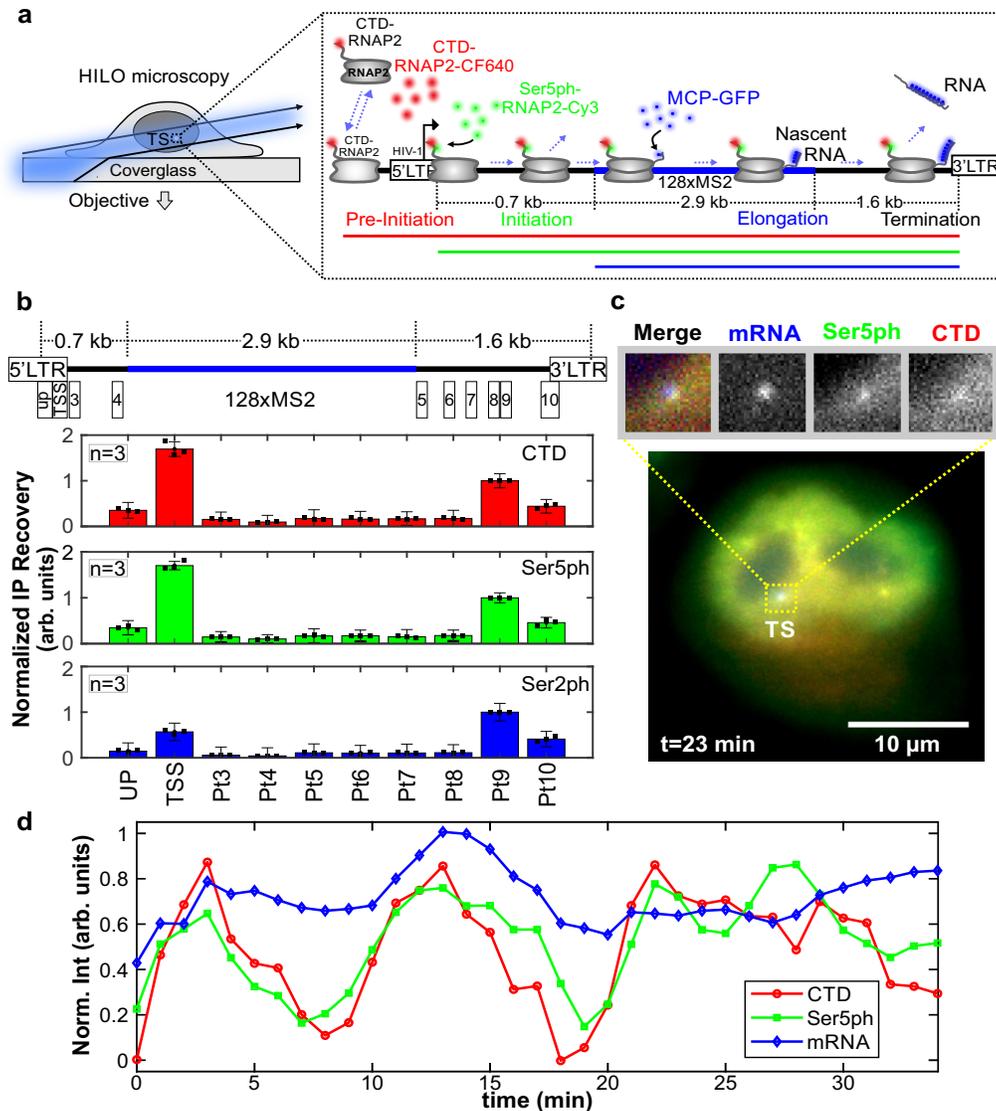


Fig. 2.1: A system for imaging the endogenous RNAP2 transcription cycle at single genes. (a) Schematic of the system. The reporter gene is controlled by the HIV-1 promoter and is tagged with a 128xMS2 cassette (blue bar). RNAP2 is represented in gray. RNA is marked by MCP-GFP that binds to the transcribed MS2 stem loops (mRNA, blue). The recruited and initiated RNAP2 are labeled by Fabs (conjugated with CF640 and Cy3) that bind unphosphorylated CTD RNAP2 heptad repeats (CTD, red) and Serine 5 phosphorylated repeats (Ser5ph, green), respectively. (b) Average chromatin immunoprecipitation occupancy of CTD-RNAP2 (red, upper panel), Ser5ph-RNAP2 (green, middle panel), and Ser2ph-RNAP2 (blue, lower panel) across the HIV-1 reporter gene (positions 1-10 are highlighted in the cartoon above). Data are presented as mean values \pm S.E.M. (c) Sample live cell showing CTD-RNAP2, Ser5ph-RNAP2, and mRNA co-localizing at the transcription site (TS), $n=9$ cells in 4 independent experiments. (d) Normalized intensity at the TS over time from the cell in (c) for CTD-RNAP2 (red circles), Ser5ph-RNAP2 (green squares), and mRNA (blue diamonds).

their respective targets via Western blotting and ELISA [27], and in ChIP-seq [30] experiments. Fab generated from these antibodies have also been shown to rapidly bind and unbind their targets, making them valuable for monitoring temporal changes in RNAP2 phosphorylation [27]. Con-

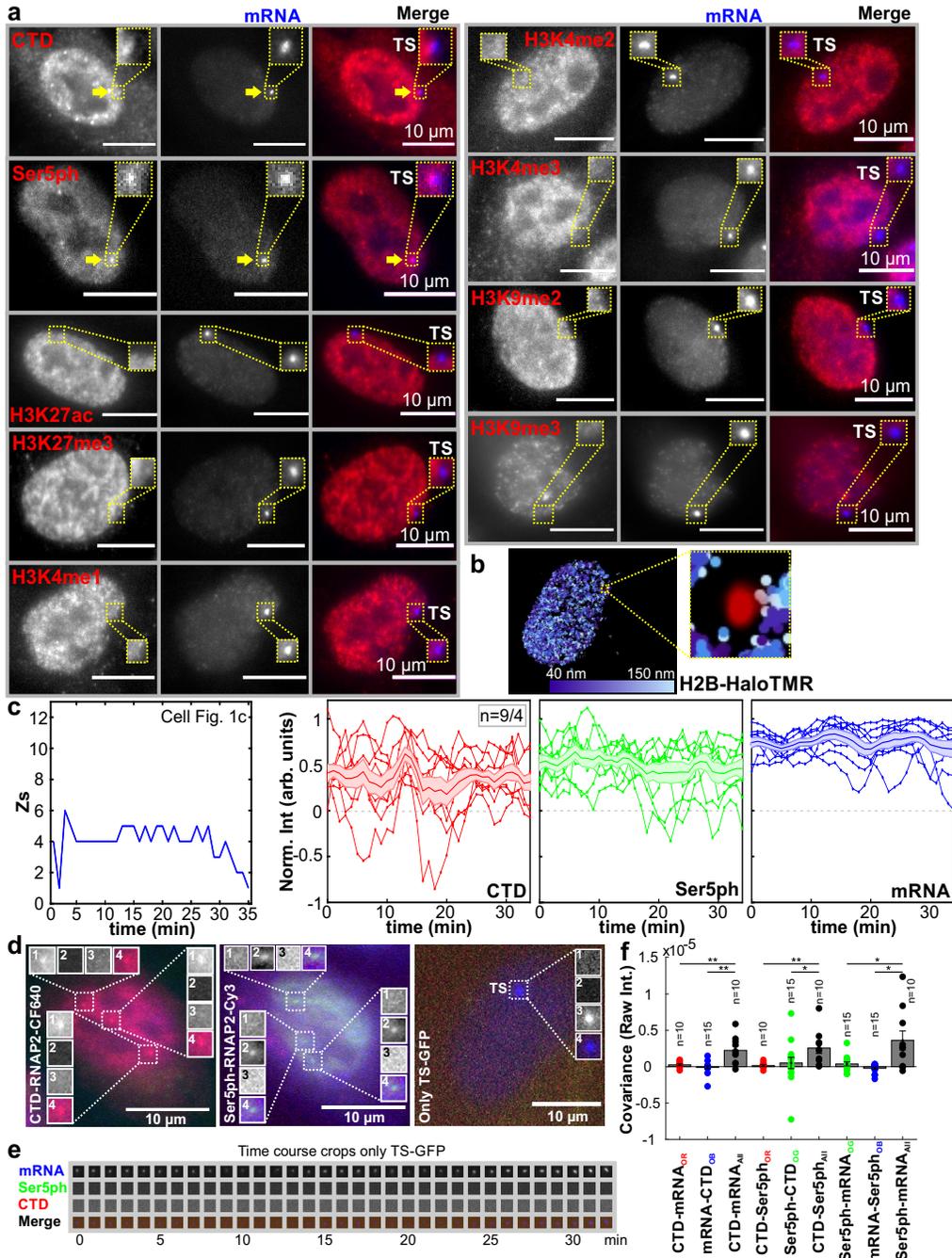


Fig. 2.2: Immunostaining, single H2B tracking, and control experiments for photobleaching and bleed through. (Continues next page.)

Fig. 2.2: (a) Immunostaining (red, left panels) of CTD-RNAP2 (n=11), Ser5ph-RNAP2 (n=10), histone H3K27ac (n=12), H3K27me3 (n=14), H3K4me1-3 (n=3) and H3K9me2-3 (n=3) at the HIV-1 transcription site (TS) marked by MCP-GFP (mRNA; blue, center panels), along with a merge (right panels). (b) Representative cell showing a mobility map of single H2B tracks. The blue scale shows the average frame-to-frame jump size (one frame every 43.86 ms) for each tracked molecule. The track corresponding to the transcription site is shown in red. The yellow dashed box displays a zoom-in around the transcription site region, where H2B is depleted (n=17 out of 28 cells in 3 total days). Control experiments for photo-bleaching showing (c) left panel, “best-Z” positions of the TS over time for the exemplary cell in Fig 2.1c; right panels, normalized intensity over time for CTD-RNAP2 (red circles), Ser5ph-RNAP2 (green squares), and mRNA (blue diamonds) for all the cells recorded as in Fig 2.1c,d. The shadow and the line in the middle represent the S.E.M and the average. (d) Images of cells from bleed-through control experiments. Left, a cell loaded with just Fab marking CTD-RNAP2 (CTD-RNAP2-CF640) displays endogenous puncta that are not the TS (designated Only Red “OR” spots; n=10); Middle, a cell loaded with just Fab marking Ser5ph-RNAP2 (Ser5ph-RNAP2-Cy3) displays endogenous puncta that are not the TS (designated Only-Green “OG” spots; n=15); Right, a cell without Fab in which the TS is marked solely by GFP-MCP binding mRNA (Only TS-GFP; Only Blue “OB”; n=15). Cropped images show the various “OR”, “OG”, and “OB” sites where the individual channels are separated and labeled as follows: (1) Red (CTD-RNAP2), (2) Green (Ser5ph-RNAP2), (3) Blue (mRNA), and (4) Merge. (e) Cropped images in a time course at an “OB” site demonstrates no bleed through of the mRNA channel into the other channels. (f) Covariance between all possible pairs of raw intensity signals is not significant at “OR”, “OG”, and “OB” sites, but is significant at the TS in cells containing all three signals (i.e. cells loaded with both Fab and expressing MCP-GFP; All). n=number of cells/independent experiments. Data are presented as mean values \pm S.E.M. Significance was tested using a two-tailed Mann-Whitney U-test with $p \leq 0.0136$ (*) and $p \leq 0.0091$ (**).

sistent with anti-CTD Fab labeling all RNAP2 and anti-Ser5ph Fab labeling a subset of RNAP2, we observed regions within the nucleus with RNAP2 enriched or depleted with Ser5ph (Fig 2.3). These capabilities allowed us to distinguish three distinct steps of the transcription cycle at the HIV-1 reporter gene: RNAP2 recruitment (marked by Fab against CTD-RNAP2), initiation (marked by both Fab against CTD-RNAP2 and Fab against Ser5ph-RNAP2), and elongation (marked by all Fab and MCP binding to mRNA), as depicted in Fig 2.1a,c. Although we attempted to also visualize Ser2ph at the locus with our Fab, signal-to-noise was insufficient to detect in living cells, presumably because the antibody is not sensitive enough to recognize this modification at the single-gene level.

Nevertheless, this setup has several advantages that collectively enhance signal-to-noise at the transcription site. First, Fab bind endogenous RNAP2, so all RNAP2 in the cell have high likelihood to be labeled without having to genetically engineer a fusion knock-in tag [18, 31] and/or alpha-amanatin resistance [5]. Second, fluorescence is naturally amplified since mammalian

RNAP2 contains 52 heptad repeats in its CTD [32], each of which can be bound by a fluorescent Fab at the transcription site. Third, Fab continually bind and unbind RNAP2, mitigating the loss of fluorescence due to local photobleaching. In combination with a multi-color, single-molecule microscope [33] employing oblique HILO illumination to enhance signal-to-noise by an order of

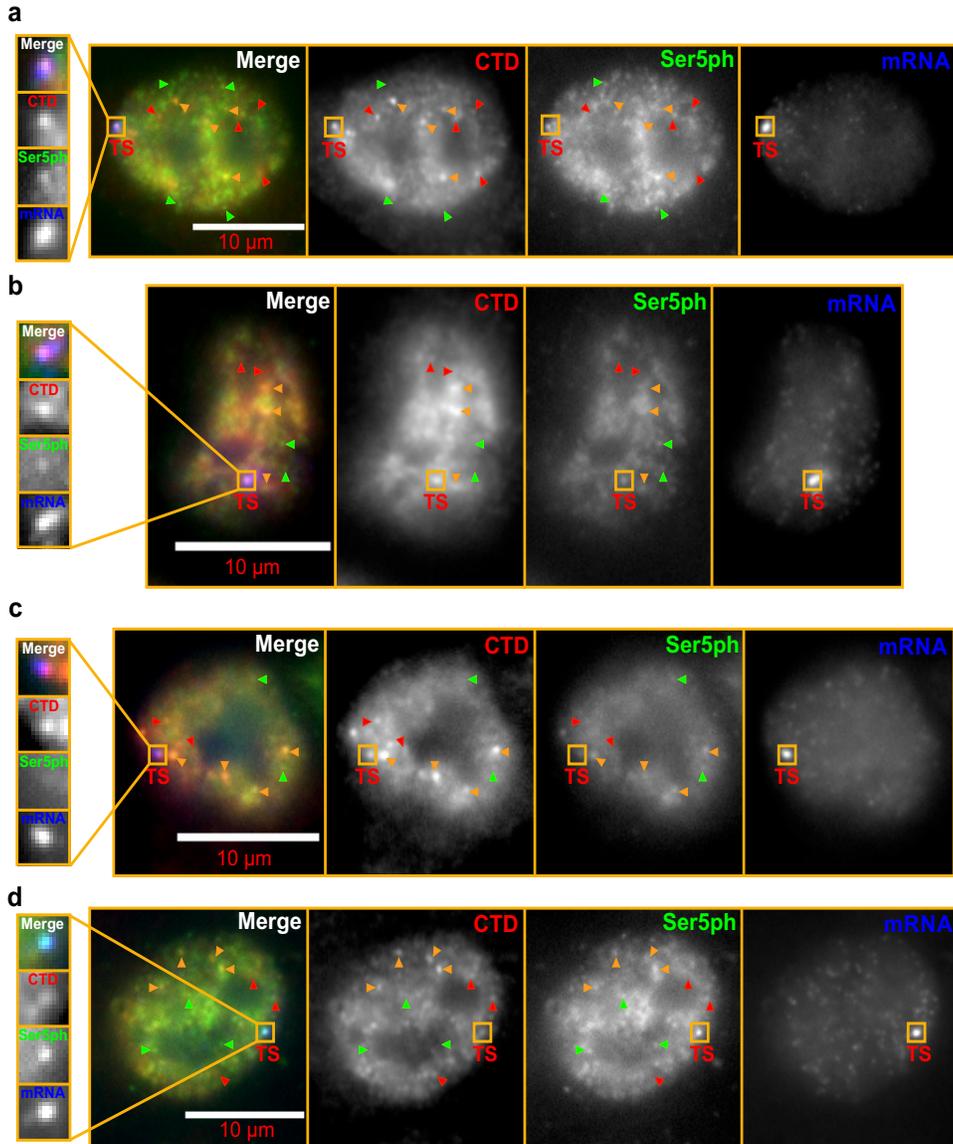


Fig. 2.3: Fixed cells stained with our CTD- and Ser5ph-specific Fab. Cells with distinct staining patterns. Some areas within cell nuclei are enriched with CTD-specific Fab (red arrows), other areas are enriched with Ser5ph-specific Fab (green arrows), while still other areas are enriched with both Fabs (orange arrows). At the HIV-1 transcription site (TS), we typically see both Fabs present (a), n=20 out of 29 cells. However, on occasion we can find TSs in which Ser5ph-RNAP2 staining is relatively dim (b & c), n=8 out of 29 cells, or, in very rare cases, CTD-RNAP2 staining is relatively dim (d), n=1 out of 29 cells. This provides evidence that the signals are fluctuating.

magnitude [13], these advantages allowed us to generate movies in which we monitored endogenous RNAP2 phosphorylation dynamics at the HIV-1 reporter gene in 3-colors.

As shown in Fig 2.1c-d, movies revealed correlated fluctuations between the mRNA signal and endogenous CTD-RNAP2 and Ser5ph-RNAP2 signals at the transcription site. To ensure correlations were not an artifact of focusing issues, we tracked the transcription site in 3D (by imaging 13 z planes per time point) to keep the MS2 signal continually in focus (Fig 2.2c, left panel). The correlations were also not caused by photobleaching, as signals fluctuated both up and down throughout the entire imaging time course, remaining on average constant (Fig 2.2c). Finally, to rule out the possibility that correlated fluctuations were caused by bleed through from one fluorescence channel to another, we re-imaged cells lacking Fab. In all cases, no bleed through was observed (Fig 2.2d-e), as quantified by the covariance between channels (Fig 2.2f). We therefore conclude the correlations reflect natural bursts in endogenous transcriptional activity at the HIV-1 reporter gene, demonstrating our ability to detect and quantify endogenous RNAP2 phosphorylation dynamics at a single-copy gene.

2.3.2 Long-term imaging of fluctuations at the reporter gene reveals temporal ordering of RNAP2 phosphorylation.

In the majority of cells, the mRNA was steadily produced by the HIV-1 reporter gene, with strong signals persisting for hours at a time. In a few cells the mRNA signal completely disappeared, indicating a loss of nearly all transcription activity. We were interested in capturing these rare events in a single time course to better discriminate the relative timing of our RNAP2 and mRNA signals. To accomplish this, we adjusted our imaging conditions to optimize detection of all three signals in single cells over a period of three hours (200 time points), as exemplified in Fig 2.4a-c (also see Fig 2.5a-c). We again imaged in z-stacks (13 planes spaced by 0.5 μm) covering the whole nucleus at each time point throughout the entire experiment. We were therefore confident that the fluctuations were due to changes in transcription activity and not related to transcription site movement into and out of the focal plane (Fig 2.5d). With these imaging conditions

we found cells in which the mRNA signal turned on and off up to four times, indicating bursts of transcription and multiple complete transcription cycles. Consistent with our previous result, signals at the transcription site were highly correlated and fluctuated generally in unison, although there were distinct periods of time when one signal could be seen for multiple frames in the absence of some other signals. This again ruled out bleed through and suggested the signals were not perfectly synchronized. To ensure the correlated fluctuations were specific to the locus and not cell-wide, we verified that covariances between the mRNA signal and the CTD or Ser5ph signals were significantly stronger when both signals were measured at the transcription site compared to when one or both signals were measured a short distance (p_1) from the transcription site (Fig 2.5e-g).

Having established a well-controlled system to examine fluctuations at a single gene, we were confident in our ability to quantify the temporal ordering of RNAP2 and mRNA throughout the transcription cycle. One thing that stood out was that peaks and troughs in the mRNA signal tended to come after the peaks and troughs in the RNAP2 signals. Although there were some exceptions due to the stochasticity of the system, in some cells this behavior was seen multiple times in even single time series (for example, see valleys at $t_{1-4} = 16, 75, 94$ and 113 min in Fig 2.4b-c). To better quantify this effect, we selected all events at which the mRNA signal dropped below a threshold value, extracted all three signal channels from seven minutes before to seven minutes after each event, and aligned all signals relative to these mRNA minima event times (Fig 2.4d). This analysis revealed two important aspects of the dynamics of our system. First, the analysis confirmed the signals were strongly correlated, since strong minima could be observed in all channels. These minima were significant compared to the results from unaligned signals (gray diamonds in Fig 2.4d, p-values of 1.72×10^{-10} for Ser5ph- and 6.39×10^{-4} for CTD-RNAP2). Such strong correlation between mRNA production at the HIV-1 reporter and endogenous RNAP2 would suggest the reporter is not part of a larger transcriptional unit containing multiple genes. Second, the analysis indicated a temporal ordering, with both RNAP2 signals coming before mRNA by 0.96 ± 0.55 min for CTD-RNAP2 (p-value 3.65×10^{-3}) and 0.88 ± 0.24 min for Ser5ph-RNAP2 (p-value

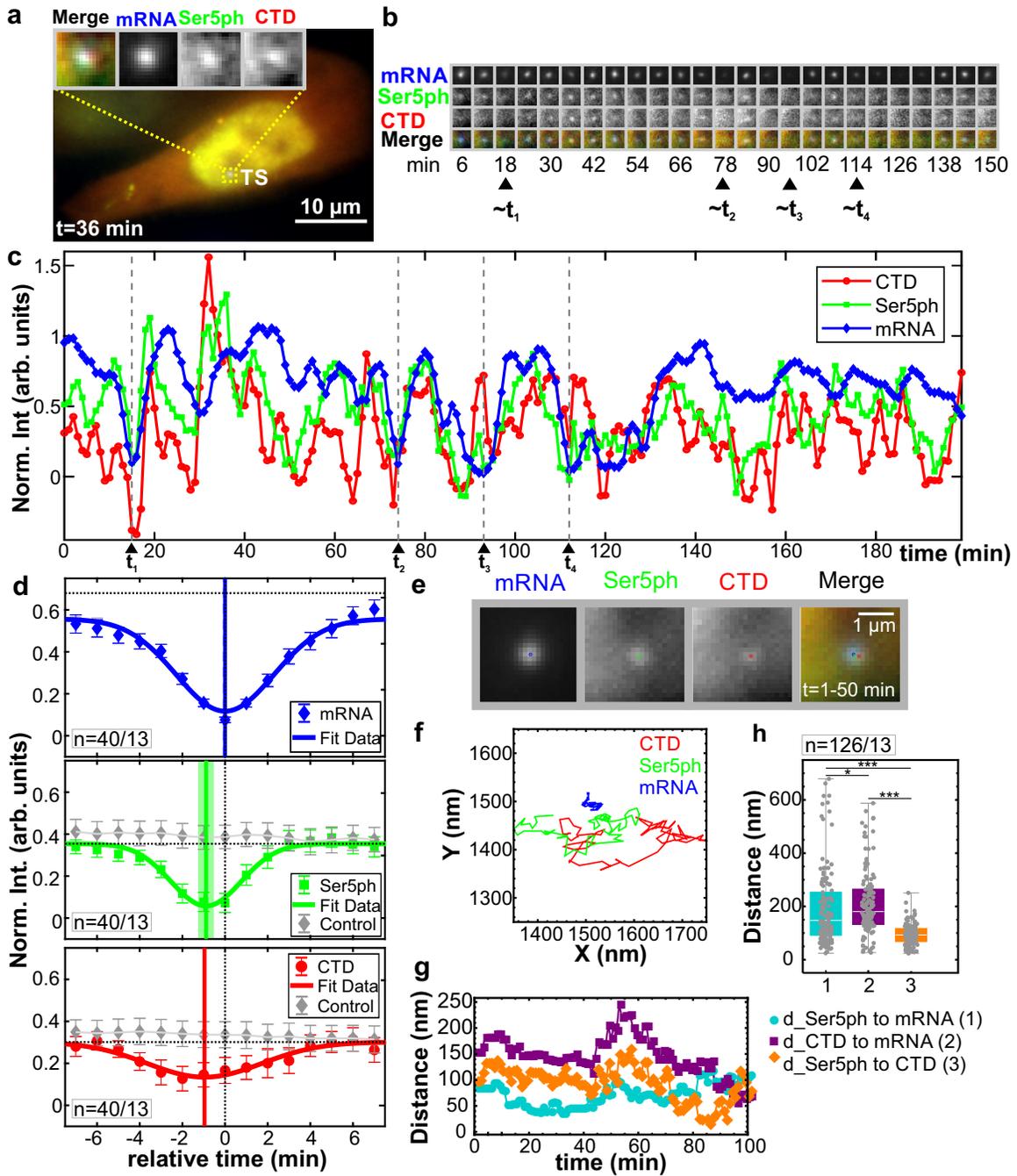


Fig. 2.4: Spatiotemporal organization of the RNAP2 CTD cycle at the HIV-1 reporter gene. (Caption continues next page)

Fig. 2.4: Spatiotemporal organization of the RNAP2 CTD cycle at the HIV-1 reporter gene. (a) Sample cell showing co-localization of CTD-RNAP2, Ser5ph-RNAP2, and mRNA signals at the transcription site (TS), n=13 out of 20 cells. (b) The TS from (a) at select times. (c) Normalized intensity fluctuations at the TS for CTD-RNAP2 (red circles), Ser5ph-RNAP2 (green squares), and mRNA (blue diamonds) versus time. Times of minimal mRNA (less than 0.20 a.u.) are marked with dashed gray lines (t_{1-4}). (d) The average normalized intensity of each signal surrounding times of minimal mRNA. Both the Ser5ph-RNAP2 and CTD-RNAP2 signals have deep minima well below the steady state value (dashed horizontal line). Solid connecting lines show the Gaussian fit, and solid vertical lines mark the minima with a lighter shadow depicting the S.E.M from the Gaussian fit. n= 40 events from 13 of 20 cells. Data are presented as mean values \pm S.E.M. When the same analysis is performed at 100 random time points, no obvious minima are seen (gray diamonds). (e) Cropped 50-frame (50 min total) moving-average image of the TS in (a) and the fitted center position for mRNA (blue), Ser5ph- (green), and CTD-RNAP2 (red), n= 126 events from 13 of 20 cells. (f) 50-frame moving-average XY position of each signal at the TS in (a) over time. Note the mRNA signal was used as the reference signal within the crop. (g) The distance between each signal in (f) over time: Ser5ph-RNAP2 to mRNA (cyan circles; (1)), CTD-RNAP2 to mRNA (Purple squares; (2)), and Ser5ph-RNAP2 to CTD-RNAP2 (orange diamonds; (3)). (h) The distribution of distances measured as in (g) at all TSs in all cells analyzed (sampled every 10 minutes). The line in the middle of each box represents the mean. The top and the bottom of the box represent the 75% and 25% quantiles, respectively. The middle region in the error bar at the bottom and the top represent the lower and upper whiskers, respectively. Significance was tested using a two-tailed Mann-Whitney U-test with $p \leq 0.0315$ (*) and $p \leq 6.969 \times 10^{-11}$ (***)).

1.28×10^{-5}). This delay makes sense because RNAP2 must escape the promoter and elongate 0.7 kb before it reaches the MS2 repeats. The CTD-RNAP2 signal also slightly preceded the Ser5ph-RNAP2 signal, although the delay was not significant at our sampling rate. This suggests nearly all RNAP2 at the locus either come in pre-phosphorylated or are rapidly phosphorylated at Serine 5 within a minute of arrival.

2.3.3 Spatial organization of CTD phosphorylation at the reporter gene.

RNAP2 is thought to be organized in phosphorylation-dependent clusters [7, 8]. To test this hypothesis, we measured the center position in X and Y of CTD-RNAP2, Ser5ph-RNAP2, and mRNA at the reporter gene over time (Fig 2.4e-g). If the hypothesis is correct, we would expect to see some spatial separation in our different RNAP2 and mRNA signals. To confirm this hypothesis, we calculated the Euclidean distance between each pair of signals. As Fig 2.4g illustrates, the distances between signals changed over time, but were spatially organized such that the RNAP2 signals were significantly separated from mRNA.

Although there was considerable variation from cell to cell, this trend could be seen in the median positions from the whole population of transcription sites we tracked (Fig 2.4h). Specifically, the median distance from mRNA to CTD-RNAP2 was ~ 181 nm compared to ~ 148 nm for Ser5ph-RNAP2 (p-value 0.032). Likewise, the median distance between the two forms of

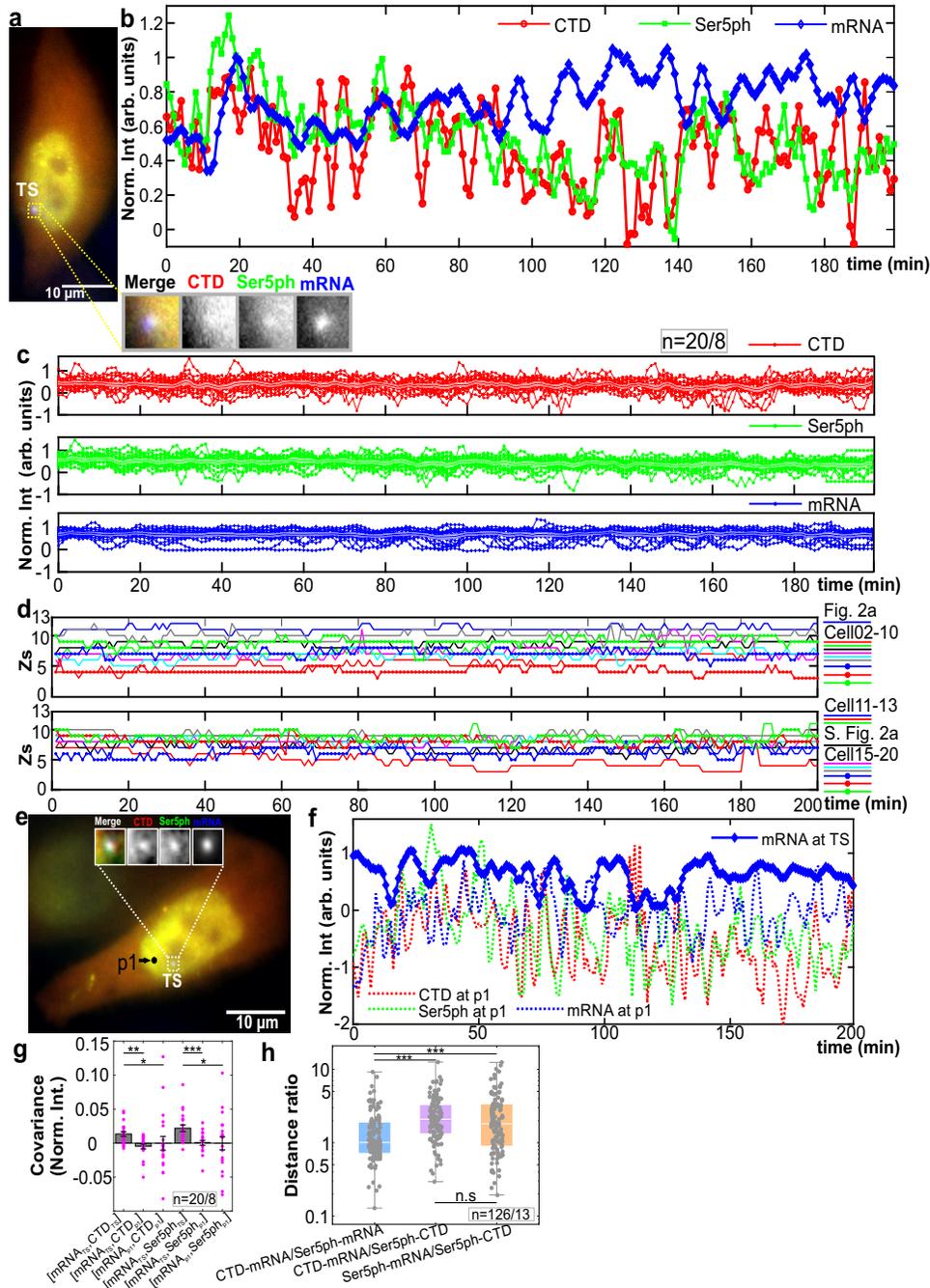


Fig. 2.5: RNAP2 fluctuations at the HIV-1 reporter locus and off target. (Continues next page.)

Fig. 2.5: (a,b) A cell with strong and persistent transcription has co-localized CTD-RNAP2 (red circles), Ser5ph-RNAP2 (green squares), and mRNA (blue diamonds) at the transcription site (TS), $n=7$ out of 20 cells. (c) Normalized signal intensities over time for all the cells analyzed as in b. The shadow and the line in the middle represent the S.E.M and the average, respectively. (d) Z-positions of all the cells quantified for transcription fluctuations. Each cell is represented with a different color/symbol (legend on the right). (e) Exemplary cell with periods of active and inactive transcription showing a control position (p1) near the transcription site, $n=13$ out of 20 cells. (f) Normalized intensity over time-target at an off-target position near the transcription site (p1; CTD-RNAP2, dashed red; Ser5ph-RNAP2 dashed green; mRNA, dashed blue) versus the mRNA signal at the transcription site (blue diamonds). (g) Covariance calculation between the normalized intensities of mRNA at the transcription site against CTD-RNAP2 or Ser5ph-RNAP2 at the transcription site and at p1. Data are presented as mean values \pm S.E.M. n =number of cells/number of independent experiments. (h) Ratiometric distribution of the euclidean distances for CTD-RNAP2 and mRNA to Ser5ph-RNAP2 and mRNA (light blue), CTD-RNAP2 and mRNA to Ser5ph-RNAP2 and CTD-RNAP2 (light purple), and Ser5ph-RNAP2 and mRNA to Ser5ph-RNAP2 and CTD-RNAP2 (light orange) in all the cells analyzed. The line in the middle of each box represents the mean. The top and the bottom of the box represent the 75% and 25% quantiles, respectively. The middle region in the error bar at the bottom and the top represent the lower and upper whiskers, respectively. n =number of events/number of cells. Significance was tested using a two-tailed Mann-Whitney U-test with $p \leq 0.05$ (*), $p \leq 0.0015$ (**), and $p \leq 9.2091 \times 10^{-4}$ (***)).

RNAP2 was just ~ 93 nm, significantly smaller than between either form of RNAP2 and mRNA (p -value $< 6.97 \times 10^{-11}$) (Fig 2.4h and Fig 2.5h). This spatial separation was consistent across the cells we analyzed (Fig 2.6) and independent of the strength of transcription as gauged by CTD-RNAP2, Ser5ph-RNAP2, and mRNA signal intensities. Together these results demonstrate that RNAP2 is spatially organized within the transcription site, with active mRNA synthesis spatially distinct from clusters of CTD-RNAP2 and Ser5ph-RNAP2.

2.3.4 Fluctuation dynamics and statistics are captured by a simple model of transcription bursting.

We wanted to obtain a more universal picture of RNAP2 phosphorylation dynamics at the HIV-1 reporter gene. We therefore performed correlation analysis [21, 34, 35] using all time points in all time series, similar to fluorescence correlation spectroscopy [36]. This technique is ideal for extracting information from noisy data provided there are a sufficient number of time series and/or time points. We began with an auto-correlation analysis, to see how long each signal remains correlated with itself given a lag time (τ) (Fig 2.7a). The auto-correlation of each signal decays

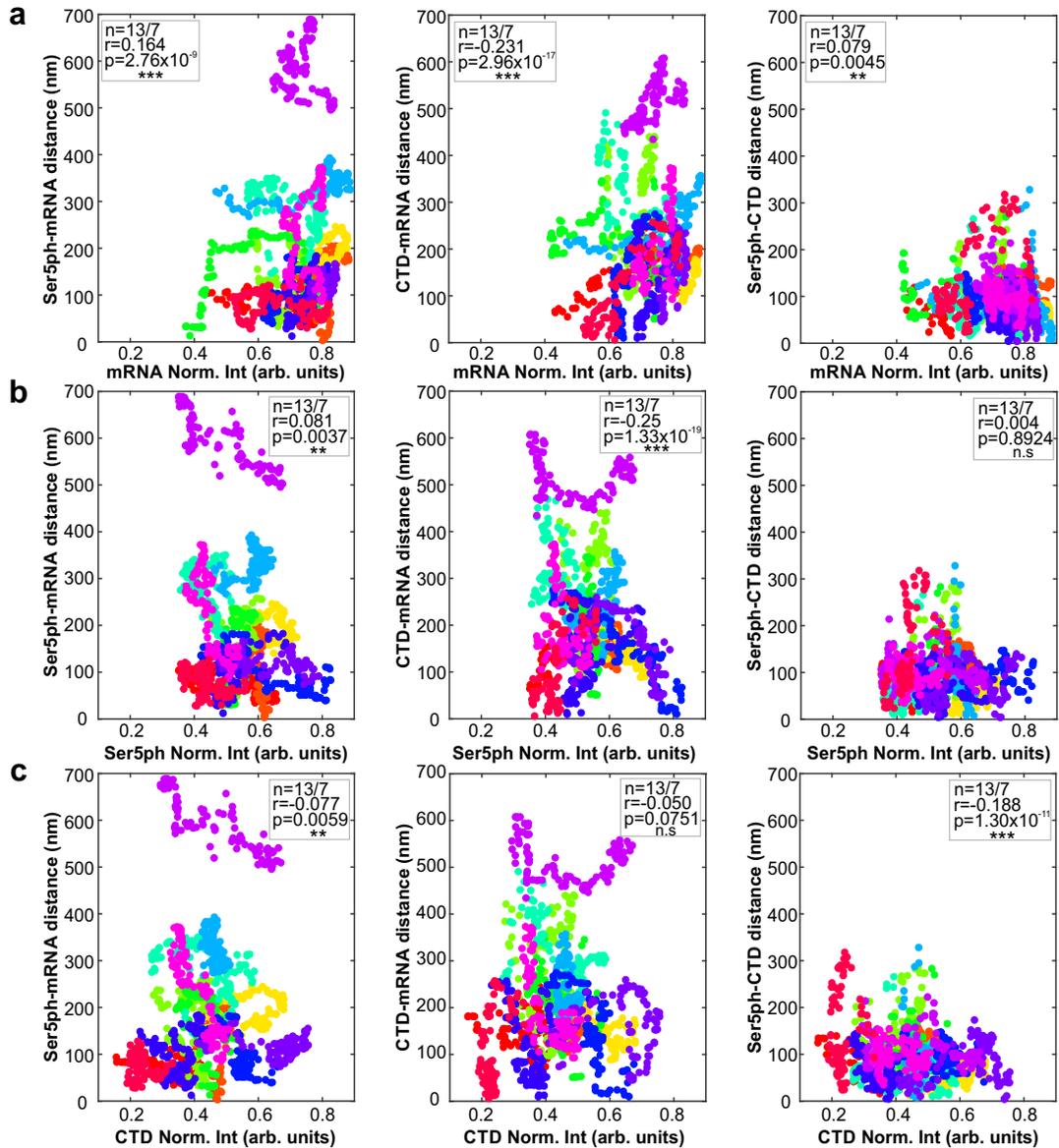


Fig. 2.6: Euclidean distances distribution versus mRNA, Ser5ph-RNAP2, and CTD-RNAP2 normalized intensities. Euclidean distance between Ser5ph-RNAP2 and mRNA (left panel), CTD-RNAP2 and mRNA (middle panel), and Ser5ph-RNAP2 and CTD-RNAP2 (right panel) versus the normalized intensities of (a) mRNA, (b) Ser5ph-RNAP2, and (c) CTD-RNAP2 for all the cells analyzed. Each cell corresponds to one color. n=number of cells/number of independent experiments. Correlation coefficient (r) and p-values (p) as, $p \leq 0.05$ (*), $p \leq 0.01$ (**), and $p \leq 0.001$ (***) were calculated using the “corrcoef” function in MATLAB.

with increasing lag time and eventually flattens out near zero. We define the dwell time as the lag time at which the auto-correlation falls below 20% of its initial zero-lag value. According to this

analysis, the two forms of RNAP2 had shorter average dwell times than mRNA, indicating RNAP2 was often unsuccessful in reaching the end of the gene and synthesizing an mRNA.

