ABSTRACT

Title of Dissertation: THE ROLE OF TRANSCRIPTION FACTOR DYNAMICS IN GENE EXPRESSION: DOES TIME MATTER?

David Alejandro Garcia Grisales,
Doctor of Philosophy in Physics
2021

Dissertation directed by: Dr. Arpita Upadhyaya
Department of Physics, UMD

Dr. Gordon L. Hager
National Cancer Institute, NIH

Different proteins and complexes work together at multiple time scales to orchestrate the activation and silencing of genes in a process called transcription. Understanding transcriptional regulation is of utmost importance to reveal mechanisms behind cell homeostasis and pathologies. The transcription machinery needs to be perfectly tuned in space and time to control the expression of genes to carry out cellular and physiological processes in the noisy and highly heterogeneous nuclear microenvironment. Transcription factors (TF), specialized proteins that bind to specific DNA sequences to regulate mRNA production, are central players in transcriptional regulation. TFs need to navigate the intricate nuclear microenvironment to bind to specific regulatory elements with binding times critically determining their regulatory functions. Recent advances in super-resolution microscopy have allowed us to investigate the dynamics of the transcriptional machinery at the single molecule level, revealing the essential features of transcriptional control. However, how TFs dynamically navigate the nuclear microenvironment and interact with chromatin to activate or silence genes remains poorly understood.

I used state of the art microscopy and genomic techniques to show that binding times of TFs to chromatin are power-law distributed. I proposed a new theoretical framework to demonstrate the broad distribution of binding affinity arises from heterogeneity in TF-chromatin interactions and the nuclear microenvironment, contrary to the current paradigm of well-defined and distinguishable TF binding times to specific and non-specific chromatin sites. These studies reconciled discrepancies between genomics, gene expression and TF mobility. I used statistical modeling to show that TFs exhibit two distinguishable low mobility states in the nucleus. One state is related to chromatin binding while the second arises due to protein-protein interactions mediated by intrinsically disordered regions of the TF and potentially controls the initiation rate of transcription. Finally, I studied transcriptional regulation on substrates of different stiffness revealing a connection between the physical properties of the cell microenvironment and TF dynamics. I demonstrated that substrate stiffness activates the estrogen receptor even in the absence of its ligand, with implications for our understanding and treatment of breast cancer. The evidence presented here shows that TF binding times are finely tuned to regulate gene expression.

THE ROLE OF TRANSCRIPTION FACTOR DYNAMICS IN GENE EXPRESSION: DOES TIME MATTER?

by

David Alejandro Garcia Grisales

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in
Physics
2021

Advisory Committee:
Professor Arpita Upadhyaya, Chair
Professor Michelle Girvan
Dr. Gordon L. Hager
Professor Wolfgang Losert
Professor Giuliano Scarcelli

Dedication

For everyone battling mental health problems. I hope you find solace during the dark night of the soul.

Acknowledgements

"Time is the substance I am made of. Time is a river which sweeps me along, but I am the river; it is a tiger which destroys me, but I am the tiger; it is a fire which consumes me, but I am the fire." Jorge Luis Borges

The work presented in this thesis would not had been possible without the immense support and assistance I have received during my graduate studies.

I would like to first thank my advisors, Prof. Arpita Upadhyaya and Dr. Gordon L Hager, whose scientific rigour, expertise and passion have inspired me to pursue the biggest challenges I have faced in science. They nourished my curiosity and supported my endeavors even in the most difficult circumstances. Thank you for believing in me.

I owe my molecular biology and genomics expertise to Prof. Diego M. Presman, Thomas Johnson and Dr. Louis R. Schiltz. I want to thank them for their patience and support. Selflessly, they shared their knowledge with me, and I will forever miss our conversations regarding biology from the eyes of a physicist.

All the present and past members in Upadhyaya's Lab and Hormone Action and Oncogenesis lab have been an important support inside and outside the lab. I already miss the passionate conversations about science, random discussions during lunch time and the camaraderie which made the lab my second home.

I never imagined a PhD to be such personally demanding journey. My parents have been the cornerstone in keeping my fire alive. I thank them for their support, unconditional love and being my light in the darkest nights. My sisters who have been a friend holding my hands while I toppled in this journey. Liyu He whose being my other self and his strength has kept my pieces together. Martha for being my PhD suffering partner and for always bringing smiles to my life. Finally, Betancur Tovar family for adopting me in a foreign land. To all my friends, thank you for always being by my side.

Table of Contents

List of Tables

List of Figures

List of Acronyms

ATAC	Assay for Transposase-Accessible Chromatin
CHIP	Transcription Factor
DBD	DNA Binding Domain
DGE	Differential Gene Expression
ECM	Extracellular Matrix
ER	Estrogen Receptor
GRE	Glucocorticoid Response Element
GR	Glucocorticoid Receptor
HILO	Highly Inclined and Laminated Optical sheet
HRE	Hormone Response Element
IDR	Intrinsically Disordered Region
LBD	Ligand Binding Domain
LCD	Low Complexity Domain
NTD	N Terminus Domain
PL	Power Law
SMT	Single Molecule Tracking
SR	Super Resolution
TADs	Topologically Associated Domains
TF	Transcription Factor
TIRF	Total Internal Reflection Fluorescence
PB	Photobleaching

Chapter 1.

Introduction

Images in Figures used with permission from the Journals.

1.1. Overview

Cellular function, fate, and survival are critically linked to the regulation of gene expression, which encompasses the transcription of DNA into RNA and the translation of RNA into proteins. Gene expression is strongly affected by environmental cues and its dysregulation is tied to many pathological disorders such cancer and diabetes. Despite decades of research on transcriptional regulation, and the discovery of the structure of DNA by Rosalind Franklin and co-workers, many questions regarding the regulation of eukaryotic gene expression remain unanswered and transcription remains a hot topic of research. The foundations of our understanding of transcription begin with early studies from bacterial models such Escherichia coli and bacteriophage λ (J. H. Miller & Reznikoff, 1980), culminating in more recent studies using advanced genomic methodologies and single molecule biophysics in eukaryotic cells (D. S. Johnson, Mortazavi, Myers, & Wold, 2007; Robb et al., 2013).

In eukaryotic cells, transcription takes place in the nucleus and is regulated by specialized machinery that finely tunes the expression of different genes in the structural and physical context of the nucleoplasm. Genomic DNA is highly compacted into chromatin which is composed of 146 base pairs of the genome wrapped around an octamer of histone proteins called nucleosomes (Kornberg & Thomas, 1974; Luger, Rechsteiner, Flaus, Waye, & Richmond, 1997). Chromatin organization remains unclear, but it is known that its structure varies across different length scales in cells providing

different organizational functions to chromatin. The nucleoplasm is divided into regions for each chromosome called chromosome territories (Meaburn & Misteli, 2008). Chromosomes are self-associated in regions dependent on cohesin mediated looping called topologically associated domains (TADs) (Bintu et al., 2018; Lieberman-Aiden et al., 2009; Rao et al., 2017). . Interphase chromosomes are organized into topologically associated domains (TADs) that are isolated from neighboring regions (Nora et al., 2012; Sexton et al., 2012) and exhibit a high frequency of inter-chromatin interactions. Moreover, membrane-less organelles whose formation is driven by dynamic protein-protein interactions further compartmentalize the nucleus at smaller scales (Klosin & Hyman, 2017), possibly in a reversible manner. TADs are enriched in chromatin binding proteins such remodelers, TFs and posttranslational modifications. Moreover, regions of chromatin are classified depending on their function such as enhancer, insulator, promoter, accessibility to protein factors (open vs closed chromatin) and compaction that regulates silencing or genes (euchromatin vs heterochromatin). Functional characterization of chromatin organization has been useful for understanding different regulatory processes in the cell, but a physical description of chromatin structure and organization is lacking **(Figure** 1-1).

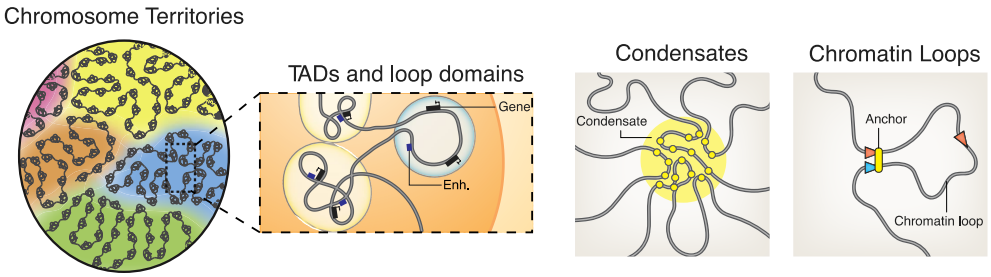


Figure 1-1 Chromatin Organization Across Length Scales
Chromatin is temporal and spatially organized at varying scales. Specific regions (Chromosome territories) occupied by particular chromosomes. Topologically Associated Domains (TADs) are Self-interacting genomic regions where DNA sequences exhibit significantly higher interaction frequency with other DNA sequences within a particular domain. Chemical interactions between proteins and nucleic acids are able to form membranelles organelles that further compartmentalized the nucleoplasm. Locally, transcription factors can stabilize interactions

between distant chromatin sequences forming chromatin loop. Adapted from (Hug & Vaquerizas, 2018; Stadhouders, Filion, & Graf, 2019).

Transcription factors, specialized proteins that control the transcription of genes by binding to specific DNA sequences, navigate this intricate genomic and nuclear landscape to find specific targets and regulate gene expression. How transcription factors find specific regulatory elements in the vast forest of genomic DNA organized within compartments with different biophysical properties and how they regulate transcription is still poorly understood. In this thesis, I use advanced microscopy and genomic techniques to explore transcription factor dynamics inside the nucleus, their interaction with chromatin and their role in gene regulation.

1.2. Transcriptional Regulation

Different proteins and complexes work together in the nucleus to regulate different genes in eukaryotic cells. How these molecules assemble, disassemble and interact with genomic DNA lies at the heart of gene regulation, but their interaction dynamics is still a mystery (Dundr et al., 2002; Hager et al., 2000). The temporal dynamics of all these different players set the time scales of the constituent processes, ranging from the order of seconds (oligomerization of transcriptional complexes, transcription factor binding) to transcriptional bursting over several minutes to hours (Chubb et al., 2006; Rodriguez et al., 2019) up to longer term patterns such as circadian oscillations (Metivier et al., 2003). Recent advances in genome wide techniques and single molecule biophysics are promoting a better understanding of transcriptional regulation *in vivo*.

TFs are master regulators of transcription; they bind to specific DNA sequences to initiate/inhibit and regulate the rate of transcription of particular genes (**Figure** 1-2). The interactions of transcription factors with chromatin are regulated at different levels inside

the cell. It is well accepted that the assembly of transcriptional machinery is initiated and stabilized by TFs and promoter-enhancer interactions. In order for the transcriptional machinery to assemble at specific gene loci, the chromatin must be remodelled and made accessible. Chromatin accessibility corresponds to the degree to which nuclear proteins are able to physically interact with genomic DNA in the face of different barriers such as DNA wrapping around nucleosomes and occupancy by different chromatin binding factors that exclude interactions by other macromolecules (Kaplan et al., 2009; Kornberg, 1974). Remodeling factors change chromatin accessibility by unwrapping nucleosomes in an ATP dependent manner while other factors change the biochemical properties of chromatin (post-translational modifications) to promote or inhibit transcription factor binding (**Figure** 1-2 **C,B,D**). The degree of accessibility is distributed broadly on a continuum across the nucleus, ranging from highly compacted and closed chromatin to highly dynamic and accessible chromatin (Poirier, Bussiek, Langowski, & Widom, 2008). The accessibility landscape is highly dynamic, reflects the regulatory capacity of the cell and is critical for transcriptional regulation and cell function. The accessible genome comprises approximately 2-3% of the total DNA sequence but it captures 90% of regions where TFs can bind (Thurman et al., 2012), providing evidence of elaborate regulatory mechanisms inside the cell.

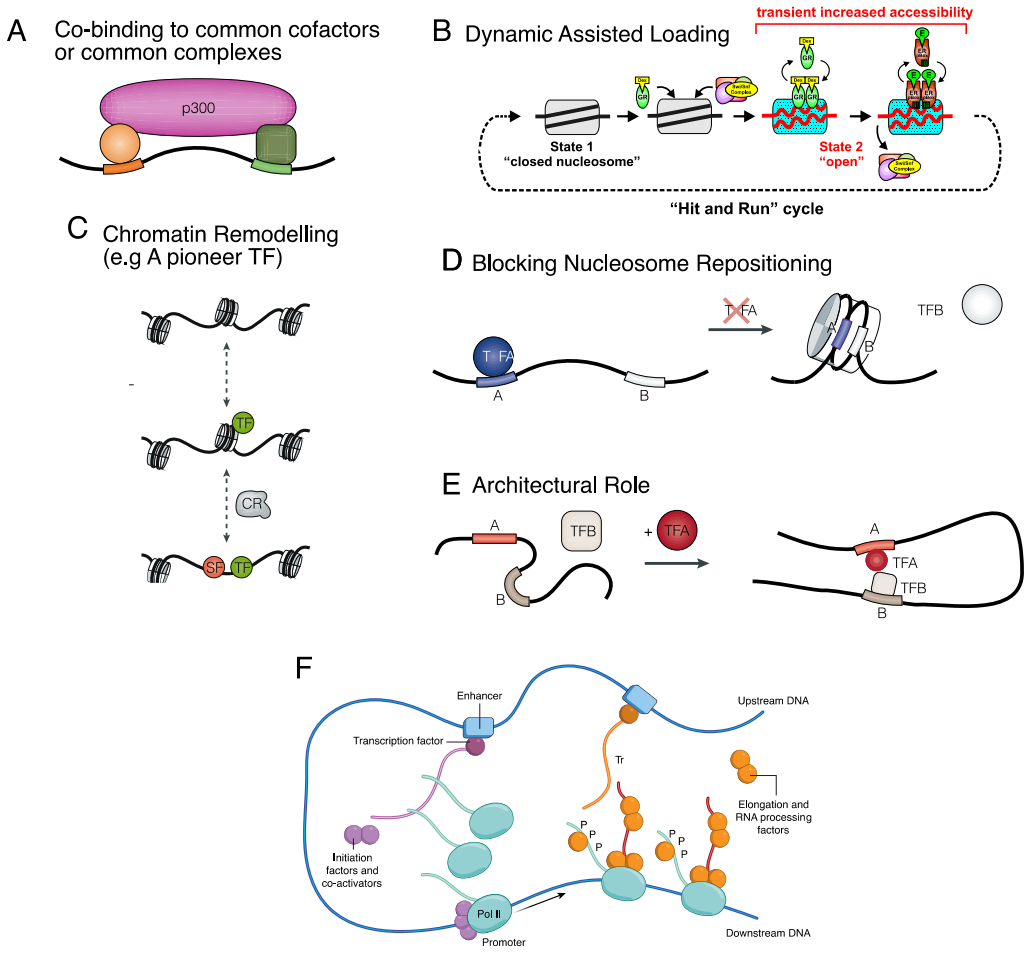


Figure 1-2 Transcription Factors and Their Interactions With Chromatin
Transcription factors exhibit different types of interactions with chromatin which affect their regulatory functions. (A) TFs can interact with multiple proteins (Cofactors) forming stable regulatory complexes. (B) Certain family of TFs are able to interact with closed chromatin and recruit remodeling factors in a dynamic fashion in a process called dynamic assisted loading. This process creates transient chromatin open states (ATP dependent) by cycling of TFs binding and recruitment of remodelers. (C) Pioneer TFs recruit remodelers at closed chromatin which increase chromatin accessibility in a process called wedging (ATP independent) (D) TFs can remain stably bound to specific chromatin sequences preventing nucleosome repositioning and facilitating binding of other TFs to the stably open site. This is a passive method of enhancer priming. (E) Architectural TFs form stable complexes between distant DNA sequences forming loop domains. (F) Through chromatin loop formation, enhancer-promoter interaction is facilitated where TFs bind at enhancers and cofactors at promoters facilitating transcriptional initiation and regulation. Adapted from (Cramer, 2019; Klemm, Shipony, & Greenleaf, 2019; Spitz & Furlong, 2012; Voss et al., 2011)

In addition to the local chromatin environment near the promoters, transcriptional regulation is finely tuned by spatiotemporal programs that span large stretches of chromatin. Transcriptional enhancers are regulatory DNA sequences with long-range interactions with their target-gene promoters spanning several thousands of bases

(**Figure** 1-2**F**) (Long, Prescott, & Wysocka, 2016; Plank & Dean, 2014). Mutations in enhancer sequences or genome rearrangement that disrupt enhancer-promoter interactions result in disease susceptibility and malformations (Lettice et al., 2003; Lupianez et al., 2015). The number of enhancers in mammalian cells far outnumber the number of genes, suggesting that multiple enhancer promoter interactions facilitate the transcription of a single gene (Sanyal, Lajoie, Jain, & Dekker, 2012), leading to combinatorial, context-dependent control. Promoters upstream of a transcriptional start site recruit transcription factors and RNA polymerase II to initiate transcription in the appropriate direction of the gene. Meanwhile, enhancers are bound by tissue specific transcription factors and cofactors such as Mediator and BRG1 (Long et al., 2016; Plank & Dean, 2014; Pombo & Dillon, 2015). The current hypothesis is that regulatory information and protein interactions are encoded in enhancer-promoter interactions to direct transcription (Carter, Chakalova, Osborne, Dai, & Fraser, 2002).

Chromatin Immunoprecipitation sequencing (CHIP-seq) provides the binding profile of proteins to chromatin at the population level at near base-pair resolution in cells and tissues (D. S. Johnson et al., 2007). Together with other genome-wide approaches (RNA-seq and ATAC-seq and HiC), comprehensive maps of regulation of different genes and transcription factors have been developed (Stamatoyannopoulos, 2012). However, these genome-wide techniques only provide a static snapshot of the highly dynamic processes that underlie transcription. Single molecule experiments of transcription factors have revealed binding kinetics to chromatin on the order of seconds to minutes and recent theoretical work suggests the use of statistical mechanics methods (broken detailed balance) to understand the asymmetry between protein-protein interactions during transcriptional regulation which has been elusive by genomic methodologies (Biddle,

Nguyen, & Gunawardena, 2019). However, transcription factor binding times are still a topic of debate.

1.3. Transcription Factors

To study the dynamic nature of transcriptional regulation, it is fundamental to understand the interactions of transcription factors with chromatin and the crowded nuclear microenvironment. Depending on their function, TFs are usually classified into four categories: 1) Architectural TFs responsible for chromatin topology through anchoring of loops (CTCF, YY1), 2) pioneer factors which remodel chromatin in an ATP independent manner, 3) complexes that require multiple TFs binding at different sites to initiate transcription and 4) TFs that dynamically interact with chromatin recruiting remodelers to reshape chromatin accessibility in an ATP dependent manner in a process called Dynamic Assisted Loading (**Figure** 1-2 **A-E**) (Goldstein et al., 2017).

One class of transcription factors that are the primary focus of this thesis are the nuclear receptors, which are inducible transcription factors with essential physiological functions. Their binding and interaction to chromatin and other cofactors can be modulated by external factors which make them a powerful tool to study the role of TFs in transcriptional regulation.

1.3.1. Nuclear Receptors

The nuclear receptor superfamily corresponds to the largest classification of eukaryotic transcription factors. It comprises 48 members in humans with a vast variety of biological functions, from controlling basic functions during homeostasis to playing crucial roles in pathologies such as cancer, immunosuppression, diabetes, autoimmune diseases, neurological disorders, cardiovascular disorders, premature ageing, metabolic

disorder, etc. (Hoffmann & Partridge, 2015; Kadmiel & Cidlowski, 2013; Lonai'd & O'Malley, 2012; Malek & Lad, 2014; Ranhotra, 2013) Therefore, understanding nuclear receptor biology remains of paramount importance in physiology and medicine.

Even though nuclear receptors have dramatically different functions in the cell, their structure is well conserved. They are composed of a highly variable amino-terminal domain referred as AF1 (activation function 1), a DNA-binding domain (DBD) and a ligand-binding domain (LBD)(Evans, 1988). The DBD is responsible for interacting with specific DNA sequences in chromatin, the LBD is usually lipophilic in nature and small ligands bind to it to change the conformation of the protein. The AF1 domain is responsible for interactions with other nuclear cofactors and further regulates the specific function of the nuclear receptor (Gustafsson, 2016; Rastinejad, Huang, Chandra, & Khorasanizadeh, 2013).

Nuclear receptors are further classified into four subfamilies based on their mode of action. Type I receptors, such as the glucocorticoid receptor, androgen receptor and progesterone receptor are anchored in the cytoplasm by chaperones (HSP90) in the absence of ligand. Upon ligand binding to the LBD, the chaperone is released, the nuclear localization signal (NLS) is exposed and the protein translocates to the nucleus, where it interacts with coregulators to alter gene expression (Sam et al., 2008). Type II receptors, such as the thyroid hormone receptor and the retinoic acid receptor, reside naturally in the nucleus in the absence of ligand. Unliganded type II receptors usually exert repressive functions through interactions with repressive complexes. Binding of the ligand leads to dissociation of corepressors and binding of coactivator complexes that facilitates opening of chromatin and promotes activation of target genes. (Glass & Rosenfeld, 2000) Type III and IV receptors function in a similar fashion to Type I receptors with the difference that type III receptors bind to Hormone Response Elements (HREs) with repeated sequence

and type IV receptors bind as monomers to half-site HREs (Mangelsdorf et al., 1995). Given the broad range of functions of nuclear receptors, their contribution to numerous diseases and their inducible function through small molecules (ligand), they provide promising therapeutic targets for engineering novel agonists and antagonists (Burris, Busby, & Griffin, 2012). However, a better understanding of their dynamics, functions and interactions is needed for future clinical applications.

1.3.2. Estrogen Receptor

Estrogens are a class of steroid hormones that regulate the estrogen receptor activity (ERα and ERβ), with ERα playing a major role throughout the body while ERβ plays a critical role in estrogen signaling in the ovary, prostate, lung and cardiovascular systems (Hamilton, Arao, & Korach, 2014). In humans, ERα is encoded by the gene ESR1. The estrogen receptor is involved in a wide range of physiological processes such as development and maintenance of reproductive organs, regulation of cardiovascular, musculoskeletal, immune, central system homeostasis and development of tissue malignancies (Gruber, Tschugguel, Schneeberger, & Huber, 2002; Nelson & Bulun, 2001). Estrogens are naturally synthesized in the ovaries, adrenal glands and adipose tissue. 17β-estradiol or estradiol (E2) is the most potent estrogen in humans and is used as a medication for menopausal hormone therapy and in hormone sensitive cancers like prostate and breast cancer (Yanfang, Hong, & Jing, 2020). Estrogens cross the plasma membrane and interact with intracellular ER (**Figure** 1-3). In the presence of estrogens, ER regulates the gene expression of target genes and the reprogramming of the chromatin landscape (**Figure** 1-3) (E.E. Swinstead et al., 2016). Recent studies have shown that ER is critical for cancer initiation and metastasis, extracellular matrix (ECM) remodeling and

drug resistance (Arpino, Wiechmann, Osborne, & Schiff, 2008; Q. Li, Gao, Yang, Wei, & Jiang, 2019; Piperigkou & Karamanos, 2020).

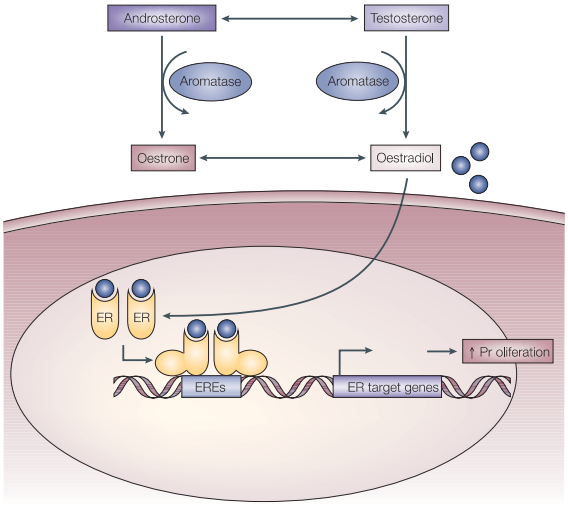


Figure 1-3 Estrogen Receptor Activation
Estradiol binds to the Estrogen Receptor promoting dimerization, conformational changes and binding to Estrogen Response Elements (EREs) to regulate gene expression of ER target genes. Estradiol is synthetized from androsterone and testosterone hormones through aromatase biosynthesis of androgenic precursors. Adapted from (Johnston & Dowsett, 2003).

ERα is composed of an amino-terminal domain (NTD) involved in transcription transactivation and binding to the target sequence due to the presence of a zinc-finger; a DBD that contributes to receptor dimerization and binding to specific sequence called estrogen response elements (ERE); and an LBD that binds to chaperones and regulates hormone activity (Kumar et al., 2011). ER also interacts with and regulates genes that do not contain EREs. This form of ER activity is collectively known as transcriptional crosstalk where ER interacts with other transcription factors and coregulators to activate and supress genes (Göttlicher, Heck, & Herrlich, 1998). Moreover, ER can induce loading of proteins such as FOXA1 to further affect gene expression (E.E. Swinstead et al., 2016).

1.3.3. Glucocorticoid Receptor

The Glucocorticoid receptor (GR) governs various aspect of inflammation, stress response, tissue development, metabolism and many other tissue specific processes (Galon et al., 2002; John et al., 2008; Z. Wang et al., 2003). GR is encoded by the nuclear receptor subfamily 3 group C member 1 gene (NR3C1) localized in the vicinity of its paralogues NR3C2, NR3C3 and NR3C4 genes that encode the mineralocorticoid receptor (MR), progesterone receptor (PR) and androgen receptor (AR) respectively. The DBD and LBD are highly conserved between GR, MR, PR and AR but the N terminus domain differs between these four proteins and conveys their specific functions (**Figure** 1-4**A**).

In its inactive form, GR is located in the cytoplasm associated with chaperone complexes containing the heat shock protein 90 (HSP90) and HSP70 (Picard et al., 1990) (**Figure** 1-4**B**). Upon binding of glucocorticoid hormones to the LBD, like cortisol in humans and exogenous drugs such as dexamethasone, GR conformation changes exposing the nuclear localization signal and it unbinds from chaperones, resulting in the nuclear translocation of the receptor (**Figure** 1-4**B**). GR activation results in the recruitment of coactivators and corepressors through protein-protein interactions and modifications of the chromatin landscape, which alter gene expression (Fryer & Archer, 1998; Muchardt & Yaniv, 1993; Spencer et al., 1997). GR is capable of regulating chromatin modifications such as acetylation or methylation and chromatin remodeling activities.

