Advanced imaging methods unveil new mechanisms transcriptional regulation in the context of nuclear organization

By

David T McSwiggen

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge
Associate Professor Xavier Darzacq, Co-chair
Professor Robert Tjian, Co-chair
Professor Donald Rio
Professor Britt Glaunsinger

Fall 2019

Abstract

Advanced imaging methods unveil new mechanisms transcriptional regulation in the context of nuclear organization

David Trombley McSwiggen

Co-chairs of the committee:

Associate Professor Xavier Darzacq and Professor Robert Tjian

Department of Molecular and Cell Biology

It has long been appreciated that organization of macromolecules within the cell is paramount to proper cellular function. As the number of tools in the molecular and cell biologist's toolbox grows, the list of ways cells achieve organization also grows. This is especially true for the process of transcription and its regulation, where the fundamental questions remain a challenge to address, but where a convergence of genetic, genomic, structural and imaging techniques offer the possibility of answers. The proper loading of an RNA polymerase on a gene requires the coordinated action of hundreds of proteins and other macro molecules, as well as cellular genome to be maintained in a topologically permissive state; yet as a system, gene regulation requires interactions to be dynamic enough to respond to the changing needs of the cell in response to internal and external ques. How a cell constructs a system that is, on the one hand robust, and on the other hand flexible remains a challenge to answer.

In Chapter 1 of this thesis, we will introduce the study of transcription regulation as a general topic, and discuss a number of fluorescence imaging techniques that have fundamentally changed the our understanding of transcription regulation. Chapter 2 will focus on one particular aspect of cellular organization which has recently come into vogue—that of membraneless compartment formation through liquid-liquid phase separation. A thorough investigation of the literature, particularly in light of the experiments which we foreshadow and introduce in more depth in Chapter 3, suggests that much more work is needed before phase separation as a means of biological organization can become a general paradigm. Chapter 3 presents a particularly poignant example of the issues raised in Chapter 2. Using Herpes Simplex Virus as a model system because of its ability to form compartments in the nucleus, we demonstrate that recruitment of Pol II and other proteins can be explained through the availability of nonspecific binding sites rather than phase separation. Lastly, Chapter 4 will present technical findings on the optimal fluorescent dyes to use for *in vivo* labeling of proteins for imaging applications.

In summary, this dissertation underscores the way in which new imaging techniques reveal new biological principles and challenge old models. It will be of great interest to watch these ideas and techniques mature beyond the work presented in this dissertation.

Table of contents

List of Figures and Tables

Acknowledgements

I use the pronoun “we” rather than “I” frequently throughout this dissertation. I know people have strong feelings about this, especially in the context of a work like a dissertation, which is supposed to chronicle the work of the single author. And yet, this feels wrong. Because while I performed the experiments you’ll read in the following pages (for the most part, anyway), the fact that I held the pipette should not diminish the major contributions that so many people have had as I wandered my way through grad school. The people that helped me, guided me, challenged me, and encouraged me deserve much more than a few lines at the beginning of a book that no one will read. And yet, lacking any better medium to convey just how important these individuals were to the work enclosed in this dissertation, I’d like to put down some words here.

First I need to thank Profs. Robert Tjian and Xavier Darzacq, without whom none of this would have been possible. Thank you for giving me the freedom to explore and follow my own interests in the lab. Thanks for letting me do the experiments that I wanted to do (even the stupid ones), and for your patience while I figured out what those experiments meant (*especially* the stupid ones). I’m sorry that the lab is now stuck with HSV1 for the foreseeable future, the irony of which is not lost on me. I count myself as lucky to have had two advisors who could put up with my bad jokes; who helped me with not only scientific but also pragmatic career advice, and who supported all the steps needed to get me there, even if they took me out of the lab at times.

Thank you, also, to the many UC Berkeley faculty that have shaped my time here, especially to thanks to Prof. Britt Glaunsinger, thesis committee member and adoptive third advisor for everything virus-related and much more besides. Thanks also to Prof. Don Rio, an invaluable thesis committee member whose breadth of expertise and interest drove many great discussions. One of the greatest parts of the Berkeley MCB community is the network of support outside of the official channels. I received advice, collaboration and expertise from Profs. Elçin Ünal, Dirk Hockemeyer, Jamie Cate, Gloria Brar, David Savage, and Mike Eisen, and these interactions transformed the way I pursued my PhD.

I’ve been incredibly fortunate through my PhD to have a number of scientific collaborators, both inside of the lab and out. A heartfelt thanks to Prof. Anders Hansen, Prof. Sheila Teves, Hervé Marie-Nelly, Jaclyn Ho, Claire Dugast-Darzacq, Mustafa Mir, Alec Heckert, Franzi Lorbeer, Jingxun Chen, Kunitoshi Chiba, Amy Tresenrider, Mia Pulos-Holmes, and Prof. Amy Lee. I have learned so much through our work together.

I also learned so much from Yvonne Hao and Kayla Umemoto, two amazing undergraduates who are, even as I write this, still manage to both impress and intimidate me with their intelligence and work ethic, albeit working for other people on their way to amazing things. Thank you for your enthusiasm and your hard work, as well as your willingness to forgive my many shortcomings as a mentor.

I have been fortunate to work with so many talented and kind people in the Tjian and Darzacq labs. Y'all were not only coworkers and colleagues, but friends. Thank you especially to the people that keep the lab running smoothly, day and night, including Claudia Cattoglio, Claire Dugast-Darzacq, Ana Robles, Astou Tangara, Carla Inouye, Janeen Lim, Mallory Haggart, Gina Dailey, Shuang Zheng, Eva Bourbon and Lilia Tomas. So much of your work goes without the thanks it deserves. Thank you also to the many post-docs and fellow grad students, particularly Jaclyn Ho, Eliza Zhang, Kevin Tsui, Sheila Teves, Alec Heckert, and Anders Hansen, who tolerated me and my constant questions, bad jokes, and worse ideas with humor and grace.

And, of course, I have made friends through my time here at Cal—so many friends in so many different and unexpected places. I could not possibly name or thank you all here, but you are no less important for all that. To Ella and Luisa, Elena and I could not have survived Cal without you and your families. Thank you both for your science advice and life advice, and for being such amazing role models for what both should look like. To Dan, Amy, and Rose – Follies with you all was a silly, ridiculous, agonizing, hilarious way to spend [what felt like] all of our third year, and remains one of the defining features of my time in grad school. Dom, Caitlin, Rob, Franzi, George, Janice, Orestes, Jamie, Helen, and Ben: Whether talking about science, or food, or about Kumi, or music, or Emmy-nominated animated short films, or medical oddities, or art, or graduate student governance, thanks for always giving me something to look forward to.

Thank you to my mother JoAnne, father Jim, and big sister Hannah. From an early age you encouraged me to experiment and to ask questions. Unfortunately, it turns out the kitchen is a great place to experiment, which makes for some delicious meals and some fun nights out, but ultimately a less-than-enviable career. Thank you for patiently supporting me, even when "the guy who thinks he's the Bobby Flay of the Longmont sushi scene" was the best way to describe me. Thank you especially to my father, Jim, who read every manuscript I wrote throughout my research career at least twice, providing not only an invaluable outside perspective but also an amazingly cost-effective copy-editing service.

Lastly, thank you to my wife Elena. Though you will likely never read it, this whole dissertation is dedicated you. I cannot imagine going through grad school without someone to come home to, to complain to, to celebrate with, and to commiserate with. You have so often been the responsible adult for the two of us, modeling good grown-up behavior so that I can try to follow along. Most of all, you get both the credit and the blame for the single thing I'm most proud of from my PhD: Our daughter. She means all the world to me, and I'm so excited to hand in this one last homework assignment so that we can move on to the next chapter, growing as a family and doing fun things along the way.

Thanks.

Chapter One: Single-molecule microscopy approaches to studying transcription: What are they, and why bother?

Abstract

Transcription is a fundamental process to all organisms, and its regulation has been an area of intense study. In recent years, a number of techniques to assay transcription and its regulation at the single-molecule level have arisen, which have fundamentally changed the field. This chapter briefly introduces transcription, and the technological developments allowing for assays with single-molecule sensitivity, including methods for achieving the signal necessary for single-molecule sensitivity. We then discuss three examples of cases where singe-molecule techniques such as single molecule RNA Fluorescence In Situ Hybridization (smFISH), Single Particle Tracking (SPT), or Single Molecule Localization Microscopy (SMLM) assays have provided unique insight into a particular aspect of transcription regulation.

A general overview of transcription

Transcription—the process of reading DNA-encoded information to produce a complementary RNA molecule containing nominally the same information—is a process fundamental to all known living organisms. Being central to nearly all downstream cellular processes, a great deal of effort has been spent trying to understand the molecular mechanism underlying this process. Broadly, this field of study can be subdivided into two types of questions. The first question is how, mechanistically, transcription is facilitated by the enzymes involved. A great deal has been learned over the past decades about the physical mechanism of transcription as a general process, and excellent reviews exist for both prokaryotic and eukaryotic systems (McClure, 1985; Ruff et al., 2015; Thomas and Chiang, 2006). Answers to the first broad question tell us that minimally an RNA polymerase needs a template as a substrate, NTPs, and some mechanism of accessing the Watson-Crick base-pairing hydrogen bonds of single-stranded DNA (Sydow and Cramer, 2009; Thomas and Chiang, 2006).

The answers to how a generic RNA polymerase can function at a generic locus have given rise to another broad class of questions: In a system where the polymerase machinery is, by necessity, agnostic to the underlying sequence it transcribes, how can a cell achieve specificity in gene expression? It is clear that much evolutionary work has occurred for the cell to solve this problem. While all of the biochemical requirements for a polymerase can be efficiently satisfied by a single polypeptide like that of the T7 bacteriophage (Zhang et al., 2014), both prokaryotes and eukaryotes have evolved multi-subunit polymerase machinery as a means to confer specificity and regulatability onto the system.

The canonical view of eukaryotic transcription regulation, as illustrated in molecular biology textbooks, identifies a large number of players (Figure 1). Broadly speaking, transcription of a given gene is orchestrated through interactions between the DNA encoding the gene and its proximal regulatory elements (e.g. DNA-encoded core promoter elements), distal regulatory elements like enhancers which may be tens to thousands of kilobases away in linear sequence, and the protein factors that recognize these sequence elements. All-told, this process requires potentially hundreds of separate molecules to converge in space. Consequently, each one of these protein complexes provides additional means through which the system may be regulated.

While the list continues to grow, some important players are the sequence-specific transcription factors, chromatin modifying enzymes, general transcription factors like the TATA-binding protein (TBP), and the polymerase itself (Thomas and Chiang, 2006; Vannini, 2012). In the case of mRNA production carried out by RNA Polymerase II (Pol II), additional important protein complexes include the Mediator complex, specific modules of the general transcription factors (GTFs) like TFIID/SAGA, and regulators of transcription that function after polymerase engagement, such as the DSIF/NELF or P-TEFb complexes(Allen and Taatjes, 2015; Ikeda et al., 2015; Izeddin et al., 2014; Levine et al., 2014; Lis, 2019; Thomas and Chiang, 2006; Vos et al., 2018). Further, nuclear components that control the three-dimensional organization of chromatin are now appreciated to play a role in determining gene expression, though the exact degree to

which nuclear organization impacts gene expression remains a subject of ongoing investigation (Hansen et al., 2018a).

It has been the work of hundreds of labs to dissect and understand the myriad ways a cell can regulate transcription through the action of different members of the transcription machinery. Many experimental tools, including genetic screens, genomic deep-sequencing experiments, biochemical and structural biology studies have been employed, and excellent reviews exist which detail historical milestones as well as the advances that have been made in recent years (Levine et al., 2014; Lis, 2019); it is not the goal of this thesis to cover these in depth. The subfield of single-molecule imaging, in particular, has made great strides which have fundamentally altered the way we envision the process of transcription. With the advances in technology of both genetically encoded and organic fluorophores, microscope optics, data processing, and clever experimental trickery, we now have a clearer vision on how gene regulation occurs in space and time. In this chapter, we will examine a few key advances, and present three case studies that highlight important insights derived from single-molecule assays.

Important advances for single molecule microscopy

"Single-molecule" is an umbrella term which applies to many different types of assays: single molecule Fluorescence In Situ Hybridizaiton (smFISH) for detecting RNA in cells; single molecule localization microscopy (SMLM) for locating individual molecules with high precision; single particle tracking (SPT) for observing how molecules move throughout the cell. And while "single-molecule" is not limited to fluorescence microscopy, here we will only address fluorescence techniques as they are the assays employed in subsequent chapters of this work.

Though these assays can be quite different in the types of molecules in question and the types of information they ultimately provide, a number of common challenges apply across these different assays. In order to visualize single molecules using fluorescence in cells, a number of technological hurdles must be overcome. Most importantly, there needs to be a way of specifically introducing a fluorescent moiety onto the molecule in question, either genetically or chemically. The fluorescence emitted from this molecule must be bright enough, or the background noise low enough, that the signal can be readily identified; and furthermore the density of fluorescent molecules in the sample must be low enough that their point spread functions (PSF; the statistical description of how photons from a point source will appear on a detector after passing through a given optical path) do not overlap so as to allow unambiguous identification (Shen et al., 2017). No one single advance has solved all of these problems, but rather separate contributions from a number of sources have provided synergistic gains in our ability to apply single molecule techniques to biological problems.

Labeling a molecule of interest is the first challenge. In fixed cells, there are a number of options available, especially if the molecule of interest is DNA or RNA. Here, one can perform smFISH using fluorescently labeled oligonucleotides to specifically hybridize with the sequence in question can (Raj et al., 2006). The signal intensity can be boosted by tiling many non-overlapping oligos to increase the number of fluorophores per molecule of DNA or RNA in

question. For visualizing proteins or other macromolecules, antibodies raised against the protein of interest can be used for immunofluorescence, fluorophores can be incorporated through copper-assisted click chemistry against noncanonical amino acids, or genetically-encoded fluorescent proteins (FP) can be appended to the protein of interest at the level of the coding sequence.

For live-cell imaging, the majority of these techniques are unavailable, and so genetically-encoded fluorescent proteins become by and large the only practical option. Proteins may be tagged simply by adding the coding sequence of the FP in-frame with the coding sequence of the protein of interest. Ideally this would be done at the endogenous locus, now that genome editing techniques are relatively straightforward and standardized (Doudna and Charpentier, 2014), though often the convenience of ectopic overexpression prevails. Nucleic acids can be labeled through the expression of a DNA- or RNA-binding protein fused to an FP—for example, Cas9 for DNA, or the MS2 coat protein binding MS2 stem loops for RNA (Qin et al., 2017; Wu et al., 2016). These systems are more involved, and as is the case for smFISH, the signal to identify individual molecules often derives from tiling or otherwise increasing the number of FPs localized to the DNA or RNA site. In all cases, careful controls need to be performed to ensure that the introduction of a tag is not affecting protein function, or causing issues with its own transcription or translation.

Unfortunately, even genetically-encoded fluorescent proteins suffer from issues which limit their utility in single molecule assays. First, they often have long maturation times between when the protein is translated and when the necessary amino acids covalently react to form the chromophore(Eason et al., 2017). Second, fluorescent proteins are rather poor in terms of their overall brightness and the fluorophore's longevity, when compared to the organic fluorophores employed in biochemical single-molecule assays (De Zitter et al., 2019). Last, most fluorescent proteins exist in their fluorescent state until they bleach, giving the observer no control over the density of fluorophores in an image. Workarounds have been developed to address some of these issues. For example through the use of the lattice light sheet microscope greatly reduces the background image noise, thereby allowing individual molecule localization (Chen et al., 2014; Mir et al., 2017, 2018); or through the discovery of fluorescent proteins like mEos and Dendra, which photoconvert upon addition of 405 nm light, thus allowing control over molecule density (Jradi and Lavis, 2019). Still, the data generated from fluorescent protein fusions often underscores the room for improvement which the invention of self-labeling protein tags and cell-compatible organic fluorophores has provided.

Self-labeling protein tags, such as HaloTag and SNAPtag, are catalytically half-dead enzymes which recognize specific chemical moieties (Keppler et al., 2003, 2004; Los et al., 2008). In the case of the mutant haloalkane dehalogenase HaloTag, a nucleophilic attack on a chloroalkane substrate forms a covalent enzyme-substrate molecule. The second half of the wildtype enzymatic reaction, which would coordinate a second nucleophilic attack of a water molecule to regenerate the enzyme and release the substrate, has been mutated away, leaving HaloTag impotent for this step (Los et al., 2008). The result is a protein tag which covalently and specifically interacts with

a molecule that does not exist endogenously in most biological systems, making it a great bioorthogonal chemical handle. In practice, this allows one to add an organic fluorophore like tetramethylrhodamine (TMR) with the chloroalkane moiety to cells expressing HaloTag, and thereby fluorescently tag a protein of interest with something much brighter than a conventional fluorescent protein, without worry over maturation time and with some degree of control over the total concentration of labeled protein. Similar mechanisms exist for SNAPtag, CLIPtag, and a growing list of self-labeling proteins (Gautier et al., 2008; Keppler et al., 2003).

The invention of self-labeling proteins also inspired developments in the dyes used to label them. Until recently, TMR was one of a small handful of molecules which were suitable to work in conjunction with HaloTag, due to limitations based on cell permeability of most organic dyes. Based on this molecular scaffold, Grimm and colleagues synthesized a TMR derivative replacing the N,N-dimethyl groups of the classic tetramethylrhodamine with azetidine rings as in the hopes of limiting pathways for nonradiative decay from the excited state (Grimm et al., 2015). This minor change—a net addition of only two carbon atoms—greatly increases the quantum yield and the extinction coefficient of the dye, generating a fluorescent molecule referred to as Janelia Fluor 549 (JF_{549}) (Grimm et al., 2015). JF_{549} provides high signal-to-noise performance and improved photostability for single molecule experiments. Further modifications to the TMR and SiTMR scaffolds have generated a broad spectrum of colors, allowing more precise matching to the optical setup. Perhaps most importantly, further modifications to make a spirocyclic diazoketone-caged version of the Janelia Fluor dyes (PA-JF_{549} and PA-JF_{646}), allowing for precise control over molecule density for super-resolution and single particle tracking experiments (Grimm et al., 2016).

Advances in molecular labeling have been accompanied by advances in microscope technology. Whereas Janelia Fluor dyes improve data by increasing the brightness of the molecules under study, alternative techniques are used to reduce noise in the data from out-of-focus molecules by only illuminating the region of the cell within the focal plane. Total Internal Reflection Fluorescence (TIRF) microscopy uses on the evanescent wave from reflected light to illuminate only those fluorophores close (<~100 nm) to the coverslip, and is often used in biochemical single molecule assays. By illuminating cells on a coverslip with a beam of coherent laser light close to the critical angle for TIRF, one can effectively generate a sheet of light which only illuminates an ~1 μm cross-section of the cell and achieve similar effects as TIRF but for molecules deeper in the cell (Tokunaga et al., 2008). At the extreme, for three dimensional cultures or whole organisms, a dithered array of Bessel beams applied to the sample from an orthogonal objective can provide flat illumination deep into a sample specifically at the imaging objective's focal plane (Chen et al., 2014). When used in conjunction with self-labeling tags, new organic dyes, and improved computational methods for analyzing data, *in vivo* single-molecule imaging has become a viable and valuable technique for understanding biological processes like transcription.

Single-molecule imaging applied to transcription

Single-molecule studies inside cells have provided new perspectives in how transcription is regulated. This is due to a number of factors, many of which are related to the fact that single-molecule studies are almost by definition also single-cell studies, and because single-molecule techniques often provide the opportunity to quantify desirable parameters. Both of these points allow one to distinguish between many potential hypotheses which could be indistinguishable from population-level assays. In the following, we will briefly address recent examples in which single-molecule data provided important or unexpected findings, and the discuss more broadly the current state of the field as revealed by similar types of experiments.

smFISH uncovers regulatory strategies

Conventional wisdom, as per the central dogma, is that the level of mRNA expression in a cell is tightly correlated with expression of the protein is encodes. Chen, Tresenrider and colleagues uncovered a striking counterexample to this general principle, in which the expression of a 5' extended mRNA isoform leads to decreased mRNA translation and feeds back to affect transcription of the gene in cis (Chen et al., 2017). In this case, the mRNA in question is that of NDC80, a central component of the kinetochore. During meiosis in budding yeast, kinetochore formation must be tightly regulated in order to ensure proper segregation of chromosomes into daughter cells, but the mechanism underlying this regulation was not understood. Through correlated RNA-seq and ribosome profiling experiments, the authors identified NDC80 as a gene whose increased expression paradoxically led to a decrease in translation of the open reading frame. Interestingly, upon entry into meiosis, ribosome profiling footprints mapped to multiple upstream open reading frames in the NDC80 transcript, at the expense of the protein-coding sequence (Figure 2A). Northern blots confirmed that cells switched from one isoform to the other upon entry into meiosis, and then back at the beginning of the meiotic divisions (Figure 2B).

Using single molecule RNA FISH (smFISH), the authors designed experiments where the 5' extended isoforms could be distinguished from the translatable isoforms by designing one set of fluorescent oligos which would hybridize to both types of mRNA, and a second set in another color which would uniquely hybridize to the extended isoform. Each mRNA in appears under the microscope as a diffraction-limited spot, and the extended isoform would have pairs of spots in the two color channels. Using this assay, the authors could conclude that transcription of the two isoforms are inversely correlated at the individual cell level (Figure 2C and D), with transcription from the upstream promoter having a repressive effect in cis on the expression of the downstream mRNA. What's more, expression from the upstream promoter occurs mutually exclusive to the downstream expression, and facilitates the repression of the downstream promoter in cis (Figure 2C and D). While these effects can also be seen in bulk population experiments, smFISH provides the crucial insight that these changes in expression are due to a switch of isoform transcription for all cells, rather than just a unique subpopulation.

Cells can regulate transcription factor binding times

A separate example of the power of single-molecule experiments comes from single particle tracking experiments, where another counterintuitive result underscores the breadth of strategies a

cell may use to regulate its processes. In this case, Ho and colleagues examined the interactions between the DNA repair complex XPC and the enzyme Thymine DNA Glycosylase (TDG) (Ho et al., 2017). Aside from their separate functions as repair enzymes in the Nucleotide Excision Repair and Base Excision Repair pathways, respectively, XPC was previously identified as an important transcriptional coactivator in stem cells and TDG was hypothesized to have an important function in maintaining the proper methylation state of regulatory elements (Fong et al., 2011). The authors find strong evidence that these two protein complexes function together to maintain the gene regulatory landscape of stem cells (Ho et al., 2017).

To this end, the authors perform single particle tracking measurements of HaloTag-labeled TDG to measure the residence time of TDG on DNA. Interestingly, they find that ectopic over-expression of XPC reduces the residence time of TDG relative to wildtype levels, while shRNA knockdown of endogenous XPC extends the residence time (Figure 3A and B). This result is counterintuitive because the biochemical interactions between XPC and TDG would lead may to expect the two would stabilize each other in binding to DNA. Rather, it appears that XPC aides in the turnover of TDG, which helps rationalize decades-old biochemical data suggesting that *in vitro* TDG binds its substrates for hours, essentially becoming a single-turnover enzyme in the absence of XPC (Waters and Swann, 1998; Waters et al., 1999). By examining the lifetimes of individual binding events, these single-molecule measurements provide the first example of such "facilitated eviction" *in vivo*, to date.

Single particle tracking reveals new insights into RNA polymerase behavior

Perhaps the most poignant example of a case where single-molecule experiments have revealed wholly unexpected results is that of the distribution of Pol II and other transcription factors in the nucleus. Though the hypothesis that hubs of gene activity, or "transcription factories", has existed for decades, recent studies have fundamentally altered our view the nucleus, towards a deeper appreciation of the role three-dimensional organization and spatial arrangements play in affecting transcription and its regulation (Cisse et al., 2013; Hansen et al., 2017; Izeddin et al., 2014; Liu et al., 2014). In conjunction with chromosome-capture sequencing methods, single-molecule microscopy has provided pivotal data to support these new models—namely that, rather than considering the nucleus as a well-mixed reaction, the different nuclear constituents are heterogeneously distributed throughout the nuclear volume, having meaningful consequences on the reactions they promote (Woringer and Darzacq, 2018; Woringer et al., 2014).

Multiple single-molecule studies have revealed that many nuclear proteins, in various eukaryotic organisms, form clusters or "hubs" whose concentration can be orders of magnitude higher than adjacent regions just a few hundred nanometers distant (Chong et al., 2018; Cisse et al., 2013; Mir et al., 2017). The exact functional role these transcriptional hubs play is unclear. They may form as a mechanism to promote the formation of productive pre-initiation complexes and efficiently load polymerases on an active gene, or they may form as a consequence of other underlying mechanisms, and distinguishing between these two interpretations remains a challenge. A recent study by Boehning, Darzacq, Rankovic and colleagues attempted to address this question

by proposing that clusters of Pol II, which had been previously shown to dynamically form in the nucleus, occurred as a result of liquid-liquid phase separation (a topic which will be discussed in detail in Chapter 2) through homotypic interactions of its unstructured C-terminal domain (Boehning et al., 2018a). Biochemically purified protein will undergo phase separation in a manner dependent on the number of heptad repeats of the canonical sequence YSPTSPS the protein contains (in yeast, Pol II has 26 degenerate repeats of this seven amino acid sequence, where as human Pol II has 52), supporting this hypothesis (Figure 4A-C).

The most important and compelling experiments of this study are those performed in cells. Human cell lines expressing either a truncated or elongated Pol II CTD (25 repeats or 70 repeats, respectively) were first analyzed using localization microscopy. Analysis of this data revealed that, while all three cell lines display clusters of Pol II in the nucleus, the size and number of clusters is dependent on the length of the CTD. Pol II with a truncated CTD had fewer, smaller clusters, whereas the 70-repeat showed much greater clustering at all length scales, and the wildtype Pol II was in between the two (Figure 4D). Moreover, single particle tracking experiments to measure how frequently Pol II binds DNA and how quickly it diffuses showed that the longer CTD increased the probability of binding DNA and decreased the speed of diffusion throughout the nucleus, and similarly the shortened CTD cause Pol II to bind less often and for it to diffuse faster (Figure 4E and F). All of these data indicate that interactions with the CTD are crucial for the formation of hubs of Pol II in the nucleus, and taken with the *in vitro* data, suggest that these interactions are primarily mediated through homotypic CTD-CTD interactions. Although some data suggest that the size or lifetime of the hub is correlated with gene expression, more work is needed to dissect the functional role these hubs play.

Concluding remarks

The invention of a number of single-molecule techniques have been transformative to the field of transcription, granting access to new types of insights into how cells might implement its regulation. This chapter has discussed examples of assays which measure gene expression with isotype resolution at the level of individual transcription in individual cells; assays which measure the binding interactions of single proteins; and assays which measure the distributions of molecules in the cell. The expanding repertoire of these and other techniques have revealed unexpected findings as to the many ways transcription can be regulated. Only a few examples have been mentioned here, but the growing interest and access to these various techniques will likely provide more detailed understanding of how these and other regulatory mechanisms are implemented.

Figures

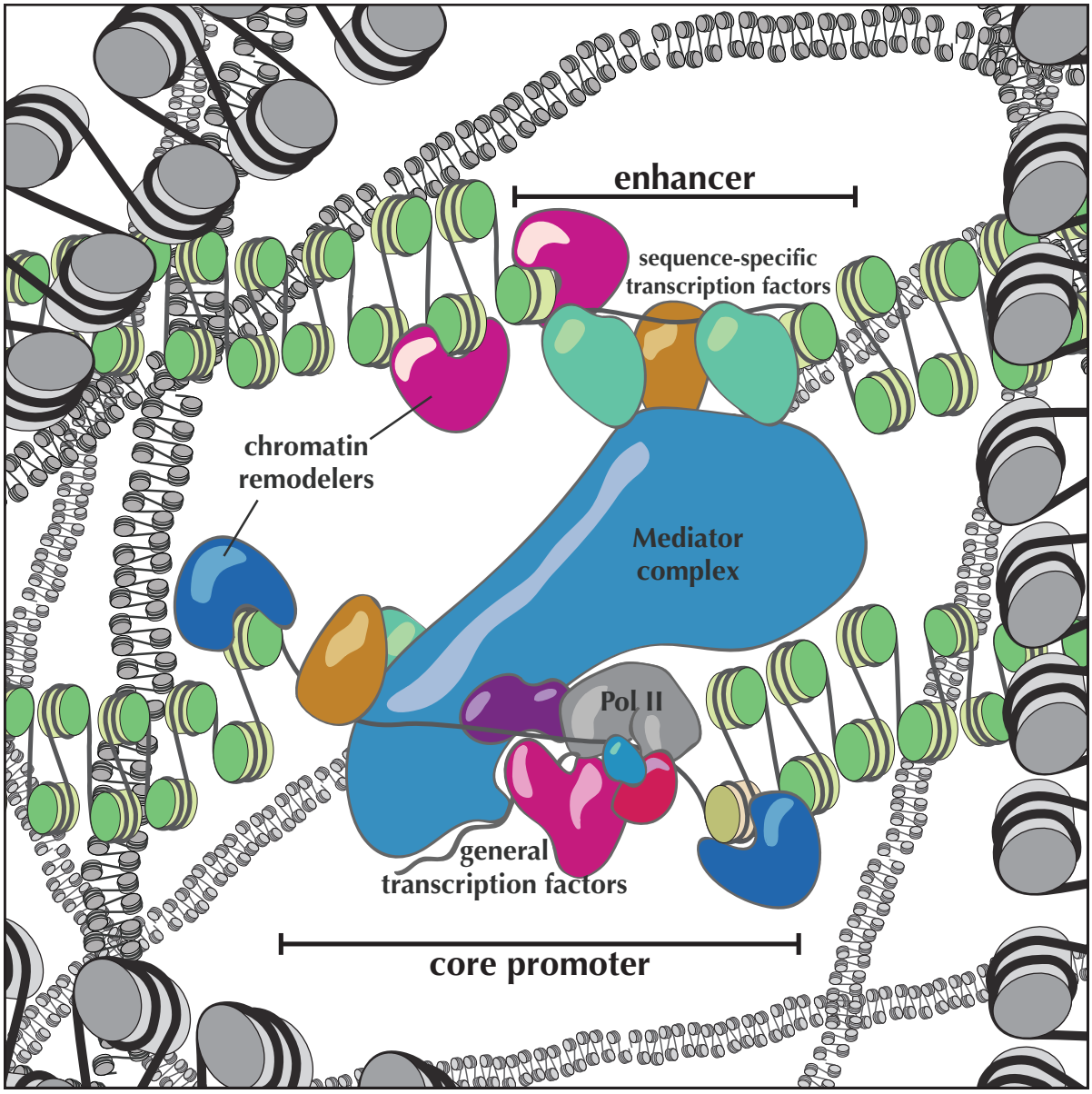


Figure 1. A schematic representation of key players in transcription regulation. Two genomic elements, one encoding the gene of interest and its core promoter and the other a regulatory enhancer (green chromatin strands) may be separated by tens-to-hundreds of kilobases in sequence space, but come into close spatial proximity. Sequence-specific transcription factors and chromatin modifying enzymes identify and open chromatin to allow the association of sequence non-specific factors. The Mediator Complex, TATA-binding protein (TBP) and the general transcription factors (GTFs) associate to facilitate RNA Polymerase II (Pol II) loading. Pol II associates with the transcription start site, opens a bubble in the DNA duplex and begins synthesizing the complementary RNA molecule. Not pictured are the positive and negative regulators of Pol II initiation and elongation.

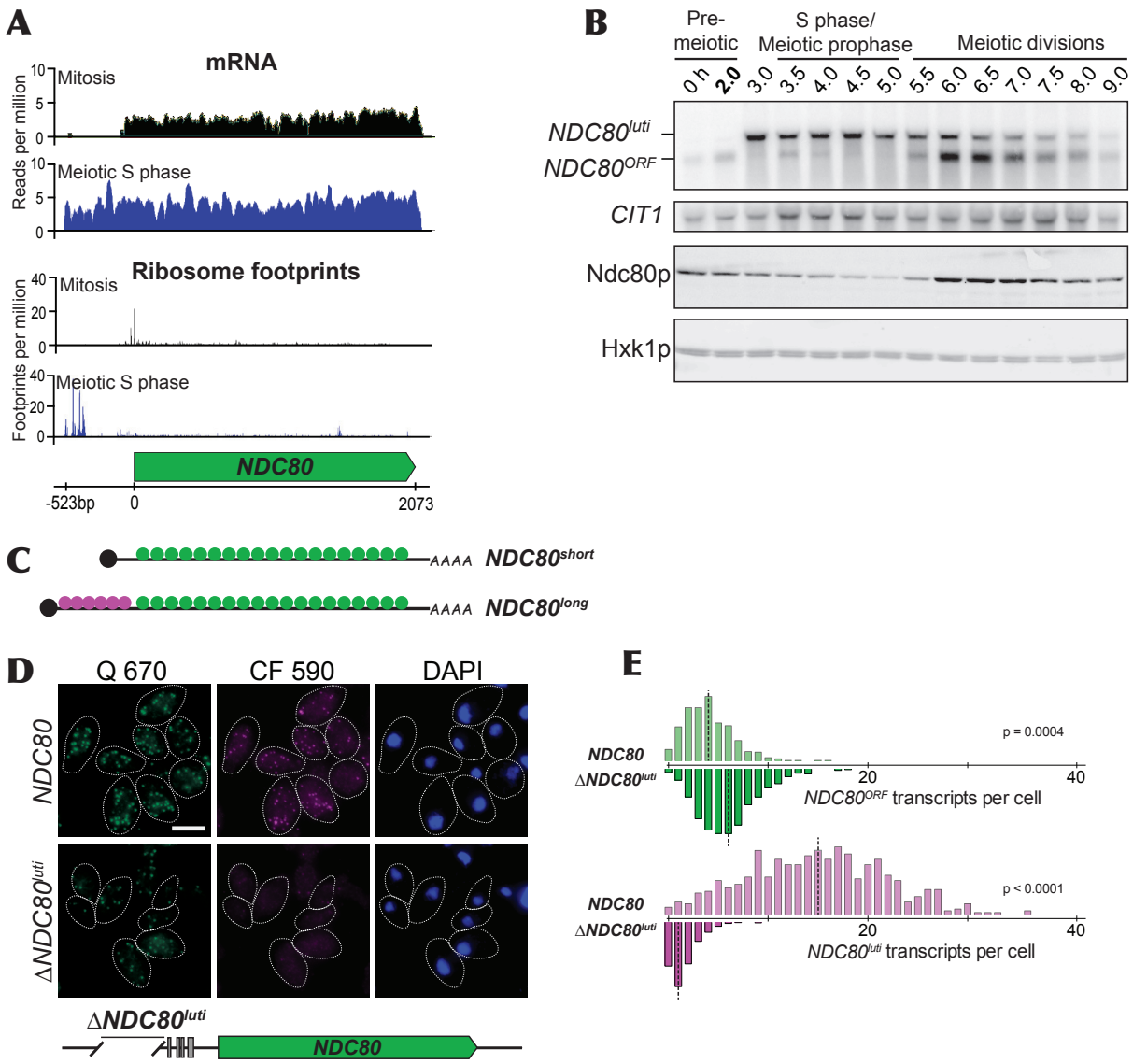


Figure 2. Cis-regulation of transcription and protein production through noncanonical RNA isoforms. **A)** The Saccharomyces cerevisiae gene NDC80 with matched RNA-seq and ribosome profiling tracks for cycling cells, and for cells beginning meiotic S phase. During S phase there is a clear shift in the transcription start site accompanied by a shift of translation to upstream open reading frames (uORFs) which do not produce a functional protein. **B)** Northern blot analysis of NDC80 isoform production in a dense timecourse through mitosis. The shift to the longer $NDC80^{luti}$ isoform is anticorrelated with NDC80 protein production, whereas $NDC80^{ORF}$ transcription increases NDB80 protein production. **C)** smFISH strategy to identify specific isoforms of NDC80. A common set of CalFluor590-labeled probes hybridize to both isoforms, while a second set of Quasar670-labeled probes hybridize specifically to $NDC80^{luti}$. A spot in the CalFluor590 channel lacking a corresponding Quasar670 spot is assigned as $NDC80^{ORF}$, while a paired spot is $NDC80^{luti}$. **D)** An example of smFISH data. Wildtype cells in mitotic S phase show a few ORF transcripts, but primarily express $NDC80^{luti}$. In mutants with a uORF deletion, only

NDC80^{ORF} transcription is seen. **E)** A quantification of the data in (D), showing that the uORF-containing region is important for repression of NDC80^{ORF} transcription in cis. Figures and data reproduced from Chen et al., 2017 with permission from the authors.

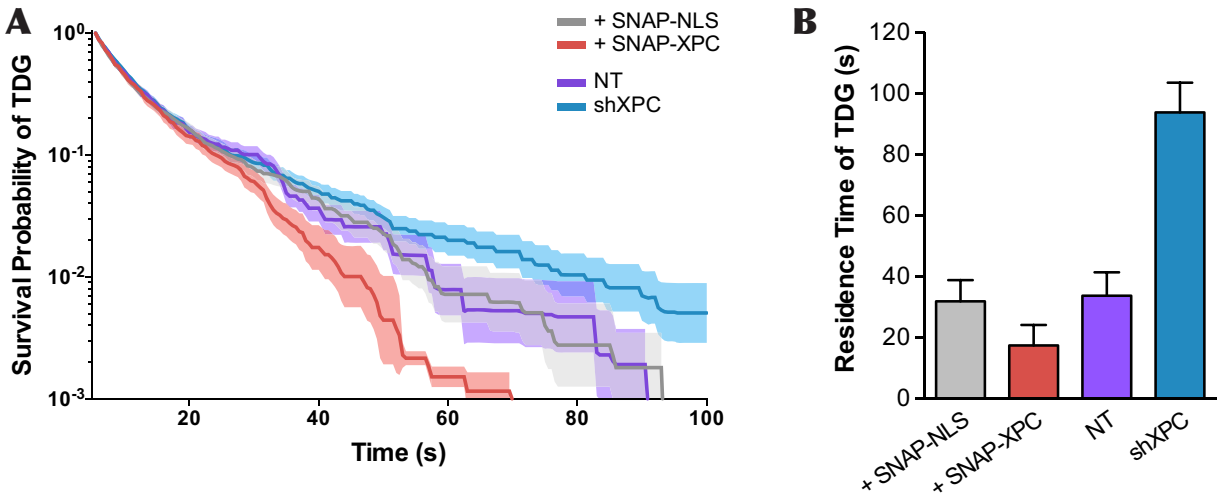


Figure 3. XPC expression level modulates TDG binding times. **A)** Curves showing the probability that a TDG molecule will remain immobilized on DNA as a function of time. Ectopic overexpression of XPC dramatically reduces binding times (red curve) relative to expression of an inert plasmid (grey curve), whereas shRNA-induced knockdown of XPC causes an increase in the probability that TDG will remain bound for long times (blue curve). **B)** Characteristic residence time calculated for TDG based on the curves in (A), after correcting for photobleaching. Figures and data reproduced from Ho et al., 2017 with permission from the authors.

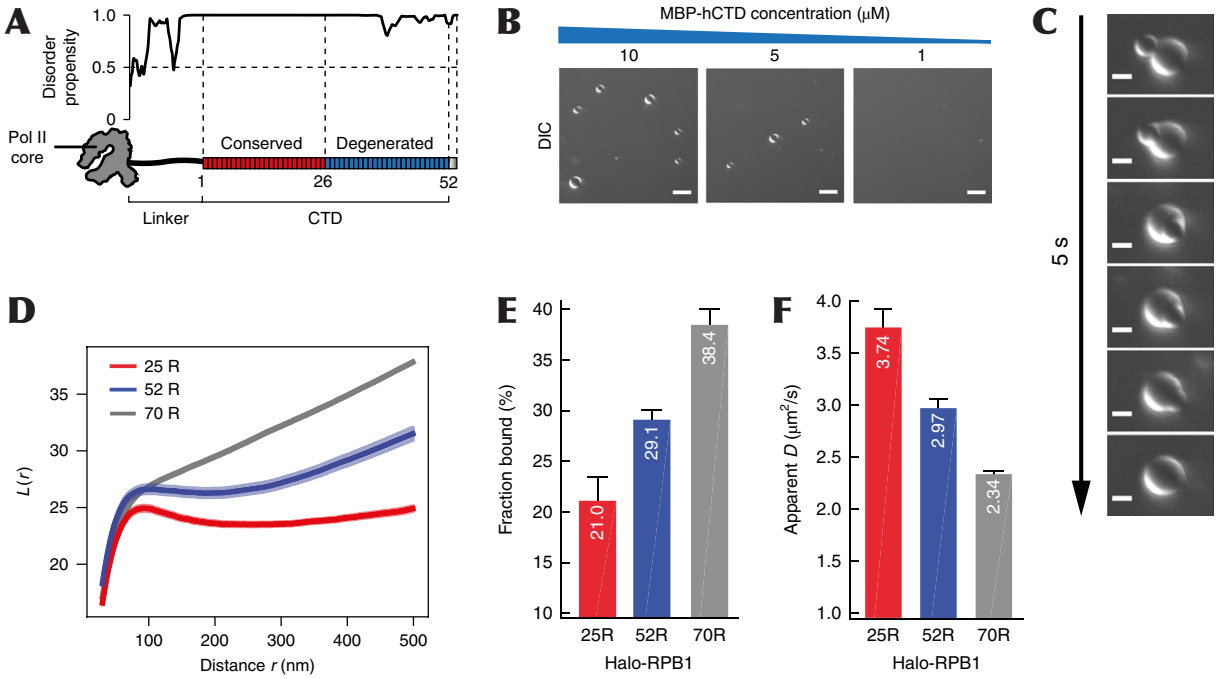


Figure 4 The unstructured C-terminal domain of Pol II imparts important behaviors to the molecule. **A)** A schematic of the well-folded RPB1 core as well as the repeats on the c-terminus of the amino acid heptad YSPTSPS. In Saccharomyces cerevisiae, the Pol II CTD has 26 repeats, whereas Homo sapiens Pol II has 52, with a higher degree of degeneracy in the second half. **B)** Purified Hs Pol II CTD forms phase-separated droplets *in vitro*, as a function of CTD concentration. **C)** Purified droplets show droplet fusion, a characteristic often associated with a liquid-like behavior. **D)** Spatial analysis of Pol II position within the nucleus using PALM. Here, Ripley's L function measures the degree to which Pol II molecules deviate from the distribution of a complete spatial randomness, with higher values indicating more clustering. Pol II with a truncated CTD shows less clustering in the nucleus at all length-scales compared to wildtype, whereas Pol II with an extended CTD shows a greater degree of clustering. **E and F)** Pol II CTD length affects Pol II binding to DNA (E) and diffusion through the nucleus (F). Cells expressing a Pol II with a truncated CTD have a lower average fraction of bound polymerases, and the freely diffusing molecules diffuse at a higher speed relative to WT. Similarly, Pol II with an extended CTD has a high average fraction bound and a slower diffusion coefficient. Figures and data reproduced from Boehning et al., 2018 with permission from the authors.