Next, we calculated the cross-correlation between signals. Consistent with our previous analysis aligning local minima, all possible pairs of signals were strongly correlated, as seen by large peaks in the cross-correlation curves near $\tau=0$ (Fig 2.7b). Measuring the precise position of each peak revealed the mRNA signal came substantially later than the CTD-RNAP2 and Ser5ph-RNAP2 signals, while the CTD-RNAP2 and Ser5ph-RNAP2 signals appeared at roughly the same time (within the 1 min sampling time of experiments). To better resolve the time delay between the CTD-RNAP2 and Ser5ph-RNAP2 signals, we re-imaged the HIV-1 transcription site in a single plane at a much faster frame rate (150 ms/frame) for a total of 1000 timepoints (150 sec). Although these higher temporal resolution experiments are much too short to capture the full auto- and cross-correlation curves, they are sufficient to resolve the short time-lag dynamics (Fig 2.9), and they revealed cross-correlation asymmetry with an off-center peak indicating that the Ser5ph-RNAP2 signal comes roughly 3-6 sec after the CTD-RNAP2 signal. The various delays we measure are consistent with the temporal ordering we saw by aligning local minima of the mRNA signal (Fig 2.4d) and provide further evidence that RNAP2 phosphorylation at Serine 5 is very rapid at the transcription site.

We next sought to find a quantitative model to unify our diverse data sets. We required that our model must simultaneously fit all three auto-correlation curves (Fig 2.7a) and all three cross-correlation curves (Fig 2.7b). To further constrain the model, we also counted mRNA at transcription sites by comparing their intensities to single mature mRNAs using FISH-quant [37] (Fig 2.7d, bottom). Consistent with an earlier report [20], we found the HIV-1 reporter contained an average of $\mu = 15.5$ mRNA with a relatively large standard deviation of $\sigma = 10.55$, and Fano Factor of $\sigma^2/\mu = 7.1$.

To unify our data, we posed several models with different levels of complexity (Fig 2.8). Each model considered a promoter with bursty expression. This was represented by specifying distinct active (ON) and inactive (OFF) promoter states with OFF-to-ON and ON-to-OFF transitions rates

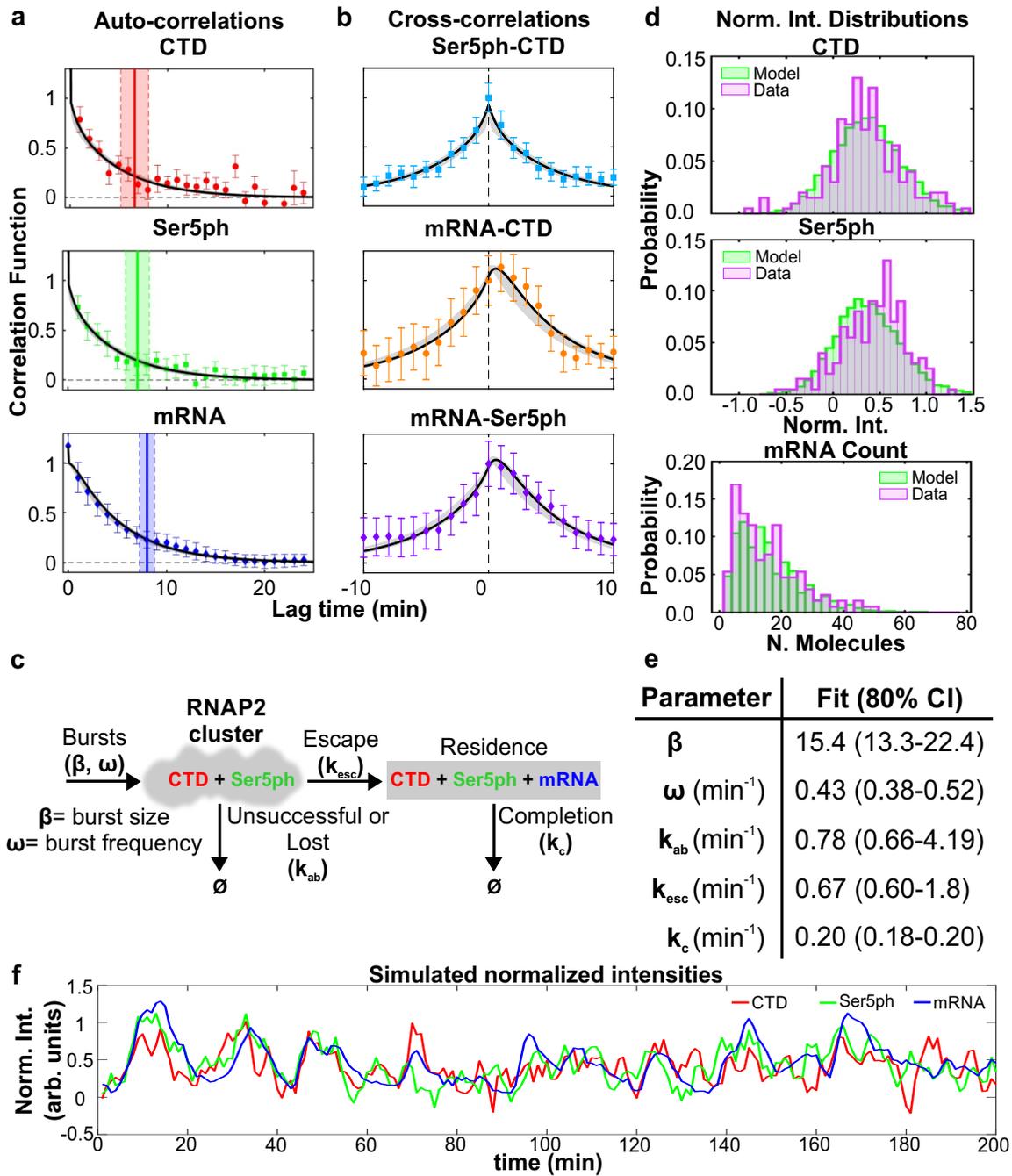


Fig. 2.7: Fluorescence auto- and cross-correlations at the HIV-1 reporter gene are well fit by a unifying model of transcription. (Continues next page)

Fig. 2.7: **(a,b)** Measured and modeled **(a)** auto-correlation functions $AC(\tau)/\overline{G(0)}$ for each signal: CTD-RNAP2 (red circles), Ser5ph-RNAP2 (green squares), and mRNA (blue diamonds). Dwell time is defined as the time at which the autocovariance dropped below 20% its zero-lag value (vertical full lines). Dwell time uncertainty is estimated from the model using SD from 400 simulated data sets, each with 20 cells over 200 min with 1 min simulation resolution (vertical dashed-lines). **(b)** cross-correlation function $CC(\tau)/\overline{G(0)}$ between signal pairs: Ser5ph-RNAP2 and CTD-RNAP2 (cyan squares), mRNA and CTD-RNAP2 (orange circles), and mRNA and Ser5ph-RNAP2 (purple diamonds) at the transcription site. Model MLE fit in black and sampled uncertainty in gray. **(c)** A simple model to capture RNAP2 fluctuation dynamics at the HIV-1 reporter gene. RNAP2 enter the transcription cluster with an average geometric burst with average burst size, β , and burst frequency, ω . Phosphorylation of Serine 5 is assumed to be fast ($\ll 1$ min) and/or the RNAP2 enter in a pre-phosphorylated form. RNAP2 can be lost from the cluster with rate k_{ab} or escape with rate k_{esc} . RNAP2 completes transcription with rate k_c . **(d)** Probability distributions for CTD-RNAP2 and Ser5ph-RNAP2 (arbitrary units of fluorescence), and mRNA (units of mature mRNA) for experimental data (purple) and model MLE predictions (green). **(e)** Maximum likelihood estimate (MLE) parameters and 95% CI range. Statistics presented for the data are sample means \pm S.E.M. n =number of cells/number of independent experiments (20/8). **(f)** Simulated trajectory (with shot noise equal to that of experiments) of CTD-RNAP2 (red), Ser5ph-RNAP2 (green), and mRNA (blue) intensities normalized to have a 95 percentile of unity.

k_{on} and k_{off} , respectively. When the promoter is ON, RNAP2 is recruited at a rate k_r [38]. Upon fitting these models to our data, the fitted burst duration was much shorter than the one minute experimental sampling time (i.e., $k_{off} \ll 1min$). This allowed us to simplify the model to one with burst frequency $\omega = 1/(1/k_{on} + 1/k_{off}) = 1/k_{on}$ and geometrically distributed bursts with average size $\beta = k_r/k_{off}$ [39]. In all models that fit our data, RNAP2 could unsuccessfully depart the promoter at rate k_{ab} or escape at rate k_{esc} . After escape, the RNAP2 would complete transcription at a combined rate k_c that includes both elongation and processing.

In the minimal model that matched all data, CTD-RNAP2 were immediately phosphorylated upon arrival at the promoter, which was consistent with the rapid (< 1 min) Serine 5 phosphorylation we observed (Fig 2.9). We also explored several more complicated models with separate steps for initiation, elongation, and processing, post-transcriptional mRNA retention [40], or with separate events describing Serine phosphorylation/initiation and de-phosphorylation/abortion (Fig 2.8). Each model was fit separately to maximize the likelihood for all observed data, but inclusion of additional mechanisms and free parameters provided only marginal improvements to the overall fit and resulted in much larger parameter uncertainties. Therefore, we used the Bayesian Information Criteria (BIC) to select our final model as the best choice given our available data

(See tables in Fig 2.8). By simultaneously fitting all six correlation plots (Fig 3a and 3b) as well as the nascent mRNA means and variances, we could estimate the best models five parameters with excellent precision (Fig 2.11). The best-fit parameter values and their uncertainties are provided in Fig 2.7e. According to the best fit, bursts of RNAP2 occur on average every $1/\omega \approx 2.3$ min

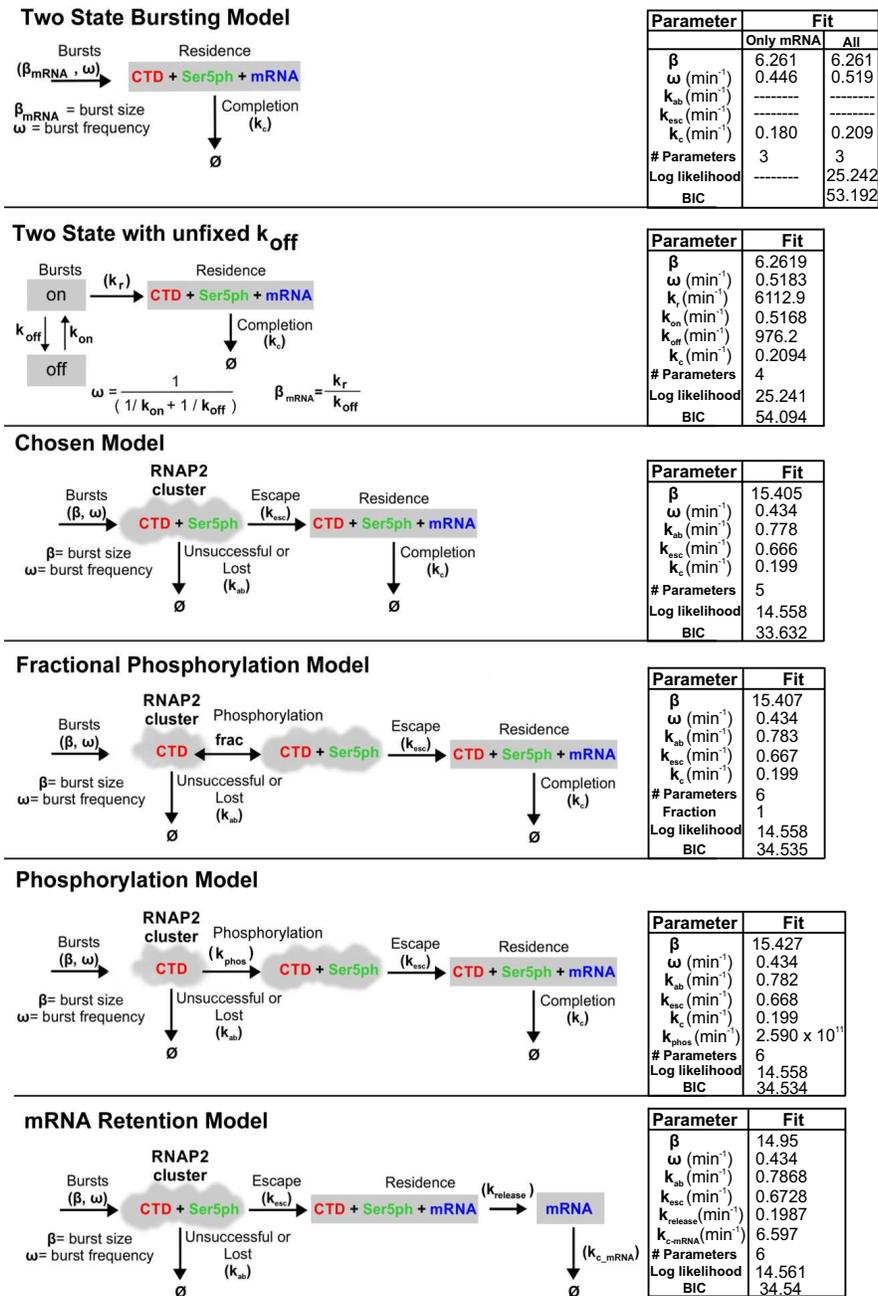


Fig. 2.8: Mathematical models tested.

Fig. 2.8: In the simplest “Two-state Bursting Model”, RNAP2 comes in with all 3 signals in a bursting fashion and leaves with a rate k_c . “The Two-State Bursting Model k_{off} unfixed” is the same as the previous model, however with the parameter k_{off} not fixed at 1000 and allowed to optimize. The “Chosen Model” is the model described in the main text and in Fig.3. This model was selected as it fits the experimental data well with a minimum amount of parameters (lowest Bayes Information Criterion, BIC, of tested models). The “Fractional Phosphorylation Model” contains an extra parameter compared to the Chosen Model, frac . Here, frac represents the fraction of unescaped RNAP2 with Ser5ph. frac is analogous to the ratio of two rates with timescales much faster than the rest of the model: CTD gaining Ser5ph and CTD+Ser5ph losing Ser5ph. In the “Phosphorylation Model”, RNAP2 binds the promoter and then becomes Ser5ph phosphorylated at a rate of k_{phos} . With this model, RNAP2 requires the Ser5ph signal to escape and transcribe to completion. The “mRNA Retention Model” allows mRNA to remain at the transcription site after RNAP2 completes transcription. All RNAP2 with Ser5ph leave with rate k_{release} and mRNA are retained at the TS. The mRNA then leave at a rate of $k_{c-\text{mRNA}}$. Next to each model, the estimated parameters, maximum likelihood, and BIC are shown.

and have an average size of about $\beta \approx 15$ molecules per burst. Of the RNAP2 that arrive at the promoter, a substantial fraction $f = k_{\text{esc}}/(k_{\text{esc}} + k_{\text{ab}}) \approx 0.46$ escape the promoter and complete transcription, leading to convoys [20] of about $f \cdot \beta \approx 7$ RNAP2 per burst. Each mRNA takes an average of $1/k_c \approx 5$ min to complete elongation and processing, meaning that on average the HIV-1 reporter contains mRNA originating from $\omega/k_c \approx 2$ consecutive bursts. Overall, the model predicts that there are an average of ~ 20 RNAP2 on the gene in steady-state, with an average of ~ 5 in the cluster near the promoter in an unphosphorylated or Ser5ph form, and ~ 15 elongating or processing near the end of the gene (See Table 1). This average picture is somewhat misleading, however, as the number of RNAP2 within the cluster fluctuates dramatically due to randomly timed bursts. According to our simulations, there are periods when as many as ~ 90 RNAP2 come in at a time interspersed by brief and random silent periods of low RNAP2 occupancy (Fig 2.12a).

After fitting the model to capture the auto- and cross-correlation functions and the mean and variance of the mRNA distribution, we verified that it also correctly predicted the full probability distributions for the number of nascent mRNA molecules and RNAP2 signal intensities at the HIV-1 transcription site (Fig 2.7d). We also simulated normalized intensities including shot noise (Fig 2.7f), and these look similar to our measured trajectories (Fig 2.4c). The shot noise was estimated directly from the experiments by comparing the observed zero-lag covariance $G(0)$ compared to an estimate for the zero-lag autocovariance found by interpolation from the short,

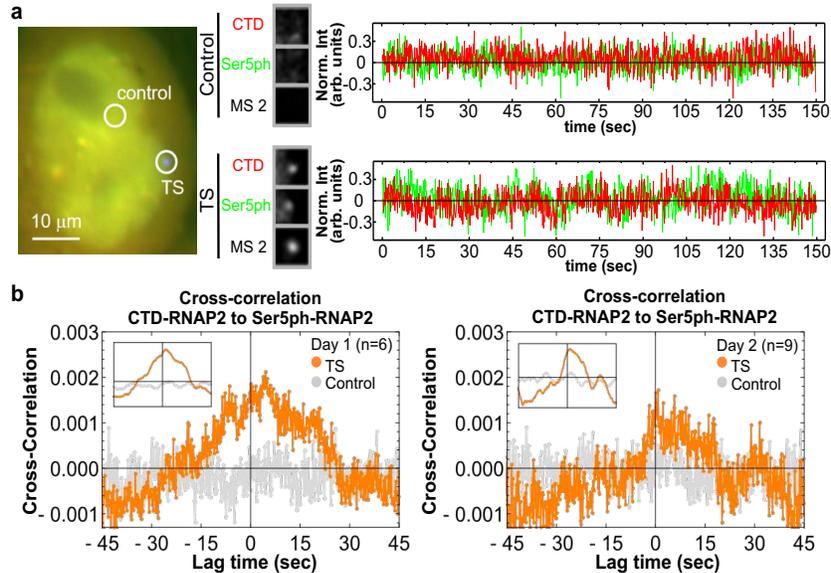


Fig. 2.9: Fast-imaging experiments revealed a 3-6 sec time delay between CTD-RNAP2 and Ser5ph-RNAP2. (a) (Left) Exemplary cell for fast imaging (150 msec/frame) for a total of 1000 time points (150 sec) in a single plane. Two positions are highlighted: the HIV-1 TS and a control nonspecific spot. (Center) Crops showing the CTD- (red), Ser5ph-RNAP2 (green), and MS2 mRNA (black) signals at the TS and control positions within the exemplary cell. (Right) Normalized intensity at the TS (bottom) and control (top) positions over time from the exemplary cell for CTD-RNAP2 (red), Ser5ph-RNAP2 (green). (b) Measured cross-correlation function $CC(\tau)$ between CTD-RNAP2 and Ser5ph-RNAP2 at the TS (orange circles) and control (gray circles) positions separated by experimental day. The inset shows a 50-frame rolling average to more easily identify the peak time delay between the two signals. In both cases, the cross-correlation peaks at a lag time of roughly 3-6 seconds.

but non-zero time lags. These shot noise standard deviations were found to be $1.98\times$, $1.42\times$, and $0.41\times$ of the standard deviation for CTD, Ser5ph, and MS2 signals, respectively. Finally, we simulated ChIP data for our single-gene reporter (Fig 2.12b-c). To do this, we assumed an elongation rate of 4.1 kb/min (measured previously at this locus by analyzing the MS2 stochastic fluorescence fluctuations [20]) and processing rate of 0.27 min^{-1} (so elongation and processing times sum to our fitted $1/k_c$ completion time). With these rates, the CTD/Ser5ph-RNAP2 simulated ChIP signals from active genes displayed strong peaks at the beginning and end of the gene, as we observed in Fig 2.1b (compare to Fig 2.12b-c). Overall, the excellent match between data and simulations indicates our best-fit model faithfully captures transcription dynamics at the HIV-1 reporter. To facilitate further exploration of our model, we provide a graphical user interface (GUI) [<https://doi.org/10.5281/zenodo.4631141>]. The GUI allows exploration of how each model param-

eter affects model predictions, including trajectories, auto- and cross-correlations, distributions of spot intensities, simulated ChIP data, and several derived quantities to describe the CTD-RNAP2, Ser5ph-RNAP2, and mRNA burst dynamics (Fig 2.13).

2.3.5 Inhibiting distinct steps of the transcription cycle provides further evidence for spatiotemporal organization of RNAP2 phosphorylation.

So far, our collective data and modeling suggest a precise temporal ordering of transcription dynamics, beginning with the recruitment of CTD-RNAP2, followed by rapid initiation in 3-6 sec (indicated by Ser5ph-RNAP2), and promoter escape and elongation within another minute or so (indicated by mRNA). Our data also provide evidence of heterogeneity in the distribution of RNAP2 along the gene, with high concentrations near the beginning and end of the gene (Fig 2.1b). To further test our system, we perturbed it by adding three different transcription inhibitors: Triptolide (TPL), THZ1, and Flavopiridol (Flav) (Fig 2.10). We began by inhibiting the earliest steps in the transcription cycle to attempt to prevent the formation of the RNAP2 cluster. To achieve this we added TPL, a small-molecule inhibitor that prevents promoter DNA opening and transcription initiation by inhibiting the DNA-dependent ATPase activity of the XPB subunit of TFIIH [4,41]. TPL has also been shown to induce RNAP2 degradation on the hours timescale [42], so we imaged for just 30 consecutive minutes to focus on the more immediate impact of TFIIH inhibition. Addition of 5 μ M TPL led to a rapid and dramatic loss of both mRNA and all RNAP2 signals at the transcription site within just \sim 10 min (Fig 2.10a-d). Consistent with our previous findings, we observed a temporal ordering in the TPL-induced run-off of RNAP2 (Fig 2.10c), with CTD-RNAP2 signals dropping earlier than Ser5ph-RNAP2, followed by mRNA. This ordering was observed in 7 out of 10 single cells we measured. Of these, 4 exhibited clear separation between the three traces (inset in Fig 2.10c). Since steps that are later in the CTD cycle necessarily take longer to respond to drugs, this ordering provides further evidence that CTD-RNAP2 slightly precedes Ser5ph-RNAP2 by less than a minute, and that both RNAP2 signals come significantly earlier than mRNA. These data also demonstrate that the opening of promoter DNA by XPB is a requirement

for the formation of RNAP2 clusters. This can work by at least two mechanisms: (1) All the Ser5ph-RNAP2 underwent initiation and abortion, but RNAP2 kept its Serine 5 phosphorylation; (2) Initiation of the first RNAP2 activates CDK7, which can phosphorylate many RNAP2 within the cluster.

We next used THZ1, which inhibits RNAP2 CTD phosphorylation at Serine 5 by targeting the TFIID kinase CDK7, thereby preventing promoter pausing, mRNA capping, and productive elongation [4, 18, 43]. In contrast to TPL, THZ1 has a slower action, so a higher concentration and longer exposure to this drug were needed to see an effect in real-time. Treatment with 15 μ M THZ1 led to a reduction in the mRNA signal at the HIV-1 reporter within 25 min (Fig 2.10e). Likewise, both CTD-RNAP2 and Ser5ph-RNAP2 levels were on average reduced. Interestingly, in some single cells we observed large, temporally ordered bursts in the levels of CTD-RNAP2 and Ser5ph-RNAP2, despite continued inhibition and overall loss of mRNA. These large bursts could even achieve RNAP2 levels that were as high as pre-treatment levels (the thicker black curve in Fig 2.10e highlights one example). Presumably these bursts occur because there is residual TFIID left in the cell that are not yet inhibited by THZ1, or because recently aborted RNAP2 retain their Ser5ph within the cluster. Since mRNA levels did not burst to the same degree, we conclude the bursts arise from clusters of RNAP2 near the promoter that initiate but fail to escape. These transient clusters near the beginning of the gene are consistent with the high concentration of RNAP2 near the promoter we observed by ChIP (Fig 2.1b) and are also consistent with the ChIP predictions of our best-fit model (Fig 2.12b-c).

We next blocked a later step in the transcription cycle using 1 μ M Flav, a drug that prevents transcription elongation and RNAP2 CTD phosphorylation at Serine 2 by inhibiting the CDK9 activity of P-TEFb [18, 44]. Like THZ1, Flav also reduced the intensity of the mRNA signal, this time within \sim 15 min (Fig 2.10f). However, CTD-RNAP2 and Ser5ph-RNAP2 signals remained relatively unchanged, exhibiting large fluctuations and a slight overall reduction on average. This difference from THZ1 can be attributed to the later action of Flav in the transcription cycle. The high levels of CTD-RNAP2 and Ser5ph-RNAP2 signals that remained post-Flav again support a

dynamic clustering model [4, 5, 7–9] in which most RNAP2 are already phosphorylated at Serine 5 and presumably make repeated attempts at initiation and promoter escape.

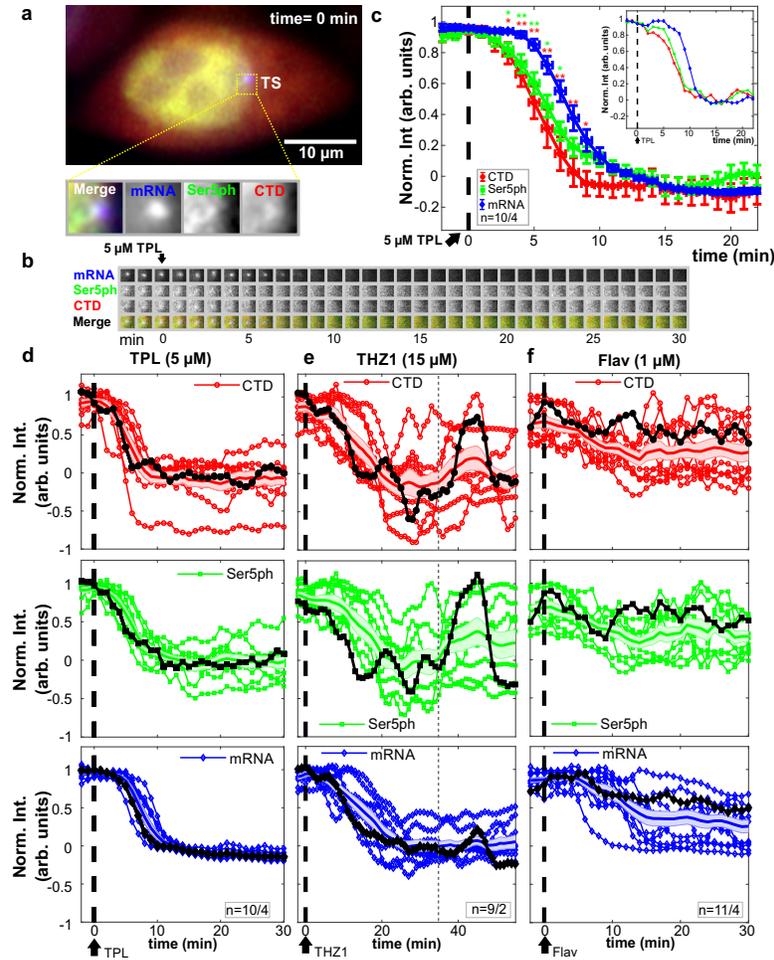


Fig. 2.10: Intensity fluctuations of CTD-RNAP2, Ser5ph-RNAP2 and mRNA in the presence of transcription inhibitors. (a) Sample cell before addition of Triptolide (TPL). The transcription site (TS) is shown in the dotted box, and the inset shows a zoom in, ($n=10/4$). (b) The TS from (a) at all times before and after addition of TPL. (c) Normalized average TS intensity over time of all the quantified cells for CTD-RNAP2 (red circles), Ser5ph-RNAP2 (green squares), and mRNA (blue diamonds) before and after application of TPL (vertical dashed line). The inset shows the signals in a representative cell. Data are presented as mean values \pm S.E.M. Significance was tested using a two-tailed Mann-Whitney U-test with $p \leq 0.05$ (*) and $p \leq 0.01$ (**). (d-f) Normalized intensity signals after application of various transcription inhibitors, including (d) TPL ($5 \mu\text{M}$), (e) THZ1 ($15 \mu\text{M}$), and (f) Flavopiridol (Flav, $1 \mu\text{M}$) for all the cells analyzed. Signals highlighted in black correspond to a sample single cell, the colored shadow and the full line in the middle of it correspond to the standard deviation and the mean in each channel, respectively. The vertical gray dashed line in (e) highlights the time point at which a burst of RNAP2 is observed in the sample cell in the presence of THZ1. n =number of cells/number of independent experiments.

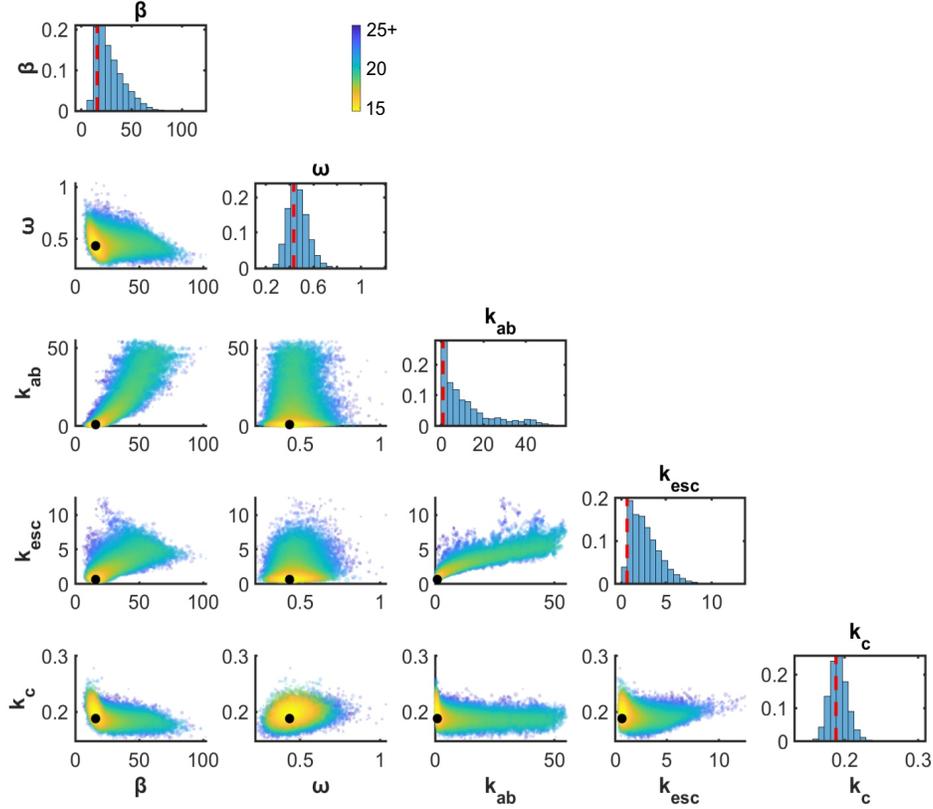


Fig. 2.11: Parameter sensitivity analysis. Metropolis-Hastings algorithm was run to determine posterior uncertainty of model parameters given the experimental data. Plots on the diagonal show the marginal posterior parameter distributions for each parameter (MLE parameter estimate denoted by red dashed line) and off-diagonal plots show the joint posterior parameter distributions for all pairs of parameters (MLE parameter combination denoted by black marker). Colors denote log-likelihood value; an upper bound of the 0.5% highest log-likelihoods was selected for coloring purposes. Any log-likelihood's color above this threshold was set to that bound). A proposal distribution of a 5 dimensional Gaussian with a standard deviation of 3% of MLE parameters was used. 20 individual chains of 250000 with a thinning rate of 20 (100 million) were used to generate the posterior distributions. 40000 points of the posterior are displayed in the figure.

Finally, we attempted to qualitatively recapitulate these perturbations using our best-fit model. To do so, we evaluated several hypothetical mechanisms in which transcription is inhibited by reducing one or more of the rates, including burst statistics (ω or β), the promoter escape rate k_{esc} , or the completion rate k_c . According to simulations, inhibiting earlier steps (ω or β) in the transcription cycle led to the sequential loss of all RNAP2 and mRNA signals at the transcription site at a rate governed by the time scale of mRNA elongation and processing (Fig 2.14a), reminiscent

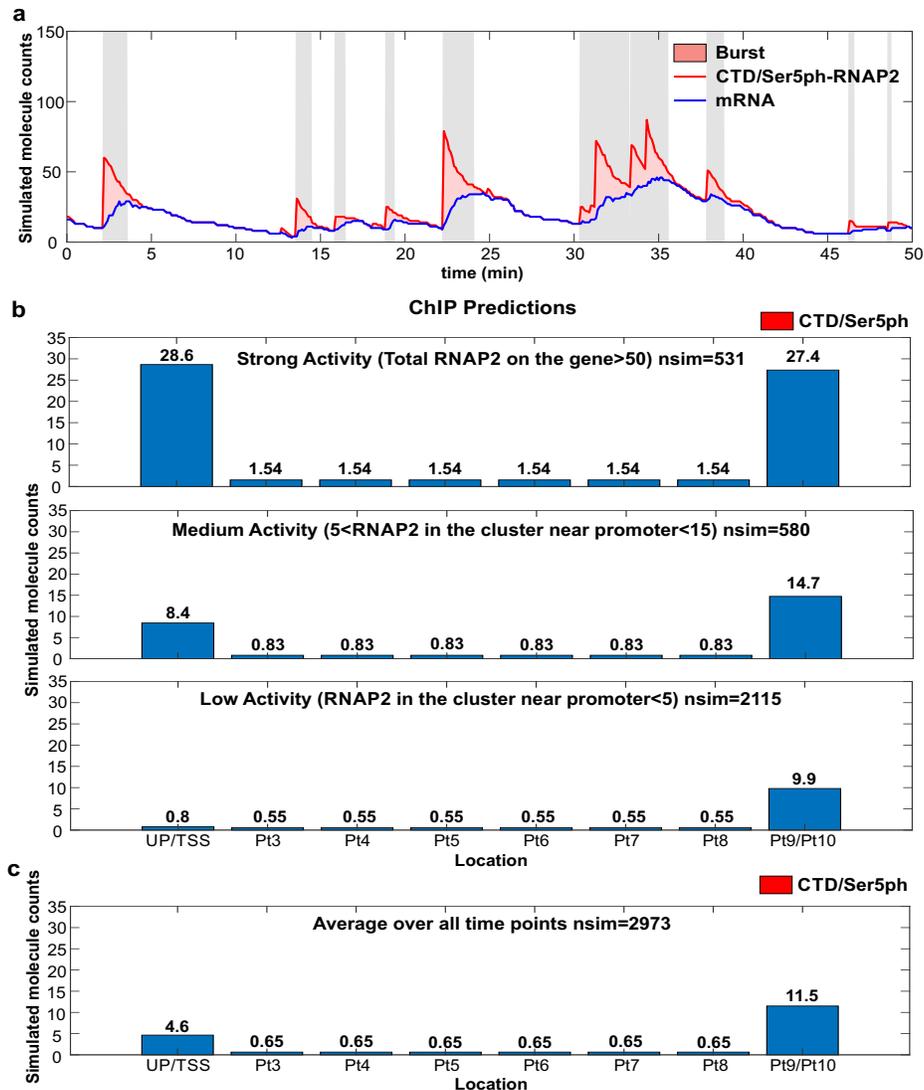


Fig. 2.12: Simulated trajectories and ChIP predictions. (a) Stochastic simulation for the number of nascent mRNA per transcription site (blue line), total number of RNAP2 at transcription site (red line), and number of RNAP2 in cluster near transcription site promoter (red shading). Periods with ≥ 10 RNAP2 at the transcription site cluster (gray shading) are classified as ‘ON’ (14.0% of total time); periods with no RNAP2 at the cluster are classified as ‘OFF’ (42.9% of time); and periods with intermediate levels of RNAP2 in the cluster are classified as ‘transient’ (43.1% of time). Note that for clarity these simulations do not include the experimental shot noise used to simulate actual measurements (as in Fig. 3f, for example). (b) Simulated ChIP data as predicted using the model for: (Top; Strong Activity) average spot during an ON period; (Middle; Medium Activity) average spot during a transient period; and (Bottom; Low Activity) average spot during an OFF period. Each stochastic simulation was run for 120,000 min and sampled at 40 min intervals to ensure de-correlated points. To estimate RNAP2 loading at the inner bins, an elongation rate of 4.1 kb/min was assumed and used to get the fraction of time spent elongating versus processing of the total RNAP2 residence time. This fraction of elongation time was then distributed from the final bin uniformly to the middle bins and is represented by the middle numbers of bins Pt3-8. (c) Average simulated RNAP2 ChIP over all times including all ON, OFF and transient periods.