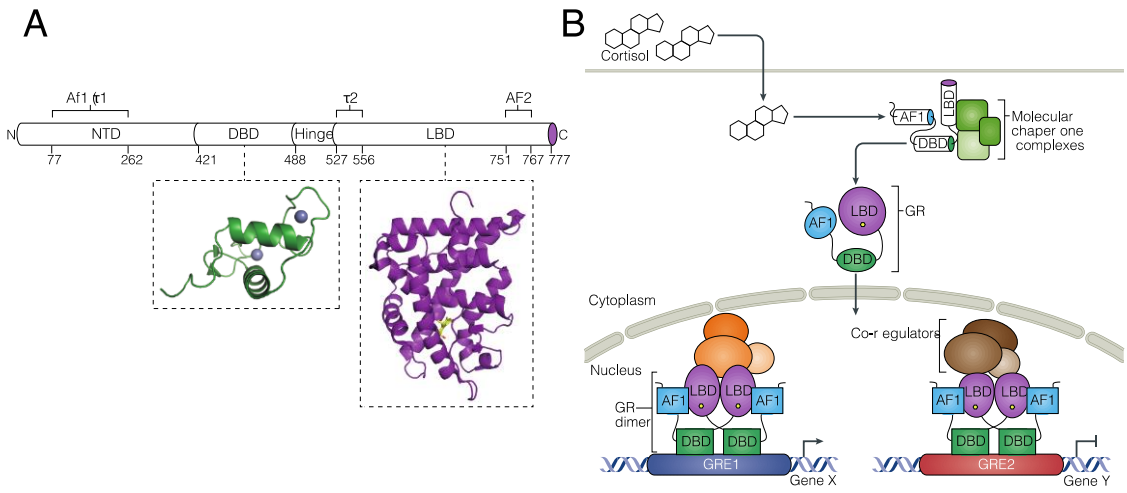


Figure 1-4 Structure and Action of the Glucocorticoid Receptor
(A) As a nuclear receptor, GR is composed by a N terminal domain with lacking 3D structure, a DNA Binding Domain (DBD) responsible for receptor oligomerization and a Ligand Binding Domain (LBD) responsible for chaperone sequestering and ligand dependent remodeling of the receptor. (B) In the absence of glucocorticoids (GR activating hormones), GR is in the cytoplasm sequestered by chaperone complexes. After binding of the hormone, GR is remodeled and translocated into the nucleus where it binds to GR Response Elements (GRE) to regulate transcription. Adapted from (Weikum et al., 2017).

GR binds to specific DNA sequences called glucocorticoid response elements (GREs) at promoter or enhancer regions to cause transcriptional induction of target genes. GR coregulates a myriad of transcription factors such as activator protein-1 (AP-1) and nuclear factor-kappa B (NF-kB) to trans-repress genes (De Bosscher, Vanden Berghe, & Haegeman, 2006; Luecke & Yamamoto, 2005; Nissen & Yamamoto, 2000). Recent studies have shown that GR is dimeric when unbound from DNA but tetrameric when bound to chromatin, and its oligomerization state plays a fundamental role in gene regulation (Presman et al., 2016).

It has been shown that unremodeled chromatin presents a barrier to transcriptional initiation by GR. It has been proposed that the chromatin landscape mediates cell and tissue specificity in response to GR activation. For instance, GR mediated transcriptional regulation is lacking in a pituitary corticotroph cell line with absent hormone inducible and constitutive sites (John et al., 2008).

It is well accepted that GR dynamics and its dynamical interaction with different complexes are of paramount importance to GR activity. However, GR dynamics remains poorly understood. GR undergoes rapid exchange at chromatin in a dynamic fashion instead of the long-term occupancy at promoter by the formation of multi-factor complexes as previously thought (McNally, Mueller, Walker, Wolford, & Hager, 2000). This rapid cycling of the GR with chromatin is actively regulated by cofactors disengaging from chromatin and ATP dependent chaperone activity (Agresti, Scaffidi, Riva, Caiolfa, & Bianchi, 2005; Becker et al., 2002; Elbi et al., 2004; Stavreva, Muller, Hager, Smith, & McNally, 2004). Recently, a clear correlation between GR dynamics and transcriptional activity at artificial genes has been demonstrated (Stavreva et al., 2019), but further experiments are needed for a complete view of the dynamical regulation of GR.

1.4. Transcription Factors Mobility and its Relation with Gene Regulation

Transcriptional activation is a dynamical process facilitated and regulated by TFs binding at regulatory elements on chromatin, changing the rate of gene transcription (**Figure** 1-5) (Shlyueva, Stampfel, & Stark, 2014). TFs binding at enhancers has been regarded as the fundamental mechanisms of gene regulation, where enhancer-promoter interactions are fundamental for the expression of the vast majority of genes (Levine, Cattoglio, & Tjian, 2014). For transcription to occur, many dynamical processes need to happen with precise temporal resolution. TFs binding to enhancers initiates the recruitment of different complexes, alteration of local chromatin accessibility, stabilization of chromatin loops, and initiation of transcription. Even though, genomic studies have shown the interplay of these different factors to orchestrate transcriptional activation (Barski et al., 2007; Buenrostro, Giresi, Zaba, Chang, & Greenleaf, 2013; Lieberman-Aiden et al., 2009), many questions remain unanswered. What is the timescale of TF

binding to regulatory elements? How are multiple TF recruited and assembled at enhancers? What are the dynamics underlying enhancer-promoter communication, etc.? (**Figure** 1-5)

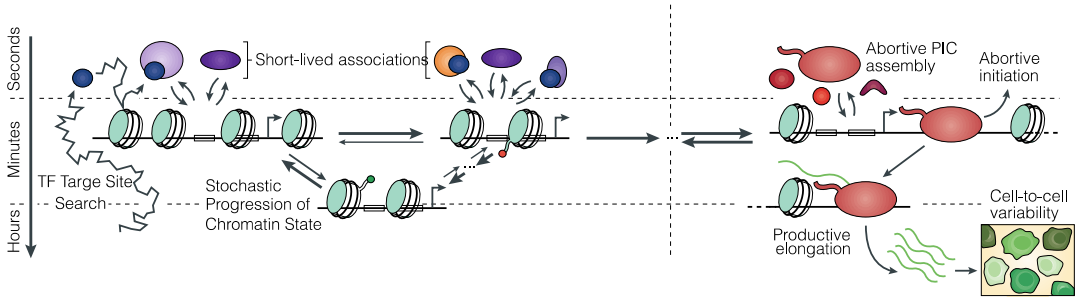


Figure 1-5 Stochasticity in Transcription Factor Dynamics and Gene Regulation
Transcription is intrinsically stochastic, and it is regulated at different time scales. TFs move across the nucleoplasm finding specific binding sites where it associates with chromatin and cofactors in the order of seconds to minutes. Interaction of TFs with other proteins promotes or inhibits recruitment of the transcriptional machinery which needs to be assemble for productive transcription. Every step in gene expression requires the stochastic interaction of proteins and molecules resulting in heterogeneity and noise in gene expression. Adapted from (Coulon, Chow, Singer, & Larson, 2013).

Seminal work using FRAP (Fluorescence Recovery After Photobleaching) showed that core histones have slow exchange rates on the order of ~1-2h with a small fraction exhibiting fast kinetics, implying that there is a population of stable core histones in nucleosomes and dynamical histone exchange takes place on the surface of active nucleosomes (Kimura & Cook, 2001). Moreover, heterochromatin was shown to display remarkably transient residence times (~20 secs) while nuclear receptors such ER and GR were shown to have extremely short binding times on the order of seconds (Hager et al., 2000; McNally et al., 2000; Voss et al., 2011). However, FRAP experiments are highly model dependent and these results have been disputed (Mazza, Stasevich, Karpova, & McNally, 2012b).

Understanding the biophysical determinants of the mobility of TFs and their interaction with the nuclear microenvironment and assessment of the role of these processes in gene regulation needs high-resolution measurements of their dynamics

complemented by the development of new theoretical frameworks to account for empirical data. Single molecule tracking (SMT) of TFs is a powerful tool to reveal their dynamics throughout the intricate nuclear environment since it allows direct visualization of TFs dynamics. SMT and super resolution microscopy have changed how we study molecular process in fundamental biological questions, from the kinetics of molecular motors to enzymatic reactions inside cells (Paakinaho et al., 2017; Sengupta, van Engelenburg, & Lippincott-Schwartz, 2014; Wedeking et al., 2015). It continues to change different paradigms in biology and physics with profound effects on our understanding of dynamical processes needed for homeostasis and cell functioning. Single molecule studies shifted the paradigm of TFs mobility and gene regulation. For decades, it was believed that TFs bind to genomic loci on the order of minutes to hours and these prolonged time scales allowed for protein-protein interactions and complex initiations which are needed for gene control (Coulon et al., 2013; Liu et al., 2014; Mueller, Stasevich, Mazza, & McNally, 2013; Thanos & Maniatis, 1995). This paradigm was compatible with population assays such as chromatin immunoprecipitation (ChIP). On the other hand, single molecule studies have shown that TFs are highly mobile, diffusing in the nucleoplasm with transient interactions with chromatin characterized by binding times on the order of seconds (Hager, McNally, & Misteli, 2009; Paakinaho et al., 2017; Yu, 2016). The latter implies a highly dynamic transcriptional initiation program with multiple TFs and cofactors exchanging on the time scale of transcriptional output.

Most dynamical studies of TFs have been studied in simplified model systems or in isolation without taking into account their biophysical interactions with the nuclear microenvironment. To obtain a fuller picture of eukaryotic gene regulation, TF dynamics needs to be studied in the context of their interaction in the complex milieu of the nuclear

microenvironment with ongoing chromatin remodeling, protein-protein interactions, nuclear compartmentalization and dynamical interactions between complexes.

1.5. Spatial Compartmentalization of the Nucleus

Compartmentalization in eukaryotic cells provides spatiotemporal control over different cellular processes from signaling pathways to metabolism. For instance, the transcriptional machinery is physically separated in the nucleus from other processes such as translation. This separation allows to cells increase control in translational output and posttranscriptional processes (Corbett, 2018). Studies have focused on compartmentalization based on membranes such as lysosomes, vesicles, endoplasmic reticulum and others. However, molecules, proteins and DNA can self-organize in membrane-less structures, which are sometimes referred as membrane-less organelles. Nucleoli, nuclear speckles, stress granules, and Cajal bodies are some examples of membrane-less organelles that have been studied for decades (Mitrea & Kriwacki, 2016). However, many questions remain in terms of their physicochemical properties, dynamics, formation and how these physical properties affect their function. The importance of membrane-less organelles in disease is being explored and recent evidence has pointed to the roles of these organelles in age-related disorders (Boeynaems et al., 2017).

Most membrane-less organelles are dynamic, assembling and disassembling over a range of time-scales. P granules (perinuclear RNA granules) were shown to have liquid-like properties and are formed by phase separation (Brangwynne et al., 2009). With this discovery, studies to understand phase separation and the formation of membrane-less organelles rapidly proliferated. Phase separation has been studied in physics and polymer chemistry for years but its properties and formation in biological materials needs further development. Even though liquid-liquid phase separation has been extensively invoked to

underlie the formation of cellular membrane-less organelles, these assemblies are not van der Waals fluids comprised of spherical particles interacting via isotropic short-range potentials. Phase separation in protein-protein, RNA-protein or DNA-protein mixtures have distinct physical properties depending on their constituents, interactions and concentrations (Feric et al., 2016). The global material properties of these aggregates range from liquid, solids and gels. For instance, recent studies showed that network elasticity regulated liquid-liquid phase separation in synthetic polymer networks. In particular, compressive stresses in a polymer network were shown to inhibit phase separation of the solvent (Rosowski et al., 2020; Style et al., 2018).

1.5.1. Intrinsic Disordered Regions and Liquid-Liquid Phase Separation

It is accepted in the field that the formation of membrane-less organelles is driven by multivalent interactions between nucleotides or amino acids (**Figure** 1-6). A large number of proteins contain polypeptide segments that are unfolded and lack well-defined tertiary structure (Wright & Dyson, 1999). These polypetide regions adopt an ensemble of different conformations and the resulting disordered states are believed to have specific biological functions. Such intrinsically disordered regions (IDRs) or low complexity domains (LCDs) do not contain sufficient amino acids to mediate cooperative folding. They typically contain a high proportion of charged amino acids that facilitates electrostatic interaction among other IDRs. Due to the structural flexibility of IDRs, many post-translational modifications are encoded and decoded in these regions and it is a driving factor in the assembly and disassembly of IDR mediated condensates (**Figure** 1-6). Phosphorylation of proteins changes the electrostatic balance of a protein due to the negative charges on the phosphate group (Aumiller & Keating, 2016). IDRs mediate important regulatory functions in proteins, facilitating interactions between proteins and

complexes, regulation of protein half-life by engaging targeted degradation by the proteasome and adoption of different conformations when binding to different partners to tune the protein function (Babu, Kriwacki, & Pappu, 2012). Recent studies have shown that most proteins in mammalian cells contain disordered regions and mutations in these segments have been related with human diseases, such as cancer (Babu, van der Lee, de Groot, & Gsponer, 2011).

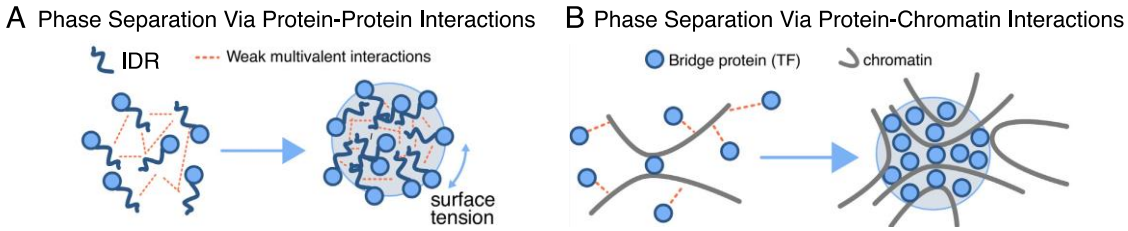


Figure 1-6 Phase Separation Via Intrinsic Disordered Regions
(A) Intrinsic Disordered Regions (IDRs) of proteins unstructured sequences interact with other IDRs through weak multivalent interactions forming phase separated droplets which compartmentalized the nucleoplasm. (B) Protein-chromatin stable binding mediates self-aggregation of chromatin regions through IDR of proteins and histones tails. Adapted from (Michieletto & Gilbert, 2019).

In vitro studies have shown that purified IDRs demix or show liquid-liquid phase separation in a concentration or temperature dependent manner (Kato et al., 2012). These droplets exhibit liquid-like properties like nucleation and surface tension similar to phase separated liquids. However, translating these in vitro experiments to biological settings have been challenging due to the complex microenvironments inside the cells, which are crowded with different molecules, macromolecules and complex structures. The properties of liquid condensates in cells have been measured indirectly using fluorescence correlation spectroscopy to quantity droplet viscosity and diffusion (M. T. Wei et al., 2017) and optogenetic tools are being developed to trigger phase transitions (Shin et al., 2017).

1.5.2. Nuclear Hubs for Transcriptional Control

Functional characterization of chromatin organization has been useful for understanding different regulatory processes in the cell, but a physical description of chromatin structure and organization is lacking. It has long been recognized that histones have disordered tails that are exposed in nucleosomes that are positively charged and serve for a myriad of posttranslational modifications (acetylation, methylation and phosphorylation). It is predicted that approximately half of the negative charges in DNA are neutralized by histones tails conveying an overall charge to nucleosomes (Kornberg & Lorch, 1999). Early in vitro experiments showed self-association of chromatin promoted by free cations on histone tails (J. C. Hansen, 2002). Moreover, reconstituted chromatin in solutions with physiological cation concentrations aggregates into spherical hubs with liquid-like properties and chemical simulations suggest compartmentalization of chromatin is driven by liquid-liquid phase separation, suggesting new physical models for chromatin organization (Falk et al., 2019; Gibson et al., 2019; Maeshima, Ide, Hibino, & Sasai, 2016).

The interaction between histone tails and DNA drives liquid-liquid phase separation (LLPS) in chromatin (**Figure** 1-6). LLPS is heavily mediated by linker histone H1 that is consistent with previous experiments where H1 depletion causes chromatin unfolding and in vitro experiments of H1-DNA condensates (Allan et al., 1981; Turner et al., 2018). H1 increases the concentration of nucleosomes within chromatin droplets and decreases their dynamics. Moreover, histone acetylation causes disassembly of the droplets, but highly acetylated chromatin coalesces through multi-bromodomain proteins binding such as the transcriptional regulator BRD4. Highly acetylated droplets do not mix with non-acetylated chromatin providing a functional compartmentalization mediated by LLPS (Gibson et al., 2019).

There is increasing evidence that LLPS is a driving force in creating chromatin microenvironments that facilitate chromatin function. It has been shown that RNA polymerase II, transcription factor FET and the mediator complex subunit 1 (Med1) form clusters in the nucleus reminiscent of LLPS (**Figure** 1-7) (Boehning et al., 2018; Chong et al., 2018; Sabari et al., 2018). These condensates are highly dynamic with selective interactions mediated by IDRs of component proteins. Surprisingly, there are families of IDRs that interact among each other and work as isolators for other IDRs (Chong et al., 2018). In terms of transcriptional initiation, recent studies suggest that Pol II CTD (C terminal domain) plays a major role in hub formation to selectively associate with transcription factors and to initiate transcription (**Figure** 1-7) (Guo et al., 2019). How TFs interact with condensates and their role in chromatin binding dynamics is still unknown. Further experiments are needed to understand how protein-protein interactions affect transcription factor mobility and binding and their relation to transcriptional control.

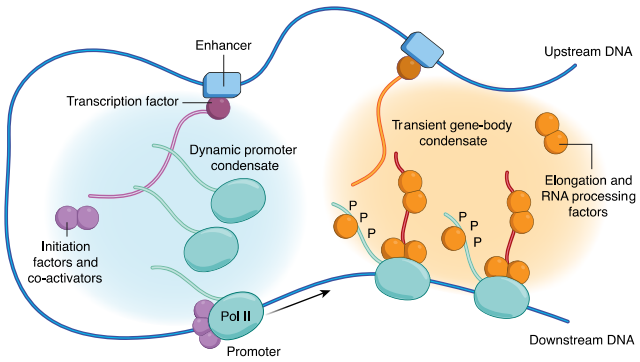


Figure 1-7 Phase Separation and Transcriptional Control
Evidence has emerged in the ability of TFs to self-interact and form hubs of increase TF density. These hubs might potentiate transcriptional initiation by increasing the probability of TF-chromatin interactions and facilitating TF and cofactor interactions through promoter enhancer action. Pol-II forms hubs as precursor of RNA processing facilitating gene expression. Adapted from (Cramer, 2019).

1.6. Mechanosensing

Adapted from: Mechanical regulation of transcription: Recent advances. K. Wagh, M. Ishikawa, D.A. Garcia, A. Upadhyaya and G. Hager. Trends in Cell Biology. (Wagh et al., 2021)

Most studies of the role of TFs in transcriptional regulation have been carried out on rigid substrates (glass and plastic) in the laboratory. However, the natural microenvironment of cells and tissues in different organs in the body is significantly softer and has a wide stiffness range. Stiffness, the extent to which a material resists deformation in response to an external force, is an important property of organs with fundamental implications in cell fate and homeostasis. Mounting evidence has shown that nuclear organization, morphology and transcription are intimately related to the physical properties of the environment. To further understand TFs mobility and their role in transcriptional regulation, studies must be carried out in substrates with physiologically relevant physical properties.

Cells in the human body are subject to a wide variety of mechanical stimuli acting at multiple scales. At the single molecule level, receptors leverage force to strengthen bonds with their cognate ligands (C. Zhu, Chen, & Ju, 2019). On the scale of single cells, stem cells rely on mechanical cues to guide cell fate decisions (Engler, Sen, Sweeney, & Discher, 2006). Finally, collective processes such as wound healing, tumorigenesis, and tissue homeostasis are intimately linked with the physical microenvironment (Cox & Erler, 2011) (Handorf, Zhou, Halanski, & Li, 2015). In order to engage in functional responses appropriate to both passive mechanical stimuli, such as stiffness or topographic features of the cellular environment, or active ones such as forces generated by cells and tissues, cells must be able to sense and measure mechanical perturbations. Different elements of the cell act in concert to maintain structural integrity and to coordinate cellular sensing of external forces and mechanical stimuli. These stimuli must then be subsequently transmitted to the nucleus leading to broad changes in chromatin structure and accessibility that is necessary for the induction of functionally appropriate gene expression programs (Uhler & Shivashankar, 2017).

The nucleus, while being the largest and stiffest organelle in the cell, is not isolated from the cytoskeleton and the forces thereof (**Figure** 1-8). The Linker of Nucleoskeleton and Cytoskeleton (LINC) complex, consisting of nesprins, and Sad1 and UNC-84 (SUN) proteins, connects the nucleus to the cytoskeleton (**Figure** 1-8). Nesprins, which contain a Klarsicht, ANC-1, Syne Homology (KASH) domain, span the outer nuclear membrane (ONM) and interact with the cytoskeleton on the cytoplasmic face of the ONM (Roux et al., 2009) (**Figure** 1-8). On the other side of the ONM, nesprins bind to SUN proteins that tether to the inner nuclear membrane and bind to the nuclear lamina. The nuclear lamina consists of A and B type lamins, whose expression levels can alter nuclear stiffness, which has been shown to scale with ECM stiffness (Swift et al., 2013) and cell geometry (Makhija, Jokhun, & Shivashankar, 2016). Forces exerted on the cell surface are transmitted to the nucleus through the LINC complex, which in turn, dictates the physical properties of the nucleus (Guilluy et al., 2014) as well as transcriptional response to mechanical stimuli (Alam et al., 2016). The role of the nuclear membrane-associated proteins as intermediaries in force transduction has been recently reviewed in (Agrawal & Lele, 2019).

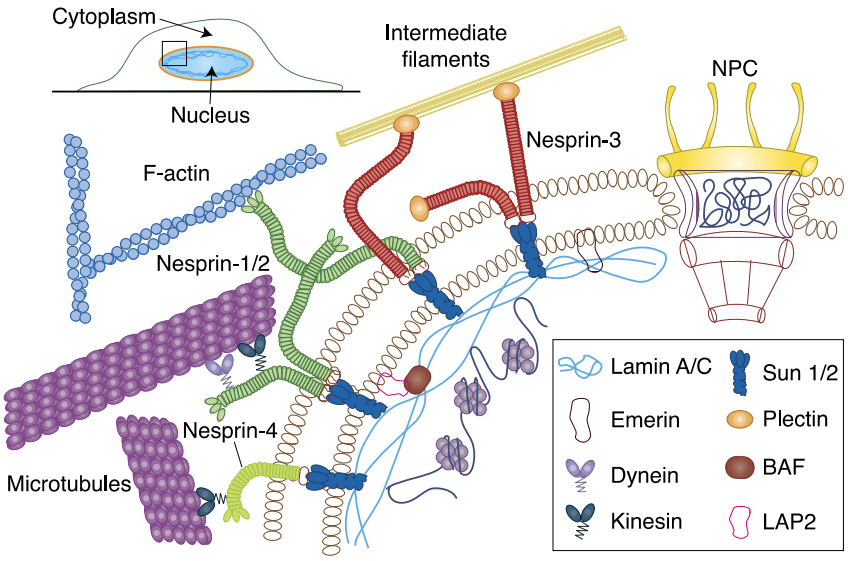


Figure 1-8 Cell Mechanosensing and Chromatin
Chromatin and nucleus is connected through the cytoskeleton to the cell substrate. The LINC complex formed by SUN and KASH proteins connect the cytoskeleton with chromatin. Microtubules, Actin polymers and intermediate filaments are interconnected with chromatin through the LINC complex with potential stress transmission during cytoskeleton remodeling. Cytoplasmic stresses might regulate chromatin accessibility by stretching of nucleosomes. Adapted from (Kirby & Lammerding, 2018).

Cells in high stiffness environments likely experiencing increased mechanical stress tend to present nuclear YAP/TAZ. Nuclear YAP/TAZ serve as co-regulators for several transcription factors leading to cell proliferation, organ growth, and tumorigenesis (Panciera, Azzolin, Cordenonsi, & Piccolo, 2017). Focal adhesion kinase (FAK) has been implicated in regulating YAP nuclear localization via the FAK-Src-PI3K pathway (N. G. Kim & Gumbiner, 2015). YAP/TAZ regulation has been extensively reviewed (Totaro, Panciera, & Piccolo, 2018) and YAP/TAZ nuclear translocation is now often used as a reporter of mechanotransduction.

Importantly, endocrine signaling pathways have recently been implicated in YAP/TAZ signaling. Sorrentino et al., showed that induction of the glucocorticoid receptor (GR) leads to increased YAP mRNA levels, and increased nuclear localization and YAP-luciferase reporter activity (Sorrentino et al., 2017). This is accompanied by increased

fibronectin deposition, further implicating the FAK-Src-PI3K axis in regulating YAP. The Hippo signaling pathway is also involved in estrogen receptor (ER) and androgen receptor (AR) signaling, which are important molecular targets for breast and prostate cancer respectively (Lit et al., 2013) (Kuser-Abali, Alptekin, Lewis, Garraway, & Cinar, 2015).

1.6.1. Mechanical Regulation of Tumorigenesis

Tumor initiation and progression are accompanied by changes in the ECM and cellular organization which develop different responses to mechanical cues (Butcher, Alliston, & Weaver, 2009). Understanding the mechanobiology of tumorigenesis is crucial to understanding its role in cancer development and metastasis. A ubiquitous process in tumorigenesis is ECM stiffening, which is a complex biochemical and biophysical process involving various types of cells, structural proteins, enzymes, and physical forces. Tumor cells carrying damaged DNA secrete inflammatory cytokines and matrix remodeling enzymes which recruit fibroblasts and immune cells to the tumor initiation site (Poltavets, Kochetkova, Pitson, & Samuel, 2018). Under physiological conditions, fibroblasts are involved in organ development and wound healing by depositing and remodeling ECM components. However, cancer associated fibroblasts (CAFs), which are activated by a variety of biophysical and biochemical stimuli in the tumor microenvironment, deposit excess ECM components, deregulate proliferation of the surrounding cells, and contribute to an imbalance in tissue homeostasis (Sahai et al., 2020). Cancer associated immune cells induce further inflammatory signaling befitting cancer's description as "wounds that do not heal" (Flier, Underhill, & Dvorak, 1986). The hypoxic environment created due to locally elevated cell density and metabolism facilitates lysyl oxidase (LOX) expression, which leads to elevated collagen crosslinking, thereby creating a dense ECM (**Figure** 1-9) (Bonnans, Chou, & Werb, 2014). In addition to morphological changes such as loss of polarity, cell-cell adhesion and acquisition of mobility, changes in gene expression during

EMT trigger deposition of ECM components including fibronectin and fibrillin, thus contributing to ECM rigidity (**Figure** 1-9) (S. C. Wei & Yang, 2016) (Baldwin et al., 2014). Actomyosin driven cell contractility can further enhance ECM stiffness leading to mechanical feedback between cells and ECM, thus elevating tension in the tissue (Yu Long Han et al., 2018) (Hall et al., 2016) (Van Helvert & Friedl, 2016).