Chapter Two: A review of liquid-liquid phase separation and the evidence for its occurrence in cells

Abstract

The idea that liquid-liquid phase separation (LLPS) may be a general mechanism by which molecules in the complex cellular milieu may self-organize has generated much excitement and fervor in the cell biology community. While this concept is not new, its rise to preeminence has resulted in renewed interest in the mechanisms that shape and drive diverse cellular self-assembly processes, from gene expression to cell division to stress responses. *In vitro* biochemical data have been instrumental in deriving some of the fundamental principles and molecular grammar by which biological molecules may phase separate, and the molecular basis of these interactions. Definitive evidence is lacking as to whether the same principles apply in the physiological environment inside living cells. In this Perspective we analyze the evidence supporting phase separation *in vivo* across multiple cellular processes. We find that the evidence for *in vivo* LLPS is often phenomenological and inadequate to discriminate between phase separation and other possible mechanisms. Moreover, the causal relationship and functional consequences of LLPS *in vivo* is even more elusive. We underscore the importance of performing quantitative measurements on proteins in their endogenous state and physiological abundance, as well as make recommendations for experiments that may yield more conclusive results.

Part of this work has been previously published in: McSwiggen DT, Mir M, Darzacq X, Tjian R. 2019. Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences. *Genes Dev* 1–16. http://genesdev.cshlp.org/lookup/doi/10.1101/gad.331520.119. It is presented here with the permission of the authors

Introduction

Fundamentally, a cell is a collection of molecules compartmentalized in a manner to modulate biochemical reactions that support diverse cellular activities. The challenges faced by a cell in managing these biochemical processes scales with organismal complexity. In eukaryotes, where some cellular tasks can require the coordinated activity of tens to hundreds of individual molecular components, elaborate mechanisms have evolved to ensure that these assemblies occur; and furthermore, that they do so on timescales relevant to their biological function. Prototypical examples of cellular organization are the membrane-bound organelles, but it has long been appreciated that many compartments exist in the cell without an enclosing membrane (Montgomery, 1898; Wilson, 1899).

In the past decade, a fresh perspective on membraneless compartments has led to a resurgence in the idea that a majority of these compartments may exist as separate liquid phases (Courchaine et al., 2016). There has been a renaissance in understanding how liquid-liquid phase separation (LLPS) might function in compartment formation and maintenance (Banani et al., 2017; Hyman et al., 2014). Perhaps the most compelling example is the nucleolus, where a convergence of studies examining its liquid-like behavior (Brangwynne et al., 2011), as well as biochemical (Feric et al., 2016; Mitrea et al., 2016), and *in vivo* experiments (Berry et al., 2015; Weber and Brangwynne, 2015) collectively support a model where the nucleolus behaves as a separate liquid phase within the nucleus. Inspired by these and other early examples of compartments with liquid-like properties (Brangwynne et al., 2009), there has been a surge of publications revisiting the formation of well-known cellular compartments through the lens of LLPS. Far from being the peculiarity it once was, phase separation now has become, for many, the default explanation to rationalize the remarkable way in which a cell achieves various types of compartmentalization.

But it is unclear how strong the evidence for *in vivo* LLPS is, particularly when applied broadly across many cellular contexts. Moreover, the current focus on LLPS as a mechanism may come at the expense of understanding alternative mechanisms by which a high local concentration of factors can be achieved in the absence of a membrane. A pertinent example of this false equivalence will be demonstrated in Chapter Three of this dissertation, where we found that Herpes Simplex Virus replication compartments derive their ability to concentrate cellular factors through transient, nonspecific binding to the viral DNA in a manner distinct from liquid-liquid phase separation (McSwiggen et al., 2019). Despite this mechanistic distinction, these replication compartments display many of the hallmarks that are often deemed sufficient to claim that such a compartment is formed via LLPS (McSwiggen et al., 2019).

Our data on replication compartments, as well as other recent studies from our group (Chong et al., 2018; Mir et al., 2017, 2018), demonstrate that there are multiple routes to establish regions with high local concentrations of specific factors inside the cell. These studies prompted us to critically reexamine the current evidence for LLPS *in vivo*. The appeal for invoking phase separation is understandable, as it presents a way to rationalize—and at least superficially explain—certain behaviors of cellular compartments. However, in light of various recent studies

and upon further analysis, we find that the evidence for LLPS occurring in the cell is often far from conclusive. This is not to imply that LLPS cannot function in biological contexts, but rather to highlight how the tests commonly employed in probing LLPS are insufficient to rule out other mechanistic interpretations.

In this Perspective, we summarize the evidence used to diagnose liquid-liquid phase separation *in vivo*. Recently, others have similarly urged caution in over-interpreting *in vivo* experiments to test LLPS (Alberti et al., 2019), but this Perspective is, to our knowledge, the first to systematically and holistically consider the evidence presented by this emerging field as a whole. We first provide a summary of the state of evidence for LLPS across multiple contexts, and address important considerations for this evidence. Second, we address the evidence for the functional consequences of LLPS in the underlying biological processes being studied. Finally, we urge the application of more stringent criteria and more appropriate experimental approaches to understand the functional role of LLPS in cellular organization.

A diagnostic problem

Phase separation arises as a result of supersaturation. When a molecular species is at or above a critical concentration based on the specific cellular conditions (temperature, pH, etc.), it becomes more thermodynamically favorable to partition into a high-concentration phase and a low-concentration phase (Banani et al., 2017). Production of more of the protein in a two-phase regime does not increase the protein concentration in either of the phases, but rather results in changes in the relative volumes occupied by the two phases (Figure 1). A simplistic example of this is the nucleation and growth of water droplets on a cold glass, but accumulating evidence suggests the potential for LLPS to occur widely with biological macromolecules as well and it has been shown that certain classes of proteins, as well as RNA and other biological polymers, readily undergo LLPS *in vitro* (Jain and Vale, 2017; Wang et al., 2018).

The topic of phase separation in biology has been extensively reviewed elsewhere, and the reader is encouraged to refer to these reviews for a more thorough explanation of the forces that drive liquid-liquid demixing (Banani et al., 2017; Brangwynne et al., 2015). Much of what we know now has foundations in early works on polymer physics (J. Overbeek and Voorn, 1957), and has been advanced by efforts to improve crystallographic methods for which phase separation was used as a means of increasing a protein's concentration without it crashing out of solution (Asherie, 2004; Lomakin et al., 1996; Vekilov, 2010). As a result, physical models exist to explain liquid demixing (Lomakin et al., 1996; Velasco et al., 2002), and for purified components like proteins or nucleic acids, there exist rigorous standards by which one may determine whether a given system is undergoing liquid-liquid demixing. Modulating the concentration of a polymer, the ionic strength of the buffer, the temperature of the system, and intra- or interpolymer interactions can all quantifiably change the propensity of the polymer to demix (Brangwynne, 2013; Lomakin et al., 1996; Vekilov, 2010; Velasco et al., 2002). Following this model, beautiful *in vitro* experiments have been performed demonstrating the ability of LLPS systems to exhibit exclusivity (Banani et al., 2016; Feric et al., 2016; Nott et al., 2015); to form and dissolve on the basis of post-

translational modifications (Li et al., 2012; Lu et al., 2018); and to exhibit changes in viscosity and to "ripen" or harden over time (Patel et al., 2015).

These studies suggest that, at least *in vitro*, LLPS is particularly pervasive for proteins containing large disordered and low-complexity domains which enable multivalent homo- and heterotypic protein-protein interactions. While elegant biochemical experiments have provided essential insights into the physical properties of macromolecules that undergo LLPS, it remains less clear to what extent LLPS is happening in the crowded milieu of the cell. The intracellular environment itself is immensely more complex by virtue of the coexistence of hundreds of thousands of other macromolecular and small-molecules species that share the same solvent in a highly confined volume. An open question is to what extent the physical models built on *in vitro* studies hold true when dealing with the innumerable possible homo- and heterotypic interactions inside the cell, each of which having the potential to facilitate or antagonize LLPS or molecular function. Furthermore, *in vivo*, there is often much less control over the various parameters that should ideally be altered to test for LLPS. For example, while it is possible to tune, to a limited degree, parameters like the concentration of a few target proteins, or the ionic strength of the solution, additional non-trivial controls are required to ensure that the resulting changes are not due to secondary effects as the cell responds to a changing environment.

An accumulation of qualitative evidence

The challenges of modulating parameters critical to validate phase transitions *in vivo* has led researchers to instead rely heavily on descriptive characteristics. Roundness as a proxy for surface tension, the ability to undergo fusion or fission, changes in refractive index, and dynamic rearrangement within the phase as measured by FRAP are all routinely used to diagnose LLPS *in vivo*, largely based on the observation that *in vitro* droplets display these same liquid-like behaviors. We examined 33 studies, collectively making claims for 50 examples of *in vivo* LLPS for a range of cellular systems and organisms (Table 1). Without drawing any specific conclusions regarding the quality of the data in a given study, we categorized evidence based on whether the study used qualitative descriptors (+) or quantitative measurements (+++) to assess a given criterion. For example, a study reporting that "the droplets were round and could be seen to "fuse" received a "+" for the "roundness" and "fusion/ripening" criteria, whereas a study which quantifies the degree of roundness or conservation of material after fusion received a "+++". If a criterion is not mentioned, or if the assay does not apply to the system under study, it received a "-", and if it cites other literature which previously reported the claim, it received a "PR".

As LLPS is critically dependent on concentration, a crucial test to determine if it is occurring is to identify a critical concentration above which droplets exist and below which they do not (Asherie, 2004). Despite this critical dependence, the majority of the studies we examined performed all of their *in vivo* measurements using ectopic overexpression (Figure 2). The degree of ectopic expression varies case-by-case, but many multi-phase systems are exquisitely sensitive to changes in concentration (indeed, this fact is often used to support the biological function of LLPS) (Alberti et al., 2019). Furthermore, at least in some cases, it has been suggested that cellular

systems exist just on the cusp of a two-phase regime, in which case even the mildest over-expression could dramatically influence the outcome and interpretation of the data (Narayanan et al., 2019). Such over-expression introduces significant caveats into the conclusions that can be made from these studies.

Another surprising feature that stands out when examining the evidence in these studies is how heavily they rely on the descriptive properties of LLPS, rather than on quantitative tests. A majority of the cases reported roundness and observations of puncta fusion (48 and 35 out of 50 proteins, respectively), but the evidence provided is often a single image or movie, whereas few (6 and 8 examples out of 50, respectively) measure these behaviors in a quantitative manner (Figure 2). Furthermore, with the exception of fluorescence recovery after photobleaching (FRAP) experiments—which have their own limits to be discussed in the next section—many studies *only* use qualitative observations or other indirect lines of evidence for LLPS *in vivo*.

One of the major considerations with these indirect measurements, particularly with roundness and fusion, is the fact that diffraction-limited features have a tendency to look round and can obscure the true underlying structure. This is especially true if the size of the features is near to, or below, the resolution of the system (Figure 2B), as is the case for many putative LLPS systems (Boija et al., 2018; Cho et al., 2018; Maharana et al., 2018; Sabari et al., 2018). When combined with creative or inconsistent contrasting of microscopy images to accentuate puncta in an image (Boija et al., 2018; Guo et al., 2019; Sabari et al., 2018), it becomes difficult to evaluate whether one is observing a separate phase. Such artifacts become less of a concern as the compartments under study become larger, but even relatively large membraneless structures can display a striking degree of structural details when examined with higher resolution (Fei et al., 2017; West et al., 2016).

Other commonly used assays test the importance of particular protein domains for phase separation, either through truncation/modification experiments, or through perturbing weak hydrophobic interactions by treatment with 1,6-hexandiol. Here again, while the experiments can be useful to identify important protein domains for protein-protein and protein-nucleic acid interactions which can give rise to puncta inside the cell, they are insufficient to diagnose whether these puncta arise through phase separation or through other mechanisms. Additionally, while hexanediol does disrupt some weak interactions that may lead to LLPS, it is also known to increase membrane permeability and can even cause the formation of aberrant puncta on its own (Kroschwald et al., 2017).

We will see in Chapter Three of this dissertation a striking case of how using only indirect, phenotypic observations can lead to the wrong conclusion. Here, viral replication compartments formed during lytic Herpes Simplex Virus infection were shown to satisfy all of the descriptive characteristics of LLPS *in vivo* (McSwiggen et al., 2019). Replication compartments are round, they undergo fusion as they grow, they clearly have a different refractive index than the surrounding nucleoplasm, and they recruit many proteins that have themselves been shown to undergo LLPS *in vitro* (Chang et al., 2011; McSwiggen et al., 2019; Taylor et al., 2003). Given

only these qualitative descriptors it would be easy to conclude that this was yet another example of a nuclear compartment generated through the generally accepted mechanisms leading to LLPS. Yet, when we performed quantitative measurements to directly assess LLPS, we were surprised to find that the replication compartments form via entirely different mechanisms. In particular, super-resolution imaging demonstrates that the compartments emerge at variable concentrations of the component molecules, and that within each compartment the concentration of such molecules is not uniform, arguing against a fluid liquid-like state. Moreover, using single particle tracking to follow molecules as they explore the replication compartments reveals no change in diffusion coefficient compared to the surrounding nucleoplasm, nor any evidence of an energetic barrier to entering or leaving the compartments (McSwiggen et al., 2019).

This counterexample underscores the importance of using quantitative assays that can more appropriately diagnose LLPS, rather than relying solely on descriptive ones. Unfortunately, only 14 of the 52 instances we examined reported data which could be said to be a necessary feature of LLPS (Table 1, Figure 2C), and in only 6 instances was the evidence quantitative. The rest share the same descriptive criteria, but cannot be said to conclusively demonstrate LLPS in favor of other explanations, particularly in light of the example seen with replication compartments. Commonly, studies first demonstrate *in vitro* that a given protein is capable of undergoing LLPS. However, care should be taken when interpreting these results, as even hemoglobin and other well-folded, purified proteins can be induced to undergo LLPS *in vitro* given the right conditions and crowding agents (Asherie, 2004; Galkin et al., 2002; Heller et al., 1996).

One cellular system in particular where current enthusiasm for LLPS has vastly outpaced the evidence is in transcription regulation mediated by enhancers, where it has been emphatically postulated by many to be dependent on a process of phase separation (Boija et al., 2018; Cho et al., 2018; Hnisz et al., 2017; Lu et al., 2018; Nair et al., 2019; Sabari et al., 2018; Shrinivas et al., 2018). Single molecule experiments tracking the behavior of clusters of molecules, thought to be located at enhancers or other active DNA elements, highlight the problems of this particular interpretation (Boehning et al., 2018a; Cisse et al., 2013; Liu et al., 2014; Mir et al., 2017, 2018). The observation that the clusters themselves appear and disappear with extremely short half-lives, and do so heterogeneously throughout the nucleus is inconsistent with our current understanding of the formation of thermodynamically-driven LLPS condensates. Indeed, transcription factor hubs in the nucleus can appear with sizes and distributions largely independent of the factor's total nuclear concentration (Mir et al., 2017), in stark contrast with the LLPS model.

While there is clearly excitement and merit in the idea that LLPS could explain long-standing questions as to how transcription factors—especially their intrinsically disordered activation domains—mechanistically drive transcription, and how this process is coordinated (Hnisz et al., 2017; Kwon et al., 2013), the evidence for LLPS formation during transcription actually occurring in cells is some of the most phenomenological. Here in particular, the experiments that can definitively support or disprove LLPS are especially challenging. Their small size and highly dynamic nature makes them prone to misinterpretation based on morphology, and their constituent

molecules' propensity to interact not only with each other, but with host genomic DNA and RNA through multiple types of interactions makes meaningful perturbations difficult. In light of the data from herpesvirus showing that nonspecific binding to DNA can evoke many of the same descriptive behaviors, and given recent evidence that accessible DNA sites are spatially clustered in the nucleus (Xie et al., 2019), it is probable that alternative models other than LLPS can better explain the data which these studies have presented.

Unless these, and the other systems represented in Table 1, can satisfy the mechanistic characteristics for LLPS *in vivo*, along with robust evidence for functional consequences, one cannot exclude the strong possibility that the compartment in question is forming through cellular processes distinct from phase separation and that alternative models need to be pursued without bias.

FRAP is not a test of "liquid-like" properties

Aside from the ability to undergo LLPS *in vitro*, our review of the literature highlighted that one of the most commonly used "gold standard" assays to diagnose a compartment as "liquid-like" is Fluorescence Recovery After Photobleaching (FRAP) (Figure 2A). In these studies, the ability of the fluorescence recovery of a labeled protein within that compartment is assumed to imply rapid reorganization or exchange of the liquid within. As a technique, FRAP has been used extensively to measure the dynamics of protein exchange and interactions in the plasma membrane, nucleus and specific organelles. In FRAP, it is assumed that the photobleached molecules will diffuse away from the bleach spot and be replaced with new fluorescent molecules, resulting in a recovery of fluorescent signal (Sprague and McNally, 2005) (Figure 3A).

Despite the prevalence of using FRAP as the "gold standard" for LLPS, there are considerable potential caveats that should be considered when interpreting the data. First and foremost, the recovery of fluorescence is not unique to freely-diffusing molecules in solution. Even proteins that engage in stable, high-affinity binding interactions can show recovery (Hansen et al., 2017; Lawrence et al., 2018; Teves et al., 2016). Fluorescence recovery has a complex dependence on several parameters, including the diffusion coefficient and the concentration of the molecule being measured; the rate of its association with binding partners and their diffusion coefficients; the number and affinity of binding partners; and the technical considerations of the microscope and detectors being used (Mueller et al., 2010). Changes in any one of these parameters can influence the rate at which a bleached spot recovers. Modelling the recovery can provide quantitative insight into the underlying molecular dynamics (diffusion, binding, etc.), but it is well known that different model choices can lead to contradictory conclusions (Mazza et al., 2012; Mueller et al., 2010; Sprague and McNally, 2005; Sprague et al., 2004).

It should therefore be of great concern—if ultimately unsurprising—that that studies measuring FRAP recoveries in supposedly phase-separated compartments have reported sub-second (Patel et al., 2015) to minutes-long recoveries (Dine et al., 2018) for droplets generated from the very same IDR (Figure 3). Indeed, the recovery times in the studies we examined spanned nearly three orders

of magnitude, and in all cases FRAP recovery was central to the argument that the compartment in question was phase separated (Figure 3B).

Setting aside concerns that the range of recovery half-lives calculated spans a huge range, these one-off diffusion measurements prove little, as there are many potential biological mechanisms that may provide the same result. In-depth treatments of how binding, diffusion, and concentration affect recovery dynamics have been compiled elsewhere (Mueller et al., 2010; Sprague and McNally, 2005; Sprague et al., 2004), including theoretical considerations for cases when molecules are not approximately homogenous in solution such as in putative LLPS condensates (Sprague et al., 2006). All of which is to say there are many physical models that can be fitted to the same fluorescence recovery curve, which makes the calculated results deeply sensitive to the chosen model. One of the major revelations that live-cell imaging has provided to biology is an appreciation for how unexpectedly dynamic molecular processes are in cells. Binding events of protein complexes which were previously expected to last in the regime of minutes to hours, based on in vitro biochemical work, actually only last for tens of seconds, even for relatively stable protein complexes (Ho et al., 2017; Teves et al., 2018). Claims of a "liquid-like rate of fluorescence recovery" (Sabari et al., 2018) therefore grossly oversimplify the potential number of models which could explain such a recovery rate.

Some groups have taken additional measures in their FRAP experiments to directly address the "liquid-like" nature of the putative compartment by partially bleaching a compartment and looking for signs of internal rearrangement, which would be suggestive of a liquid state (Patel et al., 2015). These experiments are an improvement over reporting a single recovery time, but they should still be interpreted cautiously unless control experiments are provided. One critical control, for example, is to demonstrate that the rate of recovery is dominated by diffusion rather than by binding (Sprague and McNally, 2005). This can be shown by testing whether recovery is dependent on the size of the bleach spot (Figure 3C). Further, for these experiments to be conclusive, it should be shown that the entire fluorescence signal is within the linear range of the detector, and that the recovery is only explained by internal rearrangement rather than an influx of fluorescent molecules from outside. Recent work on FRAP specifically in LLPS systems addresses some of these concerns (Taylor et al., 2019), particularly for *in vitro*, though some fundamental assumptions made in the study – such as the ability to effectively ignore binding interactions in modeling the recovery, are dubious when applied to the cell. Importantly, this work suggests some important limitations to the way FRAP recoveries can be quantified when the size of the bleach spot is close to the size of the droplet in question (Taylor et al., 2019).

As it is currently used, FRAP offers no meaningful insight into whether a compartment is a separate liquid phase for the studies we examined. If recovery rates spanning nearly three orders of magnitude can all be interpreted as LLPS, then the assay becomes meaningless. On its own, FRAP cannot distinguish between a separate liquid droplet and a collection of molecules generated by any number of alternate mechanisms. Additionally, rigorous controls should be included in any study wishing to use FRAP to diagnose LLPS on the basis of protein exchange dynamics.

Searching for the functional significance

In the previous two sections, we have discussed how the evidence for phase separation *in vivo* in any given biological system is often far from conclusive. This is not to say that the compartments in question are indeed formed by a mechanism other than LLPS, but rather to highlight the significant uncertainty that still lingers. In the fullness of time, it may come to pass that some of these different putative examples of LLPS indeed turn out to be *bona fide* examples of phase separation. Even if this were the case, there still exists the more fundamental issue regarding functional significance of LLPS.

The observation that some cellular compartments behave like separate liquid phases has prompted speculation for a number of possible functional consequences. It has been speculated that LLPS compartments might function to facilitate cellular reactions/interactions, they may work to sequester some cellular components away from an unwanted reaction/interaction, or they may buffer the effective concentration of a given component within the cell (Banani et al., 2017; Bergeron-Sandoval et al., 2016). Briefly, the rationale behind facilitating reactions is relatively straightforward: If a select set of reactants exists at higher concentrations within a particular compartment, the reactions they perform will generally occur with much faster kinetics. The contrapositive is expected if a system is acting to sequester molecules away from a given reaction. The hypothesis that LLPS may be used to effectively buffer cells from fluctuations in cellular concentrations builds on the fact that LLPS occurs at a critical concentration, above which the solution phase separates (Oltsch et al., 2019). Thus, over-production only results in the growth of droplets without further increasing the concentrations in either the dilute or concentrated phases, essentially providing a constant concentration of the molecule in these two compartments irrespective of the average concentration of the molecule inside the cell (Figure 1).

Each of the above potential functions provide tantalizing explanations for how biological systems may be regulated but concrete *in vivo* evidence substantiating these functions in an endogenous context is lacking. Some more recent work has attempted to tie phase separation to a functional outcome(Du and Chen, 2018; Franzmann et al., 2018; Reinkemeier et al., 2019; Riback et al., 2017), however these same studies provide some of the weakest evidence that the putative phase separation process they are studying is actually occurring inside the cell, instead largely relying on biochemical experiments or previously cited work. Lacking any strong evidence for phase separation *in vivo* it is imprudent to imply functional effects based on the data currently available.

A recent study may help shed light on the magnitude of effects we might see from a phase-separated system. In *in vitro* biochemical experiments, Strulson and colleagues demonstrated that inducing LLPS resulted in a boost in the enzymatic rate of the hammerhead ribozyme, a proof of principle that phase separation can help facilitate enzymatic reactions (Strulson et al., 2012). If this principle generalizes to other types of reactions, this study is helpful in understanding what sorts of effects one might expect from compartmentalization *in vivo*. The authors find that the increase in enzymatic rate scales approximately proportionally with the degree of concentration (Strulson

et al., 2012). While certainly in the minority, a few studies have endeavored to measure the critical concentration of an LLPS system *in vivo* (Berry et al., 2015; Bracha et al., 2018; Shin et al., 2017). Bracha and colleagues used ferritin "corelets" decorated with IDRs as massively multivalent over-expression constructs to robustly drive LLPS (Bracha et al., 2018). They then made precise measurements of the critical concentrations at different expression levels and valency. Their data show that the increase in concentration of the high-concentration phase is maximally around 10-fold, whereas conditions closer to physiologically relevant examples showing significantly less concentration difference between the two phases (~3-fold).

These relatively low enrichments at physiological conditions suggest a modest upper-limit to the amount of reaction acceleration that can be achieved through phase separation of a single molecular species alone, though perhaps the concentration of multiple factors may yield additional acceleration. Recent evidence in cells supports such a modest limit: Two halves of a reaction targeted through *in vivo* over-expression into droplets yielded less than a 2-fold increase in the reaction selectivity and simultaneously a marked decrease in reaction efficiency (Reinkemeier et al., 2019). Only by further promoting association through the addition of kinesin motor domains to spatially concentrate their reaction could synergistic improvements of 5- to 10-fold be achieved (Reinkemeier et al., 2019). It should be noted that, even in this case, the study did not present any evidence that phase separation itself is responsible for the increase in reaction specificity.

Current data present a similarly modest picture when considering how effective LLPS might be at sequestering a given molecule away from unwanted reactants. Because LLPS is intimately tied to the critical concentration at which a droplet forms, we can use the concentration of the dilute phase to estimate the degree to which LLPS improves protein sequestration. Again, using the FUS corelet system as an extreme example, the difference in concentration between the corelets before and after induction of LLPS in the dilute phase is modest, perhaps two-fold at most (Bracha et al., 2018). For such a system to be an effective and meaningful mode of regulation, it would need to be sequestering molecules that are exquisitely sensitive to component concentration. It is of course possible that such a system exists, but these limitations should be explicitly considered when proposing phase separation as functionally relevant for sequestering reactants.

The above points suggest that the effects of LLPS on either facilitating or sequestering reactions will likely be quite subtle, and difficult to adequately test, particularly in a physiologically relevant concentration regime. The hypothesis that phase separation serves as a means to buffer biomolecules is equally challenging to verify. One may speculate on whether there is evidence that evolution has selected for optimal LLPS behavior under a given set of conditions, but there is still too little data to begin to address these types of questions. Until clear, testable predictions are made and investigated *in vivo* under physiologically relevant conditions, the functional consequences of phase separation will remain shrouded in uncertainty.

Finding a path forward

The notion that liquid-liquid phase separation has evolved as a means of further compartmentalizing the intracellular environment to specifically regulate biochemical reactions is

a compelling one. We do not wish to suggest that phase separation can never happen inside the cell, nor that phase separation is inconsequential to certain cellular functions. To be sure, there are clearly examples where LLPS remains the most suitable interpretation of the current evidence. Rather, with the research community so intoxicated by the current crop of studies and the tantalizing promise to explain the mechanistic underpinnings of subcellular organization, it is also important to recognize the potential for other explanations and the current lack of concrete evidence to point to one interpretation or another.

It may be the case that LLPS is a pervasive phenomenon in subcellular organization, mediated by multivalent interactions through intrinsically disordered protein domains, RNA, or DNA molecules. It may also be that the various cellular systems proposed to phase separate will still stand up to greater scrutiny and to assays that can more faithfully diagnose LLPS. These assays should directly probe how the compartment responds to changes in molecular concentration, binding interaction strength, temperature, and study the effect of putative LLPS on the compartment in question's constituents. However, in the absence of these more robust data, LLPS should not be invoked as the more likely interpretation of otherwise phenomenological observations, and alternative hypotheses should be formulated and tested to provide real biological insights.

In order to advance the field as a whole, it is clear that better assays and cellular systems are needed. Unfortunately, there is unlikely to be a one-size-fits-all suite of assays which can probe LLPS, and experiments will need to be thoughtfully tailored to the system at hand. An important first step, given that LLPS is intrinsically tied to cellular protein concentration, is a concerted effort to move away from experiments that overexpress, even if only to a small degree, proteins likely to participate in LLPS, and instead take the time and effort to tag the molecule under study in the native genomic locus to ensure endogenous levels of expression and protein concentrations. It is also clear that roundness and the ability to fuse are not sufficient evidence, and similarly FRAP experiments, if used, must be held to a higher standard than they are currently and the results interpreted with caution.

Better and more creative assays are in high demand. While the appropriate experiments will clearly depend on the exact system under study, there are at least a few promising avenues. Advances in light microscopy and spectroscopy allow quantitative measurements of absolute protein abundance, with and without fluorescent labels (Cai et al., 2018; Mir et al., 2012; Walther et al., 2018; Wang et al., 2011; Wei et al., 2017). For example, even if the exact critical concentration remains elusive to quantify for a given system, theory would predict that within a cell, the putative condensates should have equivalent concentrations of the phase separation molecule (as assayed by the fluorescence intensity per volume, for example). Single molecule tracking experiments would be a much desirable substitute for FRAP and has proven to be critical in uncovering an alternative compartmentalization mechanism in the case of herpesvirus (McSwiggen et al., 2019). These results suggest that the application of single particle tracking techniques in other systems may prove fruitful for examining the effects of putative phases on

molecular behavior, as would be predicted by theories around viscoelastic materials (Elbaum-Garfinkle et al., 2015; Niewidok et al., 2018; Wei et al., 2017).

Another strategy that may more directly diagnose *in vivo* LLPS would be acute depletion using endogenously appended degron tags (Nishimura et al., 2009), which should allow one to follow the degradation of proteins to determine whether compartments follow the types of behaviors which LLPS would predict (Figure 1). Other microscopy approaches such as localization of individual molecules within the compartment (Freeman Rosenzweig et al., 2017; Narayanan et al., 2019); or super-resolution imaging analysis of compartments which reveal fine structure (Fei et al., 2017; West et al., 2016) may help in testing LLPS as a model, as well as its functional consequences. Combined with single particle tracking experiments, these and other assays might reveal the specific concentration- and state-dependent manner that LLPS predicts, as well as effects on the molecules involved such as changes in diffusive behavior or energetic penalties for crossing between one phase to another.

Lastly, the ability of a protein to undergo phase separation when purified *in vitro* is an important finding to understand intrinsic properties of the protein, but these simplified systems cannot faithfully recapitulate the richness and complexity of interactions that occur within living cells. While these experiments are very useful for defining critical reaction partners, modifications, and energetic parameters, appropriate caution should be exercised when drawing equivalencies between these reconstituted conditions and the environment of the cell *in vivo*. A protein may phase separate in a test tube, and when produced at extreme quantities may also undergo LLPS inside the cell, but perhaps the more interesting and physiologically relevant interactions are found in less extreme conditions. It should be encouraged for future studies to include a more nuanced discussion on alternative models to phase which will likely provide valuable new insights.

Conclusion

Phase separation as an organizing principle in biology has compelled us to revisit old ideas in new light, and will likely continue to do so. As we have shown, the current state of the field is rich in descriptive evidence for phase separation in cells, but in most cases lacks crucial conclusive data. Roundness, fission and fusion, and speedy fluorescence recovery may bolster support for phase separation as a model—when proper controls are also provided—but the existence of counterexamples which share these properties in the absence of LLPS emphasizes the need for more rigorous and quantitative examination in cells with proteins expressed at the endogenous level. Further, experiments demonstrating the functional impact of phase separation, both at the phenotypic and mechanistic level, remain sorely lacking. Whether or not LLPS turns out to be a general phenomenon of broad functional utility, it should be appreciated that the formation of condensates likely represents only one of many potential avenues that the cell can use to organize its contents to facilitate critical bio-molecular interactions at the right scale and temporal cadence.

Figures

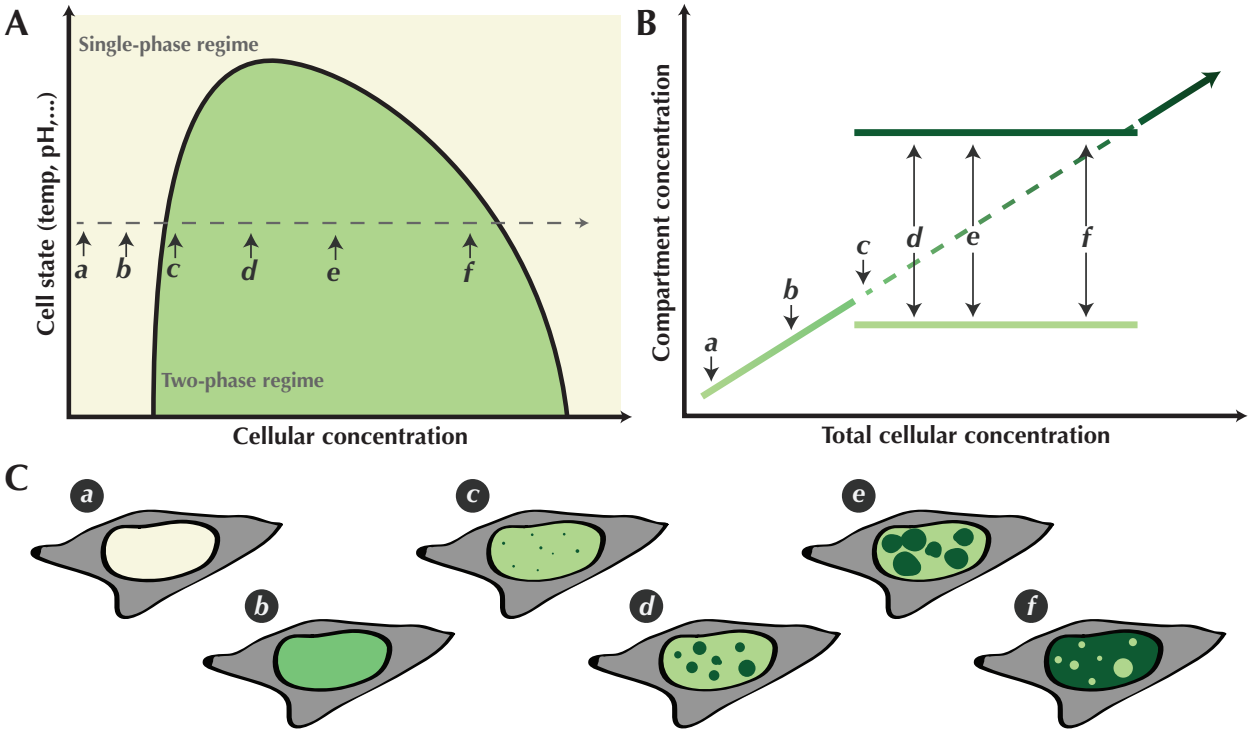


Figure 1 Liquid-liquid phase separation is a function of concentration. **A)** A schematic of a phase diagram depicting under what set of environmental conditions (temperature, salt concentration, pH, etc.) the system will remain as a single phase or spontaneously form two phases. An increase in the y-axis would represent any environmental change which would weaken monomer interactions, e.g. increasing temperature. The dashed line depicts how the system responds to increasing protein concentration, further illustrated in (B) and(C). **B)** For proteins that can phase separate, at a certain critical concentration (c), droplets form. Past this critical concentration, production of more protein increases droplet size but does not change the concentrations in either phase, until eventually the concentrated phase entirely fills the space whereupon the system returns to the one-phase regime (A). **C)** An illustration of the processes depicted in (A) and (B) as it occurs in the cell – in this hypothetical example, in the nucleus.

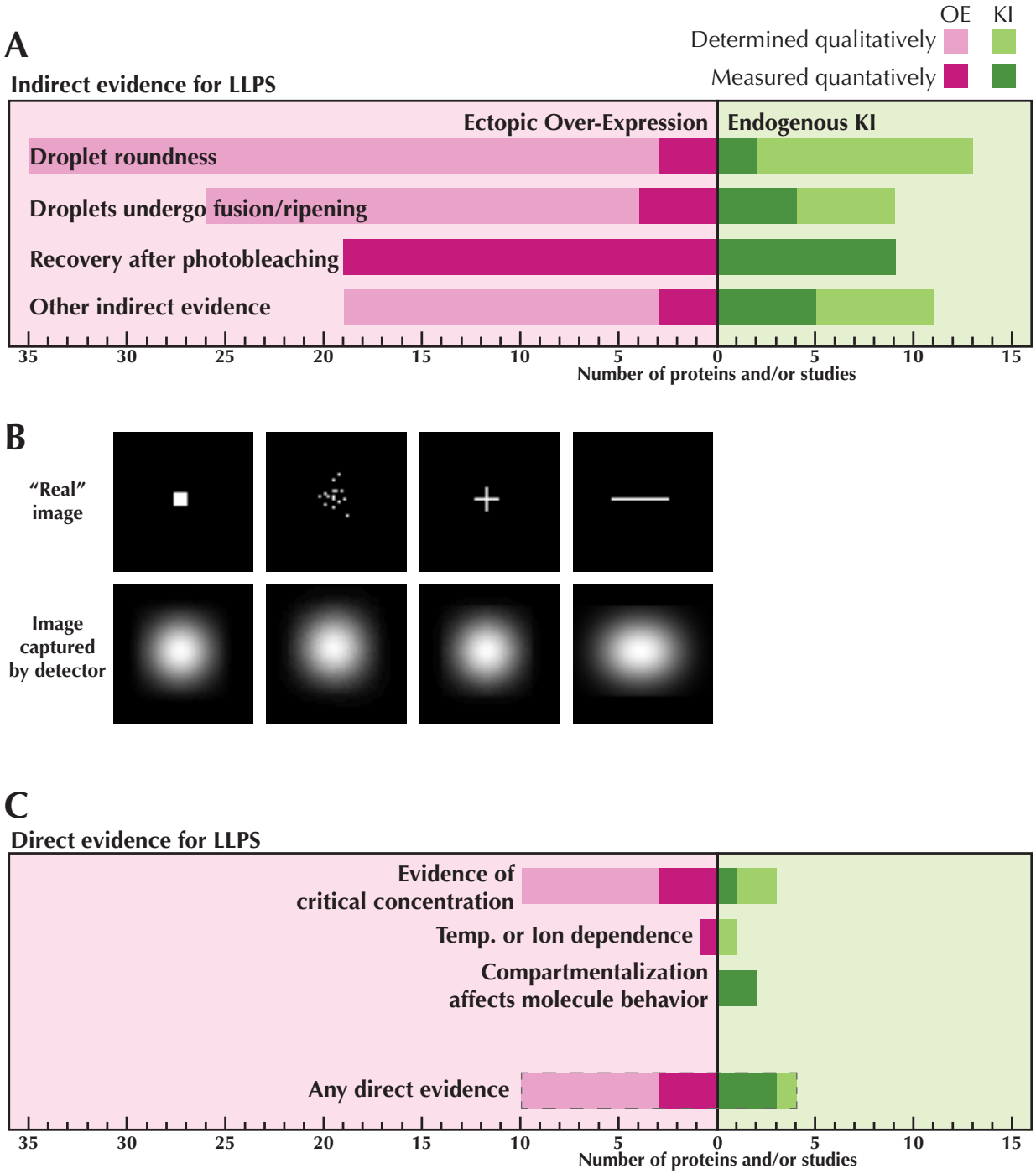


Figure 2 Evidence for LLPS in cells is largely phenomenological. **A)** Summary of the use of descriptive or phenomenological criteria in the studies from Table 1, separated into experiments that are performed on the endogenous protein (knock-in, KI) compared with those in over-expression systems (OE). The x-axis is the number of proteins from the 33 studies which were claimed to display that evidence. B) A simulated example of how diffraction-limited fluorescence imaging can obscure fine features. The top row depicts various simulated structures, and the bottom row is the image acquired by the microscope detector. C) A summary of the use of assays

which give direct evidence for LLPS in vivo. “Any direct evidence” is the sum of all the direct evidence categories. See Table 1.

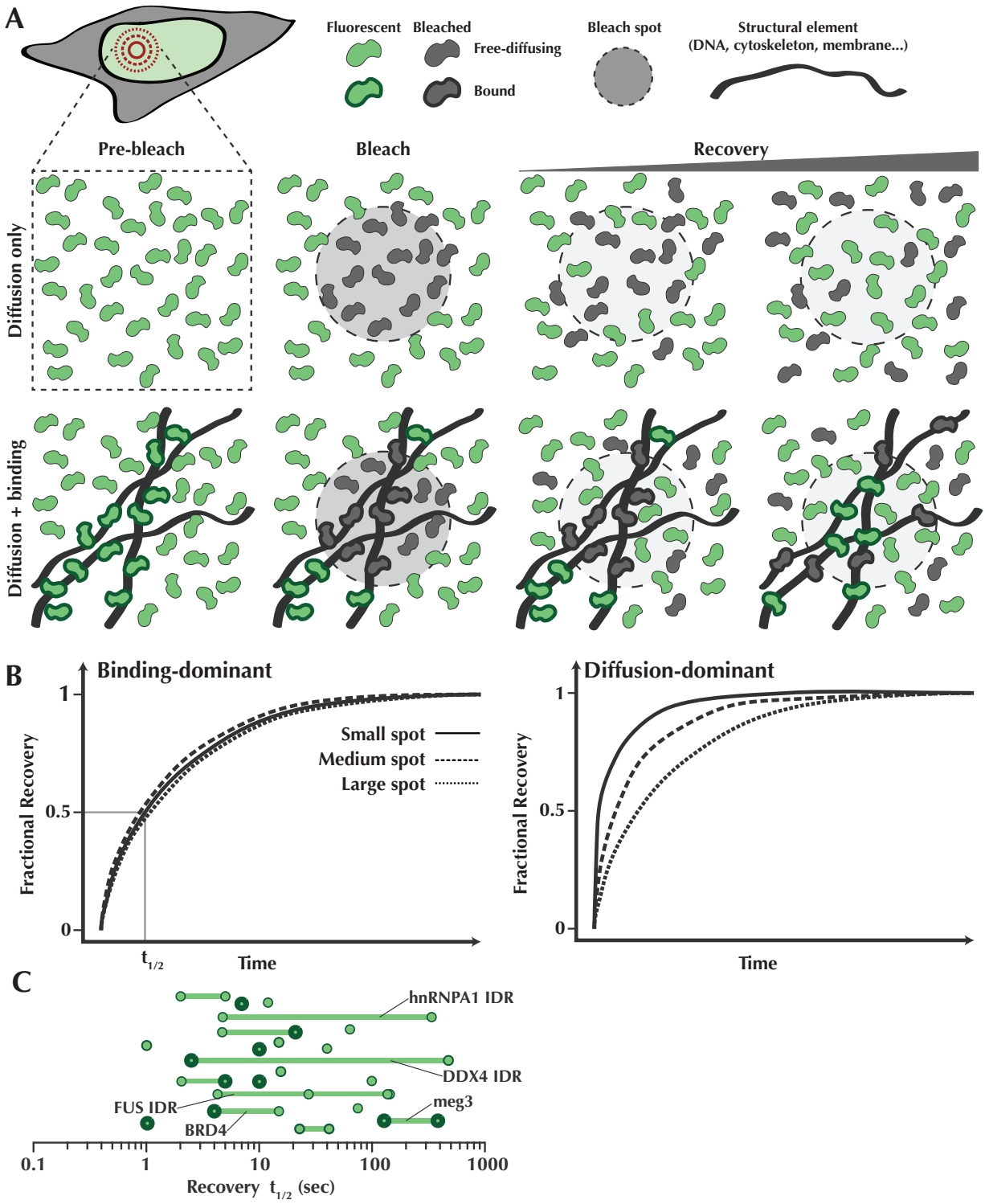


Figure 3 Fluorescence recovery is misleading as an assay for LLPS. **A)** A schematic of a Fluorescence Recovery After Photobleaching experiment. Fluorescent molecules in the cell are bleached with a strong laser in one spot and the signal is allowed to recover over time. In simple diffusion, as is expected in a liquid like a phase-separated domain, mixing of bleached and unbleached molecules is only governed by diffusion. In the case where some molecules may bind

to an immobile element, diffusing molecules will mix first before the bound molecules can unbind and exchange. **B)** Binding and diffusion have different impacts on the rate of recovery and extent of signal recovery. There are many methods to analyze FRAP data, the simplest being measuring the half-life of recovery ($t_{1/2}$). If the molecule under study has a high rate of diffusion compared to its binding rate, modulating the size of the bleach spot (dashed circles in (A)) will not affect the recovery (dashed lines). If diffusion is the limiting factor, as predicted by LLPS, then the size of the bleach spot should affect the $t_{1/2}$ of the curve. **C)** Reported $t_{1/2}$ times from the studies in Table 1. Cases where the same protein or protein domain have been measured more than once are indicated by connected lines. A few such examples have been labeled for reference. Bolded circles represent measurements on endogenous proteins concentrations while the other measurements are in over-expression conditions.