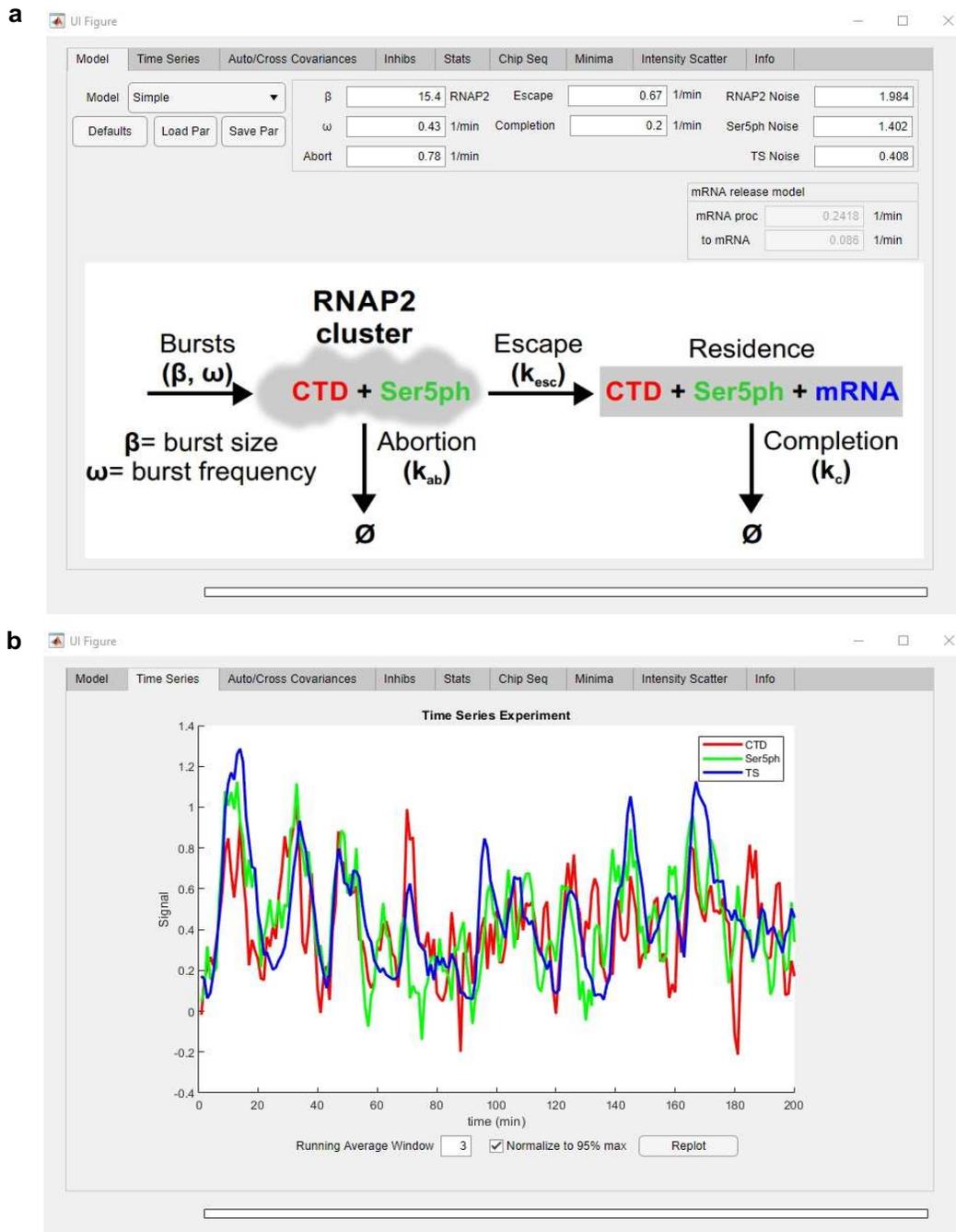


Fig. 2.13: Graphical User Interface (GUI) for the transcription model. To facilitate the simulation of transcription dynamics at a single-copy gene, the model described in the main text has been incorporated into a MATLAB toolbox. **(a)** This graphical user interface (GUI) is divided into eight upper tabs, and input boxes for specification kinetic parameters. The GUI allows the simulation of intensity trajectories in each channel. **(b)** Sample display of simulated intensities normalized to the 95th percentile and running averaged with a window of three time points. The GUI also allows for display of auto-, cross-correlations, predicted minima from the experimental data previously loaded, prediction of CHIP distributions, and perturbed intensity trajectories by blocking different steps of transcription in the model.

of our TPL experiments. In contrast, inhibiting a later step (k_{esc}), led to a retention of large numbers of RNAP2 in the cluster that undergoes relatively large and rapidly changing fluctuations (Fig 2.14b), reminiscent of our THZ1 experiments. Blocking (k_{esc}) and reducing k_c by 30 % led to a slight reduction in the mRNA signal and even less decrease in the RNAP2 signals with relatively large fluctuations (Fig 2.14c), reminiscent of our Flav experiments. We also blocked bursts (either ω or β) and reduced k_c by 30 % and obtained an overall reduction of all the signals (Fig 2.14d) that do not represent any of the inhibitors tested here. The similarity between these simulations and our experimental perturbations provide further support for our model and also provide evidence that the tested inhibitors act on distinct stages of the RNAP2 transcription cycle.

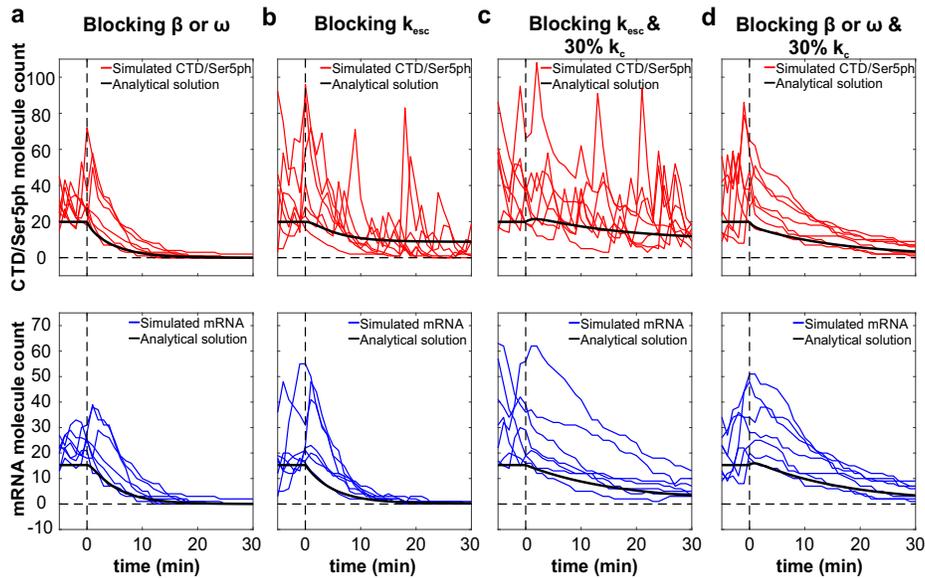


Fig. 2.14: Predicted CTD/Ser5ph-RNAP2, and mRNA signals after perturbing different steps in the mathematical model. Simulated molecule counts for CTD/Ser5ph-RNAP2 (red, upper panels), and mRNA (blue, bottom panels) after blocking: **(a)** β or ω , **(b)** k_{esc} , **(c)** k_{esc} and 30% k_c , and **(d)** β or ω and 30% k_c , and their respective analytical solution in each plot (black curve). Simulated trajectories with mRNA molecule counts above the analytical solution at time of inhibition are shown with colored lines. This was done to simulate the experimental procedure of choosing transcription sites at the beginning of an experiment where all three signals could be seen. Blocking is defined as multiplying the best fit parameter by 0.01 (99% reduction), similarly blocking 30% refers to multiplying the best fit parameter by 0.3 (70% reduction). For blocking β and ω , k_{off} was defined by setting k_{off} to 1000, effectively turning off bursting dynamics.)

2.4 Discussion

In this study, we measured the dynamics of the RNAP2 CTD transcription cycle at the single-gene level in living cells. By combining complementary antibody-based imaging probes with multi-color single-molecule microscopy and computational modeling, we were able to detect organization in both the temporal ordering and spatial distribution of endogenous RNAP2 phosphorylation along a single HIV-1 reporter gene.

We find that a large number of RNAP2 at the HIV-1 transcription site are clustered around the promoter in a region that is spatially distinct from elongating RNAP2 and mRNA synthesis (as depicted in Fig 2.15). This spatial organization supports the notion of dynamic RNAP2 clusters that form transcriptional hubs [45] or factories [46,47] that contain high concentrations of transcription machinery. In steady-state, we estimate there are an average of ~ 20 RNAP2 at the HIV-1 gene. This total number of RNAP2 is in between recent estimates of ~ 80 RNAP2 [6] clustered at the constitutively expressed beta-actin locus, ~ 17 RNAP2 at an exogenous mini-gene [17], and ~ 7.5 RNAP2 at the Pou5f1 locus [17]. Of the ~ 20 RNAP2 at our HIV-1 reporter gene, we estimate on average ~ 5 are at or near the promoter, awaiting initiation or promoter escape. During frequent bursts, however, this number can dramatically increase to as high as 90 RNAP2, with most either coming in with Serine 5 phosphorylation or rapidly acquiring Serine 5 phosphorylation within seconds (Fig 2.9). Given the limited amount of space at the promoter, it is hard to imagine all of these RNAP2 are promoter bound. Instead, we believe many are unbound and collectively this fraction helps form the transcription cluster, which remains spatially distinct from mRNA synthesis.

A major unresolved question is how RNAP2 are retained in clusters. One possibility is that RNAP2 are trapped by repeated interactions with other transcription machinery in the region. Alternatively, clusters could represent phosphorylation-dependent condensates. As others have recently shown, phase separation can be driven by phosphorylation of the unstructured RNAP2 CTD [7] and by the histidine-rich tail of P-TEFb [8]. Since Tat directly interacts with P-TEFb [28, 48], it could enhance RNAP2 recruitment and clustering at the HIV-1 reporter gene.

One possible advantage of the cluster is it retains recently aborted RNAP2 near the transcription start site so they can rapidly re-initiate. This follows from our rapid imaging experiments, which indicate initiation is very rapid (3-6 sec; Fig 2.9) compared to promoter escape (fitted $1/k_{\text{esc}} \sim 1.5$ min). The distinct timescales imply two hypotheses: First, most promoter escape attempts fail. This is consistent with earlier measurements based on FRAP that demonstrated successful promoter escape is a rare event [18, 49]. Second, a large fraction of RNAP2 in the cluster are inactive at any given time [5, 50]. Such a large fraction of inactive RNAP2 could arise from recently aborted molecules that retain their Ser5ph. Evidence for the retention of Ser5ph on RNAP2 after transcription abortion was seen in an earlier study [18], where Ser5ph-RNAP2 was detected in the soluble fraction of cells after transcription was globally inhibited via flavopiridol. The retention of RNAP2 also helps explain our model prediction that nearly half of the RNAP2 in the cluster ($k_{\text{esc}}/(k_{\text{esc}} + k_{\text{ab}}) \sim 46\%$) eventually do escape the promoter and produce a full-length transcript. Thus, local recycling of transcription machinery within clusters may play a role in HIV-1 biogenesis, where Tat expression provides a positive feedback loop to amplify transcription and facilitate the rapid production of viral proteins in host cells [51].

While the overall efficiency of transcription is relatively high at the HIV-1 reporter gene compared to other genes studied, the various kinetic rates we quantified are fairly consistent with earlier work. In particular, we found RNAP2 takes around five minutes to complete transcription after promoter escape ($1/k_c$ in Fig 2.7). This places an upper bound on the RNAP2 elongation and processing time. If we constrain the elongation rate to be 4.1 kb/min [20] (~ 1 min for the full gene), then we can assign the remaining time (~ 4 min) to RNA processing at the 3' end. Under these conditions, the model predicts a build up in RNAP2 at the 3' end of the gene because processing takes longer than elongation. This buildup is consistent with our ChIP data in Fig 2.1b. The estimated 4 min processing time is also consistent with an earlier estimate at this HIV-1 locus [20], although such relatively long processing may not be representative of other genes. Similarly, the RNAP2 initiation and promoter escape rates we quantified are consistent with earlier reports, taking between a minute and a few minutes [27, 49]. Finally, we also detected bursts in

transcription that result in convoys of RNAP2, as previously reported [20], and consistent with widespread bursting observed across the genome [38, 52]. The global agreement between studies suggests some convergence in the field, particularly given the uniqueness of our data set, which is based on fluctuations of both MS2 [21] and RNAP2 Fab signals [27].

The ability to image by fluorescence microscopy endogenous RNAP2 phosphorylation dynamics at a single-copy genes now makes it possible to estimate the RNAP2 distributions predicted by ChIP. ChIP studies of the RNAP2 CTD transcription cycle typically display heterogeneous distributions of RNAP2 that have distinct peaks of Ser5ph-RNAP2 near the promoter and Ser2ph-RNAP2 at the ends of genes [1, 3, 11]. However, based on ChIP alone, it is not clear if peaks represent the distribution of RNAP2 along single genes or instead represent a population of genes. For example, it could be that half of the genes have Ser5ph-RNAP2 paused at the beginning of the gene, while the other half have Ser2ph-RNAP2 being processed near the end of the gene. In this extreme example, no single gene would have RNAP2 at both ends. According to our best-fit model, the situation for HIV-1 is not this extreme, but the distribution of RNAP2 does depend sensitively on the timing of bursts. For example, early in a burst RNAP2 occupancy is heavily front-loaded, with all or nearly all RNAP2 at or around the promoter in a Serine 5 phosphorylated form. Since RNAP2 ChIP by design is biased towards genes with high levels of RNAP2 at the time of assay, genes that have recently burst are likely to be overrepresented in the data (Sup. Fig 5b-c). As our model demonstrates, soon after a burst, genes tend to have far more RNAP2 clustered around the promoter than the average gene (which has just five) (Fig 2.15). According to this interpretation, the large Ser5ph-RNAP2 ChIP peak we observe near the promoter could arise from rapid and repeated promoter proximal initiation and/or pausing. Given the nature of ChIP, it is also possible the peak arises from RNAP2 within clusters that are non-specifically cross-linked during the fixation step. However, this latter possibility seems unlikely as promoter proximal peaks are also observed using techniques that detect and sequence nascent mRNA, such as GRO-seq, PRO-seq, and mNET-seq [53]. In the future, it will be interesting to see to what extent dynamic cluster-

ing observed in living cells correlates with promoter-proximal RNAP2 peaks observed across the genome in populations of fixed cells [54].

Aside from HIV-1, our technology can now be used to examine RNAP2 phosphorylation dynamics at other single-copy genes. Given the high correlation between MS2 (mRNA) and RNAP2 (Fabs), in the future MS2 may not even be required. For example, by combining Fab and CARGO [55], RNAP2 phosphorylation dynamics at any endogenous gene could be visualized without extensive genome editing. Alternatively, Fab could be combined with other labeling technologies such as *lacO/lacI* [56, 57], ROLEX [58], ANCHOR [59], or post-fixation via DNA FISH [60] or CasFISH [61]. Beyond RNAP2, post-translational modifications to other proteins involved in transcription could also be studied in this way, including histones [23, 62]. However,

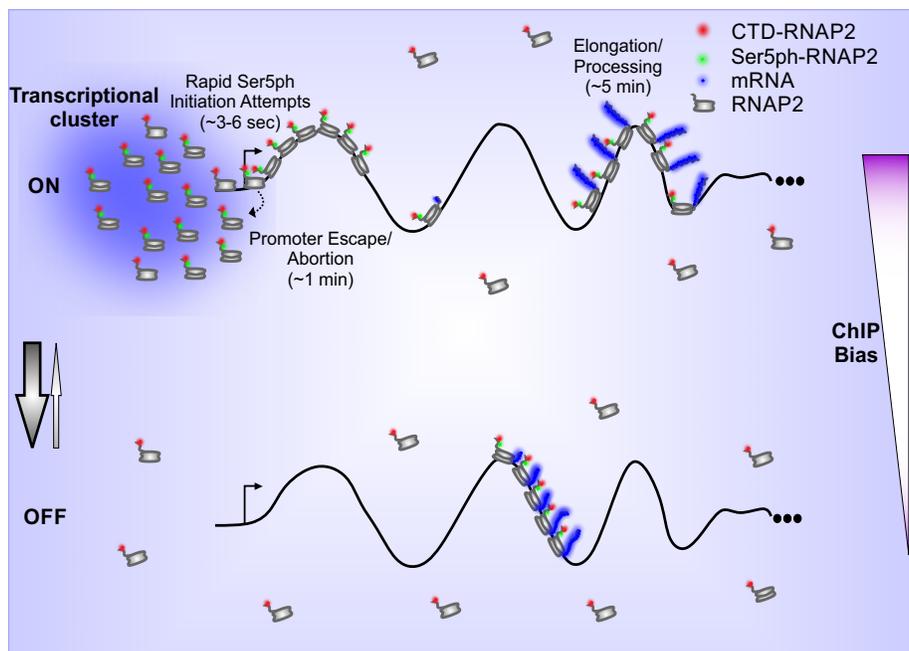


Fig. 2.15: Model depicting RNAP2 transcription dynamics in a single-copy gene. During extremely short ($\ll 1$ min) periods of the ON state, RNAP2 is recruited in bursts (~ 15 RNAP2) the HIV-1 reporter gene, creating transient (~ 1 min) of CTD-RNAP2 and Ser5ph-RNAP2 at the gene promoter, and initiating transcription in RNAP2 convoys (~ 7 RNAP2/convoy). The middle of the gene remains mostly empty due to rapid transcription, while a large number of RNAP2 (~ 15) concentrate at the end of the gene during processing ($\sim 4 - 5$ min). In OFF periods (~ 2.3 min), RNAP2 convoys that escaped the promoter during the ON state quickly elongate and complete transcription. The gene rapidly transitions back to the OFF state when ON (denoted by arrows). ChIP assays enrich for genes with lots of RNAP2, which will bias the assay towards genes with RNAP2 clusters near the promoter.

a few important caveats of Fab- or intrabody-based imaging should be kept in mind: First, if Fabs bind their targets with too low affinity, then there will be a large unbound fraction that will decrease signal-to-noise. For the CTD-RNAP2 and Ser5ph-RNAP2 Fab, the bound fraction was determined to be greater than 80% [27]. Second, if Fabs take too long to bind their targets, then very rapid processes can be entirely missed or their timescales will appear erroneously slow. According to FRAP, the vast majority of CTD-RNAP2 and Ser5ph-RNAP Fabs used in this study bind and re-bind their targets in well under 10 sec [27], meaning processes on the seconds time scale can be discerned, but anything shorter may be missed. Third, if Fabs are too numerous in a cell, they may compete with one another for binding and Fab targets could become saturated, both of which could interfere with the underlying biology. We introduce $1-3 \times 10^6$ Fab per cell [22], far less than the $\sim 1.5 \times 10^7$ RNAP2 heptad repeats [27, 63]. We therefore do not expect Fabs to compete or interfere. Together these three caveats place considerable constraints on experiments, but they are not prohibitive. With the continued development of Fab [23], scFv [24], and nanobodies [64] for live-cell imaging, finding a suitable intrabody has become significantly easier. We therefore anticipate our technology will become a valuable tool to study transcription dynamics at the single-gene level.

2.5 Methods

Cell Culture. Transcription dynamics experiments were performed in HeLa Flp-in H9 cells (H-128). The H-128 cell line generation was described previously [20]. Briefly, H-128 cells harbor an HIV-1 reporter gene tagged with an MS2X128 cassette, controlled by Tat expression. The HIV-1 reporter comprises the 5' and 3' long terminal repeats (LTRs) containing the viral promoter, polyA sites, as well as HIV-1 splice donor (SD1), splice acceptor (SA7) and Rev-responsive element (RRE). H-128 cells also stably express MS2 coat protein tagged with GFP (MCP-GFP), which binds to MS2 repeats when they are transcribed into mRNA. Cells were maintained in a humidified incubator at 37°C with 5% CO₂ in Dulbecco's modified Eagle medium (DMEM, Thermo Fisher Scientific, 11960-044) supplemented with 10% fetal bovine serum (FBS, Atlas Biologicals), 10

U/mL penicillin/streptomycin (P/S, Invitrogen), 1 mM L-glutamine (L-glut, Invitrogen) and either 400 $\mu\text{g}/\text{mL}$ Neomycin (Invitrogen) or 150 $\mu\text{g}/\text{mL}$ Hygromycin (Gold Biotechnology).

Chromatin Immunoprecipitation and quantitative-Polymerase Chain Reaction (ChIP-qPCR). ChIP was performed as described previously [65] with minor modifications. Briefly, H-128 cells grown in a 10 cm dish were fixed with 1% PFA in DMEM at room temperature for 5 min, neutralized in DMEM containing 200 mM glycine for 5 min and washed with PBS and NP40 buffer (10 mM TrisHCl, pH 8.0, 10 mM NaCl and 0.5% NP40). Fixed cells were lysed with 360 μL SDS dissolution buffer (50 mM TrisHCl, pH 8.0, 10 mM EDTA and 1% SDS) and diluted with 1440 μL ChIP dilution buffer (50 mM TrisHCl, pH 8.0, 167 mM NaCl, 1.1% Triton 100 \times and 0.11% sodium deoxycholate), supplemented with a proteinase inhibitor cocktail. After shearing chromatin using a Bioruptor UCD200 (Diagenode) at sonications of 40 sec with 50 sec intervals, eight times at high level, the median size of fragmented DNA was 200 base pairs with a range of 50500 base pairs. The supernatant, cleared by centrifugation at 20,000 $\times g$ for 10 min at 4°C, was diluted with 5.4 mL ChIP dilution buffer and then incubated with 40 μL sheep antimouse IgG magnetic beads preincubated with 1 μg mouse anti-CTDRNAP2 (MABI 0601), anti-Ser5phRNAP2 (MABI 0603) and anti-Ser2phRNAP2 (MABI 0602) monoclonal antibodies (Cosmo Bio USA) at 4°C overnight with rotation. The immune complexes were washed with lowsalt RIPA buffer (50 mM TrisHCl, pH 8.0, 1 mM EDTA, 150 mM NaCl, 0.1% SDS, 1% Triton 100 \times and 0.1% sodium deoxycholate), highsalt RIPA buffer (50 mM TrisHCl, pH 8.0, 1 mM EDTA, 500 mM NaCl, 0.1% SDS, 1% Triton 100 \times and 0.1% sodium deoxycholate) and then washed twice with TE buffer (10 mM TrisHCl, pH 8.0, and 1 mM EDTA). DNA was eluted with ChIP elution buffer (10 mM TrisHCl, pH 8.0, 300 mM NaCl, 5 mM EDTA and 0.5% SDS). After incubation at 65°C overnight to reverse the crosslinks, DNA was purified by RNase A and proteinase K treatments and recovered using a DNA purification kit (Qiagen). For ChIPqPCR, the immunoprecipitated DNA and total DNA were quantified by Power SYBR Green PCR Master Mix in a Mx3000P RealTime qPCR System (Agilent Technologies). The primers used for qPCR are listed in Sup. Table 1.

Antigen-binding fragment (Fab) generation and fluorescence conjugation. Fab preparation was performed using the same monoclonal antibodies used in ChIP experiments and the Pierce Mouse IgG1 Fab and F(ab')₂ Preparation Kit (Thermo Scientific), as described before [27]. In brief, ficin resin was equilibrated with 25mM cysteine (in HCl, pH 5.6) to digest the antibodies (CTD-RNAP2 or Ser5ph-RNAP2) into Fab. The IgG concentration used was 4 mg, and the digestion reaction was incubated for 5 h. Fab and Fc regions were separated using a Nab Protein A column (Thermo Scientific). Fabs were concentrated up to ~1 mg/mL using an Amicon Ultra 0.5 filter (10k cut-off, Millipore) and conjugated with CF640 or Cy3 (Invitrogen) dyes. For labeling Fab, 100 μ g of purified Fab and 10 μ L of 1M NaHCO₃⁻ were mixed to a final volume of 100 μ L, then 2 μ L of CF640 or 2.66 μ L of Cy3 was added, and the mixture was incubated at RT for 2 h in a rotator protected from the light. The labeled Fab sample was passed through a PD-mini G-25 desalting column (GE Health care), previously equilibrated with PBS, to remove unconjugated Fab, and then the dye-conjugated Fab was concentrated up to ~1 mg/mL with an Amicon Ultra filter 0.5 (10k cut-off). The degree of labeling (DOL) was calculated using eq. 1, where ϵ_{IgG} and ϵ_{dye} are the extinction coefficients of IgG at 280 nm and the dye (provided by the manufacturer), A_{Fab} and A_{dye} are the absorbances determined at 280 and 650 or 550 nm, and CF is the correction factor for the dye at 280 nm (provided by the manufacturer). In this study, only Fabs with a DOL between 0.75 and 1 were used for live-imaging experiments.

$$\text{DOL} = \frac{\epsilon_{\text{IgG}}}{\epsilon_{\text{dye}}} * \frac{1}{\left(\frac{A_{\text{Fab}}}{A_{\text{dye}}}\right)^{-1} - \text{CF}} \quad (2.1)$$

Loading fluorescent Fabs into living cells. Cells were cultured in glass bottom dishes (35 mm, 14 mm glass, Mat-Tek). The next day dye-conjugated Fabs were loaded into the cells through bead-loading [22, 27, 29, 66, 67], as follows: First, the fluorescent Fabs (CTD-RNAP2-CF640 and Ser5ph-RNAP2-Cy3, ~1 mg/mL, each) were mixed with PBS up to 4 μ L in the cell culture hood. Second, the medium was removed completely from the dish and stored, and the Fab mixture was added to the center of the dish. Third, glass beads (106 μ m, Sigma-Aldrich, G-4649) were immediately sprinkled on top before cells dried up and the dish was tapped ~10 times against the

bench. This tapping causes the beads to roll over cells and induce small tears into which the Fab can diffuse in. Fourth, the stored medium was quickly added back to the cells, again to prevent cells from drying out. Cells were then placed in the incubator to recover for 1-2 h. Post-recovery, the glass beads were gently washed out with phenol-free DMEM (DMEM⁻, Thermo Fisher Scientific, 31053-028), and the cells were stored in DMEM⁺ medium (DMEM⁻ supplemented with 10% FBS, 10 U/mL P/S and 1 mM L-glut) for live-imaging experiments.

Chemicals. The transcription inhibitors, Triptolide (TPL, Sigma Aldrich), Flavopiridol (Flav, Selleck Chemicals), THZ1, (Selleck Chemicals), fluorescence dyes, Cy3 (Invitrogen), CF640 (Invitrogen), and HaloTag TMR Ligand (5 mM) (Promega) were dissolved in DMSO (Sigma-Aldrich) and stored at -20°C until use. RNAP2 inhibitors were added to DMEM⁺ medium to reach the desired final concentration in the cells.

Microscopy. A custom-built widefield fluorescence microscope with highly inclined illumination was used in all experiments [13, 33]. The microscope has three excitation beams: 488, 561, and 637 nm solid-state lasers (Vortran) that are coupled and focused on the back focal plane of the objective (60×, NA 1.48 oil immersion objective, Olympus). The emission signals were split by an imaging grade, ultra-flat dichroic mirror (T6601pxr, Chroma) and detected with two aligned EM-CCD (iXon Ultra 888, Andor) cameras by focusing with a 300 mm tube lens (generating 100× images with 130 nm/pixel). Cell chambers were mounted in a stage-top incubator (Okolab) at 37°C with 5% CO₂ on a piezoelectric stage (PZU-2150, Applied Scientific Instrumentation). The focus was maintained with the CRISP Autofocus System (CRISP-890, Applied Scientific Instrumentation). The cameras, lasers and piezoelectric stage were synchronized with an Arduino Mega board. Image acquisition was performed with Micro-Manager software (1.4.22) [68]. Unless otherwise stated, the imaging size was set to 512 x 512 pixels² (66.6 x 66.6 μm²) and the exposure time set to 53.64 ms. The readout time of the cameras from the combination of the imaging size and the vertical shift speed was 23.36 ms, which resulted in an imaging rate of 13 Hz (77 ms per image).

For three-color imaging, far-red fluorescence (e.g. CF640 or Alexa Fluor 647) was imaged on one camera with an emission filter (FF01-731/137/25, Semrock), while red fluorescence (e.g. Cy3

or TMR) and green fluorescence (e.g. GFP) were alternately imaged on the other camera via a filter wheel (HS-625 HSFW TTL, Finger Lakes Instrumentation) with an emission filter for red fluorescence (593/46 nm BrightLine, Semrock) and green fluorescence (510/42 nm BrightLine, Semrock). The filter wheel position was rapidly switched during the 23.36 ms camera read-out time by the Arduino Mega board. For two-color imaging, far-red fluorescence was simultaneously imaged on one camera while red or green fluorescence was imaged on the other camera with the appropriate emission filters.

Immunofluorescence. Cells grown on glass bottom dishes (35 mm, 14 mm glass, uncoated, Mat-Tek Corporation) were fixed with 4% Paraformaldehyde (Electron Microscopy Sciences) in 1 M HEPES (Sigma-Aldrich) with or without 10% Triton 100 \times (Fisher Scientific) (pH 7.4) for 10 min at room temperature (RT), and washed with PBS (3 \times). Permeabilization (1% Triton 100 \times in PBS) and blocking (100% blocking One-P, Nacalai-USA) were performed individually, for 20 min at RT, gently rocking, and rinsing with PBS (3 \times) after each step. The cells were incubated for 2 h at RT with 1 mL of antibody solution (10% blocking One-P:90% PBS) containing 2 μ g/mL of mouse monoclonal primary antibody (CTD-RNAP2 (MABI 0601), Ser5ph-RNAP2 (MABI 0603), Ser2ph-RNAP2 (MABI 0602), as described in [27] and now available from Cosmo Bio USA, H3K27ac (MABI0309), H3K27me (MABI0321), H3K4me1-3 (MABI0302-0304), H3K9me2 (MABI0317), and H3K9me3 (MABI0318), purchased from Cosmo Bio USA). After rinsing with PBS (3 \times), the cells were incubated for 1 h at RT with 1 mL of antibody solution containing 1.5 μ g/mL of Alexa Fluor 647 Donkey Anti-Mouse IgG (Jackson ImmunoResearch) and washed with PBS (3 \times). Then, the cells were mounted using Aqua-Poly/Mount (Fischer Scientific) for imaging. Single images were acquired with laser powers at the back focal plane set to 86 μ W and 51.2 μ W for 488 nm and 637 nm, respectively.

Additionally, to show that CTD- and Ser5ph-RNAP2 Fabs stain cells distinctively, immunostaining using pre-labeled Fabs against CTD-RNAP2-CF640 and Ser5ph-RNAP2-Cy3 was performed. For this type of experiments, cells were fixed, permeabilized and blocked as described above, and then incubated in an antibody solution containing 2 μ g/mL of each pre-labeled Fab for

1 h at RT. Post Fab incubation, cells were rinsed with PBS and mounted in Aqua-Poly/Mount. The images were collected using the following laser powers at the back focal plane: 123 μW , 750 μW , and 230 μW for 488 nm, 561 nm, and 637 nm, respectively.

Single-molecule experiments using H2B-Halo. Cells were plated in glass bottom dishes at a seeding density of $\sim 10^4$ cells/cm². The next day, cells were transfected with 2.5 μg of H2B-Halo in a 1:1 (mass) ratio using Lipofectamine LTX (ThermoFisher Scientific, 15338-100). 24 h post-transfection, cells were stained with 5nM Halo-Ligand TMR pre-treated with 30 mM NaBH₄ for 30 min in the CO₂ incubator (Acros Organics) to reduce the fluorophore and induce stochastic photoblinking in live-cells [69]. After staining, the cells were washed 3 times total. Each wash consisted of 3 \times 1mL DMEM⁻, and 1mL DMEM⁺ with a 5 minute interval between washes. Cells were imaged immediately after staining and washing. For this, the imaging size was set to 256 x 256 pixels² (33.3 x 33.3 μm^2) and the exposure time set to 30 ms. This resulted in an imaging rate of 22.8 Hz (30 ms exposure + 13.86 camera readout = 43.86 ms per frame). Single z-planes were acquired for 10,000 frames total with laser powers at the objective's back focal plane set to 125 μW , and 9.93 mW for 488 nm and 561 nm, respectively. To minimize photobleaching, the 488 nm laser fired once every ten frames (to track the transcription site), while the 561 nm laser fired every frame (for tracking individual H2B).

Single molecule tracks were identified using TrackMate 3.8 with the following parameters: LoG Detector; Estimated Blob Diameter: 5.0; Pixel Threshold: 100; Sub-Pixel Localization: Enabled; Simple LAP Tracker; Linking Max Distance: 3 pixels; Gap-Closing Max Distance: 2 pixels, Gap-closing Max Frame Gap: 1 frame. Custom Mathematica code was used to calculate average Euclidean displacement for each track longer than 5 frames. Tracks were plotted with a blue-purple color distribution based upon their average Euclidean displacement. The transcription site was identified using TrackMate, and plotted in red.

Live-cell imaging of transcription at the HIV-1 reporter gene To cover the entire cell nucleus, all movies were taken using 13 z-stacks with 0.5 μm spacing. The z position was moved only after all three colors were imaged in each plane. This resulted in a total cellular imaging rate

of 0.5 Hz (2 s per volume). Note that the color scheme of the signals described in the text and figures is based on the color of the excitation lasers, CTD-RNAP2 in red (CF640), Ser5ph-RNAP2 in green (Cy3), and mRNA in blue (GFP). For shorter live-cell imaging as in Fig 2.1, each cell was scanned every 1 min for 30 min with the laser power at the objective's back focal plane set to 21.4 μW , 60.5 μW , and 21.74 μW for 488 nm, 561 nm, and 637 nm, respectively, and the exposure time was 53.64 ms. For longer live-cell imaging as in Fig 2.4, cells were imaged every 1 min for 200-time points, using weaker laser powers (1.15 μW , 15.7 μW , and 5.2 μW for 488 nm, 561 nm, and 637 nm, respectively) and longer exposure times (200 ms exposure). For faster live-cell imaging as in Fig 2.9, each cell was scanned at a much faster frame rate (150 ms/frame) for a total of 1000 time points (150 sec) in a single plane. For this, the imaging size was set to 256 x 256 pixels² (33.3 x 33.3 μm^2) and the exposure time set to 53.64 ms, with the laser power at the objective's back focal plane set to 1.2 mW, 335 μW , and 77.5 μW for 488 nm, 561 nm, and 637 nm, respectively.

Calibrating the number of mRNA per transcription site. To count the number of nascent mRNAs at the transcription site, cells were imaged for a single time point using a higher laser power for 488 nm (230 μW at the back focal plane) and a lower camera gain. These conditions allowed us to visualize both a single transcription site and single mature mRNAs. To calculate the number of mRNA per transcription site (see Fig 2.7d, bottom panel): (1) Several cells were imaged on independent days. To avoid bias, cells were chosen with the same imaging conditions used for longer live-cell experiments; (2) Images were analyzed using FISH-quant V3 [37]. Mature mRNAs were detected, localized in 3D with a Gaussian fit, and then a point-spread function was applied to discard spots that were larger than diffraction-limited spots. An image showing the average intensity of the mature mRNAs was created and compared to that of the transcription site. This ratio of these gave the number of nascent mRNA at each transcription site, from which the distribution shown in Fig 2.7d (bottom panel, purple distribution) was computed.

Quantifying signal intensities at the transcription site from live-cell imaging movies. Images were pre-processed using either Fiji [70] or custom-written batch processing Mathematica code (Wolfram Research 11.1.1) to create 2D maximum intensity projections from 3D movies.

Using Mathematica code, the 3D images were corrected for photobleaching and laser fluctuations, z-stack by z-stack, by dividing the movie by the mean intensity of the whole cell or the nucleus in each channel. The offset between the two cameras was registered using a built-in Mathematica routine FindGeometricTransform, which finds a transformation function that aligned the best fitted positions of 100 nm diameter Tetraspeck beads evenly distributed across the image field of view. 2D maximum projections and 3D image sequences from the images corrected for bleaching and laser fluctuations were then analyzed with a custom-written code in Mathematica to detect and track the transcription site. Briefly, thresholds were selected in each channel to visualize spots at the transcription site and a bandpass filter was used to highlight just the transcription site in the mRNA channel. The resulting image was binarized and used to create two masks for each time point: one marking the transcription site (transcription site mask: a mask semi-manually thresholded to cover just the transcription site within the image) and one marking the background (BG mask: a ring of width one pixel that surrounds the transcription site and is separated from the site by two pixels). The built-in Mathematica routine ComponentMeasurements-IntensityCentroid was used to find the coordinates of the transcription site in XY through time. The Z coordinate was determined by selecting the z-stack at which the particle in the XY coordinate had its maximum brightness (“best z”). If the transcription site disappeared (due to transcription turning off or inhibition), the Z was replaced by the Z coordinate of the last visible position. From the XYZ coordinates at each time point, a new 2D maximum projection was created considering the “best z” at each time point. From this, the pixel intensity values were recorded for each transcription site (TS) and background (BG) mask, representing the mean intensity values over time at the transcription site and the background, respectively. The raw and normalized intensity vectors were calculated per channel and a moving average of three time points was used to display the intensity $\text{RawInt}_{\text{Ch}}$ as a function of time, as shown in Eq. 2:

$$\text{RawInt}_{\text{Ch}} = \langle I_{\text{TS}}(t) - I_{\text{BG}}(t) \rangle_3 \quad (2.2)$$

The normalized intensity (as in Fig 2.4c) was calculated by dividing $\text{RawInt}_{\text{Ch}}$ by the average 95% intensity from all transcription sites. Occasionally, normalized intensities for CTD-RNAP2 and Ser5ph-RNAP2 dip below zero. This can be caused either by RNAP2 signals being temporally depleted at the transcription site relative to the background or by bright signals in the background due to nearby transcription in the local vicinity. To display transcription sites over time (as in Figs. 1c and 2b and Sup. Movie 1), 3 time point moving-average trims from the “best z” were created in each channel (showing CTD-RNAP2, Ser5ph-RNAP2, mRNA, and the merge). Each trim was centered on the intensity centroid of the mRNA.