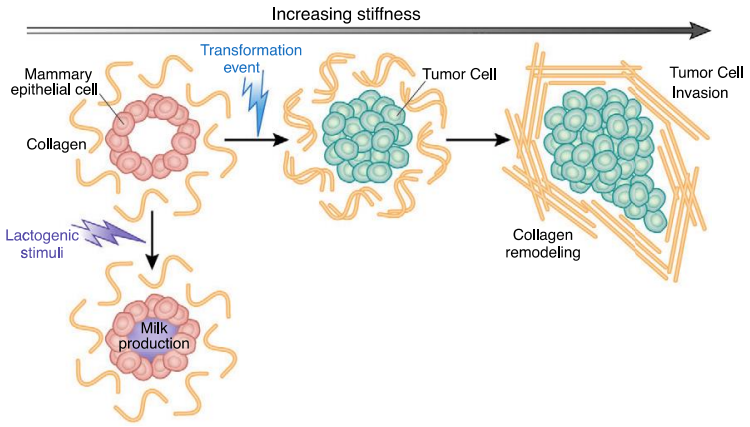


Figure 1-9 Stiffening During Breast Tumorigenesis
Mammary epithelial cells in normal breast tissue are organized in polarized acini structure with the capability of producing milk with lactogenic stimuli. During breast tumorigenesis, cells are transformed, and the polarized acini is lost. Cellular matrix is remodeled, and invasiveness is increased. Cellular transformation is correlated with matrix stiffening. (Cox & Erler, 2011).

Interestingly, stiffness measurements of human breast biopsies have revealed that the periphery of the tumor is stiffer than the core (Plodinec et al., 2012), suggesting that the gradual increase of stiffness towards the edge facilitates tumor cell invasion of the surrounding tissue (S. C. Wei & Yang, 2016). The collective cell durotaxis model, where a sheet of cells migrates towards higher stiffness gradients while maintaining cell-cell junctions, implies a systematic invasion of tumor cells into the stiffened ECM (Sunyer et al., 2016) (Sunyer & Trepat, 2020). However, different biophysical techniques have revealed that individual cancer cells are softer than healthy cultured cells (Guck et al., 2005; Rosenbluth et al., 2020). Softening of cancer cells is associated with malignant transformation and it is related with metastasis (Cross, Jin, Rao, & Gimzewski, 2007).

1.6.2. In-vitro Models for ER+ Breast Cancer

Breast cancer continues to pose a significant threat to women across the globe. About 1 in 8 women in the U.S. will develop invasive breast cancer over the course of her lifetime (Desantis, Ma, Bryan, & Jemal, 2013). For women in the U.S., breast cancer death rates are higher than those of any other cancer besides lung cancer. However, several advancements in prevention, screening, and treatment of breast cancer have been developed throughout the last decades. Despite the recent advancement in understanding breast cancer biology, breast cancer tumorigenesis and treatment are still poorly understood.

The complexity of studying breast cancer lies in its highly heterogeneous composition, encompassing a group of genetically and epigenetically distinct diseases that exhibit a broad range of clinical features (Riaz et al., 2013). Most research has been performed *in vivo* and *in vitro* using breast cancer cell lines due to the reproducibility and standardization of different models (Lacroix & Leclercq, 2004). However, cell line studies have been performed in artificially controlled environments and thus whether these studies capture the biological features of tumors remains an important issue to be resolved.

Hormone receptors are important drivers of breast cancer progression and phenotype. Clinically relevant classification of breast tumor is based on the expression levels of hormone receptors and human epidermal growth factor receptor-2 (HER2). Breast cancer is classified in three categories: hormone receptor positive (HR+), human epidermal growth factor receptor-2 amplified/overexpressed (HER2+) and triple negative (TN). Hormone receptor positive tumors include progesterone receptor positive (PR+) and/or estrogen receptor positive (ER+). HER2+ presents overexpression of HER2 receptor by dysregulation of the ERBB2 gene and TN breast cancer presents decreases

expression of ER, PR and HER2 (Jameson et al., 2018). ER+ breast cancer presents the highest incidence of breast cancer diagnosed in females (Nadji, Gomez-Fernandez, Ganjei-Azar, & Morales, 2005).

The number of cell lines widely used in breast cancer studies is small compared to the heterogeneity of breast tumors. MCF7, T47D and MDAMB231 account for more than 70% of cell lines used in the breast cancer biology studies (Lacroix & Leclercq, 2004). MCF-7 is an important breast cancer model in research for ER+ breast cancer, many subclones have been stablished representing different tumor phenotypes of varying nuclear receptor expression levels (Sweeney, Mcdaniel, Maximov, Fan, & Craig, 2013).

MCF-7 is ER and PR positive and it belongs to the luminal A subtype, it has low metastatic potential and it has produced more practical knowledge that any other breast cancer cell line. It is composed by a large number of different phenotypes with different gene expression profile, receptor expression and signalling pathway (Burdall, Hanby, Lansdown, & Speirs, 2003; Gest et al., 2013; Shirazi, 2011; Sweeney et al., 2013). MCF-7 cells are an important model in research for ER positive breast cancer biology, in particular in the investigations of anti-estrogen drug resistance since they retain ER expression even under aggressive drug treatment (Sweeney et al., 2013).

1.7. Overall Plan of The Thesis

How TFs move and interact with the complex nuclear microenvironment in different physiological contexts is of fundamental importance to understanding transcriptional regulation with overarching consequences from fundamental understanding of the principles that govern cellular fate to medical applications. I use super resolution microscopy and novel analytical methods to elucidate how TFs mobility depends on the local nuclear microenvironment, TFs properties and their interaction with other complexes

and how the physical properties of the external microenvironment regulate gene regulation and TFs dynamics. I found that TFs mobility and gene regulation is deeply affected by the nuclear and external microenvironment. Specific and non-specific binding times are power-law (PL) distributed with binding times on the order of seconds to minutes due to a broad distribution of binding affinities and the heterogeneity of the nuclear microenvironment.

In Chapter 2, I discuss general methods to study TF dynamics and gene regulation. Chapter 3 will discuss a new method that I have developed to analyze single molecule dynamics of TFs that reveals power-law distributed dwell times which has changed the current paradigm of TFs interactions with specific response elements. Chapter 4 will illustrate how this new method can be used to illustrate the complexity of the environment within the nucleus and how protein-protein interactions of TFs and interacting partners lead to a dynamic compartmentalization that alters mobility and gene regulation. In Chapter 5 I discuss my studies on the effects of stiffness in gene regulation, TFs mobility, nuclear organization and dynamics with fundamental physiological consequences. In Chapter 6, I provide general conclusions from my work and future directions.

Chapter 2. Techniques

Images in Figures used with permission from the Journals.

2.1. Genomic techniques

Mammalian chromatin is highly condensed in fundamental units called nucleosomes, composed of ~146 base pairs (bp) of DNA wrapped around a histone octamer (Kornberg, 1974). Nucleosomes are further wrapped into chromatin forming chromosomes. This dense chromatin is dynamically regulated by different transcription factors, remodelers and complexes. Chromatin is remodeled into active euchromatin and inactive heterochromatin (Falk et al., 2019), which tightly controls gene expression and guides development and differentiation (Goldberg, Allis, & Bernstein, 2007). Chromatin packaging influences transcription by allowing or preventing interactions of proteins with DNA and by modifying nucleosomes to enhance or prevent the recruitment of complexes that facilitate transcription. Therefore, understanding the regulation of chromatin states and the impact on protein binding to DNA is vital for understanding gene regulation.

With the technological advances of high-throughput sequencing, various assays have been developed to map the epigenetic landscape and the transcriptome. Techniques such as Assay of Transposase Accessible Chromatin sequencing (ATAC-seq) (Buenrostro et al., 2013; Buenrostro, Wu, Chang, & Greenleaf, 2015), DNAse I hypersensitive sites sequencing (DNAse-seq) (Song & Crawford, 2010) and Formaldehyde-Assisted Isolation of Regulatory Elements

sequencing (FAIRE-seq) (Giresi, Kim, McDaniell, Iyer, & Lieb, 2007) have been developed to map chromatin accessibility. Chromatin Immuno-Precipitation sequencing (ChIP-seq) measures DNA-protein binding (D. S. Johnson et al., 2007) while Micrococcal Nuclease sequencing (MNase-eq) detects nucleosome positioning and occupancy (Schones et al., 2008). Below, we briefly describe some of these genomic techniques we have used in this thesis.

2.1.1. ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing)

ATAC-seq has been widely used to study changes in chromatin accessibility in different physiological settings, from understanding accessibility differences in hematopoiesis and leukemia to mapping different chromatin states in schizophrenia patients (Bryois et al., 2018; Corces et al., 2016). The main advantage of ATAC-seq over DNAse-seq and FAIRE-seq is that ATAC-seq requires significantly fewer cells (~50000) compared to millions of cells for other accessibility methods.

ATAC-seq relies on a genetically engineered hyperactive Tn5 transposase that cuts open chromatin at open sites leaving a 9-bp staggered nick. The transposase assists a DNA transposon to transfer DNA sequence from one region to another (Reznikoff, 2003). DNA transposon requires open chromatin at the insertion site and the 9-bp tag left by the transposase is used to construct a library for sequencing. During the DNA transposon process, the nick is repaired, leaving a 9-bp duplication, and paired-end sequencing is performed to facilitate unique alignment rates to the open regions identified by Tn5 cutting (Tsompana & Buck,

2014). Briefly, Tn5 transposase assists the transposon process to simultaneously fragment and tag open regions of DNA with sequencing adapters (**Figure** 2-1).

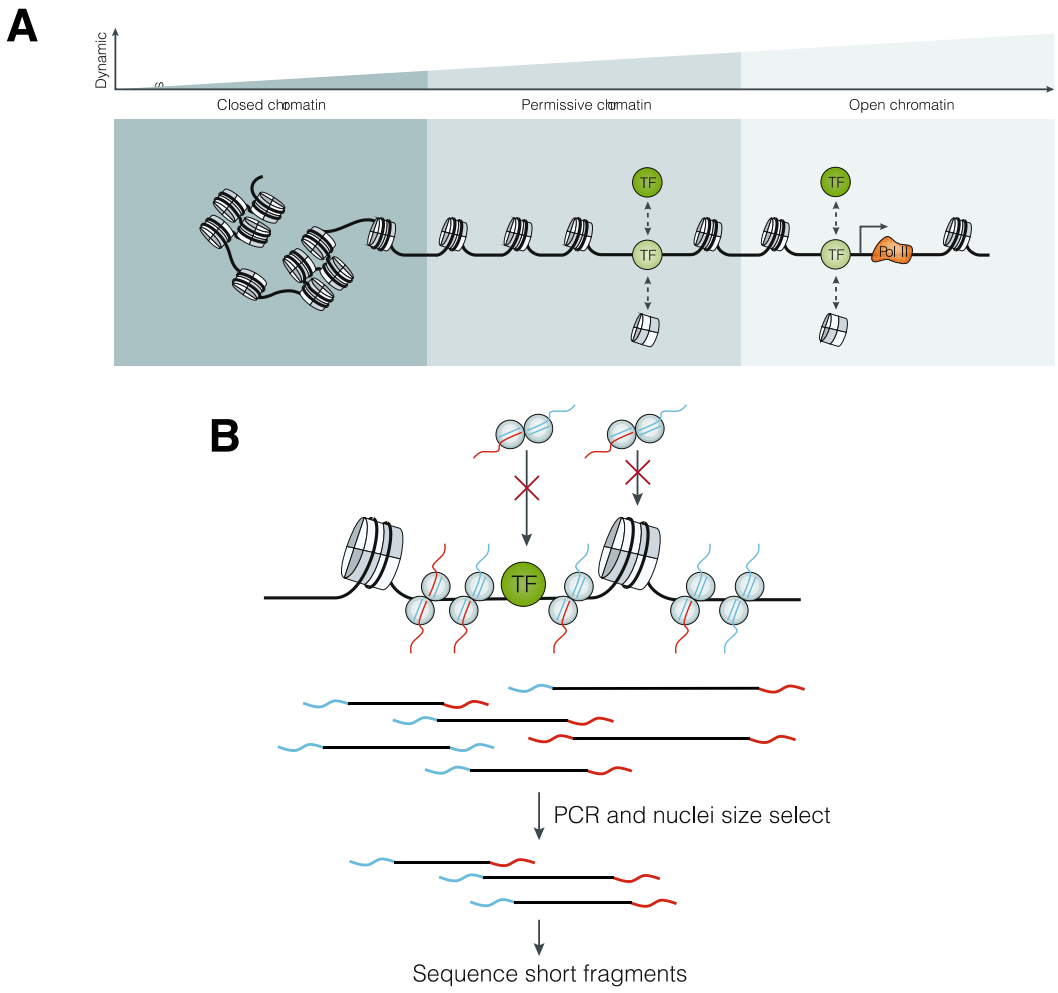


Figure 2-1 Chromatin Accessibility by ATAC-seq
(A) Chromatin accessibility is variable across the genome. Closed chromatin corresponds to DNA sequences tightly wrapped around nucleosomes where TFs and other proteins cannot physically interact with chromatin. TFs and protein are able to physically interact with open chromatin to initiate transcription. (B) TN5 transposase is able to interact and cut open chromatin. If a TF is bound to chromatin or DNA is wrapped in a nucleosome, TN5 is not able to cut the DNA sequences. After DNA fragmentation, PCR and size select is performed and fragments are sequenced to map the levels of chromatin accessibility. Taken from (Klemm et al., 2019)

The main steps in ATAC-seq consist of preparation of the nucleus, transposition and amplification (Tsompana & Buck, 2014). Cells are homogenized into single cells and single nuclei are prepared with catalysis of a lysis buffer. Re-

suspended nuclei are incubated in a transposition reaction mix containing Tn5 transposase to produce DNA fragments by the cutting mechanism of Tn5. Finally, the transposed DNA fragments are amplified to generate libraries for next generation sequencing. Quality control is performed on the ATAC-seq library to guarantee appropriate concentration of the library for sequencing. Bioinformatic tools are used to filter the sequencing data and assess sequencing quality. The adapter sequences and low-quality reads are removed from the data and the remaining reads of around 150 nucleotides are processed. A peak calling algorithm is used to map the high-quality reads to a reference genome and accessible chromatin regions such as enhancers, promoters and insulators are identified (Tsompana & Buck, 2014).

2.1.2. CHIP-seq

Chromatin immunoprecipitation sequencing (ChIP-seq) is a standard technique to identify and characterize protein-DNA interactions genome-wide (Kharchenko, Woo, Tolstorukov, Kingston, & Park, 2008; Valouev et al., 2008). ChIP-seq allows for the characterization of the genomic locations of bound proteins to chromatin such as histones and transcription factors in vivo.

In a CHIP-seq experiment, DNA fragments associated with a specific protein are enriched by immunoprecipitation through specific antibodies. Proteins interacting with DNA are crosslinked to DNA with the reversable crosslinker formaldehyde. Chromatin is sheared into small fragments using sonication to reach a size in the 200-600 bp range. The sonicated chromatin is incubated with a

cocktail of specific antibodies against the protein of interest to enrich for the DNA fragments that exhibit interactions with the protein. The sonicated chromatin is immunoprecipitated, the crosslinks are reversed, and the DNA is purified for next generation sequencing (**Figure** 2-2).

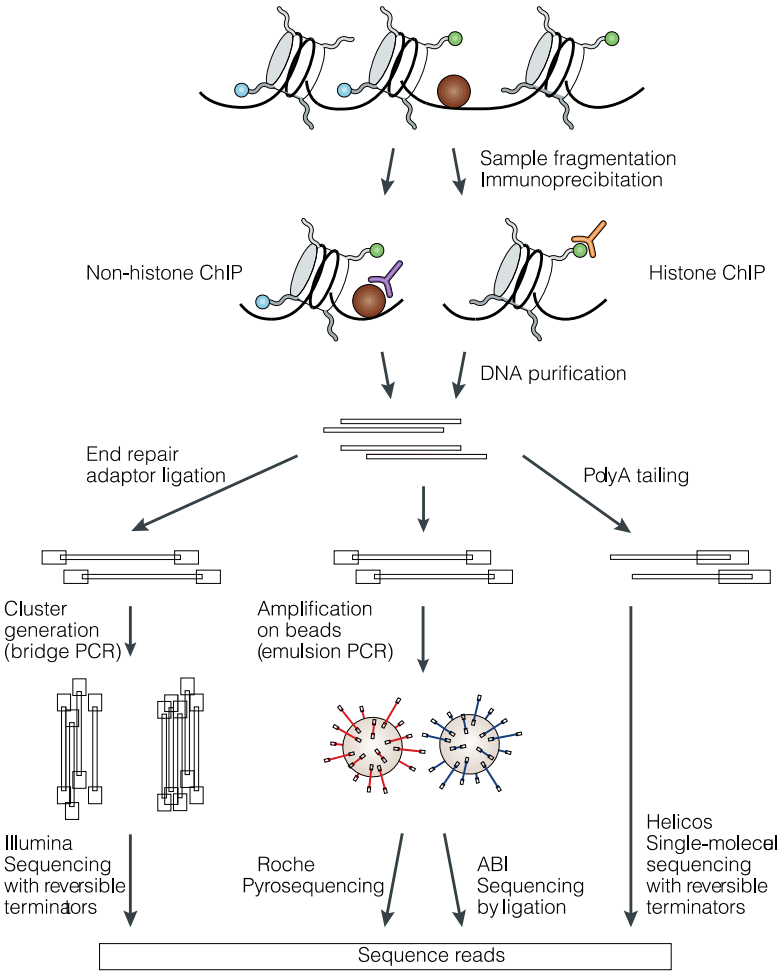


Figure 2-2 CHIP-seq Methodology
CHIP-seq quantifies the interaction of proteins to chromatin. Cells are chemically fixed; DNA is sheared, and immunoprecipitation is performed with a highly specific antibody to the protein of interest. Crosslinking is reversed, DNA purified and sent for sequencing. Enriched reads correspond to DNA sequences where the protein of interest binds to chromatin. Taken from (P. J. Park, 2009)

MNase digestion is used without crosslinking to fragment chromatin if nucleosome positioning or histone modifications need to be mapped (Voong, Xi, Wang, & Wang, 2017). MNase allows for more precise mapping of each nucleosome due to the more efficient removal of linker DNA than sonication. ChIP with crosslinking is called X-ChIP and without crosslinking N-ChIP (native) (O'Neill & Turner, 2003; Orlando, 2000).

ChIP-seq experiments are limited by the antibody quality. A sensitive and high-quality antibody is required to provide specific enrichment with respect to the background in order to identify binding events. However, antibody quality is highly variable even among batches of the same antibody. Validation of the specificity of the antibody needs to be performed prior to ChIP-seq experiments. Enrichment of ChIP-seq peaks is prone to several potential artifacts. Shearing of DNA results in non-uniform fragmentation of the genome with open chromatin regions more enriched than closed regions. Repetitive sequences might be enriched and therefore a peak in a ChIP-seq experiment needs to be compared to the same region in a control sample to determine the significance of the enrichment. The most popular control sample is a portion of DNA removed prior to immunoprecipitation, called input DNA.

After sequencing is performed, the reads are aligned to the reference genome and regions of enrichment are identified relative to the control sample. Fragments are sequenced from the 5’ end and two symmetric peaks are formed around the binding region; a smoothed profile of each strand is constructed, and the combined profile is used to identify the enriched region (Valouev et al., 2008).

A simple fold ratio is calculated with respect to the control using a Poisson or Binomial model for the tag distribution (Boyle, Guinney, Crawford, & Furey, 2008). Multiple replicates are used for reproducibility and to remove false positive enriched peaks and a *q* value is assigned to each peak which corresponds to the minimum false discovery rate (FDR) (Benjamini & Hochberg, 1995) at which a peak is deemed statistically significant. A major difficulty in enrichment analysis is the presence of different types of peaks: broad and sharp. Most algorithms are designed for sharp peaks which correspond to protein-DNA binding while broad peaks are usually associated with histone modifications. Peak calling is usually validated by performing ChIP-qPCR of the enriched regions.

2.1.3. RNA-seq

RNA-seq is mainly used for analyzing differential gene expression (DGE). It consists of extracting RNA from the sample, followed by mRNA enrichment by ribosomal RNA depletion, cDNA synthesis and preparation of a sequencing library. The library is sequenced to obtained at least 10 million reads per sample on different sequencing platforms (Z. Wang, Gerstein, & Snyder, 2009). After sequencing, the reads are aligned to a genome of reference and statistical analysis is performed to quantify changes in gene expression between different experimental groups.

Careful design of an RNA-seq experiment is needed to recover meaningful and reproducible biological data. Enough biological replicates are needed to capture the intrinsic stochasticity of biological systems. The robustness of RNA-

seq analysis is highly dependent on biological replicates even more than the read depth (total number of reads per sample) or length (Schurch et al., 2016). Biological replicates allow for identification of outlier samples to be removed or down-weighted and account for any potential differences due to biological variability.

Different bioinformatic tools and approaches exist to analyze DGE. All these methods are broadly divided into three different steps. First, the raw reads are mapped to a specific transcriptome. In step 2, the number of reads associated with each gene and transcript are calculated, low read transcripts are filtered, and the number of reads is normalized to account for technical variability between samples. Step 3 involves quantification of DGE by statistical modeling of biological replicates and covariates to calculate statistical confidence.

2.2. Imaging Techniques

2.2.1. Super-resolution Microscopy and Single Molecule Detection

In a Single Molecule experiment, individual particles (TFs, enzymes, motors, etc) are observed using different microscopy techniques. These observations are tracked over space and time to generate trajectories for each single molecule. Even though, these trajectories are inherently stochastic, the statistical analysis of an ensemble of trajectories reveals global phenomena of the properties of the underlying physical, chemical and biological process such as diffusion, binding, protein-protein interactions, trapping, folding-unfolding dynamics, enzymatic reactions and others (Manzo & Garcia-Parajo, 2015; M. J. Saxton, 2008).

Other techniques are widely used to analyze protein dynamics such as fluorescence correlation spectroscopy (FCS) and fluorescence recovery after photobleaching (FRAP)(Reits & Neefjes, 2001; Tian, Martinez, & Pappas, 2011). These techniques use the information in fluorescence fluctuations to indirectly characterize the dynamics of the molecules of interest. However, FRAP and FCS only reveal average information on dynamics and any temporal or spatial heterogeneity is lost and their spatial resolutions are diffraction limited which hinders their application for nanoscale processes.

The development of SMT has been closely related to advances in fluorescent tagging of proteins. Tags used in SMT experiments need to be small enough to not interfere with the biological function of the protein and they have to be sufficiently bright and photostable to allow tracking across different time scales. The invention of self-labeling protein tags such Halo, SNAP and CLIP-tag have allowed the used of organic dyes with improved quantum yield and photostability (Erdmann et al., 2019; Grimm et al., 2015) opening the doors to higher localization precision and longer temporal resolution in SMT experiments.

The spatial resolution of optical microscopes is naturally limited by the diffraction of light where in the limiting case of a point emitter, its image corresponds to a distribution of fluorescence intensity called the point spread function (PSF). The diffraction limit is given by $d = \frac{\lambda}{2NA}$, λ corresponds to the wavelength of the emitted light and NA the numerical aperture of the objective. For a red emitting fluorophore and a high NA objective, the diffraction limit is approximately 160nm.

However,

There are two main mathematical approximations that describe a PSF. First, the Airy disk PSF that is the most accurate approximation, but it is prohibitory computationally

expensive $PSFAiryx,y = \left[2\frac{J_1\left(k_{em}NA\sqrt{x^2+y^2}\right)}{k_{em}NA\sqrt{x^2+y^2}}\right]^2$ (1, J_1 corresponds to the first order Bessel function of the first kind, $k_{em} = \frac{2\pi}{\lambda_{em}}$ where λ_{em} is emission wavelength). Second, the Gaussian PSF which is widely used in the field and it is appropriate in most experimental situations (For high NA, a Gaussian PSF with background correction is a better approximation than the Airy PSF). It is given by a gaussian distribution with standard deviation σ_{xy} and amplitude A $PSFx,y = Ae^{\left(-\frac{x^2+y^2}{\sigma_{xy}{}^2}\right)}$ (2).

$$PSF_{Airy}(x,y) = \left[2\frac{J_1\left(k_{em}NA\sqrt{x^2+y^2}\right)}{k_{em}NA\sqrt{x^2+y^2}}\right]^2 \quad (1)$$

$$PSF(x,y) = Ae^{\left(-\frac{x^2+y^2}{\sigma_{xy}{}^2}\right)} \quad (2)$$

The Gaussian PSF is assumed as an objective function and PSF is usually fitted using least square algorithms and maximum likelihood estimation (MLE)(Abraham, Ram, Chao, Ward, & Ober, 2009). Machine learning based approaches such as convolutional neural networks for the detection of nanometer-scale particles have been developed with improved accuracy and reproducibility (Newby, Schaefer, Lee, Forest, & Lai, 2018). Combining these new detection approaches with more robust PSF fitting will improve the robustness of SMT experiments.

Another strategy to track single molecules at the super-resolution level is based on undirect excitation of the fluorophore. MINFLUX nanoscopy can attain resolutions in the order of nanometers (1-3nm) with fewer emitted photons required to reach similar resolutions with conventional microscopy and without dependence on NA and wavelength. It is based in a movable excitation beam with an intensity minimum (donut), a feedback

loop keeps the particle of interest in the center of the beam where it does not get excited. Excitation of the fluorophore triggers the feedback mechanism to identify the position of the fluorophore and recenter the beam, this microscopy technique dramatically reduces photobleaching since it requires a minimum number of photons for localization. Moreover, detection and tracking is not needed since it is naturally implemented in the microscope functionality (Gwosch et al., 2020).

2.2.2. HILO Microscopy

The ability of following single molecules in real time depends directly on the fluorescent signal of the particle. The latter is dependent on the signal to noise ratio (SNR, fluorescence of particle to background fluorescence) and motion blur due to the mobility of the molecule and the exposure time of the camera. Total Internal Reflection Fluorescence (TIRF) microscopy uses the mismatch in refractive index between the immersion index of the imaging objective and coverslip to generate an evanescent excitation wave in an optical phenomenon called total internal reflection (Lights travel from a higher refractive index medium to a lower refractive index medium) to illuminate the sample up to 200nm depth (Axelrod, 2001). By illuminating a small volume of the sample with TIRF, most of background fluorescence is removed and the SNR is drastically improved. TIRF has been applied extensively to single molecule tracking in vitro and cells (Kudalkar, Davis, & Asbury, 2016). The major drawback is the limited penetration depth and conventional TIRF is only applicable to particles closed to the coverslip.