Study	Locaion	Compartment/ process	Protein (s)	Undergoes LLPS *in vitro*	Indirect *in vivo* evidence					Direct in vivo evidence		
					Endogenous or over-expression	Roundness	Fusion/ Ripening	Other expts	FRAP τ1/2 (seconds)	Critical Conc.	Temp/Ion strength	Affects molecular behavior
Brangwynne et al., 2009	Cytplasm	P Granules	PGL-1	-	OE	+	+	-	4.7	-	-	-
Brangwynne et al., 2011	Nucleus	Nucleolus	-	-	Endo	+++	+++	+	-	-	-	-
Li et al., 2012	Cytoplasm	Cytoplasm	synthetic SH3/PRM (NCK and N-WASP)	+++	OE	+	-	-	5‡	-	-	-
Nott et al., 2015	Nucleus	Nuages (granules)	DDX-4	+++	OE	+	+++	+	2.5	+++	+++	-
Molliex et al., 2015	Cytoplasm	Stress Granule	nhRNPA1	+++	OE	+	+	-	4.2	+	-	-
Altmeyer et al., 2015	Nucleus	Cell Stress	EWS	PR	OE	+	+	+	-	+	-	-
			TAF15	PR	OE	+	+	+	-	+		
			FUS	PR	OE	+	+	+	-	+		
Berry et al., 2015	Nucleolus	Nucleolus	FIB-1	-	Endo	+	+	-	-	+++	-	-
Patel et al., 2015	Nucleus	Stress Granule	FUS	PR	OE	+++	+++	-	4‡	-	-	-
Zhang et al., 2015	Cytoplasm	Whi3 droplets	Whi-3	+++	Endo	+++	+	+++	-	-	-	-
Pak et al., 2016	Nucleus		Nephrin (NICD)	+++	OE	+	+++	-	<1	+	-	-
Feric et al., 2016	Nucleus	Nucleolus	NPM1	+++	OE	+++	+++	+	64	-	-	-
			FIB1	+++	OE	+++	+++	+	75	-	-	
Smith et al., 2016	Cytoplasm	P Granules	PGL-1	-	Endo	+	-	-	-	-	-	-
			MEG-3	+++	Endo	+	-	+	-	-	-	-
Su et al., 2016	Plasma Membrane	Plasma Membrane	LAT	+++	OE	-	+	-	12	-	-	-
Schmidt and Rohatgi, 2016	Nucleus	Splicing	TDP43	PR	OE	+++	+	-	15	+	-	-
Freeman Rosenzweig et al., 2017	Pyrenoid	Carbon fixation	Rubisco/EPYC1	-	Both	+	+	+	22-42‡	-	-	+++
Riback et al., 2017	Cytoplasm	Cell Stress	Pab1	+++	Endo	+	-	-	-	-	-	-
Larson et al., 2017	Nucleus	Heterochromatin	HP1α	+++	OE	+	-	+	-	-	-	-
Strom et al., 2017	Nucleus	Heterochromatin	HP1a	+++	Endo	+++	+++	+	2-5 ‡	+	-	+++
Woodruff et al., 2017	Cytoplasm (centrosome)	Centrosome	SPD-5	+++	OE	+	-	+	>> 100	-	-	-
Shin et al., 2017	Cytoplasm	Synthetic	Opto-FUS	PR	OE	+	+	-	137	+++	-	-
			Opto-hnRNPA1	PR	OE	+	+	-	344	-	-	-
			Opto-DDX4	PR	OE	+	+	-	476	-	-	-
Du and Chen, 2018	Cytoplasm	Innate immune	cGAS	+++	OE	+	+	+	40‡	-	-	-
Maharana et al., 2018	Nucleus	Splicing	FUS	+++	OE	+	+	+	-	-	-	-
			nhRNPA1	+++	OE	+	-	+	-	-	-	-
			TDP43	+++	OE	+	+	+	-	-	-	-
			EWSR1	+++	OE	+	-	+	-	-	-	-
			TAF15	+++	OE	+	-	+	-	-	-	-
Sabari et al., 2018	Nucleus	Transcription	MED1	+++	Endo	+	+	+++	7‡	-	-	-
			BRD4	+++	Endo	+	+	+++	4‡	-	-	-
Cho et al., 2018	Nucleus	Transcription	RPB1	PR	Endo	+	+++	+++	10‡	-	-	-
			MED19	-	Endo	+	+++	+++	10‡	-	-	-
Boija et al., 2018	Nucleus	Transcription	OCT4	+++	Both	-	-	+	-	-	-	-
Lu et al., 2018	Nucleus	Transcription	Cyclin-T1	+++	OE	+	+	+	-	-	-	-
Dine et al., 2018	Cytoplasm	Synthetic	Synthetic FUS fusion	PR	OE	+	+	-	144	+	-	-
Bouchard et al., 2018	Nucleus	SPOP droplets	SPOP/DAXX	+++	OE	+	+	+++	-	-	-	-
Shin et al., 2018	Nucleus	Synthetic	Synthetic BRD4/Cas9	PR	OE	+	+	+	25‡	-	-	-
Franzmann et al., 2018	Cytoplasm	Cell stress	Sup35	+++	Endo	+	-	-	>1	-	-	-
Bracha et al., 2018	Nucleus; cytoplasm	Synthetic	Opto-FUS	PR	OE	+	+	+++	27.5	+++	-	-
			Opto-hnRNPA1	PR	OE	+	-	-	-	-	-	-
			Opto-TDP43	PR	OE	+	-	-	-	-	-	-
			Opto-DDX4	PR	OE	+	-	-	-	-	-	-
			Opto-PGL1	-	OE	+	-	-	-	-	-	-
Nair et al., 2019	Nucleus	Transcription	ERalpha	+++	OE	+	-	+++	15.6	-	-	-
Putnam et al., 2019	Cytoplasm	P Granules	MEG-3	+++	Endo	-	-	-	128 - 384	-	-	-
			PGL-1	-	Endo	-	-	+	21	+	+	-

Table 1 Cross-study summary of evidence used for LPS A summary of 33 papers which explicitly claim in the text to observe liquid-liquid phase separation in cells in vivo. The list is by no means exhaustive, but papers were chosen from those that are considered pivotal and/or highly cited, and in an attempt to represent a broad variety of cellular systems and compartments. The shaded rows are from a single study, which may have involved multiple proteins, each of which has its own line. "Compartment/process" is either the subcellular location of the droplet or the cellular process which generates the process, depending on which is more relevant. "Undergoes LLPS *in vitro*" is scored on whether the protein in question can form droplets when biochemically purified. "Endogenous or over-expression" is whether the in-cell or in-organism experiments were through tagging the endogenous locus (Knock-in, "KI"), or whether the labeled protein was added exogenously. Any degree of expression above the endogenous protein level was considered overexpression for the purposes of this analysis. "Roundness" is whether the study measured or commented on the shape of the droplets. "Fusion/ripening" is whether the study observed or commented on fusion events; in this case a "+++" was assigned if the study showed that droplet fusions preserved the total protein content through measurements of intensity, size, etc. "Other expts" is meant to broadly capture other experiments which are used to support the evidence that the system is undergoing LLPS. Commonly this involves the use of 1,6-Hexanediol, tests of dependence on particular protein domains, tests of co-segregation with other cellular components, or others. "FRAP ($t_{1/2}$)" is the halftime of recovery from a FRAP experiment. In the case where this value was not reported but the FRAP data were shown, we estimated the $t_{1/2}$ from the plots and have marked these values with an asterisk(*). "Critical conc." is awarded if the study demonstrated that there is a unique critical concentration above which droplets form. Experiments which suggest the existence of such a concentration but could not measure it were given a "+". "Temp/ion strength" is awarded if the study shows droplet formation is dependent on the temperature of the cells, or the ion concentration of the cells. Finally, "Affects molecular behavior" is awarded if the study can show that the droplets have an effect on the behavior of molecules inside, outside, or entering/exiting the droplet through an assay other than FRAP.

Chapter Three: Evidence for DNA-mediated nuclear compartmentalization distinct from phase separation

Abstract

RNA Polymerase II (Pol II) and transcription factors form concentrated hubs in cells via multivalent protein-protein interactions, often mediated by proteins with intrinsically disordered regions. During Herpes Simplex Virus infection, viral replication compartments (RCs) efficiently enrich host Pol II into membraneless domains, reminiscent of liquid-liquid phase-separation. Despite sharing several properties with phase-separated condensates, we show that RCs operate via a distinct mechanism wherein unrestricted nonspecific protein-DNA interactions efficiently outcompete host chromatin, profoundly influencing the way DNA binding proteins explore RCs. We find that the viral genome remains largely nucleosome-free, and this increase in accessibility allows Pol II and other DNA-binding proteins to repeatedly visit nearby DNA binding sites. This anisotropic behavior creates local accumulations of protein factors despite their unrestricted diffusion across RC boundaries. Our results reveal underappreciated consequences of nonspecific DNA binding in shaping gene activity, and suggest additional roles for chromatin in modulating nuclear function and organization.

Part of this work has been previously published in: McSwiggen DT, Hansen AS, Teves SS, Marie-Nelly H, Hao Y, Heckert AB, Umemoto KK, Dugast-Darzacq C, Tjian R, Darzacq X. 2019. Evidence for DNA-mediated nuclear compartmentalization distinct from phase separation. *Elife* **8**. https://elifesciences.org/articles/47098. It is presented here with the permission of the authors

Introduction

Controlling the local concentration of molecules within cells is fundamental to living organisms, with membrane-bound organelles serving as the prototypic mechanism. In recent years, our understanding of the forces driving the formation of sub-nuclear compartments has undergone a paradigm shift. A number of studies suggest that many proteins have the ability to spontaneously form separated liquid phases in vitro (Banani et al., 2017), and recent work highlights the possibility that similar liquid compartments may occur in vivo (Shin et al., 2017). Such liquid-liquid demixing (liquid-liquid phase separation, LLPS) has been proposed to be a common mechanism in sequestering specific macromolecules within a compartment, or in increasing their local concentration and thereby facilitating molecular interactions. Formation of these structures is thought to be predominantly driven by multivalent interactions mediated through intrinsically disordered regions (IDRs), or via modular binding motifs, RNA, or DNA (Banani et al., 2017).

These observations have generated a deeper appreciation for the diversity of mechanisms that a cell may deploy so as to locally concentrate select molecular constituents. The list of proteins—particularly nuclear proteins—that can undergo phase separation in vitro continues to grow (Courchaine et al., 2016). For example, recent studies of RNA Polymerase II (Pol II) and its regulators have shown that Pol II forms dynamic hubs whose sizes depend on the number of intrinsically disordered heptad peptide repeats contained within the C-terminal domain (CTD) (Boehning et al., 2018a), and that various CTD interacting factors may form phase-separated droplets in vitro (Lu et al., 2018) as well as local concentration hubs in vivo (Chong et al., 2018). We do not, however, fully understand the nature of the molecular forces that drive compartmentalization, and we lack compelling evidence of the functional consequences of these compartments.

Herpes Simplex Virus type 1 (HSV1) lytic infection provides an attractive model system because of its ability to form nuclear compartments *de novo*. HSV1 hijacks its host's transcription machinery during lytic infection (Rice et al., 1994), transcribing its genome in three waves: immediate early, early, and late, with the latter strictly occurring only after the onset of viral DNA replication (Knipe and Cliffe, 2008). Viral replication generates subcellular structures called replication compartments (RCs) where both viral and host factors congregate to direct replication of the viral genome, continue viral transcription, and assemble new virions (Knipe and Cliffe, 2008). Recent reports highlight the ability of HSV1 to hijack host Pol II such that, once late gene transcription commences, the host chromatin is largely devoid of productively transcribing Pol II, and the majority of newly synthesized mRNAs are viral in origin (Abrisch et al., 2015; Rutkowski et al., 2015). Concomitantly, RCs show a dramatic enrichment of Pol II and other nuclear factors (Rice et al., 1994).

Given this shift in both the sub-nuclear localization of Pol II upon infection, and its effect on the transcriptional output of an infected cell, we chose to examine the mechanism of Pol II recruitment to HSV1 RCs as a model case for the generation of new subcellular compartments. We employed a combination of imaging approaches, and complemented these with genetic,

genomic, and chemical perturbation experiments while measuring Pol II behavior in infected and uninfected cells. Despite initial indications that RCs exhibit many of the macroscopic hallmarks of LLPS, we find that recruitment of Pol II and other DNA-binding proteins to RCs is achieved through a distinct compartmentalization mechanism. Pol II recruitment occurs predominantly through transient, nonspecific binding of Pol II to viral DNA. These interactions are independent of transcription initiation, relying instead on the unusual feature that the HSV1 genome is largely free of nucleosomes, and therefore hyper-accessible to DNA-binding proteins relative to host chromatin. Our findings reveal that nonspecific binding can play a key role in the recruitment and retention of Pol II during infection, and more generally in the repertoire of distinct mechanisms a cell might employ to generate membraneless compartments.

Results

Pol II recruitment to RCs exhibit hallmarks of liquid-liquid demixing

HSV1 replication compartments form *de novo* following lytic infection, making them an attractive system to dissect compartment formation at the molecular level. To determine the mechanisms leading to the hijacking of Pol II, we used a U2OS cell line in which the catalytic subunit of Pol II has been fused to HaloTag (Boehning et al., 2018a). HSV1 infection occurs rapidly, with large RCs forming within a few hours (Figure 1A). Because we were most interested in the early stages of lytic infection when Pol II is actively recruited to the RC, we focused our experiments on the period between three hours post infection (hpi), when RCs begin to emerge, and six hpi when infected cells begin to display significant cytopathic effects (Figure 1 – supplemental videos 1 and 2).

In addition to Pol II, many other viral and nuclear factors re-localize to RCs (Dembowski and DeLuca, 2015). This redistribution of proteins is so dramatic that it can be seen as a change in the refractive index of RCs (Figure 1A). RCs grow and move over the course of infection (Figure 1B), and RCs exhibit other behaviors characteristic of liquid droplets, such as fusion (Figure 1B; Figure 1 – supplemental video 1 and 2) and a spherical shape with an aspect ratio close to one (Figure 1C), reminiscent of interfaces subject to surface tension (Brangwynne et al., 2011).

Another hallmark of LLPS compartments is that they are commonly associated with enrichment in proteins with IDRs. Across all HSV1 proteins, we identified predicted IDRs based on the protein sequence (Figure 1D). When categorized by temporal class, the immediate early (IE) and viral tegument proteins—the two groups that are first available to the cell upon infection—had the highest fraction of predicted intrinsic disorder. Compared to a list of proteins known to undergo LLPS in vitro, the IE and tegument proteins are even slightly more disordered (Figure 1E). Under the working hypothesis that interactions between IDRs drive phase separation, the similarity in predicted disorder profiles between our curated list and the IE and tegument proteins suggests that IDRs in viral proteins may be as likely to undergo LLPS as experimentally validated proteins.

Based on the above descriptive observations, we hypothesized that Pol II should be recruited to RCs through interactions between its CTD and other IDR-containing proteins within the RC. To test this, we measured the Fluorescence Recovery After Photobleaching (FRAP) dynamics of Pol II in RCs. We saw a consistent slowing of recovery as infection progressed (Figure 1F), which could be interpreted as evidence that Pol II is incorporated and sequestered within the RC, an "ageing" phenotype that others have described (Shin et al., 2017). Subsequent experiments to directly test this hypothesis, however, cast doubt on this interpretation.

Hub formation by Pol II in uninfected cells occurs in a manner dependent on the length of the Pol II CTD, a prominent IDR (Boehning et al., 2018a). To test whether the Pol II CTD likewise mediates interaction with RCs, we compared Pol II accumulation in RCs using the cells generated by Boehning and colleagues: wild-type Pol II CTD (with 52 heptad repeats), and with truncated (25 repeats) or extended (70 repeats) CTDs. Despite a strong effect in uninfected cells on the distribution of Pol II (Boehning et al., 2018a), the length of the CTD had no detectable effect on Pol II incorporation into RCs (Figure 1G), suggesting that Pol II recruitment in not sensitive to CTD length.

As a further test of the role of IDR interactions in Pol II accumulation within RCs, we treated cells with 1,6-hexanediol, which disrupts weak hydrophobic interactions between IDRs that drive LLPS (Lin et al., 2016). We infected cells for five hours, and then subjected them to treatment with a high concentration (10% v/v) of 1,6-hexanediol. Despite significant morphological changes in the nucleus after treatment, consistent with widespread disruption of cellular organization (Lin et al., 2016), Pol II remained highly enriched in RCs (Figure 1H). Furthermore, other IDRs with LLPS capabilities and which are known to interact with the CTD (Chong et al., 2018) are not enriched in RCs (Figure 1 – figure supplement 1), suggesting that formation of RCs does not require interactions between the IDRs of Pol II and other host or viral proteins.

Unrestricted Pol II diffusion across RC boundaries is inconsistent with an LLPS model

The data outlined in Figure 1 present a potential contradiction, as RCs exhibit several properties commonly associated with phase separation *in vitro*, yet Pol II recruitment to RCs is clearly not dominated by homo- or heterotypic interactions through its IDR. We sought to better understand the mechanism driving the enrichment of Pol II in RCs by measuring the behavior of individual Pol II molecules. To accurately capture both immobile and freely-diffusing Pol II molecules, we used stroboscopic photo-activatable single particle tracking (spaSPT) to visualize and track molecules (Figure 2A) (Hansen et al., 2017, 2018b). We labeled Halo-Pol II with equal amounts of JF_{549} and PA-JF_{646} (Grimm et al., 2015, 2016), allowing us to accurately generate masks to then sort trajectories as either "inside" or "outside" of RCs (Figure 2B, Figure 2 – supplemental movies 1 and 2). A qualitative comparison of trajectories of single Pol II molecules in RCs shows enrichment in short, constrained jumps compared to uninfected cells (Figure 2C, red arrows).

Quantitative measurements can be made by building histograms of all the displacement distances from the trajectories, and fitting to a two-state model in which Pol II can either be freely-

diffusing ("free"), or immobile and hence presumably bound to DNA ("bound") (Figure 2D, inset). Such a two-state model gives two important pieces of information: the fraction of "bound" and "free" molecules, and the apparent diffusion coefficient of each population (Hansen et al., 2018b). It is important to note that, because this modeling approach takes the aggregate of many thousands of traces, these data cannot measure how long a particular molecule remains bound in a given binding event. Therefore, "bound" refers to both specific DNA binding events—e.g. molecules assembled at a promoter or engaged in mRNA elongation—as well as transient, non-specific binding interactions.

The difference in the behavior of Pol II inside RCs compared with the rest of the nucleoplasm is immediately apparent from examining the lengths of jumps between consecutive frames (Figure 2C, D). Surprisingly, the mean apparent diffusion coefficient of the free population was unchanged between trajectories inside of RCs compared with those outside RCs or in uninfected cells (Figure 2E; Figure 2 – figure supplement 1A - C). If RCs were a *bona fide* separate phase, one would expect differences in molecular crowding or intermolecular interactions to predominantly affect free diffusion, resulting in substantially different diffusion coefficients.

To confirm this result, we performed a fluorescence loss in photobleaching (FLIP) experiment, in which a strong bleaching laser targets the inside of an RC and loss of fluorescence elsewhere in the nucleus is measured to quantify exchange of Pol II between the nucleoplasm and the RC. Consistent with the spaSPT data, we see that Pol II molecules exchange between RCs and the rest of the nucleoplasm as fast as Pol II in uninfected cells (Figure 2F). Similar results were obtained by using Pol II tagged with the photo-convertible fluorescent protein Dendra2 (Cisse et al., 2013) and photo-converting, rather than bleaching, molecules in the RC (Figure 2 – figure supplement 2A). Unlike the FRAP data, the rate of photobleaching does not change as a function of time after infection (Figure 2 – figure supplement 2B - C). Thus, Pol II molecules freely diffuse out of the RC, rather than remain sequestered within RCs.

An LLPS model predicts that a diffusing Pol II molecule within an RC should be more likely to remain within the RC than to exit when it reaches the compartment boundary. We tested this prediction by examining all trajectories for events in which a molecule crosses from inside RCs to outside, or vice versa, to look for evidence of such a boundary constraint. Comparing the distribution of displacements for a particle going from inside the RC to outside, we see no difference in the distribution of displacements, either entering or leaving RCs, when compared to uninfected cells in which mock RC annotations were randomly imposed *in silico* (Figure 2G; Figure 2 – figure supplement 3). Indeed, we cannot detect any evidence of a boundary for molecules entering or leaving RCs, further arguing that RCs do not consist of a distinct liquid phase.

While the two-state model shows no change in diffusion coefficient of Pol II, the fraction of molecules in the "bound" state doubles inside RCs, reaching ~70% (Figure 2H). We verified that this was not an artifact of the masking process by randomly shuffling RC annotations around *in silico* (Figure 2 – figure supplement 3C, D), and that diffusion coefficients of the bound population

are consistent with those of chromatin (Hansen et al., 2018b), and thus reflect DNA binding (Figure 2 – figure supplement 1D). The increase in the fraction of bound molecules is further supported by slowed recovery in the FRAP data (Figure 1F). The striking shift in the fraction of DNA-bound molecules, even while the FLIP decay rates remain unchanged, argues that this is due to an increase in the rate of Pol II binding rather than a decrease in the rate of Pol II unbinding. Thus, the mechanism driving Pol II recruitment to RCs is dominated by DNA binding rather than unbinding, which argues against the "aging" phenomenon that others have observed (Shin et al., 2017).

Pol II recruitment to RCs occurs independent of transcription initiation

One possible explanation for the increased fraction of bound Pol II in RCs would be a high level of active transcription in these compartments. Multiple lines of evidence suggest that transcription derived from the viral genome is activated to a much greater extent than transcription of even the most highly transcribed host mRNAs (Rutkowski et al., 2015), and this may be sufficient to explain the increase in DNA-bound Pol II.

To test whether active transcription is driving Pol II recruitment to RCs, we treated infected cells with either Triptolide or Flavopiridol, small molecules that selectively inhibit stable Pol II promoter binding or transcription initiation, respectively (Figure 3A) (Bensaude, 2011). HSV1 requires the expression of immediate-early and early genes to generate its DNA replication machinery, so we allowed the infection to progress for four hours before treating with either compound. Cells at this time point have well-formed RCs, and Pol II binding is already greatly increased (Figure 2H). We treated these cells with either drug for 15, 30, or 45 minutes to inhibit *de novo* transcription and allow any elongating polymerases to finish transcribing (Figure 3B). RNA fluorescence in situ hybridization (FISH) against an intronic region showed significantly reduced nascent transcripts after 30 minutes of drug treatment (Figure 3C, D). Remarkably, even after 45 minutes of treatment, ~80% of the Pol II signal remains within RCs (Figure 3E, F). These data suggest that the recruitment of Pol II to RCs occurs largely independently of transcription, and without stable engagement with gene promoters.

By spaSPT, in uninfected cells, Triptolide or Flavopiridol treatment both reduce the fraction of bound Pol II by half, to ~15% (Figure 3G), similar to what others have reported (Boehning et al., 2018a; Teves et al., 2018). Nevertheless, inhibition of transcription with Flavopiridol reduced the bound fraction inside of RCs by only ~5% (Figure 3G). Even treatment with Triptolide, which prevents stable engagement with TSS-proximal DNA, only reduced the fraction bound by ~12% (Figure 3G). Given this result, we conclude that the majority of binding events we measure are independent of viral transcription.

HSV1 infection appears also to confer some resistance to the effects of these drugs on Pol II binding to host chromatin, despite the fact that these inhibitors are sufficient to abrogate transcription (Figure 3C-F). Given the inherent limitation of spaSPT for inferring the length of binding events, we wanted to confirm that drug treatment prevented stable Pol II binding. Indeed, FRAP experiments in cells treated with Triptolide show a dramatically faster recovery rate for both uninfected and infected cells (Figure 3H). For the infected samples, this means that the "bound"

molecules measured by SPT do not remain bound for long times, as one would expect from high affinity protein-protein or protein-DNA interactions at cognate sites. Instead, the majority of the bound fraction is comprised of transient binding events independent of transcription. The fact that infected cells show increased DNA binding outside of RCs after drug treatment may be a result of other viral mechanisms that occur during infection, such as aberrant Pol II CTD phosphorylation (Rice et al., 1994) or termination defects (Rutkowski et al., 2015). Still, our results suggest that viral DNA and/or DNA-associated proteins mediate rapid and predominantly nonspecific interactions with Pol II in RCs.

It has been reported that the viral protein ICP8 interacts with the CTD of Pol II through a bridging interaction by the viral protein ICP27 (Zhou and Knipe, 2002). Others have used ICP27 truncation mutants to suggest that this ICP27-mediated mechanism is responsible for Pol II recruitment into RCs (Dai-Ju et al., 2006). Thus, we tested HSV1 mutant strains n504 and n406, which carry nonsense mutations in ICP27 that weaken or abrogate (respectively) the Pol II-ICP8 interaction, and should be defective for Pol II recruitment to RCs (Rice and Knipe, 1990; Zhou and Knipe, 2002). While these mutant strains generally show a deficiency in forming RCs and producing virus, we found that in cells where RCs do form, Pol II is recruited as efficiently as in cells infected with a WT virus (Figure 3 – figure supplement 1A), and the FRAP recovery dynamics are indistinguishable from WT virus-infected cells (Figure 3 – figure supplement 1B) suggesting it is unlikely that this specific viral complex is the major player in recruiting Pol II to RCs.

HSV1 DNA is much more accessible than host chromatin to Pol II

The finding that Pol II molecules remain bound—however transiently—to the viral DNA, even in the absence of transcription or other interactions involving viral proteins, suggests that the DNA itself could plays a dominant role in Pol II enrichment in RCs. Knowing the amount of viral DNA contained in any one RC may be crucial to understand the role viral DNA may play in RC formation and function, but to our knowledge, this has not been determined. We therefore sought to measure the amount of DNA in RCs using DNA FISH by targeting fluorescent probes to two specific regions of the viral genome (Figure 4A). Fluorescence intensities from infected samples were compared at different times post infection to samples that were infected in the presence of phosphonoacetic acid (PAA), an inhibitor of viral DNA replication that ensures there is only one copy of the viral genome per punctum (Figure 4B; Figure 4 – figure supplement 1A) (Eriksson and Schinazi, 1989).

The number of genomes within an RC correlates well with the time post infection (Figure 4C), and there is also a strong correlation between RC size and genome copy number (Figure 4 – figure supplement 1B). Based on these data, we calculate that the average RC at 6 hpi has a DNA concentration of 3.9 $x10^4$ bp/μm^3, approximately 240 times less concentrated than average host chromatin (Monier et al., 2000). The totality of viral DNA in an average cell after 6 hours of infection corresponds to just ~0.2% of total DNA in karyotypically normal human nuclei (Table 1). Yet, despite its 100-fold lower DNA concentration, inhibition of viral DNA replication with

PAA caused the fraction of bound Pol II molecules inside the pre-replication foci to decrease to ~50% (Figure 4D).

Since most of the observed Pol II binding events that we observe inside of RCs appear to be unrelated to transcription, but are clearly dependent on viral DNA replication, we wondered what might be different about the viral genome relative to host chromosomes. A likely candidate is the chromatin state of the viral DNA. There is presently no consensus about the organization of viral DNA during lytic infection, but mass spectrometry studies have failed to detect histones associated with viral DNA (Dembowski and DeLuca, 2015). Moreover, infection of a cell line constitutively expressing Histone H2B fused to HaloTag is not incorporated into RCs (Figure 4E).

To measure histone occupancy on HSV1 DNA, and get a measure of its accessibility, we turned to ATAC-seq, which gives signal proportional to the accessibility of the DNA at a given locus (Buenrostro et al., 2013). Based on the amount of viral DNA present in an infected cell, we calculated the fraction of reads one would expect to map to the virus relative to the host. At six hpi, by DNA FISH the viral DNA represents an average 0.2% of total nuclear DNA content. Yet under the same conditions at this time point, 24.2% of reads mapped to the virus on average, showing that viral DNA is at least 100-fold more accessible (Table 1).

The ATAC-seq fragment length distributions (Figure 4F; Figure 4 – figure supplement 1C) showed a much faster decay for reads mapping to the virus at all times post infection, and with no evidence of nucleosomal laddering, in stark contrast to reads that map to the host genome. When we visualized the position of all HSV1-mapped reads along the viral genome, the profiles were strikingly flat and featureless (Figure 4G). An average of all annotated human mRNA genes, centered at the TSS, shows a characteristic peak of accessibility at the TSS for reads with a length corresponding to inter-nucleosomal distances (<100 bp), and a characteristic trough of mono-nucleosome sized fragments (180 – 250 bp) (Figure 4H). By contrast, TSS averages mapped to the viral genome for either short or mono-nucleosome fragments show no changes in accessibility. Even averaging over all viral transcripts, it is clear that the entire viral DNA remains equally accessible (Figure 4I). Taken together, these data indicate that the HSV genome is maintained in a largely nucleosome-free state, and thus highly accessible to DNA binding proteins like Pol II.

Transient DNA-protein interactions drive Pol II hub formation through repetitive exploration of the replication compartment

Knowing that the DNA inside RCs is vastly more accessible to nuclear factors than host chromatin, we next asked what emergent properties of this accessible DNA might help explain Pol II recruitment. Using an HSV1 strain that allows incorporation of nucleotide analogs, (Dembowski and DeLuca, 2015), we fluorescently labeled DNA, imaged it at super-resolution, and found that, within a given RC, viral DNA shows variability in local density of nearly three orders of magnitude (Figure 5A).

The greater accessibility and higher variability in local density of viral DNA lend themselves to a possible mechanism by which Pol II becomes enriched. Recent theoretical work has shown

that a polymer like DNA, which has many binding sites in close proximity, can induce an interacting protein to revisit the same or adjacent sites repetitively during its exploration of the nucleus (Amitai, 2018) (Figure 5B). In such a case, we should be able to see signatures in our spaSPT dataset of Pol II continually revisiting adjacent sites on the viral DNA. To check, we calculated the angle formed by consecutive displacements and compiled these angles into a histogram (Figure 5C) (Izeddin et al., 2014). For particles experiencing ideal Brownian motion, the angular histogram will be isotropic. Anisotropy can arise through a variety of mechanisms, such as adding the aforementioned "traps," thereby giving the particle a greater probability of revisiting proximal sites before diffusing away (Amitai, 2018).

In uninfected cells, and in infected cells outside of RCs, Pol II displays diffusion that is largely isotropic. In stark contrast, inside RCs Pol II diffusion is highly anisotropic, particularly around 180° (Figure 5D, Figure S3). To compare across samples, we computed the likelihood of a backward translocation (180° ± 30°) relative to a forward translocation (0° ± 30°). Analyzed this way, Pol II inside RCs has a 1.7-fold greater chance of making a backward step for every forward step it takes (Figure 5E). In cells treated with Triptolide, where stable binding is inhibited, the effect created by transient binding events is further amplified (Figure 5E, F), which helps explain the dramatic retention of Pol II inside RCs, even 45 minutes after inhibition of transcription (Figure 3E). These data are most consistent with a model in which Pol II repetitively visits the highly accessible viral genome via multiple weak, transient binding events which likely result in Pol II hopping or sliding along the DNA. The sharp anisotropy of the molecular exploration within the compartment means that a given Pol II molecule within an RC is more likely to visit the same or proximal sites multiple times before either finding a stable binding site or diffusing away.

The heterogeneous distribution of viral DNA within RCs, and the anisotropic way Pol II explores RCs, is also borne out in the distribution of Pol II molecules. Similar to the viral DNA, super-resolution photo-activated localization microscopy (PALM) renderings of infected nuclei revealed a heterogeneous Pol II distribution within RCs (Figure 5G). A key prediction of the formation of phase condensates is that LLPS compartments should form at a characteristic critical concentration, and that molecules within the high concentration phase should return to homogeneity within the phase(Bracha et al., 2018). The highly heterogeneous nature of Pol II within the RCs provides yet further evidence that these compartments are not derived through an LLPS process. We used Ripley's L-function to measure how the Pol II distribution deviates from spatial randomness, with values greater than zero indicating a concentration higher than predicted for complete randomness at that given radius (Figure 5H) (Ripley, 1977). We find that the curve remains well above zero, and increases, for all radii up to one micron. This suggests that Pol II forms hubs within RCs at multiple length scales, consistent with the behavior of Pol II in uninfected cells (Boehning et al., 2018a), and inconsistent LLPS driving the constitution of RCs.

Nonspecific interactions with viral DNA license recruitment of other proteins

Seeing that Pol II is recruited to RCs via transient and nonspecific binding to the viral genome made us wonder whether this effect was specific to Pol II, or whether DNA accessibility can

generally drive the recruitment of any DNA-binding proteins to RCs. Certainly, many other DNA-binding proteins are recruited to RCs (Dembowski and DeLuca, 2015). To assess whether nonspecific DNA binding could be responsible for their accumulation as well, we looked to an extreme example: The tetracycline repressor (TetR), and the Lac repressor (LacI). Both proteins are sequence-specific bacterial transcriptions factor, the consensus sites for which are absent in both human and HSV1 genomes. If proteins like TetR and LacI can be recruited to RCs despite lacking cognate binding sites, this is strong evidence that nonspecific DNA association is the driving mechanism for recruitment.

Expression of TetR-Halo and LacI-Halo shows enrichment within RCs (Figure 6), in stark contrast to Halo-NLS or HaloTag-fused IDRs (Figure 1 – figure supplement 1). Furthermore, a comparison of the jump lengths measured in single particle tracking of TetR-Halo also reveals an enrichment in short translocations inside of RCs, consistent with higher fraction of bound TetR-Halo molecules (Figure 6 – figure supplement 1). Thus, while IDR-based interactions alone are unable to generate strong enrichment in the RCs (Figure 1 – figure supplement 1), even modest nonspecific DNA-binding affinity appears sufficient to do so.

These data suggest a model in which viral Pol II recruitment consists of transient, nonspecific binding/scanning events of the highly exposed viral genome (Figure 7A). A DNA-binding protein exploring the nucleus (uninfected, or infected but outside of RCs) may encounter some occasions for nonspecific interaction with duplex DNA, but because of the nucleosome-bound nature of the host chromatin, these binding/scanning events are necessarily spatially dispersed and infrequent (Figure 7B). Within RCs, many copies of the unprotected HSV1 DNA are present, allowing nonspecific events to happen much more frequently, with fewer and shorter 3D excursions between DNA contacts (Figure 7C). Thus, transient protein-DNA interactions drive enrichment of DNA-binding proteins within RCs.

Discussion

Multiple routes to create high local concentrations

Here we have demonstrated that Herpes Simplex Virus type 1 accumulates Pol II in replication compartments because the virus' unusually accessible DNA genome provides many potential nonspecific binding sites, acting as a molecular sink which causes a net accumulation of Pol II even in the absence of transcription. Such a mechanism for locally concentrating proteins is revealing, as it neither requires the formation of stable macromolecular structures nor produces any behaviors at the single-molecule level suggesting a separate liquid phase. Instead, by virtue of the fact that the viral genome appear to act as a single polymer globule (Figure 5A), from the macroscopic view Pol II recruitment to RCs appears to share many of the behaviors commonly attributed to liquid-liquid phase separation, and yet RCs are clearly a distinct class of membraneless compartment that operate on principles very different from an LLPS model.

We cannot completely rule out the possibility that some form of LLPS-like mechanism contributes to our observations in Figure 1. However, our data demonstrate that even if this is the

case, it does not contribute to the enrichment of Pol II or the other proteins that we have tested. It is also difficult to rationalize how RCs could exist as a phase condensate without having any measurable impact on the free diffusion (Figure 2E), distribution (Figure 5G, H) or exchange of molecules that diffuse within and between compartments (Figure 2F, G; Figure 6 – figure supplement 1). Our results prompt the need for a better characterization of *bona fide* phase separation, with a focus on its functional consequences *in vivo*, and suggest that caution should be exercised before assigning LLPS as the primary assembly mechanism based on criteria such as those applied in Figure 1. Likewise, significant caution should be exercised before interpreting the functional role of an LLPS-like system solely based on macroscopic behaviors.

We recently showed that the CTD of Pol II and other Pol II interacting partners can undergo LLPS in vitro and can form hubs in vivo (Boehning et al., 2018a; Lu et al., 2018). Given the data presented above, there appears a contradiction between this and our previous findings. We emphasize that our current results do not mean that interactions between IDRs are not important. Rather, our results suggest an "upper limit" for the potency Pol II CTD-mediated interactions to facilitate recruitment to RCs. While ectopic over-expression or in vitro preparations of IDRs may spontaneously create droplet-like structures (Figure 1 – figure supplement 1E), these condensates do not become enriched in RCs either through heterotypic interactions with the Pol II CTD, or with other viral IDRs.

Multiple viral proteins are known to interact with Pol II or other preinitiation complex components. While we tested the most prominent of these interactions, and found that Pol II remains recruited to the viral DNA in the absence of interactions with the viral protein ICP27 (Figure 3 – figure supplement 1), we cannot—nor do we wish to—rule out the possibility that other viral proteins may help facilitate this process. Importantly, our results do not contradict any of these unique mechanisms, but rather they provide a unifying rationalization for how they may work. As we demonstrated in Figure 6, even proteins that would never have been exposed to HSV1 over evolutionary time can still be recruited to RCs, provided they have some nonspecific affinity for DNA. In this way, any protein complex, be it solely viral or host or a composition of both, should be recruited to RCs provided it contains a DNA-binding domain.

Nonspecific DNA binding is an important feature for nuclear exploration

Our data also reveal a previously underappreciated aspect of how a DNA-binding protein finds its target site within the nucleus. It has long been recognized that nonspecific binding to DNA could accelerate the target search process by sliding in 1D; reducing the search space and empowering faster-than-diffusion association kinetics (Berg et al., 1981). The data we present here offer a new perspective on the importance of nonspecific low-affinity binding. When HSV1 replicates its genome, the newly synthesized viral DNA representing just 0.2% of the host chromosome load, is nevertheless, much more accessible to DNA-binding proteins than the totality of host chromatin (Table 1).

The finding that Pol II recruitment to RCs is independent of its CTD is reminiscent of RNA Polymerase I (Pol I) transcription of rDNA in the nucleolus. Pol I, lacking the long unstructured

CTD that its homolog Pol II contains, is nevertheless robustly recruited to the nucleolus and transcribes rDNA into ribosomal precursors at prodigious rates. While there are certainly differences in the structure and stability of nucleoli and RCs, it has been shown that nucleolar components indeed exchange with the rest of the nucleoplasm rapidly (Chen and Huang, 2001). It is tempting to speculate that recruitment of some nucleolar proteins may benefit from the same mechanism of non-specific DNA binding that drives recruitment of Pol II and other DNA-binding proteins to viral RCs. We speculate that nonspecific protein:nucleic-acid interactions could also be a general mechanism used in other contexts. In particular, many RNA-binding proteins have been reported to undergo apparent LLPS (Courchaine et al., 2016), and it will be interesting to explore if these RNA-binding proteins share a similarities to what we observe here.

Mechanism of Pol II recruitment may explain robust transcription of late genes

An unresolved question in the study of herpesviruses is how genes with seemingly weak promoter elements can sustain such robust transcription (Rutkowski et al., 2015). While it is clear that other regulatory components also play a role in regulating late gene transcription (Davis et al., 2015), our data may at least help shed light on how the virus robustly transcribes these late genes. After replication onset, when there are many copies of the viral genome present in a single RC, the compartmentalization of Pol II (and the other general transcription factors) mediated through nonspecific binding could greatly favor assembly of PICs at otherwise weak late gene promoters. In this way, the virus can conserve precious sequence space in its genome to encode other important features, relying on fundamental mechanisms of nuclear exploration for Pol II and other components of the transcription machinery while still providing sufficiently robust gene expression for these essential late genes.

Revisiting insights into chromatin function

DNA accessibility in eukaryotes has long been recognized as a critical parameter for gene regulation (Paranjape et al., 1994; Weintraub and Groudine, 1976), and many chromatin remodelers have been shown to play a role in modulating nucleosome occupancy at promoters and enhancers. *In vivo* experiments using sequence-specific eukaryotic transcription factors find that a given factor will spend approximately half its search time undergoing 3D diffusion, and the other half bound nonspecifically, presumably scanning in 1D (Normanno et al., 2015); that it may visit as many as 10^5 non-cognate sites during its search. These experiments highlight the challenge a cell faces ensuring that endogenous regulatory sequences are able to effectively compete for cognate DNA binding factors without becoming adversely influenced by non-target DNA sites. In this context, our results suggest that a less obvious—but critical—function of nucleosomes may involve the passivation of genomic DNA to minimize nonspecific interactions so as to maintain an active pool of freely diffusing nuclear factors, less hindered by their intrinsic propensity for nonspecific binding.

We postulate that a fine balance between the total amount of DNA binding proteins and the degree of accessible DNA content in the cell is critically important. Nucleosomes, in addition to their obvious structural role in DNA compaction and cis-repression, could serve to uncouple

cellular DNA content from the expression level of binding proteins. This mechanism of DNA passivation may be necessary in eukaryotes where the gene density and coding capacity is sparse, but total genomic load is very high; an essential step enabling the evolution of large genomes concomitant with the appearance of chromatin.