Covariance analysis in Sup. Figs. 1f and 2g. To test for covariance between intensity signals from control spots and the transcription site, signal covariance was calculated using the “cov” function in MATLAB. For quantification of bleed through, the covariance was calculated between all possible pairs of raw intensities (CTD-mRNA, CTD-Ser5ph, and Ser5ph-mRNA) in normal vs. bleed-through control conditions. For quantification of signals off-target, the covariance was calculated between all possible pairs of normalized intensities on-target at the transcription site vs. off-target at a random site p1. Significance was calculated using the Mann-Whitney U-test.

Analysis of minima signal in Fig 2d. The local minima in the mRNA signal of each cell was detected using the `islocalmin` function in MATLAB. The cells that exhibited minimas below a threshold (normalized intensity ≤ 0.20 a.u.) were selected by the algorithm. Then 7-time points before and after the mRNA valley were considered, including the minimum, in each channel. All the traces in each channel were averaged and fitted with a Gaussian using a 95% confidence interval to determine the minima and maximum steady state of the average trace in each channel.

To confirm that the minima were true and not an artifact of our analysis, the analysis was repeated at hundreds of random time points. Significance was calculated using the Mann Whitney U-test. The p-values for the magnitude of the minima and their time delays were calculated by comparing the magnitude of the minima to the control and the time lag to minute zero in each signal, respectively.

Analysis of transcription site spatial organization in Fig. 2.4e-h and Fig 2.5h. Moving average (50 time points) movies were generated to accurately determine the mean XY position of the transcription site in each channel. As described in Quantifying signal intensities at the transcription site from live-cell imaging movies subsection, the built-in Mathematica routine ComponentMeasurements-IntensityCentroid was used. Once the XY positions for each signal were obtained, the Euclidean distance between each pair over time was calculated, from which distributions were calculated. Significance between signals was calculated using the Mann-Whitney U-test.

Auto- and cross-correlation analysis. The auto- and cross-correlation functions were calculated for each time trace obtained from the longer movies (like in Fig 2.4c, but without performing a 3-time point moving-average), as previously described [34, 35]. The covariance function is defined:

$$G(\tau) = \langle \delta a(t) \delta b(t + \tau) \rangle, \quad (2.3)$$

where $\langle \cdot \rangle$ indicates the temporal mean, and $\delta a(t)$ denotes the deviation about the mean, i.e., $\delta a(t) = (a(t) - \langle a(t) \rangle)$. Signals $a(t)$ and $b(t)$ can be the same signal or two different signals. In the first case, $a = b$ and $G(\tau)$ represent the auto-covariance, which is symmetric about $\tau = 0$; In the second case, $G(\tau)$ represents the cross-covariance and may be asymmetric. To calculate the cross-correlation between the CTD-RNAP2 and Ser5ph-RNAP2 signals in the fast imaging experiments in Fig 2.9, the intensities of tracked transcription sites through time were quantified as described above, with a couple of minor modifications: First, because imaging was in a single plane, the rate of photobleaching in the plane was not captured by the rate of photobleaching in the cell. For this reason, each signal exponentially decayed. This was corrected by dividing out a single-exponential fit to each curve. Second, we did not perform any moving average on the signals to maintain the highest possible temporal resolution.

For fitting and data analysis, the normalized covariances, $G(\tau)/\overline{G(0)}$, were used for all signals, where $\overline{G(0)}$ denotes the zero lag auto- or cross-covariance averaged over all time points and all biological replicas. To quantify and remove shot noise from the zero-lag auto-covariances, $G(0)$

was estimated for each biological replica assuming a linear interpolation from the three shortest non-zero lag times (1, 2, 3 minutes) prior to averaging over all replicas. The standard error of the mean normalized covariance functions, denoted $\text{SEM}_{\overline{G}}(\tau)$, was computed as the standard deviation of $G(\tau)/\overline{G(0)}$ divided by \sqrt{N} .

A quantitative model for transcription. The derivation of the bursting model for RNAP2 recruitment and nascent transcription simple model begins with the specification of three variables: $x_1(t)$ describes the promoter state, $x_2(t)$ describes the number of RNAP2 in the cluster, and $x_3(t)$ describes the number of RNAP2 engaged in active transcription. Six reactions can occur: (1) a promoter can become temporarily active with propensity equal to the burst frequency, $\omega \sim k_{\text{on}}$; (2) the active promoter can deactivate at a rate k_{off} ; (3) the active promoter can recruit and phosphorylate RNAP2 at Serine 5 (Ser5ph-RNAP2) at a rate $\beta \cdot k_{\text{off}}$; (4) Ser5ph-RNAP2 can be lost from the cluster at rate k_{ab} ; (5) Ser5ph-RNAP2 can escape at rate k_{esc} ; and (6) escaped RNAP2 can complete transcription with rate k_c . We solve the model for the first and second order statistical moments as previously described [71]. First, we combine the stoichiometry vectors for all six reactions into the stoichiometry matrix, \mathbf{S} as follows:

$$\mathbf{S} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}, \quad (2.4)$$

and we write the linear propensity functions in vector form as:

$$\mathbf{w} = \begin{bmatrix} -k_{\text{on}} & 0 & 0 \\ k_{\text{off}} & 0 & 0 \\ \beta \cdot k_{\text{off}} & 0 & 0 \\ 0 & k_{\text{ab}} & 0 \\ 0 & k_{\text{esc}} & 0 \\ 0 & 0 & k_{\text{c}} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} k_{\text{on}} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad (2.5)$$

$$= \mathbf{W}_1 \mathbf{x} + \mathbf{w}_0 \quad (2.6)$$

With this notation, the expected mean dynamics of $\mathbb{E}\{\mathbf{x}\}$ are described by the ordinary differential equation:

$$\frac{d\mathbb{E}\{\mathbf{x}\}}{dt} = \mathbf{S} (\mathbf{W}_1 \mathbb{E}\{\mathbf{x}\} + \mathbf{w}_0). \quad (2.7)$$

From this expression, the steady state expected mean can be calculated as the solution to the algebraic expression:

$$\mathbf{S} \mathbf{W}_1 \mathbb{E}_{\text{SS}}\{\mathbf{x}\} + \mathbf{S} \mathbf{w}_0 = 0; \quad (2.8)$$

the steady state co-variance, Σ_{SS} , can be calculated as the solution of the algebraic Lyapunov equation:

$$\mathbf{S} \mathbf{W}_1 \Sigma_{\text{SS}} + \Sigma_{\text{SS}} \mathbf{W}_1^T \mathbf{S}^T + \mathbf{S} \text{diag} (\mathbf{W}_1 \mathbb{E}_{\text{SS}}\{\mathbf{x}\} + \mathbf{w}_0) \mathbf{S}^T = 0; \quad (2.9)$$

and the auto- and cross-covariance functions versus time lag, $\Sigma(\tau)$ can be calculated as the solution of the ODE:

$$\frac{d\Sigma_x(\tau)}{d\tau} = \mathbf{S} \mathbf{W}_1 \Sigma_x(\tau), \quad (2.10)$$

with initial condition $\Sigma_x(0) = \Sigma_{\text{SS}}$ given as the solution to (2.9).

To convert these above expressions, which are in terms of x_1 , x_2 , and x_3 , into quantities reflecting the total RNAP2 at the transcription site ($y_1 = x_2 + x_3$) and number of transcribing RNAP2

($y_2 = x_3$), we define a simple linear transformation:

$$\mathbf{y} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x}, \quad (2.11)$$

$$= \mathbf{c}\mathbf{x}. \quad (2.12)$$

Under this transformation, $\mathbb{E}\{\mathbf{y}\} = \mathbf{c}\mathbb{E}\{\mathbf{x}\}$ and $\Sigma_{\mathbf{y}}(\tau) = \mathbf{c}\Sigma_{\mathbf{x}}(\tau)\mathbf{c}^T$.

We note that this version of the model does not distinguish between RNAP2 and Ser5ph-RNAP2. These two distinct forms as well as other configurations are easily incorporated by extending \mathbf{x} to include a fourth or more states. In such cases, each new state adds two reaction stoichiometry vectors to Eqn. (4), two reaction terms to Eqn (5), and one additional row to the output matrix \mathbf{c} in Eqn (11), but the rest of the analysis remains unchanged.

Using this model formulation, it is straightforward to solve for the steady state moments (Eqs. 2.8 and 2.9) and the auto- and cross-correlations (Eq. 2.10) for any combination of parameters. However, upon fitting this model to the data, we observed that estimates for k_{off} tended to very large values ($k_{\text{off}} \gg \omega$) and with substantial estimation uncertainty. Under these excessively large rates for k_{off} , each ‘on’ period is extremely short-lived and attracts a geometric number of RNAP2 with mean β (this model reduction is equivalent to the strategy in models that use geometric bursts of protein to replace translation of short-lived mRNA as described previously elsewhere [39]). Therefore, to reduce the number of free parameters required by the model, we fixed k_{off} at 1000 min^{-1} such that each burst would be very short lived on the time scale of the experimental measurements. This choice led to simpler model, but had no discernible effect on the fit of the model to the data.

All codes, including graphical user interface are available at Zenodo

[<https://doi.org/10.5281/zenodo.4631141>].

Model Parameter search.

Parameters were found using maximum likelihood estimation (MLE) considering several data types as follows. First, errors in the measurement of the normalized auto- and cross-covariances were assumed to be normally distributed with the measured standard error, $\text{SEM}_{\bar{G}}(\tau)$, such that their log-likelihood functions are written

$$\log L_G(\theta) = C_G - \frac{1}{2} \frac{\sum_{n=1}^{N_t} (\bar{G}_D(\tau_n) - \bar{G}_M(\tau_n, \theta))^2}{(\text{SEM}_{\bar{G}}(\tau_n))^2} \bar{G}_D, \quad (2.13)$$

where θ is the set of parameters, $\bar{G}_D(\tau_n)$ is the measured covariance function in the data (D), $\bar{G}_M(\tau_n, \theta)$ is the predicted covariance function of the model (M) at a time lag of τ_n , and C_G is a normalization constant that does not depend on the parameters. The summation is over the first 15 lag times for the three auto-covariance functions and the 21 smallest lag times (i.e., -10min to 10min) for the three cross-covariance functions.

The model was further constrained to match the mean and variance for the measured number of mRNA per transcription site as estimated in units of mature mRNA as calibrated using FISH-quant. Assuming the central limit theorem, the log-likelihood of matching the observed sample mean was estimated as:

$$\log L_\mu(\theta) = C_\mu - \frac{1}{2} \frac{(\mu_D - \mu_M(\theta))^2}{\text{SEM}_D^2}, \quad (2.14)$$

where μ_D is the sample mean levels of mRNA from the data, $\mu_M(\theta)$ is the mean number of mRNA predicted by the model, and $\text{SEM}_D = 0.93$ is the standard error of mean level of mRNA from the data. Similarly, the log-likelihood of the measured variance, σ_D^2 given the model was estimated as

$$\log L_{\sigma^2}(\theta) = C_{\sigma^2} - \frac{1}{2} \frac{(\sigma_D^2 - \sigma_M^2(\theta))^2}{\text{SEM}_{\sigma^2}^2}, \quad (2.15)$$

where $\sigma_M^2(\theta)$ is the mRNA variance predicted by the model, SEM_{σ^2} is standard error for the mRNA variance, and C_{σ^2} is a constant that does not depend on the parameters. The standard error of the

sample variance was estimated using a Gaussian approximation such that:

$$\text{SEM}_{\sigma^2} = \text{SEM}_D^2 \sqrt{\frac{2}{N-1}} = 14. \quad (2.16)$$

Under the assumption of independence between the different data types, the total log likelihood to match all data was the sum of the individual likelihoods:

$$\log L_{\text{Total}}(\theta) = \log L_G(\theta) + \log L_\mu(\theta) + \log L_{\sigma^2}(\theta). \quad (2.17)$$

Maximum likelihood estimates were found using iterated rounds of MATLAB's `fminsearch` until convergence.

To compare multiple models with different numbers of mechanisms and parameters, we computed the Bayesian Information Criteria (BIC) as:

$$\text{BIC} = k \log(n) - 2 \log L_{\text{Total}}(\theta). \quad (2.18)$$

In this formulation, the value for the number of independent experiments, n , was estimated at $n = 8$, which conservatively assumes one data degree of freedom for each of the six different auto- and cross-correlation signals estimated from the time lapse experiments, and one each for the measurement of the mean and variance of mRNA per transcription site as estimated imaging a single frame using higher laser power to visualize single mature mRNAs. The number of parameters, k , disregards the directly measured shot noise magnitudes and any parameter that was fixed at a large value (e.g., k_{out} when that values was fixed to 1000 s^{-1}). This leaves $k = 5$ parameters for the selected model: β , ω , k_{ab} , k_{esc} , and k_c . The Fractional Phosphorylation or Phosphorylation models each have one additional parameter, *fraction* or k_{phos} , respectively. The mRNA retention model has one additional parameter (i.e., k_c was replaced with $k_{c-\text{mRNA}}$, and k_{release} was added). The numbers of parameters, maximum likelihood values and parameter estimates, and BIC results for all examined models are listed in Fig 2.8. We note that our low conservative estimate for $n = 6$

(rather than basing n on the much larger number of independent experiments) was chosen to avoid biasing the model selection toward simpler models - larger choices of n would result in much stronger rejection of the more complex models.

Transcription inhibition experiments. For the transcription inhibition experiments in Fig 2.10, cells were imaged every 1 min for 5 time points before applying the inhibitor ($t=0$), TPL (5 μM), THZ1 (15 μM), or Flav (1 μM). Cells were then imaged every 1 min for 30 min total after addition of TPL or Flav, and for 55 min total after addition of THZ1. Here, laser power at the objective's back focal plane were set to 21.4 μW , 60.5 μW , and 21.74 μW for 488 nm, 561 nm, and 637 nm, respectively, and the exposure time was 53.64 ms.

To quantify time delays in the TPL-runoff assay, TPL signals were further analyzed as follows: (1) To account for cell variability and experimental conditions, the decays curves from each cell were aligned. This was achieved by subtracting the time at which each cell reached half of the decay after TPL addition. This time was obtained by an inverse hyperbolic tangent fit applied to each channel in every cell (Fig 2.10c); (2) After the alignment, all the traces in each channel were averaged together, and the standard error of the mean (S.E.M) was calculated. Finally, to determine the time delays between CTD-RNAP2, Ser5ph-RNAP2, and mRNA, an inverse tanh fit was applied, and weighted with respect to the variance of each signal.

Software All images were acquired with Micro-Manager software (1.4.22). Image preprocessing was made using ImageJ (2.0.0-rc-67/1.52e Java 1.8.0_66, 64-bit).

Images were analyzed with a custom Wolfram Mathematica (11.1.1) code. For the fast movies, tracking of the spots was performed using the ImageJ plugin, TrackMate (3.8.0). Final plots, modeling, and the GUI were made using MATLAB R2019b; Figures were assembled together using CorelDraw 2020 (64-bit).

Bibliography

- [1] Buratowski, S. Progression through the RNA polymerase II CTD cycle. *Mol. Cell* **36**, 541–546 (2009).
- [2] Schüller, R. et al. Heptad-specific phosphorylation of RNA polymerase II CTD. *Mol. Cell* **61**, 305–314 (2016).
- [3] Harlen, K .M. & Churchman, L. S. The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat Rev Mol Cell Biol.* **18**, 263–273 (2017).
- [4] Cramer, P. Organization and regulation of gene transcription. *Nature* **573**, 45–54 (2019).
- [5] Cissé, I. et al. Real-Time Dynamics of RNA Polymerase II Clustering in Live Human Cells. *Science* **341**, 664–667 (2013).
- [6] Cho, WK. et al. RNA Polymerase II cluster dynamics predict mRNA output in living cells. *Elife* **5**, e13617–1–31 (2016).
- [7] Boehning, M. et al. RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat. Struct. Mol. Biol.* **25**, 833–840 (2018).
- [8] Lu, H. et al. Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. *Nature* **558**, 318–323 (2018).
- [9] Nagashima, R. et al. Single nucleosome imaging reveals loose genome chromatin networks via active RNA polymerase II. *J. Cell Biol.* **218**, 1511–1530 (2019).
- [10] Sabari, B. R. et al. Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, eaar3958–1–11 (2018).
- [11] Heidemann, M., Hintermair, C., Voß, K. & Eick, D. Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *BBA Gene Regulatory Mechanisms* **1829**, 55–62 (2013).

- [12] Coulon, A., Chow, C. C., Singer, R. H. & Larson, D. R. Eukaryotic transcriptional dynamics: from single molecules to cell populations. *Nat. Rev. Genet.* **14**, 572–84 (2013).
- [13] Tokunaga, M., Imamoto, N., & Sakata-Sogawa, K. Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nat. Methods* **5**, 1–7 (2008).
- [14] Mazza, D., Abernathy, A., Golob, N., Morisaki, T. & McNally, J. G. A benchmark for chromatin binding measurements in live cells. *Nucleic acids res.* **40**, e119–1–13 (2012).
- [15] Chen, BC. et al. Lattice light-sheet microscopy: imaging molecules to embryos at high spatiotemporal resolution. *Science* **346**, 12579981–12 (2014).
- [16] Chen, J. et al. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* **156**, 1274–1285 (2014).
- [17] Li, J. et al. Single-Molecule Nanoscopy Elucidates RNA Polymerase II Transcription at Single Genes in Live Cells. *Cell* **178**, 1–16 (2019).
- [18] Steurer, B. et al. Live-cell analysis of endogenous GFP-RPB1 uncovers rapid turnover of initiating and promoter-paused RNA Polymerase II. *PNAS* **115**, E4368–E4376 (2018).
- [19] Bertrand, E. et al. Localization of ASH1 mRNA Particles in Living Yeast. *Mol. Cell* **2**, 437–445 (1998).
- [20] Tantale, K. et al. A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting. *Nat. Commun.* **7**, 1–14 (2016).
- [21] Larson, D., Zenklusen, D., Wu, B., Chao, J. & Singer, R. H. Real-time observation of transcription initiation and elongation on an endogenous yeast gene. *Science* **332**, 475–8 (2011).
- [22] Hayashi-Takanaka, Y., Yamagata, K., Nozaki, N. & Kimura, H. Visualizing histone modifications in living cells: Spatiotemporal dynamics of H3 phosphorylation during interphase. *J. Cell Biol.* **187**, 781–790 (2009).

- [23] Kimura, H., Hayashi-Takanaka, Y., Stasevich, T. J., & Sato, Y. Visualizing posttranslational and epigenetic modifications of endogenous proteins in vivo. *Histochem. Cell Biol.* **144**, 101–109 (2015).
- [24] Lyon, K. & Stasevich T. J. Imaging translational and post-translational gene regulatory dynamics in living cells with antibody-based probes. *Trends in Genetics* **33**, 322–335 (2017).
- [25] Conic, S. et al. Imaging of native transcription factors and histone phosphorylation at high resolution in live cells. *J. Cell Biol.* **217**, 1537–1552 (2018).
- [26] Sato, Y. et al. Histone H3K27 acetylation precedes active transcription during zebrafish zygotic genome activation as revealed by live-cell analysis. *Development* **146**, 1–10 (2019).
- [27] Stasevich, T. J. et al. Regulation of RNA polymerase II activation by histone acetylation in single living cells. *Nature* **516**, 272–275 (2014).
- [28] Taube, R., Lin, X., Irwin, D., Fujinaga, K. & Peterlin, B. M. P-TEFb regulation of transcription termination factor Xrn2 revealed by a chemical genetic screen for Cdk9 substrates. *Mol. Cell. Biol.* **22**, 321–331 (2002).
- [29] Hayashi-Takanaka, Y. et al. Tracking epigenetic histone modifications in single cells using Fab-based live endogenous modification labeling. *Nucleic Acids Res.* **39**, 6475–6488 (2011).
- [30] Nojima, T. et al. Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* **161**, 526–540 (2015).
- [31] Cho, WK. et al. Super-resolution imaging of fluorescently labeled, endogenous RNA Polymerase II in living cells with CRISPR/Cas9-mediated gene editing. *Sci. Rep.* **6**, 1–8 (2016).
- [32] Zaborowska, J., Egloff, S., & Murphy, S. The pol II CTD: new twists in the tail. *Nat. Struct. Mol. Biol.* **23**, 771–777 (2016).
- [33] Morisaki, T. et al. Real-time quantification of single RNA translation dynamics in living cells. *Science* **352**, 1425–9 (2016).

- [34] Coulon, A. et al. Kinetic competition during the transcription cycle results in stochastic RNA processing. *Elife* **3**, 1–22 (2014).
- [35] Coulon, A. & Larson, D.R. *Meth. Enzymol. Fluctuations Analysis: Dissecting Transcriptional Kinetics with Signal Theory*, chap. 7 (New York, Academic Press, New York, 2016).
- [36] Bacia, K., Kim, S. & Schwille S. Fluorescence cross-correlation spectroscopy in living cells. *Nat. Methods* **3**, 83–89 (2006).
- [37] Müller, F. et al. FISH-quant: automatic counting of transcripts in 3D FISH images. *Nat. Methods* **10**, 277–278 (2013).
- [38] Munsky, B., Neuert, G. & Van Oudenaarden, A. Using Gene Expression Noise to Understand Gene Regulation. *Science* **336**, 183–187 (2012).
- [39] Kumar, N., Singh, A., Kulkarni, R. V. Transcriptional Bursting in Gene Expression: Analytical Results for General Stochastic Models. *PLoS Comput. Biol* **11**, e1004292 (2015).
- [40] Brody, Y. et al. The In Vivo Kinetics of RNA Polymerase II Elongation during Co-Transcriptional Splicing. *PLOS Biol.* **9**, e1000573 (2011).
- [41] Titov, D. V. et al. XPB, a subunit of TFIIH, is a target of the natural product triptolide. *Nat. Chem. Biol.* **7**, 182–8 (2011).
- [42] Wang, Y., Lu, J-J., He, L., & Yu, Q. Triptolide (TPL) Inhibits Global Transcription by Inducing Proteasome-Dependent Degradation of RNA Polymerase II (Pol II). *PLoS One* **6**, e23993 (2011).
- [43] Kwiatkowski, N. et al. Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. *Nature* **511**, 616–620 (2014).
- [44] Chao, SH. et al. Flavopiridol Inhibits P-TEFb and Blocks HIV-1 Replication. *J. Biol. Chem.* **275**, 28345–28348 (2000).

- [45] Cho, WK. et al. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412–415 (2018).
- [46] Edelman, L. B. & Fraser, P. Transcription factories: genetic programming in three dimensions. *Curr. Opin. Genet. Dev.* **22**, 110–114 (2012).
- [47] Feuerborn, A. & Cook, P. R. Why the activity of a gene depends on its neighbors. *Trends Genet.* **31**, 483–490 (2015).
- [48] Tahirov, T. H. et al. Crystal structure of HIV-1 Tat complexed with human P-TEFb. *Nature* **465**, 747–751 (2010).
- [49] Darzacq, X. et al. In vivo dynamics of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **14**, 796–806 (2007).
- [50] Guo, Y. E., et al. Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* **572**, 543–548 (2019).
- [51] Barboric, M. & Peterlin, B. M. A new paradigm in eukaryotic biology: HIV Tat and the control of transcriptional elongation. *PLoS biology* **3**, 200–203 (2005).
- [52] Lionnet, T. & Singer, R. H. Transcription goes digital. *EMBO Rep.* **13**, 313–321 (2012).
- [53] Core, L., & Adelman, K. Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev.* **33**, 960–982 (2019).
- [54] Adelman, K. & Lis, J. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet* **13**, 720–731 (2012).
- [55] Gu, B., et al. Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. *Science* **359**, 1050–1055 (2018).
- [56] Straight, A., Belmont, A. S., Robinett, C. C., & Murray, A. W. GFP tagging of budding yeast chromosomes reveals that protein–protein interactions can mediate sister chromatid cohesion. *Curr. Biol.* **6**, 1599–1608 (1996).

- [57] Viollier, P. et al. Rapid and sequential movement of individual chromosomal loci to specific subcellular locations during bacterial DNA replication. *PNAS* **101**, 9257–9262 (2004).
- [58] Ochiai, H., Sugawara, T. & Yamamoto, T. Simultaneous live imaging of the transcription and nuclear position of specific genes. *Nucleic Acids Res.* **43**, e127–1–12 (2015).
- [59] Mariamé, B. et al. Real-time visualization and quantification of human Cytomegalovirus replication in living cells using the ANCHOR DNA labeling technology. *J. Virol.* **92**, e00571–18 (2018).
- [60] Takei, Y. et al. Multiplexed dynamic imaging of genomic loci by combined CRISPR imaging and DNA sequential FISH. *Biophys. J.* **112**, 1773–1776 (2017).
- [61] Deng, W. et al. CASFISH: CRISPR/Cas9-mediated in situ labeling of genomic loci in fixed cells. *PNAS* **112**, 1870–11875 (2015).
- [62] Sato, Y. et al. Genetically encoded system to track histone modification in vivo. *Sci. Rep.* **3**, 1–7 (2013).
- [63] Kimura, H., Tao, Y., Roeder, R.G., Cook, P.R. Quantitation of RNA Polymerase II and Its Transcription Factors in an HeLa Cell: Little Soluble Holoenzyme but Significant Amounts of Polymerases Attached to the Nuclear Substructure. *Mol. Cell. Biol.* **19**, 5383–5392 (1999).
- [64] Rothbauer, u. et al. Targeting and tracing antigens in live cells with fluorescent nanobodies. *Nat Methods* **3**, 887–889 (2006).
- [65] Kimura, H., Hayashi-Takanaka, Y., Goto, Y., Takizawa, N., & Nozald, N. The organization of histone H3 modifications as revealed by a panel of specific monoclonal antibodies. *Cell Struc. Funct.* **33**, 61–73 (2008).
- [66] McNeil, P. L. & Warder, E. Glass beads load macromolecules into living cells. *J. Cell Sci.* **88**, 669–78 (1987).

- [67] Manders, E. M. M., Kimura, H. & Cook, P. R. Direct imaging of DNA in living cells reveals the dynamics of chromosome formation. *J. Cell Biol.* **144**, 813–821 (1999).
- [68] Edelstein, A. et al. Advanced methods of microscope control using microManager software. *J. Biol. Methods.* **1**, 1–18 (2014).
- [69] Carlini, L. et al. Reduced dyes enhance single-molecule localization density for live super-resolution imaging. *ChemPhysChem* **15**, 750–755 (2014).
- [70] Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–82 (2012).
- [71] Aguilera, LU. et al. Computational design and interpretation of single-RNA translation experiments. *PLoS Comput Biol.* **15**, 1–27 (2019).

Table 2.1: Derived Quantities and Confidence Intervals Resulting from Model Fit to Data

Derived Quantity	Formula	Value (95% CI Range)
Average RNAP2 burst size	β	15.4 (11.76-61.83)
Average RNAP2 burst frequency	ω	0.43 (0.33-0.66) min^{-1}
Average RNAP2 arrival rate	$r = \omega \cdot \beta$	6.622 (5.74-27.43) min^{-1}
Average time RNAP2 spends in cluster	$\tau_{\text{Cluster}} = 1/(k_{\text{esc}} + k_{\text{ab}})$	0.692 (0.02-0.79) min
Average RNAP2 in cluster	$\mu_{\text{Cluster}} = r/(k_{\text{esc}} + k_{\text{ab}})$	4.624 (0.46-5.26)
Probability of RNAP2 in cluster to escape	$f = k_{\text{esc}}/(k_{\text{esc}} + k_{\text{ab}})$	0.4615 (0.10-0.56)
Average mRNA burst size	$\beta_{\text{mRNA}} = f \cdot \beta$	7.11 (4.90-8.26)
Average mRNA production rate	$r_{\text{mRNA}} = \mu_{\text{Cluster}} \cdot k_{\text{esc}}$	3.083 (2.49-3.52) min^{-1}
Average actively transcribing RNAP2	$\mu_{\text{mRNA}} = \mu_{\text{Cluster}} \cdot k_{\text{esc}}/k_{\text{c}}$	15.512 (13.69-17.34)
Average total RNAP2 at transcription site	$\mu_{\text{Total}} = \mu_{\text{Cluster}} + \mu_{\text{mRNA}}$	20.136 (14.62-21.28)
Average time for mRNA completion	$\tau_{\text{mRNA}} = 1/k_{\text{c}}$	5.032 (4.59-5.91) min

Table 2.2: List of primers for ChIP-qPCR as shown in Figure 1. All primer sequences are 5' to 3'.

Name	Forward Primer	Reverse Primer
Up	GGATGACCCGGAGAGAGAAGTG	AAGCAGCTGCTTATATGCAGG
TSS	GCATATAAGCAGCTGCTTTTTGCC	TCCCTGTTCGGGCGCCACT
Pt3	AAGGGAAACCAGAGGAGCTCTC	CCCATCTCTCTCCTTCTAGCCTC
Pt4	CCATCCCTTCAGACAGGATCAGAAG	GCTCTTCCTCTATCTTGTCTAAAGCTTCC
Pt5	GGAGGAGATATGAGGGACAATTGGA	AAGGAACAAAGCTCCTATTCCCACT
Pt6	TGTCTGGTATAGTGCAGCAGCAG	GCTGTTGATCCTTTAGGTATCTTTCCAC
Pt7	CTGTGCCTTGGAATGCTAGTTGGA	TTCTTGCTGGTTTTGCGATTCTTCA
Pt8	TGGGCAAGTTTGTGGAATTGGT	ATGGTGAATATCCCTGCCTAACTCT
Pt9	CAGGCCCGAAGGAATAGAAGAAGA	ATTGCTACTTGTGATTGCTCCATGT
Pt10	GACCAATGACTTACAAGGCAGCT	CCCTGGTGTGTAGTTCTGCCA

Chapter 3

The relationship between histone acetylation, transcription initiation, and chromatin dynamics, imaged in living-cells

3.1 Summary

Post-translational protein modifications play an important role in the regulation of gene dynamics. Certain modifications, such as histone acetylation and RNA polymerase II phosphorylation, are associated with transcriptionally active chromatin. However, the spatial and temporal relationship between chromatin and post-translational protein modifications, and how these dynamics facilitate selective gene expression, remain poorly understood. In this study, we address this problem by developing a general methodology for quantifying in live cells the dynamics of chromatin across multiple time and length scales in the context of residue-specific protein modifications. By combining Fab-based labeling of endogenous protein modifications with single-molecule imaging, we track the dynamics of chromatin enriched with histone H3 Lysine-27 acetylation (H3K27ac) and RNA polymerase II Serine-5 phosphorylation (RNAP2-Ser5ph). Our analysis reveals chromatin enriched with H3K27ac is separated from chromatin enriched with RNAP2-Ser5ph. Furthermore, in these separated sites, we show the presence of the two modifications are inversely correlated with one another on the minutes timescale. We then track single nucleosomes in both types of sites on the sub-second timescale and again find evidence for distinct and opposing changes in their diffusive behavior. While nucleosomes diffuse 15% faster in chromatin enriched with H3K27ac, they diffuse 15% slower in chromatin enriched with RNAP2-Ser5ph. Taken together, these results argue that high levels of H3K27ac and RNAP2-Ser5ph are not often present together at the same

place and time, but rather each modification marks distinct sites that are transcriptionally poised or active, respectively.²

3.2 Introduction

Transcription in eukaryotic cells occurs in the context of nucleosomes, the fundamental unit of chromatin. Nucleosomes are composed of a strand of DNA wrapped around an octet of core histones [1]. This wrapped configuration provides an additional layer of gene regulation as it limits the accessibility of DNA to transcription machinery [2]. To make genes more or less accessible, the tails of histones are post-translationally modified at a large number of unique residues. These post-translational modifications (PTMs) which include acetylation, methylation, ubiquitination, and phosphorylation, among others change the local chemical environment of chromatin to alter the ease by which nucleosomes can slide across DNA and interact with one another [3,4]. They also create binding substrates for chromatin remodelers or transcription factors. Collectively, histone PTMs are thought to create a code or language that fine-tunes chromatin spatio-temporal dynamics [35].

Besides histones, the transcription machinery itself is also subject to a variety of PTMs that regulate gene expression. A critical example is the phosphorylation of RNA polymerase II (RNAP2) the complex responsible for the transcription of the vast majority of protein-coding genes. RNAP2 is phosphorylated at key residues along a conserved heptad sequence that is repeated (52 times in humans) within the C-terminus of the large, catalytic RPB1 subunit. Phosphorylation occurs sequentially and is thought to delineate major events in the RNAP2 transcription cycle [69]. In general, similar to histone PTMs, RNAP2-specific PTMs create binding substrates to recruit specific factors that facilitate transcription. As an additional regulatory mechanism, RNAP2 phosphorylation alters the local chemical environment, leading to functionally distinct clusters [1012].

²The following chapter is based upon work in which I am first author, currently under review in Science Advances. Author Contributions: Conceptualization: M.S. and T.J.S. Performed experiments / collected data: M.S. and T.M. Antibodies and Fab Preparation: M.S. and H.K. Cell Culture: M.S. Software development and implementation: M.S., D.K., T.M., T.J.S. Diffusion Analysis: M.S. and D.K. Original draft: M.S. and T.J.S. Review and edits of drafts: M.S., T.M., D.K., H.K. and T.J.S. Resources, supervision and funding: T.J.S.

Histone and RNAP2 PTMs are tightly coupled during gene activation. Histone acetylation at Lysine 27 (H3K27ac), in particular, is strongly correlated with RNAP2 Serine 5 phosphorylation (RNAP2-Ser5ph), a marker of transcription initiation and pausing at promoters and enhancers [13,14]. A wealth of classic biofractionation, immunoprecipitation, and structural studies have led to a standard textbook model for this correlation [3,15-17]. In the model, acetylation neutralizes the positive histone charge and reduces hydrogen bond formation. These processes weaken interactions with negatively charged DNA and other nucleosomes, so individual nucleosomes can move more easily. Acetylation also creates binding substrates for chromatin-remodelers and other trans-factors with acetyl-binding domains [18]. The end result is decondensed, loose, and more mobile chromatin where enhancers and promoters can more easily come into contact and RNAP2 can more easily be recruited and initiated for efficient transcription. Super-resolution imaging of histones in intact, fixed cells further support this model of chromatin decondensation by histone acetylation [19], and numerous genome-wide studies have confirmed acetylation is co-enriched with RNAP2-Ser5ph near gene and enhancer transcription start sites [7,14].

Although the textbook model provides a clear snapshot of the impact histone acetylation has on chromatin and transcription, the dynamics underlying the model remain poorly understood and have not been validated. This gap arises because PTM dynamics are hard to capture experimentally. Histone acetylation is a dynamic mark whose levels can rapidly change through the fine tuning of the activities of lysine deacetylases (KDAC) and acetyltransferases (KHAT) [15,20]. RNAP2 phosphorylation also rapidly changes as RNAP2 progresses through the various stages of the transcription cycle [7,8]. Resolving the spatiotemporal relationships between highly dynamic acetylation and phosphorylation is therefore especially challenging. Current knowledge of PTM relationships has therefore relied on fixed cell assays, including chromatin immunoprecipitation, immunostaining, and western blotting. However, in fixed-cell experiments, the dynamics are blurred by the experimental necessity of fixation (which mixes signals in time) and, in many cases, cell population averaging (which mixes signals in space).

To better resolve PTM dynamics, several live-cell imaging techniques have been developed over the last decade. First, we and others have developed technology to directly image and quantify endogenous histone and RNAP2 PTMs in living cells [21-26]. In these studies, fluorescent antibody-based probes are used to rapidly bind and light up residue-specific PTMs in distinct colors. This technology makes it possible to record the temporal fluctuations of the various modifications and better understand their dynamic relationships. Second, individual genetic loci, genes, and nucleosomes can now be tracked in living cells, making it possible to directly measure their mobilities. For example, DNA fluorescence amplification tags can be used to track single genetic loci [27-30], RNA fluorescence amplification tags can be used to track single-gene transcription³¹, and both DNA and RNA amplification tags can be combined [32-34]. More recently, advances in single-molecule fluorescence microscopy [35,36] and fluorescent dye development [37] have made it possible to track and quantify the mobility of single nucleosomes [35,38-42]. Together, these new technologies are revealing a high degree of heterogeneity in chromatin mobility, both within and across cells.

Of the various live-cell imaging studies to date, discrepancies have begun to emerge that raise several fundamental questions about the dynamic relationship between chromatin, gene activation, and histone acetylation. Although there seems to be consensus that dense, constitutive heterochromatin is less mobile than average [38,39,41,43], the mobility of less dense euchromatin remains debatable. According to the transcription factory [44,45] or hub [46] model, the transcription machinery is thought to be relatively immobile, causing nearby chromatin to be constrained or anchored. Recent support for this model comes from several studies that tracked specific genetic loci in living cells and showed their mobility goes down when associated with transcription [32-34,47]. Additional support comes from single-nucleosome tracking experiments in which a variety of global inhibitors and perturbations of transcription were used to show transcription generally confines single-nucleosome movements [38,48]. Nevertheless, direct imaging of RNAP2 [49] and other components of transcriptional hubs [50,51] has revealed they are highly dynamic and transient structures, leaving it questionable to what degree they can immobilize chromatin.

Furthermore, some genes and regulatory elements near genes have been observed to become more dynamic upon transcription, providing support for an opposing model whereby transcription stirs up chromatin [52], in opposition to the factory concept. Finally, the same single-nucleosome tracking studies that showed transcription slows down chromatin also showed generalized acetylation speeds it up [38]. These results lead to a paradox: if histone acetylation and transcription initiation act together at the same place and time in the nucleus, as the standard textbook model predicts, how can active chromatin be both more and less mobile?

Here we directly confront this conundrum by combining single molecule tracking [35,36] with Fab-based imaging of live-endogenous modifications (FabLEM) [21,22,24,53]. The combination of these technologies allow us to simultaneously label and directly monitor (1) chromatin dynamics, (2) transcription initiation (RNAP2-Ser5ph), and (3) histone H3 Lysine 27 acetylation (H3K27ac), all without the use of broad-acting inhibitors or perturbations. Using this approach, we show that chromatin enriched with H3K27ac is for the most part physically separated from chromatin enriched with RNAP2-Ser5ph. We furthermore show that the two modifications localize to regions that are oppositely correlated with one another and the mobilities of nucleosomes therein are substantially different. Taken together, our data support a model wherein histone acetylation and transcription initiation are enriched in functionally separate chromatin regions with distinct physical behavior.