To overcome the limitations of TIRF, Highly Inclined and laminated optical sheet (HILO) microscopy was invented to reduce background fluorescence at longer excitation depths (Tokunaga, Imamoto, & Sakata-Sogawa, 2008). Before the excitation laser reaches its critical angle (Incidence angle needed for total internal reflection), the light is

refracted and a thin optical sheet of length $dz = \frac{R}{\tan(\theta)}$, where θ corresponds to the incident angle and R the diameter of the illuminated area of the specimen (taken from (Tokunaga et al., 2008), **Figure** 2-3). The maximum of illumination is formed at the specimen plane minimizing the illumination divergence away from the focal plane and therefore significantly decreasing out of focus fluorescence and improving SNR. HILO has become a powerful and cost-effective approach for single molecule experiments. A major drawback is the thick sheet of light (~3-7um) and non-uniform excitation intensity at the focal plane.

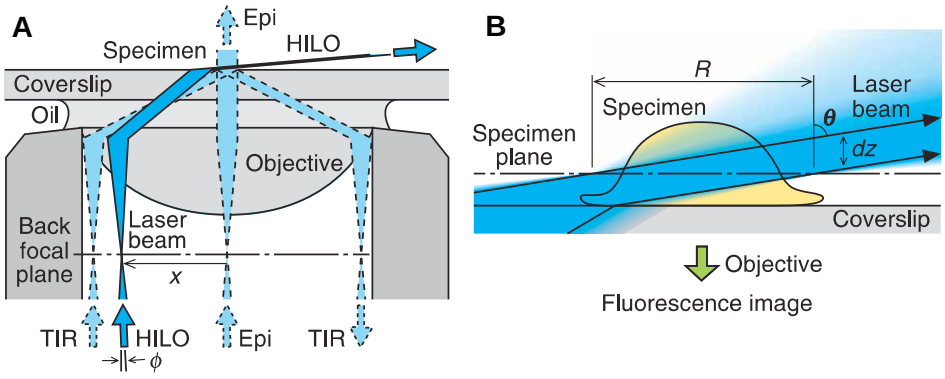


Figure 2-3 HILO Schematics
(A) Different illumination modes depending on the laser beam angle incidence with the coverslip. In EPI illumination, the laser beam goes through the entire specimen maximizing the background fluorescence. In TIRF, an evanescent wave is produced illuminating the surface of the specimen and HILO uses an intermediate incident angle to create an inclined sheet of light (B). Taken from (Tokunaga et al., 2008).

2.2.3. Absorbing Boundary Conditions for the Time Domain

In an SMT experiment, the protein of interest is tagged with a fluorescent probe and imaged. Binding events are then associated with stationary particles in the focal plane. The final experimental information that can be recovered is the time that a protein can be detected in the imaging volume before it bleaches or moves out of the focal plane. From these observations, one can calculate a dwell time for transcription factors (TFs) which is defined as the time interval between a single molecule transitioning from a diffusive state

to a bound state and its subsequent unbinding from DNA and return to the diffusive state (Mazza, Abernathy, Golob, Morisaki, & McNally, 2012a). The dwell time distribution is obtained by calculating the ensemble distribution of bound times for a specific TF in different cells in the experiment after photobleaching correction. The survival distribution is then calculated as 1-CDF, where CDF is the empirical cumulative distribution function of dwell times.

Calculation of dwell time distributions is a first-passage time problem in stochastic analysis and these distributions have been widely used to characterize kinetic properties of molecular motors and ion channels (Liao, Spudich, Parker, & Delp, 2007). In cases involving simple kinetic schemes, the dwell time distributions can be calculated analytically but for more complex schemes, a number of methods have been utilized. One particularly powerful approach is to assign one or more states to act as an absorbing boundary and then solve the associated first-order kinetic equations to obtain dwell time distributions. We assume that the diffusive (unbound) state corresponds to an absorbing boundary state, since the measurement ends with such transitions, because the particle either photobleaches, disappears from the focal plane or begins diffusing; any rebinding of the TF is considered as an independent event. This assumption implies that the population of particles in the absorbing boundary state increases with time. At the end of every experimental measurement, all the observed TFs transition to the absorbing state since the experiments are continued until most particles are bleached.

For a general process, a TF can be found in any state i such as diffusing around the nucleus, confined in a microenvironment, bound to a particular specific or non-specific site of the DNA. When a TF transitions to a diffusive state, it cannot be observed experimentally, and this state plays the role of an absorbing boundary state. We observe the system over a time interval from $t' = 0$ to $t' = \tau$, during which individual TFs may

undergo transitions between different states $\mathrm{i} \in \{1, \ldots, \mathrm{n}\}$. When a transcription factor is in a "bound" state, it can be observed experimentally as a trace as displayed in **Figure** 2-4**a**. Each TF in any of these bound states will be experimentally recorded from a certain time interval t_1 to t_2 with $t_2 - t_1 < \tau$. t_1 corresponds to the time when the transcription factor transitions to a "bound" state and t_2 corresponds to the time when the TF enters an absorbing state (diffusion in the nucleus). All the traces will be shifted by a time t_1 to a new aligned time $t = t' - t_1$ so that all the TFs begin in a bound state at $t = 0$ (**Figure** 2-4**b**). During the experimental time τ, a finite number of TFs (N, equal to the number of traces) will be observed. When a TF transitions to a diffusive state, it cannot longer be observed experimentally, and this state plays the role of an absorbing boundary state.

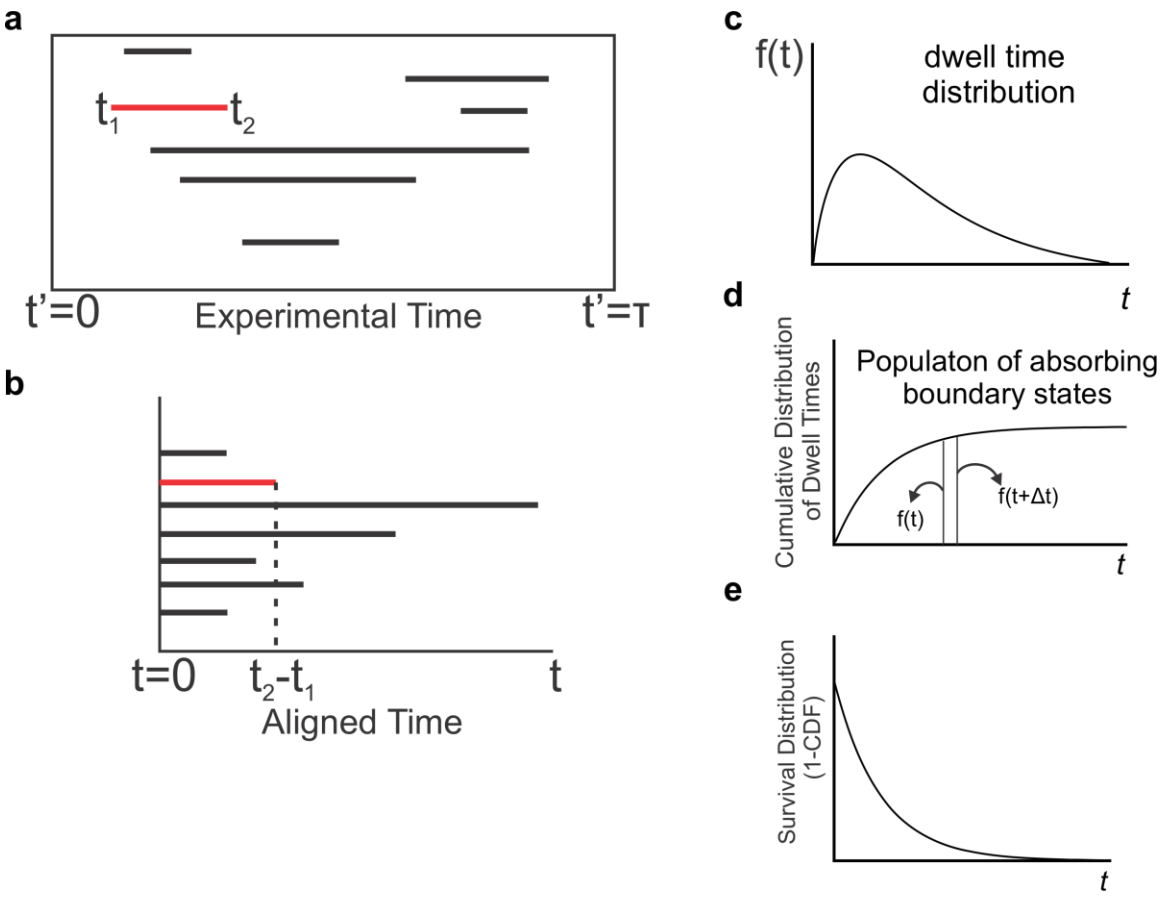


Figure 2-4 Survival Distribution Calculation
(A) Experimentally, slow events are seeing as traces in a kymograph. (B) These traces are aligned on a new time t and the distribution of lengths corresponds to the dwell time distribution $f(t)$ (C) Shown in red a

sample trace before and after time alignment. (D) The CDF of $f(t)$ corresponds to the normalized population of an absorbing boundary state. (E) (1-CDF) corresponds to the survival distribution $\widehat{D}$. Taken from (Garcia et al., 2021b).

Let $p_i(t)$ correspond to the probability of being in state i at time t.$\sum_{i\in boundaries} p_i(t)$ corresponds to the population of all absorbing states. To calculate number of unbinding events over a certain time interval ($f(t)$ dwell-time distribution, **Figure** 2-4**c** - adapted from (Liao et al., 2007)), we take the time derivative of this population, $f(t)$ can be visualized as the probability distribution of experimental track lengths of TFs entering a bound state and evolving independently from a registered time $t = 0$, until they transition to an absorbing boundary state, at which time they leave the bound state(Van Kampen, 1992). The cumulative distribution of $f(t)$ is calculated ($f(t) = \frac{d}{dt}\sum_{i\in boundaries} p_i(t)$ (3), **Figure** 2-4**d**) and 1-CDF corresponds to the survival distribution ($\widehat{D}$, **Figure** 2-4**e**).

$$f(t) = \frac{d}{dt}\sum_{i\in boundaries} p_i(t) \quad (3)$$

A double exponential model is widely used in the analysis of SMT of transcription factors to describe their binding with chromatin (Darzacq et al., 2009; A. S. Hansen, Pustova, Cattoglio, Tjian, & Darzacq, 2017; Paakinaho et al., 2017; E.E. Swinstead et al., 2016). This model, a TF can bind either to a specific site or a non-specific site in chromatin with transitions between specific and non-specific binding forbidden. The biological reasoning is based on that TF have a binary binding behavior with well-defined binding energies. The survival distribution for a double exponential model is given by:

$$\boxed{\widehat{D} = \frac{1}{k_1+k_2}\left(k_1 e^{-k_1' t} + k_2 e^{-k_2' t}\right)} \quad (4)$$

The first exponential component corresponds to specific binding to chromatin with an average binding time of $\frac{1}{k_1}$ and the second component corresponds to non-specific binding to chromatin with an average binding time of $\frac{1}{k_2}$. Non-specific binding has been

reported in the order of milliseconds to seconds and specific chromatin binding for transcription factors have been reported in the order of seconds (7.25 secs for GR activated with Dexamethasone (Paakinaho et al., 2017)).

Chapter 3. Power-law Behavior of Transcription Factor Dynamics at the Single-Molecule Level Implies a Continuum Affinity Model

This chapter has been adapted from Garcia et al., Nucleic Acid Research, 2021 (Garcia et al., 2021b). I was primarily responsible for the experimental design, single molecule tracking experiments (with help from Diego Presman, Ville Paakinaho and Gregory Fettweis), modeling and data analysis.

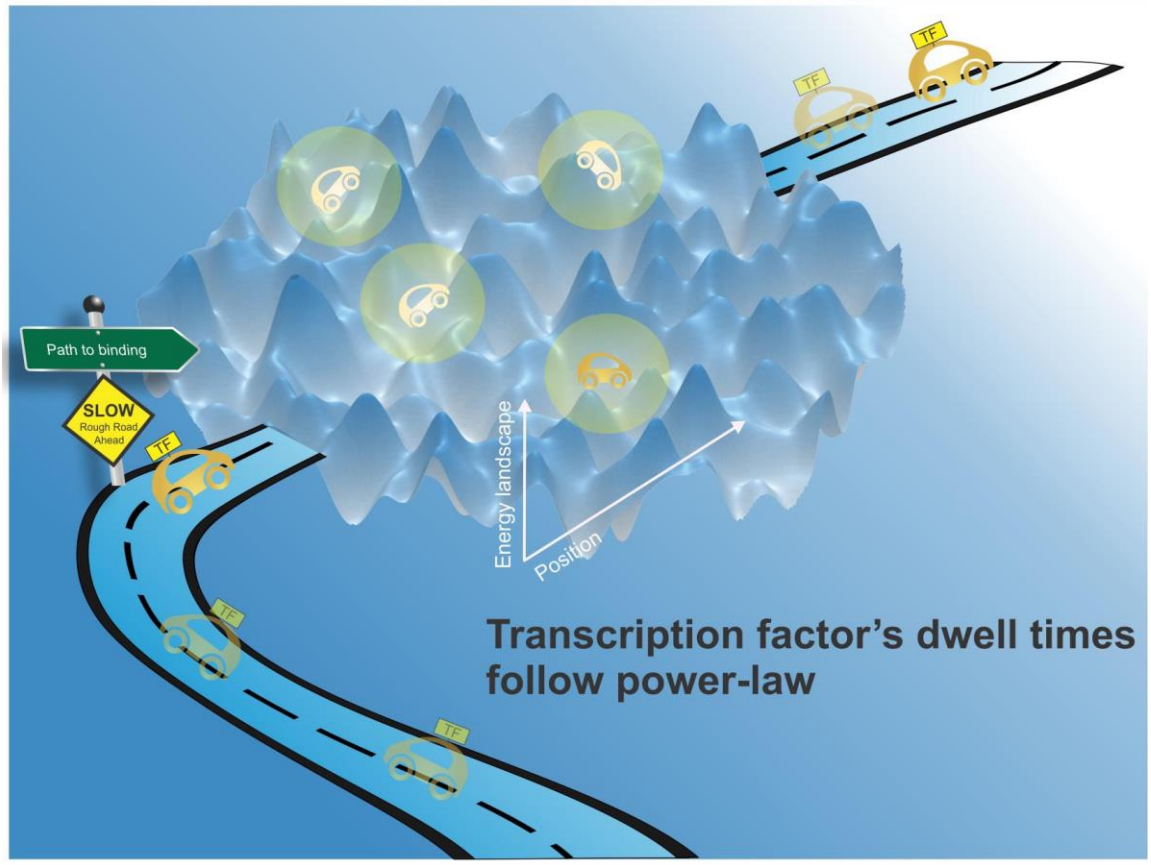


Figure 3-1 Graphical Summary
Transcription factors searching on chromatin for specific response elements is depicted as a road representing chromatin. TFs interact with a broad distribution of binding affinities which lead to power-law distributed dwell times.

3.1. Summary

Single-molecule tracking (SMT) allows the study of transcription factor (TF) dynamics in the nucleus, giving important information regarding the diffusion and binding behaviour of these proteins in the nuclear environment. Dwell time distributions obtained by SMT for most TFs appear to follow bi-exponential behavior. This has been ascribed to two discrete populations of TFs - one non-specifically bound to chromatin and another specifically bound to target sites, as implied by decades of biochemical studies. However, emerging studies suggest alternate models for dwell-time distributions, indicating the existence of more than two populations of TFs (multi-exponential distribution), or even the absence of discrete states altogether (power-law distribution). Here, we present an analytical pipeline to evaluate which model best explains SMT data. We find that a broad spectrum of TFs (including glucocorticoid receptor, estrogen receptor, FOXA1, CTCF) follow a power-law distribution of dwell-times, blurring the temporal line between non-specific and specific binding, suggesting that productive binding may involve longer binding events than previously believed. From these observations, we propose a continuum of affinities model to explain TF dynamics, that is consistent with complex interactions of TFs with multiple nuclear domains as well as binding and searching on the chromatin template.

3.2. Introduction

Transcription factors (TFs) are key regulatory proteins responsible for turning genes “on” and “off” by binding to enhancer or promoter elements across

the genome (Lambert et al., 2018). The current consensus describes TFs as being able to transition between three different states: 1) unbound from DNA (diffusing in the nucleus), 2) non-specifically bound and 3) specifically bound to chromatin (i.e. interacting with specific response elements) (Goldstein et al., 2017). However, biochemical studies and live-cell imaging experiments appear to disagree on the timescale that eukaryotic TFs can remain bound to chromatin, ranging from seconds to several hours or even days (Brouwer & Lenstra, 2019; Coulon et al., 2013; Gurdon, Javed, Vodnala, & Garrett, 2020; Hager et al., 2009; Lerner et al., 2020; Stasevich & McNally, 2011).

Advances in fluorescence microscopy have revolutionized our understanding of how TFs search and interact with chromatin (Liu & Tjian, 2018). Single-molecule tracking (SMT), which is based on detecting and following through time the traces produced by the light emitted from a single fluorophore, allows the characterization of protein dynamics in live cells. When applied to the study of TFs, important information regarding the search and binding dynamics of these proteins can be extracted (Liu & Tjian, 2018). SMT has been applied to over a dozen TFs, and has revealed that the time TFs remain bound to chromatin (i.e. residence time) is relatively short (seconds) and follows a bi-exponential distribution (reviewed in (Goldstein et al., 2017)). The bi-exponential behavior is consistent with decades of biochemical studies, indicating that the DNA-bound population of molecules are composed of two distinct subpopulations: a short-lived fraction (‘fast stops’) and a longer-lived fraction (‘slow stops’). The fast fraction has been interpreted to represent non-specific binding to chromatin while the slow fraction is thought to be

consistent with specific binding at enhancers or promoters (Ball et al., 2016; Chen et al., 2014; Kilic, Bachmann, Bryan, & Fierz, 2015; Morisaki, Muller, Golob, Mazza, & McNally, 2014). Experiments wherein TFs were mutated in their DNA-binding domains seem to confirm this model as the longer binding events were reported to be dramatically reduced (Callegari et al., 2019; Chen et al., 2014; Morisaki et al., 2014; D. M. Presman et al., 2017; Sugo et al., 2015).

However, this view is at odds with our current understanding of the nuclear environment. Far from being homogenous, the nucleus is highly compartmentalized and can impose constraints on the motion of many transcription-related molecules (Finn & Misteli, 2019; S. Kim & Shendure, 2019; Lerner et al., 2020). For example, the presence of nuclear bodies, liquid-liquid condensates and distinct chromosomal architectures can critically affect the searching process of TFs for their target sites (Brouwer & Lenstra, 2019; Woringer & Darzacq, 2018), implying that TF dynamics should exhibit dynamics beyond the bi-exponential model.

Recently, studies fitting TF dynamics to a three-exponential model have found longer residence times for the serum response factor (SRF) (over 4 minutes) (Hipp et al., 2019) or CCCTC-binding factor (CTCF) (~15 min) (Agarwal, Reisser, Wortmann, & Gebhardt, 2017) than would be expected from a bi-exponential model. Moreover, a multi-exponential model was used to explain the dynamics of the TF CDX2 (Reisser et al., 2020). Finally, the bacterial proteins, tetracycline repressor (TetR) and LacI, with no known endogenous targets in mammalian cells, show power-law behavior when heterologously expressed (Caccianini, Normanno,

Izeddin, & Dahan, 2015; Normanno et al., 2015). In fact, these non-specific binding events could be as long as specific ones (Normanno et al., 2015).

A random variable t follows a power-law (Newman, 2005) for $t > t_{min}$ if $f(t) = At^{-\beta}$, where *A* is a constant and β is the exponent or scaling parameter. Power-laws are heavy tailed (right-skewed), which makes rare events more likely to occur than for exponential distributions; and β is a measure of the skewness. Many natural and artificial systems have been found to follow power-law distributions (Newman, 2005). For proteins interacting with chromatin, it would mean that the frequency of binding events of a given TF will be inversely proportional to the residence time of said TF. In fact, binding times orders of magnitude longer than the average are likely to be observed. More importantly, for mammalian TFs that follow a power-law distribution, assigning discrete residence times for specific and non-specific binding would not be feasible. Whether this phenomenon occurs for endogenous mammalian TFs remains an open question. While these discordant results regarding TF binding dynamics could reflect the underlying biology, they may also arise due to the lack of consensus in the field regarding tracking algorithms, photobleaching (PB) correction methods, and model fitting.

Here, we revaluated some of the core aspects of the SMT technique, focusing on PB correction methods. We then derived theory-based models for TF dynamics and a principled method to obtain optimal model parameters from empirical residence time distributions, using Bayesian statistics. With these methods, we analyzed the dynamics of several TFs, including the glucocorticoid

receptor (GR), the estrogen receptor (ER), the “pioneer factor” forkhead box A1 (FOXA1), the chromatin remodeler BRG1 (SMARCA4) as well as the architectural protein CTCF. Our data is consistent with power-law behavior for all tested proteins. We further discuss theoretical considerations for how the observed power-law distribution can arise from broad effective distributions of binding affinities can result in the observed power-law distribution. We suggest that TF dynamics is not explained by a simple separation between non-specific and specific binding but rather reflects the heterogeneous nature of chromatin structure and binding strengths.

3.3. Results

3.3.1. Photobleaching correction methods and their effect on survival distributions

When tracking TFs at the single-molecule level, the experimental information that is recovered is the time the molecule “remains” visible before it bleaches or moves out of the focal plane. Thus, binding events can be observed as stationary spots (**Figure** 3-2**A-C**). From these observations, one can obtain a local dwell time for TFs, which is defined as the time interval between a single molecule transitioning from a diffusive state to a bound state and its subsequent return to diffusion. The dwell time distribution is generated by integrating the ensemble-averaged distribution of bound times (**Figure** 3-2**D** and **Appendix A 1.1**). Most often, a “survival” distribution, defined as 1-CDF, where CDF is the empirical cumulative distribution function of dwell times, is used for further analysis [**Figure** 3-2**E,** GR dynamics adapted from (Paakinaho et al., 2017)]. This plot

represents the probability P that a molecule will last t number of time points, or longer. This survival distribution is fit to a bi-exponential distribution [**Figure** 3-2**E** and reviewed in (Goldstein & Hager, 2018a)], and interpreted as the "three population model" (i.e. diffusive, fast bound or non-specific binding, slow bound or specific binding) as illustrated in **Figure** 3-2**F**. However, as can be seen in **Figure** 3-2**E**, the data shows a distinct departure from the bi-exponential fit, especially at longer dwell times.

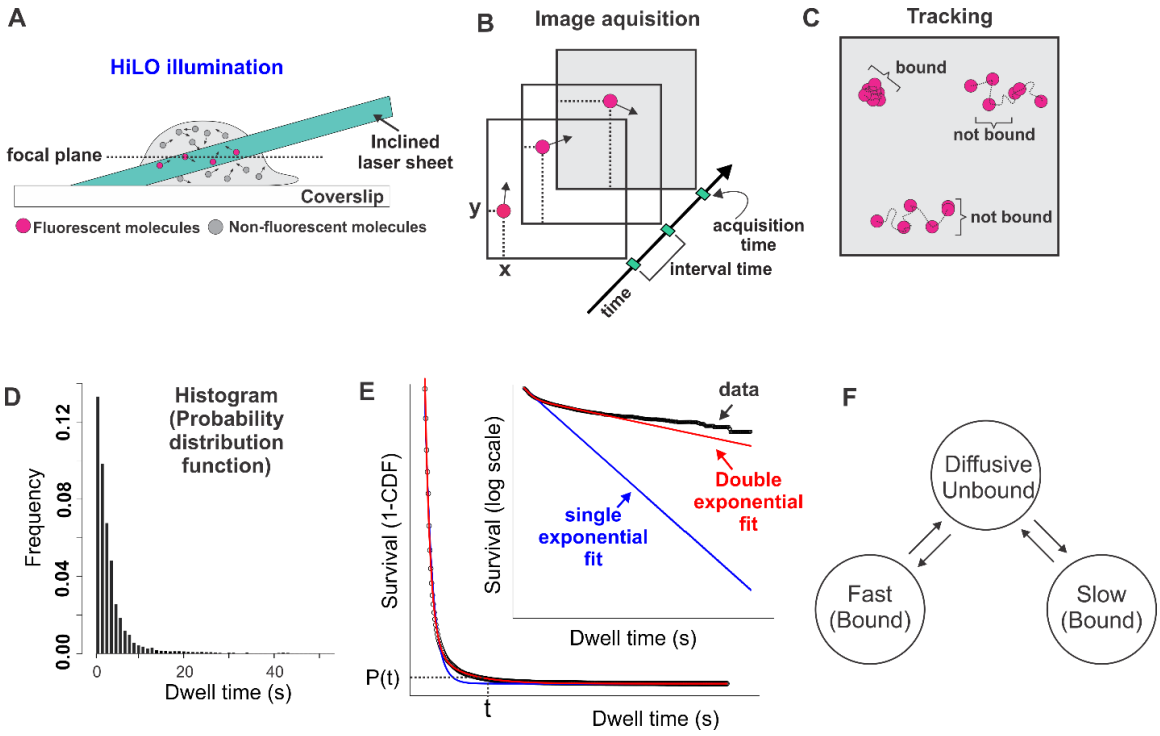


Figure 3-2 The current SMT pipeline and interpretation of TF dynamics.
(A) A HiLO set-up is most commonly implemented to increase signal-to-noise ratio. A laser beam is tilted and hits the sample creating a thin illumination layer in the focal plane. (B) Several images are taken at specific yet variable acquisition and interval time conditions. (C) A tracking algorithm is used to follow individual molecules and classify them as either bound or unbound. (D) Histogram plotted from the bound population showing the frequency of TF molecules that are bound for a specific time (dwell time). Data acquired at 200ms interval for HaloTag-GR activated with its natural ligand corticosteron. (E) Fitting of the survival distribution (1-CDF; cumulative distribution function) calculated from the data shown in D (circles) is fit to single-exponential (blue line) or bi-exponential (red line). Inset shows semi-log plot of the same. (F) Schematic showing the bi-exponential model according to which TFs occupy three different states: unbound from the DNA (diffusing in the nucleus), specifically bound (slow stops), and non-specifically bound (fast stops).