This may also point to a less obvious function for the observed increase in accessibility around promoters and enhancers, as a mechanism for effectively funneling DNA-binding proteins into the correct sites. The data presented above suggest that maintaining enhancers and promoters depleted of nucleosomes and accessible to DNA-binding proteins may contribute critically to facilitating the local accumulation of Pol II and other PIC components for transcription activation, without the need to invoke LLPS. In the case of RCs and the recruitment of Poll II, even well-established interactions between IDRs seem to be dispensable, underscoring the diversity of mechanisms driving local hub formation and functional compartments.

Figures

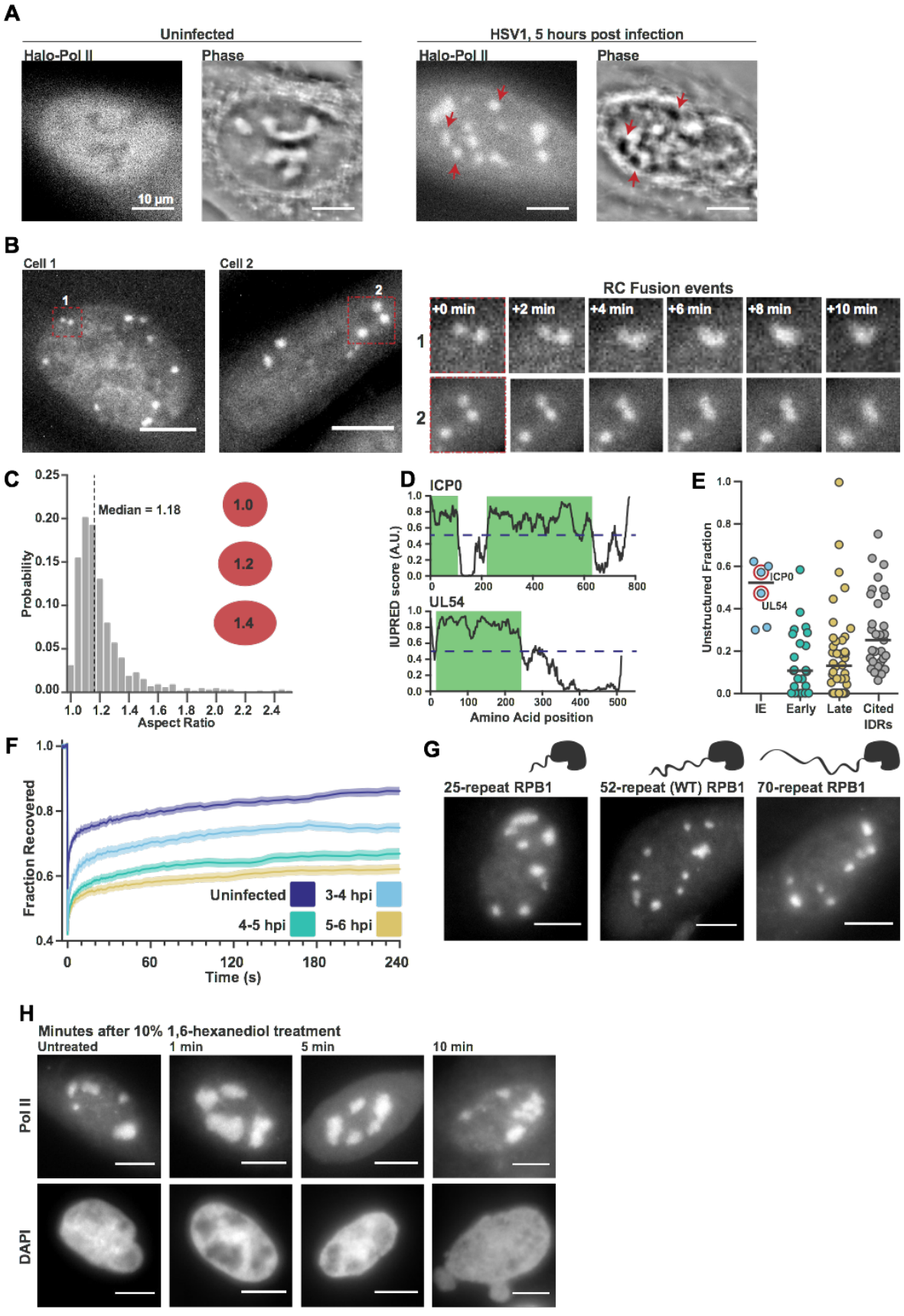

A
Uninfected
Halo-Pol II
Phase
10 µm
HSV1, 5 hours post infection
Halo-Pol II
Phase
B
Cell 1
Cell 2
1
2
RC Fusion events
+0 min
+2 min
+4 min
+6 min
+8 min
+10 min
C
Median = 1.18
Probability
Aspect Ratio
1.0
1.2
1.4
D
ICP0
UL54
IUPRED score (A.U.)
Amino Acid position
E
Unstructured Fraction
ICP0
UL54
IE
Early
Late
Cited IDRs
F
Fraction Recovered
Time (s)
Uninfected
3-4 hpi
4-5 hpi
5-6 hpi
G
25-repeat RPB1
52-repeat (WT) RPB1
70-repeat RPB1
H
Minutes after 10% 1,6-hexanediol treatment
Untreated
1 min
5 min
10 min
Pol II
DAPI

Figure 1. Pol II recruitment to Replication Compartments exhibits hallmarks of liquid-liquid demixing. **A)** Representative fluorescence and phase images in uninfected and infected cells. RCs shows a different phase value compared with the surrounding nucleoplasm. Red arrows show matched examples of RCs in the two channels. **B)** Time-lapse images of Pol II recruitment to RCs. Zoom in shows RC fusion events. See also Figure 1 – Video 1 and 2. **C)** Aspect ratios (max diameter/min diameter) of RCs from 817 RCs in 134 cells, 3 to 6 hpi. Red ellipses provided a guide to the eye of different aspect ratios. **D)** IUPred scores for two Immediate Early viral proteins, ICP0 and UL54, as a function of residue position. Green boxes are predicted IDRs. **E)** The fraction of each protein in the viral proteome that is unstructured, separated by kinetic class. HSV1 proteins are compared to a curated list of proteins containing IDRs known to drive phase separation (Cited IDRs). **F)** FRAP curves of Pol II in RCs from 3-4 hpi, 4-5 hpi, and 5-6 hpi (n= 24, 33, and 33), compared with uninfected cells (n= 31). Shown is the mean flanked by SEM. **G)** Infected HaloTag-RPB1 cell lines with a C-terminal domain containing different numbers of heptad repeats. **H)** Pol II localization 1, 5 and 10 minutes after 10% 1,6-hexanediol treatment. All scale bars are 10 μm. Source data for of the list of IDRs in the HSV genome as well as previously cited IDRs can be found in Figure 1 – source data 1 and 2, respectively.

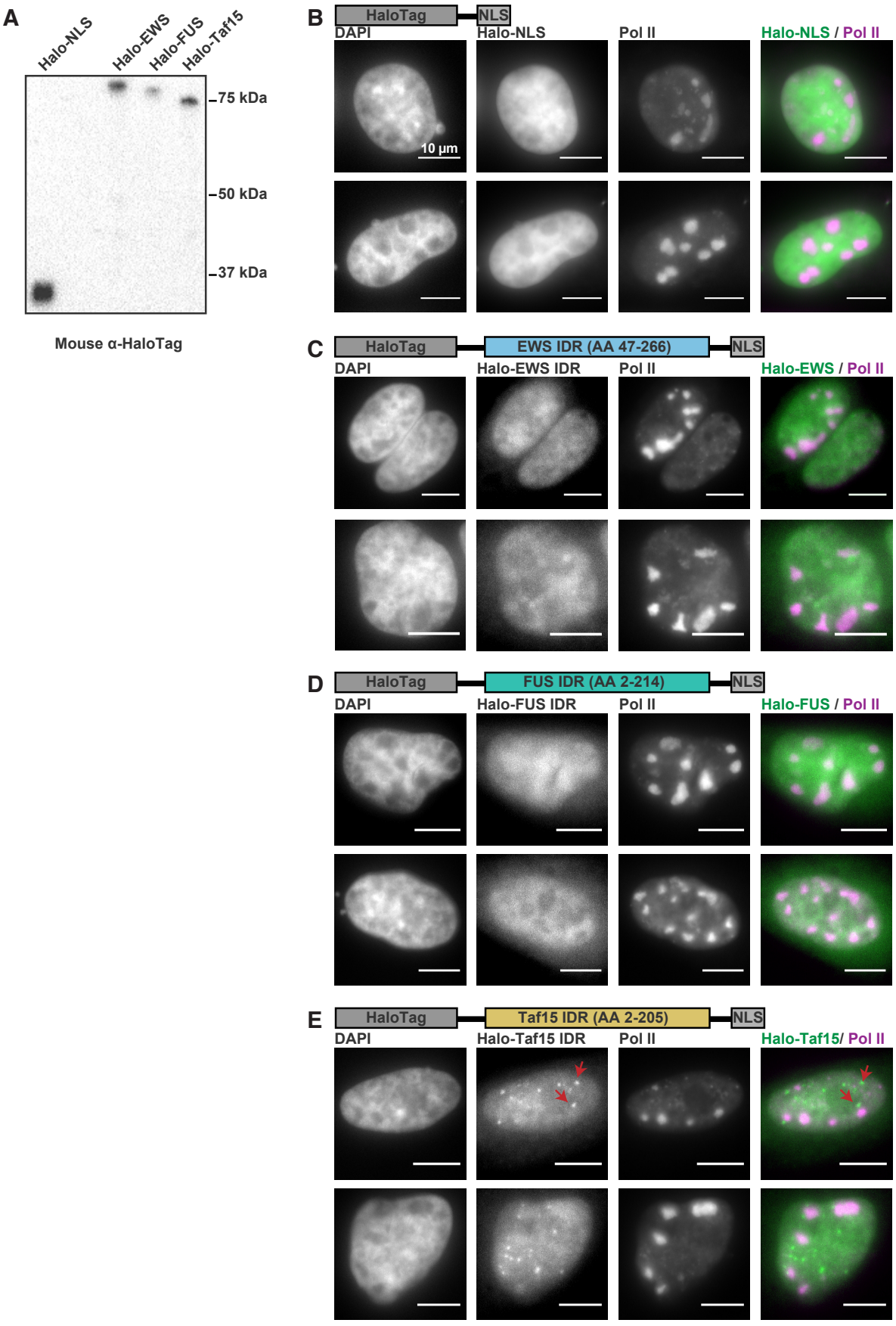

A
Halo-NLS
Halo-EWS
Halo-FUS
Halo-Taf15
75 kDa
50 kDa
37 kDa
Mouse α-HaloTag
B
HaloTag
NLS
DAPI
Halo-NLS
Pol II
Halo-NLS / Pol II
10 µm
C
HaloTag
EWS IDR (AA 47-266)
NLS
DAPI
Halo-EWS IDR
Pol II
Halo-EWS / Pol II
D
HaloTag
FUS IDR (AA 2-214)
NLS
DAPI
Halo-FUS IDR
Pol II
Halo-FUS / Pol II
E
HaloTag
Taf15 IDR (AA 2-205)
NLS
DAPI
Halo-Taf15 IDR
Pol II
Halo-Taf15/ Pol II

Figure 1 – figure supplement 1. FET family IDRs are not recruited to RCs. **A)** Western blot of whole cell extracts of U2OS cells transfected with Halo-NLS, Halo-EWS(LC), Halo-FUS(LC), and Halo-Taf15(LC). **B-E)** Two representative SNAPtag-Pol II cells expressing Halo-NLS (B), and the HaloTag fused to the IDRs from EWS (C), FUS (D), and Taf15 (D) (Chong et al., 2018). Cells were fixed 5 hours post infection. The Taf15 IDR has strong enough homotypic interactions to form puncta in nuclei (red arrows), but no IDR was enriched in RCs. All scale bars are 10 μm.

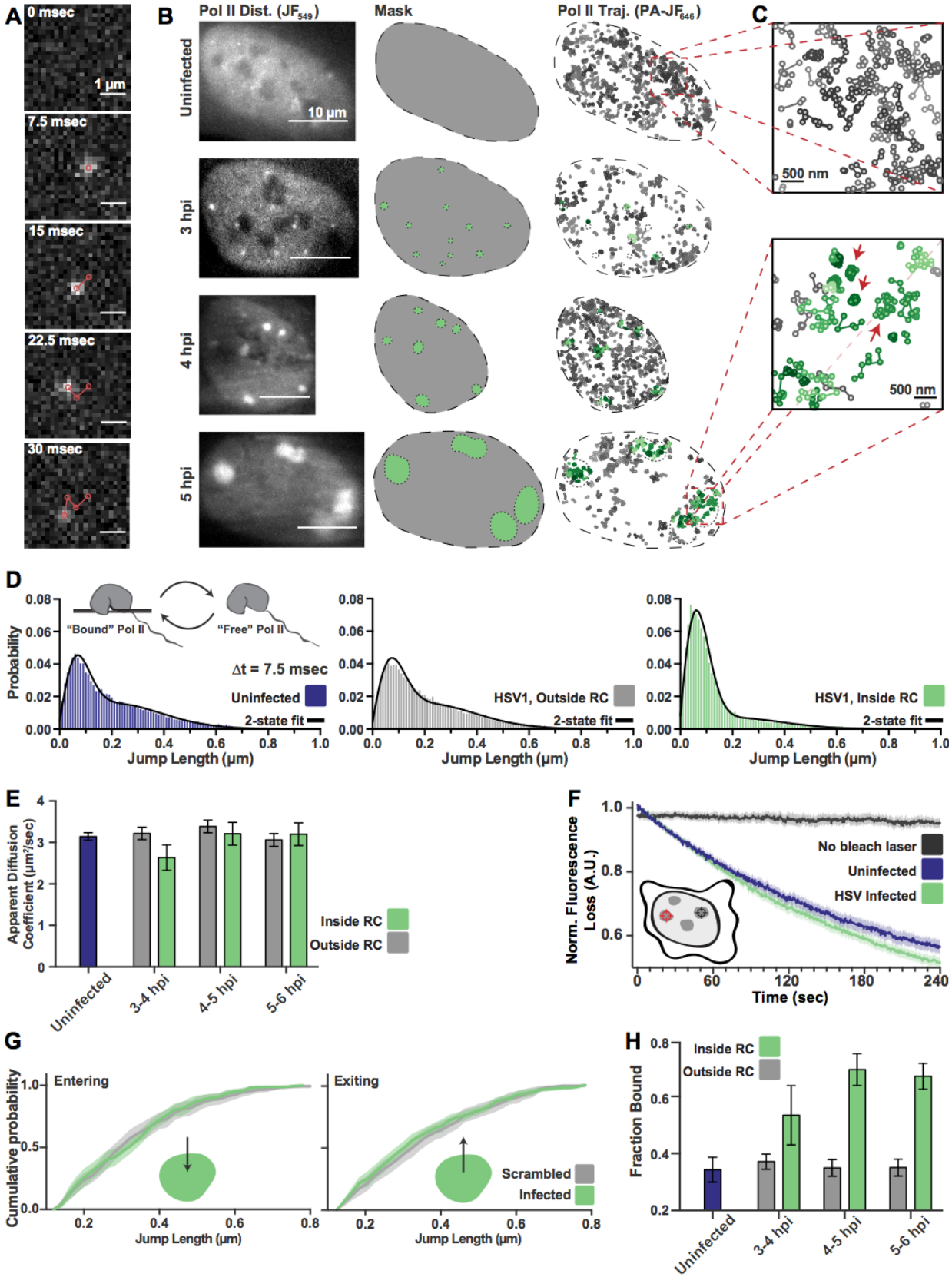

A
0 msec
1 μm
7.5 msec
15 msec
22.5 msec
30 msec
B
Pol II Dist. (JF549)
Mask
Pol II Traj. (PA-JF646)
Uninfected
10 μm
3 hpi
4 hpi
5 hpi
C
500 nm
500 nm
D
"Bound" Pol II
"Free" Pol II
Probability
Δt = 7.5 msec
Uninfected
2-state fit
Jump Length (μm)
HSV1, Outside RC
HSV1, Inside RC
E
Apparent Diffusion Coefficient (μm²/sec)
Inside RC
Outside RC
Uninfected
3-4 hpi
4-5 hpi
5-6 hpi
F
Norm. Fluorescence Loss (A.U.)
No bleach laser
Uninfected
HSV Infected
Time (sec)
G
Cumulative probability
Entering
Exiting
Scrambled
Infected
Jump Length (μm)
H
Fraction Bound
Inside RC
Outside RC
Uninfected
3-4 hpi
4-5 hpi
5-6 hpi

Figure 2. spaSPT of Pol II in infected cells shows no change in diffusion but an increase in binding. **A)** Example frames from spaSTP localization and tracking. Scale bar is 1 μm. **B)** spaSPT experiments in infected cells at different times post infection. RCs are identified using Pol II fluorescence and used to make masks for sorting trajectories (green inside RCs; grey outside). **C)** Zoom-in of trajectories in infected and uninfected cells. Red arrows show examples of traces with restricted movement. **D)** Jump length distributions between consecutive frames of spaSPT trajectories. Histograms pooled from uninfected cells (n= 27), or HSV1 infected cells between 4 and 6 hpi (n= 96). Each distribution is fit with a 2-state model. Inset shows depiction of two-state model where Pol II can either be freely diffusing or DNA-bound. **E)** Mean apparent diffusion coefficient from the 2-state fit in (D). Error bars are the standard deviation of the mean, calculated as described in Methods. **F)** FLIP curves comparing the rate of fluorescence loss after photobleaching Pol II in uninfected and HSV1 infected cells. Schematic shows location of bleaching laser (red crosshairs) and the region measured (black crosshairs). **G)** Cumulative distribution function of the mean flanked by the SEM for jump lengths of molecules entering (left) or exiting (right) RCs. The distribution for HSV1 infected cells is compared to the distribution of jump lengths when RC annotations have been shuffled randomly. **H)** Mean fraction of bound molecules from the 2-state fit in (D). Error bars are the standard deviation of the mean, calculated as described in Methods.

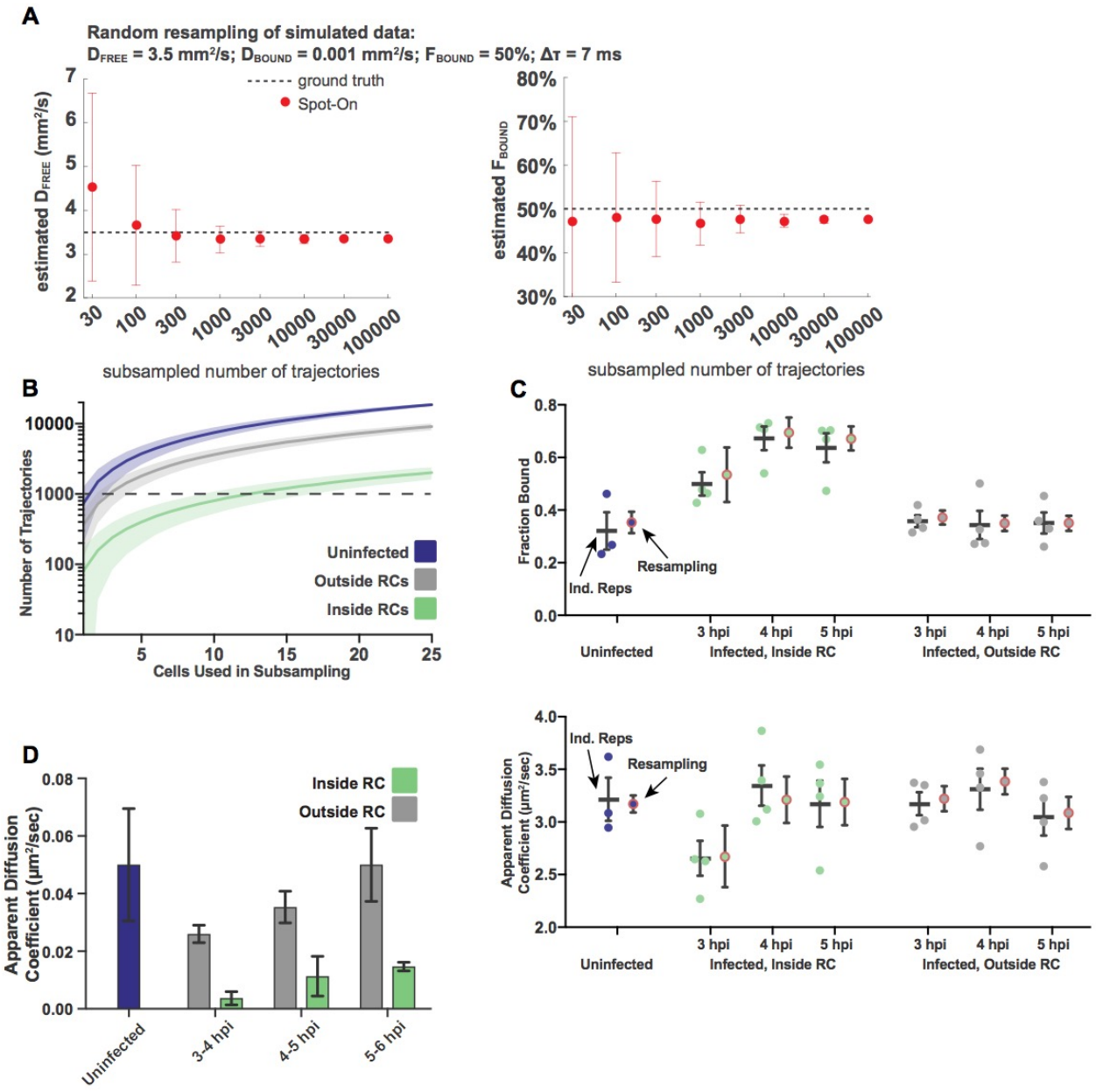


Figure 2 – figure supplement 1. Sampling statistics and quality measurements of spaSPT. Related to Figure 2. **A)** Measurements of the goodness of fit for Spot-On as a function of the number of trajectories sampled, using Monte Carlo simulation. Data taken from simulations performed in Hansen et al., 2018. **B)** The number of trajectories in the data set for uninfected and infected cells as a function of the number of cells randomly sampled from the data set. Plot shows the mean flanked by the standard deviation. Dashed line demarcates 1000 trajectories. **C)** Comparison of treating data as biological replicates versus using random subsampling. For each condition, the left bar shows the mean and SEM from at least three biological replicates, whereas the right shows the mean and standard deviation of the mean calculated from 100 resampling iterations. Either approach gives values within measurement error of each other. **D)** Mean diffusion coefficient of the Bound population determined through 2-state model fitting for uninfected cells,

and for cells at different times post infection, both inside and outside of RCs. In all data sets, the calculated diffusion coefficient is well below the upper bound set for the fitting, consistent with diffusion coefficients of chromatin (Hansen et al., 2018b). Error bars are the standard deviation of the mean, calculated from 100 iterations of randomly subsampling 15 cells without replacement and fitting with the model.

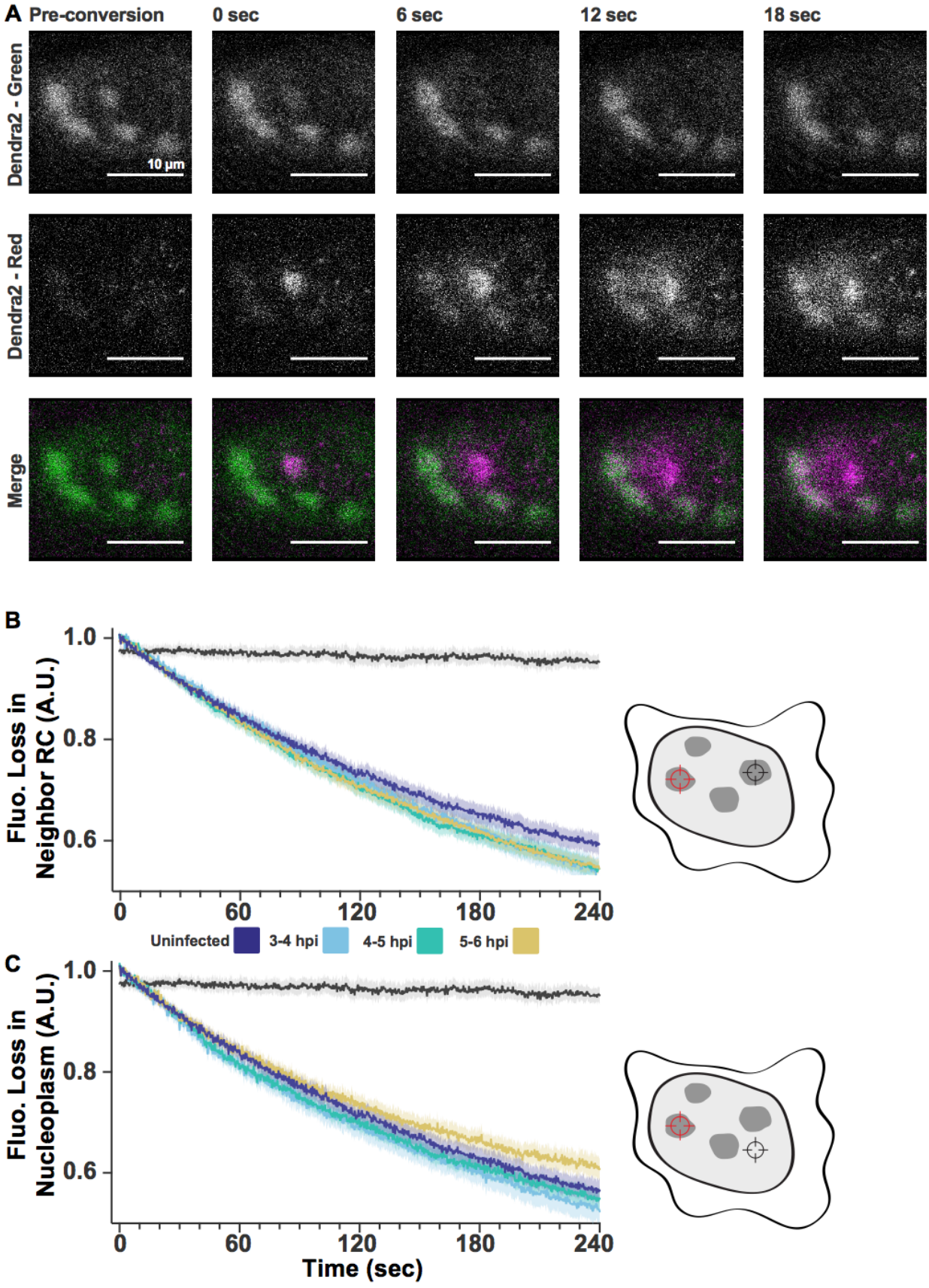

A
Pre-conversion
0 sec
6 sec
12 sec
18 sec
Dendra2 - Green
Dendra2 - Red
Merge
10 µm
B
Fluo. Loss in Neighbor RC (A.U.)
1.0
0.8
0.6
0
60
120
180
240
Uninfected
3-4 hpi
4-5 hpi
5-6 hpi
C
Fluo. Loss in Nucleoplasm (A.U.)
1.0
0.8
0.6
0
60
120
180
240
Time (sec)

Figure 2 – figure supplement 2. FLIP shows exchange within and between RCs. Related to Figure 2. **A)** Dendra2 photoconversion shows Pol II exchanges with nucleoplasm. Cells stable expressing Dendra2-Pol II were infected with HSV1. Fluorescence was monitored in both the green channel (pre-conversion), and red channel (post-conversion). A 1μm spot of 405nm light was used to convert one RC from green to red, alternating between photoconversion and frame acquisition. All scale bars are 10 μm. **B)** FLIP measurements as in Figure 2F, except separated by time post infection. All times after infection show the same decay coefficients. **C)** FLIP measurements of the loss of fluorescence in the nucleoplasm outside of RCs. All times after infection show similar decay coefficients.

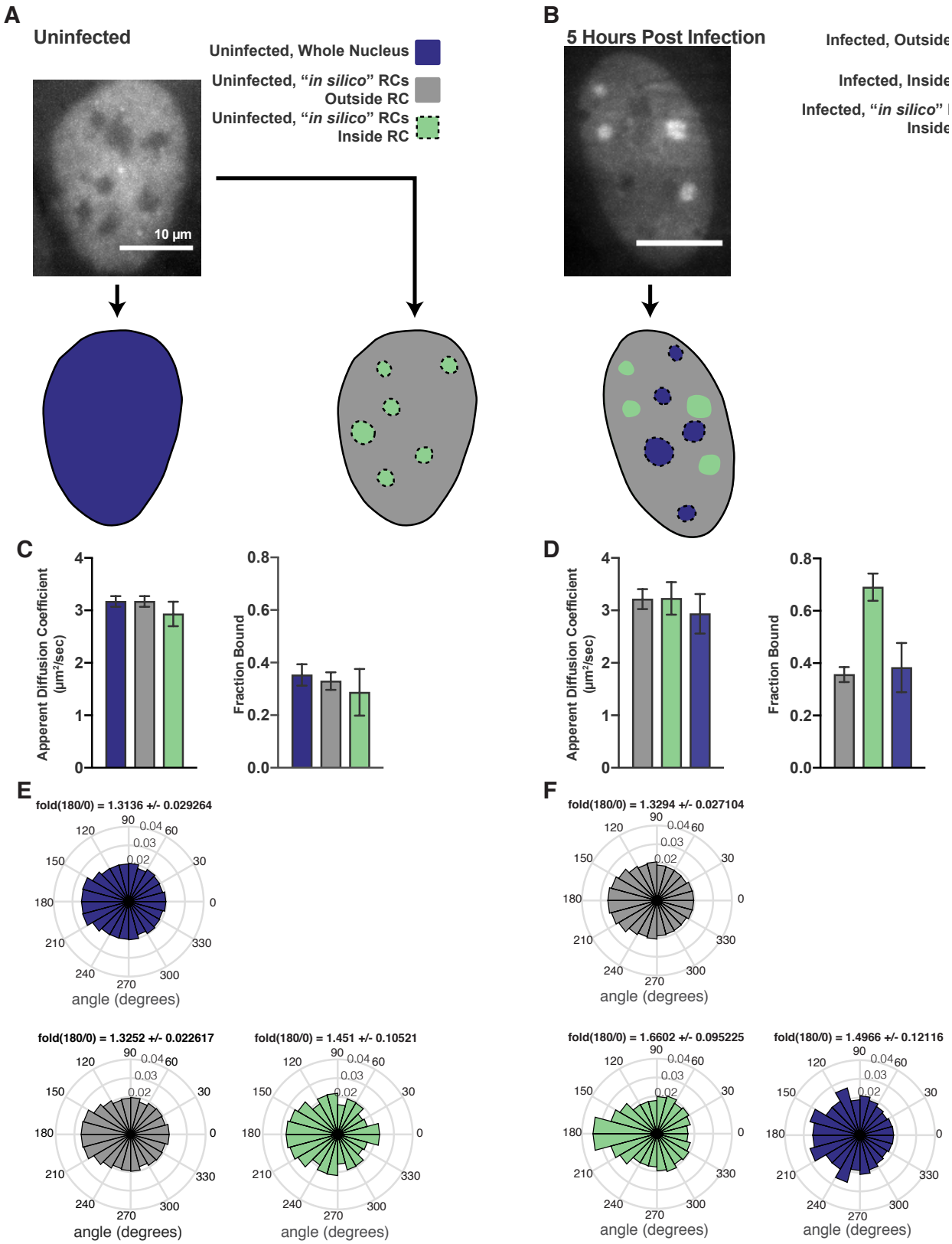

A
Uninfected
Uninfected, Whole Nucleus
Uninfected, "in silico" RCs Outside RC
Uninfected, "in silico" RCs Inside RC
10 µm
B
5 Hours Post Infection
Infected, Outside
Infected, Inside
Infected, "in silico"
Inside
C
Apperent Diffusion Coefficient (µm²/sec)
Fraction Bound
D
Apperent Diffusion Coefficient (µm²/sec)
Fraction Bound
E
fold(180/0) = 1.3136 +/- 0.029264
fold(180/0) = 1.3252 +/- 0.022617
fold(180/0) = 1.451 +/- 0.10521
angle (degrees)
F
fold(180/0) = 1.3294 +/- 0.027104
fold(180/0) = 1.6602 +/- 0.095225
fold(180/0) = 1.4966 +/- 0.12116
angle (degrees)

Figure 2 – figure supplement 3. Comparison of bona fide RCs with RCs generated in silico. Related to Figure 2. **A)** Example workflow for uninfected cells, where either just the nucleus was masked (left), or the nucleus was masked and RC-sized annotations were randomly placed inside the nucleus (right). **B)** Example workflow for HSV1 infected cells, where both the correct annotations based on the widefield image and randomly shuffled RCs were generated for all measured cells. **C)** Spot-On measurements of trajectories after inside/outside classification in uninfected cells. *In silico* shuffling of RC positions has very little effect on either the measured apparent diffusion coefficient or the fraction bound. Error bars are the standard deviation of the mean, calculated from 100 iterations of randomly subsampling 15 cells without replacement and fitting with the model. **D)** Similar to (C), but for infected cells. Real RCs show an increase in fraction bound, whereas *in silico* shuffled compartments show no difference with trajectories outside RCs. **E)** Angular distributions of Pol II trajectories in the regions marked in (A) Fold(180/0) is the mean plus/minus the standard deviation, calculated from 100 iterations of randomly subsampling 15 cells without replacement and fitting with the model. **F)** Angular distributions of Poll II trajectories in the regions marked in (B). Fold(180/0) is the mean plus/minus the standard deviation, calculated from 100 iterations of randomly subsampling 15 cells without replacement and fitting with the model. All scale bars are 10μm.

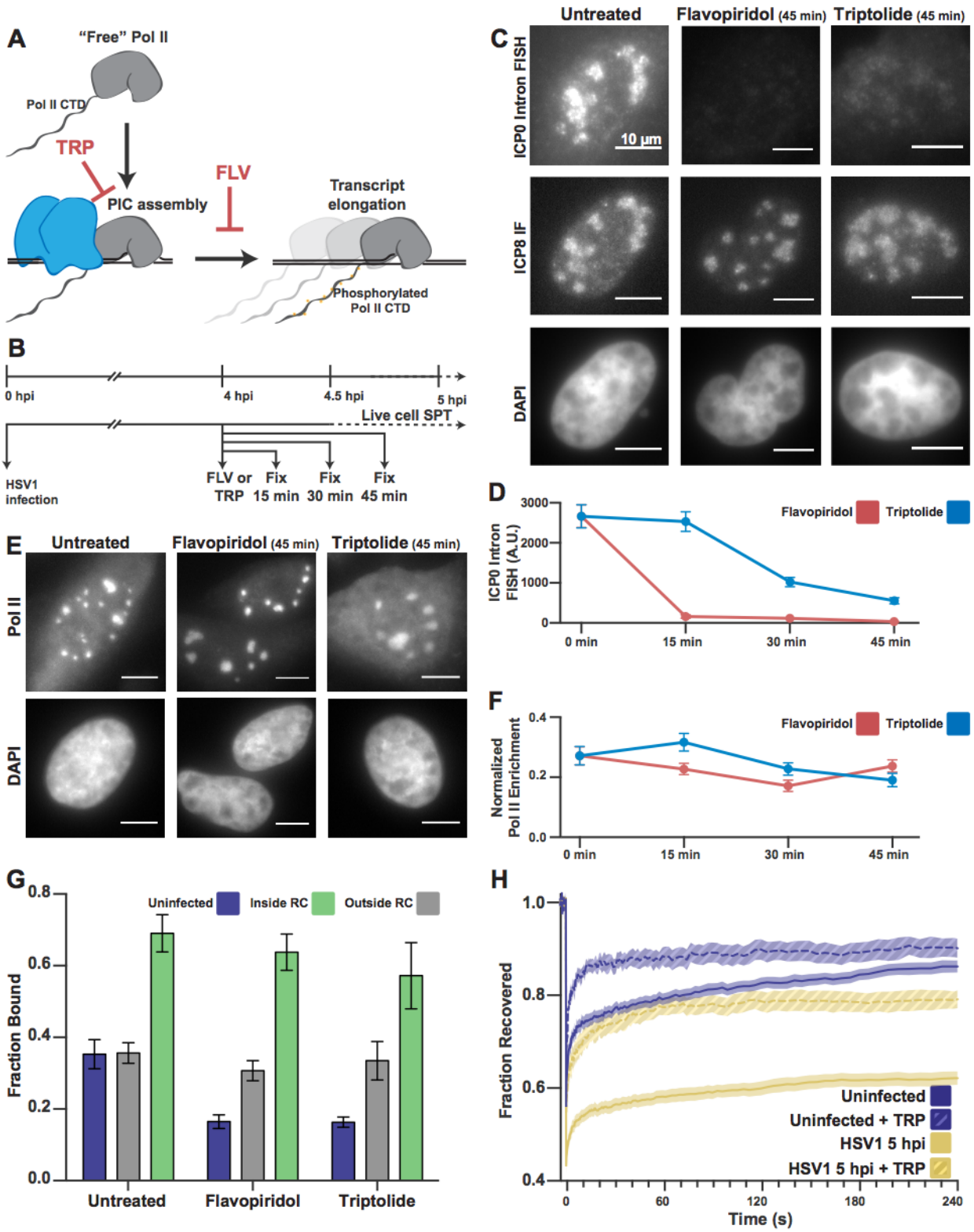


Figure 3. Pol II recruitment to RCs occurs independent of active transcription. **A)** Schematic of Pol II-mediated transcription inhibition. **B)** Schematic of the experiment regimen for imaging infected cells after transcription inhibition. **C)** RNA FISH against the ICP0 intron to measure

nascent transcription after Flavopiridol or Triptolide treatment. ICP8 marks viral RCs. **D)** Quantification of the ICP0 intron signal in untreated cells (n = 170 RCs) those treated with TRP(n = 192, 171, 191 RCs, respectively) and FLV(n = 158, 238, 153 RCs, respectively). Error bars are standard error of the mean. **E)** Halo-Pol II distribution after 45 minutes of Triptolide or Flavopiridol treatment. All scale bars are 10 μm. **F)** Quantification of the total fraction of Pol II recruited to RCs in untreated cells (n = 29) with TRP(n =33, 24, 33, respectively) and FLV(n = 36, 24, 38, respectively). Error bars represent standard error of the mean. **G)** Mean fraction bound measured from spaSPT of Halo-Pol II, after transcription inhibition. Error bars are the standard deviation of the mean, calculated as described in Methods. **H)** FRAP recovery curves of Pol II with (hashed) and without (solid) Triptolide treatment, for uninfected cells (n = 31, 9 respectively) and cells infected with HSV1, 5hpi (n = 32, 12 respectively).

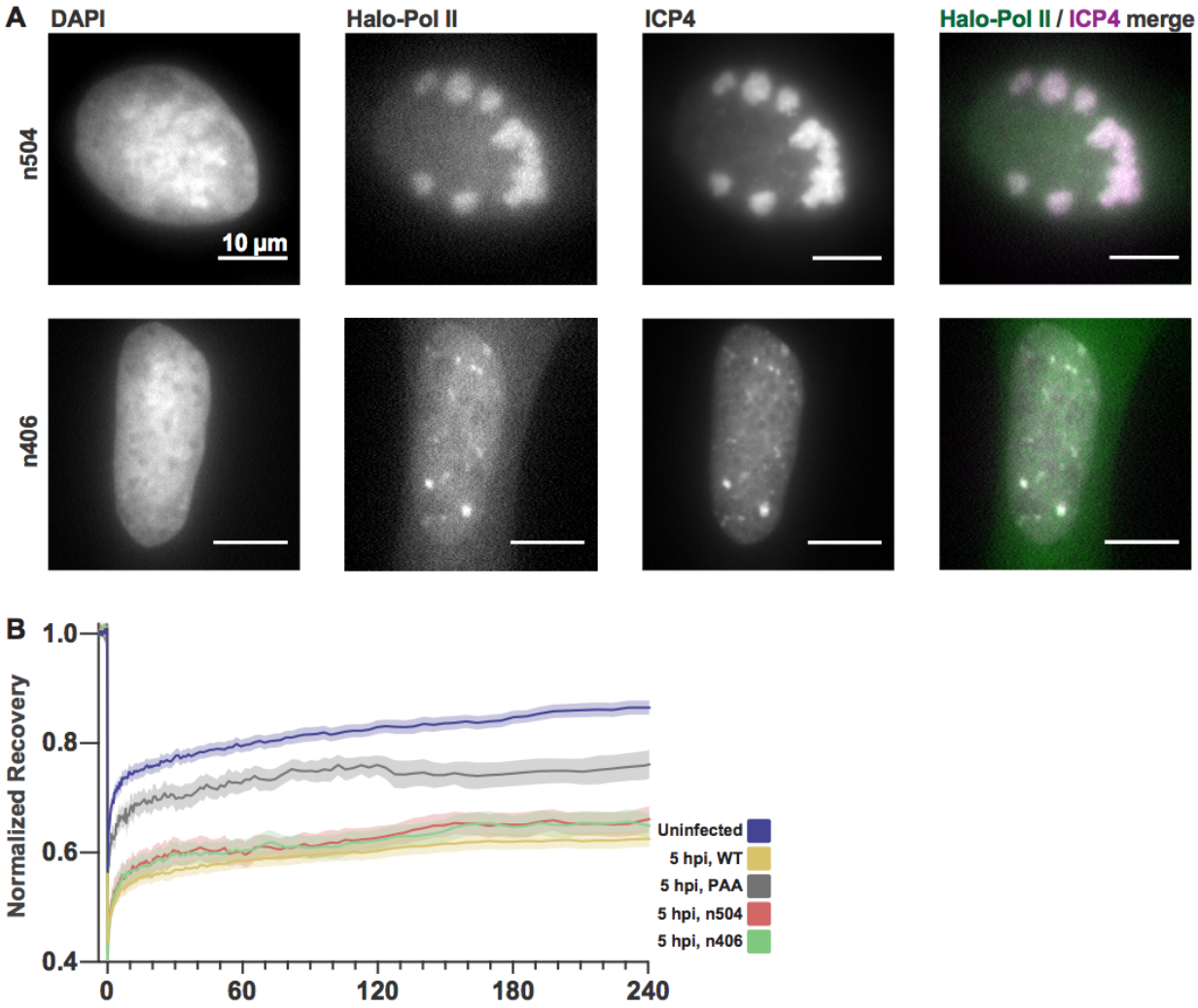


Figure 3 – figure supplement 1. HSV1 mutants affect neither Pol II recruitment nor binding dynamics. Related to Figures 2 and 3. **A)** n504 and n406 mutants containing nonsense mutations in the UL54 gene causing premature termination of the ICP27 protein. Representative fluorescence images of HaloTag-Pol II cells infected with either mutants. Immunofluorescence against ICP4 marks viral RCs. Both virus mutants accumulate Pol II in RCs. **B)** FRAP measurements of n504 and n406 mutants at 5 hpi, plotted with the uninfected and WT infected cells from Figure 1F, and WT infected cells treated with 300 µg/mL PAA to inhibit viral genome replication. Curves show the mean flanked by the SEM (n504 n = 10 cells, n406 n = 10 cells, WT n = 33 cells, WT PAA n = 8, Uninfected n = 31 cells).