3.3 Results

3.3.1 Monitoring global chromatin dynamics in the context of histone acetylation and RNAP2 phosphorylation

To quantitatively explore the relationship between histone acetylation, RNAP2 phosphorylation, and chromatin dynamics, we first created an RPE1 cell line that stably expresses Halo-tagged histone H2B (H2B-Halo; Fig. 3.1A). With this cell line, we investigated the dynamics of chromatin at multiple time and length scales depending on the concentration of added Halo ligand. At higher concentrations, we could label a significant fraction of H2B, allowing us to measure the dynamics

of subcellular chromatin regions on the minutes to hours timescale (Fig. 3.1B); At lower concentrations, we could label a tiny subset of H2B, allowing us to measure the dynamics of individual nucleosomes on the sub-second to minutes timescale (Fig. 3.1C,D).

To place these measurements within the local context of histone acetylation and RNAP2 phosphorylation, we used FabLEM23 to co-image endogenous histone H3 Lysine 27 acetylation (H3K27ac) - marking active genes and enhancer DNA [4,54-56] - and endogenous RNAP2 Serine 5 phosphorylation (RNAP2-Ser5ph) - marking transcription initiation sites [8]. We were motivated by our previous experiments in which we imaged the same modifications at an artificial tandem gene array activated by the synthetic hormone dexamethasone [22]. One of our main findings was that H3K27ac can facilitate gene activation and chromatin decondensation, leading to more efficient RNAP2 recruitment and promoter escape. Due to the artificial nature of the tandem gene array, however, we worried about the generalizability of our findings and wondered if more natural and less repetitive chromatin would behave the same.

To directly test this hypothesis, we co-loaded our stable H2B-Halo cells with the same Fab (fragmented antigen binding regions) we used in the earlier study and performed 3-color confocal time lapse microscopy. As illustrated in Fig. 3.2A, we imaged endogenous H3K27ac using an AF488-conjugated Fab (green), endogenous RNAP2-Ser5ph using a CF640-conjugated Fab (purple), and H2B-Halo using JFX-554-labeled Halo ligand57 (orange). Unlike standard fluorescent protein fusion tags like GFP, which are permanently attached, Fabs only bind their targets transiently. This reduces interference with the underlying biology and allows Fabs to rapidly respond to changes in the post-translational protein modification landscape [24,53]. To further minimize interference, we kept the concentration of Fab relatively low in cells so target modifications were far from saturated, as we had in our earlier study [22]. By imaging cells every 2-3 minutes for 100 time points total (representing 200 or 300 minutes, respectively), we generated a complementary pair of datasets that we could mine to precisely quantify the spatiotemporal relationship between H3K27ac and RNAP2-Ser5ph across the entire nucleus (Fig. 3.2B, Sup. Movie 1).

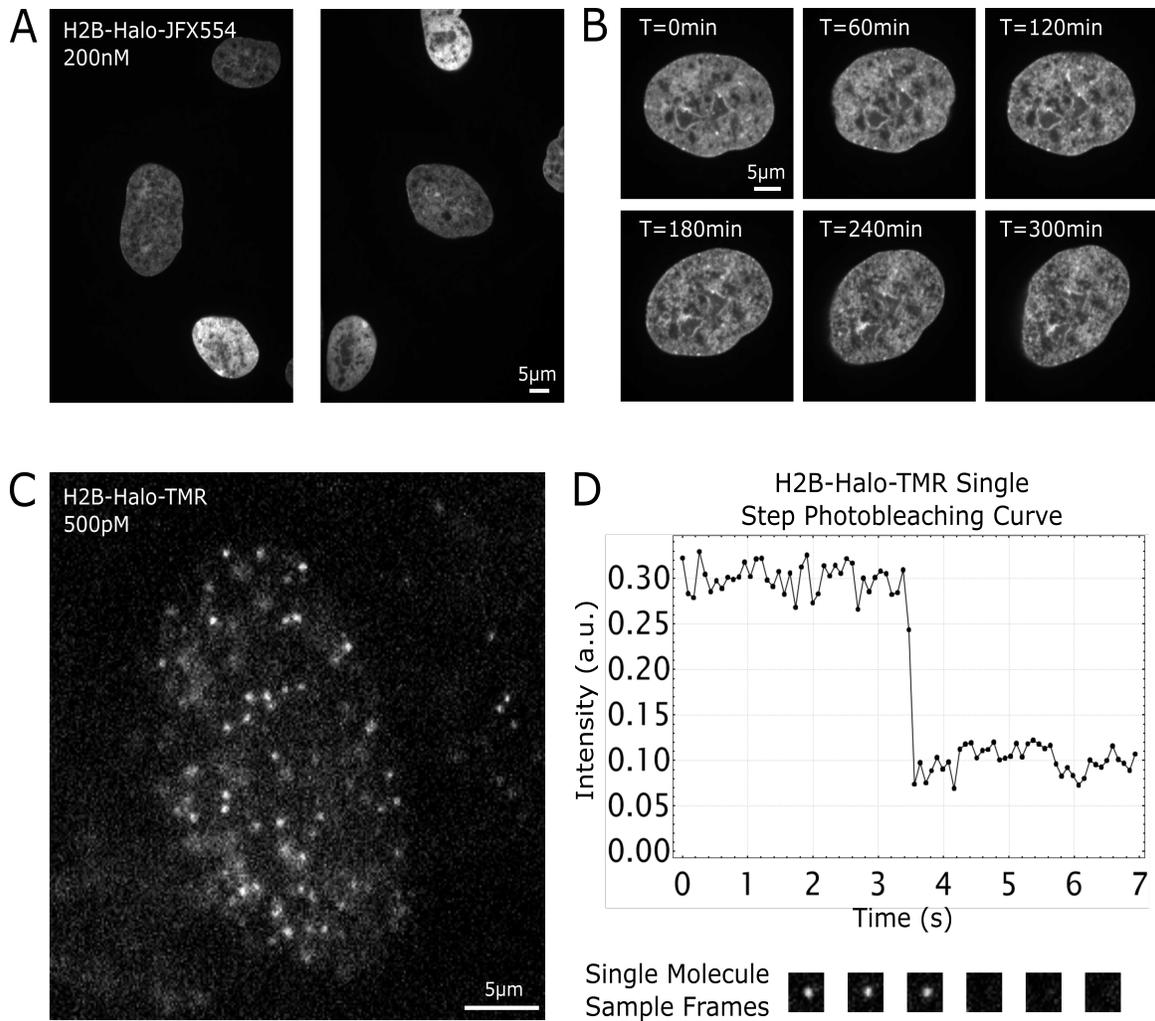


Fig. 3.1: H2B-Halo Stable RPE1 Cell Line Allows For Tracking Chromatin at Multiple Timescales. (a) Sample images of multiple RPE1 cells in a single frame, stably expressing H2B-Halo, stained with Halo-JFX554 dye at 200nM concentration. (b) Single cell H2B intensity and dynamics can be tracked for hours without significant photobleaching (c) Same RPE1 cells as in A), except stained with 500pM Halo-TMR. Now, single nucleosomes can be identified and tracked in real time with sub-second resolution. (d) Single-step photobleaching curve for a sample H2B track, along with frames of a tracked H2B before and after photobleach occurrence.

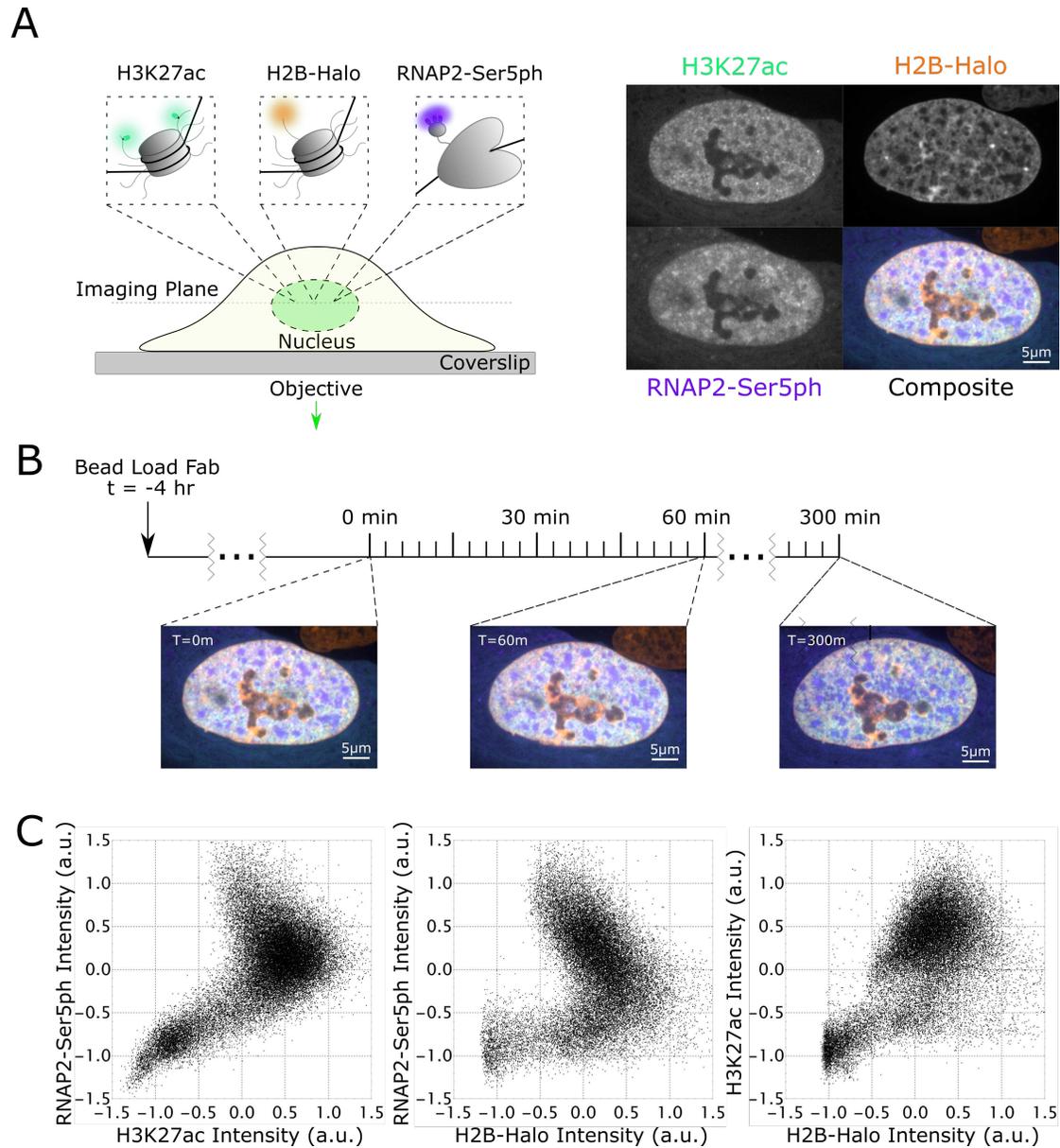


Fig. 3.2: Simultaneous imaging of RNAP2-Ser5ph, H2B, and H3K27ac in living RPE1 cells. (a) Left) Schematic of imaging system. H2B-Halo is stably expressed in RPE1 cells and stained with Halo-JFX-554 ligand. AF488-H3K27ac and CF640-RNAP2-Ser5ph specific Fab are beadloaded into cells, marking each modification in real-time. Right) Sample images and composite for all channels. (b) Experimental time course. Cells are beadloaded with Fab 4 hours prior to imaging, and then H3K27ac, RNAP2-Ser5ph, and H2B channels are collected in a single axial z-plane every 3 minutes for 300 minutes. (c) Scatterplots of renormalized H3K27ac, RNAP2-Ser5ph and H2B signal intensities. Each plotted point is the paired intensity values of the same pixel inside of a single nucleus at a single time point, with 3 representative plots chosen, one for each pairing of channels.

To begin to quantify these data in a simple manner, we asked to what degree the H3K27ac and RNAP2-Ser5ph signals were spatially correlated. On the one hand, standard immunoprecipitation assays have long observed a strong positive correlation between histone acetylation and transcription [15,58]. According to this body of work, genes with higher H3K27ac levels should have higher RNAP2-Ser5ph levels [4]. On the other hand, immunoprecipitation assays are typically performed in a fixed population of cells, so data represent an average cellular state. Thus, it is unclear if two modifications are actually present at a single site at the same time in individual cells. For example, it could be feasible that histone acetylation precedes RNAP2 phosphorylation, as we observed at the tandem gene array [22]. In this case, the two signals could be spatially anticorrelated at any given time, i.e., they would be spatially segregated.

To test if our data supported either of these two scenarios, we generated scatterplots from the signal intensities found within the nuclear pixels of all imaged cells. Here, intensity is rescaled from -1 to 1, with $I_{RS} = (I - \langle I \rangle) / (I_{97.5} - I_{2.5})$, where $\langle I \rangle$ is the average intensity and $I_{97.5}$ and $I_{2.5}$ are the 97.5 and 2.5 intensity quantiles, respectively (Fig. 3.2C). This analysis revealed a complex relationship between the various signals. For example, while the signals were overall positively correlated, especially the H3K27ac and H2B signals (Fig. 3.2C, right), the correlations were not linear. In particular, in the H3K27ac versus RNAP2-Ser5ph scatterplot (Fig. 3.2C, left), we observed a negative correlation in the upper-right quadrant, where both signals are relatively enriched. The H2B and RNAP2-Ser5ph had a strong negative correlation, consistent with dechromatinization being coupled to transcription activity (Fig. 3.2C, middle). Thus, there is a mixture of distinct correlations between signals within our dataset, some positive and some negative. This would suggest the relationship between chromatin, histone acetylation, and RNAP2 phosphorylation is heterogenous and most likely context dependent.

3.3.2 Tracking histone acetylation and RNAP2 phosphorylation dynamics at thousands of endogenous chromatin sites reveals two distinct dynamics

To better assess the complex relationships we saw emerging in our scatterplots, we next zoomed in on individual chromatin sites to track signals at those sites through time. To achieve this, we created an image processing pipeline consisting of 10 steps (Fig. 3.3A). Briefly, after averaging over five frames and correcting for cell movement (Fig. 3.3B,C), we used a local adaptive binarization filter to mask, identify, and track thousands of individual chromatin sites that were enriched with either H3K27ac or RNAP2-Ser5ph (Fig. 3.3D-F). An example is shown in Fig. 3.5A, where two individual sites (of many) from a sample cell are tracked through time, the first enriched with H3K27ac (green) and the second enriched with RNAP2-Ser5ph (purple).

In total, we tracked 15,295 non-specific endogenous chromatin sites spread across the nuclei of 27 cells, 10,290 enriched with H3K27ac and 5,005 enriched with RNAP2-Ser5ph (Fig. 3.4). The sites varied in size, with a median cross sectional area of $1.12 \pm 0.96 \mu\text{m}^2$ (SD) for HK27ac-enriched sites and $0.92 \pm 0.70 \mu\text{m}^2$ (SD) for RNAP2-Ser5ph-enriched sites (Fig. 3.4C). These domains were too big to be single TADs [59,60] or clutches [19], but rather more on the order of chromatin territories [61]. Consistent with the negative correlation we saw in the upper quadrant of our RNAP2-Ser5Ph vs H3K27ac scatterplot, the two sites tended to border one another but were practically independent, with an overlap that occupied 5 percent of the nucleus or less (see, for example, the combined masks and quantification in Fig. 3.4A,B). In other words, the data confirmed regions of chromatin most enriched with H3K27ac had only moderate amounts of RNAP2-Ser5ph, and vice versa (Fig. 3.4D).

Given the spatially independent nature of chromatin sites enriched with H3K27ac or RNAP2-Ser5ph, we next wondered if the two types of sites displayed different dynamics. To address this question, we quantified the rescaled fluorescence intensities of each track through time (Fig. 3.5B,C). Analyzing these tracks in detail revealed a heterogeneous track-to-track population with a number of interesting behaviors and trends. First, many regions of chromatin maintained their

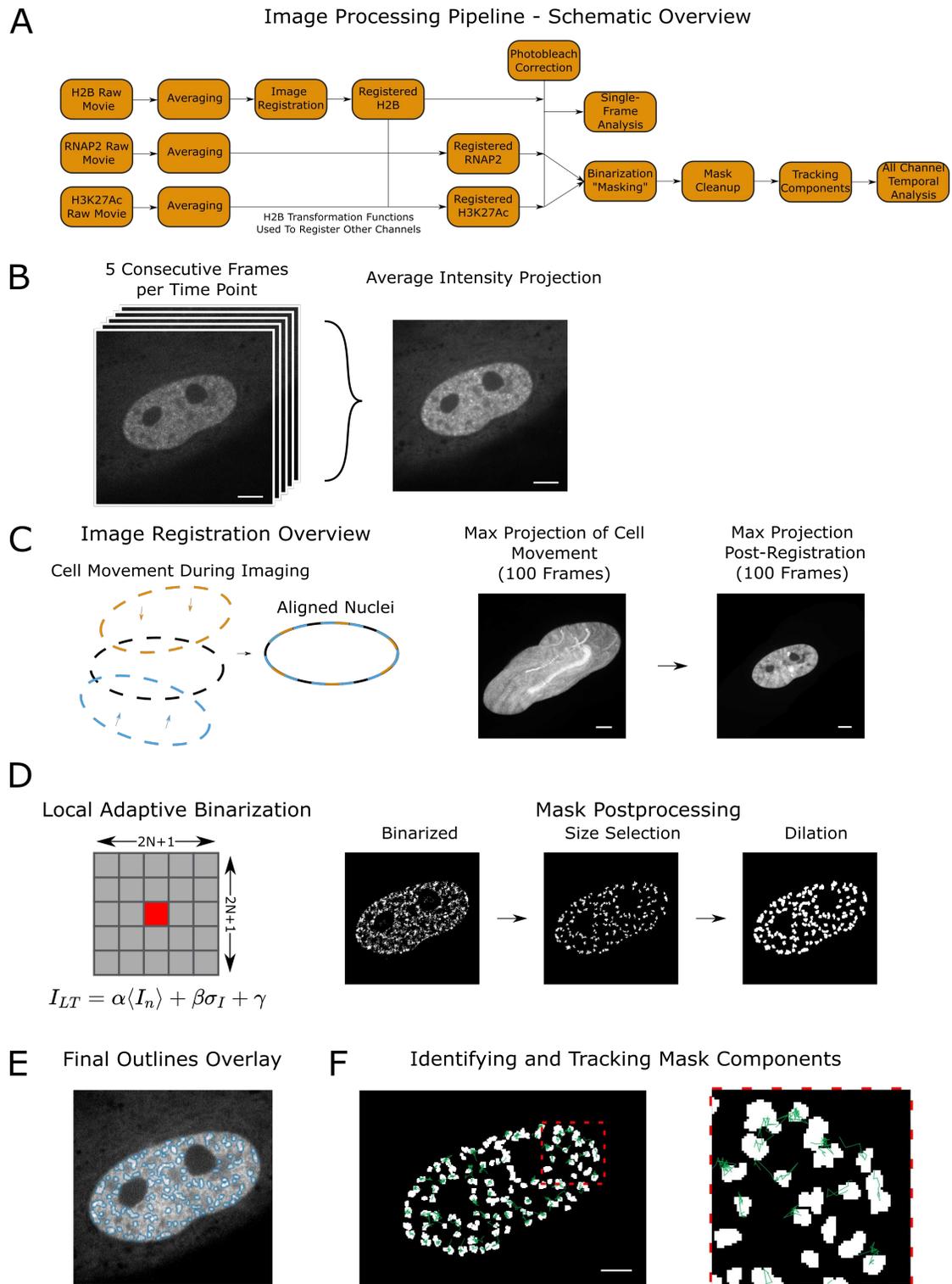


Fig. 3.3: Analysis pipeline for confocal imaging of RNAP2-Ser5ph, H2B, and H3K27ac in living RPE1 cells. (Continued on next page)

Fig. 3.3: Analysis pipeline for confocal imaging of RNAP2-Ser5ph, H2B, and H3K27ac in living RPE1 cells. (a) Schematic overview of the analysis pipeline used to process confocal images and track regions of H3K27ac and RNAP2-Ser5ph. (b) Five frames are collected per time point and averaged to reduce noise from freely diffusing background Fab. (c) Each cell's nucleus is individually registered to the H2B channel to correct for any cell movement during imaging. Pre- and Post-registration max projections of an example cell are shown. (d) Local adaptive binarization is used to identify regions of a nucleus enriched for H3K27ac or RNAP2-Ser5ph. For additional information see materials and methods. The same binarization parameters are applied to all cells in all cases. After binarization, components less than 10 pixels are removed to reduce speckle noise, and then dilated by a single pixel to fill in gaps and smooth edges. (e) An outline of final binarization components (blue) overlaid on the original Fab image. (f) Demonstration of tracking individual H3K27ac chromatin regions through time. Components with a line trace (green) show the detected regions centroid movement over time. Components without a shown line trace did not meet the minimum requirement of 10 frame length for tracking and were discarded.

enriched status (in either H3K27ac or RNAP2-Ser5ph) for half an hour or longer, with only minor fluctuations (Fig. 3.3C, bottom). When fluctuations did occur inside of H3K27ac enriched areas, we noticed that all three signals tended to shift in sync (Fig. 3.5B). Conversely, inside of RNAP2-Ser5ph enriched chromatin, increases in transcription signal tended to be accompanied by decreases in the H3K27ac and H2B signals and vice versa (Fig. 3.5C).

To assess the generalizability of these trends, we aligned all H3K27ac and RNAP2-Ser5ph signal peaks in both sets of tracks (see small arrows in Fig. 3.5B,C for a few sample peaks). We then examined the behavior of all signals when one signal peaked. For example, how H2B and RNAP2-Ser5ph responded when H3K27ac reached a local maximum. As expected, this revealed the dynamics of H3K27ac and RNAP-Ser5ph differed in the two types of sites. In H3K27ac-enriched sites, local peaks in one modification were predictive of small local peaks in the other modification, whereas in RNAP2-Ser5ph enriched sites, local peaks in one modification were predictive of small local troughs in the other modification (Fig. 3.5D,E). This opposing behavior, while subtle, was consistent across experimental replicates (Fig. 3.6A,B) and independent of whether we aligned signals by peaks (Fig. 3.5D,E, Sup. Fig 4A,B) or troughs (Fig. 3.6A,B). To further confirm the opposing behavior, we calculated the cross-correlation between the two modifications using signals from all tracks at all timepoints. In line with our peak alignment results, this demonstrated the

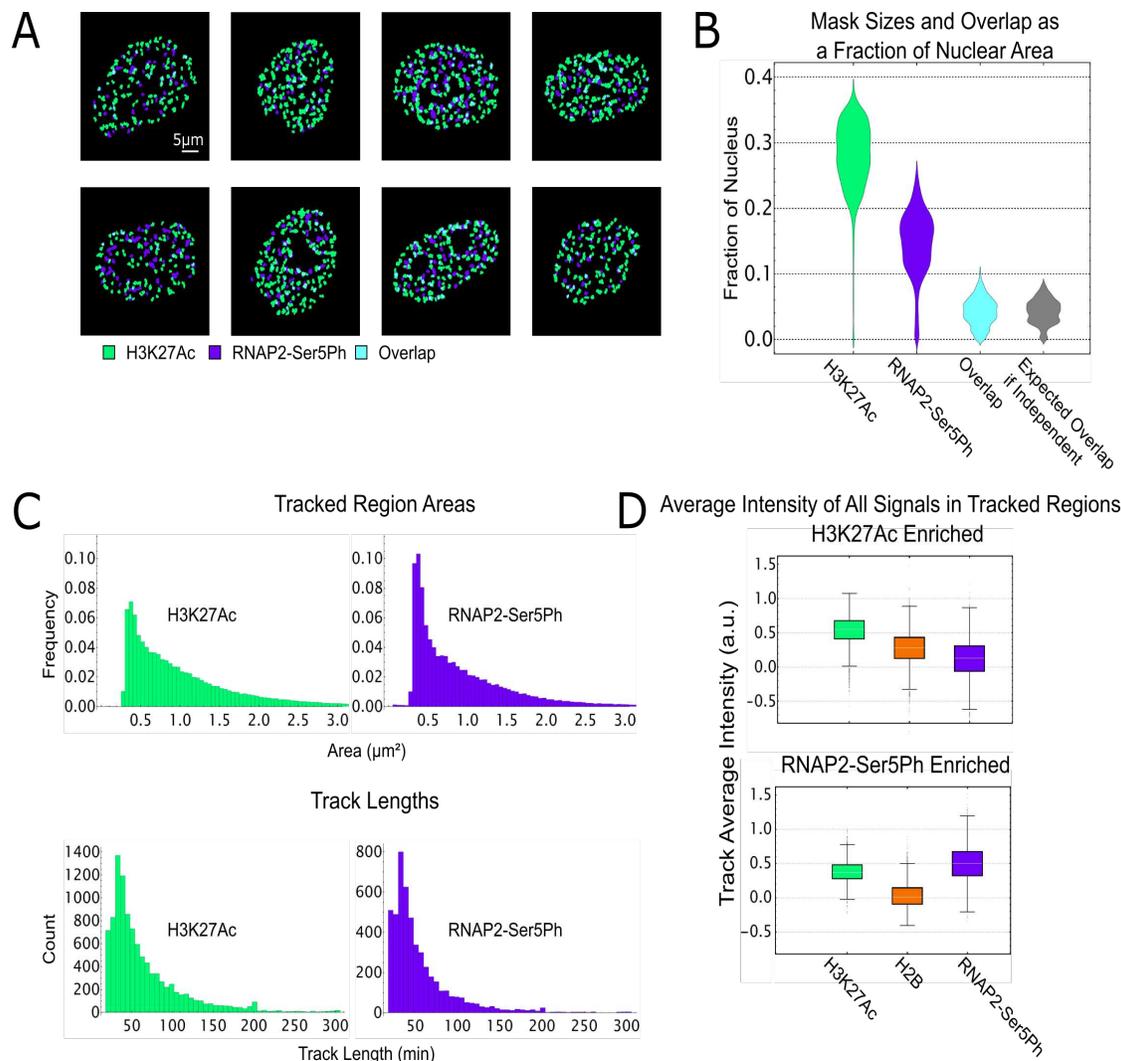


Fig. 3.4: Characterization of H3K27ac and RNAP2-Ser5ph Regions: Overlap, Nuclear Area, and Intensity. (a) Sample images from 8 cells demonstrating the location and overlap between H3K27ac and RNAP2-Ser5ph masks. (b) Distribution of mask area as a fraction of nuclear area, with the overlap between masks also shown. Expected overlap if independent represents the amount of expected overlap if each masks area were independent and uniformly distributed throughout the nucleus. $n=27$ cells, 2700 masks. (c) Top) Histogram of area of detected regions for both H3K27ac and RNAP2-Ser5ph masks. $n=228,610$ regions for H3K27ac, $n=143,854$ regions for RNAP2-Ser5ph. Bottom) Histogram of track lengths for both H3K27ac and RNAP2-Ser5ph regions. (d) Distribution of average intensities in all signals for all H3K27ac and RNAP2-Ser5ph tracks (i.e. each tracks signal is averaged, and contributes one data point to the distribution).

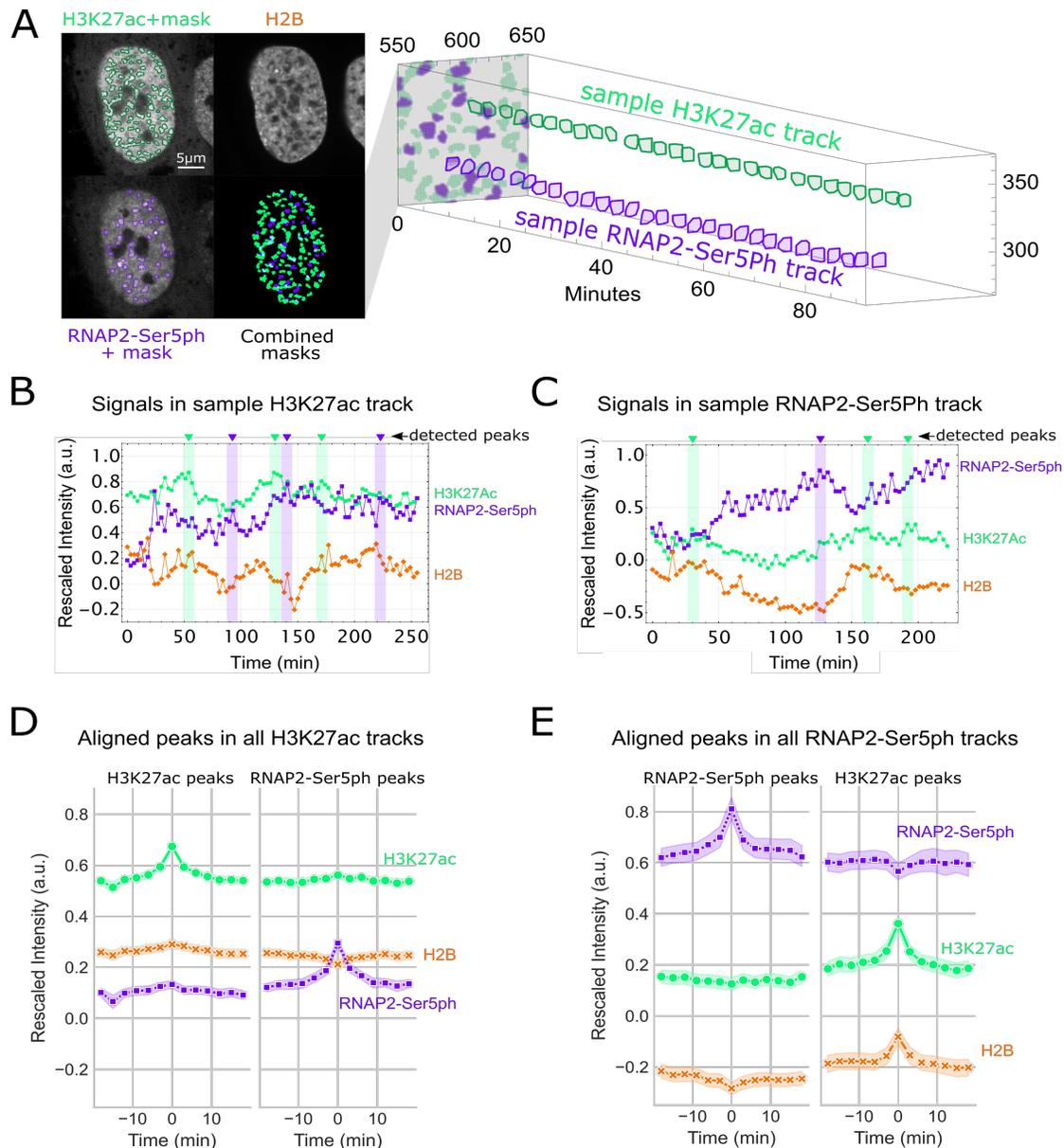


Fig. 3.5: Hours-long tracking of chromatin regions enriched for H3K27ac and RNAP2-Ser5ph. (a) Left) A local adaptive binarization filter is used to identify and isolate regions of chromatin enriched for specific post-translational modifications. Right) Enriched regions are tracked through time. An individual sample of H3K27ac enriched and RNAP2-Ser5ph enriched tracks are shown. (b) Time traces from a sample track of chromatin enriched for H3K27ac, with average levels of H3K27ac, RNAP2-Ser5ph, and H2B within the region plotted. Identified peaks within H3K27ac and RNAP2-Ser5ph signals are labeled in green and purple, respectively. (c) Same as B, for RNAP2-Ser5ph enriched region. (d) Left) Aligned H3K27ac ($n=2228$ peaks) peaks from all H3K27ac enriched chromatin tracks ($n=10,290$ tracks total; $n = 2743$ greater than 30 frames). All signals were aligned based upon the detected peaks timepoints. Right) Aligned RNAP2-Ser5ph detected peaks ($n=2387$) in the same tracks. (e) Left) Aligned RNAP2-Ser5ph peaks ($n=874$ peaks) from all RNAP2-Ser5ph enriched chromatin tracks ($n=5,005$ tracks total; $n = 965$ greater than 30 frames). All signals were aligned based upon the detected peaks timepoints. Right) Aligned H3K27ac peaks ($n=683$) in the same tracks.

two modifications are positively correlated with one another in sites enriched with H3K27ac but negatively correlated with one another in sites enriched with RNAP2-Ser5ph (Fig. 3.6C).

Based on these observations we can conclude three things. First, the fact that peaks did not always align with peaks (and troughs did not always align with troughs) confirmed our analysis was free of focus drift or illumination issues that would cause signals to artifactually dim or brighten in sync. Second, despite H2B and H3K27ac being present together within nucleosomes, our peak analysis found the two signals can be decoupled when acetylation levels are high. Specifically, when we aligned RNAP2-Ser5ph peaks occurring inside highly acetylated tracks, a distinct trough in H2B appeared despite an increase in H3K27ac (Fig. 3.5D, right). Conversely, when we aligned RNAP2-Ser5ph troughs in the same tracks, a distinct peak in H2B appeared despite a drop in H3K27ac (Fig. 3.6A, bottom-right). These data therefore provide direct evidence that H3K27ac can act independently of and even counter changes in nucleosome density. Third, the temporal correlation we measured between H3K27ac and RNAP-Ser5ph dramatically changed depending on which modification was enriched. When H3K27ac was enriched, the signals fluctuated in a correlated manner, whereas when RNAP2-Ser5ph was enriched, the signals fluctuated in an anti-correlated manner (Fig. 3.6C). We therefore conclude chromatin enriched with either H3K27ac or RNAP2-Ser5ph display distinct dynamics from one another on the minutes timescale.

3.3.3 Single Particle Tracking of H2B at Transcriptionally Enriched Sites Reveals Chromatin Slowdown

Given the differences in the temporal correlations we observed between sites enriched with H3K27ac and RNAP2-Ser5ph, we were curious if the two types of sites also exhibited different dynamics on shorter length and timescales. In particular, we wondered if chromatin within each site exhibited unique microscopic dynamics that could help explain or predict our longer-term observations. Our experimental system was uniquely poised to address this question because we could adjust our imaging setup in a straight-forward manner to enable the tracking of individual nucleosomes on the milliseconds time scale while still co-imaging PTMs using FabLEM (Fig.

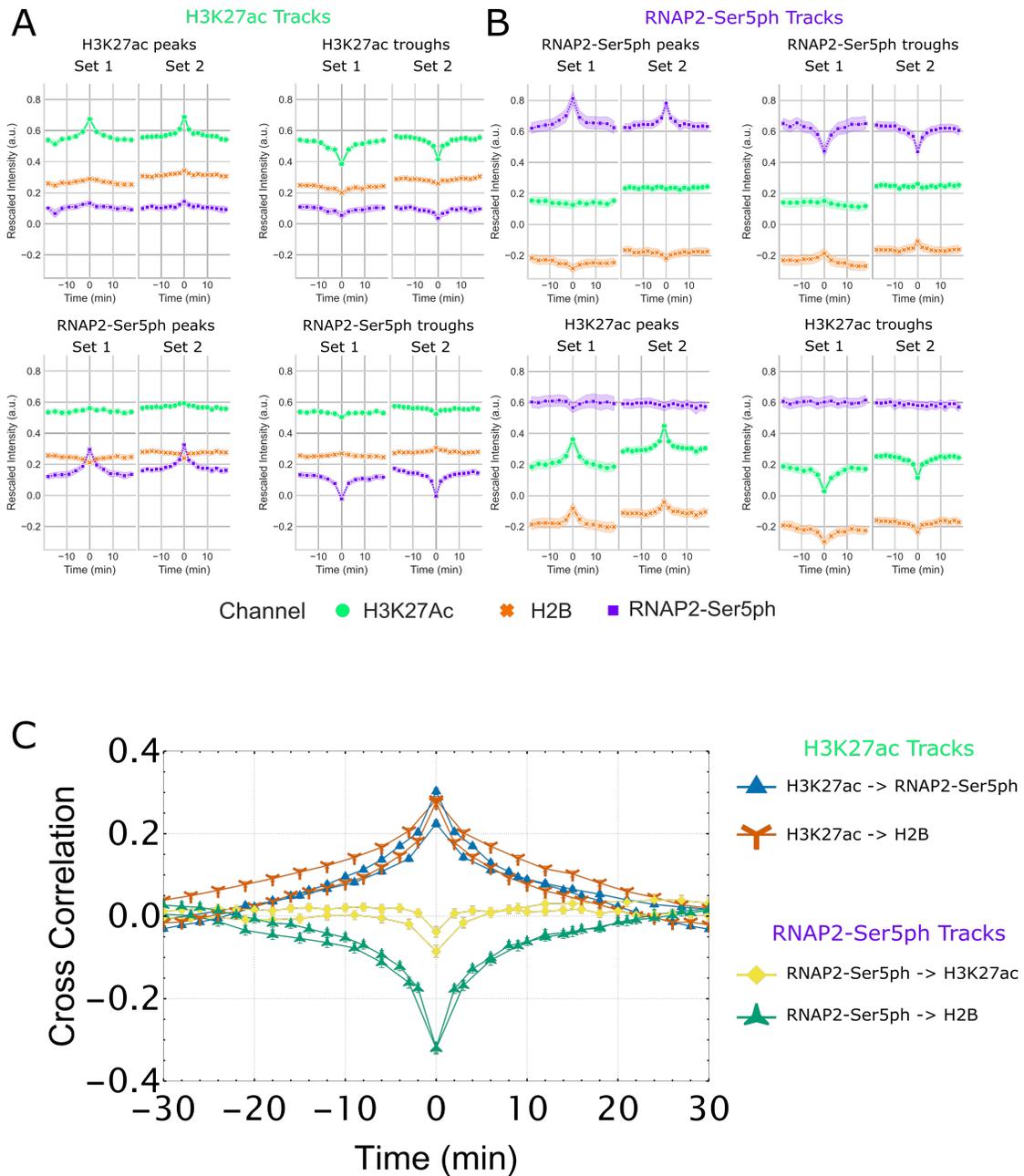


Fig. 3.6: Peak and Trough Detection Combined with Cross-Correlation Analysis Verify H3K27ac, RNAP2-Ser5ph and H2B Dynamic Behavior. (a) Detected peaks and troughs in all H3K27ac enriched tracks for two data sets of $n=13$ cells and $n=15$ cells. Both RNAP2-Ser5ph and H3K27ac peaks and troughs were detected, and all 3 signals were aligned based upon those peaks and troughs. (b) Same as (A), for RNAP2-Ser5ph enriched regions. (c) Cross correlation of various signals inside of H3K27ac and RNAP2-Ser5ph enriched regions, calculated for all tracks of length 20 frames or greater. Error bars are standard deviation obtained via bootstrapping ($n=1000$).