The upper temporal limit in SMT experiments is ultimately determined by the intrinsic photostability of the chosen fluorophore (Liu & Tjian, 2018). When the affinity of bound TFs results in dwell times longer than those resulting from the average photostability of their fluorescent dyes, residence times cannot be resolved. Importantly, even when bound molecules have relatively lower affinities, they will appear to have shorter experimental dwell times due to photobleaching (PB) bias. To illustrate this known phenomenon, we conducted single-molecule imaging by transiently transfecting 3617 mouse mammary adenocarcinoma cells with the GR, a ligand-dependent transcription factor (Presman & Hager, 2017), tagged with HaloTag-Janelia Fluor 549 (JF549) (Paakinaho et al., 2017) and stimulated with GR's natural ligand corticosterone (Cort, 600 nM). When we artificially modulated the PB conditions by changing acquisition parameters (exposure time, imaging interval, laser power), the resulting kymographs have different typical lengths (**Figure** 3-3) and thus appear to have originated from different TFs. Therefore, PB must be properly corrected to prevent artifacts in the analysis of SMT data (Mazza, Ganguly, & McNally, 2013). Since PB correction methods vary widely among research groups (Callegari et al., 2019; Chen et al., 2014; Gebhardt et al., 2013; A. S. Hansen et al., 2017; Mazza et al., 2012a; Morisaki et al., 2014; Normanno et al., 2015; Teves et al., 2016; Zhen et al., 2016) there is no standard approach to overcome the photobleaching bias of SMT strategies. Therefore, we decided to test the most common methods and our proposed approach by comparing how well they can correct the artifacts generated in GR dynamics measured with different acquisition conditions.

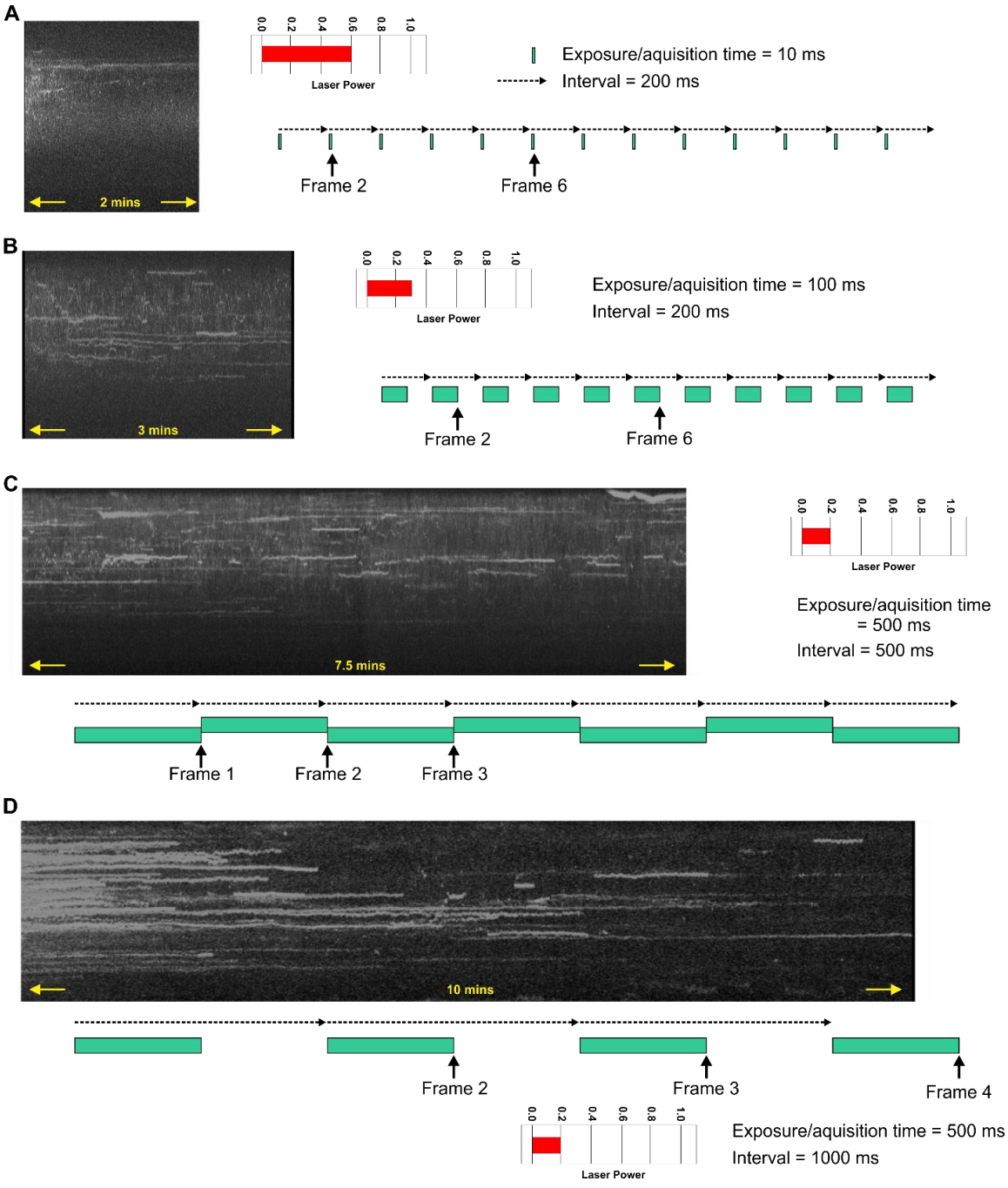


Figure 3-3 Qualitative illustration of photobleaching artifacts in SMT experiments. Single-molecule tracking data of the glucocorticoid receptor (GR) activated with corticosterone (Cort). The figure shows representative kymographs of GR molecules taken at different acquisition conditions (A-D), as indicated. The figure shows that track lengths are dependent on photobleaching kinetics, artificially modifying the apparent dwell time of GR. Without further analysis, the kymographs resemble different TFs.

First, we tested the approach of estimating PB rates by counting, frame-by-frame, the number of particles of the TF of interest in the focal plane, then fitting the time-dependent decay of the molecule count (which is taken as a proxy for PB) to a bi-exponential model (Ball et al., 2016; Loffreda et al., 2017; Mazza et al., 2012a; Morisaki et al., 2014; Paakinaho et al., 2017). This bi-exponential fit is finally used to normalize the survival distribution of the TF of interest, in this case, GR (**Figure** 3-4**A-B, method #1**). However, this method underestimates PB because most of the "counted molecules" are diffusive ones, and as such they are exposed to less laser illumination than bound molecules at the focal plane. Accordingly, this method fails to correct the apparent differences in GR survival distributions obtained from different acquisition conditions (**Figure** 3-4**B**).

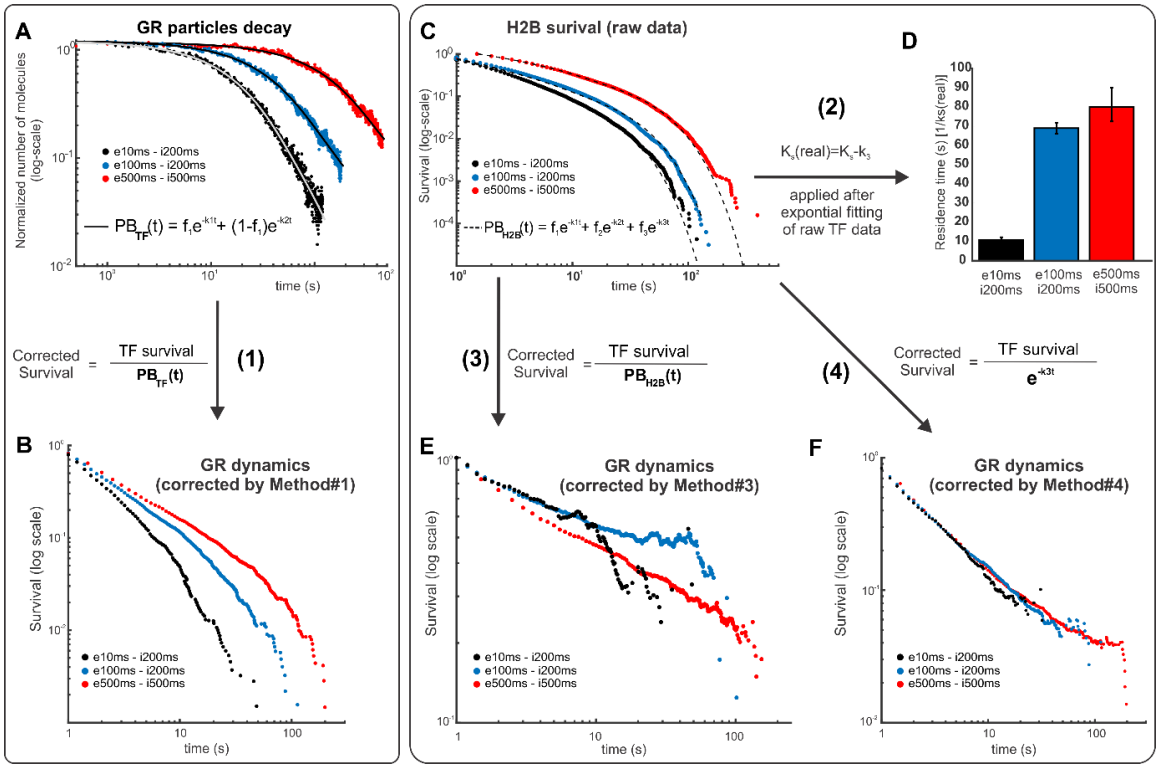


Figure 3-4 Effect of different photobleaching correction methods on the survival distribution of GR.

(A) The number of particles (normalized to the initial number of particles for each cell at time zero) from frame-to-frame as a function of time shown for three different acquisition conditions as indicated in the legend ('e' denotes exposure time and 'i' denotes inter-frame interval). In method #1, this is taken as a proxy for photobleaching (PB), which is fitted to a bi-exponential function (black lines). (B) Effect of method #1 on GR dynamics at different acquisition conditions (e, exposure time; i, interval time). The corrected survival is obtained by dividing the observed TF survival to the bi-exponential distribution obtained in A. Number of cells/number of tracks are 67/9374 for GRe10ms/i200ms; 65/23172 for GRe100ms/i200ms; and 34/37953 for GRe500ms/i500ms. (C) In methods #2-4, the survival of H2B, taken under the same acquisition conditions as the TF, is used as a proxy for PB, which is fitted to either two or three family of exponentials. Number of cells/number of tracks are 100/36625 for H2Be10ms/i200ms; 63/40652 for H2Be100ms/i200ms; and 36/20307 for H2Be500ms/i500ms. (D) Method #2 does not correct the entire TF survival distribution but rather uses the slowest rate of the histone survival fitting (k3) to correct by subtraction the rate of the TF fitting (ks), thus obtaining the "real" rate (ks(real)). The panel shows the residence time (1/ ks) for the three experimental acquisition conditions. (E) Method #3 is similar to method #1, except that it uses H2B survival as a proxy for PB correction. The panel show GR dynamics at different acquisition conditions. (F) In Method #4, the exponential distribution of the slowest component in H2B survival is used as a proxy for PB correction. The panel show GR dynamics at different acquisition conditions. See Table 3-2 and Table 3-3 for more data points details.

Another family of methods uses histones as a proxy for obtaining PB rates (**Figure** 3-4**C)**. Histones are a good representation of stably bound proteins because, after integration into chromatin, their residence time is much longer than

the photostability of any currently available organic fluorophore (Kimura & Cook, 2001). Therefore, by measuring the residence time of histones, one can obtain, in principle, a direct representation of PB for particles in the focal plane, since the disappearance of a long-lived particle will most likely represent a PB event. Different methodologies have been used under the umbrella of histone PB correction, ranging from measuring “bulk” histone levels and fitting the mean nuclear fluorescence (Chen et al., 2014), to variants of measuring histone dynamics at the single-molecule level (Callegari et al., 2019; A. S. Hansen et al., 2017; Teves et al., 2016; Zhen et al., 2016). We will focus on the latter methods, as they use the same acquisition conditions as the TF of interest. One variant (A. S. Hansen et al., 2017; Teves et al., 2016) fits the histone data to an exponential family (usually two components). However, instead of using the information of the entire histone survival distribution, only the decay rate of the longest component is used to correct the residence time of the TF by subtraction (**method #2**), effectively assuming that both TF and photobleaching dynamics follow exponential forms. Unfortunately, this method still gives different residence times for different acquisition conditions (**Figure** 3-4**D**). Another variant (Callegari et al., 2019; Zhen et al., 2016) is similar to method #1, but uses the survival distribution from histones instead of the number of molecules to normalize the TF data (**Figure** 3-4**E, method #3**). Although much better than method #1 (and #2), it fails to normalize GR distributions obtained with different acquisition conditions (**Figure** 3-4**E)** because the survival distribution of histones still has a significant population of diffusive molecules that are not incorporated into chromatin.

We therefore propose a modification to the previous PB correction methods, by combining the best of the three methodologies (**Figure** 3-4**F, method #4,** see methods for details). First, as in method #2 and #3, we fit the histone (HaloTag-H2B) SMT data, taken under the same imaging conditions as the TF of interest, to a family of exponentials. Second, we use the exponential distribution of the longer component (the entire exponential distribution rather than just the rate of the exponential) to normalize the TF survival data. In this way, we only correct for photobleaching by taking into account the bound histone population, without making any *a priori* assumptions about the survival distribution of a TF, as done in method #2. Using this modified version of PB correction, we find that GR survival time distributions obtained under different imaging conditions fall along the same curve as they should (**Figure** 3-4**F**). Taken together, our analysis suggests that this method more accurately corrects for photobleaching bias, as we obtain similar survival distribution of the TF irrespective of the photobleaching kinetics.

3.3.2. GR dwell time distribution deviates from bi-exponential behaviour

We had previously used method #1 and characterized GR's survival distribution as bi-exponential (D. M. Presman et al., 2017; E. E. Swinstead et al., 2016). Similarly, other groups have characterized binding times for other TFs as bi-exponentially distributed using their own PB correction methods [For example (Chen et al., 2014; A. S. Hansen et al., 2017; Lerner et al., 2020; Sugo et al., 2015; Zhen et al., 2016)]. Remarkably, when we apply our newly proposed method (method #4) of PB correction to the dwell time distribution derived from SMT data

of HaloTag-GR activated with corticosterone (Cort, 600 nM), we find that the distribution now deviates from a bi-exponential distribution (**Figure** 3-5**A**, **Table** 3-2 **and Table** 3-3). The data look strikingly linear on a log-log plot (**Figure** 3-5**B**), which suggests power-law behavior. The deviation from exponential is not due to an artifact of HaloTag, as dynamics of HaloTag-alone remain bi-exponentially distributed with no detectable “bound” molecules longer than 20 seconds (**Figure** 3-5**C**). To rule out artifacts from the imaging of exogenously expressed GR, we performed SMT on a GR knock-out cell line stably expressing Halo-GR at endogenous levels (Paakinaho et al., 2017). The results are indistinguishable from the exogenous Halo-GR (**Figure** 3-5**D**), validating our transient expression strategy. Our data thus far suggests that a bi-exponential function does not describe GR dynamics at the single-molecule level, and a power-law might better explain the data.

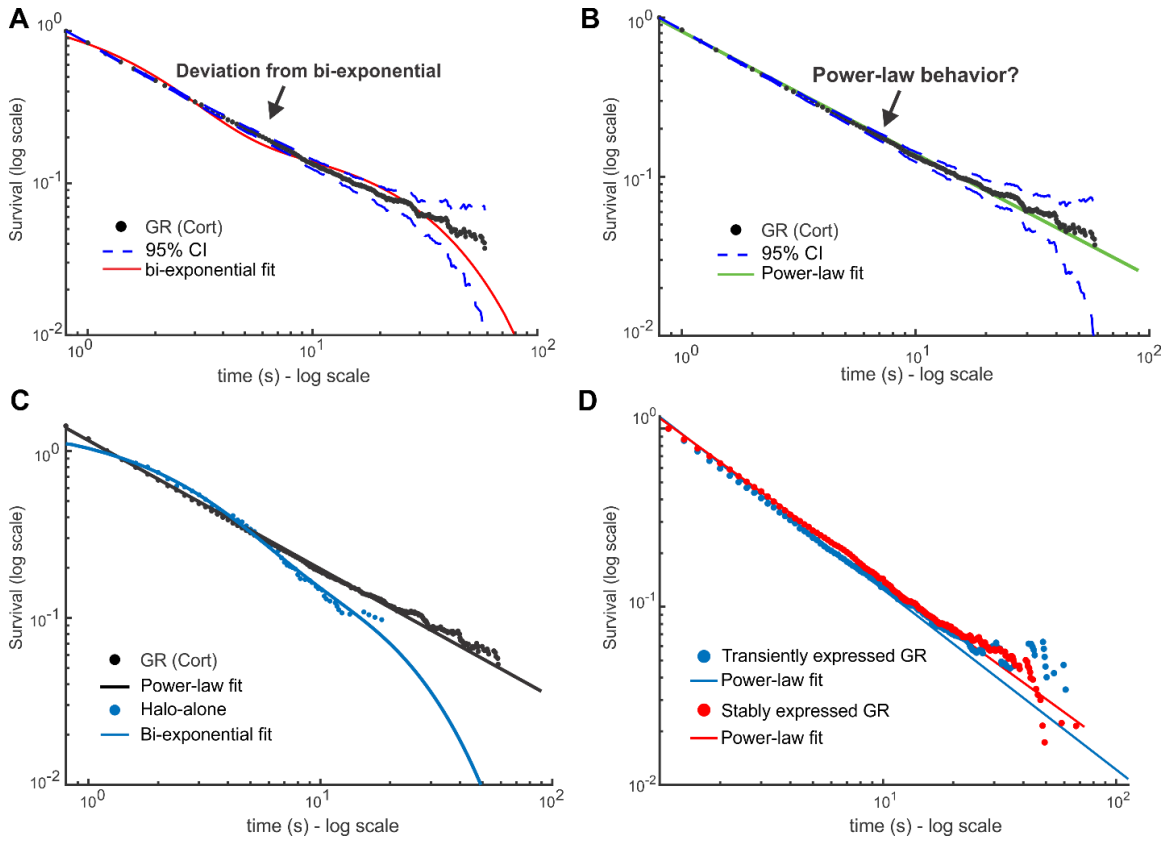


Figure 3-5 Impact of photobleaching correction on GR dynamics.
(A-B) Single-molecule tracking data of GR activated with corticosterone (Cort). Data was acquired at 100 ms exposure time with 200ms interval. The survival distribution is shown (black), fit to a bi-exponential (A) or a power-law (B) function. Dashed lines show 95% confidence intervals (CI). Number of cells = 65; number of tracks = 23172. (C) Comparison of survival distributions of HaloTag-alone (blue) with a bi-exponential fit and HaloTag-GR (black) with a power-law fit. Data was acquired at 100 ms exposure time with 200ms interval. Number of cells = 64; number of tracks = 19436. (D) Survival distributions of HaloTag-GR, treated with Dex, either transiently transfected in 3617 cells (blue) or stably integrated in a GR knock-out subclone (red), expressed at endogenous levels. Data was acquired at 10ms exposure time with 200ms interval. Number of cells = 60; number of tracks = 7068 for GR transient. Number of cells = 60; number of tracks =16450 for GR stable. Coloured lines show power-law fits. See **Table** 3-2 **and Table** 3-3 for details on fits.

3.3.3. Theoretical models for TFs kinetics to interpret SMT data

Deviation from bi-exponential behavior and the emergence of power-law behavior has been described previously using heterologous expression of bacterial proteins (TetR and LacI) into mammalian cell lines (Caccianini et al., 2015; Normanno et al., 2015). Moreover, a multi-exponential model has recently been proposed for the TF CDX2 (Reisser et al., 2020) and SRF (Hipp et al., 2019).

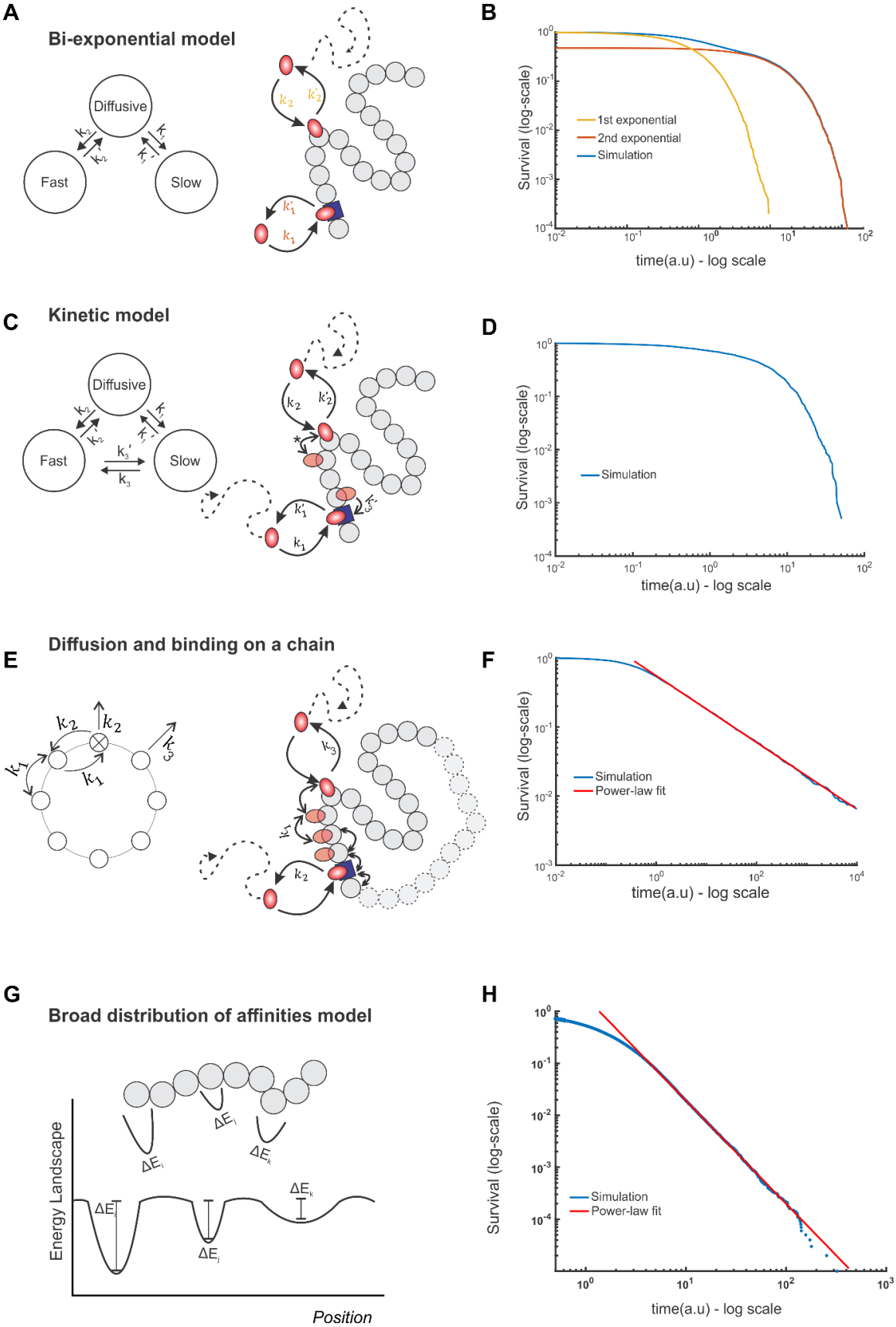
To better understand the link between TF binding and the observed residence time distributions, we explored different theoretical models that may explain the emergence of different behaviors of the survival distribution.

Calculation of dwell time distributions is a first-passage time problem in stochastic analysis and has been widely used to characterize the kinetic properties of molecular motors and ion channels (Liao et al., 2007). When simple kinetic schemes are involved, dwell time distributions can be calculated analytically. However, for more complex systems, other methods must be used. One particularly powerful approach is to assign one or more states to “act” as an absorbing boundary, and then solve the associated first-order kinetic equations to obtain dwell time distributions (Van Kampen, 1992) (**Appendix A 1.1**). We assume that the diffusive state (unbound) corresponds to an absorbing boundary state since tracked particles end with such transitions. The single molecule either photobleaches, disappears from the focal plane or begins diffusing. Any rebinding of the TF is considered an independent event.

We first examined the widely used bi-exponential model under this framework (**Figure** 3-6**A**). According to this model, TFs can occupy three different states: diffusive, slow and fast. In our analytic framework, the diffusive state plays the role of an absorbing boundary state, since particles entering the state are no longer tracked. The slow and fast states correspond to the empirically observed specific and nonspecific binding, respectively (reviewed in (Goldstein & Hager, 2018b)). With this assumption of a well separated and narrow distribution of affinities, the expected behavior of the survival distribution corresponds to a bi-

exponential with the exponential parameters determining the average residence time of each state, as determined by analytic calculation (see **Appendix A1.2)** and confirmed with stochastic simulations (**Figure** 3-6**B**) using the Gillespie algorithm (Gillespie, 1977). We note that this model does not allow for transitions between fast and slow states, which can be hard to interpret biologically, as searching (fast) should lead to specific binding (slow).

Figure 3-6 Theoretical models for TF kinetics.
(A) State diagram (left) and schematic (right) of the bi-exponential model. TFs (orange oval) can bind to specific sites (blue square) or non-specific sites (grey circles) with rate constants k1 and k2 or unbind and return to the diffusive state with rate constants, k'1 and k'2 respectively (A). Transitions between specific and non-specific sites are forbidden. (B) Numerical simulation showing the emergence of bi-exponential behaviour for the model in A. The first and second exponential components are also shown as indicated. (C) State diagram (left) and schematic (right) of the Kinetic model. In addition to binding/unbinding to/from specific and non-specific sites, TFs can transition from specific sites to non-specific sites (with rate constant k3) and vice versa (with rate constant k'3). Transitions between non-specific sites are considered indistinguishable (denoted by *). (D) Simulation results showing survival distributions arising from the kinetic model. (E) State diagram (left) of the continuum of affinities model, showing that transitions from a non-specific site to any other site occur with rate constant k1 and from a specific to a non-specific site with rate constant k2. Transitions to the diffusive state from the specific site occur with rate constant k2 and from a non-specific site with rate constant k3. Schematic (right) illustrating that a TF arrives at a random site and scans the DNA until it finds a specific site from which it can subsequently unbind. (F) Simulation of (E) to calculate the dwell time, which is defined as the time spent on the DNA, either bound or sliding, showing the emergence of power-law behaviour (red line, PL exponent 0.5, k1=10 a.u, k2=1 a.u, k3=10 a.u). (G) Schematic of the energy landscapes, representing the different binding affinities and the local microenvironment denoted as potential wells with different depths. (H) Numerical simulation of (G) showing the emergence of power-law behaviour (blue line). See also **Appendix A** for details.