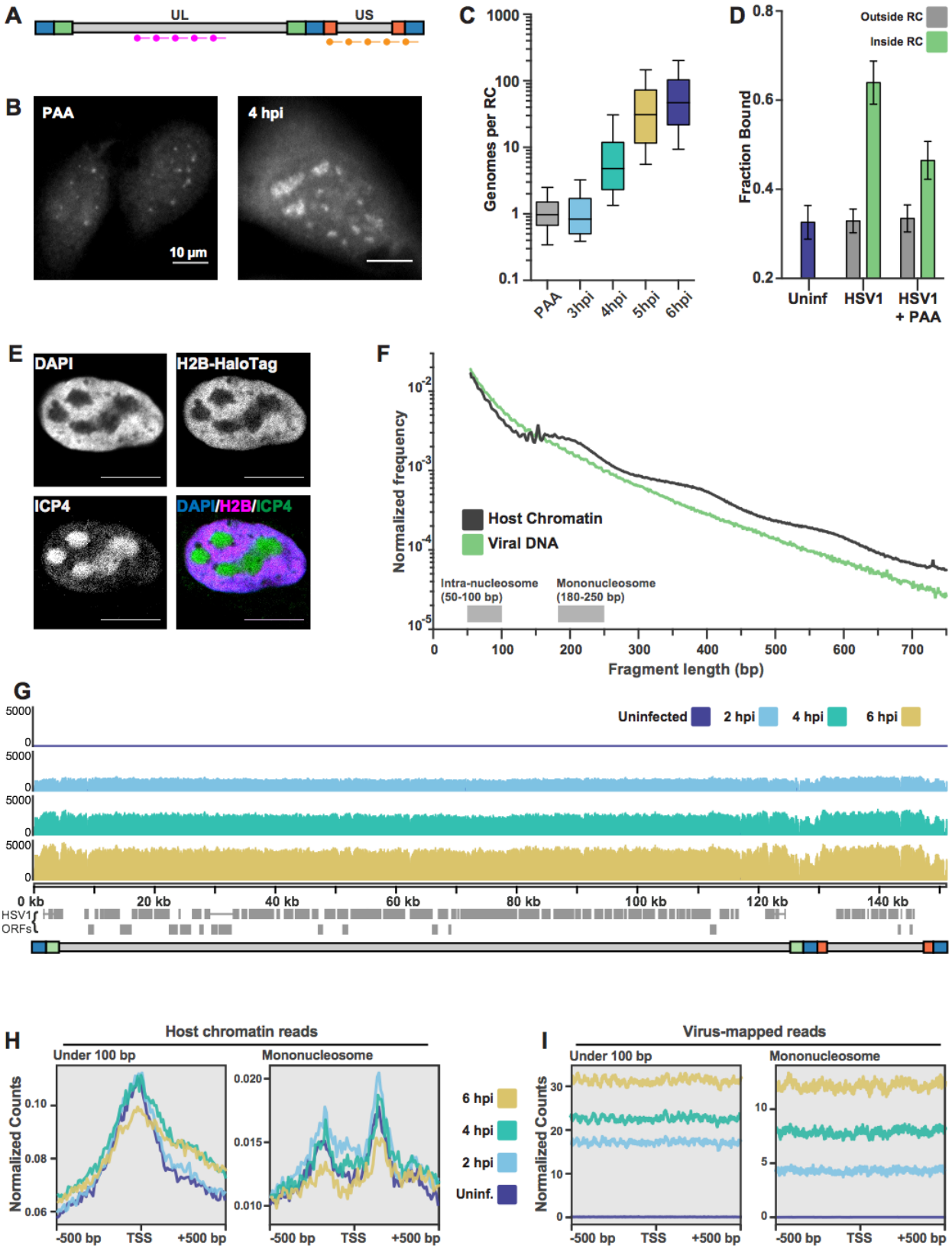

A
UL
US
B
PAA
4 hpi
10 µm
C
Genomes per RC
1000
100
10
1
0.1
PAA
3hpi
4hpi
5hpi
6hpi
D
Fraction Bound
0.8
0.6
0.4
0.2
Outside RC
Inside RC
Uninf
HSV1
HSV1 + PAA
E
DAPI
H2B-HaloTag
ICP4
DAPI/H2B/ICP4
F
Normalized frequency
Host Chromatin
Viral DNA
Intra-nucleosome (50-100 bp)
Mononucleosome (180-250 bp)
Fragment length (bp)
G
Uninfected
2 hpi
4 hpi
6 hpi
0 kb
20 kb
40 kb
60 kb
80 kb
100 kb
120 kb
140 kb
HSV1
ORFs
H
Host chromatin reads
Under 100 bp
Mononucleosome
Normalized Counts
-500 bp
TSS
+500 bp
6 hpi
4 hpi
2 hpi
Uninf.
I
Virus-mapped reads
Under 100 bp
Mononucleosome
Normalized Counts

Figure 4. ATAC-seq reveals HSV1 DNA is much more accessible than chromatin. **A)** Schematic of the Oligopaint targets for DNA FISH. Separate probe sets target regions in the Unique Long (UL) arm and the Unique Short (US) arm. **B)** Representative images of DNA FISH of cells 4 hpi, infected in the presence (PAA, left) or absence (4hpi, right) of the replication inhibitor PAA. Pixel intensity values are the same for the two images. Scale bars are 10 μm. **C)** Fluorescence intensity of DNA FISH signal in RCs after infection. 5%-95% intervals are shown, with inner quartiles and median. Data are normalized to the median intensity value of PAA-treated infected cells. Medians: PAA = 1.0, 3 hpi = 0.8, 4 hpi = 4.8, 5 hpi = 31.1, 6 hpi = 47.0. **D)** Mean fraction bound for Pol II in infected cells with and without PAA. Error bars are the standard deviation of the mean, calculated as described in Methods. **E)** H2B-Halo cells show histone H2B is not incorporated into RCs. Innumofluorescence against ICP4 marks RCs. **F)** Fragment length distribution of ATAC-seq data for cells 4 hpi. Lengths corresponding to intra-nucleosomal DNA (50–100 bp) and mononucleosomal DNA (180-250 bp) are marked as a reference. **G)** ATAC-seq read density plotted across HSV1 genomic coordinates. **H)** ATAC-seq analysis of intra-nucleosomal DNA (50–100 bp) and mononucleosomal DNA (180-250 bp). Global analysis of all human Pol II-transcribed genes, centered at the transcription start site (TSS). **I)** The same analysis as in (G), but centered at the TSS of HSV1 genes.

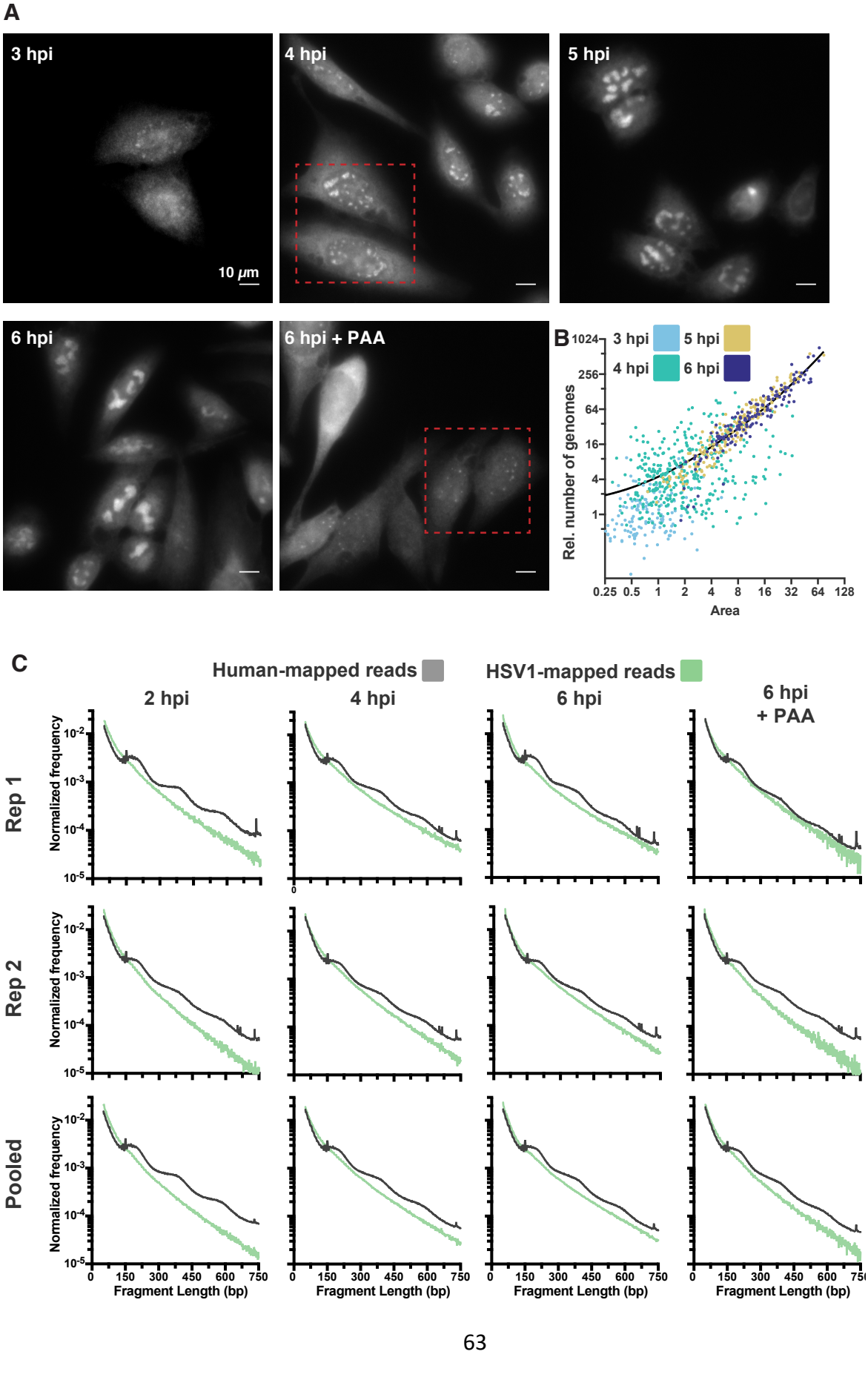

A
3 hpi
10 μm
4 hpi
5 hpi
6 hpi
6 hpi + PAA
B
3 hpi
5 hpi
4 hpi
6 hpi
Rel. number of genomes
1024
256
64
16
4
1
0.25 0.5 1 2 4 8 16 32 64 128
Area
C
Human-mapped reads
HSV1-mapped reads
2 hpi
4 hpi
6 hpi
6 hpi + PAA
Rep 1
Rep 2
Pooled
Normalized frequency
10-2
10-3
10-4
10-5
0 150 300 450 600 750
Fragment Length (bp)

Figure 4 – figure supplement 1. Quantification of DNA content and chromatin state in HSV1 RCs. Related to Figure 4. **A)** Representative full fields of view from DNA FISH hybridization. Dashed red boxes indicate the regions displayed in Figure 4B. **B)** Dot plot showing the individual RC values from Figure 4C; normalized to the median of cells infected in the presence of PAA. Medians are indicated by a solid line (PAA = 1, 3 hpi = 0.83, 4 hpi = 4.77, 5 hpi = 31.13, 6 hpi = 46.95). **C)** Fragment length distributions of all conditions tested after HSV1 infection, for two individual replicates as well as for the pooled data. The green line indicates the lengths of fragments mapping to the viral genome, the gray line indicates lengths of fragments mapping to the human genome. All data are normalized to the total number of mapped reads to the respective genome, per condition.

	Genome Size (bp)	Genome Copy number[c]	Total DNA (bp)	Percent of Total DNA[c]	Concentration (bp/μm³)[d]	ATAC-seq read percentage[e]	Fold enrichment over expected[f]
Host Genome[a]	3.2×10^9	2	6.4×10^9	99.8 (± 0.2)	9.4 (± 1.6) $\times 10^6$	75.8 (± 10.4)	0.8 (± 0.1)
Viral DNA	1.5×10^5	82 (± 105)	1.3 (± 1.6) $\times 10^7$	0.2 (± 0.2)	3.9 (± 5.8) $\times 10^4$	24.2 (± 10.4)	130 (± 170)
Rel. Diff.[b]	2.1×10^4		513 (± 658)		240 (± 369)		

All values are the Mean (± S.D.).
a. Assuming karyotypically normal human cell; b. relative difference = Human / HSV1; c. Under experimental conditions of MOI = 1; d. Concentration assuming nucleus volume taken from Monier et al., 2000; e. based on total reads mapped from each organism, n = 3; f. Fold enrichment = ATAC-seq read percentage / Percent of Total DNA.

Table 1 Quantitative measurements of HSV1 DNA inside of RCs. Related to Figure 4. Using the values obtained through DNA FISH and ATAC-seq, we can make estimates of the copy number, concentrations, and relative enrichment of the viral DNA compared to the host. All values are calculated based on measurements of cells 6 hpi.

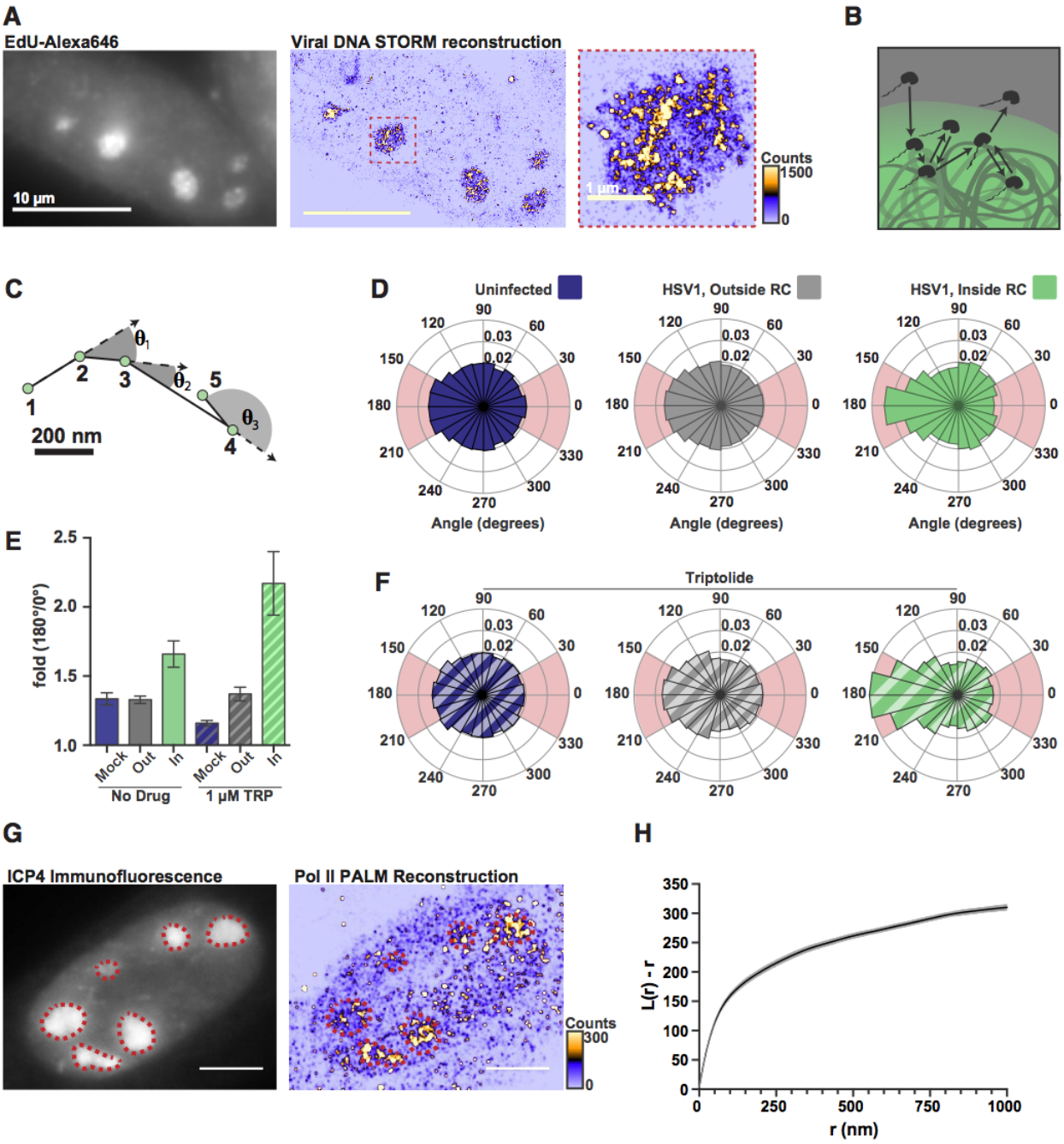


Figure 5. DNA-binding alters Pol II exploration of RCs. **A)** STORM image of fluorescently labeled HSV1 DNA. Zoom-in shows one RC, and the heatmap shows the number of fluorophore localizations in each rendered pixel. **B)** Schematic of Pol II exploring an RC and randomly sampling the viral DNA. **C)** Example spaSPT trace, marking the angles between consecutive steps. **D)** Angular distribution histograms extracted from Halo-Pol II in uninfected cells, and HSV1 infected cells 4-6 hpi, inside and outside of RCs. **E)** Quantification of the relative probability of moving backward compared to forward (180° ± 30° / 0° ± 30°). Error bars are the standard deviation of the mean, calculated as described in Methods. **F)** Same as in (D), except that cells were treated with Triptolide at least 30 minutes prior to imaging. Quantification of this data is also

show in (E). **G)** Representative PALM image of Halo-Pol II. ICP4 marks viral RCs. Heatmap corresponds to the number of detections per rendered pixel. **H)** L-modified Ripley Curve (L(r)-r) for Halo-Pol II inside of RCs in cells 5 hpi (n = 13 cells). Graph shows the mean flanked by the SEM. All scale bars are 10 μm.

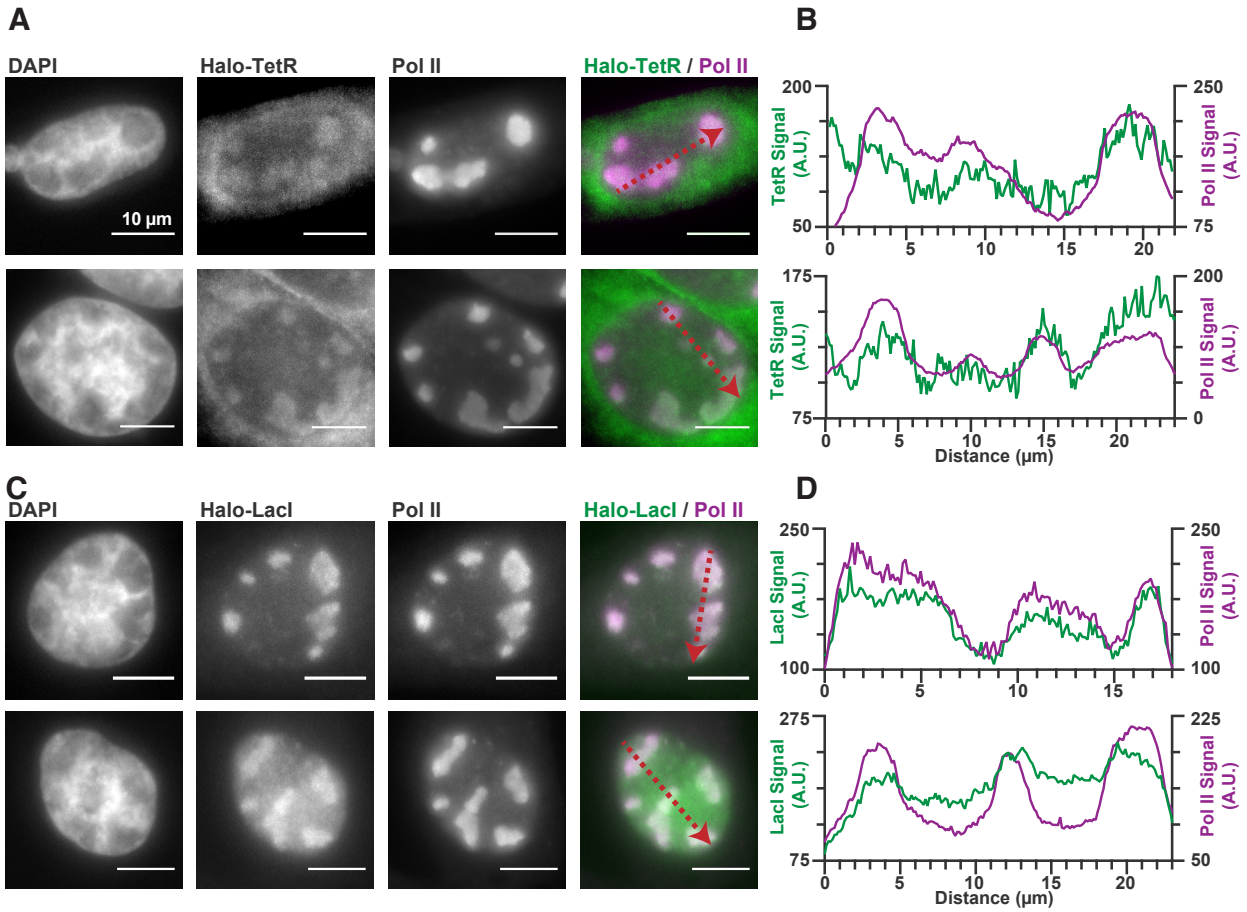


Figure 6. Nonspecific DNA binding drives accumulation of other factors in RCs. **A and C)** Two representative cells from SNAPtag-RPB1 cells expressing TetR-Halo (A) and LacI-Halo (C), showing that both bacterial transcription factors are enriched in RCs. **B and D)** Pixel line scans of images in (A) and (C). Red arrows give the direction of the x-axis. Left y-axis is the intensity of TetR-Halo or LacI-Halo fluorescence, right y-axis is the intensity of SNAPtag--Pol II fluorescence. All scale bars are 10 µm. Also see Figure 6 – figure supplement 1.

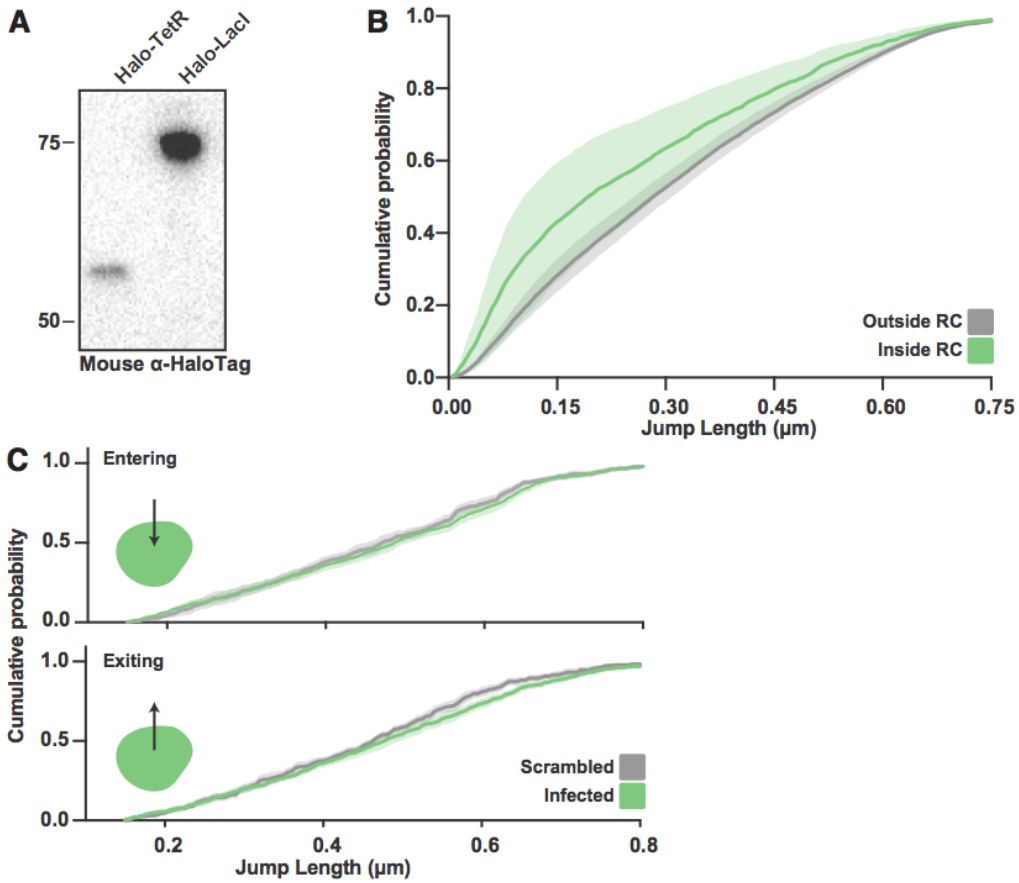


Figure 6 – figure supplement 1. SPT of Halo-TetR in infected cells. Related to Figure 6. **A)** Anti-HaloTag western blot from SNAPtag-Pol II cells expressing Halo-TetR and Halo-LacI. **B)** CDF of Halo-TetR displacements inside and outside RCs. Curve shows the mean flanked by the standard deviation as calculated by random resampling (see STAR methods). SPT data for TetR-Halo were not well fit by the two state model in Spot-On, however a qualitative assessment can be made from the CDF curves. The shift to the short displacements inside of RCs is a strong indication of an increase in binding events. **C)** CDF of displacements from spaSPT for Halo-TetR entering (top) or exiting (bottom) RCs. Data for jumps in and out of RCs is compared to jumps in cells where the annotations have been randomly shuffled. Curve shows the mean flanked by the standard deviation as calculated by random resampling (see Methods).

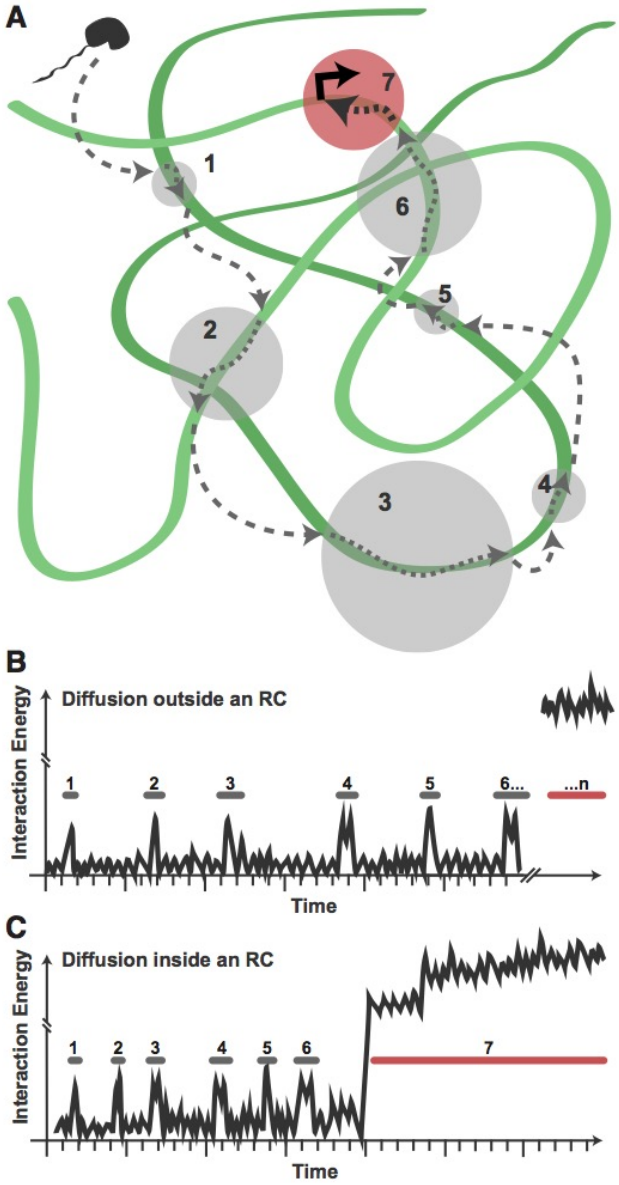


Figure 7. Model for Pol II exploration of RCs. **A)** A Pol II molecule encounters the accessible viral DNA multiple times along one potential route to eventually bind at a promoter. 3D diffusion through the RC is interrupted by binding interactions with the viral DNA (gray circles). **B)** Hypothetical comparison of nuclear exploration outside RCs as a function of time and binding energy. A DNA-binding protein in the chromatinized nucleus will encounter nucleosome-free DNA sporadically, making multiple low-affinity interactions before eventually finding a high affinity site. **C)** Inside an RC, the high DNA accessibility might shorten the length of 3D excursions before a DNA-binding protein encounters another region of viral DNA in a low affinity, nonspecific interaction. This, in turn, may reduce the distance a molecule might diffuse before its next binding event, and increases both the chances of that molecule remaining in close proximity and the chances that it will find a high binding energy interaction.

Figure 1- source data 1. List of putative IDRs in the HSV1 genome identified by IUPred. Related to Figure 1. Each protein listed was analyzed as described in the Methods section, and regions with an IUPred score of greater than 0.55 were recorded.

gene name	function	Class	Protein Length	Region Start	Region End	IDR Length
RS1	transcription regulation	alpha	1294	0	79	79
				94	273	179
				594	615	21
				717	820	103
				884	893	9
				1214	1236	22
				1276	1294	18
US12	transcription regulation	alpha	88	35	88	53
US11	tegument	tegument	149	0	149	149
US10	tegument	tegument	300	23	92	69
				124	156	32
US9	membrane/glycoprotein	beta	57	0	16	16
				54	57	3
US8	membrane/glycoprotein	beta	190	0	5	5
				31	104	73
				172	190	18
US8A	membrane/glycoprotein	beta	550	161	215	54
				392	410	18
				480	550	70
US7	membrane/glycoprotein	beta	383	195	253	58
				325	383	58
US6	membrane/glycoprotein	beta	394	277	319	42
				381	394	13
US4	membrane/glycoprotein	gamma	239	33	170	137
US3	kinase/phosphatase	beta	481	19	170	151
US2	unknown	gamma	291	254	291	37
US1.5	unknown	alpha	250	0	10	10
				137	212	75
US1	replication	alpha	420	0	14	14
				15	181	166
				307	382	75
RL2	transcription regulation	alpha	776	0	106	106
				222	568	346
				595	628	33
				761	776	15
UL56	membrane/glycoprotein	beta	234	0	101	101
				151	187	36
UL55	unknown	beta	186	179	186	7
UL54	transcription regulation	alpha	512	0	243	243
UL52	replication	beta	1058	0	12	12
				380	389	9
				480	497	17
				699	735	36
				1055	1058	3
UL51	tegument	tegument	244	180	244	64
UL50	replication	beta	371	0	7	7
				155	159	4
				333	371	38
UL49A	membrane/glycoprotein	beta	91	29	44	15
UL49	tegument	tegument	301	0	181	181
				270	301	31
UL48	transcription regulation	gamma	490	0	46	46
				454	456	2
				487	490	3
UL47	tegument	tegument	693	0	26	26
				51	125	74
				150	180	30

				680	693	13
UL46	tegument	tegument	719	0	8	8
				435	540	105
				561	606	45
				667	694	27
				706	719	13
UL45	membrane/glycoprotein	gamma	172	0	10	10
UL44	membrane/glycoprotein	gamma	511	39	125	86
				314	330	16
UL43	membrane/glycoprotein	gamma	415	0	17	17
				218	255	37
UL42	replication	beta	489	0	25	25
				333	448	115
				477	488	11
				110	148	38
UL41	tegument	tegument	489	286	367	81
UL40	replication	beta	340	0	13	13
UL39	replication	beta	1137	0	30	30
				125	157	32
				176	309	133
UL38	capsid	gamma	465	0	56	56
				70	85	15
				144	155	11
				357	388	31
UL37	tegument	tegument	1123	0	46	46
				971	977	6
				1057	1123	66
UL36	tegument	tegument	3136	0	22	22
				268	382	114
				396	495	99
				749	770	21
				948	971	23
				1254	1282	28
				1911	1925	14
				2267	2291	24
				2489	2534	45
				2553	2701	148
				2728	2984	256
				3029	3066	37
UL35	capsid	gamma	112	0	10	10
				40	46	6
				103	112	9
UL34	membrane/glycoprotein	gamma	275	0	11	11
UL33	packaging	gamma	130	0	14	14
UL32	packaging	gamma	596	0	5	5
				77	107	30
				227	237	10
UL31	other	gamma	306	0	41	41
UL30	replication	beta	1235	0	16	16
				50	61	11
				644	693	49
				1099	1134	35
				1233	1235	2
UL29	replication	beta	1196	0	8	8
				288	307	19
				1158	1196	38
UL28	packaging	gamma	785	265	288	23
				435	491	56
				778	785	7
UL27	membrane/glycoprotein	beta	904	45	99	54
				469	492	23
				819	836	17
				881	904	23
UL26.5	capsid	gamma	329	0	15	15
				108	171	63
				211	295	84

UL26	capsid	gamma	635	0	11	11
				270	318	48
				414	477	63
				517	601	84
UL25	packaging	gamma	580	108	132	24
UL24	unknown	gamma	269	0	3	3
				190	242	52
				265	269	4
UL23	replication	beta	376	0	41	41
				263	280	17
				371	376	5
UL22	membrane/glycoprotein	gamma	838	175	210	35
				293	317	24
UL21	tegument	tegument	535	206	231	25
				246	277	31
UL20	membrane/glycoprotein	gamma	222	0	10	10
UL19	capsid	gamma	1374	0	7	7
UL18	capsid	gamma	318	0	2	2
UL17	packaging	gamma	703	202	250	48
UL16	tegument	tegument	373	0	32	32
			373	154	175	21
UL15	packaging	gamma	735	0	3	3
UL14	tegument	tegument	219	0	7	7
				164	219	55
UL13	tegument	tegument	518	0	115	115
UL12	replication	beta	626	0	125	125
				603	626	23
UL11	tegument	tegument	96	48	96	48
UL10	membrane/glycoprotein	gamma	473	0	14	14
				366	427	61
				444	473	29
UL9	replication	beta	851	0	19	19
				263	283	20
UL7	tegument	tegument	296	0	6	6
UL6	capsid	gamma	676	0	17	17
				384	401	17
				629	676	47
UL5	replication	beta	882	0	34	34
				602	629	27
UL4	unknown	gamma	199	149	160	11
UL3	unknown	gamma	235	42	88	46
				144	158	14
				233	235	2
UL2	DNA repair	beta	334	0	94	94
UL1	membrane/glycoprotein	gamma	224	164	224	60
RL1	transcription regulation	late	240	0	130	130
				144	168	24
				236	240	4
LRP1	latency	N/A	300	0	99	99
				256	300	44
LRP2	latency	N/A	60	0	60	60

Figure 1- source data 2. List of proteins reported to undergo phase separation. Related to Figure 1. Gene name, organism of origin, size, and the fraction of the protein that scores as an IDR according to the analysis described in the Methods section. References and the citation within and provided.

Protein	Organism[a]	Protein Length (AA)	Total IDR length (AA)	IDR Fraction	Reference[b]
FUS	Hs	526	397	0.75475285	Chong et al., 2018
Taf15	Hs	592	274	0.46283784	
EWS	Hs	656	420	0.6402439	
SP1	Hs	785	220	0.28025478	

Huntingtin	Hs	3142	311	0.09898154	Bergeron-Sandoval et al., 2016
PML	Hs	882	145	0.16439909	
PGL-1	Ce	730	94	0.12876712	
RPB1	Hs	1970	119	0.06040609	Boehning et al., 2018
DDX4	Hs	724	144	0.19889503	Courchaine et al., 2016
eIF4GII	Sc	907	229	0.25248071	
Fibrillarin	Hs	321	89	0.27725857	
hnRNPA1	Hs	320	48	0.15	
Laf1	Ce	708	232	0.32768362	
Lsm4	Sc	187	92	0.49197861	
RBM14	Hs	669	85	0.12705531	
SRSF2	Hs	221	135	0.61085973	
TDP-43	Hs	414	82	0.19806763	
Tia1	Hs	386	35	0.09067358	
Whi3	Ag	729	359	0.49245542	
PUB1	Hs	453	213	0.47019868	
HP1a	Dm	213	68	0.31924883	Strom et al., 2017
DAXX	Hs	740	407	0.55	Banani et al., 2017
PGL-3	Ce	693	124	0.17893218	
NPM1	Hs	294	125	0.42517007	
hRNPAB	Hs	332	93	0.28012048	Aguzzi and Altmeyer, 2016
hnRNPA3	Hs	378	56	0.14814815	
hnRNPA2B1	Hs	353	41	0.11614731	
hnRNPD	Hs	355	40	0.11267606	
hnRDL	Hs	420	127	0.30238095	
NUP145	Sc	1317	219	0.16628702	Schmidt and Görlich, 2015
a) Organism abbreviations: Hs, Homo sapiens; Ce, Caenorhabditis elegans; Sc, Saccharomyces cerevisiae; Dm, Drosophila melanogaster; Ag, Ashbya gossypii b) May include citations within reference.					

Figure 1 – Video 1 and 2. Time lapse movies of HaloTag-Pol II after HSV1 infection. Related to Figure 1. Cells were identified 3 hpi, and followed until they moved out of the focal plane.

Figure 2 – Video 1 and 2. Examples of SPT data. Related to Figure 2. Example 500 frames, played at $1/10^{th}$ normal speed, from SPT data collected for the cells shown in Figure 2B from an uninfected cell (Video 1) and a cell infected for 4 hours (Video 2). Examples were taken from data sets with relatively high densities of localizations per frame to illustrate tracking and sorting into compartments, but in general the localization density was kept much lower, at approximately 0.5 localizations per frame.