3.1C,D), and single-particle tracking and the corresponding analysis of trajectories is a useful way to decode the dynamics of intracellular components [62-65].

We first focused on transcriptionally active regions marked by RNAP2-Ser5ph. We hypothesized nucleosomes in these regions would either diffuse faster or slower than normal. In support of slower nucleosomes, several recent reports [33,38,66] have shown an anti-correlative relationship between chromatin mobility and transcriptional activity, consistent with the presence of hypothesized transcription factories or hubs that lock down chromatin [44]. In support of faster nucleosomes, on the other hand, it has been observed that some enhancers and promoters are more mobile after differentiation-induced transcription activation in embryonic stem cells, leading to the proposition that RNAP2 transcription activity stirs the local chromatin environment [52].

To see which of these two opposing scenarios is dominant, we again bead-loaded RPE1 cells stably expressing H2B-Halo with AF488 labeled Fab specific to RNAP2-Ser5ph. To enhance fluorescence signal-to-noise and enable long-term tracking of single fluorophores, we switched from confocal to highly inclined laminated optical (HILO) sheet microscopy [36] (Fig. 3.7A). To acquire many tracks with minimal crossing events, we sparsely labeled H2B-Halo with a reduced form of TMR ligand that is stochastically photoactivated during imaging [67]. Finally, we adjusted our imaging rate to a little over ten frames per second to exclusively track chromatin-incorporated H2B, an indicator for single nucleosomes [35,38] (Fig. 3.7B).

Using this experimental setup, we collected tens of thousands of tracks from 31 cells across 3 experimental replicates and created an analysis pipeline (Fig. 3.8A). To isolate tracks associated with RNAP2-Ser5ph, we again applied a local adaptive binarization filter to the nuclei of each cell to highlight all sub-nuclear regions enriched with the specific PTM (Fig. 3.7B, Fig. 3.8B,C). We then selected all tracks that remained inside of the mask for at least ten consecutive frames (Fig. 3.7C, Fig. 3.8D). This left us with hundreds to thousands of single-nucleosome tracks per cell to quantify chromatin mobility in the context of RNAP2-Ser5ph.

To ensure that our masking process was not creating a selection bias amongst our tracks, we created a control mask. For this, we isolated and identified each region of connected pixels in

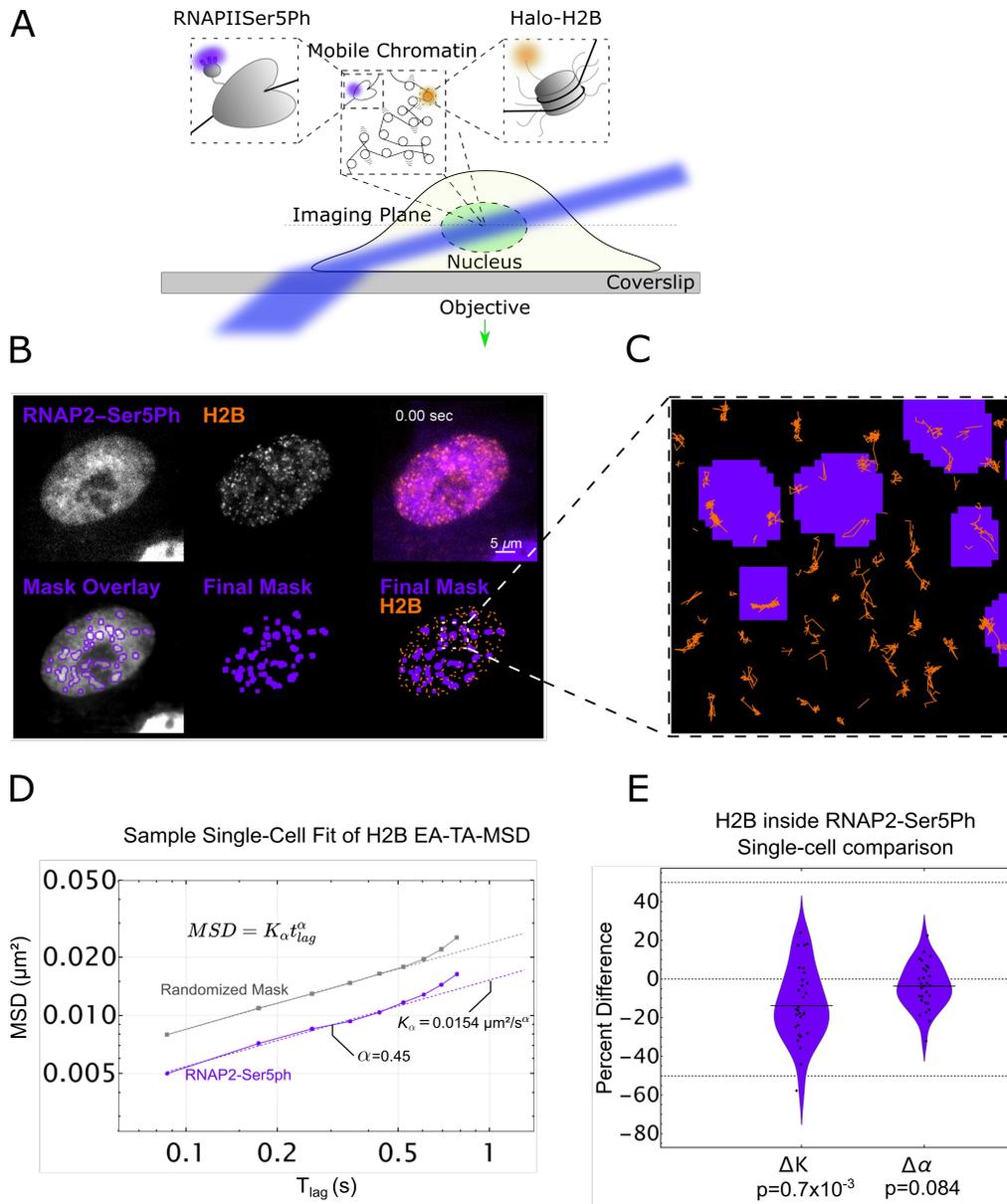


Fig. 3.7: Single-particle tracking of H2B located within transcriptionally active regions marked by RNAP2-Ser5ph (Continued on next page).

Fig. 3.7: Single-particle tracking of H2B located within transcriptionally active regions marked by RNAP2-Ser5ph. (a) Imaging scheme for single-particle H2B tracking experiments utilizing HILO microscopy setup. A combination of H2B-Halo and Fab specific for RNAP2-Ser5ph are utilized to track single molecules of H2B in the context of transcription. (b) Top Row: Left) Sample single frame of RNAP2-Ser5ph Fab in living RPE1 cells. Middle) Sample single frame of single-molecule H2B-Halo in the same cell. Right) Colored composite of RNAP2-Ser5ph and H2B channels in purple and orange, respectively. Bottom Row: Left) Outlines of RNAP2-Ser5ph local adaptive binarization mask overlaid upon the 100-frame average image from which the mask was extracted. Middle) Single local adaptive binarization mask of RNAP2-Ser5ph enriched regions. Right) RNAP2-Ser5ph mask with overlaid H2B-Halo tracks in orange. (c) Selected zoom from (B) showing sample full traces of Halo-H2B tracks occurring both inside and outside RNAP2-Ser5ph enriched regions. (d) Sample single-cell fit for the ensemble MSD of all H2B tracks lasting 10 frames or longer, localized in a randomized mask control (gray, n=5860 total tracks across 10 randomizations) or in RNAP2-Ser5ph enriched, transcriptionally active regions (purple, n=463 tracks). Determined diffusion coefficient (K_α) and alpha coefficient (α) for this fit are labeled. (e) Difference in K_α and α for RNAP2-Ser5ph enriched regions vs a randomized mask control. Each data point represents the difference in fit coefficients for a single cell fit (n=31 cells), relative to the mean value, with each cell comprising hundreds to thousands of H2B tracks. Significance determined via students t-test.

each mask, and each region was then separately moved by a randomized vector. This procedure created a random mask with a similar morphology to our original mask, but randomly dispersed throughout the nucleus (Fig. 3.8E,F). In so doing, we could compare nucleosome dynamics inside the original RNAP2-Ser5ph mask to average nucleosome dynamics inside a large collection of morphologically identical random masks. This process is critical because restricting tracks to any masked region inherently selects for slower tracks and the amplitude of this selection bias depends on the precise size, shape, and morphology of the mask in question (Fig. 3.8G).

To quantitatively compare dynamics in our original and random masks, we fit the ensemble and time-averaged mean-squared displacement (MSD) of nucleosomes to a model of anomalous diffusion [68,69] such that $MSD = K_\alpha t_{lag}^\alpha$. Here, K_α is the so-called generalized diffusion coefficient and α is the anomalous exponent (Fig. 3D). Similar to other reports [35,38,70], our fitted K_α was in the range of $0.03 \pm 0.01 \mu m^2/sec$, while our fitted α was consistently $\sim 0.45 \pm 0.1$ (even after correcting for both static and dynamic localization errors and eliminating their effect using a resampling approach [71]). Normal diffusion would be characterized by $\alpha=1$. In our case the exponent α indicates the nucleosomes in the cell nucleus are subdiffusive, i.e., their MSD is sub-linear in lag time.

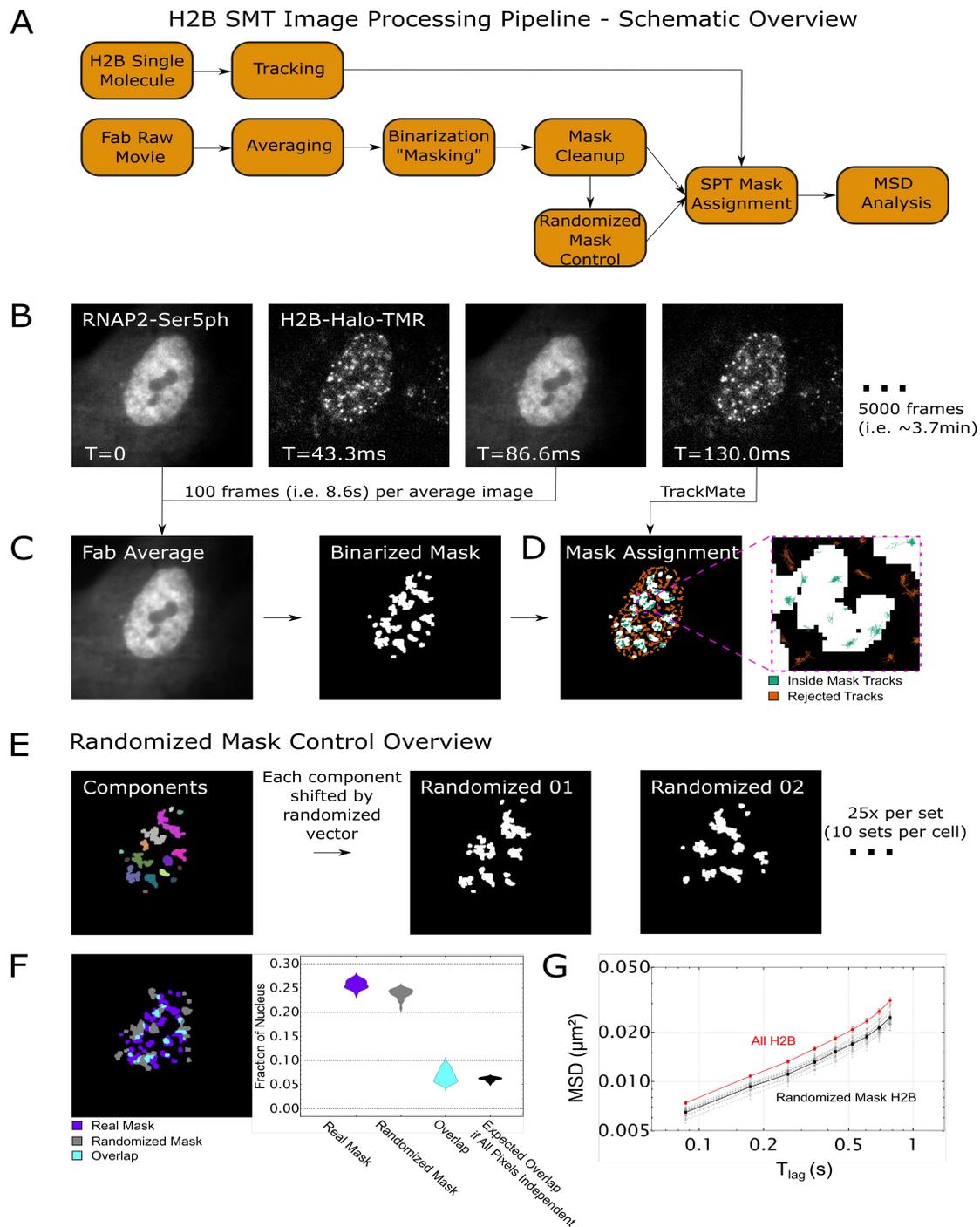


Fig. 3.8: Analysis Pipeline for Tracking Single Nucleosomes in the Context of Post-Translational Modifications. (Continued on next page).

Fig. 3.8: Analysis Pipeline for Tracking Single Nucleosomes in the Context of Post-Translational Modifications. (a) Schematic overview of the image processing pipeline for H2B single-molecule tracking experiments. (b) Fab and single-molecule H2B are imaged in an alternating fashion, for 5000 total frames (3.7min). (c) 100 Fab images are used to create an average intensity image, reducing the noise generated by freely diffusing Fab. A local adaptive binarization algorithm is then applied to create a mask of enriched regions. (d) Determination is made whether single nucleosome tracks are inside or outside of the mask, and tracks outside of a mask are not used in subsequent analysis. (e) To create a randomized mask, first the connected pixels in a mask are segmented into separate regions, or components. Entire components are shifted by an individually randomized vector, creating a mask with random distribution of components, but whose morphology is analogous to the original mask. (f) Overlap between real and randomized masks. Distribution of overlap is quantified and compared to an in-silico control, the expected overlap between masks if all pixels in the nucleus were randomized individually. n=25 real and randomized masks (g) Log-log plot of all nuclear H2B MSD in a single cell vs those inside of 10 sets of 25 randomized masks for the same cell. The EA-TA-MSD for H2B inside randomized masks is always slightly decreased, demonstrating the need for a randomized mask control.

To control for cell-to-cell variability, we subtracted the fitted K_α and α values of nucleosomes inside randomized masks from those inside RNAP2-Ser5ph enriched masks. This generated a $\Delta\alpha$ and ΔK for each cell, allowing us to directly compare chromatin dynamics at a single-cell level and controlling for any cell-to-cell variability. As shown in Fig. 3E, the percent change in these differences provided a convenient single-cell measure of the impact RNAP2-Ser5ph enrichment has on single-nucleosome mobilities.

According to our fits, nucleosomes diffusing near chromatin sites enriched with RNAP2-Ser5ph experienced no significant change in the type of diffusion they underwent ($\Delta\alpha \sim 0$, Fig. 3E, right). Despite that, the nucleosome diffused 15 percent more slowly, characterized by a marked decrease in the diffusion coefficient. This substantial reduction in the mobility of nucleosomes near sites of transcription initiation supports the transcription factory/hub model [44]. Furthermore, because our measurements were at random sites in a random set of cells, our data suggest nucleosome slowdown near transcription initiation is a common event that does not depend on specific perturbations or drug treatments.

3.3.4 Enrichment of H3K27ac Predicts an Increased Nucleosome Diffusion

We next turned our attention to histone acetylation, applying the same methodology to chromatin sites enriched with H3K27ac. According to the textbook model, histone acetylation should weaken DNA-nucleosome interactions, theoretically opening up chromatin structure to facilitate nucleosome mobility. In principle, we would therefore predict acetylated nucleosomes have a higher mobility than average. On the other hand, we just observed that less dense, transcriptionally active chromatin can have reduced mobility. The direct relationship between chromatin density and mobility is therefore not obvious. So far the only study [38] that observed increased single-nucleosome mobility in the context of acetylation relied on the use of Trichostatin-A (TSA), a broad inhibitor of HDAC proteins that have many non-histone targets [72]. It is therefore unclear how chromatin will behave in regions enriched for H3K27ac.

To more directly address this question, we performed the same experiment as above, except we now loaded cells with AF488 labeled Fab specific for H3K27ac. As before, we tracked single nucleosomes, created real and randomized masks for H3K27ac enrichment, and isolated tracks inside of those regions (Fig. 3.9A, Sup Movie 3). We then fit the ensemble and time-averaged MSD from each cell inside the masks to extract K_α and α (Fig. 3.9B), calculated ΔK and $\Delta\alpha$ for each cell, and plotted the distributions together (Fig. 3.9C).

As we saw with RNAP2-Ser5ph, our data indicate nucleosomes in regions enriched with H3K27ac did not experience any significant change in the type of diffusion they underwent ($\Delta\alpha \sim 0$, Fig 4D, right). However, in stark contrast to what we saw with RNAP2-Ser5ph, we now observed a significant increase in the diffusion coefficient K_α . Specifically, our data suggest nucleosomes diffuse ~ 15 percent faster than normal when associated with chromatin enriched for H3K27ac ($\Delta K > 0$, Fig 4D, left). These data therefore provide direct live-cell support that acetylated nucleosomes move more freely in 4D space, as the standard model of histone acetylation would predict. Furthermore, the fact that we observed a speed up in nucleosomes in this case rather than a slow down suggests our general strategy for measuring nucleosome dynamics in the context of specific PTM masks is unbiased. Taken together, our data support a model whereby H3K27ac and RNAP2

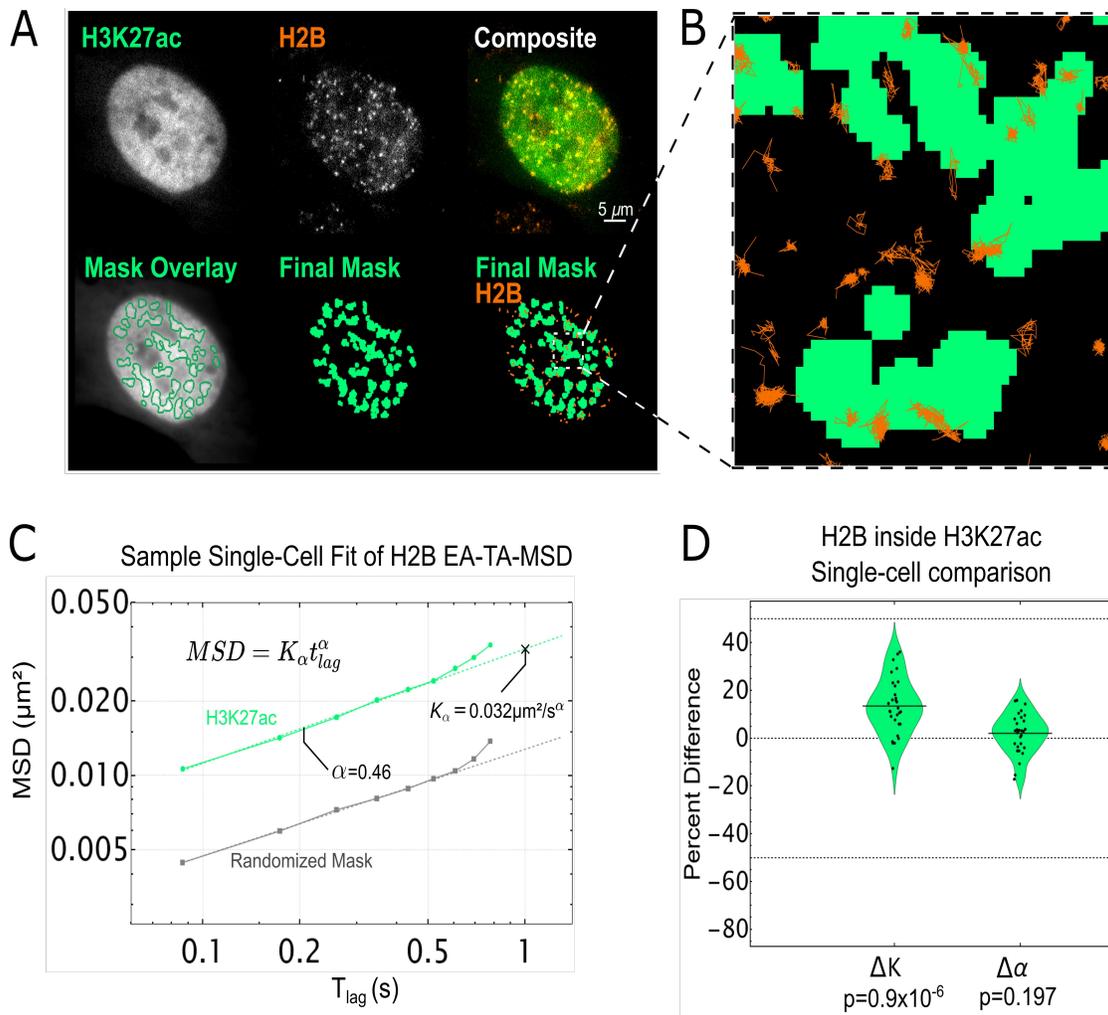


Fig. 3.9: Single-particle tracking of H2B inside of H3K27ac enriched chromatin regions. (a) Top Row: Left) Sample single frame of H3K27ac Fab in living RPE1 cells. Middle) Sample single frame of single-molecule Halo-H2B in the same cell. Right) Colored composite of H3K27ac and H2B channels in green and orange, respectively. Bottom Row: Left) Outlines of H3K27ac local adaptive binarization mask overlaid upon the 100-frame average image from which the mask was extracted. Middle) Single local adaptive binarization mask of H3K27ac enriched regions. Right) H3K27ac mask with overlaid Halo-H2B tracks in orange. (b) Selected zoom from (A) showing sample full traces of Halo-H2B tracks occurring both inside and outside RNAP2-Ser5ph enriched regions. (c) Sample single-cell fit for the ensemble MSD of all H2B tracks lasting 10 frames or longer, localized in a randomized mask control (gray, $n=1892$ total tracks across 10 randomizations) or in H3K27ac enriched chromatin (green, $n=445$ tracks). Determined diffusion coefficient (K_{α}) and alpha coefficient (α) for this fit are labeled. (d) Difference in K_{α} and α for RNAP2-Ser5ph enriched regions from a randomized mask control. Each data point represents the difference in fit coefficients for a single cell fit ($n=30$ cells), relative to the mean, with each cell comprising thousands of H2B tracks. Significance determined via student's t-test.

Ser5ph mark opposing ends of the nucleosome mobility landscape, with H3K27ac marking sites of increased mobility and RNAP2-Ser5ph marking sites of decreased mobility.

3.4 Discussion

Despite the well known correlation between histone acetylation, chromatin structure, and gene activity, the dynamic relationship between these factors has remained largely unexplored in living cells. In this study we addressed this issue by combining single-molecule tracking with live-cell imaging of PTMs. Using this unique combination of technologies, we quantified the dynamic interplay between endogenous histone acetylation marked by H3K27ac, endogenous transcription initiation marked by RNAP2-Ser5ph, and chromatin mobility marked by stably expressed H2B-Halo.

Our work revealed several fundamental relationships (Fig. 3.10). First, we showed chromatin that is highly enriched with H3K27ac was generally segregated from chromatin that was highly enriched with RNAP2-Ser5ph. This was a bit surprising given the strong and positive genome-wide correlation between H3K27ac and RNAP2-Ser5ph regularly observed via chromatin immunoprecipitation. A straightforward explanation for this apparent discrepancy is that individual chromatin sites oscillate between two states, one enriched with H3K27ac and one enriched with RNAP2-Ser5ph. Over time or across a cellular population this would result in an average co-enrichment in signals. Alternatively, given the apparent temporal stability of many of the regions we tracked, a single genetic locus could move back and forth between stable, preformed sites enriched with either H3K27ac or RNAP2-Ser5ph. Such models are not unprecedented as we observed a similar phenomenon at the MMTV tandem gene array [22]. In that study, H3K27ac levels at the array were high prior to activation, but the levels rapidly dropped post-activation when RNAP2-Ser5ph rapidly rose. Similarly, in another study with zebrafish embryos, high H3K27ac were observed at a developmental gene prior to its activation, but levels again dropped once RNAP2 was recruited²⁴. Beyond these specific examples, it is now generally accepted that the vast majority of genes display bursty transcription [73-75], transitioning between active and inactive states. In light of this,

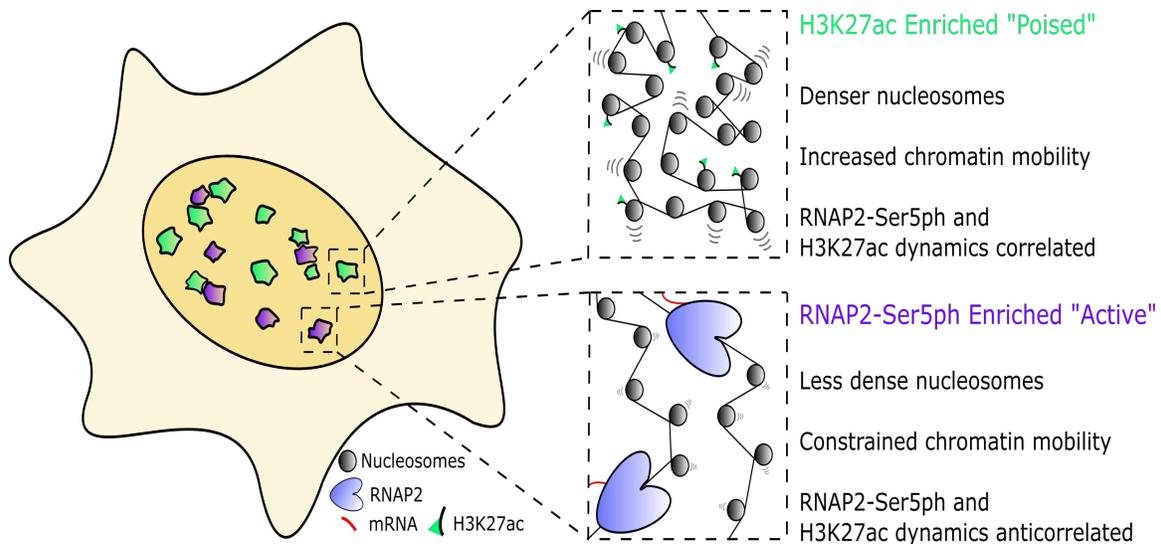


Fig. 3.10: Two state model of RNAP2-Ser5ph and H3K27ac enriched chromatin domains. Inside the nucleus, two distinct sites are characterized by enrichment for RNAP2-Ser5ph or H3K27ac with little overlap. H3K27ac enriched regions represent poised chromatin, which is slightly more dense than average chromatin but has increased single nucleosome mobility. In combination, these factors reinforce 3D target searching, increasing the probability of contact between genetic elements or interacting proteins. RNAP2-Ser5ph enriched regions are areas of high transcriptional activity, where chromatin is decondensed but stabilized by the presence of bound transcription machinery.

it is tempting to speculate that bursty genes also oscillate between PTM states, having high levels of H3K27ac at one time followed by high levels of RNAP2-Ser5ph at another time. Unfortunately, we were unable to confirm any state-transitions in this study because we could only track a specific site while it remained enriched. In the future this shortcoming could be reconciled by directly labeling and tracking single genetic elements while co-imaging H3K27ac and RNAP2-Ser5ph.

Second, we showed the H3K27ac and RNAP2-Ser5ph signals were positively correlated through time in sites enriched with H3K27ac, but negatively correlated in sites enriched with RNAP2-Ser5ph. These distinct time-dependent spatial correlations persisted for up to 20 minutes (Sup Fig. 3.9C), a timescale on par with the transcription cycle of average-sized genes according to RNAP2 FRAP and MS2-based gene tracking experiments [76-78]. The data therefore suggest the relationship between H3K27ac and RNAP2-Ser5ph may change throughout the transcription cycle. In sites poised for transcription (which we define here as high H3K27ac and low RNAP2-Ser5ph), increases in H3K27ac levels help recruit RNAP2 and vice versa (green sites in Fig. 3.10). On the

other hand, in transcriptionally active sites (high RNAP2-Ser5ph, lower H3K27ac; purple sites in Fig. 3.10), increases in H3K27ac lead to a loss in RNAP2-Ser5ph and vice versa, most likely due to enhanced RNAP2 promoter escape or pause release by H3K27ac [13,22] and the decondensation of nucleosomes by active RNAP2. This dual behavior is consistent with what we observed at the MMTV tandem gene array [22]. In that study, we showed H3K27ac at the array prior to its activation was predictive of (1) the efficiency of transcription factor recruitment immediately after hormone induction and (2) later RNAP2 promoter escape and chromatin decondensation. Whereas the former would positively correlate with RNAP2-Ser5ph, the latter would negatively correlate. Thus, our data support a model wherein H3K27ac plays multiple roles in tuning gene activation efficiency.

Third, we showed that single nucleosomes diffused $\sim 15\%$ slower than normal in transcriptionally active sites enriched with RNAP2-Ser5ph and $\sim 15\%$ faster than normal in sites enriched with H3K27ac. Taken together, these data paint a heterogeneous picture of the chromatin mobility landscape [43], one that is dynamically colored by local enrichments in specific PTMs (Fig. 3.10). While the molecular details of this two-state model have yet to be clarified, it does answer the conundrum between standard model predictions in acetylated and transcriptionally active regions. Namely, H3K27ac and RNAP2 can have separate impacts on chromatin landscape and mobility because they are acting in different places and times in living cells.

Regarding the slowdown of nucleosomes we observed in the context of transcription, our results are consistent with previous reports tracking single-genomic loci [33,66], as well as with a similar study tracking single nucleosomes in the presence of transcriptional inhibitors [38]. Going beyond those previous studies we have now linked the slowdown to RNAP2-Ser5ph, a specific PTM associated with transcription initiation/pausing [7,14]. Furthermore, we did so without the use of drugs or broad-acting perturbations. Our data therefore support a model wherein transcription initiation is processed predominantly in factories [44] or hubs [50,51] that tend to lockdown nearby chromatin rather than stir it up [52]. While the lockdown we observe is consistent with traditional notions of transcription factories, we cannot rule out other models of lockdown. For

example, crowding within phase-separated condensates [79] or within stiff transcriptional loops [80] could also restrict chromatin mobility. Similarly, we cannot exclude the possibility that some genetic loci behave differently. This is because we had no way of labeling the genomic context of each track. In other words, we could not say if a subset of fast or slow tracks were associated with a specific gene or a specific chromosomal locus. Furthermore, because we focused exclusively on RNAP2-Ser5ph, we may have missed some portion of transcription sites. Recently, there is growing evidence that transcription initiation (marked by RNAP2-Ser5ph) and elongation (marked by RNAP2-Ser2ph) are spatially separated [12,81]. Thus, it is possible that chromatin enriched with actively elongating RNAP2 may still associate with highly mobile nucleosomes [81]. In the future it will therefore be interesting to compare nucleosome mobilities within the context of a wider range of RNAP2 PTMs.

Regarding the enhanced mobility of nucleosomes we observed in the context of H3K27ac, our results are again consistent with a previous report that tracked nucleosomes in the presence of inhibitors [38]. As before, we go beyond this work by directly tracking nucleosomes in the context of a specific PTM without the use of inhibitors, which have many off-target and non-specific effects. Furthermore, by focusing on the H3K27ac modification, a well known marker of promoters and enhancers [55], our data suggest these elements are especially dynamic prior to transcription activation, in line with the predictions of the standard model of gene activation by histone acetylation. Although we did not examine other residue-specific histone acetylation marks, such as H3K18ac or H3K9ac, we expect many will behave similarly, not only because of the earlier experiments with TSA [38], but also from the classic viewpoint that acetylation acts redundantly to neutralize histone charge [3]. In particular, we expect H3K27ac is most likely redundant because it is written by the promiscuous p300/CBP lysine acetylase subfamily [82] and its complete removal has little impact on fly development and mouse embryonic stem cell differentiation [83,84]. Nevertheless, it will be interesting to examine other acetylations one by one. H3K9, in particular, would be an interesting comparison since it plays a distinct role in nuclear receptor transactivation and is written independently of H3K27ac by the Gcn5/PCAF lysine acetylase subfamily [85].

In summary, we have combined single-nucleosome tracking with live-cell, multicolor PTM imaging to reveal two types of distinct intranuclear regions that are enriched with RNAP2-Ser5ph or H3K27ac and that display distinct dynamics across multiple time and length scales. We believe this imaging strategy, combined with computational tools to track, extract, and analyze dynamics within masked regions of cells in an unbiased fashion, create a powerful experimental platform to better elucidate the role residue-specific PTMs play in gene expression networks. The techniques we have developed here can now be used to study a suite of other modifications. For example, beyond acetylation, it will be interesting to mask chromatin marked by other histone PTMs, both individually and in combination. Our methodology to track nucleosomes in the context of PTMs can also be generalized to track other proteins, such as transcription factors or chromatin remodelers. As the number of PTM-protein combinations is huge, our general method can be broadly applied to help elucidate the hidden roles PTMs play in diverse gene and chromatin regulatory networks.

3.5 Materials and Methods

Cell culture and care

Human retinal epithelial cells (hTERT-RPE1, ATCC) were grown in 10% (v/v) FBS (Atlas) / DMEM (Thermo Scientific). Media was supplemented with 1 mM L-glutamine (Gibco) and 1% (v/v) Pen/Strep (Invitrogen/Gibco) and cells were grown at 37°C in 5% CO₂. Cell density was maintained between 15-80%. RPE1 cells were purchased from ATCC and were authenticated via STR profiling by ATCC and tested negative for mycoplasma contamination. Cells were stored at -80°C in Cellbanker1 (Amsbio LLC), and thawed >1 week prior to any imaging.

Stable cell line creation

RPE1 cells were transfected with a Halo-H2B plasmid containing a Neomycin-resistance selection site. Transfection was performed with an LTX Lipofectamine with Plus Reagent kit (Thermo Fisher), per the manufacturers instructions. Briefly, an 80% confluency, 35mm MatTek Chamber was washed and the media was replaced with 1.75 mL Opti-MEM (Thermo Scientific) directly

before transfection. The transfection solution included 2.5 μg DNA plasmid, 7.5 μl Plus reagent, and 7.5 μl Lipofectamine, and the remainder Opti-MEM for a total solution volume of 250 μl . This solution was incubated for 515 min at room temperature before being added to the cell chamber. Cells were incubated in this transfection solution for 24 h before the media was changed back to 10% FBS-DMEM. Cells were then transferred to a 10cm dish and allowed to replicate for 14 days under selection at 750 $\mu\text{g}/\text{ml}$ Geneticin G418 Sulfate (Gibco). Cells were then stained with Halo-JF646 at 500nM for 30 min, immediately washed with 1X PBS, and sorted via flow cytometry (FACS) to obtain a consistent population of stable H2B-Halo expressing RPE1 cells. The stable line was maintained under a constant selection of 400 $\mu\text{g}/\text{ml}$ Geneticin.

Fab generation and dye conjugation

RNAP2-Ser5Ph and H3K27ac specific Fab were generated and affinity purified as previously described [21]. Briefly, Fab were generated from monoclonal antibodies using the Pierce Mouse IgG1 Fab and F(ab')₂ Preparation Kit (Thermo Fisher). Antibodies were first digested into Fab in a Zeba Desalt Spin Column (Thermo Fisher) containing immobilized Ficin while gently rotating for 35 h at 37°C. Fab were purified from the digest by centrifugation in a NAb Protein A column. Eluted Fab were concentrated to >1 mg/mL and stored at 4°C. Fab labeling with AF488 and CF640 was performed in small batches using 100 μg Fab. The dye was an AlexaFluor488 ester (Invitrogen) or CF640 (Biotium) dissolved in DMSO and either used immediately or stored at 20°C. For labeling, 100 μg of Fab was dissolved in a final volume of 100 μl of 100 mM NaHCO₃ (pH 8.5) plus 4 μl of AF488 or 2 μl of CF640 dye. Fabs were incubated with dye for ~2 h at room temperature with constant gentle rotation and agitation. The Fab were separated from unconjugated dye in an equilibrated PD MiniTrap G-25 desalting column (GE Healthcare). Fab were concentrated in an Amicon Ultra-0.5 Centrifugal Filter Unit (NMWL 10 kDa; Millipore) to >1 mg/ml. The degree of labeling (DOL) was calculated using the following equation:

$$\text{DOL} = \frac{\epsilon_{\text{IgG}}}{\epsilon_{\text{dye}}} * \frac{1}{\left(\frac{A_{\text{Fab}}}{A_{\text{dye}}}\right)^{-1} - \text{CF}} \quad (3.1)$$

Fab is the extinction coefficient of Fab ($70,000 \text{ M}^{-1} \text{ cm}^1$), dye is the extinction coefficient of the dye used for conjugation ($71,000 \text{ M}^{-1} \text{ cm}^1$ for AF488 and $105,00 \text{ M}^{-1} \text{ cm}^1$ for CF640), A280 and Adye are the measured absorbances of dye-conjugated Fab fragments at 280 nm and at the peak of the emission spectrum of the dye (488 nm for AF488, 637 nm for CF640), respectively, and CF is the correction factor of the dye (the ratio of the absorbances of the dye alone at 280 nm to at the peak). If the DOL was <0.7 , this protocol was repeated on the same Fab to increase their DOL to ~ 1 . Fab were then stored at 4°C for later use.