A
Bi-exponential model
Diffusive
Fast
Slow
B
Survival (log-scale)
time(a.u) - log scale
1st exponential
2nd exponential
Simulation
C
Kinetic model
Diffusive
Fast
Slow
D
Survival (log-scale)
time(a.u) - log scale
Simulation
E
Diffusion and binding on a chain
F
Survival (log-scale)
time(a.u) - log scale
Simulation
Power-law fit
G
Broad distribution of affinities model
Energy Landscape
Position
H
Survival (log-scale)
time(a.u) - log scale
Simulation
Power-law fit

We next extended the bi-exponential model to allow for transitions between the slow and fast components, which we call the kinetic model (**Figure** 3-6**C**). This model is a generalization of the bi-exponential model above. We note that due to the resolution limit (~30nm), any transitions between specific and non-specific bound states cannot be distinguished experimentally. We found that for this extended model, the resulting survival distribution again corresponds to a bi-exponential distribution, with the exponential parameters as the eigenvalues of the transition matrix (**Appendix A 1.3**). Stochastic simulations were performed as before, and the resultant distribution, displayed in **Figure** 3-6**D**, again clearly demonstrates exponential behavior. An implication of the kinetic model is that simple interpretations of the exponential parameters as kinetic transition rates in either of the exponential models is not straightforward, since each rate constant might represent transitions between multiple hidden states and therefore the average dwell time may not necessarily represent the characteristic timescale of a particular interaction with chromatin.

Several theoretical studies have posited that TF searching for and “final” binding to its cognate site on the DNA involves a combination of bulk diffusion in the nucleus, 1D sliding along the DNA, hopping and translocation, and the theoretical search times for the TF to find specific sites in this framework have been estimated (M. Bauer & Metzler, 2012; Berg & Blomberg, 1976; Marklund et al., 2013). In this model, TFs will have a multiplicity of short-lived bound states that must be accounted for in the analysis of dwell time data. To do so, we modelled TF movement on the DNA as hopping on a circular chain composed of specific

and non-specific sites (**Figure** 3-6**E**). The main assumption in this model (**Appendix A 1.4**) is that the number of non-specific sites on the DNA is much larger than the number of specific sites. This is biologically reasonable as only a few to tens of thousands of specific sites are bound by any TF according to genome wide studies (J. Wang et al., 2012), while the entire genome contains millions of "other" potential chromatin sites. Since the length of time spent bound to the DNA depends on the number of non-specific sites visited before binding to and dissociating from the specific site, this will manifest itself as a continuum of effective binding affinities. An analytical solution can be found for the simplest case in which there is a single specific binding site and the TF can only unbind from this specific site (**Appendix A 1.4.3**). Biologically, this situation represents the case in which the TF finds the specific site and stays bound or rebinds rapidly upon dissociation. This has been hinted at by evidence of asymmetric diffusion prior to TFs binding (A. S. Hansen, Amitai, Cattoglio, Tjian, & Darzacq, 2020) and protein-protein interaction mediated phase separation of different transcription factors (Chong et al., 2018). A simulation based on the model gives rise to asymptotic power-law behavior at time scales compatible with specific binding, for a number of representative parameter values (**Figure** 3-6**F**).

Finally, TFs can bind chromatin regions with varying physical microenvironments and motif degeneracy (Brodsky et al., 2020; Schöne et al., 2016; F. Zhu et al., 2018). These local properties affect the binding affinity of the TF. Given the heterogeneities in local organization and nuclear structure, TF binding sites on chromatin can be viewed as a collection of traps with a distribution

of trap depths (**Figure** 3-6**G**), analogous to binding affinities. If the binding affinities across different nuclear microenvironments and response elements are broadly and continuously distributed (for instance, exponentially distributed binding affinities), we can analytically demonstrate that the dwell times will asymptotically approach a power-law (Bouchaud & Georges, 1990), as confirmed by simulations (**Figure** 3-6**H** and **Appendix A 1.5**). In summary, we present phenomenological models that give us a framework to evaluate possible outcomes in SMT data.

3.3.4. Dwell time distributions of GR and other TFs follow power-law behaviour

Having developed a theoretical framework to evaluate TF dynamic behaviour, we next explored which model better explains GR dynamics. We fit the survival distributions of GR activated with Corticosterone (GR-Cort, **Figure** 3-7**A**), or with dexamethasone (Dex, 100 nM), a more potent, synthetic hormone (GR-Dex, **Figure** 3-7**B**) to bi-exponential, kinetic and power-law models. As evident from the distributions, the bi-exponential and kinetic models show qualitative deviations from the data. We then used metrics based on the Bayesian information criterion (BIC) (Schwarz, 1978) test to choose the best predictive model (**see Methods**). Indeed, our statistical analysis confirms that a power-law corresponds to the best predictive model based on these metrics over a fit to the bi-exponential or kinetic model [Delta-BIC1 is 114423 (1047.3) for GR-cort; 13572 (942.8) for GR-Dex for the power-law fit compared to kinetic model (bi-exponential model)]. Moreover, the power-law fits were also superior to a tri-exponential fit (Agarwal et al., 2017; Hipp et al., 2019) (see **Table** 3-2 **and Table** 3-3 for all statistical

comparisons). The apparent deviation in the tail of the distribution is due to the low number of data points, as shown by the increased confidence intervals. Surprisingly, we find that GR-Dex has a larger power-law exponent (β) than GR-Cort (*c.f.* **Figure** 3-7**A and Figure** 3-7**B**), suggesting longer dwell times for the less potent ligand (Cort). This counterintuitive result is nevertheless consistent with a previous report correlating residence times with transcriptional bursting, wherein longer residence times (GR-Cort) correspond to a larger burst size, while overall transcriptional output is greater in GR-Dex due to a higher bound fraction (Stavreva et al., 2019).

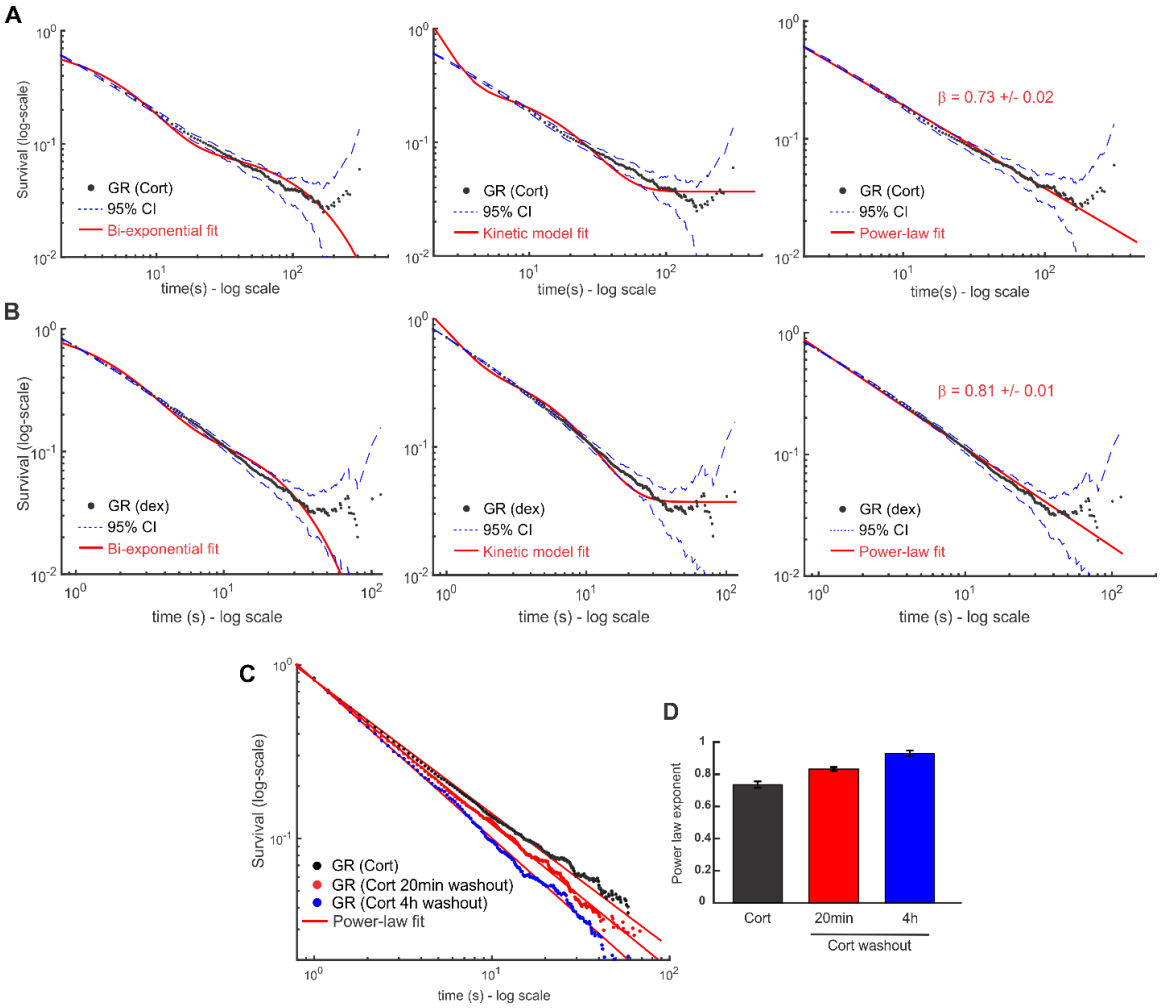


Figure 3-7 Dwell time distribution of GR follows power-law behaviour.
(A-B) Survival distribution of GR activated with corticosterone (Cort, panel A, acquired with 500ms exposure time and 1000ms interval) and dexamethasone (Dex, panel B, acquired with 100ms exposure time and 200ms interval) obtained from SMT data. Number of cells/number of tracks are 30/15732 for GR (Cort); 40/29211 for GR (dex). Red lines show the best fit obtained for the bi-exponential model (left), the kinetic model (center) and a power-law (right). Dashed lines show the 95% confidence intervals (CI). (C) Survival distribution of GR activated by Cort (black symbols) or following washout of the hormone under a 20 min (red) or a more stringent 4h washout protocol (blue). Solid lines show fits to power-law model. Data acquired with 100ms exposure time and 200ms interval. Number of cells/number of tracks are 65/23172 for GR (Cort); 62/22530 for GR (Cort 20 min washout); 61/16611 for GR (Cort 4h washout). (D) Aggregate data for power-law exponents of fits to survival distribution of GR following stimulation by Cort, 20 min washout following stimulation and 4 h following washout. Errors represent 95% confidence interval. See **Table** 3-2 **and Table** 3-3 for details on data acquisition and statistical comparisons.

We found that the power-law model better describes the data independent of the acquisition conditions (**Figure** 3-8**A**), yielding the same exponent under different PB rates (**Figure** 3-8**B**). In contrast, a fit to a triple-exponential model

showed that the parameters are dependent on acquisition times (**Figure** 3-8 **C-D**). Further, the survival distributions obtained using a different tracking software [uTrack, (Jaqaman et al., 2008)] were very similar to our tracking algorithm, ruling out any artifacts due to tracking (**Figure** 3-8**E**). Finally, the power-law behavior of the survival distribution of GR is conserved even if we use a different tag such as SNAP-Tag (Gautier et al., 2008) (**Figure** 3-8**F**).

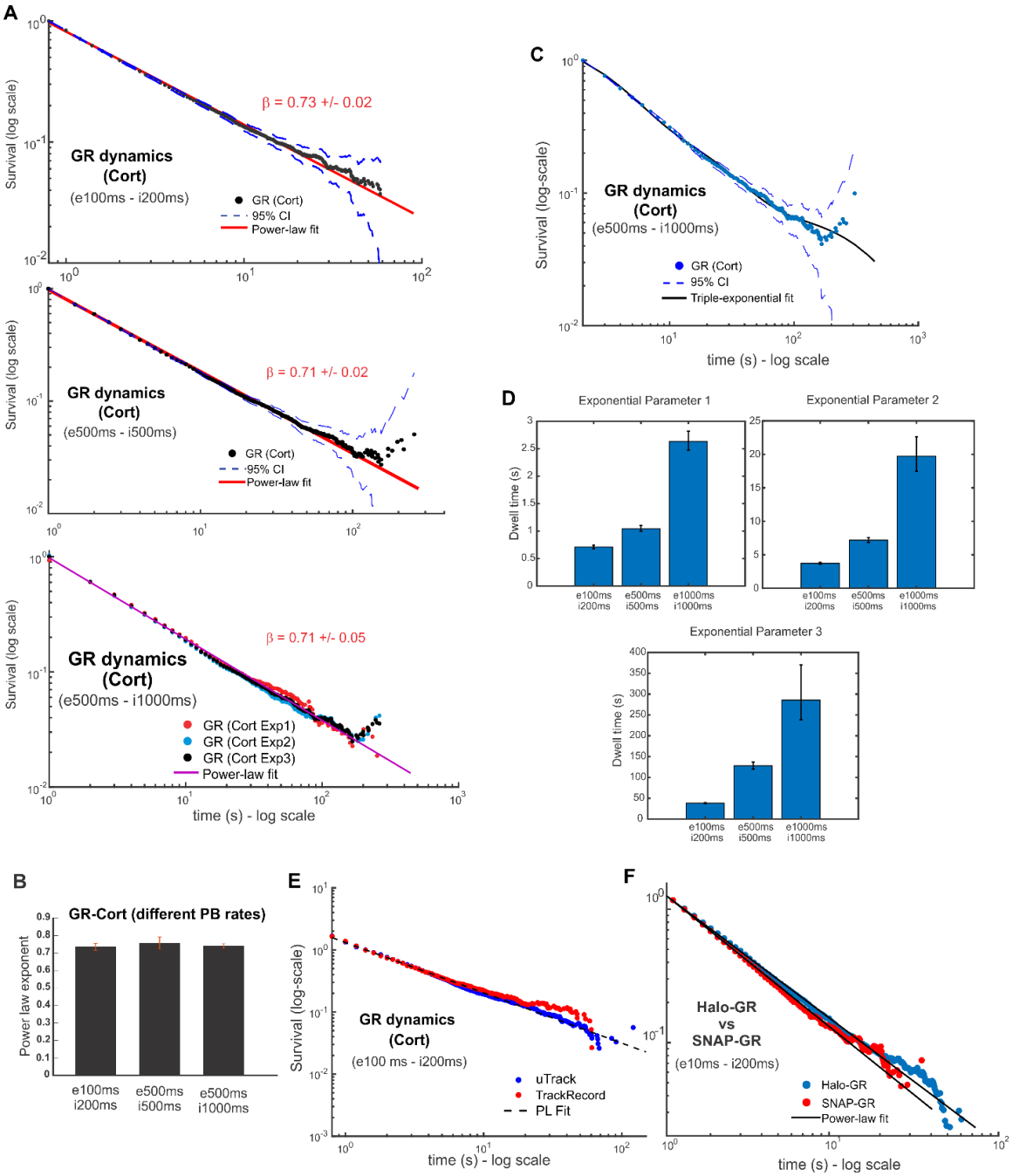


Figure 3-8 GR dwell times follow power-law distribution.

(**A**) Power-law fits (red line) to the survival distribution of Corticosterone (Cort) activated GR (black symbols) from SMT data acquired at the indicated exposure (e) and interval times (i). Number of cells/number of tracks are 65/23172 for GRe100ms/i200ms; 34/37953 for GRe500ms/i500ms; and 30/15732 for GRe500ms/i1000ms. Bottom panel shows independent replicates of Cort-treated GR data acquired at 1000ms intervals, exemplifying reproducibility between SMT experiments. (**B**) Power-law exponent of GR-Cort SMT data under different photobleaching rates, generated by modulation of acquisition conditions, as indicated. Error bars represent 95% confidence intervals. (**C**) Survival distribution of GR-Cort fit to a triple exponential function. (**D**) The three exponential parameters of the triple exponential fit for different acquisition conditions as indicated shows that these parameters depend on acquisition conditions. (**E**) Survival distribution of GR dynamics tracked with uTrack (bue symbols) or TrackRecord (red symbols) software packages. Dashed line

shows the power-law fit for uTrack. Number of cells/number of tracks are 65/23172 for TrackRecord; 40/11890 for uTrack. (**F**) Survival distribution of GR either tagged with HaloTag (Halo-GR, blue symbols) or SnapTag (SNAP-GR, red symbols) and the corresponding power law fits (black line). Number of cells/number of tracks are 67/9374 for Halo-GR; 50/7023 for SNAP-GR. See **Table S1** for details on data points and statistics.

Previous work has largely assumed that the dynamics of non-specific binding is well described by a single exponential component with a much shorter dwell time than specific binding (Chen et al., 2014; A. S. Hansen et al., 2017; Mazza et al., 2012a; Morisaki et al., 2014; D. M. Presman et al., 2017). However, heterologous proteins have also been reported to show power-law behavior for the dwell times (Caccianini et al., 2015; Normanno et al., 2015). To examine the dynamics of non-specific binding, we inactivated GR by washing out the hormone for 20 minutes, which greatly reduces specific binding as measured by chromatin immunoprecipitation (Stavreva et al., 2015). Interestingly, GR still exhibits power-law behaviour both for brief (20 min) washout, as well as longer washouts (four hours) (**Figure** 3-7**C**), although with shorter dwell times as indicated by a larger power-law exponent (**Figure** 3-7**D**).

To further establish the generality of our observations, we tested the dwell time distributions of different proteins previously characterized as bi-exponentially distributed by SMT (A. S. Hansen et al., 2017; Paakinaho et al., 2017; E. E. Swinstead et al., 2016). As with GR, both ER and FOXA1 exhibit power-law distributions (**Figure** 3-9**A**), with similar dynamics ($\beta = 0.698 \pm 0.005$ for ER and 0.742 ± 0.003 for FOXA1) but slower compared to GR ($\beta = 0.828 \pm 0.004$). This remains consistent with our previous observations wherein GR was more dynamic than ER and FOXA1 (E. E. Swinstead et al., 2016). Similarly, one of the major

ATPase subunits of the SWI/SNF chromatin remodelling complex, SMARCA4, also exhibits a residence time distribution compatible with power-law behavior ($\beta = 0.845 \pm 0.005$, **Figure** 3-9**B**). Surprisingly, even the dynamics of the 11-zinc finger DNA-binding protein CTCF, involved in genome architecture among other functions (A. S. Hansen et al., 2017), is better described by a power-law ($\beta = 0.55 \pm 0.02$, **Figure** 3-9**B**). Taken together, our results indicate that the bi-exponential model might not properly reflect the dynamics of a wide range of chromatin interacting factors, and that it underestimates TF dwell times on chromatin. Thus, the power-law distribution emerges as a better descriptor of single-molecule dynamics, at least for the proteins tested.

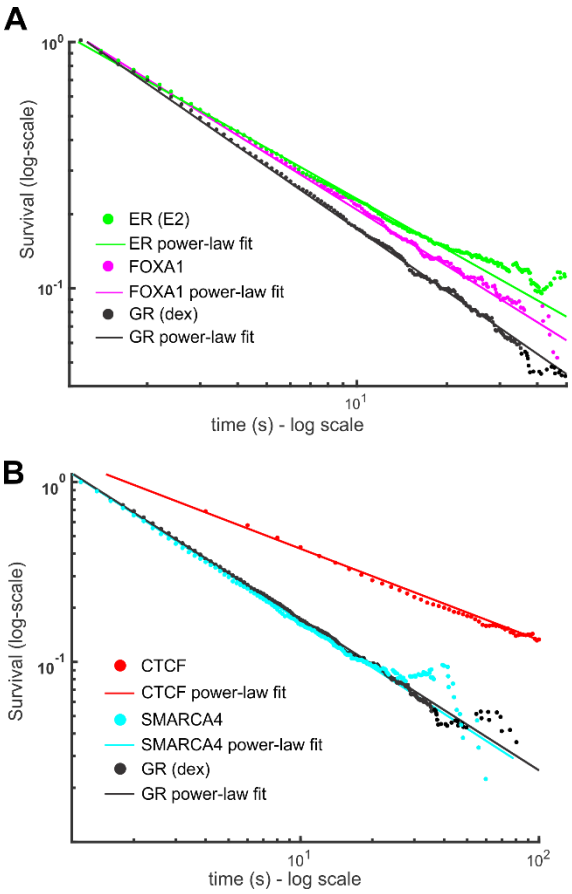


Figure 3-9 Dwell time distributions of TFs and other chromatin associated proteins show power-law behaviour.

(A-B) Survival distribution calculated from SMT data of the Halo-Tagged oestrogen receptor (ER, activated with oestradiol, E2) (green), FOXA1 (magenta), CTCF (red) and SMARCA4 (cyan). GR (activated with dex, black) is shown for comparison in both plots. Data was acquired at 10ms exposure time with 200ms interval. Number of cells/number of tracks are 60/17823 for ER; 41/12864 for FOXA1; 50/7023 for SMARCA4, 40/29211 for GR (dex). CTCF data was acquired with a 10 ms exposure time and a 2000ms interval. Number of cells/number of tracks are 48/11606 for CTCF. Symbols are SMT data and solid lines are power-law fits to the data (see **Table** 3-2 **and Table** 3-3 for comparison and number of data points).

In conclusion, our analysis reveals hitherto unobserved features of the distribution of mammalian TF residence times (power-law vs. bi-exponential). This, in turn, suggests that specific and non-specific binding cannot be identified as two distinct populations with discrete (and measurable) residence times.

3.4. Discussion

In the present study, we propose a modified SMT pipeline with an improved photobleaching correction method to prevent bias of the dwell time distribution of TFs, and test underlying models using different statistical metrics. We are now able to reconcile data acquired under different experimental conditions whereas previous attempts were not successful (Paakinaho et al., 2017).

We find that GR, as well as other TFs (ER and FOXA1), the chromatin remodeler SMARCA4 (also known as BRG1), and the insulator protein CTCF, all appear to exhibit power-law dynamics. It is generally accepted that to confirm this distribution, at least two orders of magnitude (both in x and y axes) should behave linearly on a log-log plot (Stumpf & Porter, 2012). This would require measuring TF binding up to several minutes (> 10 min), which is not currently feasible by SMT. Nevertheless, while there is a possibility that the power-law truncates at some point for really long binding times, we have enough statistical evidence to conclude that the power-law fit is a better predictor than a bi-exponential model over the observable experimental timescales. Examining whether more (or all) of the TFs originally characterized by bi-exponential behavior are better described by a power-law exceeds the scope of this work and needs to be evaluated on a TF-by-TF basis.

Our observation of power-law behavior of GR residence times suggests a model with a continuum of DNA-bound states rather than discrete non-specific/specific binding times (Figure 7). Consistent with this model, inactivation of GR by washing-out of the hormone revealed that the dwell time distribution also

follows a power-law, indicating no apparent dynamical differences between specific and non-specific binding, as previously observed for bacterial proteins expressed in mammalian systems (Caccianini et al., 2015; Normanno et al., 2015). Nevertheless, the overall residence times decrease when the receptor is less active, suggesting that a majority of the longer events observed with the fully activated receptor are associated with productive transcription as previously reported (Callegari et al., 2019; Chen et al., 2014; Mazza et al., 2012a; Morisaki et al., 2014; Paakinaho et al., 2017).

If all binding affinities lie on a smooth continuum, is it possible to distinctly define non-specific binding or assign such states to a set of sequences that can be bound but are kinetically indistinguishable? TFs can interact non-specifically with chromatin through electrostatic interactions with predicted short binding times and can also diffuse along DNA *in vitro* (Dahirel, Paillusson, Jardat, Barbi, & Victor, 2009). On the other hand, non-specific protein-DNA interactions can be broadly distributed with high binding energies at the tail of the distribution due to symmetric sequences in DNA that might facilitate long search times on chromatin (Afek, Schipper, Horton, Gordân, & Lukatsky, 2014). In an SMT experiment, due to resolution limitations and natural thermal fluctuations, the kinetics of diffusion along DNA, rapid binding/rebinding to non-specific sites, and transient trapping of a protein due to protein-protein interactions may appear indistinguishable from specific binding (kinetic model, **Figure** 3-6**C**). Therefore, non-specific binding is defined by the limits of the measurement, and not necessarily discernible as a

different mode of binding. However, any sequence with higher affinity (with long dwell times) is likely to be specific.

An important characteristic of power-law distributions is that for exponents lower than or equal to one (as in our case), the mean is not a well-defined quantity (Newman, 2005). This implies that the mean can vary enormously from one measurement to the next and it is a limited measure of the process. Interestingly, the heavy tails of power-law distributions imply that the probability of long-lived events is not negligible. This raises the possibility that productive binding events, although rare, may have dwell times much longer than previously appreciated, as indicated by the right-skewness of the distribution. We have recently shown a temporal correlation between GR dwell times and bound fraction with the length and frequency of transcriptional bursting (Stavreva et al., 2019). A similar behavior has been observed in yeast with the Gal4/GAL3 model (Donovan et al., 2019). However, non-specific binding can also result in TF binding events with long residence times, the implications of which are still not known. Critical efforts are required to investigate whether the slow(er) stops seen in SMT are matched exclusively to specific interactions with chromatin (Lerner et al., 2020). For example, GR binding precedes RNA synthesis by ~3 min (Stavreva et al., 2019). Alternatively, a sub-population of these “stops” could correspond to microscopic regions in the nucleus where diffusion is severely limited, due to transient interaction with “clustered” structures such as foci observed for GR (Stortz et al., 2017), or another hitherto unknown mechanism.

The emergence of power-law might reflect the wide distribution of binding affinities in the nucleus. This broad distribution of affinities is puzzling but may be explained by a diverse set of non-mutually exclusive mechanisms. First, the intrinsic affinity of TFs for DNA likely follows the dwell time power-law model, ranging smoothly from “non-specific” to the highest affinity. Indeed, it has been shown that TF-DNA binding affinities ranges from low-affinity, not necessarily detectable by ChIP, to high affinity, corresponding to strong CHIP-seq peaks (Rastogi et al., 2018). In addition, microfluidic studies (k-MITOMI) of the mouse TF Zif268 reported binding times in the 0.2s-200s range, consistent with our observation of *in vivo* dwell times (Geertz, Shore, & Maerkl, 2012). Moreover, tag density in ChIP-seq experiments has been correlated to TF affinity, and are also power-law distributed, at least for CTCF and a few TFs (Jothi, Cuddapah, Barski, Cui, & Zhao, 2008). While this suggests a potential connection between ChIP-seq data, occupancy, and dynamics of TF binding, further experiments are needed to demonstrate causation.