Materials and Methods

Key Resources Table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
antibody	anti-ICP8 (mouse monoclonal)	Abcam	Cat# ab20193, RRID:AB_445413	2 µg/mL
antibody	anti-ICP4 (mouse monoclonal)	Abcam	Cat# ab6514, RRID:AB_305537	2 µg/mL
antibody	anti-ICP27 (mouse monoclonal)	Abcam	Cat# ab31631, RRID:AB_732867	2 µg/mL
antibody	anti-HaloTag (mouse monoclonal)	Promega	Cat# G9211, RRID:AB_2688011	1:1000
antibody	Donkey anti-Mouse IgG (H+L) Secondary	Thermo Fisher	Cat# A-31571, RRID:AB_162542	1 µg/mL

	Antibody, Alexa Fluor 647			
antibody	Goat anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	Thermo Fisher	Cat# A-11001, RRID:AB_2534069	1 µg/mL
antibody	Goat anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 546	Thermo Fisher	Cat# A-11003, RRID:AB_141370	1 µg/mL
antibody	Goat anti-Mouse IgG (H+L) Secondary Antibody, HRP	Thermo Fisher	Cat# 31430, RRID:AB_228307	1:10000
Strain (Herpes Simplex Virus Type 1)	KOS	PMID: 26676778	KOS	
Strain (Herpes Simplex Virus Type 1)	UL2/50	PMID: 26018390	UL2/50	Deletion of the UL2 and UL50 genes to facilitate incorporation of noncanonical nucleotides into viral genome. A generous gift from Neal DeLuca.
Strain (Herpes Simplex Virus Type 1)	n504	PMID: 2157053	n504	Early termination of the ICP27 (UL54) gene. Propagated in V27 cells, both virus strain and complementing cell line a generous gift from Stephen Rice.
Strain (Herpes Simplex Virus Type 1)	n406	PMID: 2157053	n406	Early termination of the ICP27 (UL54) gene. Propagated in V27 cells, both virus strain and complementing cell line a generous gift from Stephen Rice.
chemical compound, drug	Alexa Fluor® 647 Azide, Triethylammonium Salt	Thermo Fisher	Cat# A10277	
chemical compound, drug	Alexa Fluor® 555 Azide, Triethylammonium Salt	Thermo Fisher	Cat# A20012	
chemical compound, drug	Triptolide from Tripterygium wilfordii, ≥98% (HPLC), solid	Sigma Aldrich	Cat# T3652-1MG	
chemical compound, drug	Flavopiridol	Santa Cruz Biotech	Cat# sc-202157; CAS 146426-40-6	
chemical compound, drug	Phosphonoacetic acid - 98%	Sigma Aldrich	Cat# 284270-10G	
chemical compound, drug	alpha-Amanitin,from Amanita phalloides, >=90% (HPLC), powder	Sigma Aldrich	Cat# A2263-1MG	
chemical compound, drug	Catalase from bovine liver,lyophilized powder, >=10,000 units/mg protein	Sigma Aldrich	Cat# C40-100MG	

chemical compound, drug	Glucose Oxidase from Aspergillus niger Type VII, lyophilized powder, ≥100,000 units/g solid (without added oxygen) Type VII, lyophilized powder, ≥100,000 units/g solid (without added oxygen)	Sigma Aldrich	Cat# G2133-250KU	
chemical compound, drug	(±)-6-Hydroxy-2,5,7,8-tetramethylchromane-2-carboxylic acid (Trolox)	Sigma Aldrich	Cat# 238813-1G	
chemical compound, drug	Cyclooctatetraene (98%)	Sigma Aldrich	Cat# 138924-1G	
chemical compound, drug	HaloTag PA-JF549 ligand	PMID: 27776112		A generous gift from Luke Lavis
chemical compound, drug	HaloTag PA-JF646 ligand	PMID: 27776112		A generous gift from Luke Lavis
chemical compound, drug	HaloTag JF549 ligand	PMID: 25599551		A generous gift from Luke Lavis
chemical compound, drug	HaloTag JF646 ligand	PMID: 25599551		A generous gift from Luke Lavis
chemical compound, drug	SNAPtag JF646 ligand	PMID: 25599551		A generous gift from Luke Lavis
chemical compound, drug	SNAPtag JF549 ligand	PMID: 25599551		A generous gift from Luke Lavis
chemical compound, drug	HaloTag TMR ligand	Promega	Cat# G8251	
chemical compound, drug	DAPI 4',6-Diamidine-2'-phenylindole dihydrochloride	Sigma Aldrich	Cat# 10236276001	
chemical compound, drug	Hygromycin B, 50 mg/ml Solution	Thermo Fisher	Cat# 10687-010	
commercial assay or kit	Click-iT RNA Alexa Fluor 594 Imaging Kit	Thermo Fisher	Cat# C10330	
commercial assay or kit	Maxima H Minus Reverse Transcriptase (200 U/µL)	Thermo Fisher	Cat# EP0752	
commercial assay or kit	Kapa2G HotStart PCR Kit 500U	Kapa Biosystems	Cat# kk5517	
commercial assay or kit	HiScribe™ T7 Quick High Yield RNA Synthesis Kit	New England Biolabs	Cat# E2050S	
commercial assay or kit	Nextera DNA Library Preparation K	Illumina	Cat# FC-121-1030	

commercial assay or kit	Stellaris® RNA FISH Wash Buffer A	Biosearch Technologies	Cat# SMF-WA1-60	
commercial assay or kit	Stellaris® RNA FISH Wash Buffer	Biosearch Technologies	Cat# SMF-WB1-20	
commercial assay or kit	Stellaris® RNA FISH Hybridization Buffer	Biosearch Technologies	Cat# SMF-HB1-10	
commercial assay or kit	Cell Line Nucleofector Kit V	Lonza	Cat # VVCA-1003	
commercial assay or kit	Fugene 6	Promega	Cat# E2692	
cell line (Homo sapiens)	Halo-RPB1	PMID: 30127355	U2OS HaloTag-RPB1	U2OS (15 y/o female osteosarcoma, RRID: CVCL_0042) expressing HaloTag-RPB1(N792D) selected for using alpha-amanitin.
cell line (Homo sapiens)	Halo-RPB1(25R)	PMID: 30127355	U2OS HaloTag-RPB1(25R)	U2OS (15 y/o female osteosarcoma, RRID: CVCL_0042) expressing HaloTag-RPB1(N792D) with the c-terminal domain truncated to 25 heptad repeats instead of the wildtype 52, selected for using alpha-amanitin.
cell line (Homo sapiens)	Halo-RPB1(70R)	PMID: 30127355	U2OS HaloTag-RPB1(70R)	U2OS (15 y/o female osteosarcoma, RRID: CVCL_0042) expressing HaloTag-RPB1(N792D) with the c-terminal domain extended to include 20 additional heptad repeats in addition to the WT 52, selected for using alpha-amanitin.
cell line (Homo sapiens)	Dendra2-RPB1	PMID: 23828889	U2OS Dendra2-RPB1	U2OS (15 y/o female osteosarcoma, RRID: CVCL_0042) expressing Dendra2-RPB1(N792D) selected for using alpha-amanitin.
cell line (Homo sapiens)	SNAP-RPB1	This paper	U2OS SNAPtag-RPB1	U2OS (15 y/o female osteosarcoma, RRID: CVCL_0042) expressing SNAPtag-RPB1(N792D) selected for using alpha-amanitin.
cell line (Homo sapiens)	Halo-TetR	This paper	U2OS SNAPtag-RPB1, HaloTag-TetR	U2OS (15 y/o female osteosarcoma, RRID: CVCL_0042) expressing HaloTag-RPB1(N792D) selected for using alpha-amanitin, further expressing HaloTag-TetR and selected for with Hygromycin
cell line (Homo sapiens)	Halo-LacI	This paper	U2OS SNAPtag-RPB1, HaloTag-LacI	U2OS (15 y/o female osteosarcoma, RRID: CVCL_0042) expressing HaloTag-RPB1(N792D) selected for using alpha-amanitin, further expressing HaloTag-LacI and selected for with Hygromycin
cell line (Homo sapiens)	Halo-NLS	This paper	U2OS SNAPtag-RPB1, HaloTag-NLS	U2OS (15 y/o female osteosarcoma, RRID: CVCL_0042) expressing HaloTag-RPB1(N792D) selected for using alpha-amanitin, further expressing HaloTag-fused to three copies of the SV40 nuclear localization signal and selected for with Hygromycin
cell line (Homo sapiens)	Halo-EWS	This paper	U2OS SNAPtag-RPB1, HaloTag-EWS	U2OS (15 y/o female osteosarcoma, RRID: CVCL_0042) expressing HaloTag-RPB1(N792D) selected for using alpha-amanitin, further expressing HaloTag-EWS(AA 47-266)-NLS and selected for with Hygromycin
cell line (Homo sapiens)	Halo-FUS	This paper	U2OS SNAPtag-RPB1, HaloTag-FUS	U2OS (15 y/o female osteosarcoma, RRID: CVCL_0042) expressing HaloTag-RPB1(N792D)

				selected for using alpha-amanitin, further expressing HaloTag-FUS(AA 2-214)-NLS and selected for with Hygromycin
cell line (Homo sapiens)	Halo-TAF15	This paper	U2OS SNAPtag-RPB1, HaloTag-TAF15	U2OS (15 y/o female osteosarcoma, RRID: CVCL_0042) expressing HaloTag-RPB1(N792D) selected for using alpha-amanitin, further expressing HaloTag-TAF15(AA 2-205)-NLS and selected for with Hygromycin
cell line (Homo sapiens)	H2B-SNAP-Halo	PMID: 29300163	U2OS Histone H2B-SNAPtag-HaloTag	U2OS (15 y/o female osteosarcoma, RRID: CVCL_0042) expressing Histone H2B-SNAPtag-HaloTag and maintained in selection with G418
cell line (Cercopithecus aethiops)	Vero	ATCC	ATCC CCL-81; RRID:CVCL_0059	
cell line (Cercopithecus aethiops)	V27	PMID: 2157053	V27	Vero cells stable expressing ICP27 under selection of G418. A generous gift from Septhen Rice.
sequence-based reagent	Common DNA FISH forward primer: 5'-GACACGTGATCCGCG ATACGATGAAAGCGC GACGTCAGGTCGGCC-3'	Integrated DNA Technologies	N/A	
sequence-based reagent	Common DNA FISH forward primer: 5'-GACACGTGATCCGCG ATACGATGAAAGCGC GACGTCAGGTCGGCC-3'	Integrated DNA Technologies	N/A	
sequence-based reagent	Common DNA FISH reverse primer: 5'-CTCGCTAATACGACTC ACTATAGCCGGCTCCA GCGG -3'	Integrated DNA Technologies	N/A	
sequence-based reagent	Alexa Fluor 647-labeled RT primer: 5'-TCGCGCTTTCATCGTA TCGCGGATCACGTGTC -Alexa647-3'	Integrated DNA Technologies	N/A	
sequence-based reagent	Alexa Fluor 555-labeled RT primer: 5'-TCGCGCTTTCATCGTA TCGCGGATCACGTGTC -Alexa555-3'	Integrated DNA Technologies	N/A	
recombinant DNA reagent	pSNAP-RPB1(N792D) (plasmid)	This paper		RPB1 carrying N792D mutation for alpha-amanitin resistence inserted downstream of SNAPtag with the TEV protease sequence as a linker reagion.
recombinant DNA reagent	pHalo-TetR (plasmid)	This paper		The Tet repressor inserted downstream of HaloTag with the TEV proease site as a short linker.
recombinant DNA reagent	pHalo-LacI (plasmid)	This paper		The Lac repressor inserted downstream of HaloTag with the TEV proease site as a short linker and a single SV40 NLS at the c-terminus.
recombinant DNA reagent	pHaloTag-3xNLS (plasmid)	PMID: 28467304		

recombinant DNA reagent	pHalo-TEV-EWS LC-NLS (plasmid)	PMID: 29930090		
recombinant DNA reagent	pHalo-TEV-FUS LC-NLS (plasmid)	PMID: 29930090		
recombinant DNA reagent	pHalo-TEV-Taf15 LC-NLS (plasmid)	PMID: 29930090		
software, algorithm	Custom implementation of Spot-On and graphical analysis	PMID: 29300163; this paper	Spot-On	The source code is freely available at https://gitlab.com/dmcswiggen/mcswiggen_et_al_2019
software, algorithm	Matlab versions 2014b, 2017a	Mathworks	2014b, 2017a	
software, algorithm	IUPred 2A	PMID: 15769473; 15955779	IUPred	This tool is available at: https://iupred2a.elte.hu/download
software, algorithm	Bowtie2	PMID: 22388286	Bowtie	This tool is availabe at: http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
software, algorithm	SamTools	PMID: 19505943	SamTools	This tool is available at: http://samtools.sourceforge.net
software, algorithm	deepTools2	PMID: 19505943	deepTools	This tool is available at: https://deeptools.readthedocs.io/en/develop/
software, algorithm	Integrative Genomics Viewer 2.4.4	PMID: 21221095	IGV	This tool is available at: https://software.broadinstitute.org/software/igv/ReleaseNotes/2.4.x
software, algorithm	R version 3.5.1	R project	R	
software, algorithm	ADS R package	DOI: 10.18637/jss.v063.i06	ADS R package	This tool is available at: https://cran.r-project.org/web/packages/ads/index.html
software, algorithm	vbSPT	PMID: 23396281	vbSPT	This tool is available at http://vbspt.sourceforge.net
software, algorithm	Adobe Illustrator CC2017	Adobe Inc		
software, algorithm	Prism 7	GraphPad		

Tissue Culture

Human U2OS cells (female, 15 yr old, osteosarcoma; STR verified) were cultured at 37°C and 5% CO2 in 1 g/L glucose DMEM supplemented with 10% Fetal Bovine Serum and 10 U/mL Penicillin-Streptomycin, and we subcultivated at a ratio of 1:3 – 1:6 every two to four days. Stable cell lines expressing the exogenous gene product α-amanitin resistant HaloTag-RPB1(N792D), SNAPf-RPB1(N792D) or Dendra2-RPB1(N792D) were generated using Fugene 6 (Promega) following the manufacturer's protocol, and selection with 2 μg/mL α-amanitin. Stable colonies were pooled and maintained under selection with 1 μg/mL α-amanitin to ensure complete replacement of the endogenous RPB1 pool, as described previously (Boehning et al., 2018a; Cisse et al., 2013). Cells co-expressing SNAPf-RPB1 and Halo-TetR were generated using the previously described SNAP-RPB1 cell line, and transfecting with TetR-HaloTag and a linearized Hygromycin resistance marker using Fugene 6 following the manufacturer's protocol. Cells were selected and maintained with 100 μg/mL Hygromycin B. Fluorescent cells were selected by

labeling the TetR-Halo with 500 nM JF_{549} and using Fluorescence Activated Cell Sorting to identify and keep the fluorescent clones.

Vero cells (*Cercopithecus aethiops* kidney cells; STR verified), were cultured for the growth and propagation of HSV1. Vero cells were cultured at 337°C and 5% CO2 in 4.5 g/L glucose DMEM supplemented with 10% Fetal Bovine Serum and 10 U/mL Penicillin-Streptomycin. Cells were subcultivated at a ratio of 1:3 – 1:8 every two to four days.

Virus infection

HSV1 Strain KOS was a generous gift from James Goodrich and Jennifer Kugel (Abrisch et al., 2015). UL2/50 was a generous gift from Neal DeLuca (Dembowski and DeLuca, 2015). All virus strains were propagated in Vero cells as previously described (Blaho et al., 2005). Briefly, cells were infected by incubation at an MOI ~ 0.01 in Medium 199 (Thermo) for one hour. 36-48 hpi, cells were harvested by freeze-thawing, pelleted, and sonicated briefly, and then centrifuged to clear large cellular debris. Because we were interested in the early events in infection, approximate titers were first determined by plaque formation assay in Vero cells (Blaho et al., 2005). More accurate MOI were determined by infecting U2OS cells plated on coverslips with the same protocol as would be using for imaging experiments. Cells were washed once with PBS, and then 100 µL of complete medium containing 1:10 – $1:10^5$ dilutions of harvested virus were added dropwise onto the coverslip to form a single meniscus on the coverslip. Infection was allowed to proceed for 15 minutes at 37 °C. Samples were then washed once with PBS and returned to culturing medium and incubated for 8 hours before fixation. To measure the MOI, immunofluorescence for the expression of ICP4 using an anti-ICP4 primary antibody (Abcam), and counting the number of infected versus uninfected cells. MOI was then calculated, assuming a Poisson distribution of infection events, as $P(k_{inf}) = \frac{MOI^{k_{inf}} e^{-MOI}}{k_{inf}!}$, where k_{inf} is the number of infection events per cell. When counting the uninfected cells, this simplifies to $MOI = -\ln(f_{uninfected})$. All experiments were performed from the same initial viral stock, with care taken so that each experiment was done with virus experiencing the same total number of freeze/thaw cycles to ensure as much consistency as possible.

Transient Transfection

For experiments where transiently transfected cells were also infected with HSV1, nucleofection was used to achieve more consistent infection across the coverslip. $1x10^6$ cells were trypsinized and resuspended in Kit V buffer plus supplement (Lonza) with 500 ng plasmid, and nucleofected using program X-001, per the manufacturer's instructions. Cells were plated on coverslips and allowed to recover for 48 hours prior to HSV1 infection.

Live cell imaging

Cells were plated on plasma-cleaned 25 mm circular No. 1.5H cover glasses (Marienfeld High-Precision 0117650) and allowed to adhere overnight. For experiments with HaloTag-expressign cells, cells were incubated with 5 – 500 nM fluorescent dye (e.g. JF_{549}) conjugated with the

HaloTag ligand for 15 minutes in complete medium. Cells were washed once with PBS, and the media replaced with imaging media (Fluorobrite media (Invitrogen) supplemented with 10% FBS and 10 U/mL Penicillin-Streptomycin). For experiments with cells expressing SNAP-RPB1, cells were labeled with 250 nM fluorescent dye (e.g. JF_{549}) conjugated with the cpSNAP ligand for 30 minutes. After labeling, cells were washed for 30 minutes in complete medium. Prior to imaging, coverslips were mounted in an Attofluor Cell Chamber filled with 1 mL of imaging medium. Cells were maintained at 37 °C and 5 % CO2 for the duration of the experiment. For long term time course imaging experiments, cells were plated in 35mm No. 1.5 glass-bottomed imaging dishes (MatTek), infected with HSV1 at an MOI of ~1, and labeled with $JF_{549,}$ and finally the media exchanged for imaging media before placing in a pre-warmed Biostation (Nikon). At 3 hours post infection, infected cells were identified and imaged were taken every 30 seconds for 5 hours. For phase images, cells were plated and labeled as above, and imaged on a custom-built widefield microscope with a SLIM optics module (PhiOptics) placed in the light path directly before the camera.

Fluorescence Recovery After Photobleaching (FRAP)

FRAP experiments were performed as previously described, with modifications. HaloTag-RPB1 cells labeled with 500 nM JF_{549} were imaged on an inverted Zeiss LSM 710 AxioObserver confocal microscope with an environment chamber to allow incubation at 37°C and 5% CO_2. JF_{549} was excited with a 561 nm laser, and the microscope was controlled with Zeiss Zen software. Images were acquired with a 63x Oil immersion objective with a 3x optical zoom. 1200 total frames were acquired at a rate of 250 msec per frame (4 Hz). Between frames 15 and 16, an 11-pixel (0.956 μm) circle was bleached, either in the center of a RC, or in a region of the nucleus far from the nuclear periphery or nucleoli.

FRAP movies were analyzed as previously described (Hansen et al., 2017). Briefly, the center of the bleach spot was identified manually, and the nuclear periphery segmented using intensity thresholding that decays exponentially to account for photobleaching across the time of acquisition. We measured the intensity in the bleach spot using a circle with a 10 pixel diameter, to make the measurement more robust to cell movement. The normalized FRAP values were calculated by first internally normalizing the signal to the intensity of the whole nucleus to account for photobleaching, then normalizing to the mean value of the spot in the first 15 frames. We corrected for drift by manually updating a drift-correction vector with the stop drift every ~40 frames. FRAP values from individual cells were averaged across replicates to generate a mean recovery curve, and the error displayed is the standard error of the mean.

Fluorescence Loss in Photobleaching (FLIP)

FLIP experiments were performed on the same microscope described above for FRAP. Rather than bleach an 11-pixel spot a single time, in FLIP the spot is bleached with a 561 nm laser (or in the case of Dendra2, photoconverted with a 405 nm laser) between each acquisition frame. Movies were collected for 1000 frames at 250 msec per frame (4 Hz), or 1 frame per second (1 Hz) for Dendra2.

FLIP movies were analyzed using the same core Matlab code as the FRAP data, except that fluorescence intensities from another 10-pixel circle were recorded to measure the loss of fluorescence elsewhere in the nucleus. This analysis spot was chosen to be well away from the bleach spot, either at a neighboring RC in infected samples or somewhere else in the nucleoplasm far away from both the nuclear periphery and nucleoli. Instead of internally correcting for photobleaching, photobleaching correction was based on an exponential decay function empirically determined to be at a rate of $e^{-0.09}$ per frame. FLIP data from multiple cells were averaged together to determine the mean and standard error for a given condition.

RNA Fluorescence In Situ Hybridization (FISH) and immunofluorescence (IF)

RNA FISH was used to measure the transcription output for a given RC. To ensure we were measuring nascent transcription, we chose to tile the intronic region of RL2, one of the few HSV1 transcripts with an intron. The 25 oligonucleotide probes were synthesized conjugated with a Cal Fluor 610 dye (Biosearch Technologies; for a full list of oligo sequences see Supplementary File 1). FISH was performed based on the manufacturer's protocol. Briefly, cells were plated on 18 mm No. 1.5 coverslips (Marienfield) and infected. At the desired time point, cells were fixed in 4% Paraformaldehyde diluted in PBS for 10 minutes. After two washes with PBS, coverslips were covered with 70% v/v ethanol and incubated at -20 °C for 1 hour up to 1 week.

For hybridizations, coverslips were removed from ethanol and washed in freshly-prepared Wash Buffer A (2 volumes 5x Wash Buffer A, 1 volume formamide, 7 volumes H_2O) (Bioseach Technologies). Hybridization buffer (10% v/v Dextran Sulfate, 300 mM Sodium Chloride, 30 mM Sodium Citrate, 400, 10% Formamide v/v, and 12.5 nM pooled fluorescent probes) was prepared freshly before each hybridization. A hybridization chamber was prepared with moistened paper towels laid in a 15cm tissue culture plate. A single sheet of Parafilm was laid over the moistened paper towel. 50 µL of hybridization buffer was pipetted onto the parafilm, and a coverslip inverted into the hybridization buffer. The chamber was sealed with parafilm and placed in a dry 37 °C oven for 4-16 hours. After hybridization, coverslips were placed back into a 12-well plate containing 1 mL Wash Buffer A and incubated twice for 20 minutes in a dry oven at 37 °C, with the second wash containing 300 nM DAPI. In a final wash step, cells were washed in Wash Buffer B (Biosearch Technologies). Coverslips were mounted on glass microscope slides in Vectashield mounting medium (Vector Laboratories) and the edges sealed with clear nail polish (Electron Microscopy Sciences). For experiments with combined immunofluorescence and FISH, primary antibody was added to the hybridization buffer at a concentration of 2 µg/mL. An additional wash step with Wash Buffer A containing 1µg/mL anti-mouse polyclonal antibody conjugated to AlexaFluor 647 was performed before DAPI staining and incubated at 37°C for 20 minutes.

Samples were imaged on a custom-built epifluorescence Nikon Eclipse microscope equipped with piezoelectric stage control and EMCCD camera (Andor), as well as custom-built filter sets corresponding to the wavelength of dye used. All samples were imaged the same day after hybridaztion and/or incubation with secondary antibody, and all samples to be quantitatively

compared across coverslips were imaged on the same day using exactly the same illumination and acquisition settings to minimize coverslip-to-coverslip variation.

Single Particle Tracking (spaSPT)

Single particle tracking experiments were carried out as previously described (Hansen et al., 2017), but are described here in brief. After overnight growth, U2OS cells expressing Halo-RPB1 were labeled with 50 nM each of JF_{549} and PA-JF_{646}. Single molecules imaging was performed on a custom-built Nikon Ti microscope fitted with a 100x/NA 1.49 oil-immersion TIRF objective, motorized mirror are to allow HiLo illumination of the sample, Perfect Focus System, and two aligned EM-CCD cameras. Samples were illuminated using 405- nm (140 mW, OBIS coherent), 561-nm (1 W, genesis coherent), and 633-nm (1 W, genesis coherent) lasers, which were focused onto the back pupil plane of the objective via fiber and multi-notch dichromatic mirror (405-nm/488-nm/561-nm/633-nm quad-band; Semrock, NF03-405/488/532/635E-25). Excitation intensity and pulse width were controlled through an acousto-optic transmission filter (AOTF nC-VIS-TN, AA Opto-Electronic) triggered using the camera's TTL exposure output signal. Fluorescence emissions were filtered with a single bandpass filter in front of the camera (Semrock 676/37 nm bandpass filter). All of the components of the microscope, camera, and other hardware were controlled through NIS-Elements software (Nikon).

For all spaSPT experiments, frames were acquired at a rate of 7.5 msec per frame (7 msec integration time plus 0.447 msec dead time). In order to obtain both the population-level distribution of the molecules for masking and the single trajectories, we used the following illumination scheme: First 100 frames with 561 nm light and continuous illumination were collected; then 20,000 frames with 633 nm light at 1-2 msec pulses per frame and 0.4 msec pulses of 405 nm light during the camera dead time; then 100 frames with 561 nm light and continuous illumination were collected. 405 nm illumination was optimized to achieve a mean density of ~0.5 localizations per camera frame, a density sufficiently low to unambiguously identify trajectories, even in dense regions like RCs. Data were collected over multiple courses of infection and two to four separate days for each condition in order to ensure a sufficiently large sample size.

ATAC-seq sample preparation

ATAC-seq experiments were performed as previously described (Buenrostro et al., 2013). Briefly, 100,000 U2OS cells stably expressing HaloTag-RPB1 were plated and allowed to grow overnight. The following day, cells were infected as described above, and incubated either in complete medium, or complete medium supplemented with 300 µg/mL phosphonoacetic acid (PAA). Infections were timed such that all cells were harvested at once. All of the infected cell lines were then trypsinized, and 100,000 cells were transferred to separate eppendorff tubes. Cells were briefly centrifuged at 500 xg for 5 minutes at 4°C, and the supernatant discarded. After one wash with ice-cold PBS and another 5-minute spin at 500 xg and 4°C, cells were resuspended directly in tagmentation buffer (25µL 2x Buffer TD, 22.5 µL nuclease free water, 2.5µL Tn5 (Illumina)) and incubated for 30 minutes at 37 °C. DNA extraction and amplification with barcodes were performed as previously described, with 10-16 total cycles amplification. Barcoded samples

were pooled in equimolar amounts and sequenced using a full flow-cell of an Illumina Hi-Seq 2500 per replicate. Three replicates were performed, though the first replicate was deemed to have been over-amplified during the PCR step, and thus was omitted from the analysis.

Oligopaint on infected cells

For DNA FISH experiments, custom pools of fluorescently labeled DNA oligos were generated using previously published protocols (Boettiger et al., 2016). Briefly, oligo sequences tiling a 10,016 bp region in the Unique Long arm (JQ673480 position 56,985 to 66,999) and a 7703 bp region in the Unique Short arm (JQ673480 position 133,305 to 141,007) were manually curated using oligo BLAST (NCBI) against the HSV1 and human genomes with the following settings, following guidelines for Tm, GC-content, and length from previous Oligopaint protocols (Boettiger et al., 2016). Individual oligos were purchased commercially (the sequences for these oligos can be found in Supplementary file 2 and pooled. PCR was used to introduce a common T7 promoter on the 3' end of the final probe sequence, then the PCR products were gel purified before *in vitro* transcription to generate ssRNA complimentary to the hybridization sequence. Finally, the entire RNA pool was reverse transcribed in a single reaction using Maxima RT (ThermoFisher) using either AlexaFluor-647 or AlexaFluor-555 5'-labeled oligos as the reverse transcription primer. After acid hydrolysis to remove the RNA, oligos were purified using high binding capacity oligo cleanup columns (Zymo) and resuspended in TE.

Cells were plated on 18 mm coverslips and infected as described above. Infected was allowed to progress for between 3 and 8 hours in the presence or absence of phosphonoacetic acid, then fixed with 4% paraformaldehyde for 15 minutes. Coverslips were washed twice with PBS, then incubated with 100mM Glycine in PBS for 10 minutes. Samples were permeabilized for 15 minutes with 0.5% Triton-X100 in PBS, then washed twice with PBS. After permeabilization, samples were treated with 100 mM HCl for 5 minutes, then washed twice with PBS. Prior to hybridization, samples were washed twice with 2X SSC (300 mM NaCl, 30 mM Sodium Citrate), and then incubated at 42 °C for 45 minutes in 2X SSC with 50% v/v Formamide. Coverslips were inverted onto a slide containing 25 µL hybridization buffer (300 mM NaCl, 30 mM Sodium Citrate, 20% w/v Dextran Sulfate, 50% v/v Formamide, and 75 pmol of fluorescently-labeled oligos) and sealed with rubber cement. Samples were denatured at 78 °C on an inverted heat block for 3 minutes, then incubated in a humidified chamber at 42°C for 16 hours. Samples were then removed from the glass slides and washed twice to 60 °C with pre-warmed 2x SSC for 15 minutes, then washed twice with 0.4x SSC at room temperature for 15 minutes. Finally, coverslips were mounted on glass slides with Vectashield mounting medium.

DNA FISH samples were imaged on the same microscope as described above for immunofluorescence and RNA FISH. Z-stack images were collected from all the way below the focal plane to all the way above the focal plane, with a step size of 100 nm. All samples were imaged on the same day using the same illumination and acquisition settings to minimize coverslip to coverslip differences.

PALM of Pol II in RCs

For PALM experiments to precisely localize Pol II molecules within RCs, cells were labeled with 500 nM PA-JF_{549}, and then infected as described above. Cells were fixed in 4% Paraformaldehyde in PBS, washed twice with PBS. Fluorescent 100 nm and 200 nmTetraspek beads were mixed in a 9:1 ratio then diluted 1000-fold in PBS. 100 µL was added to each coverslip and allowed to settle for 5 minutes, followed by 5 minutes of washing while rocking. Coverslips were mounted in Attofluor Cell Chambers and covered with PALM imaging buffer (50 mM NaCl, 50 mM Tris pH 7.9, 2 mM Trolox) to reduce triplet-state blinking.

Samples were imaged on a custom-built Nikon Ti microscope equipped similarly to the microscope for single particle tracking, with some differences described here. An Adaptive Optics module (MicAO) and a removable cylindrical lens were placed in the light path ahead of the EM-CCD (Andor iXon Ultra 897) cameras in the left and right camera ports (respectively) of the microscope. Astigmatism for precise 3D localization was introduced using the Adaptive Optics system. The Adaptive Optics system was controlled through the MicAO software and calibrated on 200 nM Tetraspek beads based on the total photon yield and point spread function shape after iterative tuning of the deformable mirror. After optimization, a slight astigmatism in the vertical Zernike mode (Astigmatism 90° = 0.060) was added, and several z-stacks of 100 nM Tetraspek beads with 10 nm between slices to calibrate the PSF shape with the Z-position. 30,000 frames were acquired with the 561 nm laser line and increasing amounts of 405 nm illumination in order to keep the number of single molecules consistent across the duration of acquisition.

STORM on infected cells

For STORM experiments to visualize both RNA Polymerase II and the viral DNA, U2OS cells stably expressing Halo-RPB1 were plated on coverslips, labeled with 300 nM JF_{549}, and infected with the UL2/50 virus strain (Dembowski and DeLuca, 2015) as described above. After infection incubation with virus, cells were transferred into complete medium containing 300 µg/mL PAA for two hours to prevent replication. After two hours, cells were released from inhibition by exchanging the culture medium with complete medium containing 2.5 µM 5-Ethynyldeoxyuridine for 4 hours. Cells were fixed with 4% Paraformaldehyde in PBS for 10 minutes, then permeabilized with 0.5% Triton X100 in PBS for 10 minutes. Copper(1)-catalyzed alkyne-azide cycloaddition was performed with the ClickIT imaging kit following the manufacturer's protocol (Thermo). Coverslips were mounted in Attofluor Cell Chambers and covered with freshly-made STORM buffer (50 mM NaCl, 50 mM Tris pH 7.9, 10% D-glucose, 10 mM DTT, 700 µg/mL Glucose Oxidase (Sigma), and 4 µg/mL catalase). STORM experiments were performed on the same microscope described for PALM.

IUPred disorder prediction

Disorder predictions were preformed using a custom built python script to implement the IUPred intrinsic disorder prediction program (Dosztanyi et al., 2005; Dosztányi et al., 2005). Specific protein sequences were placed in a table and this was fed into the script. All protein sequences were downloaded from the reference organism at uniport.org. The resulting traces were smoothed by a rolling mean of 8 residues to remove noise and prevent single low-energy residues

from splitting single large IDRs into multiple apparent IDRs. Contiguous substrings of residues with centered-mean IUPred disorder likelihood greater than 0.55 were annotated as "disordered regions" (Fig. 1E), and those contiguous regions larger than 10 amino acids were included in the calculation of “fraction IDR”.

spaSPT data processing

SPT data sets were processed in 4 general steps using a custom-written Matlab (Mathworks): 1) Masks for RCs were annotated manually, 2) the masks were corrected for drift throughout the sample acquisition, 3) particles were localized and trajectories constructed, and 4) trajectories were sorted as “inside” compartments or “outside”.

First, the 100 frames at the beginning and the end of each movie were separately extracted and a maximum-intensity projection used to generate “before” and “after” images of the cell or cells in the field of view. These images would be used to correct for movement of the cell as well as the individual RCs. For each cell, the nucleus was annotated in the “before” image, and then again in the “after” image. We assumed that the cell movement over the ~4 minutes of acquisition was approximately linear and calculated the drift-corrected nuclear boundary for every frame in the stack of SPT images. The same procedure was applied to each of the replication compartments. Particle localization and tracking were implemented based on an adapted version of the Multiple Target Tracking (MTT) algorithm, available at https://gitlab.com/tjian-darzacq-lab/SPT_LocAndTrack. In the first step, particles were identified with the following input parameters: Window = 9 px; Error Rate = $10^{-6.25}$; Deflation Loops = 0. Following detection, a mask generated from the drift-corrected nuclear boundary was applied to discard any detections not within the nucleus. Trajectories were reconstructed with the following parameters: Dmax = 10 μm^2/sec; Search exponent factor = 1.2; Max number of competitors = 3; Number of gaps allowed = 1.

Finally, after trajectories have been reconstructed, they were sorted as “inside” RCs or “outside”. To minimize the potential for bias in calling trajectories inside of compartments, we only required a single localization in a trajectory to fall within a compartment for that trajectory to be labeled as “inside”. As is discussed in the main text, we tested this sorting strategy for implicit bias by computationally generating mock RCs in uninfected or infected samples (Figure S3). To do this, all of the annotations for RCs from the infected samples (n = 817), as well as the distribution of number of RCs per infected cell, were saved in a separate library. We then took the uninfected cells and, in a similar process as described above, annotated the nuclear boundary and nucleoli. We then randomly sampled from distribution of RCs per cell a number of RCs to place in the nucleus, and then from the library of annotations randomly chose these RCs and placed them in the nucleus by trial-and-error until all of the chosen RCs could be placed in the nucleus without overlapping with each other, a nucleolus, or the nuclear boundary (Figure S3A). The SPT data were then analyzed as above—drift-correction, followed by localization, building of trajectories, and sorting into compartments—using the exact same parameters. We also followed this same

procedure of randomly choosing and placing artificial RCs in infected cells, this time avoiding previously annotated RCs instead of nucleoli (Figure S3B).

Two-state kinetic modeling using Spot-On

We employed the Matlab version of Spot-On (available at https://spoton.berkeley.edu) in our analysis and embedded this code into a custom-written Matlab routine. All data for a given condition were merged, and histograms of displacements were generated for between 1 and 7 Δt. These histograms were fitted to a two-state kinetic model which assumes one immobile population and one freely diffusing population: Localization Error = 45 nm; D_{free} = [0.5 μm^2/sec, 25 μm^2/sec]; D_{bound} = [0.0001 μm^2/sec, 0.08 μm^2/sec]; Fraction Bound = [0, 1]; UseWeights = 1; UseAllTraj = 0; JumpsToConsider = 4; TimePoints = 7; dZ = 0.700. Trajectory CDF data were fit to a two-state model as first outlined by Mazza and colleagues, then expanded with implementation in Hansen and colleagues.

Spot-On has been shown to robustly estimate all of the fitted parameters, provided there is sufficient data—at a minimum 1000 trajectories for a 2-state fit of a model protein with diffusion characteristics similar to Pol II (50% bound, Dfree = 3.5 μm^2/sec) (Figure S2A) (Hansen et al., 2018). Because of the sparsity of the data we collected per cell, we found that we could not reliably generate single-cell statistics, particularly within RCs where the total number of trajectories per cell fell well below the 1000-trajectory threshold (Figure S2B). In order to robustly fit our data and simultaneously estimate its variability, we first calculated the number of cells we would need to confidently fit all compartments and found 15 cells to optimal (Figure S2B). We then implemented a random subsampling approach where 15 cells from a particular condition were randomly chosen and analyzed. The D_{free}, D_{bound}, and Fraction Bound were calculated iteratively for trajectories inside and outside of RCs. This random resampling was repeated 100 times, and the median values and standard deviations calculated and reported. When compared to the values that would have been obtained for taking the mean and standard deviation of the individual biological replicates, our subsampling approach agreed with these means within the measurement error (Figure S2C).

Analysis of angular distribution

Angular distribution calculations were performed using a custom written routine in Matlab, implementing a previous version of this analysis (available at https://gitlab.com/anders.sejr.hansen/anisotropy). To analyze the angular distribution of trajectories in different conditions, we started with the list of trajectories generated above, annotated as either "inside" or "outside" of RCs. A trajectory of length N will have N-2 three-localization sets that form an angle, and so we built a matrix consisting of all consecutive three-localization sets. It is crucially important that only diffusing molecules be considered in the analysis, as localization error of bound molecules would skew all of the data to be highly anisotropic. To address this, we used two criteria. First, we only applied a Hidden-Markov Model based trajectory classification approach to classify trajectories as either diffusing or bound (Persson et al., 2013), and kept only the trajectories that were annotated as diffusing. Second, we

applied a hard threshold that both translocations (1 to 2, 2 to 3) had to be a minimum of 150 nm, which ensured that we could accurately compute the angle between them. Because a particle may diffuse into or outside of the annotated region, we counted a trajectory as "inside" only if the vertex of the angle occurred within an annotated region.

ATAC-seq analysis

Sequenced reads were mapped separately to hg19 genome using Bowtie2 (Langmead and Salzberg, 2012) with the following parameters: --no-unal --local --very-sensitive-local --no-discordant --no-mixed --contain --overlap --dovetail --phred33. Reads were separately mapped to the HSV1 genome, JQ673480, using Bowtie2 with the following parameters: --no-unal --no-discordant --no-mixed --contain --overlap --dovetail --phred33. The bam files were converted to bigwig files and visualized using IGV (Robinson et al., 2011). TSS plots were generated using Deeptools suite (bamCoverage, computeMatrix, plotHeatmap tools) using UCSC TSS annotations for hg19 genome, and using a highly refined map of the gene starts in HSV1 kindly provided by Lars Dölken (University of Cambridge, to be published separately).

Analysis of Immunofluorescence, RNA, and DNA FISH

All cells were analyzed using a custom Matlab script. First, a single image for each color channel was generated by automatically identifying the focal plane of the stack, and then integrating the pixel intensity for all pixels 1 μm above and below the focal plane. Nuclei were automatically segmented, but replication compartments could not reliable by detected using simple thresholding, and so each was manually annotated. A region of the image was selected to represent the black background, and the mean pixel value of this region was subtracted from every pixel in the image. After segmentation, the pixel values for each nucleus were recorded, as well as every RC within a given nucleus, and these were used to measure the signal within the RC, as well as the fraction of signal within compared to the rest of the nucleus (immunofluorescence only).

Quantification of DNA content within RCs

DNA FISH data were compared with ATAC-seq data for the 6 hpi timepoint. Despite the fact that U2OS are hypertriploid, we based all the calculations on the DNA content of a diploid cell. As such, the values presented here likely represent an upper bound on the relative concentrations of host and HSV1 gDNA for our experiments. Volume estimates for nuclei were based on data from Monier et al., 2000; volumetric measurements for RCs were taken directly from the annotations of the DNA FISH data.

PALM Spatial Statistics

Spatial statistics were collected on cells using previously published methods (Boehning et al., 2018a). First, cell boundaries and replication compartments were annotated as for spaSPT experiments (above). Particularly for small objects like RCs, edge correction is crucial for accurate spatial point pattern statistics. Given a set of detections P, we used the estimator *f* to correct for biases generated by points near the RC boundary:

$$f(i,j,r) = \begin{cases} 0, & if\ d(i,j) > r \\ \dfrac{2\pi\, d(i,j)}{C_{in}}, & otherwise \end{cases}$$

where *d(i,j)* is the distance between points *i* and *j* for *i,j*∈P, and C_{in} is arclength of the part of the circle of *d(i,j)* centered on *i* which is inside the annotated region (Goreaud and Pélissier, 1999). We then calculated N(r), the local neighborhood density:

$$N(r) = \frac{1}{N_p} \sum_{i \in P} \sum_{i \neq j} f(i,j,r)$$

where N_p is the total number of detections within the region (Goreaud and Pélissier, 1999).

The modified L-function is compared to complete spatial randomness (CSR), a homogenous Poisson process with intensity λ, equal to the density of detections in the region of interest A. The K-Ripley function is defined as:

$$K(r) = \frac{N(r)}{\lambda}$$

(Ripley, 1977). We estimated the modified L-function given by:

$$L(r) - r = \sqrt{\frac{K(r)}{\pi}} - r$$

(Goreaud and Pélissier, 1999). For the modified L-function, a spatial distribution with CSR remains at 0 for all radii. To implement this analysis, we used a previously published python script and the ADS R package to estimate the spatial statistics (Boehning et al., 2018a; Pélissier and Goreaud, 2015). In order to estimate the error in our measurements, for each cell we performed random subsampling of the data, before annotation, to randomly select 25,000 detections 100 times, and fed these subsampled data to the R script computing the statistic. For very small radii, a high L(r)-r value is likely due to blinking and other photo-physical artifacts (Annibale et al., 2011), but at length scales larger than localization error the method becomes robust.

Data and software availability

The GEO accession number for the ATAC-seq data is: GSE117335. The SPT trajectory data are available via Zenodo at DOI:10.5281/zenodo.1313872. The software used to generate these data is available at https://gitlab.com/tjian-darzacq-lab.

Chapter Four: Measuring the performance of rhodamine-derived dyes for live-cell imaging

Abstract

The rhodamine scaffold has served as the foundation for a number of technological advanced in dye chemistry related to live-cell and single-molecule imaging. While these advances have allowed unprecedented access to biological questions through microscopy, no dye is yet perfect. Improvements to the brightness, photostability, or usability of these fluorescent dyes is highly desirable. Here we first introduce the JFX series of dyes, in which rational addition of deuterium atoms in the place of common hydrogen improve the photophysical properties of the dye. We find that these isotope effects indeed improve molecule brightness and photostability by limiting the pathways for non-radiative decay. We also discovered that these deuterated dyes label live cells more efficiently than their hydrogenated counterparts. Through these experiment, we uncover striking differences in cell permeability between different dye molecules, despite the fact that they only differ structurally by a few atoms. These results are reported with the intent of aiding the reader in choosing the proper dye for any given imaging application.

Part of this work has been performed in collaboration with Luke Lavis, Jonathan Grimm, Heejun Choi, and Thomas Binns of Janelia Research Campus. It is presented here with the permission of the authors

Introduction

Advances in fluorescence microscopy offer a suite of new tools a biologist might employ. As the last chapters have outlined, these new advances—particularly those enabling studies in live cells or organisms— are revealing new biological insights that are inaccessible through other means. With the invention of self-labeling tags (Keppler et al. 2003, 2004; Los et al. 2008) and the improvements to dye performance (Grimm et al. 2015, 2016, 2017), historically challenging assays like single particle tracking (SPT) are now more approachable than ever. These technological advances in dye chemistry, in particular, have facilitated imaging with higher spatial and temporal resolution, as well as fine control over parameters critical to the quality of the data. It is these successes in improving the quality and ease of SPT data acquisition through improved dye chemistry that served as motivation to further improve the performance of the Janelia Fluor dye family.

From the perspective of an SPT experiment, an ideal dye is one that never photobleaches; that has a perfect quantum yield and a high extinction coefficient; that labels its corresponding protein tag rapidly and with high specificity; and that is spectrally tuned to work with the available optical filters of the microscope. Of course, such an ideal dye does not exist, but improvements in any of these different areas (brightness, lifetime, labeling kinetics, or spectra) should generally result in a high quality SPT dataset.

The photophysics and initial bleaching steps of rhodamines are well understood because of their importance as biological probes and laser dyes (Zheng and Lavis 2017). Absorption of a photon excites the tetramethylrhodamine (TMR) molecule from ground state (**1**-S_0) to the first excited state (**1**-S_1) and following excitation the molecule may relax back to the ground state through a variety of processes (Figure 1A). Emission of a photon (fluorescence) is the most desirable outcome for the microscopist, but other non-radiative relaxation processes compete with fluorescence. For example, TMR is thought to relax to S_0 through a twisted internal charge transfer (TICT) process where electron transfer from the aniline nitrogen to the xanthene system gives a charge-separated species (**1**-TICT) that rapidly decays back to the ground state without emitting a photon (Grimm et al. 2015). Alternatively, the excited dye can undergo or intersystem crossing to the first triplet excited state (**1**-T_1), where it can sensitize singlet oxygen ($^1O_2$), returning to the ground state (**1**-S_0). The resulting $^1O_2$ can then react with the ground state of the dye, oxidizing the aniline nitrogen to radical cation **2**, which can undergo deprotonation to a carbon-centered radical (**3**) that ultimately results in dealkylation of the dye. Subsequent photobleaching of the dealkylated products (*e.g.*, **4**) proceeds rapidly through a still-elusive photochemical mechanism (Zheng and Lavis 2017).

Here we first describe a new family of Janelia Fluor dyes, the JFX series, which use isotope effects through deuteration to improve the performance of the rhodamine-derived Janelia Fluor dyes. We find that these deuterated JFX dyes outperform their non-deuterated counterparts in both brightness and photostability. In making these measurements in live cells, we were surprised to discover that deuteration also appears to improve the labeling kinetics of the dye, remarkable for

so seemingly small a change to the dye's composition. These results prompted us to more systematically measure the available rhodamine derivatives used for live cell imaging, and in doing so report a number of striking differences affected by subtle changes in dye structure. Our goal is to provide assays that may be used to compare dye species' performance as the list of possible dyes continues to grow, as well as to provide a rationale for choosing one over another for a given experiment.

Results

Deuteration changes photophysical properties of rhodamine-based dyes

We were curious if we could use isotope effects introduced by exchanging key hydrogens for deuterium atoms, to increase brightness and photostability of small molecule fluorophores such as **1**. Deuterated alkylamines exhibit higher ionization potentials relative to their hydrogen-containing analogs (Hull et al., 1967), suggesting that deuteration could decrease the efficiency of the TICT process and therefore increase quantum yield. This higher ionization potential could also affect the initial electron abstraction step in photobleaching to form intermediate **2** (Figure 1A). Moreover, deuteration could increase photostability as the stronger C–D bond could decrease the rate of the deprotonation step in the photobleaching pathway to form photobleaching intermediate **3**–a manifestation of the well-known kinetic isotope effect (Wiberg, 1955).