Bead loading of Fab and Halo-ligand staining prior to imaging

To bead load RNAP2-Ser5Ph and H3K27ac specific Fab, cells were plated at 75% confluency on a 35 mm glass-bottom chamber (MatTek) 24 hours before imaging. Cells were bead loaded as described previously⁸⁶. Briefly, $1.5 \mu\text{g}$ of each Fab required for an experiment is added together and diluted to $5 \mu\text{l}$ in 1x PBS. Cell media is removed, and the $5 \mu\text{l}$ mixture of Fab was pipetted directly on top of the cells, followed by a sprinkling of $100 \mu\text{m}$ glass beads to form a monolayer on top of the cells (Sigma Aldrich). The entire chamber is then lifted and firmly tapped on the worktop 6-8 times inside of the culture hood. The media was then immediately replaced and cells were returned to the incubator. After 4 h of recovery, cells were washed three times in phenol-free DMEM with 10% FBS and 1 mM L-glutamine. For single molecule experiments, HaloTag-TMR ligand was first incubated in 100mM NaBH₄ / 1X PBS for 10 minutes, and then diluted to 500pM and cells were stained. For non-single molecule experiments, HaloTag-JaneliaFlour-X554 (JFX554) was used at 200nM concentration. In all cases cells were incubated with HaloTag ligand for 30 minutes, then followed by 3x (3 washes in DMEM with 10% FBS and 1 mM L-glutamine followed by 5 minutes of incubation). Cells were moved to the microscope stage-top incubator for imaging 4h post-bead loading.

Live cell imaging with the confocal microscope

Live cell images were acquired on an Olympus (IX83) inverted spinning disk confocal microscope equipped with a Cascade II EMCCD camera. A 100x oil immersion objective with a pixel size of $0.096 \mu\text{m}$ was used for all images. Cells were plated 24 hours before imaging, beadloaded

4-5 hours before imaging, and stained with Halo-JFX554 30 minutes before imaging, as described above. The chamber was allowed to acclimate in the stage-top incubator for 2-3 hours before final image collection to prevent focal shift. 488 nm, 561 nm, and 637 nm lasers were used in all cases. 5 frames were acquired in each channel, for each cell, every 2 or 3 minutes.

Confocal image registration

A sequence of post-processing functions was applied to all confocal images prior to any binarization or tracking of specific subnuclear regions. First, for each channel, the 5 images collected at each time point were averaged together. To correct for cell movement, a maximum intensity projection of all 100 time points of H2B images was created for each cell, and a mask of nucleus movement was drawn by hand around the maximum projection. The mask of nucleus movement was multiplied by all images, leaving only the nucleus of interest. The resultant nuclear H2B images were fed into the FIJI [87] plugin Register Virtual Stack Slices [88], which aligned each nuclear image. This created a list of transformation functions for each time point for each cell, which were then used in conjunction with FIJI plugin Transform Virtual Stack Slices to register the Fab channels for each cell, ensuring that the exact same transformation is applied to all channels.

Identification of sub-nuclear enriched regions with local adaptive binarization

To isolate sub-nuclear regions enriched for RNAP2-Ser5Ph or H3K27ac, custom code was written which utilized a built-in Mathematica function, LocalAdaptiveBinarize. First, a nuclear mask was drawn by hand, and multiplied by all images to isolate only the nucleus. Each nuclear image had its intensities normalized from 0 to 1, and a function was applied to each pixel:

$$I_{lt} = \alpha \langle I_n \rangle + \beta \sigma_n + \gamma \quad (3.2)$$

where I_{lt} is the local threshold, $\langle I_n \rangle$ and σ_n represent the local mean and standard deviation in the neighborhood of n pixels. The values α , β , γ are user-defined constants. If a pixels value is higher than the resultant function, it is set to 1, else it is set to 0. For all masks created, the same parameters were used. These are: $n=441$ (21x21 window), $\alpha=0.94$, $\beta=0.6$, $\gamma=0.05$. The resulting

binary image had speckle noise eliminated by applying a 10-pixel size filter, and the remaining regions were dilated by 1 pixel to smooth and fill in gaps.

Tracking of RNAP2-Ser5Ph and H3K27ac enriched regions

The local adaptive binarization masks from above were then tracked using FIJI plugin Track-Mate [89] (v7.6.1) using the following parameters: LoG Detector; Estimated Blob Diameter: 15.0; Pixel Threshold: 0.01; Sub-Pixel Localization: Enabled; Simple LAP Tracker; Linking Max Distance: 10 pixels; Gap-Closing Max Distance: 3 pixels, Gap-closing Max-Frame Gap: 2 frame. The resulting tracks were fed into custom Mathematica code which finds the region associated with each x-y coordinate for each track. A region was defined as the set of connected pixels with a value of 1 that are also within 5 pixels in any direction from the X-Y centroid. The values of each channels raw data pixels (H3K27ac, H2B, RNAP2-Ser5Ph) were then rescaled for each nuclear image, with -1 and 1 set to the 2.5 and 97.5 quantiles, respectively, to account for outliers. All 3 channels pixel intensities within the tracked region were then extracted and averaged together, resulting in a single value for H3K27ac, H2B, and RNAP2-Ser5Ph for each tracked region, for each time point. If the resulting time traces had missing data points due to track gaps, these values were linearly interpolated from adjacent data (less than 3% of total data).

Correlation and Cross-Correlation Analysis

To calculate the correlation of H3K27ac, H2B, and RNAP2-Ser5Ph rescaled intensity traces, the Pearson correlation coefficient was calculated. The mean time-lag cross correlations were obtained by calculating the Pearson correlation coefficient for each track individually at various time lags, and then weighting each time lag by the length of the track used to calculate it. Then all tracks cross correlation curves were averaged. The error of cross correlation curves was obtained via bootstrapping. To bootstrap, intensity traces were randomly selected from the pool of data, with replacement, equal to the size of the dataset. Then, the mean cross correlation is calculated as above, and the process is repeated 1000 times. The error is reported as the standard deviation of all 1000 bootstrapped cross correlation curves.

Peak and trough calling analysis

To find peaks in the H3K27ac and RNAP2-Ser5Ph rescaled intensity tracks, the `scipy` [90] function in Python `scipy.signal.find_peaks()` was used using a prominence of 0.15, a width of 2 and distance of 10. Troughs were found the same way after signals were inverted. For this analysis, tracks of length < 30 time points were ignored. Also, if peaks were within 10 time points of the beginning or end of a track, they were ignored (to facilitate easier alignment of all peaks).

Live cell imaging on the HILO microscope

All live-cell imaging was performed on a custom built widefield fluorescence microscope with a highly inclined thin illumination scheme described previously. Briefly, the microscope equips three solid-state laser lines (488, 561, and 637 nm from Vortran) for excitation, an objective lens (60X, NA 1.49 oil immersion, Olympus), an emission image splitter (T660lpxr, ultra-flat imaging grade, Chroma), and two EMCCD cameras (iXon Ultra 888, Andor). Achromatic doublet lenses with 300 mm focal length (AC254- 300-A-ML, Thorlabs) were used to focus images onto the camera chips instead of the regular 180 mm Olympus tube lens to satisfy Nyquist sampling (this lens combination produces 100X images with 130 nm/pixel). A single camera was used to capture both Fab (AF488, Thermo-Fisher) and Halo-H2B stained with Halo-ligand (TMR, Promega). A high-speed filter wheel (HS-625 HSFW TTL, Finger Lakes Instrumentation) is placed in front of the camera to minimize the bleed-through between the red and the green signals (593/46 nm BrightLine for the red and 510/42 nm BrightLine for the green, Semrock). All single molecule imaging was performed on a single focal plane for each cell. The laser emission, the camera integration, and the emission filter wheel position change were synchronized by an Arduino Mega board (Arduino). Image acquisition was performed using open source Micro-Manager software [91] (1.4.22).

Live RPE1 cells were placed into a stage-top environmental chamber at 37°C and 5% CO₂ (Okolab) to equilibrate for at least 30 min before image acquisition. Imaging size and exposure time were set to 256 by 256 pixels and 30 msec, respectively. The resultant imaging rate was 23hz, (43.34 ms per frame). Fab and single-molecule images are alternated, resulting in 86.68 ms per

frame for each channel. Laser powers were measured at the back focal plane to be 150 μ W for 488 nm and 3.4 mW for 561 nm.

Single-molecule H2B tracking and Mask Assignment

Single-molecule tracks were identified using TrackMate 5.0.1 [92] with the following parameters: LoG Detector; Estimated Blob Diameter: 5.0; Pixel Threshold: 100; Sub-Pixel Localization: Enabled; Simple LAP Tracker; Linking Max Distance: 3 pixels; Gap-Closing Max Distance: 2 pixels, Gap-closing Max-Frame Gap: 1 frame.

Movies were 5000 total frames, 2500 of Fab and 2500 of single molecule H2B in an alternating fashion. Masks were generated by first averaging 100 frames of Fab images, then applying the local adaptive binarization algorithm as above, thereby generating 25 masks which updated every 8.2 seconds. Single-molecule H2B tracks were then assigned to a mask based upon maximum residence time. Tracks were determined to be inside or outside of H3K27ac, RNAP2-Ser5Ph, or randomized masks based upon their sub-pixel X-Y coordinates. Tracks which crossed between mask borders were split, and track segments which resided in an area for at least 10 consecutive frames were kept.

Diffusion analysis (Generating EA-TA-MSD curves, truncating tracks based on length, and fitting for K and alpha)

TrackMate files were fed into a custom MatLab script for analysis. To calculate the time-averaged mean square displacement (TAMSD) for each track, all tracks were first truncated to include only their first 10 frames. TAMSD was calculated as follows [93]:

$$\overline{\delta^2(n\Delta t)} = \frac{1}{N-n} \sum_{j=1}^{N-n} (r[(j+n)\Delta t] - r[j\Delta t])^2 \quad (3.3)$$

where r is xy position as a function of time, the overbar indicates a temporal average, t is the frame time, $t_{lag} = n\delta t$ is the lag time, and N is the number of data points in the trajectory. The ensemble averaged TAMSD (EA-TA-MSD) was calculated by averaging all single-nucleosome TAMSD results for each cell. A localization error of 30nm was roughly determined from the intercept of the EA-TA-MSD curve, and a correction was applied to all curves prior to fitting. The

first five points of the logarithm of EA-TA-MSD vs logarithm of time curve for each cell were used in a linear fit to obtain K (generalized diffusion coefficient) and α (anomalous exponent) according to:

$$\langle \overline{\delta^2(t_{lag})} \rangle \cong K_{\alpha} t_{lag}^{\alpha} \quad (3.4)$$

To obtain K_{α} and α for each individual cell, K_{α} and α obtained using randomized masks were subtracted from K_{α} and α obtained using the real-mask.

Randomized mask generation

A randomized mask was created from the local adaptive binarization of real data through the following process. Each subnuclear component, defined as the series of connected foreground pixels within the local adaptive binarization, had its pixel locations identified. For each component, a randomized vector was created with value between -20 and 20 for both the x and y directions. Each component was then shifted in its entirety by pixels equal to the value of the randomized vector. The result was a randomized, binarized mask which has a morphological nature based on the data from which it was created, but with each subnuclear component randomly placed.

Resampling MSD Curves to Calculate α Without Static and Dynamic Errors

MSD data obtained from single particle tracking are prone to static and dynamics errors, making it challenging to obtain a reliable estimation of the anomaly exponent α [94,95]. In order to eliminate both types of errors, we employ a resampling approach as proposed by Weiss [71]. Namely, we analyzed nucleosome trajectories within a real and control mask with minimum lengths 20 or 30 frames. We employed two different trajectory lengths to ensure that the approach is robust in terms of number of frames. For each dataset, we resampled the trajectories, taking only even or odd positions, resulting in the two trajectories with frame times $2\Delta t$ for each original one with frame time Δt . The TA-MSD was calculated for the original and resampled data, $\delta^2(t_{lag})$. Then a translated MSD functional is computed for each $v(t_{lag}, \Delta t) = \delta^2(t_{lag}) - \delta^2(\Delta t)$, so that the exponent α is found from the relation:

$$v(t_{lag}, 2\Delta t) = 2^\alpha v(t_{lag}, \Delta t) \quad (3.5)$$

where $v(t_{lag}, 2\Delta t)$ is the functional for the resampled data. The exponent α is thus:

$$\alpha = \frac{1}{\log 2} \left\langle \frac{\log[v(t_{lag}, 2\Delta t)]}{\log[v(t_{lag}, \Delta t)]} \right\rangle \quad (3.6)$$

after averaging over different lag times. The from three independent data sets were then averaged to find the anomaly exponent for H3K27ac and RNAP2-Ser5ph data, and for their associated control masks.

Bibliography

- [1] K. Luger, A.W. Mäder, R.K. Richmond, D.F. Sargent, and T.J. Richmond. Crystal structure of the nucleosome core particle at 2.8 a resolution. *Nature*, 389:251260.
- [2] A.J. Andrews and K. Luger. Nucleosome structure(s) and stability: Variations on a theme. *Annu. Rev. Biophys.*, 40:99117.
- [3] P.B. Talbert and S. Henikoff. The yin and yang of histone marks in transcription. *Annu. Rev. Genomics Hum. Genet.*, 22:147170.
- [4] H. Kimura. Histone modifications for human epigenome analysis. *J. Hum. Genet.*, 58:439445.
- [5] B.D. Strahl and C.D. Allis. The language of covalent histone modifications. *Nature*, 403:4145.
- [6] S. Buratowski. Progression through the rna polymerase ii ctd cycle. *Mol. Cell*, 36:541546.
- [7] K.M. Harlen and L.S. Churchman. The code and beyond: transcription regulation by the rna polymerase ii carboxy-terminal domain. *Nat. Rev. Mol. Cell Biol.*, 18:263273.
- [8] M. Heidemann, C. Hintermair, K. VoSS, and D. Eick. Dynamic phosphorylation patterns of rna polymerase ii ctd during transcription. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.*, page 5562.
- [9] N.J. Fuda, M.B. Ardehali, and J.T. Lis. Defining mechanisms that regulate rna polymerase ii transcription in vivo. *Nature*, 461:186192.
- [10] M.L. Nosella and J.D. Forman-Kay. Phosphorylation-dependent regulation of messenger rna transcription, processing and translation within biomolecular condensates. *Curr. Opin. Cell Biol.*, 69:3040.
- [11] R.S. Isaac and L.S. Churchman. The long and the short of the rna polymerase c-terminal domain and phase separation. *Mol. Cell*, 73:10871088.

- [12] H. Kimura and Y. Sato. Imaging transcription elongation dynamics by new technologies unveils the organization of initiation and elongation in transcription factories. *Curr. Opin. Cell Biol*, 74:7179.
- [13] T. Henriques. Widespread transcriptional pausing and elongation control at enhancers. *Genes Dev*, 32:2641.
- [14] L. Core and K. Adelman. Promoter-proximal pausing of rna polymerase ii: a nexus of gene regulation. *Genes Dev*, 33:960982.
- [15] A.L. Clayton, C.A. Hazzalin, and L.C. Mahadevan. Enhanced histone acetylation and transcription: A dynamic perspective. *Mol. Cell*, 23:289296.
- [16] V.G. Allfrey, R. Faulkner, and A.E. Mirsky. Acetylation and methylation of histones and their possible role in the regulation of rna synthesis. *Proc. Natl. Acad. Sci*, 51:786794.
- [17] A.J. Bannister and T. Kouzarides. Regulation of chromatin by histone modifications. *Cell Res*, 21:381395.
- [18] G.A. Josling, S.A. Selvarajah, M. Petter, and M.F. Duffy. The role of bromodomain proteins in regulating gene expression. *Genes*, 3:320343.
- [19] J. Otterstrom. Super-resolution microscopy reveals how histone tail acetylation affects dna compaction within nucleosomes in vivo. *Nucleic Acids Res*, 47:84708484.
- [20] N.T. Crump. Dynamic acetylation of all lysine-4 trimethylated histone h3 is evolutionarily conserved and mediated by p300/cbp. *Proc. Natl. Acad. Sci*, 108:78147819.
- [21] Y. Hayashi-Takanaka. Tracking epigenetic histone modifications in single cells using fab-based live endogenous modification labeling. *Nucleic Acids Res*, 39:64756488.
- [22] T.J. Stasevich. Regulation of rna polymerase ii activation by histone acetylation in single living cells. *Nature*, 516:272275.

- [23] H. Kimura, Y. Hayashi-Takanaka, T.J. Stasevich, and Y. Sato. Visualizing posttranslational and epigenetic modifications of endogenous proteins in vivo. *Histochem. Cell Biol*, 144:101109.
- [24] Y. Sato, M. Nakao, and H. Kimura. Live-cell imaging probes to track chromatin modification dynamics. *Microscopy*, 70:415422.
- [25] S. Conic. Imaging of native transcription factors and histone phosphorylation at high resolution in live cells. *J. Cell Biol*, 217:15371552.
- [26] S. Conic. Visualization of endogenous transcription factors in single cells using an antibody electroporation-based imaging approach. In Y. Shav-Tal, editor, *Imaging Gene Expression*, volume 2038, page 209221. Springer New York.
- [27] H. Ma. Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. *Nat. Biotechnol*, 34:528530.
- [28] B. Chen. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*, 155:14791491.
- [29] I. Tasan. CRISPR/Cas9-mediated knock-in of an optimized tetO repeat for live cell imaging of endogenous loci. *Nucleic Acids Res*, 46:100 100.
- [30] A.F. Straight, A.S. Belmont, C.C. Robinett, and A.W. Murray. GFP tagging of budding yeast chromosomes reveals that protein-protein interactions can mediate sister chromatid cohesion. *Curr. Biol*, 6:15991608.
- [31] X. Darzacq. Imaging transcription in living cells. *Annu. Rev. Biophys*, 38:173196.
- [32] T. Germier, S. Audibert, S. Kocanova, D. Lane, and K. Bystricky. Real-time imaging of specific genomic loci in eukaryotic cells using the anchor DNA labelling system. *Methods*, 142:1623.

- [33] H. Ochiai, T. Sugawara, and T. Yamamoto. Simultaneous live imaging of the transcription and nuclear position of specific genes. *Nucleic Acids Res*, 43:127.
- [34] H. Chen. Dynamic interplay between enhancerpromoter topology and gene activity. *Nat. Genet*, 50:12961303.
- [35] D. Mazza, A. Abernathy, N. Golob, T. Morisaki, and J.G. McNally. A benchmark for chromatin binding measurements in live cells. *Nucleic Acids Res*, 40:119 119.
- [36] M. Tokunaga, N. Imamoto, and K. Sakata-Sogawa. Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nat. Methods*, 5:159161.
- [37] J.B. Grimm. A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nat. Methods*, 12:244250.
- [38] T. Nozaki. Dynamic organization of chromatin domains revealed by super-resolution live-cell imaging. *Mol. Cell*, 67:282293 7.
- [39] P.A. Gómez-García. Mesoscale modeling and single-nucleosome tracking reveal remodeling of clutch folding and dynamics in stem cell differentiation. *Cell Rep*, 34, 108614.
- [40] S. Hihara. Local nucleosome dynamics facilitate chromatin accessibility in living mammalian cells. *Cell Rep*, 2:16451656.
- [41] S. Ide, S. Tamura, and K. Maeshima. Chromatin behavior in living cells: Lessons from single-nucleosome imaging and tracking. *BioEssays*, 44:2200043.
- [42] M. Lakadamyali. Single nucleosome tracking to study chromatin plasticity. *Curr. Opin. Cell Biol*, 74:2328.
- [43] J. Lerner. Two-parameter mobility assessments discriminate diverse regulatory factor behaviors in chromatin. *Mol. Cell*, 79:677688 6.
- [44] P.R. Cook. A model for all genomes: The role of transcription factories. *J. Mol. Biol*, 395:110.

- [45] P.R. Cook. *Principles of Nuclear Structure and Function*. Wiley-Liss.
- [46] B. Lim and M.S. Levine. Enhancer-promoter communication: hubs or loops? *curr. Opin. Genet. Dev*, 67:59.
- [47] H. Ma. Cell cycle and genomic distancedependent dynamics of a discrete chromosomal region. *J. Cell Biol*, 218:14671477.
- [48] R. Nagashima. Single nucleosome imaging reveals loose genome chromatin networks via active rna polymerase ii. *J. Cell Biol*, 218:15111530.
- [49] W.-K. Cho. Rna polymerase ii cluster dynamics predict mrna output in living cells. *eLife*, 5:13617.
- [50] W.-K. Cho. Mediator and rna polymerase ii clusters associate in transcription-dependent condensates. *Science*, 361:412415.
- [51] S. Chong. Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science*, 361:2555.
- [52] B. Gu. Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. *Science*, 359:10501055.
- [53] K. Lyon and T.J. Stasevich. Imaging translational and post-translational gene regulatory dynamics in living cells with antibody-based probes. *Trends Genet. TIG*.
- [54] O. Bogdanovi. Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. *Genome Res*, 22:20432053.
- [55] M.P. Creighton. Histone h3k27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci*, 107:2193121936.
- [56] A. Rada-Iglesias. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, 470:279283.

- [57] J.B. Grimm. A general method to improve fluorophores using deuterated auxochromes. *JACS Au*, 1:690696.
- [58] J.H. Shin, R.W. Li, Y. Gao, R. Baldwin, and C. Li. Genome-wide chip-seq mapping and analysis reveal butyrate-induced acetylation of h3k9 and h3k27 correlated with transcription activity in bovine cells. *Funct. Integr. Genomics*, 12:119130.
- [59] B. Bintu. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, 362:1783.
- [60] Q. Szabo, F. Bantignies, and G. Cavalli. Principles of genome folding into topologically associating domains. *Sci. Adv*, 5:1668.
- [61] J.C. Hansen, K. Maeshima, and M.J. Hendzel. The solid and liquid states of chromatin. *Epigenetics Chromatin*, 14:50.
- [62] C. Manzo and M.F. Garcia-Parajo. A review of progress in single particle tracking: from methods to biophysical insights. *Rep. Prog. Phys*, 78:124601.
- [63] H. Shen. Single particle tracking: From theory to biophysical applications. *Chem. Rev*, 117:73317376.
- [64] A. Sabri, X. Xu, D. Krapf, and M. Weiss. Elucidating the origin of heterogeneous anomalous diffusion in the cytoplasm of mammalian cells. *Phys. Rev. Lett*, 125:058101.
- [65] G. Muñoz-Gil. Objective comparison of methods to decode anomalous diffusion. *Nat. Commun*, 12:6253.
- [66] T. Germier. Real-time imaging of a single gene reveals transcription-initiated local confinement. *Biophys. J*, 113:13831394.
- [67] L. Carlini, A. Benke, L. Reymond, G. Lukinavicius, and S. Manley. Reduced dyes enhance single-molecule localization density for live superresolution imaging. *ChemPhysChem*, 15:750755.

- [68] R. Metzler, J.-H. Jeon, A.G. Cherstvy, and E. Barkai. Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking. *Phys. Chem. Chem. Phys.*, 16:2412824164.
- [69] D. Krapf. Chapter five - mechanisms underlying anomalous diffusion in the plasma membrane. In A.K. Kenworthy, editor, *Current Topics in Membranes*, volume 75, page 167207. Academic Press.
- [70] G.A. Lovely. Direct observation of rag recombinase recruitment to chromatin and the igh locus in live pro-b cells. Preprint at.
- [71] M. Weiss. Resampling single-particle tracking data eliminates localization errors and reveals proper diffusion anomalies. *Phys. Rev. E*, 100:042125.
- [72] M.A. Glozak, N. Sengupta, X. Zhang, and E. Seto. Acetylation and deacetylation of non-histone proteins. *Gene*, 363:1523.
- [73] T.L. Lenstra, J. Rodriguez, H. Chen, and D.R. Larson. Transcription dynamics in living cells. *Annu. Rev. Biophys.*, 45:2547.
- [74] A. Raj, C.S. Peskin, D. Tranchina, D.Y. Vargas, and S. Tyagi. Stochastic mrna synthesis in mammalian cells. *PLoS Biol.*, 4:309.
- [75] Z.S. Singer. Dynamic heterogeneity and dna methylation in embryonic stem cells. *Mol. Cell*, 55:319331.
- [76] L.S. Forero-Quintero. Live-cell imaging reveals the spatiotemporal organization of endogenous rna polymerase ii phosphorylation at a single gene. *Nat. Commun.*, 12:3158.
- [77] X. Darzacq. In vivo dynamics of rna polymerase ii transcription. *Nat. Struct. Mol. Biol.*, 14:796806.
- [78] E. Bertrand. Localization of ash1 mrna particles in living yeast. *Mol. Cell*, 2:437445.

- [79] M.-T. Wei. Nucleated transcriptional condensates amplify gene expression. *Nat. Cell Biol*, 22:11871196.
- [80] S. Leidescher. Spatial organization of transcribed eukaryotic genes. *Nat. Cell Biol*, 24:327339.
- [81] S. Uchino. Live imaging of transcription sites using an elongating rna polymerase iispecific probe. *J. Cell Biol*, 221:202104134.
- [82] R. Marmorstein and M.-M. Zhou. Writers and readers of histone acetylation: Structure, mechanism, and inhibition. cold spring harb. *Perspect. Biol*, 6:018762.
- [83] T. Zhang, Z. Zhang, Q. Dong, J. Xiong, and B. Zhu. Histone h3k27 acetylation is dispensable for enhancer activity in mouse embryonic stem cells. *Genome Biol*, 21:45.
- [84] A. Sankar. Histone editing elucidates the functional roles of h3k27 methylation and acetylation in mammals. *Nat. Genet*, 54:754760.
- [85] Q. Jin. Distinct roles of gcn5/pcaf-mediated h3k9ac and cbp/p300-mediated h3k18/27ac in nuclear receptor transactivation. *EMBO J*, 30:249.
- [86] C.A. Cialek, G. Galindo, A.L. Koch, M.N. Saxton, and T.J. Stasevich. Bead loading proteins and nucleic acids into adherent human cells. *J. Vis. Exp*, e62559.
- [87] J. Schindelin. Fiji: an open-source platform for biological-image analysis. *Nat. Methods*, 9:676682.
- [88] I. Arganda-Carreras. Consistent and elastic registration of histological sections using vector-spline regularization. In R.R. Beichel and M. Sonka, editors, *Computer Vision Approaches to Medical Image Analysis*, page 8595. Springer.
- [89] D. Ershov. Trackmate 7: integrating state-of-the-art segmentation algorithms into tracking pipelines. *Nat. Methods*, 19:829832.

- [90] P. Virtanen. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nat. Methods*, 17:261272.
- [91] A.D. Edelstein. Advanced methods of microscope control using manager software. *J. Biol. Methods*, 1:10.
- [92] J.-Y. Tinevez. Trackmate: An open and extensible platform for single-particle tracking. *Methods*, 115:8090.
- [93] A.V. Weigel, B. Simon, M.M. Tamkun, and D. Krapf. Ergodic and nonergodic processes coexist in the plasma membrane as observed by single-molecule tracking. *Proc. Natl. Acad. Sci*, 108:64386443.
- [94] T. Savin and P.S. Doyle. Static and dynamic errors in particle tracking microrheology. *Biophys. J*, 88:623638.
- [95] F. Etoc. Non-specific interactions govern cytosolic diffusion of nanosized objects in mammalian cells. *Nat. Mater*, 17:740746.

Chapter 4

Bead loading proteins and nucleic acids into adherent human cells

4.1 Summary

Bead loading is an inexpensive assay for loading membrane, impermeable particles into live adherent cells. This protocol has proven extremely useful for loading probes for single cell or a single molecule or fluorescence microscopy experiments. Bead loading allows researchers to quickly load all sorts of molecules, including proteins, DNA, RNA, synthetic particles, or a combination of these simultaneously into single cells. ³

4.2 How to introduce plasmids and proteins into cells

Loading macromolecules into mammalian cells necessitates methodology which allows them to cross the cells plasma membrane (Stewart et al., 2016). Several methods can introduce plasmids into mammalian cells through transfection, including liposomal transfection (Felgner et al., 1987) and DEAE-dextran transfection (Schenborn Goiffon, 2000). However, methods for loading proteins or membrane-impermeable particles into cells are more limited.

Several techniques have bypassed this difficult hurdle using various strategies. First, microinjection delivers particles through a micropipette into cells live on a microscope (Celis, 1984). While arguably the most controlled and least invasive method, this technique is relatively low throughput because cells must be loaded one-by-one. Further, microinjection requires specialized

³Chapter 4 is adapted from Cialek, C. A., Galindo, G., Koch, A. L., Saxton, M. N., Stasevich, T. J. Bead Loading Proteins and Nucleic Acids into Adherent Human Cells. *J. Vis. Exp.* (172), e62559, doi:10.3791/62559 (2021). Author contributions: C.A.C. acquired/analyzed all data, did all experiments, and made all graphs, and made Table 4.1 (except imaging HeLa and RPE1 cells in 4.4A,B, contributed by M.N.S.). A.L.K., C.A.C. M.N.S. and T.J.S. wrote the manuscript. G.G. and M.N.S. designed and created the figures. A.L.K., C.A.C., G.G., M.N.S. and T.J.S. edited the manuscript and figures. Dr. Phil Fox and Dr. Linda Forero gave valuable advice on bead loading different cell types.

equipment and is technically demanding. Second, electroporation is a way to electro-inject proteins into cells via voltage-induced membrane disruption (Chakrabarti et al., 1989; Potter, 2003; Wilson et al., 1991). However, this method again requires specialized, expensive equipment and the shock can cause cell stress and mortality. Further, cells must be trypsinized before electroporation and subsequently replated, limiting the timeframe at which cells can be investigated post-electroporation. Third, cell membranes may be chemically modified for temporary reversible permeabilization (Fawell et al., 1994; Prior et al., 1992). Steptolysin-O (SLO) loading inserts an endotoxin into cell membranes which forms temporary pores, allowing exogenous membrane-impermeable particles, including proteins and DNA plasmids, to enter cells (Walev et al., 2001). After a 2 hour recovery, about half the cells repair these pores and halt internalizing particles from the solution. However, this technique requires a long recovery time and is incompatible with cell types that cannot tolerate endotoxins. Fourth, mechanical disruption loads particles into cells through physical perturbation of the cell membrane (Pitchiaya et al., 2012). This can be done in multiple ways, including scratching, scraping, and rolling beads atop cells (McNeil et al., 1984; Ortiz et al., 1987). As early as 1987, beads have been used to mechanically load proteins into cells (McNeil Warder, 1987b). More recently, we have optimized and adapted the bead loading technique beyond proteins to include loading of plasmids and RNA, as described here.

Bead loading is an easy, inexpensive, and fast method for loading protein and plasmids into adherent human cells. Glass beads are briefly rolled atop cells, temporarily disrupting their cellular membrane. This allows particles in solution to enter. Since bead loading has low efficiency, it is best suited for single-molecule or single-cell microscopy experiments. Bead loading can introduce a wide variety of proteins, including fragmented antibodies (Fab; (Hayashi-Takanaka et al., 2011a; T. Morisaki et al., 2016)) purified proteins like scFvs (Tanenbaum et al., 2014), intrabodies (Jedlitzke Mootz, 2021), or mRNA coat proteins like MCP (Coulon et al., 2014; Pichon et al., 2020). Plasmid expression vectors can also be added to the protein solution and bead loaded simultaneously (Koch et al., 2020a; Moon, 2020; Moon et al., 2019). Beyond proteins and plasmids, molecules as large as 250 nm polystyrene beads have been introduced into cells via bead loading

(personal correspondence). Bead loading is incredibly inexpensive, costing less than a penny per experiment in materials and requiring no additional expensive equipment. The cost is further reduced by minimizing the amount of probes used per experiment, since only the cells in the center 14 mm-diameter microwell of an imaging chamber are loaded. It should be noted that the limited loading area means that bead loading is not ideal for bulk-cell loading.

In this chapter, we demonstrate the bead loading process, including both how to construct the bead loading apparatus and how to perform an experiment (Fig. 4.1). We show that proteins, RNA, and DNA can be loaded into various cell types. Further, we demonstrate both by representative images and quantification that two different proteins simultaneously bead loaded have highly correlated cellular concentrations and relatively low variance. We discuss variations in the protocol based on cell type and loading of protein, plasmid or RNA. Though beads are thought to perforate and disrupt the cell membrane, when performed properly, the bead loading process dislodges only a small number of cells from the bottom of the imaging chamber and after a short recovery period cells continue to grow and divide. This methodology is ideal for live-cell microscopy experiments, including single-molecule protein and RNA tracking, post-translational modification detection, observation of dynamic cellular mechanisms, or subcellular localization monitoring (Forero-Quintero et al., 2020; Hayashi-Takanaka et al., 2011a; Koch et al., 2020a; Lyon et al., 2019b; T. Morisaki et al., 2016).

4.3 Bead loading proteins and plasmids into adherent human cells

The most common application of bead loading is to introduce one or more types of protein into adherent human cells. To illustrate this, we bead loaded cells with a solution of a Cy3- and Alexa488-conjugated Fab protein. Though not every cell in the microwell was bead loaded, the cells that were loaded almost always had both Cy3- and Alexa488-labeled proteins together (Fig 4.2A). According to an earlier estimate, when 1 μl of 0.5 mg/mL (2.5 μm) Fab is bead loaded

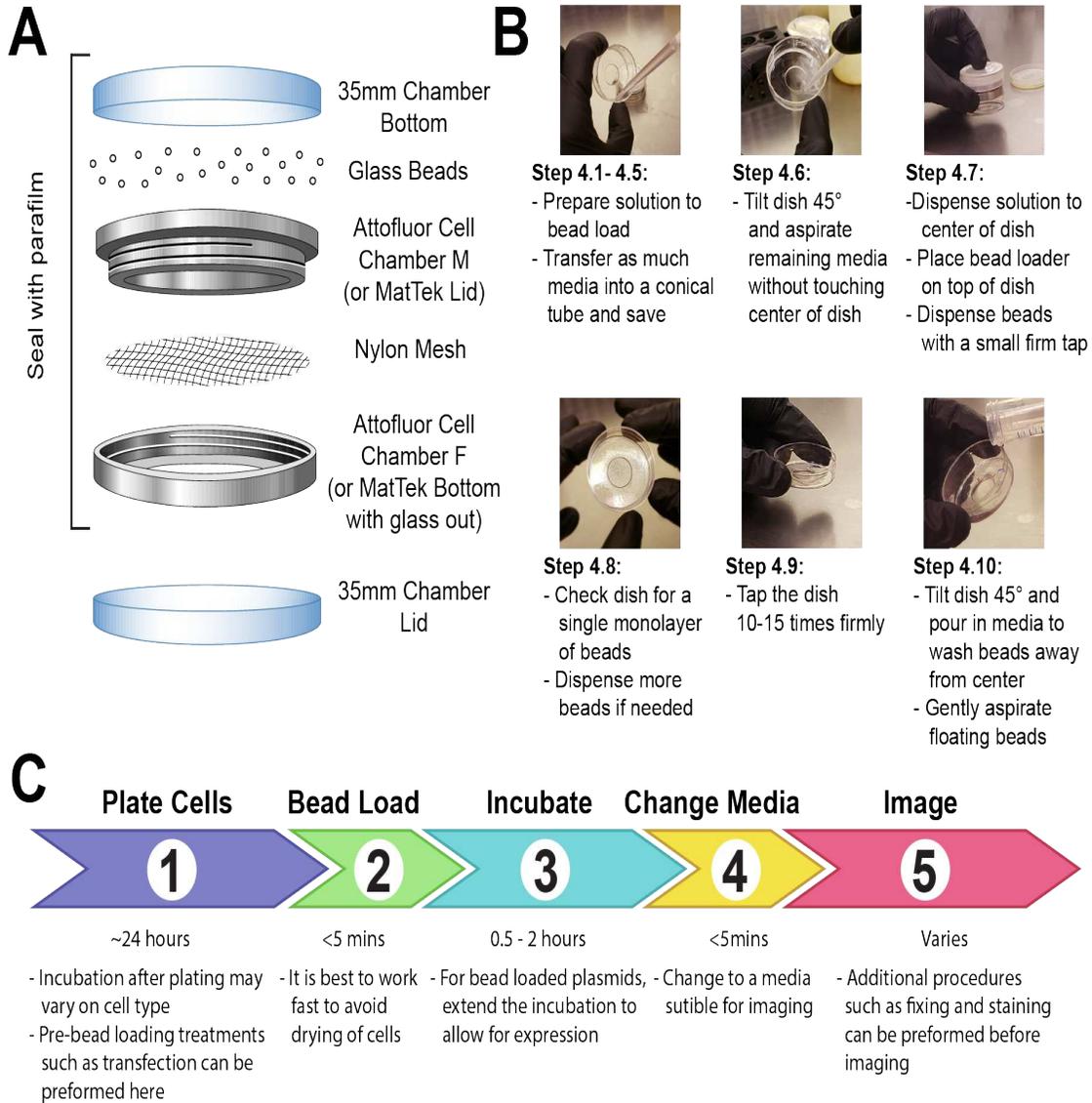


Fig. 4.1: Bead loading apparatus, technique, and timeline..

(Hayashi-Takanaka et al., 2009), as in Fig. 4.2A, each cell is loaded with roughly a million Fab molecules.

Plasmid DNA encoding GFP (1 μg of plasmid DNA, 1.8 microliters of 557 $\text{ng}/\mu\text{l}$) added with 0.5 μg Cy3-labeled protein was also introduced into cells via bead loading and subsequently expressed and visualized (Fig. 4.2B). The GFP fluorescence represented that the GFP-encoded plasmid was not only loaded into cells but also expressed. Thus, in the same cell, bead loading can introduce a protein probe (like Cy3-labeled Fab) and reporter plasmid (like GFP), as performed in our lab previously.(Koch et al., 2020a; Moon, 2020; Moon et al., 2019) We quantified 40 percent of cells were bead loaded with Fab protein and 21 percent of the bead-loaded cells expressed the co-loaded plasmid, as shown in the representative fields-of-view in Fig. 4.2B. We typically load each chamber with 1-2 μg of plasmid, approximately the same amount as lipofections. Bead-loaded cells express widely varying levels of plasmids (Fig. 4.2C,D). To specifically measure this, we used the Fisher Ratio test to compare the distributions of protein and plasmid intensity data. We found that though the protein 1 and 2 had similar intensity distributions to one another ($p = 1$) each protein had a significantly smaller distribution than the plasmid ($p = 3.2\text{e-}6$ and $1.8\text{e-}5$). Though this could be from variability in how many plasmids are loaded per cell, we suspect that the greater source of variability comes from the many steps required for plasmid expression – including being imported into the cell nucleus, transcription, and translation – that are likely to vary greatly between cells. In contrast, the levels of bead-loaded proteins had little cell-to-cell variance, and the levels of two simultaneously loaded proteins highly correlated with each other (Fig. 4.2C-E). Plasmid expression can be seen as early as 2-4 hours post bead loading but may occur later depending on when optimal plasmid expression is obtained. We recommend performing a time course to determine the best window of expression for a specific plasmid spanning 2-24 hours post bead loading. This can be done in one chamber with long timeframe imaging or by bead loading and staggering multiple chambers.