Second, nuclear structure and the chromatin environment is known to be highly heterogeneous (Finn & Misteli, 2019; Liu & Tjian, 2018). Thus, TFs will encounter a wide variety of chromatin states (compacted fibers, different nucleosome modification conditions, etc.). Moreover, affinities for the thousands of alternative binding sites in response elements likely vary significantly. Furthermore, recent work points to the presence of transcriptional hubs and liquid-liquid phase separation domains (Chong et al., 2018; Garcia et al., 2021a; Hnisz, Shrinivas, Young, Chakraborty, & Sharp, 2017; H. Lu et al., 2018; Sabari et al.,

2018; Stortz, Pecci, Presman, & Levi, 2020) that contribute to the complexity of nuclear organization. If TFs exhibit different dynamical properties in these structures, it would not be surprising to find a broad variation in binding affinities. Third, power-law distributed dwell time distributions can emerge as a consequence of the molecular kinetics of the protein itself, as recently reported *in vitro* by RNA polymerase II in bacteria (Janissen, Eslami-Mossallam, Artsimovitch, Depken, & Dekker, 2020). Fourth, the heterogeneity in the searching mechanism of TFs may affect the effective affinity constant observed in SMT experiments (M.J. Saxton, 2020). In support of the latter, while heterologous expression of TetR in mammalian cells showed power-law behaviour for non-specific binding, it could still be described as an exponential on an artificially (and single) specific DNA binding array (Normanno et al., 2015). Thus, the intrinsic nature of the searching mechanism of any DNA-binding protein in native chromatin may be governed by power-law dynamics. In addition, the heterogeneity of dwell times in the thousands of response elements for an endogenous TF could explain why GR can exhibit power-law tails as opposed to TetR, which can only bind to one artificial array site. Interestingly, a study in yeast (Mehta et al., 2018) reports that both the TF Ace1p and the chromatin remodeler RSC binding follow a bi-exponential binding distribution in cells containing a natural tandem of ten CUP1 (Ace1p responsive) genes. This dynamic and discrete behaviour, in contrast with our GR data, can be explained by the particular and homogeneous chromatin environment of a single array of specific sites. Consequently, we speculate that a broad distribution of binding affinities due to a whole population of different binding sites (thousands in

the case of GR) may result in power-law behaviour (**Figure** 3-10). In this sense, a few defined states and the continuum may just be two ends of a spectrum. Thus, we might need to revisit the classification of non-specific TF binding solely as static interactions with random DNA sequences but rather arising from a dynamical process involving biophysical properties of the nuclear microenvironment, chromatin, and protein-protein interactions. Consistently, we have recently described “binding events” which are independent of chromatin interactions, power-law distributed, and depend on intrinsically disordered regions (IDRs) (Garcia et al., 2021a).

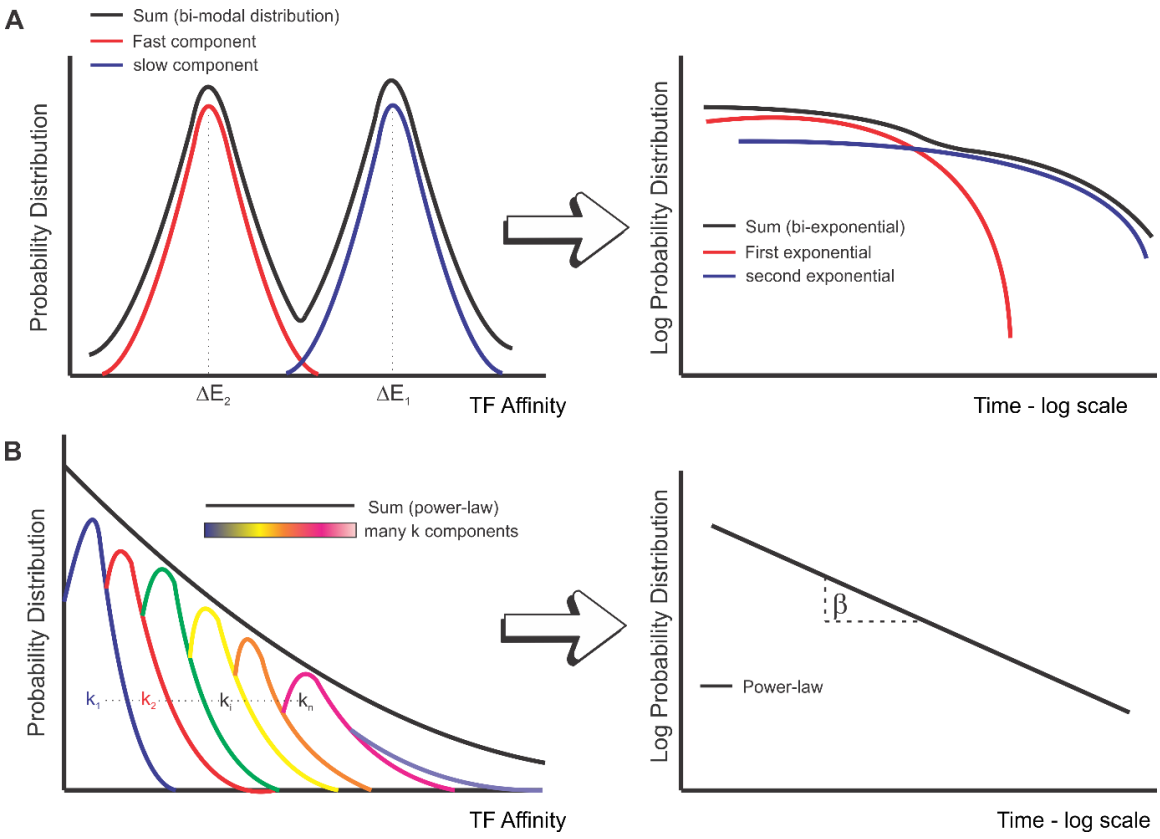


Figure 3-10 Heterogeneity in binding affinities can lead to a power-law behaviour of survival time distributions.

A) Schematic of the binding affinity distributions for a bi-exponential model. In this model, specific sites (blue) and non-specific sites (red) have a well separated and narrow distribution of affinities (ΔE, left graph), which results in a bi-exponential behaviour of the overall survival distribution (right graph, black curve). (B) Schematic showing a broad distribution of TF affinities (black line) which

arises as a superposition of multiple sites with closely spaced affinity distributions (depicted in different colours in the left graph). Note that the distributions get progressively wider. This distribution of affinities may explain the emergence of power-law behaviour (characterized by the exponent, β) in the residence time of TFs (right graph).

While SMT methodology gives us the opportunity to study TF dynamics with unprecedented temporal and spatial resolution, it still has some major drawbacks. The sparse labelling conditions needed to resolve individual molecules severely limit the possibility of following all functional TFs at a time, and therefore may affect the implementation of a two-color version where two different proteins interact at the single-molecule level. In addition, we still do not have direct measurements of the affinity at specific sites which makes it difficult to functionally distinguish between specific and non-specific binding. Nevertheless, the current major limitation in SMT is the photostability of the fluorophore, which limits the dynamical range of experiments and prevents accurate analysis of long TF trajectories that sample over different binding and/or diffusive events. Our temporal measurement window will improve with better, more stable fluorophores. Until then, our proposed pipeline allows us to have better estimates on the dynamics and the residence time distribution of TFs.

In summary, by incorporating an improved PB correction method and testing different models, we showed that the survival distribution of GR and other TFs dwell times does not follow an exponential model. Ultimately, if there is a way to define or distinguish non-specific from specific binding, our results indicate that it cannot be based on their global residence times. However, the data is consistent with a power-law distribution, which we suggest may arise generically due to

heterogeneities in TF interactions with DNA or in the diffusive environment in the nucleus. Thus, the slope of the residence time distribution does provide an estimate of the overall affinity and can be used to compare TFs and their function under different conditions.

3.5. Methods

3.5.1. Plasmid constructs

The pHaloTag–GR has been previously described (Morisaki et al., 2014). The construct expresses rat GR fused with HaloTag protein (Promega, Madison, WI, USA) in the C-terminal domain under the CMVd1 promoter. The pHaloTag-H2B has also been previously described (Mazza et al., 2012a). The N-terminus of H2B is fused with the HaloTag. pHaloTag-H3 and-H4 were purchased from Promega (pFN21AE1298 and pFN21AE0273, respectively). The pHaloTag-ER and pHaloTag-FoxA1 has been previously described (E. E. Swinstead et al., 2016). The pHalo-CTCF expresses the mouse CTCF with HaloTag fused in the C-Terminal domain. It has been generated by PCR amplification from the pCTCF-GFP vector (Nakahashi et al., 2013) and sub cloned into the pHalo-GR previously cut with PvuI and XhoI restriction enzymes (New England Biolabs). The pHalo-SMARCA4 was purchased from Promega (pFN21AE0798). The pSNAP and pSNAP-GR have been previously described (D. M. Presman et al., 2017).

3.5.2. Cell culture and transfection

The 3617 mouse mammary adenocarcinoma cell line used in this study as well as the GR knock-out subclone expressing Halo-GR has been previously described (Paakinaho et al., 2017; D. M. Presman et al., 2017). Cells were routinely cultured in high glucose DMEM supplemented with 10% fetal bovine serum and 2 mM L-glutamine at 37°C in a CO2-controlled humidified incubator. The cell line contains a stable integration of the rat GFP–GR under tetracycline regulation (Hager et al., 2000). To prevent expression of GFP–GR, the cells were grown in the presence of 5 µg/ml tetracycline (Sigma-Aldrich, St. Louis, MO, USA).

5 million cells were electroporated using BTX T820 Electro Square Porator (Harvard Apparatus, Holliston, MA, USA) in 100ul of DPBS with 2.5 ug of plasmid. 25 ms pulses of 120v were used and cells were resuspended in fresh media. Single-molecule imaging experiments were set up as follows: 100,000 electroporated cells were seeded onto each well of a 2-well Lab-Tek chamber (1.5 German borosilicate coverglass, Thermo Fisher, Waltham, MA, USA) in high glucose DMEM supplemented with 10% FBS (Life Technologies), 2mM L-glutamine, 5 µg/ml tetracycline, and cultured overnight. The media was then replaced with high glucose DMEM supplemented with 10% charcoal stripped FBS (Life Technologies), 2mM L-glutamine, 5 µg/ml tetracycline, and incubated at 37°C for at least 24 hours before labeling.

3.5.3. Fluorescent labeling of Halo-tagged molecules and hormone treatments

Transfected cells were incubated with 5 nM JF549-HaloTag or 10 nM cpSNAP-tag (Grimm et al., 2015) ligand for 20 min at 37°C. Stably integrated Halo-GR cells were incubated with 0.5 nM JF549-HaloTag for 20 min at 37°C. Free ligand was depleted by washing three times with phenol red free DMEM media (supplemented with 10% charcoal-stripped FBS and 5 µg/ml tetracycline) in 15 min intervals at 37°C. Next, cells were treated with 600 nM Corticosterone (Cort) (Sigma-Aldrich) or 100 nM Dexamethasone (Dex) (Sigma-Aldrich), or 100 nM estradiol (Sigma-Aldrich) and incubated for 20 min at 37°C before imaging. For wash-out experiments, cells were washed with media three times for 4 different intervals (every 15 minutes for 1 hour or every hour for 4 hours) after 20 minutes of hormone treatment and finally imaged.

3.5.4. Image acquisition for single-molecule tracking and analysis

A custom Highly Inclined and Laminated Optical sheet (HiLO) microscope was used as previously described in detail (D. M. Presman et al., 2017), with an objective heater to reduce drifting. Briefly, the custom-built microscope from the CCR, LRBGE Optical Microscopy Core facility is controlled by µManager software (Open Imaging, Inc., San Francisco, CA.), equipped with an Okolab stage top incubator for CO2 (5%) and temperature control (37°C), a 150X 1.45 numerical aperture objective (Olympus Scientific Solutions, Waltham, MA), a 561nm laser (iFLEX-Mustang, Excelitas Technologies Corp., Waltham, MA), and an acousto-

optic tunable filter (AOTFnC- 400.650, AA Optoelectronic, Orsay, France). Images were collected on an EM-CCD camera (Evolve 512, Photometrics). Tracking was performed in MATLAB (version 2016a, The MathWorks, Inc., Natick, MA) with the custom software TrackRecord [version 6, originally developed elsewhere (Mazza et al., 2013) and updated in-house]. For step-by-step instructions, please refer to the User Manual file in the supplemental files. Briefly, in TrackRecord, to analyze each time series, data were filtered using top-hat, Wiener, and Gaussian filters, then particles were detected, fitted to two dimensional gaussian function for "super resolution" and finally tracked using a nearest neighbor algorithm (D. M. Presman et al., 2017). Particle trajectories are divided into mobile and immobile. The displacements of histones H2B characterize the thermal jiggling of the DNA and from it, two parameters are extracted called R_{min} and R_{max}. R_{min} corresponds to the maximum displacement of 99% of histones at a time-lag of 2 frames (frame to frame displacement) while R_{max} corresponds to the maximum displacement of 99% of histones at a time-lag of shortest track. The shortest track is calculated using the diffusion coefficient of GR (~5 $\mu m^2/s$) to minimize tracking errors as explained elsewhere (Mazza et al., 2012a). The immobile tracks are used to calculate the survival distribution using the Kaplan-Meier estimate. The 95% confidence interval was estimated using Greenwood's Formula. All fits performed to the data were implemented with the nonlinear least square method using bi-square weights due to the noise on the tail of the survival distribution. Parameters used for acquisition conditions and analysis are shown in Table 3-1.

Exported tracking data was further analyzed in MATLAB by a custom-made script (see User Manual for details). For comparison and control purposes, we also performed tracking using u-Tack (Jaqaman et al., 2008). Briefly, we used the "Gaussian Mixture-Model Fitting" under default parameters for particle detection and localization. The tracking was then performed with the following values: Problem dimensionality = 2; Maximum Gap to close = 2; Minimum Length of Track Segment from the First Step = 4; Do segment merging = checked; Do segment splitting = checked. Finally, we chose the "Cost functions" and "Kalman Filter functions" to the "Brownian + Directed motion" model.

3.5.5. Photobleaching correction

The modified correction method is based on histone data as a proxy for the fluorophore stability as originally performed elsewhere (Callegari et al., 2019; A. S. Hansen et al., 2017; Teves et al., 2016; Zhen et al., 2016). One caveat still common to all methods described and applied here is the assumption of homogenous illumination, which unfortunately does not occur in HiLO set ups, as the laser hits the sample at an inclined angle [discussed elsewhere (D. M. Presman et al., 2017)]. A first step involves SMT of histones under the same conditions that the TF of interest will be imaged, as previously described (Callegari et al., 2019; A. S. Hansen et al., 2017; Teves et al., 2016; Zhen et al., 2016). We tracked individual H2B, H3 or H4 molecules using HiLO by sub-optimal transient transfection of HaloTag-fused histones, labeled with JF549 HaloTag ligand. The three histone variants we tested presented statistically similar dynamics (**Figure** 3-11**A**). We continued with H2B for all further experiments. Histone genes are

primarily transcribed upon entry into S-phase of the cell cycle (Ewen, 2000). Due to our transient transfection approach, HaloTag-H2B proteins will be translated during interphase and therefore some histones will not be incorporated into chromatin at the time of acquisition. Hence, the survival distribution of H2B will be composed of PB kinetics and a diffusive/transient binding component. To account for this behavior and assuming PB kinetics at the single-molecule level are exponentially distributed, the survival distribution of H2B is fit to an exponential family with three components (**Figure** 3-11**B**). This constitutes the second step in the protocol, which only differs thus far from previous work in the fitting to three exponential rather than two-exponentials (Callegari et al., 2019; A. S. Hansen et al., 2017; Teves et al., 2016); or fitting to two exponential with an offset (Zhen et al., 2016). The faster components characterize the dynamics of histones that have not been stably incorporated into chromatin, while the third (slower) component describes the PB kinetics of the fluorophore. The invariance of the first two components to photobleaching conditions strongly suggest they are indeed due to the dynamics of unincorporated histones, tracking errors and shortest track selection (**Figure** 3-11**C**). To confirm that the third component quantifies PB kinetics and not the intrinsic dynamics of H2B, we calculated PB lifetimes using histones H3 and H4 with the same statistical results (**Figure** 3-11**D**). Finally, the third step for corrects the binding dynamics of the TF by using the experimental (observed) TF distribution and the PB dynamics. It is the ideal measurement where neither photobleaching nor sample drift occur. The novelty of our approach is that we use the third exponential distribution of H2B as a proxy for photobleaching,

while other methods use the entire H2B distribution (Callegari et al., 2019). In this sense, no assumption regarding the survival distribution of the TF is made, and the empirical survival distribution is corrected by the third exponential component of the H2B survival distribution.

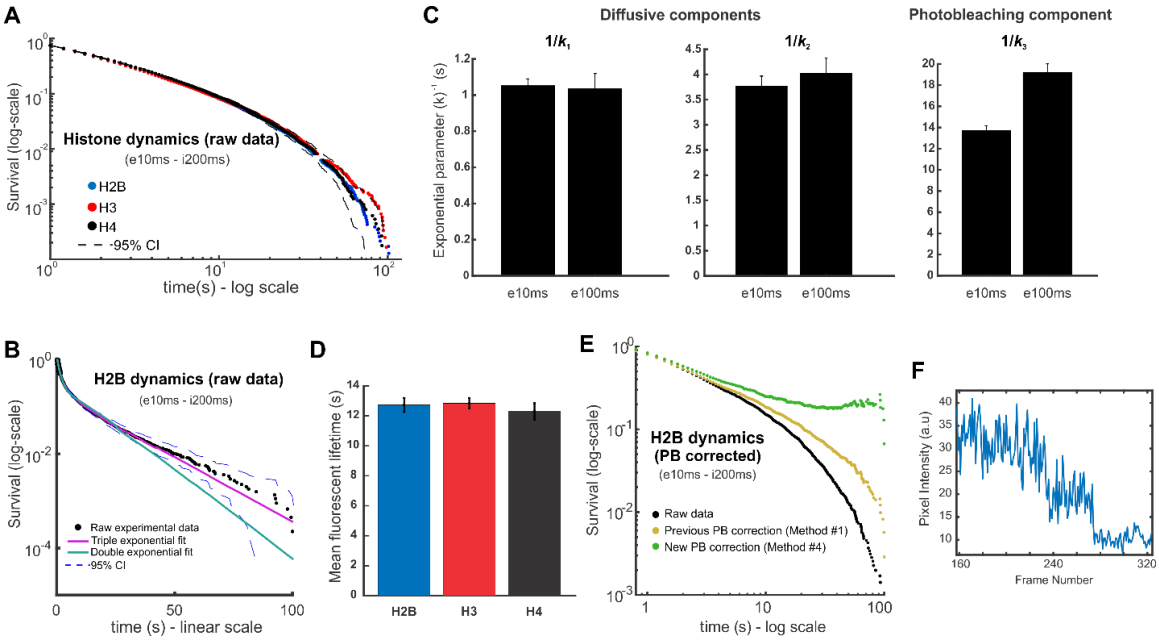


Figure 3-11 Histone dynamics as a proxy for photobleaching correction.
(A) Survival distribution of histones H2B, H3 and H4 acquired under the same acquisition parameters as indicated (e, exposure time, i, interval). Number of cells/number of tracks are 100/36625 for H2B; 59/11708 for H3; 43/15601 for H4. (B) Fit of H2B survival distribution to a double exponential and triple exponential. A triple exponential better represents the experimental data where the slower component corresponds to the photobleaching rate in the focal plane. CI is the confidence interval. (C) Fitting the H2B data from two different exposure conditions (10ms and 100ms) to a triple exponential model gives the exponents k1, k2, k3. The bar graph shows the mean +/- 95% confidence interval. (D) Mean fluorescence lifetime calculated as 1/ k3 where k3 is the slowest rate of the triple exponential. H2B (12.73 +/- 0.46 s), H3 (12.84 +/- 0.35 s) and H4 (12.30 +/- 0.55 s). Errors represent 95% confidence interval. (E) Survival distribution of H2B dynamics (black) corrected with our previous correction method (yellow), or the upgraded one (green). Note that after correction with method #1, H2B still has a finite dwell time. However, after correction with method #4, H2B presents two different regimes: stably incorporated histones that have very long residence times (plateau) and a dynamic regime representing unincorporated histones nonspecifically interacting in the nucleus. (F) Representative intensity profile of a histone particle selected from the tail of the distribution, with a cumulative probability of less than 1%. These very long events (outliers) usually present multiple photobleaching steps, indicating multiple particles at the same point-spread function. This might explain the deviations in the long tail of the H2B distribution after photobleaching correction. See **Table** 3-2 **and Table** 3-3 for data points details

More formally, let $P(\tau_{TF} \geq t)$, $P(\tau_p \geq t)$, $P(\tau_{TFreal} \geq t)$ be the survival distribution of an experimental particle, photobleached particle and a dynamic particle, respectively. The survival distribution of an experimental particle is the one typically measured in a SMT experiment; the survival distribution of a dynamic particle corresponds to the ideal measurement where neither photobleaching nor sample drift occur. We are interested in $P(\tau_{TFreal} \geq t)$

$$P(\tau_{TF} \geq t) = P(\tau_p \geq t;\ \tau_{TFreal} \geq t) \quad \text{(eq. 1)}$$

If a molecule is observed to live longer than t then it neither photobleached nor unbound from the DNA. These two processes are independent:

$$P(\tau_{TF} \geq t) = P(\tau_p \geq t)\, P(\tau_{TFreal} \geq t) \rightarrow\ P(\tau_{TFreal} \geq t) = \frac{P(\tau_{TF} \geq t)}{P(\tau_p \geq t)} \quad \text{(eq. 2)}$$

If the empirical survival distribution of photobleaching at the focal plane is available, then the dynamic survival distribution can be extracted from the microscopy data.

$P(\tau_p \geq t)$ is estimated by fitting the survival distribution of H2B by a triple exponential function of the form:

$$P(\tau_{his} \geq t) = f_1 e^{-\gamma_1 t} + f_2 e^{-\gamma_2 t} + f_3 e^{-\gamma_3 t} \quad \text{(eq. 3)}$$

where γ_3 corresponds to the photobleaching and γ_1; γ_2 the parameters of the dynamics of diffusive and/or unincorporated histones. The survival distributions are normalized with respect to the shortest track, for a shortest track of 6 frames and an acquisition interval of 200ms, the survival distribution is set up to $P(\tau \geq 1.2\ s) = 1$.

Finally, assuming that the third component of $P(\tau_{his} \geq t)$ corresponds to photobleaching:

$$P(\tau_p \geq t) = e^{-\gamma_3 t} \quad \text{(eq. 4)}$$

$$P(\tau_{TFreal} \geq t) = \frac{P(\tau_{TF} \geq t)}{e^{-\gamma_3 t}} \quad \text{(eq. 5)}$$

If we correct the H2B survival distribution with this method, we observe a predictable upward shift of the distribution (**Figure** 3-11**E**), in contrast to our previous methodology (Paakinaho et al., 2017), wherein H2B data still artifactually resembles the dynamics of a TF. The high fluctuations at the tail of the distribution are likely due to noise in the data and the appearance of multiple particles within the point spread function, as illustrated in **Figure** 3-11**F**.

3.5.6. Quantification and statistical analysis

For statistical analysis, all the parameters are reported by the ensemble average, standard deviation (s.d.) and number of observations. At least three biological replicates of SMT experiments were performed per condition. Two sample K-S test on the survival distribution were performed between replicates to confirm statistical reproducibility. Between 10 and 20 cells were imaged per SMT replicate. Each condition has at least 15000 tracks after analysis of SMT experiments. For survival distribution analysis, a statistical threshold of 5 tracks were implemented for visualization purposes only. Any point in the survival distribution with less than 5 cumulative tracks was not displayed in the figure. Data was not removed for fitting purposes. Fitting was done using non-linear least

squares, initially a best local fit was found and then 50 iterations were run to find a global solution.

Simulations were written in MATLAB to numerically verify the different models of TF using the Gillespie algorithm (Gillespie, 1977). Graphical inspection was used to qualitatively determine if a straight line was observed for multiple decades in the case of a power-law fit in a log-log plot. Two different metrics were used to determine the difference between exponential models and power-law models. The first metric corresponds to Bayesian information criterion (BIC) using the probability distribution function (PDF) corrected for photobleaching as the likelihood function. BIC is a criterion for model selection that penalizes for model complexity (number of free parameters in the model). The PDF of TF dwell times was normalized between the minimum and maximum observation range (BIC1). BIC1 is given by (James, Witten, Hastie, & Tibshirani, 2017):

$$BIC1(M) = kln(n) - 2ln\big(P(D|\hat{\theta}, M)\big) \quad \text{(eq. 6)}$$

where *M* corresponds to the model (power-law, bi-exponential or triple-exponential), *k* corresponds to the number of parameters of the model, $\hat{\theta}$ corresponds to the model parameters found by fitting, *D* the observed data and *n* the number of observations. $P(x|\hat{\theta}, M)$ corresponds to the realization probability of x given the model PDF with parameters $\hat{\theta}$. For SMT, *D* is the set of independent and identically distributed discrete experimental events and $P\big(D\big|\hat{\theta}, M\big)$ is calculated as follows:

$$P\big(D\big|\hat{\theta}, M\big) = \prod_{x \in D} \int_{x-\frac{\delta t}{2}}^{x+\frac{\delta t}{2}} p\big(x\big|\hat{\theta}, M\big)\, dx \quad \text{(eq. 7)}$$

Where $p(x|\hat{\theta}, M)$ corresponds to the PDF of the model after photobleaching correction. For instance, the bi-exponential PDF is given by:

$$p(x|\hat{\theta} = (\alpha, \beta), M_{DE}) = C(f_1 \gamma \alpha e^{-(\gamma+\alpha)t} + (1 - f_1)\gamma \alpha e^{-(\gamma+\beta)t}) \quad \text{(eq. 8)}$$

where α, β are the exponential parameters, γ the photobleaching rate and C a normalization constant.

The second metric, the evidence, in decibels (dB), for a particular model given the observed data and priors, was calculated to compare the alternative models explored. The evidence measures the probability of a particular model being the best predicting model in comparison with another model. For instance, for the power-law model (M_{PL}) the evidence versus the bi-exponential model (M_{DE}) and the triple exponential model (M_{TE}) is given by (Jaynes & Bretthorst, 2019):

$$E(M_{PL}|D, \hat{\theta}, A) = E(M_{PL}|A) + 10log_{10}\left[\frac{P(D|M_{PL}, \hat{\theta})}{P(D|M, \hat{\theta})}\right] \quad \text{(eq. 9)}$$

Where $M = M_{DE} \cup M_{TE}$

$$E(M_{PL}|A) = 10log_{10}\frac{P(M_{PL}|, A)}{P(M|A)} \qquad \text{(eq. 10)}$$

$$P(D|M, \hat{\theta}) = \frac{P(D|M_{DE}, \hat{\theta})P(M_{DE}|, A) + P(D|M_{TE}, \hat{\theta})P(M_{TE}|, A)}{P(M_{TE}|, A) + P(M_{DE}|, A)} \qquad \text{(eq. 11)}$$

where *A* corresponds to the priors; *P*, *D* and $\hat{\theta}$ as defined for BIC1. Uniform priors were used for all model comparisons. For instance, an evidence of 30 dB corresponds to a probability higher than 0.999 that the power-law model better describes the data in comparison with an alternative model tested. In general, a positive value of the evidence indicates that the corresponding model is a better

predictor of the data in comparison to the other tested models. If the evidence was not high enough to reach a conclusion about the comparison between the different models, more data was acquired until the evidence reached a satisfactory value.