To test this hypothesis, we first synthesized the matched pair of TMR (**1**) and its deuterated analog **1-d$_{12}$** using a cross-coupling approach. We measured remarkably similar absorption maximum (λ_{max}) and fluorescence emission maximum (λ_{em}; Figure 1B) for the two dyes and we observed no change in the shape of the absorption peak (Figure 1C). Deuteration did affect the brightness of the dye, however, with **1-d$_{12}$** showing a ~20% increase in both the extinction coefficient (ε) and Φ (Figure 1B). We also observed slower rate of photobleaching for **1-d$_{12}$** (Figure 1D).

Based on this result with the parent tetramethylrhodamine (**1**) we synthesized matched pairs of rhodamine dyes with H- or D-containing *N*-alkyl groups (**5-9**, Figure 1B). The piperdine-containing dye **10** showed observed substantial increases in both ε and Φ but the inclusion of an oxygen into the cyclic amine structure to give morpholine-contining dye **11** caused a much smaller increase in brightness. The pyrrolidinyl-rhodamine **12** also showed substantial increases in ε and Φ. Interestingly, the azetidine containing rhodamine **13**, showed no improvement in ε or Φ. We previously showed that incorporation of an azetidine substituent into rhodamines could improve the brightness and photostability of the dyes, likely due to its higher ionization potential (Grimm et al., 2015). This result suggests that the azetidine and deuterium substitutions affect the same nonradiative decay pathways and that further improvements in fluorophore brightness will require modulation of other photophysical mechanisms. Deuteration of the azetidine-containing Si-rhodamine **14** also showed no effect on the brightness of the dye again suggesting that the azetidine and deuterium substitutions modulate the same photophysical pathways

Based on these spectroscopy results we then synthesized the H- and D-containing pairs of HaloTag ligands of the pyrrolidine-containing dye **7** and the azetidinyl-rhodamines **8** and **9** (**10-12** hereafter referred to as JFX_{549} and JFX_{646}, respectively; Figure 1E). Interestingly, in contrast with our initial spectroscopy experiments, we found that both the quantum yield (Φ, Figure 1F) and fluorescence lifetime (τ, Figure 1G) were significantly higher in the deuterated derivative when conjugated to the HaloTag protein, with the Si-rhodamine dye pair (**11** and **11d**) showing the more dramatic improvement. This result suggests that deuteration suppresses a protein-bound-specific mode of nonradiative decay. Finally, we evaluated the singlet oxygen ($^1O_2$) quantum yield and found that although JF_{549} and JF_{646} have measurable $^1O_2$ production, the deuterated derivatives JFX_{549} and JFX_{646} exhibit no measurable $^1O_2$ generation, suggesting that deuterium modulates either the intersystem crossing to T_1 or relaxation from T_1 (τ, Figure 1H).

Single-molecule imaging reveals unexpected differences between JF_{549} and JF_{646}

The result that deuteration of the azetidine ring caused improvements in fluorescence quantum yield and fluorescence lifetime, and the reduction of singlet oxygen production was intriguing. We thought that single molecule imaging applications would be the most likely to show benefit from these changes in photophysics, and so compared deuterated and non-deuterated analogs in living cells. Cells expressing a chimeric version of histone H2B fused to both HaloTag and SNAPtag (Figure 2A) were labeled with 10 pM of both JF_{549} and JF_{646}, or with JFX_{549} and JFX_{646}. This resulted in cells sparsely labeled enough to visualize individual molecules (Figure 2B). As expected, both deuterated dyes gave higher intensity signals compared to their non-deuterated counterparts (Figure 2C), again with the SI-containing JFX_{646} earning a greater boost in signal consistent with the in vitro results (Figures 1F-H).

Perhaps more surprisingly, however, were the significant differences between the JF_{549} and JF_{646}, or the deuterated counterparts, in labeling density. Given that these samples were labeled with the same dye concentration, and given the structural similarities between JF_{549} and JF_{646}, this was quite surprising. This result replicated in multiple subsequent experiments, and suggested that there were still important differences between these two molecules which remained to be explored.

Because our initial experiments suggested that small structural changes in the dye could affect both brightness and labeling kinetics, we devised an assay to allow us to measure different HaloTag ligands in a systematic way. We took advantage of the inherent variability in expression level of the H2B-HaloTag-SNAPtag transgene in U2OS cells; pulsing in different HaloTag-liganded fluorophores into cells that had also been labeled with excess of a SNAPtag dye of a non-overlapping color (Figure 2D). We then used flow cytometry to measure thousands of cells, and thus determine the slope of the HaloTag ligand signal relative to the signal of the SNAPtag ligand. Consistent with our expectations, increasing the length of the HaloTag fluorophore pulse results in an increase in slope (Figure 2E). Plotted all of together and normalized, the data are well described by a single exponential association model (Figure 2E). Here, the rate constant (k) considers all of the processes related to the appearance of signal, and the scaling coefficient (A) contains information about the brightness of the dyes. Just as we observed under the microscope,

JF_{549} shows labeling kinetics at least six times faster than JF_{646} (Figure 2E). Such a dramatic difference between such similar seeming ligands suggested that a more fine-grained analysis would be useful for determining the source of the differences.

Subtle structural changes have dramatic effects on cell permeability

Our flow cytometry assay allowed us to compare JF_{549} to closely analogous fluorophores like Tetramethylrhodamine (TMR), JFX_{549} (deuterated azetidine ring), rhodamine-pyrolamine (**7**, RhP; 5-member pyrolamine rings), and JFX_{554} (**7-d_{16},** deuterated pyrolamine) (Figure 3A). For each dye, the brightness values were normalized to the maximum brightness of JF_{549}, so as to determine how the other dyes compared relative to the dye we currently consider as standard. All of the JF_{549}-like molecules we tested rapidly labeled cells with a labeling half-time of less than a minute – too fast to accurately measure given the spacing between timepoints in our experiment (Figure 3A, Table 1). Consistent with previous studies, JF_{549} greatly outperforms its parent molecule, TMR, for brightness (Grimm et al., 2015). Additionally, similar to the *in vitro* experiments, other variants showed modest changes in brightness, though deuteration of either the azetidine or pyrolamine ring were brighter than their non-deuterated counterpart (Table 1).

We wondered whether the labeling time we measured was determined by the dye's permeability into the cell, or whether it was limited by the reaction rate with the HaloTag protein. To test this, cells expressing HaloTag-CTCF were labeled with an excess of JF_{549} or JFX_{549} and incubated until all of the CTCF binding sites were saturated. We then pelleted the cells, resuspended them in fresh media lacking dye, and immediately transferred them to the flow cytometer, measuring the fluorescence of the cells as a function of time. Fluorescence intensity measurements show an decrease in signal that is well fit by a single exponential decay, consistent with a single-rate process. The washout of either dye has a half-time of approximately 2 minutes, in relatively good agreement with the labeling kinetics we measured, suggesting that cell penetration is the rate-limiting factor in labeling the HaloTag protein.

We made the same brightness and kinetics comparison of JF_{646}-like dyes, such as SiTMR, JFX_{646}, SiRhP, and JFX_{650} (Figure 4). Here, the results were much more striking and more surprising. Unlike the JF_{549}-spectrum dyes, we observed marked differences in both the brightness of the deuterated far red dyes (Table 1). Deuteration of either the four-member azetidine ring or the five-member pyrolamine ring resulted in an approximately 25% increase in intensity compared to the non-deuterated. Given our brightness measurements on dye free in solution, the improved brightness upon deuteration of JF_{646} suggests non-radiative decay mechanisms unique to the photophysics of the HaloTag conjugate. This is consistent with other recent experiments showing similar results with other rhodamine-scaffolded dyes in the presence of HaloTag (Grimm et al., 2017).

Perhaps more surprising is the finding that deuteration appears to substantially improve the labeling kinetics of the dye (Figure 4, Table 1). Moreover, there is a clear difference between the azetidine- and pyrolamine-containing dyes with respect to the speed of labeling. While the pyrolamine dyes are on par with the labeling speed of the JF_{549}-family dyes—too fast to accurately

measure given the time resolution of this assay—the azeditine-containing at least five times longer despite them differing by a total of two carbon and four hydrogen (or deuterium) atoms. This result demonstrates that subtle changes to the dye scaffold can have substantial and non-intuitive consequences on the performance of the dye. More broadly, these results demonstrate that not all fluorescent dyes can be treated equally, and that ideal labeling conditions must be established for each dye species.

Photoactivatable dyes are much less permeable than their standard equivalents

Our findings that small changes in dye composition can have dramatic consequences for labeling kinetics caused us to critically reexamine some of the most common assays we use these dyes for. Perhaps the most notably, photoactivatable versions of the JF dyes were recently reported (Grimm et al., 2016), which have been used as the pivotal technology allowing the robust single particle tracking and PALM assays discussed in Chapter 3 (Hansen et al., 2017, 2018b; McSwiggen et al., 2019; Teves et al., 2016, 2018). Though these dyes differ from their non-photoactivatable counterparts by a few atoms in composition, they are initially locked as a spirocyclic diazoketone that is non-fluorescent. Only upon addition of ultraviolet light does the dye molecules uncage, forming the fluorescent phenylacetic acid (**14**) or methyl-substituted (**15**) JF_{549}-like products, along with some non-fluorescent products (**16**) (Figure 5A) (Grimm et al., 2016).

We wondered whether photoactivatable versions of the JF dyes would label cells with the same kinetics as their conventional counterparts. Because these molecules naturally exist in a dark state prior to UV light addition, our original flow cytometry scheme would be unable to measure the addition of a non-fluorescent ligand. To address this, we took advantage of the high stability of DNA-incorporated histones like Histone H2B. We performed a time course pulsing in photoactivatable versions of JF_{549}, JFX_{549}, and JF_{646}, but at the end of the labeling time, the media was replaced and the cells incubated on an LED illuminator plate equipped with 450 nm LEDs (Figire 5B). The photolabile bond of the PA dyes absorbs most strongly near 405 nm, so the 45 nm bathochromic shift reduced the efficiency of photoconversion greatly. However after approximately 10 hours of photoactivation, we were able to achieve saturation of the photoactivated dye signal (Figure 5C). We confirmed with fluorescence microscopy that the HaloTag signal was still properly localized to the nucleus and that the photoconversion step did not have obvious deleterious consequences on the health of the cells (Figure 5D-F).

Shockingly, the photoactivated dyes all showed at least two orders of magnitude slower labeling kinetics than the equivalent conventional dyes (Fig 5G-I). Indeed, the rate of labeling was so much lower that, even at three hours it was unclear whether we had reached a labeling plateau, making calculation of an exact rate impossible with this data.

All of the rhodamine dyes exist in an equilibrium between a zwitterionic form, which is fluorescent, and a lactone that brakes the π-electron conjugation thereby rendering the molecule dark in this state. Recent work from the Lavis lab suggests that the equilibrium between these two states may affect the membrane permeability of the dye (Grimm et al., 2017; Zheng et al., 2019).

Given that the photoactivatable compounds are essentially rhodamine dyes locked into a lactone configuration (compound **13**), we hypothesized that it was this fluorophore geometry that was responsible for the dramatic increase in labeling time. To test this hypothesis, we photoactivated each of the dyes using a strong 405 nm laser prior to labeling cells, which should predominantly generate compounds **14** and **15.** We then followed the same labeling protocol as with the photoactivatable dyes. Interestingly, pre-photoactivation of the dye did not greatly change the overall kinetics of labeling, showing still significantly slower kinetics than the conventional dyes (Figure 5G-I).

Lastly, we asked whether it was cell permeation or interaction with HaloTag protein that was impaired by the photoactivatable dyes. We measured the rate at which pre-photoactivated PA-JF_{549} washed out of cells as a means of measuring the dye's cell permeability. Unlike conventional JF_{549} or JFX_{549}, PA-JF_{549} washed out much slower (Fig 5J), suggesting that this compound exists the cell much more slowly. These results seem to suggest that the lactone-to-zwitterion equilibrium may contribute to the rate limiting step in labeling, though they do not point to a clear alternative to explain why the photoactivatable dyes appear to penetrate the cell so inefficiently despitse quite modest structural difference.

SNAPtag dyes have lower permeability than HaloTag

We wanted to extend our study to dyes outside of the HaloTag ligand family. Other self-labeling proteins like SNAPtag and CLIPtag have been suggested to also work for live cell imaging (Gautier et al., 2008; Keppler et al., 2004; Sun et al., 2011), but anecdotal evidence suggested that these SNAPtag-dye conjugates do not perform as well as HaloTag. To test whether SNAPtag dyes truly label their cognate proteins slower than HaloTag, we performed the same time course experiments described for HaloTag, but with the roles of the dyes reversed: An excess of HaloTag dye in a different color channel was incubated with the H2B-HaloTag-SNAPtag cells throughout the experiment, while SNAPtag ligand dye was pulsed in for different lengths of time. We compared the rate of Halo-JF_{549} with a benzylguanine JF_{549} (bgSNAP-JF_{549}) ligand and a chloropyrimidine JF_{549} (cpSNAP-JF_{549}) ligand. The former bgSNAP substrate is a direct mimic of the enzymes natural substrate (Keppler et al., 2003, 2004), whereas the cpSNAP ligand was developed more recently as a means to improve cell permeability (Cole, 2013).

Consistent with recently reported findings comparing between the two self-labeling systems (Erdmann et al., 2019), we observed a significant difference in the labeling kinetics and overall labeling efficiency of the two tags. Whereas Halo-JF_{549} can label cells in a matter of minutes, the half-time of labeling for bgSNAP-JF_{549} was approximately an hour, and even the more cell permeable cpSNAP-JF_{549} ligand had a labeling half-time of 16 minutes—at least an order of magnitude longer than the Halo-JF_{549} ligand (Figure 6A). What's more, even after reaching a plateau the total fluorescence intensity was only a fraction of the brightness (4% and 60% as bright for the BG and CP ligands, respectively) compared to the saturated HaloTag signal. Given that this is a ratiometric measurement between the SNAPtag signal and the HaloTag signal for a given set of dyes, further experiments are needed to more accurately measure differences across ligands.

As with HaloTag, it is possible that the delayed labeling kinetics of SNAPtag derive from slower reactrion kinetics with the SNAPtag protein, or because of poorer cell permeability. Like the photoactivatable HaloTag ligand, the cpSNAP-JF_{549} ligand washed out of cells very slowly suggesting that poor cell permeability is a potential rate-limiting step in the labeling reaction (Figure 6B). Taken together, these data lend credence to the anecdotal evidence that HaloTag is generally better to use than SNAPtag, and should be the first choice for self-labeling tag, all other considerations being considered equal.

Discussion

The invention of self-labeling protein tags has been hailed as a major advance in cellular imaging, not the least of which because of the perceived ease of choosing the dye that is best suited to the needs of the experiment. The data here suggest that, while it is still the case that one may choose different self-labeling proteins in conjunction with a wide array of organic fluorophores (or other ligands), these molecules are not all created equal. Small structural changes—just a handful of neutrons in some cases—can dramatically alter the behavior and performance of the dye in question. As a consequence, one must check to identify the optimal way to utilize the dye, to ensure that cells are labeled appropriately.

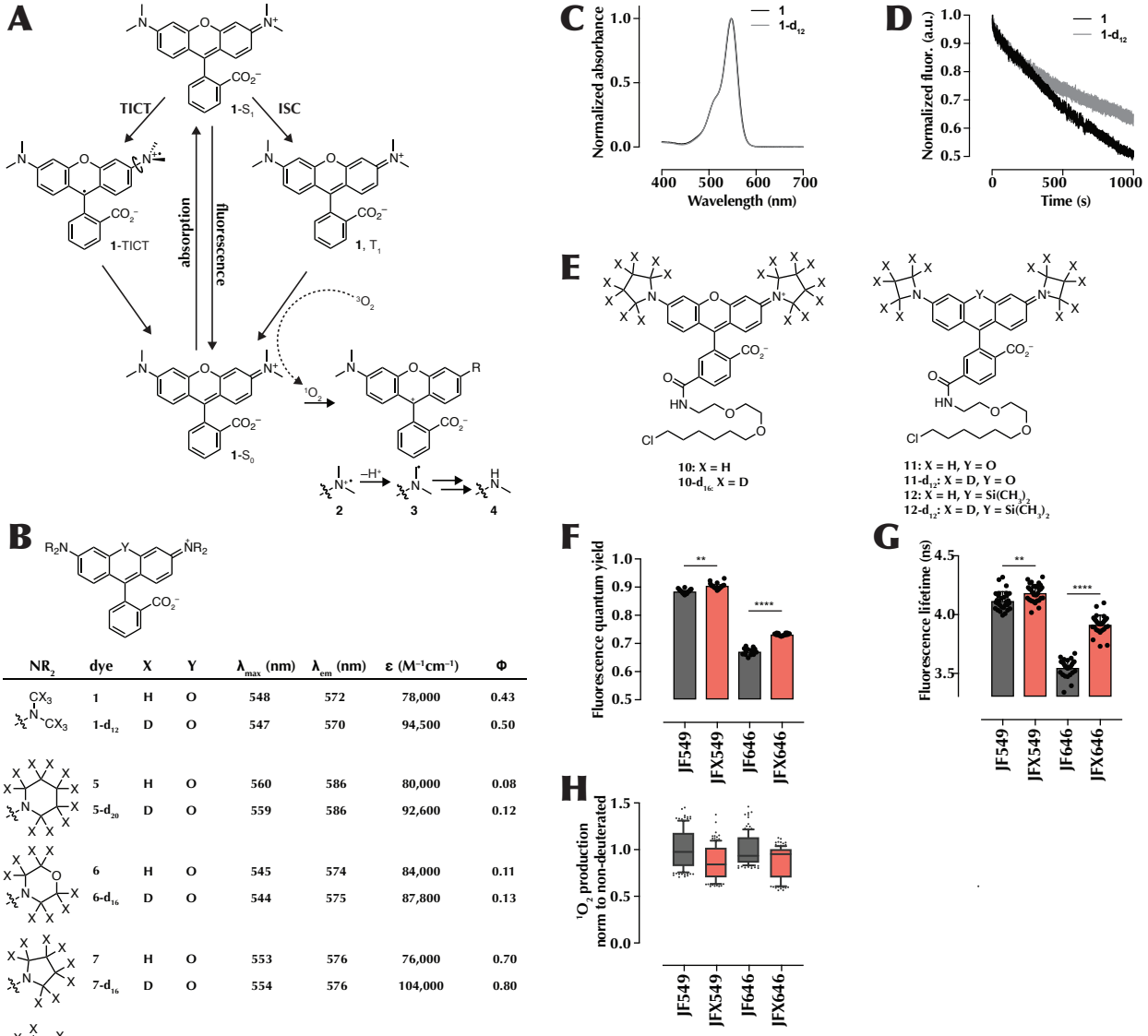
The different behavior of structurally similar dyes has significant consequences for researchers wishing to employ them in fluorescence microscopy. Perhaps the most striking example is the finding that photoactivatable dyes do not enter (or exit) the cell with nearly the same rate and their non-photoactivatable analogs. Thus, one must optimize labeling conditions with the PA dye itself. These data also suggest that cells are more often incompletely labeled than they are labeled to saturation, which is a potential caveat to consider when trying to use the PA dyes for quantitative PALM experiments where counting molecules is the goal (Lee et al., 2012). A corollary of the same issue related to PA dyes is that additional stringent washing is important to ensure all unreacted dye molecules have been removed before imaging experiments are initiated.

The above data suggest a few bits of practical advice for those thinking of using self-labeling tags. Harnessing isotope effects by deuterating carbons adjacent to the rhodamine aniline ring appears to improve brightness by reducing the probability of non-radiative TICT. This is especially true for the silicon-containing far-red dyes. Moreover, although the five-member pyrolamine-constining JFX_{650} is slightly less bright than the workhorse JF_{646}, its substantially improved labeling kinetics suggest it would be the better dye to use in most scenarios. In cases where absolute brightness of every molecule are of utmost importance, the deuterated JFX_{646} delivers the most photons per molecule for this wavelength. And, of course while the effects were less dramatic for the TMR-spectrum dyes, it is still clear the deuteration improves brightness overall.

The above data also highlight the performance differences between HaloTag and SNAPtag. When only one tag is needed, HaloTag is the clear choice. In the event that two tags are necessary, the chloropyrimidine moiety clearly outperforms the more common benzylguanine. We suggest that SNAPtag might be made significantly better through more concerted efforts to improve the

permeability of the molecule, but more testing is needed to understand the molecular forces at play affecting permeability of either HaloTag- or SNAPtag- liganded dyes.

Figures



NR_2	dye	X	Y	λ_{max} (nm)	λ_{em} (nm)	ε ($M^{-1}cm^{-1}$)	Φ
CX_3	1	H	O	548	572	78,000	0.43
CX_3	1-d_{12}	D	O	547	570	94,500	0.50
	5	H	O	560	586	80,000	0.08
	5-d_{20}	D	O	559	586	92,600	0.12
	6	H	O	545	574	84,000	0.11
	6-d_{16}	D	O	544	575	87,800	0.13
	7	H	O	553	576	76,000	0.70
	7-d_{16}	D	O	554	576	104,000	0.80
	8	H	O	549	571	101,000	0.88
	8-d_{12}	D	O	548	570	96,700	0.86
	9	H	$Si(CH_3)_2$	646	664	5,600 (152,000)[a]	0.54
	9-d_{12}	D	$Si(CH_3)_2$	645	662	8,600 (155,000)[a]	0.54

[a]Extinction coefficient measured in EtOH containing 0.1% (v/v) TFA

Figure 1. Deuterated rhodamine dyes have different photophysical properties. **A)** Excited state reaction schema of rhodamine dyes. **B)** Photophysical properties of dyes **1, 5-9** and their deuterated analogs. **C)** Absorbance spectra of **1** and **1-d12**. **D)** Photobleaching of **1** and **1-d12**. **E)** Chemical structures of HaloTag ligands **10, 11** (aka JF_{549}), and **12** (aka JF_{646}) and deuterated analogs. **F)** Fluorescence quantum yield of JF_{549} and J JF_{646}, and deuterated analogs attached to purified HaloTag protein. **G)** Fluorescence lifetime of the HaloTag conjugates of JF_{549} and JF_{646}, and deuterated analogs in live cells. **H)** Singlet oxygen quantum yield of JF_{549} and JF_{646}, and deuterated analogs attached to HaloTag protein, normalized to the singlet oxygen production of the hydrogenated dye.

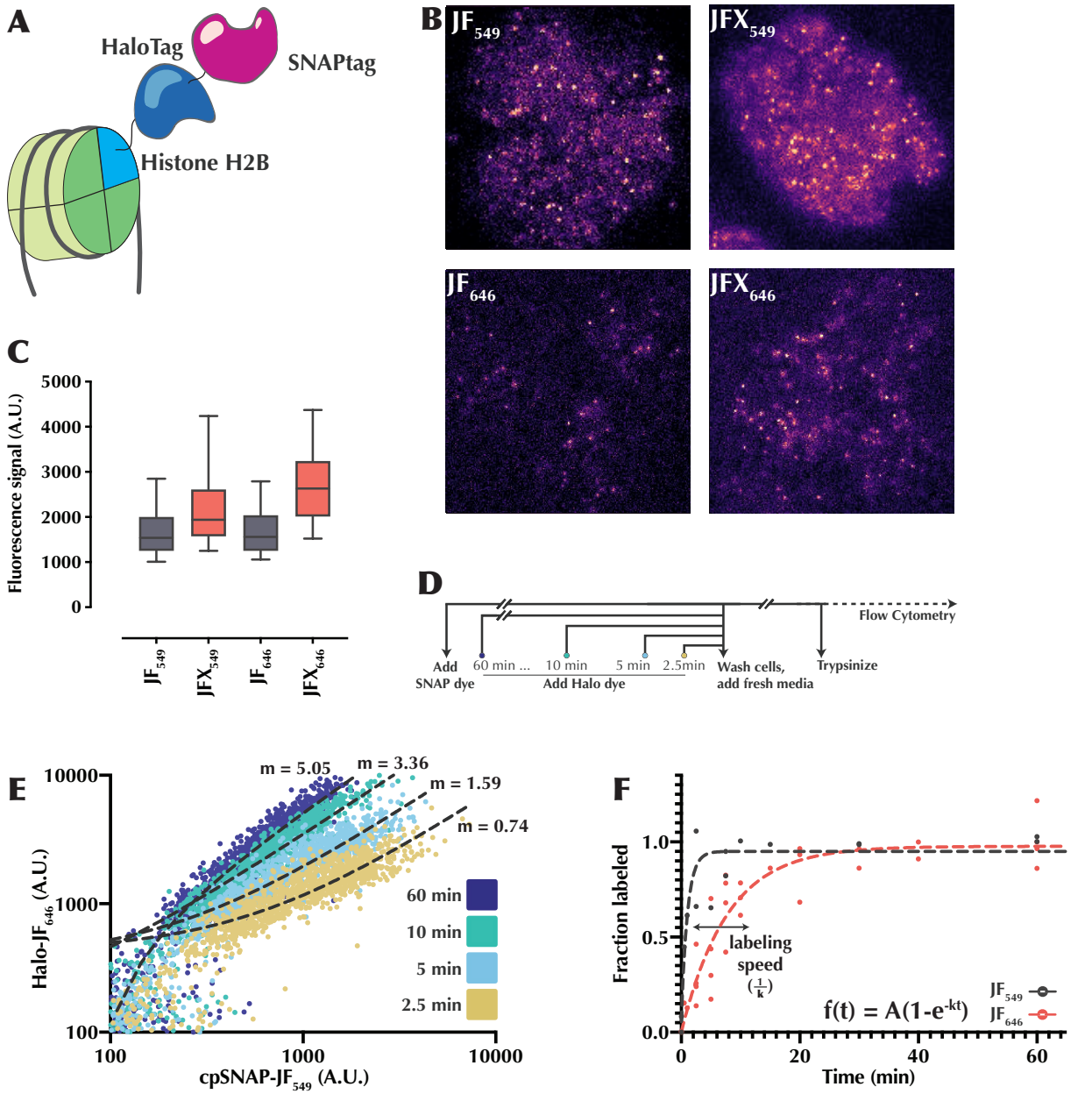


Figure 2. Measuring deuterated and non-deuterated dyes in cells. **A)** Schematic of the Histone H2B-HaloTag-SNAPtag fusion protein. **B)** Representative images of cells labeled with limited concentrations of JF_{549}, JFX_{549}, JF_{646}, or JFX_{646} to measure brightness of the single molecule level. **C)** Quantification of the spot brightness from the experiment shown in (B). **D)** Experimental scheme for flow cytometry labeling kinetics experiments. **E)** The ratio of HaloTag signal to SNAPtag signal (the slope of a linear fit) as a function of labeling time. **F)** Individual slopes extracted from the experiment in (E) fit with a single exponential curve to extract the labeling speed and the relative brightness.

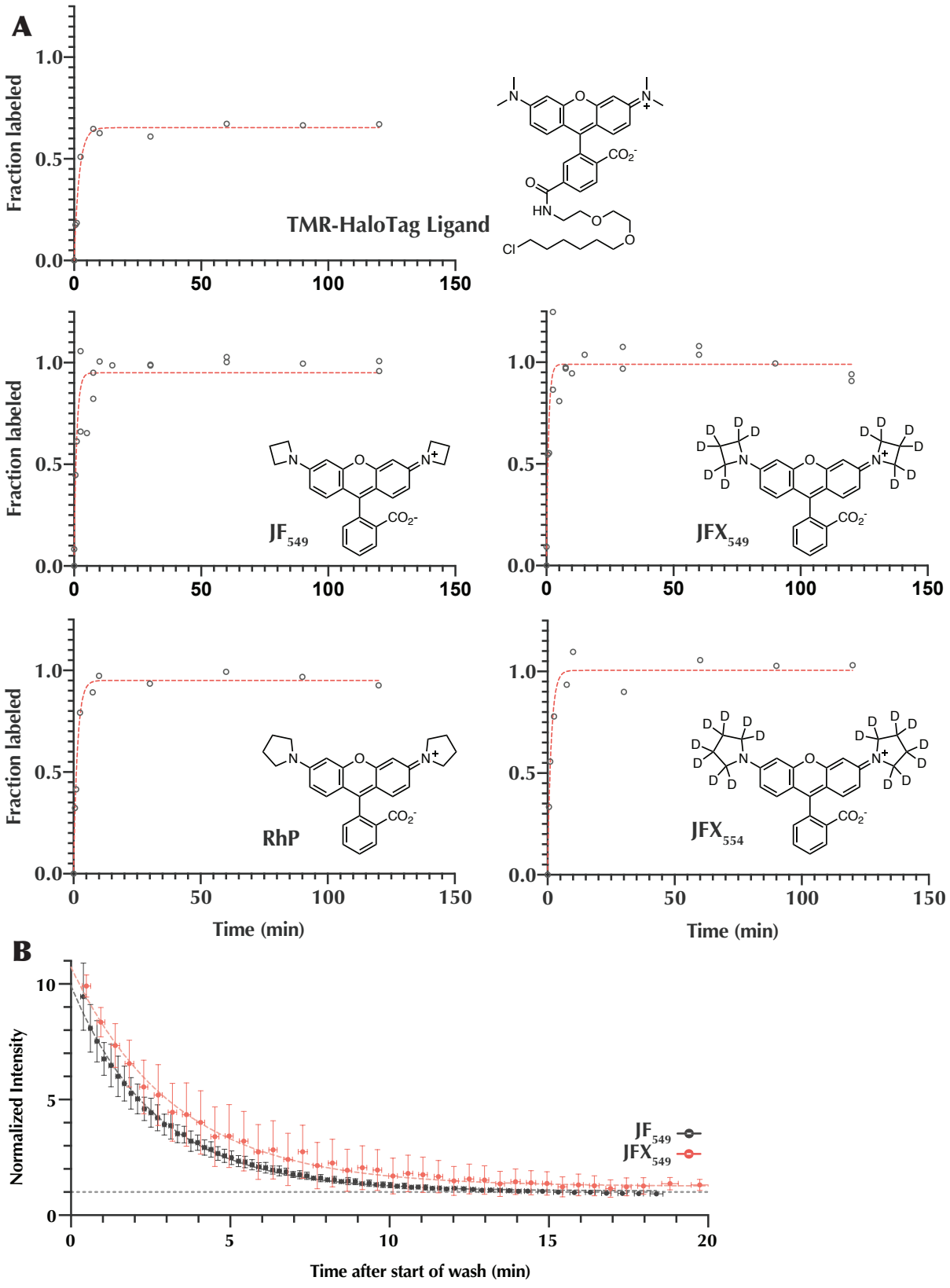


Figure 3. Summary of TMR-spectrum HaloTag ligand labeling kinetics **A)** Labeling kinetics measurements from five different TMR-derived HaloTag variants labeling cells expressing

Histone H2B-HaloTag-SNAPtag using flow cytometry and assessed as a slope relative to the SNAPtag fluorescence signal. Dashed red line represents a single exponential association fit. **B)** Flow cytometry data of Halo-CTCF cells measuring the relative amount of JF_{549} (grey points) and JFX_{549} (red points) inside cells after addition of fresh media. Data are the average of three technical replicates.

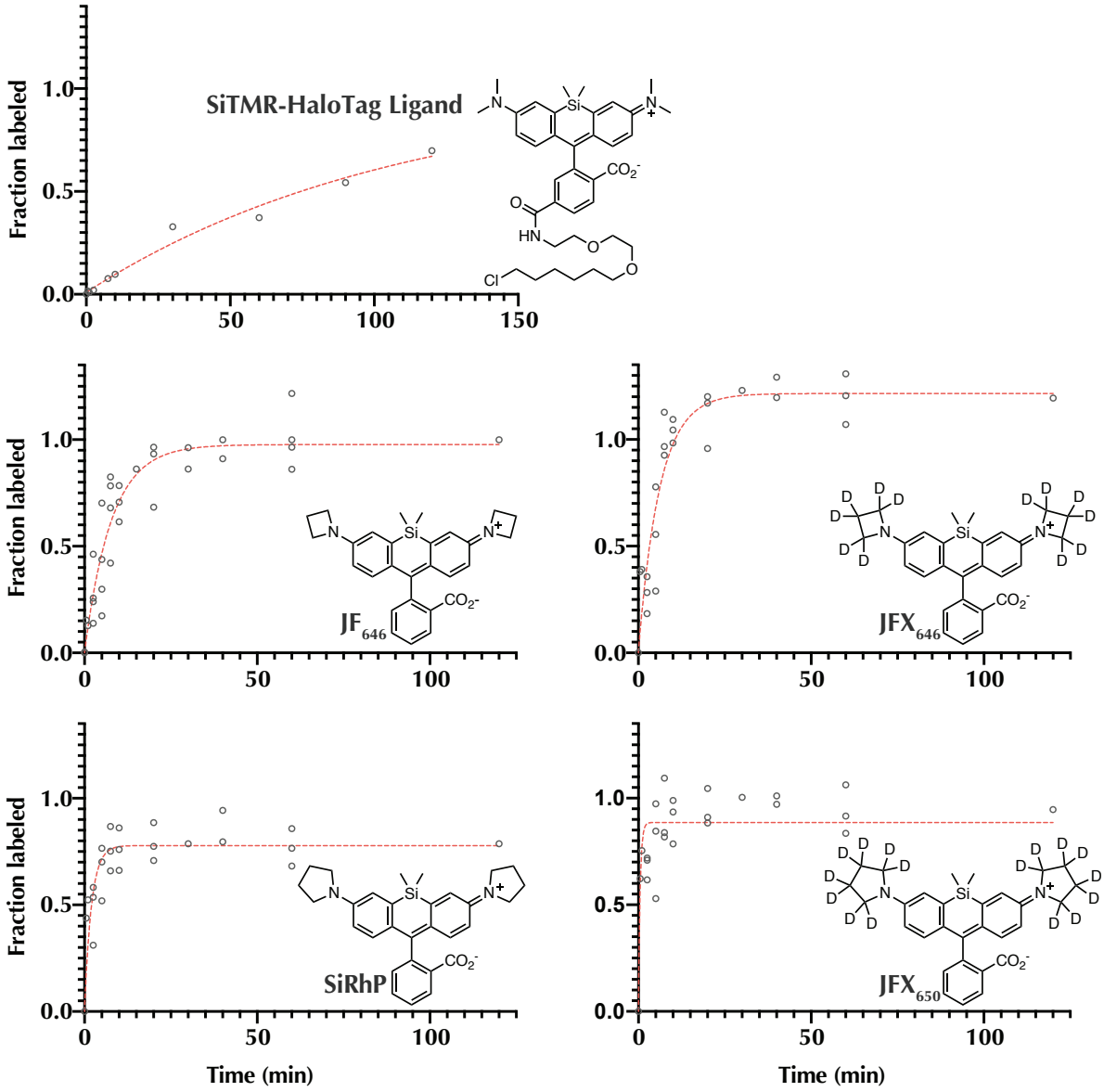


Figure 4. Summary of SiTMR-spectrum HaloTag ligand labeling kinetics. Labeling kinetics from cells expressing Histone H2B-HaloTag-SNAPtag using flow cytometry and assessed as a slope relative to the SNAPtag fluorescence signal. Dashed red line represents a single exponential association fit.

	JF_{549}-like dyes					**JF_{646}-like dyes**				
	JF_{549}	TMR	JFX_{549}	RhP	JFX_{554}	JF_{646}	SiTMR	JFX_{646}	SiRhP	JFX_{650}
Labeling $t_{1/2}$ (min)	0.72*	1.37*	0.58*	1.05*	0.98*	6.30	75	4.75	0.98*	0.34*
Relative Brightness (A.U.)	1.00	0.69	1.04	1.00	1.06	1.00	0.69	1.24	0.79	0.89

Table 1. Summary of the brightness and labeling times of the HaloTag ligands tested. All brightness are reported with respect to JF_{549} in the case of the TMR-spectrum dyes, and JF_{646} for the SiTMR-spectrum dyes. Values marked with an asterisk (*) were too fast to adequately measure, and should be interpreted as an upper bound on the real labeling time.

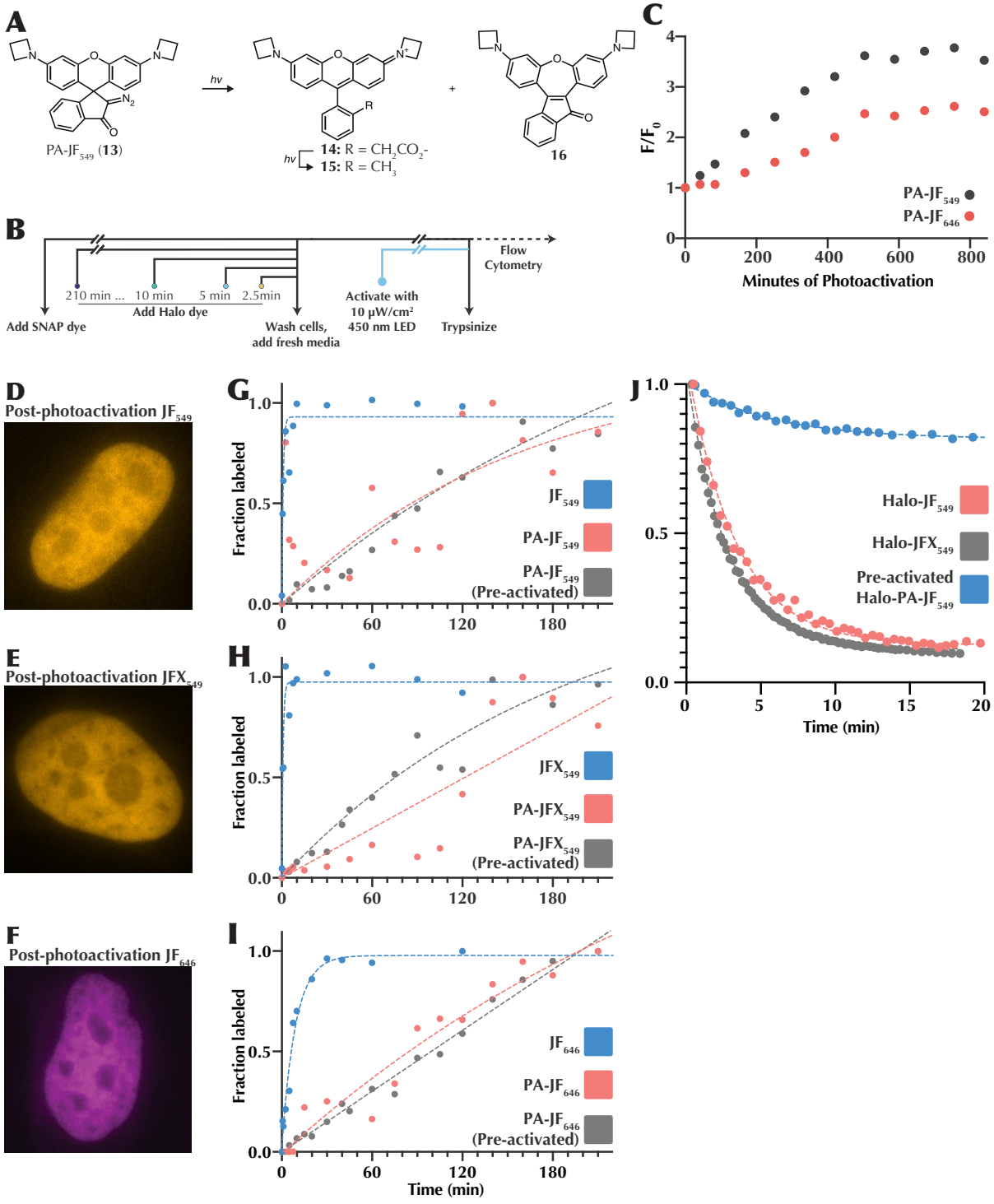


Figure 5. Measuring labeling kinetics of photoactivatable JF dyes **A)** Photoactivation reaction of PA-JF_{549} to produce fluorescent compounds **14** and **15** and non-fluorescent compound **16**. Schematic reproduced with permission from (Grimm et al. 2016). **B)** Experimental schematic for labeling and uncaging of PA dyes prior to flow cytometry. **C)** Degree of photoactivation as a function of illumination time. **D-F)** Representative images of Histone H2B-HaloTag-SNAPtag

cells labeled with PA dyes after 600 minutes of photoactivation. **G-I)** Labeling kinetics measurement of conventional dyes (blue points) and their matched photoactivatable counterparts. Dyes were either added to cells prior to photoactivation (red points) or the dye we photoactivated in solution using a strong 405 nm light source and then added to cells (grey points). **J)** Washout kinetics of pre-photoactivated PA-JF_{549} compared to JF_{549} and JFX_{549}.

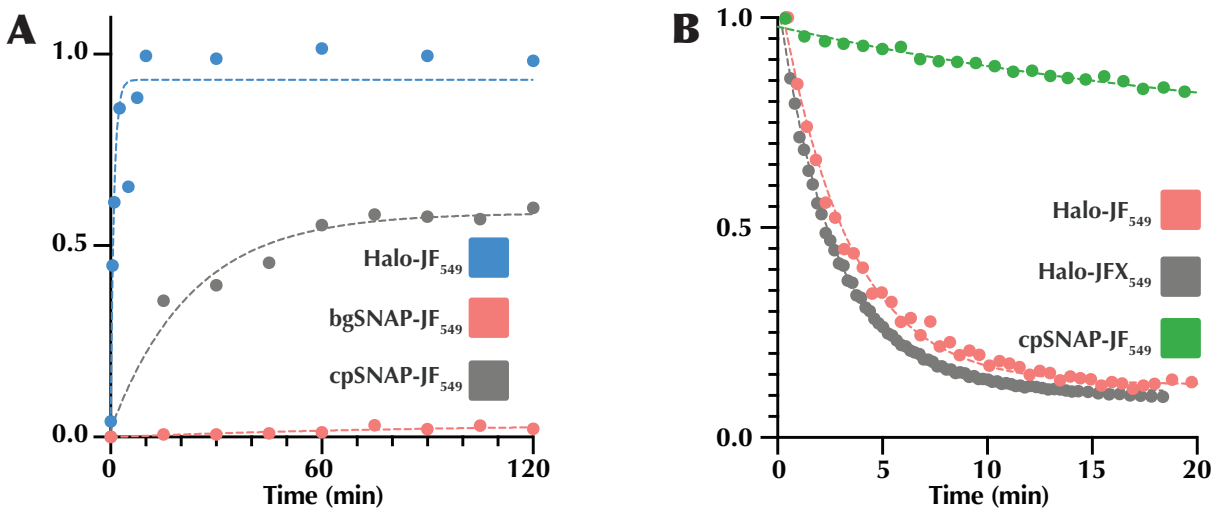


Figure 6. Comparing SNAPtag ligands to HaloTag ligands. **A)** Labeling kinetics of bgSNAP-JF_{549} (red points) and cpSNAP-JF_{549} (grey points) compared to HaloTag-JF_{549} (blue points). **B)** Washout kinetics of cpSNAP-JF_{549} (green points) compared to JF_{549} and JFX_{549}.