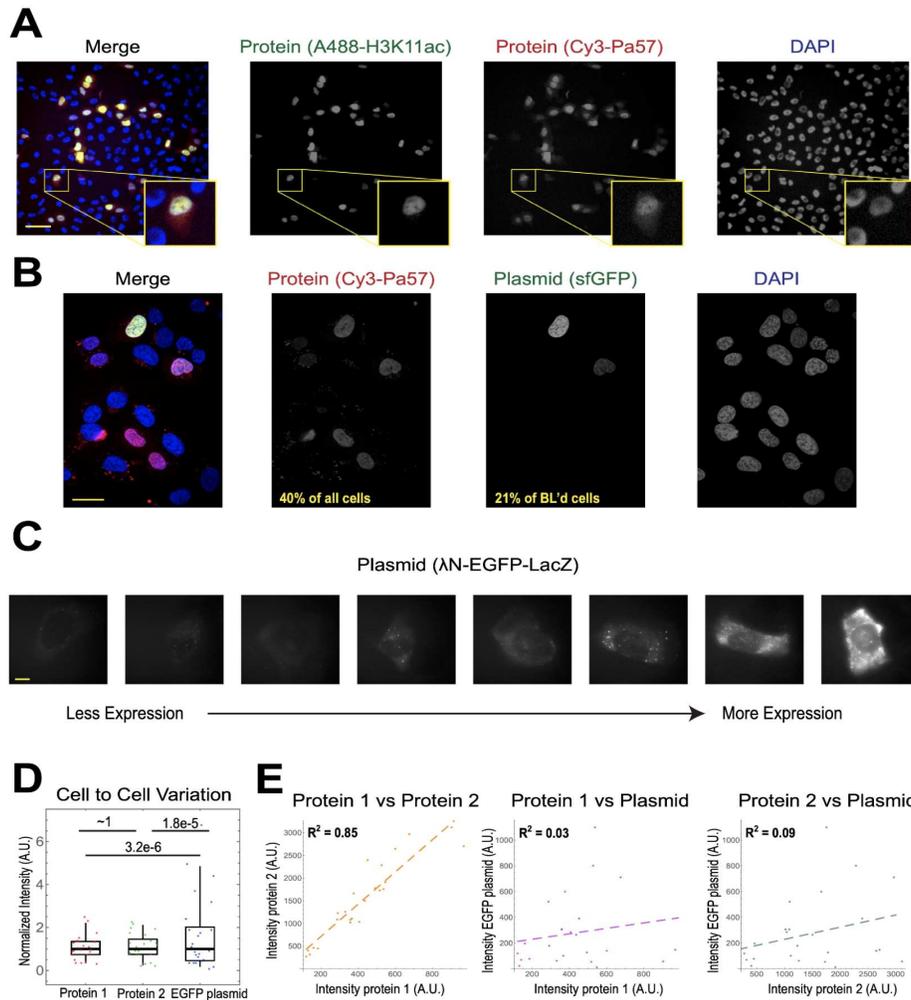


Fig. 4.2: Bead loading introduces low variability in protein concentration but high variability in plasmid expression. (a) Cells were bead loaded with $0.5 \mu\text{g}$ of each of the proteins Alexa488-conjugated anti-H3K27Acetyl Fab (green) and Cy3-conjugated anti-RNAPII-Serine5phosphorylation Fab (red) in $4 \mu\text{l}$ of bead loading solution. Cells were DAPI-stained (blue) then live-imaged immediately. Scale bar measures $20 \mu\text{m}$. (b) Cells were bead loaded with $0.5 \mu\text{g}$ of Fab protein (Cy3-conjugated anti-RNAPII-Serine5phosphorylation, red) and $1 \mu\text{g}$ of plasmid encoding superfoldGFP-H2B (green) in $4 \mu\text{l}$ of bead loading solution. After 24 hours, cells were DAPI-stained (blue) and imaged live. Scale bar measures $30 \mu\text{m}$. (c-e) Protein 1 (JF646-HaloLigand-labeled HaloTag-MCP), protein 2 (Cy3-conjugated anti-FLAG Fab) and a plasmid encoding EGFP (N-EGFP-LacZ) were bead loaded together into cells. The total intensity in each fluorescent channel was measured in a $1.3 \times 1.3 \mu\text{m}$ patch in the cytoplasm of each cell. $N = 25$ cells. (c) Representative cells expressing the bead loaded plasmid, N-EGFP-LacZ. The same imaging conditions and intensities were used for all cells. Spots are aggregates of the expressed protein. Scale bar measures $10 \mu\text{m}$. (d) The chart shows each cells total intensity of either protein 1, protein 2, or EGFP expressed from the plasmid. Each channel was normalized to the median. Bonferroni-corrected P values were calculated by the Fisher Ratio test, comparing if the distribution of protein or plasmid intensity data has the same variability. Each point represents a cell. (e) The total intensities for either both proteins, protein 1 and the plasmid, or protein 2 and the plasmid are plotted against each other. Calculated R^2 values are displayed. Each point represents a cell.

4.4 Bead loading and cell health

Bead-loaded cells remain adherent and are healthy enough to grow and divide. Bead-loaded human U2OS cells were imaged directly before, directly after, and 24 hours after bead loading. Proper bead loading had almost no noticeable effect on the number of cells or their morphology, as shown in this representative field of view (Fig. 4.3A, left, middle). In contrast, poor bead loading with too many beads and excessive tapping force is depicted in Fig. 4.3B. This caused much cell loss (large patches of the coverglass without cells and detached, floating, out-of-focus cells), poor cell morphology (cells appearing rounded up and poorly adhered), and clusters of beads remaining on the coverglass after bead loading. Though cells are thought to undergo mechanical damage during bead loading, cells grew and proliferated in the properly bead loaded chamber, as evidenced by the increased number of cells 24 hours after bead loading (Fig. 4.3A, right). The magnitude of the effect to cell viability can be assessed through a variety of assays, such as a MTT assay, to compare bead-loaded to mock loaded cells (Kumar et al., 2018). Further, we show here and previously that the bead loaded cells, themselves, undergo cell divisions (Fig. 4.3C) and the timing of mitosis was not affected by bead loading (Stasevich et al., 2014), which serves as further evidence to sustained cell health after bead loading.

4.5 Loading different cell types and materials via bead loading

Bead loading is a versatile technique, accommodating several types of adherent cell lines and various types of macromolecules. We demonstrate this variety by loading RPE1 and HeLa cell lines with Fab (Fig. 4.4A, B). Table 4.1 provides further examples of bead loading in different cell lines, in our lab and beyond, and points out some of the nuanced differences between bead loading protocols from other labs. Of note, the diameter of glass beads used for loading varies greatly between labs, though most efficient loading was found for small, 75 μm diameter beads in several cell lines (McNeil Warder, 1987b). Further, our lab has begun bead of RNA as well (currently unpublished). Fig. 4.4C displays a representative U2OS cell bead-loaded with a Cy5-RNA 9mer and Cy3-DNA 28mer together.

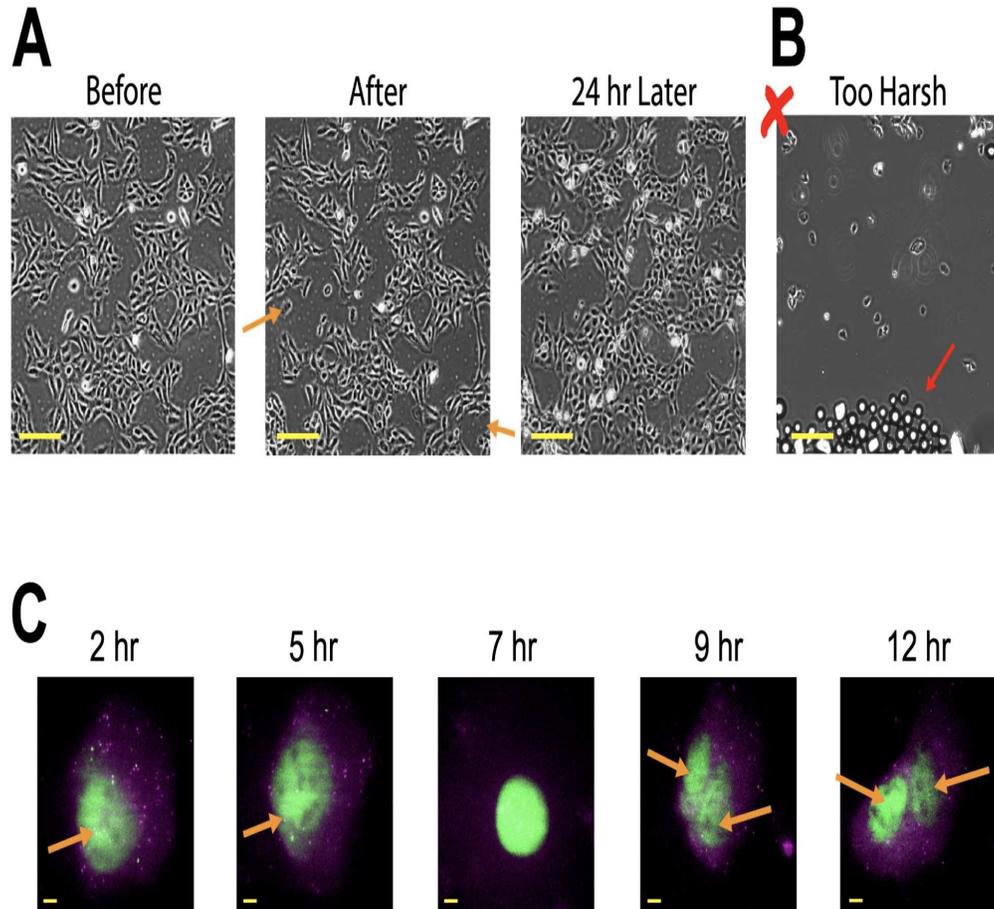


Fig. 4.3: Bead loaded cells remain adherent and are healthy enough to grow and divide.. (a) Cells were bead loaded with $0.5 \mu\text{g}$ of each of the proteins Alexa488-conjugated anti-H3K27Acetyl Fab (green) and Cy3-conjugated anti-RNAPII-Serine5phosphorylation Fab (red) in $4 \mu\text{l}$ of bead loading solution. Cells were DAPI-stained (blue) then live-imaged immediately. Scale bar measures $20 \mu\text{m}$. (b) Representative image of U2OS cells bead loaded with components from A but with harsh tapping and too many beads. The red arrow identifies extra glass beads. Scale bar measures 2mm . (c) U2OS cells were loaded with $1.5 \mu\text{g}$ of the 14.4 kbp plasmid smFLAG-KDM5B-15xBoxB-24xMS2, $0.5 \mu\text{g}$ of Cy3 conjugated anti-FLAG Fab (green), 130ng HaloTag-MCP (magenta) in $8 \mu\text{l}$ of bead loading solution. Directly before imaging, the HaloTag was stained with JF646-HaloLigand. The MS2 stem loops of the mRNA transcribed from the reporter plasmid are labeled by MCP (magenta spots), and FLAG-tagged translated reporter protein is labeled via anti-FLAG Fab (green colocalization to mRNA). Mature Fab-labeled protein localizes to the nucleus. This cell was imaged 4-15 hours after bead loading. Yellow arrows identify the cell nucleus before and nuclei after cell division. Scale bar measures $5 \mu\text{m}$.

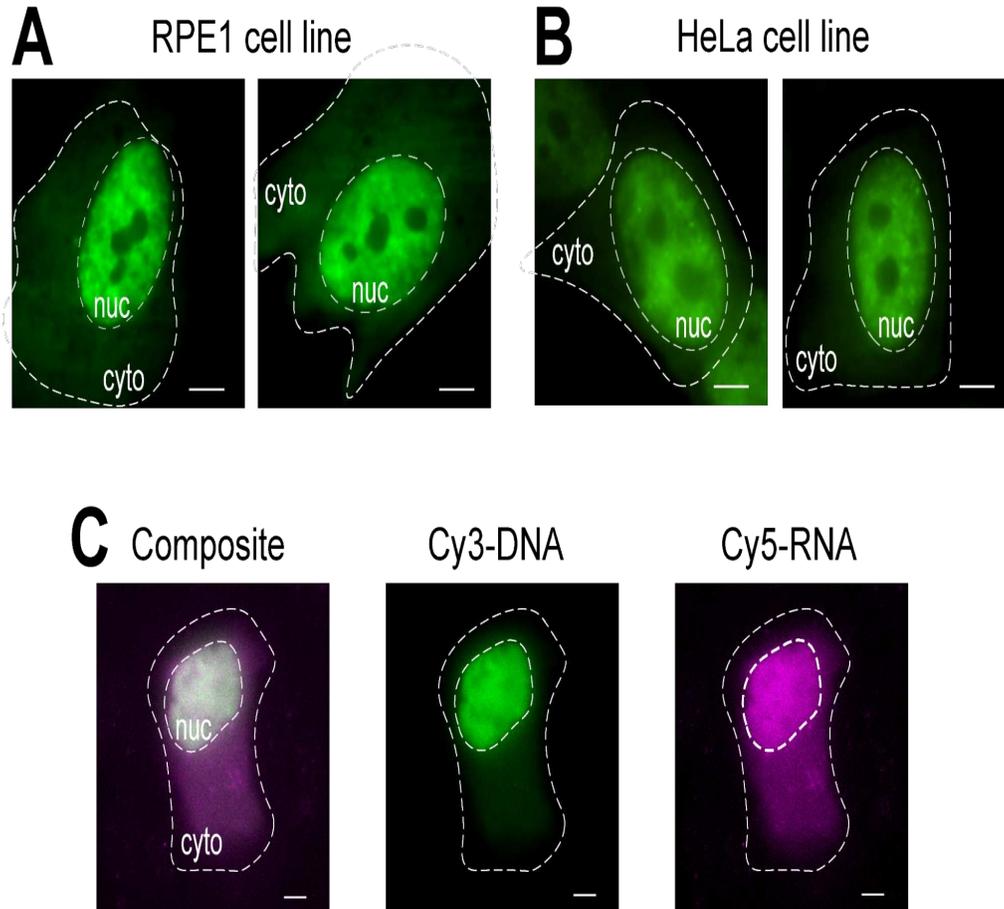


Fig. 4.4: Bead loaded cells remain adherent and are healthy enough to grow and divide.. (a-b) RPE1 (A) and HeLa (B) cells were bead loaded with $1.5 \mu\text{g}$ of a nuclear Fab protein (anti-RNAPII-Serine5-phosphorylation) in a $4 \mu\text{l}$ loading solution. Each cells nucleus (nuc) and cytoplasm (cyto) are marked. Cells were imaged 6 hours after being bead loaded. Scale bars measure $5 \mu\text{m}$. **(c)** Human U2OS cells were bead loaded with both Cy5-RNA 9mer (magenta) and Cy3-DNA 28mer (green) oligos, 10 picomoles of each, in $4 \mu\text{l}$ bead loading solution. Cells were imaged 4 hours after being bead loaded. All cell nuclei are highlighted by a dashed line. The cells nucleus (nuc) and cytoplasm (cyto) are marked. Scale bar measures $5 \mu\text{m}$.

4.6 Bead loading of live, adherent cells is inexpensive and versatile

The bead loading technique described here is a cost-effective and time-efficient method for introducing macromolecules and other types of particles into adherent cells. This versatile process can load protein (Fig. 4.2A; (Forero-Quintero et al., 2020; Hayashi-Takanaka et al., 2011a; Lyon et al., 2019b; T. Morisaki et al., 2016)), a combination of protein and plasmids (Fig. 4.2B-E; (Koch et al., 2020a)), RNA (Fig. 4.4C), 100 and 250 nm polystyrene beads (Ashok Prasad lab, unpublished), synthetic dyes (Jones et al., 2011) or quantum dots (Sabri, 2020; T. Emerson et al., 2014). Bead loading may have the capability to load other types of membrane-impermeable particles as well. Its most-used application is for loading antibodies or Fabs to target endogenous epitopes, like post-translational modifications (PTMs), in live cells. Targets like PTMs are often difficult to label in live cells without established PTM-specific genetically encoded tags, such as a mintbodies (Sato et al., 2013, 2018). In contrast, bead loading can introduce multiple types of probes, reporters, or other molecular tools together into the same cell for monitoring multiple readouts simultaneously. We anticipate that bead loading will be a useful technique for loading a variety of macromolecules or particles.

A major advantage of bead loading is the low cost: each experiment costs less than a penny. A bead loader apparatus can be made easily using inexpensive materials costing in total about \$150, which is significantly less expensive than any other cell loading method. The cost of a bead loader apparatus can be further reduced to under \$10 by replacing the metal reusable chamber with a plastic one. For this, either drill a hole in a 35 mm chamber or remove the glass from a 35 mm glass-bottom chamber, then securely fasten the mesh in place with tape. In lieu of an apparatus, bead loading can even be performed using a wide-bore p-1000 pipette tip to scoop and sprinkle beads onto cells, though this variation adds variability in the size and number of beads.

A major benefit of bead loading is that cells retain normal overall morphology, recover rapidly, and continue to grow and divide for the U2OS, RPE1, and HeLa cells studied here and other

cell lines studied elsewhere (Fig. 4.3, Fig. 4.4A,B, Table 4.1; (Stasevich et al., 2014)). During bead loading, cells undergo physical stress and sometimes dislodge and peel (5 percent of cells peel under optimal conditions, but greater cell loss can happen if bead loading is performed too forcefully or too many glass beads are loaded atop the cells, depicted in Fig. 4.3B). However, cells appear healthy and can be imaged as soon as 30 minutes after bead loading (Fig. 4.3A). We generally give cells a 30-minute recovery period but anticipate that imaging sooner post-bead loading is feasible. A major drawback of this technique is that the cells need to be capable of withstanding minor physical stress during loading and remain securely adhered to the coverslip. Poorly/non-adherent cell lines or cells grown on coated plates (e.g. HEK and stem cells) detach upon the gentle tapping during bead loading. Further, in our experience, primary neurons are too sensitive for bead loading.

Bead loading is best for single-cell or single-molecule experiments due to its low efficiency. In our experience, bead loading demonstrated around 20-40 percent protein loading efficiency, and about 20 percent of the bead loaded cells also expressed a co-loaded plasmid (Fig. 4.2A,B). We speculate that bead loaded plasmids had low efficiency because plasmids must not only enter cells but also be expressed. Increasing the amount of plasmid in the bead loading solution may increase how many cells express the plasmid, yet also could increase the level of expression per transfected cell and cause overexpression. The low efficiency of bead loaded plasmid expression can be circumvented by using alternative transfection protocols, such as lipofection, prior to bead loading the proteins or probes (Lyon et al., 2019b; T. Morisaki et al., 2016). Additionally, we noticed that incubating cells in optimal media for 30 minutes before bead loading may assist plasmid expression. Due to low plasmid expression, we do not use bead loading as an alternative method to lipofection exogenous expression. However, for cells that are unresponsive or intolerant to lipofection, bead loading may provide an alternate, albeit low-efficiency, method for transient plasmid expression.

4.7 Protocol

4.7.1 Clean, sterilize and dry glass beads so that they do not clump and can spread evenly atop cells.

1.1 Sterilize approximately 5 mL glass beads in NaOH. Measure the beads in a 50 mL conical tube. Add 25 mL of 2 M NaOH and mix gently using a shaker or rotator for 2 hours.

1.2 Decant the NaOH, retaining as many beads as possible. If the beads are in suspension, spin down the tube of beads briefly in a centrifuge (1 minute at 1000 rcf).

1.3 Wash the beads thoroughly with cell culture-grade water until the pH is neutral (use pH test strip on the eluent to confirm a neutral pH). Decant the wash water each time, as before.

1.4 Wash the beads thoroughly with 100% ethanol 2-3 times. Decant the ethanol each time, as before.

1.5 Dry the beads. Sprinkle the beads to form a thin layer inside a sterile container (such as a 10 cm petri dish). Leaving the container open, let the beads air dry in a biosafety cabinet overnight. Ensure that the beads are completely dry by tapping or gently shaking the container and checking that the beads have a sandy texture with no clumping or flaking.

1.6 UV-sterilize the dry beads for 15 minutes.

4.7.2 Assemble the bead loader apparatus.

2.1 Fasten a patch of mesh (polypropylene or equivalent material, 105 μm openings to allow beads to pass through) to cover the entire opening of the beads holding chamber with either tape or clamping the mesh between male and female ends of a metal reusable imaging chamber (Fig. 4.1A).

2.3 UV-sterilize the apparatus for 15 minutes.

2.4 Add beads to the apparatus and seal tightly with parafilm. NOTE: It is essential that the beads are completely clean and dry at this step. They should be loose and look sandy with no clumps. If they do not appear so, re-wash and completely dry the beads.

2.5 Store the apparatus in a sealed, dry container desiccated by silica gel or other desiccant medium. If the beads become damp, which will be apparent by bead clumping, thoroughly dry and sterilize the bead loader and replace with fresh beads. This will prevent any mold or bacteria from growing on or around the beads within the bead loader. NOTE: The bead loader apparatus can be made in different ways. See details in Chapter 4.5.

4.7.3 Prepare glass-bottom chambers of adherent cells

3.1 Seed adherent mammalian cells onto a 35 mm glass-bottom chamber. Ensure cells are approximately 80 percent confluent at the time of bead loading. (See Table 4.1 for more information on various cell types and notes on the effectiveness of bead loading in different cell types.) NOTE: Cells can be seeded in only the microwell in the center of the chamber to conserve how many cells are used.

3.2 Incubate the cells under normal conditions until they are completely adhered to the glass. NOTE: It is essential that cell density is high enough and that the cells are securely adhered to the glass. If these requirements are not met, cells will likely peel off during bead loading. The timeline between cell seeding and bead loading can be lengthened to ensure proper cell adhesion and confluency.

4.7.4 Bead loading cells

Optional: Wash cells briefly with PBS then add 2 mL optimal media. Incubate for at least 30 m.

4.1 Make a solution of 3-8 μ l containing desired plasmids, protein, and/or particles. Use 1 μ g (0.1 - 1 pmol) of each type of plasmid and 0.5 μ g (0.01 nmol) protein, depending on experimental requirements. Use a low-retention tube for proteins so that they are not left behind on tube walls. Bring the solution up to a minimum of 3 μ l with PBS. Adjust solution volume as necessary to coat the entire area of cells to be loaded (i.e. chambers microwell, Fig. 41B).

4.2 Mix the solution thoroughly by pipetting up and down and/or flicking the tube.

4.3 Briefly spin the solution down to the bottom of the tube in a tabletop microfuge.

4.4 Move bead loading solution and chamber of cells into a tissue culture hood. Perform the remaining steps in the tissue culture hood using sterile technique.

4.5 Remove media from cells and temporarily store it in a sterile tube.

4.6 Gently suck off all media from around the edges of the chamber. Tilt the chamber at approximately a 45° angle and suck off the remaining drop of media in the center microwell. During media removal, make sure not to let the pipette tip touch the glass, which may result in cell peeling and loss. Move quickly to the next step so that the cells are not dry for long.

4.7 Gently pipette the bead loading solution onto the glass microwell in the center of the chamber. Optional: Incubate with gentle rocking for 30 seconds, without allowing the chamber to dry up completely.

4.8 Gently disperse a monolayer of glass beads on top of the cells, preferably using a bead loading apparatus (Fig. 4.1A). The beads should cover the cells in the glass-bottom microwell completely.

4.9 Pinching the chamber with two fingers, tap it against the hood surface by lifting it 2 inches and bringing it down firmly. Use a force approximately equivalent to dropping the dish from that height. Repeat for 10 taps total. Note: Ensure that taps do not substantially peel the cells. Tapping can be optimized for cell type. If cells load poorly, tap harder, but if many cells peel off, tap more lightly.

4.10 Gently add media back into the chamber by pipetting slowly onto the plastic side of the chamber. Try to suck off any floating beads without disturbing cells. Add more pre-warmed media at this step if too much was removed.

4.11 Incubate cells for 0.5 - 2 hours in the incubator.

4.12 UV-sterilize the bead loader for 15 minutes before returning it to storage under desiccating conditions.

4.13 Add dye (e.g. DAPI or HaloTag ligand stain, if required by experiment) as per manufacturer's recommended protocol.

4.14 Wash the cells 3 times with media before imaging to remove beads and excess loading components in solution. Avoid pipetting directly onto cells to keep them from peeling.

4.7.5 Imaging the bead loaded cells

5.1 Image the cells immediately or when required by the experiment. Use a microscope capable of capturing fluorescence (lasers or monochromatic light source). Ensure that the excitation wavelengths are appropriate for the chosen fluorophore or dye (e.g. 488 nm wavelength light for GFP). Note: Bead loaded proteins may be imaged once the cells have recovered (as soon as 30 minutes post loading for the cell lines described here). Plasmid expression takes 2 or more hours depending on expression vector elements (Fig. 4.1C, and further explanation in the Chapter 4.5). Note: Imaging of bead-loaded cells can be performed on any microscope equipped with the appropriate fluorescent sources associated with loaded probes, a camera capable of capturing fluorescence images, such as an EMCCD or sCMOS camera, and an incubator to control temperature, humidity and carbon dioxide.

Bibliography

- [1] M.P. Stewart. In vitro and ex vivo strategies for intracellular delivery. *Nature*, 538(7624):183192.
- [2] P.L. Felgner. Lipofection: a highly efficient, lipid-mediated dna-transfection procedure. *Proceedings of the National Academy of Sciences of the United States of America*, 84(21):74137417.
- [3] E.T. Schenborn and V. Goiffon. Deae-dextran transfection of mammalian cultured cells. *Methods in Molecular Biology*, 130:147153.
- [4] J.E. Celis. Microinjection of somatic cells with micropipettes: comparison with other transfer techniques. *Biochemical Journal*, 223(2):281291.
- [5] R. Chakrabarti, D.E. Wylie, and S.M. Schuster. Transfer of monoclonal antibodies into mammalian cells by electroporation. *Journal of Biological Chemistry*, 264(26):1549415500.
- [6] A.K. Wilson, J. Horwitz, and P. Lanerolle. Evaluation of the electroinjection method for introducing proteins into living cells. *American Journal of Physiology*, 260(2):355363.
- [7] H. Potter. Transfection by electroporation. *Current Protocols in Molecular Biology*, 62(1):16.
- [8] S. Fawell. Tat-mediated delivery of heterologous proteins into cells. *Proceedings of the National Academy of Sciences of the United States of America*, 91(2):664668.
- [9] T.I. Prior, D.J. FitzGerald, and I. Pastan. Translocation mediated by domain ii of pseudomonas exotoxin a: transport of barnase into the cytosol. *Biochemistry*, 31(14):35553559.
- [10] I. Walev. Delivery of proteins into living cells by reversible membrane permeabilization with streptolysin-o. *Proceedings of the National Academy of Sciences of the United States of America*, 98(6):31853190.

- [11] S. Pitchiaya, J.R. Androsavich, and N.G. Walter. Intracellular single molecule microscopy reveals two kinetically distinct pathways for microRNA assembly. *EMBO Reports*, 13(8):709715.
- [12] P.L. McNeil, R.F. Murphy, F. Lanni, and D.L. Taylor. A method for incorporating macromolecules into adherent cells. *The Journal of Cell Biology*, 98(4):15561564.
- [13] D. Ortiz, M.M. Baldwin, and J.J. Lucas. Transient correction of genetic defects in cultured animal cells by introduction of functional proteins. *Molecular and Cellular Biology*, 7(8):30123017.
- [14] P.L. McNeil and E. Warder. Glass beads load macromolecules into living cells. *Journal of Cell Science*, 88(5):669678.
- [15] Y. Hayashi-Takanaka. Tracking epigenetic histone modifications in single cells using fab-based live endogenous modification labeling. *Nucleic Acids Research*, 39(15):64756488.
- [16] T. Morisaki. Real-time quantification of single rna translation dynamics in living cells. *Science*, 352(6292):14251429.
- [17] M.E. Tanenbaum, L.A. Gilbert, L.S. Qi, J.S. Weissman, and R.D. Vale. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell*, 159(3):635646.
- [18] N. Zhao. A genetically encoded probe for imaging nascent and mature ha-tagged proteins in vivo. *Nature Communications*, 10(1):2947.
- [19] B. Jedlitzke and H.D. Mootz. Photocaged nanobodies delivered into cells for light activation of biological processes. *ChemPhotoChem*, 5(1):2225.
- [20] A. Coulon. Kinetic competition during the transcription cycle results in stochastic rna processing. *eLife*, 3:03939.
- [21] New generations of ms2 variants and mcp fusions to detect single mrnas in living eukaryotic cells. *Methods in Molecular Biology*, 2166:121144.

- [22] A. Koch. Quantifying the dynamics of ires and cap translation with single-molecule resolution in live cells. *Nature Structural Molecular Biology*, 27:10951104.
- [23] S.L. Moon. Multicolor single-molecule tracking of mrna interactions with rnp granules. *Nature cell biology*, 21(2):162168.
- [24] S.L. Moon. Coupling of translation quality control and mrna targeting to stress granules. *Journal of Cell Biology*, 219(8).
- [25] C.A. Cialek. Imaging translational control by argonaute with single-molecule resolution in live cells.
- [26] L.S. Forero-Quintero. Live-cell imaging reveals the spatiotemporal organization of endogenous rna polymerase ii phosphorylation at a single gene.
- [27] K. Lyon, L.U. Aguilera, T. Morisaki, B. Munsky, and T.J. Stasevich. Live-cell single rna imaging reveals bursts of translational frameshifting. *Molecular Cell*, 75(1):172183.
- [28] JoVE. Introduction to fluorescence microscopy. general laboratory techniques. *JoVE Science Education Database*, JoVE.
- [29] Y. Hayashi-Takanaka, K. Yamagata, N. Nozaki, and H. Kimura. Visualizing histone modifications in living cells: spatiotemporal dynamics of h3 phosphorylation during interphase. *Journal of Cell Biology*, 187(6):781790.
- [30] P. Kumar, A. Nagarajan, and P.D. Uchil. *Analysis of cell viability by the MTT assay*, volume 6. Cold Spring Harbor Protocols.
- [31] T.J. Stasevich. Regulation of rna polymerase ii activation by histone acetylation in single living cells. *Nature*, 516(7530):272275.
- [32] Export of a toxoplasma gondii rhostry neck protein complex at the host cell membrane to form the moving junction during invasion. *PLOS Pathogens*, 5(2).

- [33] A.P. Gilmore and L.H. Romer. Inhibition of focal adhesion kinase (fak) signaling in focal adhesions decreases cell motility and proliferation. *Molecular Biology of the Cell*, 7(8):12091224.
- [34] Mechanodelivery of nanoparticles to the cytoplasm of living cells. *Nanoscale*, 6(9):45384543.
- [35] E.M.M. Manders, H. Kimura, and P.R. Cook. Direct imaging of dna in living cells reveals the dynamics of chromosome formation. *Journal of Cell Biology*, 144(5):813822.
- [36] S. Memedula and A.S. Belmont. Sequential recruitment of hat and swi/snf components to condensed chromatin by vp16. *Current Biology*, 13(3):241246.
- [37] T. Becker, A. Volchuk, and J.E. Rothman. Differential use of endoplasmic reticulum membrane for phagocytosis in j774 macrophages. *Proceedings of the National Academy of Sciences of the United States of America*, 102(11):40224026.
- [38] C. Molenaar. Visualizing telomere dynamics in living mammalian cells using pna probes. *The EMBO Journal*, 22(24):66316641.
- [39] S. A. Jones, S. Shim, J. He, and X. Zhuang. Fast, three-dimensional super-resolution imaging of live cells. *Nature Methods*, 8(6):499505.
- [40] A. Sabri, X. Xu, D. Krapf, and M Weiss. Elucidating the origin of heterogeneous anomalous diffusion in the cytoplasm of mammalian cells. *Physical Review Letters*, 125(5):053901.
- [41] Y. Sato. Genetically encoded system to track histone modification in vivo. *Scientific Reports*, 3:2436.
- [42] Y. Sato, T.J. Stasevich, and H. Kimura. Visualizing the dynamics of inactive x chromosomes in living cells using antibody-based fluorescent probes. In *X-Chromosome Inactivation. Methods in Molecular Biology. 1861, Humana*, page 91102. New York, NY, USA.

Chapter 5

Final Conclusions and Future Directions

Overall, each chapter of this thesis has a significant portion of space dedicated to discussion of the work therein, including interpretation of the data and implications for the field as a whole. I do not wish to rehash those conclusions, however, there some components of this work that deserve further focused attention. I will lay out my general thoughts as well as considerations for future experimentation here.

5.1 The masking process: applications and considerations

5.1.1 Masking as a methodology

In Chapter 3, a process for utilizing local adaptive binarization on fluorescence microscopy images to characterize the location of a specific cellular activity is described. In this case, transcription initiation and histone acetylation. Given the multiplexed nature of fluorescence microscopy, this allowed for tracking of single nucleosomes in a cellular context inside of single, living cells. This is a powerful technique, combining the powerful spatiotemporal resolution of single-molecule imaging with the specificity of antibody-based labeling (i.e. Fab, intrabodies etc.). However, there are some important considerations when applying this type of masking.

First, it is important that there is significant signal to noise ratio in the image to be masked. With any type of thresholding, if the image is noisy then there will be significant noise in the mask. As a mask is a binary image, you thereby lose the gradation of intensity values, and there is less certainty that all components of the mask are reflective of truly enriched regions. Given that we utilized Fab, which bind transiently, there is a significant number of freely diffusing Fab which contribute to background noise. In both image modalities utilized in Chapter 3, signal-to-noise ratio was increased by averaging a series of images together. This totaled 100 images

with relatively low exposure when doing fast, single-molecule imaging, or 5 images with higher exposure time when doing long time course imaging.

It is also important to balance the need for signal with how frequently a mask is updated. Update frequency should be primarily calculated based upon how often significant changes in the signal are expected. For example, if transcription has a typical lifetime of minutes, mask frequency may not be very important if imaging quickly, and masks could even be spread out to preserve cell health and prevent photobleaching. In the case of acetylation, turn over can be quite rapid, meaning that frequent imaging is crucial. In each case, it is also important to remember auxiliary factors such as chromatin drift.

When considering co-localization from two imaging channels within a mask, it is also important to consider imaging volume. That is, the combination of axial and lateral resolution in each pixel. In this case the same general principles apply as colocalizing two point sources. Meaning that if a component of a mask is very small, there is much less certainty that the factor being masked is colocalized with another channel. For this reason, it is important to eliminate speckle noise by applying a size or speckle filter. This has the added benefit of also reducing speckle noise originating from low signal to noise ratio, as low SNR tends to result in mask components with a relatively small area.

Lastly, it is important to consider the effects of masking when tracking individual molecules in another channel. When performing diffusion calculations on particles found within a mask, the mask will create a natural bias which selects for immobile particles. Although we believed this would likely not be an issue with chromatin, one of the most immobile molecules in the nucleus, we still found a significant effect. Thus, it is crucial to create a control mask when performing diffusion analysis. This was done by moving each component of a mask by a randomized vector, essentially creating a new nuclear mask with the same general morphology as the original mask. Extra caution should be taken if applying this methodology to fast moving molecules (i.e. transcription factors), as a mask would naturally select for more immobile elements and thus may be artefactually interpreted as increased binding.

All combined, however, as long as the above can be properly controlled and accounted for, multi-color imaging with local adaptive binarization masking is a promising technical solution for defining enrichment in fluorescent images. This technique therefore has many broad and interesting potential applications, some of which I will discuss below.

5.1.2 Histone modifications and masking

In Chapter 3.3, the mobility of individual nucleosomes inside of regions enriched for H3K27ac was measured. It was found that this chromatin was 15% more mobile. Now that this technology has been validated and developed, the next logical step is to expand this analysis to other histone post-translational modifications and characterize chromatin dynamics within enriched regions. It would be particularly interesting to examine H3K27-trimethyl (H3K27me3). Unlike H3K27ac, H3K27me3 is associated with gene silencing and the formation of compact, transcriptionally repressive chromatin domains. Given that H3K27ac and H3K27me3 are mutually exclusive modifications, it would be interesting to compare the two directly inside of living cells.

As this work primarily focused on characterizing transcriptionally active chromatin (H3K27ac and RNAP2-Ser5ph enriched), examining H3K27me3 enriched regions would provide insight into the dynamics of heterochromatic regions. It would be predicted, based upon standard models of heterochromatin, that H3K27me3 enriched chromatin should be relatively immobile, to prevent interactions with TFs and to restrict enhancer promoter interactions. However, H3K27me3 associated genes are not constitutively silenced, so they may exhibit unique behavior. This behavior could be compared to regions of known permanent silencing, such as the inactive X-locus, to test for functional difference in heterochromatic states.

Lastly, many histone PTM writing proteins are known to interact with specific modifications. It would be interesting to examine the binding behavior of these enzymes within PTM enriched regions, as they may exhibit specific behavior or demonstrate a propensity to perpetuate specific marks. Ultimately, however, masking technology provides a promising platform for interrogating the behavior and relationship between many combinations of post-translational modifications.

5.1.3 Other applications and considerations

One primary limitation of expanding the analysis of PTMs is the nature of how they are detected. Simply, Fabs are created from antibodies which must themselves be created. This can be a very involved process, involving extensive testing and validation to insure proper specificity. As a secondary consideration, even if commercially available, the amount of antibody necessary to create a significant portion of Fab can be expensive. While there is a suite of already developed antibodies for a host of post-translational modifications, there are other potential processes for antibody development and production. Specifically, if the polypeptide sequence of an antibody is already known, then the complementary determining region (CDR) sequences can be combined with mouse or rabbit antibody scaffolds, and then expressed and harvested from HEK293 cells. While some preliminary progress has been made on this end, further testing and validation is required for histone PTM specific antibodies.

Fortunately, however, the masking process is applicable to biological questions beyond post-translational modifications. As long as a region of biological activity can be marked, the power of single-molecule tracking can be leveraged to interrogate the behavior of protein or RNA associated with any targetable cellular activity. Ultimately, the combination of masking and single-molecule microscopy is a powerful and versatile tool for studying the nature of highly dynamic cellular life.