Materials and Methods

Cell culture

U2OS (osteosarcoma from a 15 y.o. female) expressing Histone H2B-HaloTag-SNAPtag, or endogenously tagged at the CTCF locus have been previously described. Cells were cultured in DMEM with 10% FBS and 100 U/mL Penecillin and Streptomycin. Cells were maintained at 37°C and 5% CO_2 and subcultured every 2-3 days.

Microscopy

Single particle tracking experiments were carried out as previously described *(Hansen et al. 2017)*, but are described here in brief. After overnight growth, U2OS cells expressing H2B-HaloTag-SNAPtag were labeled with 10 pM each of JF_{549} and JF_{646}, or JFX_{549} and JFX_{646}. Single molecule imaging was performed on a custom-built Nikon Ti microscope fitted with a 100x/NA 1.49 oil-immersion TIRF objective, motorized mirror are to allow HiLo illumination of the sample, Perfect Focus System, and two aligned EM-CCD cameras. Samples were illuminated using 405-nm (140 mW, OBIS coherent), 561-nm (1 W, genesis coherent), and 633-nm (1 W, genesis coherent) lasers, which were focused onto the back pupil plane of the objective via fiber and multi-notch dichromatic mirror (405-nm/488-nm/561-nm/633-nm quad-band; Semrock, NF03-405/488/532/635E-25). Excitation intensity and pulse width were controlled through an acousto-optic transmission filter (AOTF nC-VIS-TN, AA Opto-Electronic) triggered using the camera's TTL exposure output signal. Fluorescence emissions were filtered with a single bandpass filter in front of the camera (Semrock 676/37 nm bandpass filter). All of the components of the microscope, camera, and other hardware were controlled through NIS-Elements software (Nikon). Image data were analyzed using a MATLAB implementation of the Multiple Target Tracking algorithm *(Sergé et al. 2008)*, and the intensity values from fitting molecules were used to interpret molecule brightness.

Flow Cytometry

Histone H2B-HaloTag-SNAPtag cells were seeded in 12-well plates. 0.5 mL of media containing 500 nM cpSNAP-JF549 (for experiments measuring SiTMR-spectrum dyes) or SNAP-JFX650 (for experiments measuring TMR-spectrum dyes) was added to samples at -2 hr relative to the wash. A master mix containing 500 nM each of SNAPtag ligand and 500 nM HaloTag ligand was prepared an maintained at 37 °C, and added into the correct wells at the corresponding time point, to make a final concentration of 500 nM SNAP ligand and 250 nM HaloTag ligand. Once time zero was reached, all wells were aspirated, washed with prewarmed PBS, and then incubated with fresh DMEM for 30 minutes. Sample were trypsinized and strained to remove any large cell debris.

For experiments with photoactivatable dyes, H2B-Halo-SNAP cells were plated on 96-well plates, and four wells were collectively treated for each condition. After the last addition of dye in the time course, cells were washed with PBS, then covered with fluorobright DMEM

supplemented with 10% FBS. The 96-well plate was then placed on a blue-light illuminator, generously provided by Nicole Repina, and illuminated with 10 μW/cm^2 450 nm light for 2-10 hours. Thankfully, the turnover of CAN-incorporated histones is quite low, making such long delays between labeling and analysis possible. After illumination, samples were trypsinized and separate wells of the same condition were combined prior to analysis.

Cells were analyzed on an LSR Fortessa Analyzer using the PE-YG and APC channels to measure the TMR-spectrum and SiTMR-spectrum, respectively. Voltages were set and verified using unlabeled and single-channel controls independently for each biological replicate, but converged to voltages of ~250 volts for each. 3000 – 10000 cells were recorded for each sample. Slopes were calculated from the raw intensity values after selecting for singlet events using forward and side scatter, as well as removing cells not expressing H2B-Halo-SNAP.

Washout rate measurements

To measure the rate at which dyes wash out of cells, U2OS cells expressing HaloTag-CTCF at the endogenous locus (Hansen et al. 2017) were grown to ~90% confluence in 10 cm tissue culture plates, and then labeled with 500 nM dye for 2 hours. We chose HaloTag-CTCF because careful measurements have subsequently demonstrated the number of molecules per cell (Cattoglio et al. 2019), reasoning that it would be preferable for the fluorescent signal in the cells to approach a known brightness value rather than a more poorly-defined zero point in the flow cytometry plot. After 2 hours, the cells were quickly washed with PBS and trypsinized, and resuspended in media containing 500 nM dye. Cells were then centrifuged at 200g for 5 minutes to pellet, the old media aspirated, and 3 mL fresh media without dye was added. This point is considered time zero. The cell suspension was then placed on the flow cytometer and continuously measured for 10-30 minutes at a flow rate of ~500 cells/second. Data were binned into 2000-cell bins and the average and standard deviation measured from these binned events. Unlabeled cells as well as "wildtype" U2OS cells not expressing HaloTag were used to set voltages and to confirm that free dye in the media does not significantly increase the fluorescence signal in the cytometer. Only dye contained within the cell contributes measurable signal.

References

Abrisch, R.G., Eidem, T.M., Yakovchuk, P., Kugel, J.F., and Goodrich, J.A. (2015). Infection by Herpes Simplex Virus Type-1 Causes Near-Complete Loss of RNA Polymerase II Occupancy on the Host Cell Genome. J. Virol. JVI.02665-15.

Aguzzi, A., and Altmeyer, M. (2016). Phase Separation: Linking Cellular Compartmentalization to Disease. Trends Cell Biol. *26*, 547–558.

Alberti, S., Gladfelter, A., and Mittag, T. (2019). Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. Cell *176*, 419–434.

Allen, B.L., and Taatjes, D.J. (2015). The Mediator complex: a central integrator of transcription. Nat. Rev. Mol. Cell Biol. *16*, 155–166.

Altmeyer, M., Neelsen, K.J., Teloni, F., Pozdnyakova, I., Pellegrino, S., Grøfte, M., Rask, M.-B.D., Streicher, W., Jungmichel, S., Nielsen, M.L., et al. (2015). Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose). Nat. Commun. *6*, 8088.

Amitai, A. (2018). Chromatin Configuration Affects the Dynamics and Distribution of a Transiently Interacting Protein. Biophys. J. *114*, 766–771.

Annibale, P., Vanni, S., Scarselli, M., Rothlisberger, U., and Radenovic, A. (2011). Identification of clustering artifacts in photoactivated localization microscopy. Nat. Methods *8*, 527–528.

Asherie, N. (2004). Protein crystallization and phase diagrams. Methods *34*, 266–272.

Banani, S.F., Rice, A.M., Peeples, W.B., Lin, Y., Jain, S., Parker, R., and Rosen, M.K. (2016). Compositional Control of Phase-Separated Cellular Bodies. Cell *166*, 651–663.

Banani, S.F., Lee, H.O., Hyman, A.A., and Rosen, M.K. (2017). Biomolecular condensates: organizers of cellular biochemistry. Nat. Rev. Mol. Cell Biol. *18*, 285–298.

Bensaude, O. (2011). Inhibiting eukaryotic transcription. Transcription *2*, 103–108.

Berg, O.G., Winter, R.B., and von Hippel, P.H. (1981). Diffusion-Driven Mechanisms of Protein Translocation on Nucleic Acids. 1. Models and Theory. Biochemistry *20*, 6929–6948.

Bergeron-Sandoval, L.-P., Safaee, N., and Michnick, S.W. (2016). Mechanisms and Consequences of Macromolecular Phase Separation. Cell *165*, 1067–1079.

Berry, J., Weber, S.C., Vaidya, N., Haataja, M., and Brangwynne, C.P. (2015). RNA transcription modulates phase transition-driven nuclear body assembly. Proc. Natl. Acad. Sci. *112*, E5237–E5245.

Blaho, J.A., Morton, E.R., and Yedowitz, J.C. (2005). Herpes Simplex Virus: Propagation, Quantification, and Storage. Curr. Protoc. Microbiol. 0–23.

Boehning, M., Dugast-Darzacq, C., Rankovic, M., Hansen, A.S., Yu, T.-K., Marie-Nelly, H., McSwiggen, D.T., Kokic, G., Dailey, G.M., Cramer, P., et al. (2018a). RNA polymerase II clustering through carboxy-terminal domain phase separation. Nat. Struct. Mol. Biol. *25*, 833–840.

Boehning, M., Dugast-Darzacq, C., Rankovic, M., Hansen, A.S., Yu, T.-K., Marie-Nelly, H., Kokic, G., Dailey, G.M., Cramer, P., Darzacq, X., et al. (2018b). RNA polymerase II

clustering through CTD phase separation. BioRxiv.

Boettiger, A.N., Bintu, B., Moffitt, J.R., Wang, S., Beliveau, B.J., Fudenberg, G., Imakaev, M., Mirny, L.A., Wu, C., and Zhuang, X. (2016). Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. Nature *529*, 418–422.

Boija, A., Klein, I.A., Sabari, B.R., Dall'Agnese, A., Coffey, E.L., Zamudio, A. V., Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M., et al. (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell *175*, 1842-1855.e16.

Bouchard, J.J., Otero, J.H., Scott, D.C., Szulc, E., Martin, E.W., Sabri, N., Granata, D., Marzahn, M.R., Lindorff-Larsen, K., Salvatella, X., et al. (2018). Cancer Mutations of the Tumor Suppressor SPOP Disrupt the Formation of Active, Phase-Separated Compartments. Mol. Cell *72*, 19-36.e8.

Bracha, D., Walls, M.T., Wei, M.T., Zhu, L., Kurian, M., Avalos, J.L., Toettcher, J.E., and Brangwynne, C.P. (2018). Mapping Local and Global Liquid Phase Behavior in Living Cells Using Photo-Oligomerizable Seeds. Cell 1467–1480.

Brangwynne, C.P. (2013). Phase transitions and size scaling of membrane-less organelles. J. Cell Biol. *203*, 875–881.

Brangwynne, C.P., Eckmann, C.R., Courson, D.S., Rybarska, A., Hoege, C., Gharakhani, J., Julicher, F., and Hyman, A.A. (2009). Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. Science *324*, 1729–1732.

Brangwynne, C.P., Mitchison, T.J., and Hyman, A.A. (2011). Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. Proc. Natl. Acad. Sci. *108*, 4334–4339.

Brangwynne, C.P., Tompa, P., and Pappu, R. V. (2015). Polymer physics of intracellular phase transitions. Nat. Phys. *11*, 899–904.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Cai, Y., Hossain, M.J., Hériché, J.-K., Politi, A.Z., Walther, N., Koch, B., Wachsmuth, M., Nijmeijer, B., Kueblbeck, M., Martinic-Kavur, M., et al. (2018). Experimental and computational framework for a dynamic protein atlas of human cell division. Nature *561*, 411–415.

Chang, L., Godinez, W.J., Kim, I.-H., Tektonidis, M., de Lanerolle, P., Eils, R., Rohr, K., and Knipe, D.M. (2011). Herpesviral replication compartments move and coalesce at nuclear speckles to enhance export of viral late mRNA. Proc. Natl. Acad. Sci. U. S. A. *108*, E136–E144.

Chen, D., and Huang, S. (2001). Nucleolar components involved in ribosome biogenesis cycle between the nucleolus and nucleoplasm in interphase cells. J. Cell Biol. *153*, 169–176.

Chen, B.-C., Legant, W.R., Wang, K., Shao, L., Milkie, D.E., Davidson, M.W., Janetopoulos, C., Wu, X.S., Hammer, J.A., Liu, Z., et al. (2014). Lattice light-sheet microscopy: Imaging molecules to embryos at high spatiotemporal resolution. Science *346*, 1257998–1257998.

Chen, J., Tresenrider, A., Chia, M., McSwiggen, D.T., Spedale, G., Jorgensen, V., Liao, H., van Werven, F.J., and Unal, E. (2017). Kinetochore inactivation by expression of a repressive mRNA. Elife *6*.

Cho, W.-K., Spille, J.-H., Hecht, M., Lee, C., Li, C., Grube, V., and Cisse, I.I. (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science *361*, 412–415.

Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G.M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X., et al. (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. Science *2555*, eaar2555.

Cisse, I.I., Izeddin, I., Causse, S.Z., Boudarene, L., Senecal, A., Muresan, L., Dugast-Darzacq, C., Hajj, B., Dahan, M., and Darzacq, X. (2013). Real-time dynamics of RNA polymerase II clustering in live human cells. Science *341*, 664–667.

Cole, N.B. (2013). Site-Specific Protein Labeling with SNAP-Tags. In Current Protocols in Protein Science, (Hoboken, NJ, USA: John Wiley & Sons, Inc.), pp. 30.1.1-30.1.16.

Courchaine, E.M., Lu, A., and Neugebauer, K.M. (2016). Droplet organelles? EMBO J. *35*, 1603–1612.

Dai-Ju, J.Q., Li, L., Johnson, L.A., and Sandri-Goldin, R.M. (2006). ICP27 Interacts with the C-Terminal Domain of RNA Polymerase II and Facilitates Its Recruitment to Herpes Simplex Virus 1 Transcription Sites, Where It Undergoes Proteasomal Degradation during Infection. J. Virol. *80*, 3567–3581.

Davis, Z.H., Verschueren, E., Jang, G.M., Kleffman, K., Johnson, J.R., Park, J., Von Dollen, J., Maher, M.C., Johnson, T., Newton, W., et al. (2015). Global Mapping of Herpesvirus-Host Protein Complexes Reveals a Transcription Strategy for Late Genes. Mol. Cell *57*, 349–360.

Dembowski, J. a., and DeLuca, N.A. (2015). Selective Recruitment of Nuclear Factors to Productively Replicating Herpes Simplex Virus Genomes. PLOS Pathog. *11*, e1004939.

Dine, E., Gil, A.A., Uribe, G., Brangwynne, C.P., and Toettcher, J.E. (2018). Protein Phase Separation Provides Long-Term Memory of Transient Spatial Stimuli. Cell Syst. *6*, 655-663.e5.

Dosztanyi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics *21*, 3433–3434.

Dosztányi, Z., Csizmók, V., Tompa, P., and Simon, I. (2005). The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. J. Mol. Biol. *347*, 827–839.

Doudna, J.A., and Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. Science *346*, 1258096–1258096.

Du, M., and Chen, Z.J. (2018). DNA-induced liquid phase condensation of cGAS activates innate immune signaling. Science *361*, 704–709.

Eason, M.G., Damry, A.M., and Chica, R.A. (2017). Structure-guided rational design of red fluorescent proteins: towards designer genetically-encoded fluorophores. Curr. Opin. Struct.

Biol. *45*, 91–99.

Elbaum-Garfinkle, S., Kim, Y., Szczepaniak, K., Chen, C.C.-H., Eckmann, C.R., Myong, S., and Brangwynne, C.P. (2015). The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. Proc. Natl. Acad. Sci. *112*, 7189–7194.

Erdmann, R.S., Baguley, S.W., Richens, J.H., Wissner, R.F., Xi, Z., Allgeyer, E.S., Zhong, S., Thompson, A.D., Lowe, N., Butler, R., et al. (2019). Labeling Strategies Matter for Super-Resolution Microscopy: A Comparison between HaloTags and SNAP-tags. Cell Chem. Biol. *26*, 584-592.e6.

Eriksson, B.F., and Schinazi, R.F. (1989). Combinations of 3'-azido-3'-deoxythymidine (zidovudine) and phosphonoformate (foscarnet) against human immunodeficiency virus type 1 and cytomegalovirus replication in vitro. Antimicrob. Agents Chemother. *33*, 663–669.

Fei, J., Jadaliha, M., Harmon, T.S., Li, I.T.S., Hua, B., Hao, Q., Holehouse, A.S., Reyer, M., Sun, Q., Freier, S.M., et al. (2017). Quantitative analysis of multilayer organization of proteins and RNA in nuclear speckles at super resolution. J. Cell Sci. *130*, 4180–4192.

Feric, M., Vaidya, N., Harmon, T.S., Mitrea, D.M., Zhu, L., Richardson, T.M., Kriwacki, R.W., Pappu, R. V., and Brangwynne, C.P. (2016). Coexisting Liquid Phases Underlie Nucleolar Subcompartments. Cell *165*, 1686–1697.

Fong, Y.W., Inouye, C., Yamaguchi, T., Cattoglio, C., Grubisic, I., and Tjian, R. (2011). A DNA repair complex functions as an Oct4/Sox2 coactivator in embryonic stem cells. Cell *147*, 120–131.

Franzmann, T.M., Jahnel, M., Pozniakovsky, A., Mahamid, J., Holehouse, A.S., Nüske, E., Richter, D., Baumeister, W., Grill, S.W., Pappu, R. V., et al. (2018). Phase separation of a yeast prion protein promotes cellular fitness. Science *359*, eaao5654.

Freeman Rosenzweig, E.S., Xu, B., Kuhn Cuellar, L., Martinez-Sanchez, A., Schaffer, M., Strauss, M., Cartwright, H.N., Ronceray, P., Plitzko, J.M., Förster, F., et al. (2017). The Eukaryotic CO2-Concentrating Organelle Is Liquid-like and Exhibits Dynamic Reorganization. Cell *171*, 148-162.e19.

Galkin, O., Chen, K., Nagel, R.L., Hirsch, R.E., and Vekilov, P.G. (2002). Liquid-liquid separation in solutions of normal and sickle cell hemoglobin. Proc. Natl. Acad. Sci. *99*, 8479–8483.

Gautier, A., Juillerat, A., Heinis, C., Corrêa, I.R., Kindermann, M., Beaufils, F., and Johnsson, K. (2008). An Engineered Protein Tag for Multiprotein Labeling in Living Cells. Chem. Biol. *15*, 128–136.

Goreaud, F., and Pélissier, R. (1999). On explicit formulas of edge effect correction for Ripley's K -function. J. Veg. Sci. *10*, 433–438.

Grimm, J.B., English, B.P., Chen, J., Slaughter, J.P., Zhang, Z., Revyakin, A., Patel, R., Macklin, J.J., Normanno, D., Singer, R.H., et al. (2015). A general method to improve fluorophores for live-cell and single-molecule microscopy. Nat. Methods *12*, 244–250.

Grimm, J.B., English, B.P., Choi, H., Muthusamy, A.K., Mehl, B.P., Dong, P., Brown, T.A., Lippincott-Schwartz, J., Liu, Z., Lionnet, T., et al. (2016). Bright photoactivatable fluorophores for single-molecule imaging. Nat. Methods *13*, 985–988.

Grimm, J.B., Muthusamy, A.K., Liang, Y., Brown, T.A., Lemon, W.C., Patel, R., Lu, R., Macklin, J.J., Keller, P.J., Ji, N., et al. (2017). A general method to fine-tune fluorophores for live-cell and in vivo imaging. Nat. Methods *14*, 987–994.

Guo, Y.E., Manteiga, J.C., Henninger, J.E., Sabari, B.R., Dall'Agnese, A., Hannett, N.M., Spille, J.-H., Afeyan, L.K., Zamudio, A. V., Shrinivas, K., et al. (2019). Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. Nature *572*, 543–548.

Hansen, A.S., Pustova, I., Cattoglio, C., Tjian, R., and Darzacq, X. (2017). CTCF and cohesin regulate chromatin loop stability with distinct dynamics. Elife *6*, 1–33.

Hansen, A.S., Cattoglio, C., Darzacq, X., and Tjian, R. (2018a). Recent evidence that TADs and chromatin loops are dynamic structures. Nucleus *9*, 20–32.

Hansen, A.S., Woringer, M., Grimm, J.B., Lavis, L.D., Tjian, R., and Darzacq, X. (2018b). Robust model-based analysis of single-particle tracking experiments with Spot-On. Elife *7*, 1–33.

Heller, M.C., Carpenter, J.F., and Randolph, T.W. (1996). Effects of phase separating systems on lyophilized hemoglobin. J. Pharm. Sci. *85*, 1358–1362.

Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A Phase Separation Model for Transcriptional Control. Cell *169*, 13–23.

Ho, J.J., Cattoglio, C., McSwiggen, D.T., Tjian, R., and Fong, Y.W. (2017). Regulation of DNA demethylation by the XPC DNA repair complex in somatic and pluripotent stem cells. Genes Dev. *31*, 830–844.

Hull, L.A., Davis, G.T., Rosenblatt, D.H., Williams, H.K.R., and Weglein, R.C. (1967). Oxidations of Amines. III. Duality of Mechanism in the Reaction of Amines with Chlorine Dioxide. J. Am. Chem. Soc. *89*, 1163–1170.

Hyman, A.A., Weber, C.A., and Jülicher, F. (2014). Liquid-Liquid Phase Separation in Biology. Annu. Rev. Cell Dev. Biol. *30*, 39–58.

Ikeda, K., Steger, D.J., Eberharter, A., and Workman, J.L. (2015). Activation Domain-Specific and General Transcription Stimulation by Native Histone Acetyltransferase Complexes. Mol. Cell. Biol. *19*, 855–863.

Izeddin, I., Récamier, V., Bosanac, L., Cisse, I.I., Boudarene, L., Dugast-Darzacq, C., Proux, F., Bénichou, O., Voituriez, R., Bensaude, O., et al. (2014). Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus. Elife e02230.

J. Overbeek, M.J., and Voorn (1957). Phase Separation in Polyelectrolyte Solutions. Theory of Complex Coacervation. J. Cell. Physiol. 7–26.

Jain, A., and Vale, R.D. (2017). RNA phase transitions in repeat expansion disorders. Nature *546*, 243–247.

Jradi, F.M., and Lavis, L.D. (2019). Chemistry of Photosensitive Fluorophores for Single-Molecule Localization Microscopy. ACS Chem. Biol. *14*, 1077–1090.

Keppler, A., Gendreizig, S., Gronemeyer, T., Pick, H., Vogel, H., and Johnsson, K. (2003). A general method for the covalent labeling of fusion proteins with small molecules in vivo. Nat.

Biotechnol. *21*, 86–89.

Keppler, A., Pick, H., Arrivoli, C., Vogel, H., and Johnsson, K. (2004). Labeling of fusion proteins with synthetic fluorophores in live cells. Proc. Natl. Acad. Sci. *101*, 9955–9959.

Knipe, D.M., and Cliffe, A. (2008). Chromatin control of herpes simplex virus lytic and latent infection. Nat. Rev. Microbiol. *6*, 211–221.

Kroschwald, S., Maharana, S., and Simon, A. (2017). Hexanediol: a chemical probe to investigate the material properties of membrane-less compartments. Matters 1–7.

Kwon, I., Kato, M., Xiang, S., Wu, L., Theodoropoulos, P., Mirzaei, H., Han, T., Xie, S., Corden, J.L., and McKnight, S.L. (2013). Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. Cell *155*, 1049–1060.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Larson, A.G., Elnatan, D., Keenen, M.M., Trnka, M.J., Johnston, J.B., Burlingame, A.L., Agard, D.A., Redding, S., and Narlikar, G.J. (2017). Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. Nature *547*, 236–240.

Lawrence, R.E., Cho, K.F., Rappold, R., Thrun, A., Tofaute, M., Kim, D.J., Moldavski, O., Hurley, J.H., and Zoncu, R. (2018). A nutrient-induced affinity switch controls mTORC1 activation by its Rag GTPase–Ragulator lysosomal scaffold. Nat. Cell Biol. *20*, 1052–1063.

Lee, S.-H., Shin, J.Y., Lee, a., and Bustamante, C. (2012). Counting single photoactivatable fluorescent molecules by photoactivated localization microscopy (PALM). Proc. Natl. Acad. Sci. *109*, 17436–17441.

Levine, M., Cattoglio, C., and Tjian, R. (2014). Looping back to leap forward: transcription enters a new era. Cell *157*, 13–25.

Li, P., Banjade, S., Cheng, H.-C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J. V., King, D.S., Banani, S.F., et al. (2012). Phase transitions in the assembly of multivalent signalling proteins. Nature *483*, 336–340.

Lin, Y., Mori, E., Kato, M., Xiang, S., Wu, L., Kwon, I., and McKnight, S.L. (2016). Toxic PR Poly-Dipeptides Encoded by the C9orf72 Repeat Expansion Target LC Domain Polymers. Cell *167*, 789-802.e12.

Lis, J.T. (2019). A 50 year history of technologies that drove discovery in eukaryotic transcription regulation. Nat. Struct. Mol. Biol.

Liu, Z., Legant, W.R., Chen, B.-C., Li, L., Grimm, J.B., Lavis, L.D., Betzig, E., and Tjian, R. (2014). 3D imaging of Sox2 enhancer clusters in embryonic stem cells. Elife 1–29.

Lomakin, A., Asherie, N., and Benedek, G.B. (1996). Monte Carlo study of phase separation in aqueous protein solutions. J. Chem. Phys. *104*, 1646–1656.

Los, G. V, Encell, L.P., McDougall, M.G., Hartzell, D.D., Karassina, N., Zimprich, C., Wood, M.G., Learish, R., Ohana, R.F., Urh, M., et al. (2008). HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. ACS Chem. Biol. *3*, 373–382.

Lu, H., Yu, D., Hansen, A.S., Ganguly, S., Liu, R., Heckert, A., Darzacq, X., and Zhou, Q. (2018). Phase-separation mechanism for C-terminal hyperphosphorylation of RNA

polymerase II. Nature *558*, 318–323.

Maharana, S., Wang, J., Papadopoulos, D.K., Richter, D., Pozniakovsky, A., Poser, I., Bickle, M., Rizk, S., Guillén-Boixet, J., Franzmann, T.M., et al. (2018). RNA buffers the phase separation behavior of prion-like RNA binding proteins. Science *360*, 918–921.

Mazza, D., Abernathy, A., Golob, N., Morisaki, T., and McNally, J.G. (2012). A benchmark for chromatin binding measurements in live cells. Nucleic Acids Res. *40*, e119–e119.

McClure, W.R. (1985). Mechanism and Control of Transcription Initiation in Prokaryotes. Annu. Rev. Biochem. *54*, 171–204.

McSwiggen, D.T., Hansen, A.S., Teves, S.S., Marie-Nelly, H., Hao, Y., Heckert, A.B., Umemoto, K.K., Dugast-Darzacq, C., Tjian, R., and Darzacq, X. (2019). Evidence for DNA-mediated nuclear compartmentalization distinct from phase separation. Elife *8*.

Mir, M., Bhaduri, B., Wang, R., Zhu, R., and Popescu, G. (2012). Quantitative Phase Imaging (Elsevier Inc.).

Mir, M., Reimer, A., Haines, J.E., Li, X.Y., Stadler, M., Garcia, H., Eisen, M.B., and Darzacq, X. (2017). Dense bicoid hubs accentuate binding along the morphogen gradient. Genes Dev. *31*, 1784–1794.

Mir, M., Stadler, M.R., Ortiz, S.A., Hannon, C.E., Harrison, M.M., Darzacq, X., and Eisen, M.B. (2018). Dynamic multifactor hubs interact transiently with sites of active transcription in Drosophila embryos. Elife *7*, 1–27.

Mitrea, D.M., Cika, J.A., Guy, C.S., Ban, D., Banerjee, P.R., Stanley, C.B., Nourse, A., Deniz, A.A., and Kriwacki, R.W. (2016). Nucleophosmin integrates within the nucleolus via multi-modal interactions with proteins displaying R-rich linear motifs and rRNA. Elife *5*, 1–33.

Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A.P., Kim, H.J., Mittag, T., and Taylor, J.P. (2015). Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. Cell *163*, 123–133.

Monier, K., Armas, J.C., Etteldorf, S., Ghazal, P., and Sullivan, K.F. (2000). Annexation of the interchromosomal space during viral infection. Nat. Cell Biol. *2*, 661–665.

Montgomery, T.H. (1898). Comparative cytological studies, with especial regard to the morphology of the nucleolus. J. Morphol. *25*, 266–584.

Mueller, F., Mazza, D., Stasevich, T.J., and McNally, J.G. (2010). FRAP and kinetic modeling in the analysis of nuclear protein dynamics: What do we really know? Curr. Opin. Cell Biol. *22*, 403–411.

Nair, S.J., Yang, L., Meluzzi, D., Oh, S., Yang, F., Friedman, M.J., Wang, S., Suter, T., Alshareedah, I., Gamliel, A., et al. (2019). Phase separation of ligand-activated enhancers licenses cooperative chromosomal enhancer assembly. Nat. Struct. Mol. Biol. *26*, 193–203.

Narayanan, A., Meriin, A., Andrews, J.O., Spille, J.-H., Sherman, M.Y., and Cisse, I.I. (2019). A first order phase transition mechanism underlies protein aggregation in mammalian cells. Elife *8*, 1–26.

Niewidok, B., Igaev, M., Pereira da Graca, A., Strassner, A., Lenzen, C., Richter, C.P., Piehler, J., Kurre, R., and Brandt, R. (2018). Single-molecule imaging reveals dynamic biphasic

partition of RNA-binding proteins in stress granules. J. Cell Biol. *217*, 1303–1318.

Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T., and Kanemaki, M. (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. Nat. Methods *6*, 917–922.

Normanno, D., Boudarène, L., Dugast-Darzacq, C., Chen, J., Richter, C., Proux, F., Bénichou, O., Voituriez, R., Darzacq, X., and Dahan, M. (2015). Probing the target search of DNA-binding proteins in mammalian cells using TetR as model searcher. Nat. Commun. *6*, 7357.

Nott, T.J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T.D., Bazett-Jones, D.P., Pawson, T., Forman-Kay, J.D., et al. (2015). Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. Mol. Cell *57*, 936–947.

Oltsch, F., Klosin, A., Julicher, F., Hyman, A.A., and Zechner, C. (2019). Phase separation provides a mechanism to reduce noise in cells. BioRxiv 524231.

Pak, C.W., Kosno, M., Holehouse, A.S., Padrick, S.B., Mittal, A., Ali, R., Yunus, A.A., Liu, D.R., Pappu, R. V., and Rosen, M.K. (2016). Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein. Mol. Cell *63*, 72–85.

Paranjape, S.M., Kamakaka, R.T., and Kadonaga, J.T. (1994). Role of chromatin structure in the regulation of transcription by RNA polymerase II. Annu. Rev. Biochem. *63*, 265–297.

Patel, A., Lee, H.O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M.Y., Stoynov, S., Mahamid, J., Saha, S., Franzmann, T.M., et al. (2015). A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. Cell *162*, 1066–1077.

Pélissier, R., and Goreaud, F. (2015). ads Package for R : A Fast Unbiased Implementation of the K -function Family for Studying Spatial Point Patterns in Irregular-Shaped Sampling Windows. J. Stat. Softw. *63*.

Persson, F., Lindén, M., Unoson, C., and Elf, J. (2013). Extracting intracellular diffusive states and transition rates from single-molecule tracking data. Nat. Methods *10*, 265–269.

Putnam, A., Cassani, M., Smith, J., and Seydoux, G. (2019). A gel phase promotes condensation of liquid P granules in Caenorhabditis elegans embryos. Nat. Struct. Mol. Biol. *26*, 220–226.

Qin, P., Parlak, M., Kuscu, C., Bandaria, J., Mir, M., Szlachta, K., Singh, R., Darzacq, X., Yildiz, A., and Adli, M. (2017). Live cell imaging of low- and non-repetitive chromosome loci using CRISPR-Cas9. Nat. Commun. *8*, 1–10.

Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA Synthesis in Mammalian Cells. PLoS Biol. *4*, e309.

Reinkemeier, C.D., Girona, G.E., and Lemke, E.A. (2019). Designer membraneless organelles enable codon reassignment of selected mRNAs in eukaryotes. Science *363*, eaaw2644.

Riback, J.A., Katanski, C.D., Kear-Scott, J.L., Pilipenko, E. V., Rojek, A.E., Sosnick, T.R., and Drummond, D.A. (2017). Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response. Cell *168*, 1028-1040.e19.

Rice, S., and Knipe, D. (1990). Genetic evidence for two distinct transactivation functions of the herpes simplex virus alpha protein ICP27. J. Virol. *64*, 1704–1715.

Rice, S. a, Long, M.C., Lam, V., and Spencer, C. a (1994). RNA polymerase II is aberrantly phosphorylated and localized to viral replication compartments following herpes simplex virus infection. J. Virol. *68*, 988–1001.

Ripley, B.D. (1977). Modelling Spatial Patterns. J. R. Stat. Soc. Ser. B *39*, 172–212.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.

Ruff, E., Record, M., and Artsimovitch, I. (2015). Initial Events in Bacterial Transcription Initiation. Biomolecules *5*, 1035–1062.

Rutkowski, A.J., Erhard, F., L'Hernault, A., Bonfert, T., Schilhabel, M., Crump, C., Rosenstiel, P., Efstathiou, S., Zimmer, R., Friedel, C.C., et al. (2015). Widespread disruption of host transcription termination in HSV-1 infection. Nat. Commun. *6*, 7126.

Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A. V, Manteiga, J.C., et al. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. Science *361*, eaar3958.

Schmidt, H.B., and Görlich, D. (2015). Nup98 FG domains from diverse species spontaneously phase-separate into particles with nuclear pore-like permselectivity. Elife *4*, 1–30.

Schmidt, H.B., and Rohatgi, R. (2016). In Vivo Formation of Vacuolated Multi-phase Compartments Lacking Membranes. Cell Rep. *16*, 1228–1236.

Shen, H., Tauzin, L.J., Baiyasi, R., Wang, W., Moringo, N., Shuang, B., and Landes, C.F. (2017). Single Particle Tracking: From Theory to Biophysical Applications. Chem. Rev. *117*, 7331–7376.

Shin, Y., Berry, J., Pannucci, N., Haataja, M.P., Toettcher, J.E., and Brangwynne, C.P. (2017). Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets. Cell *168*, 159-171.e14.

Shin, Y., Chang, Y.-C., Lee, D.S.W., Berry, J., Sanders, D.W., Ronceray, P., Wingreen, N.S., Haataja, M., and Brangwynne, C.P. (2018). Liquid Nuclear Condensates Mechanically Sense and Restructure the Genome. Cell *175*, 1481-1491.e13.

Shrinivas, K., Sabari, B.R., Coffey, E.L., Klein, I.A., Boija, A., Zamudio, A. V., Schuijers, J., Hannett, N.M., Sharp, P.A., Young, R.A., et al. (2018). Enhancer features that drive formation of transcriptional condensates. BioRxiv.

Smith, J., Calidas, D., Schmidt, H., Lu, T., Rasoloson, D., and Seydoux, G. (2016). Spatial patterning of P granules by RNA-induced phase separation of the intrinsically-disordered protein MEG-3. Elife *5*, 1–18.

Sprague, B.L., and McNally, J.G. (2005). FRAP analysis of binding: Proper and fitting. Trends Cell Biol. *15*, 84–91.

Sprague, B.L., Pego, R.L., Stavreva, D. a, and McNally, J.G. (2004). Analysis of Binding Reactions by Fluorescence Recovery after Photobleaching. Biophys. J. *86*, 3473–3495.

Sprague, B.L., Müller, F., Pego, R.L., Bungay, P.M., Stavreva, D. a, and McNally, J.G. (2006). Analysis of Binding at a Single Spatially Localized Cluster of Binding Sites by Fluorescence Recovery after Photobleaching. Biophys. J. *91*, 1169–1191.

Strom, A.R., Emelyanov, A. V., Mir, M., Fyodorov, D. V., Darzacq, X., and Karpen, G.H. (2017). Phase separation drives heterochromatin domain formation. Nature *547*, 241–245.

Strulson, C.A., Molden, R.C., Keating, C.D., and Bevilacqua, P.C. (2012). RNA catalysis through compartmentalization. Nat. Chem. *4*, 941–946.

Su, X., Ditlev, J.A., Hui, E., Xing, W., Banjade, S., Okrut, J., King, D.S., Taunton, J., Rosen, M.K., and Vale, R.D. (2016). Phase separation of signaling molecules promotes T cell receptor signal transduction. Science *352*, 595–599.

Sun, X., Zhang, A., Baker, B., Sun, L., Howard, A., Buswell, J., Maurel, D., Masharina, A., Johnsson, K., Noren, C.J., et al. (2011). Development of SNAP-tag fluorogenic probes for wash-free fluorescence imaging. ChemBioChem *12*, 2217–2226.

Sydow, J.F., and Cramer, P. (2009). RNA polymerase fidelity and transcriptional proofreading. Curr. Opin. Struct. Biol. *19*, 732–739.

Taylor, N.O., Wei, M.T., Stone, H.A., and Brangwynne, C.P. (2019). Quantifying dynamics in phase-separated condensates using fluorescence recovery after photobleaching. Biophys. J.

Taylor, T.J., McNamee, E.E., Day, C., and Knipe, D.M. (2003). Herpes simplex virus replication compartments can form by coalescence of smaller compartments. Virology *309*, 232–247.

Teves, S.S., An, L., Hansen, A.S., Xie, L., Darzacq, X., and Tjian, R. (2016). A dynamic mode of mitotic bookmarking by transcription factors. Elife *5*, 066464.

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

Thomas, M.C., and Chiang, C.-M. (2006). The general transcription machinery and general cofactors. Crit. Rev. Biochem. Mol. Biol. *41*, 105–178.

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

Vannini, A. (2012). A structural perspective on RNA polymerase I and RNA polymerase III transcription machineries. Biochim. Biophys. Acta - Gene Regul. Mech. *1829*, 258–264.

Vekilov, P.G. (2010). Phase transitions of folded proteins. Soft Matter *6*, 5254–5272.

Velasco, E., Mederos, L., and Navascués, G. (2002). Phase Diagram of Colloidal Systems. Langmuir *14*, 5652–5655.

Vos, S.M., Farnung, L., Urlaub, H., and Cramer, P. (2018). Structure of paused transcription complex Pol II–DSIF–NELF. Nature *560*, 601–606.

Walther, N., Hossain, M.J., Politi, A.Z., Koch, B., Kueblbeck, M., Ødegård-Fougner, Ø., Lampe, M., and Ellenberg, J. (2018). A quantitative map of human Condensins provides new insights into mitotic chromosome architecture. J. Cell Biol. *217*, 2309–2328.

Wang, J., Choi, J.-M., Holehouse, A.S., Lee, H.O., Zhang, X., Jahnel, M., Maharana, S., Lemaitre, R., Pozniakovsky, A., Drechsel, D., et al. (2018). A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. Cell *174*, 688-699.e16.

Wang, Z., Millet, L., Mir, M., Ding, H., Unarunotai, S., Rogers, J., Gillette, M.U., and Popescu,

G. (2011). Spatial light interference microscopy (SLIM). Opt. Express *19*, 797–797.

Waters, T.R., and Swann, P.F. (1998). Kinetics of the action of thymine DNA glycosylase. J. Biol. Chem. *273*, 20007–20014.

Waters, T.R., Gallinari, P., Jiricnyl, J., and Swann, P.F. (1999). Human thymine DNA glycosylase binds to apurinic sites in DNA but is displaced by human apurinic endonuclease 1. J. Biol. Chem. *274*, 67–74.

Weber, S.C., and Brangwynne, C.P. (2015). Inverse Size Scaling of the Nucleolus by a Concentration-Dependent Phase Transition. Curr. Biol. *25*, 641–646.

Wei, M.T., Elbaum-Garfinkle, S., Holehouse, A.S., Chen, C.C.H., Feric, M., Arnold, C.B., Priestley, R.D., Pappu, R. V., and Brangwynne, C.P. (2017). Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles. Nat. Chem. *9*.

Weintraub, H., and Groudine, M. (1976). Chromosomal subunits in active genes have an altered conformation. Science *193*, 848–856.

West, J.A., Mito, M., Kurosaka, S., Takumi, T., Tanegashima, C., Chujo, T., Yanaka, K., Kingston, R.E., Hirose, T., Bond, C., et al. (2016). Structural, super-resolution microscopy analysis of paraspeckle nuclear body organization. J. Cell Biol. *214*.

Wiberg, K.B. (1955). The Deuterium Isotope Effect. Chem. Rev. *55*, 713–743.

Wilson, E.B. (1899). American Association for the Advancement of Science. Science *10*, 33–45.

Woodruff, J.B., Ferreira Gomes, B., Widlund, P.O., Mahamid, J., Honigmann, A., and Hyman, A.A. (2017). The Centrosome Is a Selective Condensate that Nucleates Microtubules by Concentrating Tubulin. Cell *169*, 1066-1077.e10.

Woringer, M., and Darzacq, X. (2018). Protein motion in the nucleus: from anomalous diffusion to weak interactions. Biochem. Soc. Trans. *46*, 945–956.

Woringer, M., Darzacq, X., and Izeddin, I. (2014). Geometry of the nucleus: a perspective on gene expression regulation. Curr. Opin. Chem. Biol. *20*, 112–119.

Wu, B., Eliscovich, C., Yoon, Y.J., and Singer, R.H. (2016). Translation dynamics of single mRNAs in live cells and neurons. Science aaf1084.

Xie, L., Dong, P., Qi, Y., Marzio, M. De, Chen, X., Banala, S., and Wesley, R. (2019). Super-resolution Imaging Reveals 3D Structure and Organizing Mechanism of Accessible Chromatin. BioRxiv 1–36.

Zhang, H., Elbaum-Garfinkle, S., Langdon, E.M., Taylor, N., Occhipinti, P., Bridges, A.A., Brangwynne, C.P., and Gladfelter, A.S. (2015). RNA Controls PolyQ Protein Phase Transitions. Mol. Cell *60*, 220–230.

Zhang, Z., Revyakin, A., Grimm, J.B., Lavis, L.D., and Tjian, R. (2014). Single-molecule tracking of the transcription cycle by sub-second RNA detection. Elife *2014*, 1–20.

Zheng, Q., Ayala, A.X., Chung, I., Weigel, A. V, Ranjan, A., Falco, N., Grimm, J.B., Tkachuk, A.N., Wu, C., Lippincott-Schwartz, J., et al. (2019). Rational Design of Fluorogenic and Spontaneously Blinking Labels for Super-Resolution Imaging. ACS Cent. Sci. acscentsci.9b00676.

Zhou, C., and Knipe, D.M. (2002). Association of herpes simplex virus type 1 ICP8 and ICP27 proteins with cellular RNA polymerase II holoenzyme. J. Virol. *76*, 5893–5904.

De Zitter, E., Thédié, D., Mönkemöller, V., Hugelier, S., Beaudouin, J., Adam, V., Byrdin, M., Van Meervelt, L., Dedecker, P., and Bourgeois, D. (2019). Mechanistic investigation of mEos4b reveals a strategy to reduce track interruptions in sptPALM. Nat. Methods *16*, 707–710.