Refer to Table 3-2 and Table 3-3 for all statistical results. Table 3-2 and Table 3-3 lists the evidence in dB for the models, the difference of BIC1 (denoted as Delta-BIC1) between the power-law model and bi/triple exponential models. A positive value of the difference in BIC1 implies a preference for the power-law model over the bi- or triple-exponential models.

Model Selection was performed in the following manner: Graphical inspection for linearity of the survival distribution on a Log-Log plot for at least 1.5 decades for power-law model consideration, the model with an evidence higher than 30 dB and a difference of BIC1 in accordance with the model (a positive value for power-law as a better model, negative value for bi/triple exponential models) was chosen. Evidence does not take into account model complexity and therefore the model selection is done jointly with BIC1.

3.5.7. Availability

Tracking was performed in MATLAB (version 2016a) with the custom software TrackRecord (version 6). The latest version of this software is freely available in:

https://github.com/davidalejogarcia/PL_HagerLab

Exported tracking data was further analyzed in MATLAB by a custom-made script also available at the provided link. For step-by-step instructions, please refer to the User Manual (Appendix B).

3.5.8. Tables

Interval (ms)	Exposure (ms)	Laser power (mW)	Frame Number	Maximum (Pixels)	Shortest track (frames)	Gap (frames)	R_{min} (µm)	R_{max} (µm)
200	10	0.96	600	4	4	2	0.23	0.31
200	100	0.29	900	4	4	2	0.21	0.29
500	500	0.16	800	4	2	2	0.23	0.29
1000	500	0.16	800	4	2	1	0.29	0.33

Table 3-1 Parameters used for each acquisition condition, and analysis of SMT data.

					Evidence			Delta-BIC1	
	Acquisition Interval (ms)	Exposure (ms)	Number of Tracks	Number of Cells	Bi-exponential	Power-Law	Triple-exponential	Bi-exponential	Triple-exponential
Halo-GR (Cort)	200	10	9374	67	-119.2	65.7	-69.2	62.78	55.62
Halo-GR (Cort)	200	100	23172	65	-996.1	76.0	-79.5	466.9	62.9
Halo-GR (Cort)	500	500	37953	34	-2255.1	2251.6	-6428.3	1047.3	2988.3
Halo-GR (Cort)	1000	500	15732	30	-701.7	698.1	-1049.0	331.1	508.5
Halo-GR (dex)	200	100	29211	40	-2028.5	2005.6	-2009.1	942.8	952.6
Halo-GR (Cort Wash 20 min)	200	100	22530	62	-601.4	597.9	-8662.1	285.0	4014.8
Halo-GR (Cort Wash 4h)	200	100	16611	61	-503.3	374.2	-377.8	239.5	198.7
Snap-GR (Cort)	200	10	9630	50	-316.4	190.4	-193.9	153.8	115.2
Halo Alone	200	10	19436	64	62.0	-65.5	-213.6	-21.5	91.8
Halo-ER	200	10	17823	60	-1455.1	1451.5	-1497.2	678.6	716.5
Halo-FOXA1	200	10	12864	41	-615.8	57.4	-60.9	291.9	54.7
Halo-SMARCA4	200	10	7023	50	-374.7	371.1	-679.8	180.0	337.0
Halo-CTCF	2000	10	11606	48	-47.8	38.7	-43.6	29.8	45.1
Halo-GR (dex, stable)	200	10	16450	60	-1107.1	1103.6	-11671	518.4	5402.2
Halo-GR (dex, transient)	200	10	7068	60	-408.8	405.3	-4211.3	195.9	1963.8
Halo-GR (Cort) uTrack	100	200	11890	40	-817.9	349.8	-353.3	384.9	186.9

Table 3-2 Data acquisition and statistical results TFs.
Table shows for the indicated TF: the acquisition interval (in ms), exposure time (in ms), number of tracks, the evidence for the models (in dB), and the difference of BIC1 (denoted as Delta-BIC1) between the power-law model and bi/triple exponential models. Since the kinetic model is a special case of the bi-exponential model, model comparison was limited to the power-law, bi-exponential and triple-exponential models.

	Acquisition Interval (ms)	Exposure (ms)	Number of Tracks	Number of Cells	Evidence	Delta-BIC1
					Triple-exponential	Triple-exponential
Halo-H2B	200	10	36625	100	936.7	411.710824
Halo-H2B	200	100	40652	63	1451.2	648.223959
Halo-H2B	500	500	20307	36	1993.8	898.953197
Halo-H2B	1000	500	15823	20	4372.2	1994.155983
Halo-H2B	2000	10	11047	66	0.1	-18.204499
Halo-H3	200	10	11708	59	594.7	256.287697
Halo-H4	200	10	15601	43	496.5	210.797995

Table 3-3 Data acquisition and statistical results Histones
Table shows the same information as Table 3-2 for H2B, H3, and H4 data. For histones, model comparisons were done between the bi-exponential and triple exponential models only.

Chapter 4. An Intrinsically Disordered Region Mediated Confinement State Contributes to the Dynamics and Functions of Transcription Factors

This chapter has been adapted from Garcia et al., Molecular Cell, 2021 (Garcia et al., 2021a). I was primarily responsible for the experimental design, single molecule tracking experiments (with help from Diego Presman, Gregory Fettweis and Ville Paakinaho), super resolution microscopy, cloning and genomics (with help from Thomas Johnson), data analysis and modeling.

4.1. Summary

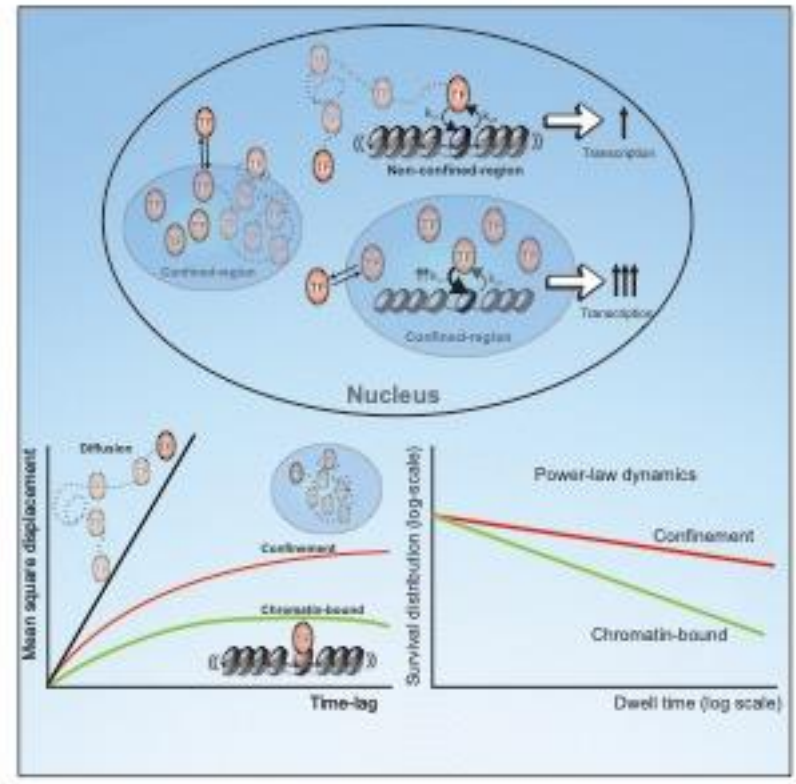


Figure 4-1 Graphical Abstract
We use a systems-level approach to analyze single-molecule tracks of the glucocorticoid receptor. In addition to the known chromatin-bound state, the authors characterized an IDR-mediated, long-lived confined state consistent with liquid condensates that can amplify transcriptional output by increasing the local concentration of TFs at enhancer sites.

Transcription factors (TFs) regulate gene expression by binding to specific consensus motifs within the local chromatin context. The mechanisms by which TFs navigate the nuclear environment as they search for binding sites remain unclear. Here, we used single-molecule tracking and machine-learning based classification to directly measure the nuclear mobility of the glucocorticoid receptor (GR) in live cells. We revealed two distinct and dynamic low-mobility populations. One accounts for specific binding to chromatin, while the other represents a confinement state that requires an intrinsically disordered region (IDR), implicated in liquid-liquid condensate subdomains. Further analysis showed that the dwell times of both subpopulations follow a power-law distribution, consistent with a broad distribution of affinities on the GR cistrome and interactome. Altogether, our data link IDRs with a confinement state that is functionally distinct from specific chromatin binding and modulates the transcriptional output by increasing the local concentration of TFs at specific sites.

4.2. Introduction

The specific binding of transcription factors (TFs) to regulatory sites embedded within promoter-proximal elements and enhancers guides the assembly of the transcription apparatus and ensures the expression of target genes (Lazar, 2017). Fluorescent imaging of TFs and coactivators in living cells has revealed that they are dynamic and only transiently interact with chromatin targets (Hager et al., 2009).

Single-molecule tracking (SMT) has made it possible to observe individual TF molecules in live cells (Brouwer & Lenstra, 2019). A number of studies have used SMT to explore the kinetics of mammalian TFs [reviewed in (Liu & Tjian, 2018)] and their interactions with nuclear structures and the nuclear architecture as they search for specific binding sites on the genome (Benichou, Chevalier, Meyer, & Voituriez, 2011; Izeddin et

al., 2014; Kent et al., 2020; Normanno et al., 2015; Reingruber & Holcman, 2011). More recent work has revealed that in addition to diffusion, TF kinetics are indicative of complex interactions beyond specific and nonspecific binding to DNA (Garcia et al., 2021b; A. S. Hansen et al., 2020; Hipp et al., 2019; Lerner et al., 2020; Reisser et al., 2020; Stavreva et al., 2019). When traveling through the crowded nucleus, TFs are likely to interact with other proteins and coregulators, chromatin, diverse RNA species, and may also be sequestered in various nuclear compartments. Indeed, previous experiments in mammalian cells have pointed to a plethora of possible interactions in the nucleus (Grünwald, Spottke, Buschmann, & Kubitscheck, 2006; Normanno et al., 2015) but their role in transcription remains unclear.

Nuclear receptors (NRs) are ligand-regulated TFs that recognize and bind their cognate regulatory sites throughout the genome upon activation. Interactions of all NRs are mediated by well-structured DNA-binding domains (DBD) and one or more activation function domains (ADs), which bind to coactivators and corepressors via protein-protein interactions to regulate gene expression (Dasgupta, Lonard, & O'Malley, 2014). The ADs of the NRs frequently contain intrinsically disordered regions (IDRs) of low-complexity amino acid sequences that assume multiple different conformations (Kumar & Litwack, 2009). The glucocorticoid receptor (GR) is a typical member of the NR family harboring an IDR (Voss & Hager, 2014). While the structures of the GR's DBD and ligand binding domain (LBD) have been elucidated by X-ray crystallography (Bledsoe et al., 2002; Luisi et al., 1991), the structure of its N-terminal domain activation function 1 (AF1) is not well understood, despite being a major region for control of GR's transcriptional activity (Khan et al., 2012; Simons & Kumar, 2013). The fact that ADs of NRs and other TFs are so poorly characterized, limits our understanding of their interactions with the Mediator complex and coactivators (Allen & Taatjes, 2015; Reiter, Wienerroither, & Stark, 2017).

It was recently demonstrated that the ADs of diverse TFs, including the estrogen receptor, can form heterotypic condensates with the IDR of the MED1 subunit of the Mediator complex *in vitro,* and this process requires the TFs' IDRs (Boija et al., 2018). These IDR-IDR interactions can result in the formation of phase-separated condensates (Alberti, 2017; Banani, Lee, Hyman, & Rosen, 2017; Hyman, Weber, & Jülicher, 2014; Shin et al., 2017). Such condensates are membrane-less micron-scale compartments organized through liquid-liquid phase separation driven by multivalent macromolecular interactions. These compartments are prevalent in eukaryotic cells and are implicated in many biological processes (summarized by (Banani et al., 2017)).

The recruitment of the transcription machinery at genomic sites is also driven by liquid-liquid phase separation (LLPS) (Boehning et al., 2018; Chong et al., 2018; Hnisz et al., 2017; H. Lu et al., 2018; Sabari et al., 2018). Many TFs (e.g., FET family TFs, OCT4, SP1, including GR), co-activators (e.g., Mediator and BRD4), and RNAPII contain IDRs, which can drive their phase separation, leading to formation of discrete nuclear foci in mammalian cells sensitive to short-chain aliphatic alcohols, which can dissolve these membrane-less structures (Boehning et al., 2018; Cho et al., 2018; Chong et al., 2018; H. Lu et al., 2018; Sabari et al., 2018; Stortz et al., 2020). Despite sustained interest in the role of IDRs in the formation of macromolecular condensates, their influence on TF diffusion and the consequent impact on transcription are unclear. Moreover, whether the kinetics of an IDR-containing and IDR-less TF are qualitatively different remains to be explored.

Advances in imaging and statistical analysis of TF dynamics are required to distinguish between the diverse diffusive properties of single molecules and link their kinetics to the underlying function(s). Here, we used a machine learning based method called perturbation expectation-maximization [pEM, (Koo, Weitzman, Sabanaygam, van

Golen, & Mochrie, 2015)] to classify individual trajectories of TFs tracked using SMT. Our analysis uncovered two distinct states with limited mobility. One of these low-mobility states accounts for the expected specific GR binding to chromatin, while the other represents a novel confined state, mediated by the IDR regions of the receptor. The dwell times of both chromatin-bound and confined states follow a power-law distribution. In the case of the chromatin-bound population, the power-law emerges as a result of the heterogeneity of binding to response elements with different motif strength. On the other hand, the power-law in the confined state likely emerges as a consequence of the broad distribution of effective binding affinities due to IDR-mediated protein-protein interactions. We propose that the confined state can amplify transcriptional output by increasing the local concentration of TFs at specific sites, thus providing a functional link between confinement and gene regulation.

4.3. Results

4.3.1. The glucocorticoid receptor exhibits four distinct populations within the nucleus

Imaging studies have revealed a high degree of heterogeneity in nuclear architecture (Finn & Misteli, 2019; Lerner et al., 2020). How the diversity of interactions between TFs and the nuclear environment influences the dynamics of TFs within the nucleus and its subsequent effect on transcription is unclear.

The glucocorticoid receptor (GR) is a ligand-regulated transcription factor. In the absence of ligand, the GR remains inactive and only after ligand binding does the receptor bind chromatin and regulate gene expression (D. M. Presman et al., 2017), providing a platform to test the behavior of a physiologically relevant TF. We performed SMT

experiments under Highly Inclined and Laminated Optical sheet (HiLO) illumination (Tokunaga et al., 2008) on HaloTag-fused wild type GR (GRwt-Halo) (Paakinaho et al., 2017) conjugated with the photo-activatable PA-JF_{549} fluorophore (Grimm et al., 2016) and activated with dexamethasone (Dex), a synthetic glucocorticoid agonist. We imaged continuously using 12 ms exposure times for an optimal balance between fast acquisition, a good signal-to-noise ratio, and minimization of localization noise. The trajectories of localized particles from a representative cell are shown in **Figure** 4-2**A**.

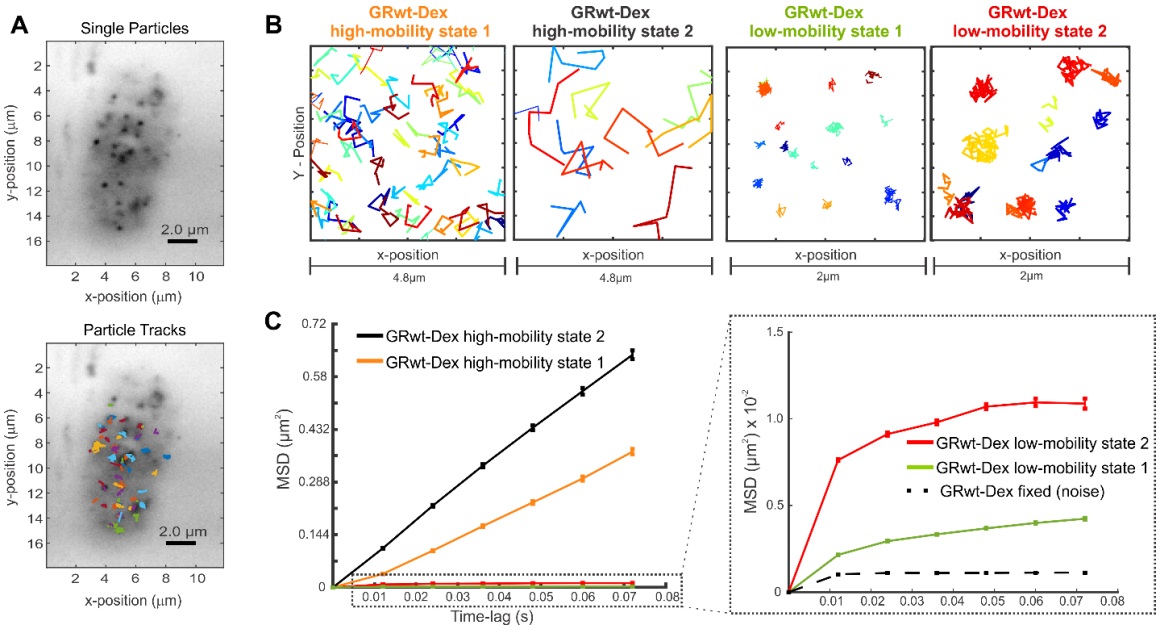


Figure 4-2 pEM based MSD analysis reveals four types of GR movement within the nucleus.

(A) Representative temporal projection image of an SMT experiment via HiLO imaging (top) with superimposed particle trajectories sampled over 84 ms with continuous acquisition (12 ms exposure, GRwt-Halo) (bottom). (B) Representative examples of particle trajectories of the observed populations classified by pEM. (C) MSD versus lag time for the four families of trajectories exhibited by GRwt-Halo conjugated with PA-JF549 and treated with 100 nM Dex (15-120 minutes prior to imaging). Right panel shows a zoomed in section of the same plot. The noise floor was calculated by imaging GRwt-Halo in fixed cells (GR-fixed, black dotted line). MSDs are calculated from 7-frame tracks. Number of cells/ tracks are 109/33,377. Error bars denote standard error measure (SEM).

For any particle trajectory, the Mean Squared Displacement (MSD) is a measure of the movement of the particle (in our case a single TF molecule) over time and can be used to elucidate the type of motion that the particle undergoes (Levi & Gratton, 2007).

When the MSD of an ensemble of particles is proportional to time [$\mathrm{MSD}(\tau) \equiv \langle|\mathrm{r}(\mathrm{t}+\tau) - \mathrm{r}(\mathrm{t})|^2\rangle = 2\mathrm{n}\mathrm{D}\tau + \mathrm{c}$; where n is the dimensionality, D the diffusion coefficient and c a constant, Einstein Relation], the particles are said to follow Brownian motion (i.e. simple diffusion). When the MSD is non-linear ($\mathrm{MSD}(\tau) \propto \tau^b$) and scales faster ($b > 1$) or slower ($b < 1$) with time, the particles are said to undergo super-diffusion or sub-diffusion, respectively. Super-diffusion may indicate directed motion while sub-diffusion implies restrictions to movement such as binding or confinement (Ben-Avraham & Havlin, 2000; Metzler, Jeon, Cherstvy, & Barkai, 2014).

To classify the single-molecule trajectories based on their diffusive properties, we applied a systems-level algorithm, perturbation-Expectation Maximization [pEM (Koo & Mochrie, 2016)], which uses unsupervised machine learning together with Bayesian Inference Criterion (BIC) (Schwarz, 1978). We first segmented the tracks into 7 frame intervals to remove length effect artifacts and decrease the probability of transitions between diffusive states (See **Figure** 4-3**A-D** and methods). This analysis revealed that activated GRwt exhibits four different types of diffusive motion (**Figure** 4-2**B-C**). Particle tracks that explore the most space are consistent with diffusive behavior **(Figure** 4-2**B,** GR high mobility states). Indeed, both states resemble 2D Brownian motion based on the linear behavior of their MSD (**Figure** 4-2**C)**, with diffusion coefficients of 0.73 ± 0.02 $\mu m^2/s$ and 2.11 ± 0.05 $\mu m^2/s$ for the high mobility states 1 and 2, respectively. Although these estimates fall within the range previously reported by other methods (Mikuni, Tamura, & Kinjo, 2007; Mikuni, Yamamoto, Horio, & Kinjo, 2017; Stasevich et al., 2010), it is surprising to find two distinct diffusive populations. However, as tracking TFs only in two dimensions (i.e. in a single focal plane) poses a limitation for accurate classification of faster diffusion modes since tracks may disappear from the focal plane faster than the tracking rate, we will not investigate these populations further in this work. The two

rightmost panels in **Figure** 4-2**B** show trajectories that explore significantly restricted regions of space. MSD analyses revealed that these correspond to two distinct low mobility states that appear sub-diffusive (**Figure** 4-2**C**, red/green lines), which we will explore further. Both states are distinguishable from GR fixed data under identical acquisition conditions (dashed black line), which represent the sensitivity threshold of the technique. We obtained similar results when tracks are segmented at 15 frames (**Figure** 4-3**E**).

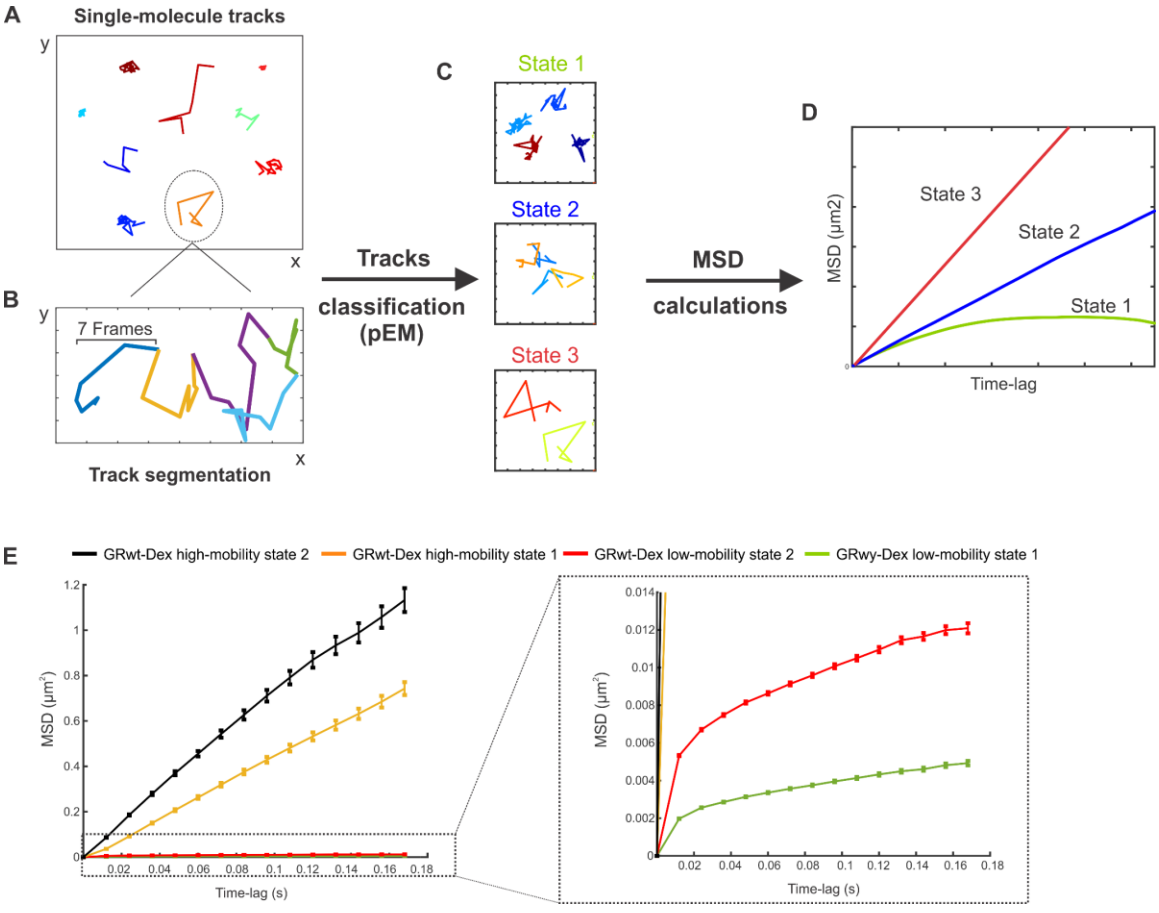


Figure 4-3 Workflow for MSD analyses of tracks.
(A) Tracks are collected in a single focal plane by single molecule tracking using HiLO illumination. (B) Tracks are segmented into 7-frame intervals. The last segment may resample frames from a previous track to include 7 frames. (C) Track segments are classified using pEM into different diffusive states. (D) The ensemble MSD for each diffusive state is calculated using MATLAB custom scripts (see methods) from the 7-frame tracks. (E) MSD versus lag time for the four families of trajectories exhibited by GR-halo conjugated with PA-JF549 and treated with 100 nM Dex (15-120 minutes prior to imaging). Right panel shows a zoomed in section of the same plot. Plot is obtained from 15-frame tracks with continuous acquisition (12 ms exposure). #cells/#tracks are 109/33,377. In all cases error bars denote SEM.

4.3.2. Specific chromatin binding accounts for the first GR low mobility state

We next analyzed the properties of the two low mobility states and correlated the MSD analyses to the activity of GR. To confirm the existence of both states over a wider temporal window, we repeated the SMT experiments on GRwt-Halo conjugated with the JF_{549} fluorophore with a longer acquisition interval (200ms) but similar exposure time (10ms). As with the continuous 12 ms acquisition conditions, pEM analysis of tracks segmented at 7 frame intervals confirmed two low mobility populations but only one higher mobility population (**Figure** 4-5**A).** The inability to discern the two high mobility states observed with the fast imaging is expected, as the sampling time of 200 ms is too slow to resolve them. Accordingly, the observed higher mobility state (blue curve in **Figure** 4-5**A**) represents only a small proportion of the tracks (**Figure** 4-5**B**) and will not be further explored. The two low mobility populations show qualitatively different degrees of restricted mobility (**Figure** 4-4**A**), consistent with the quantification by MSD plots (**Figure** 4-4**B**).

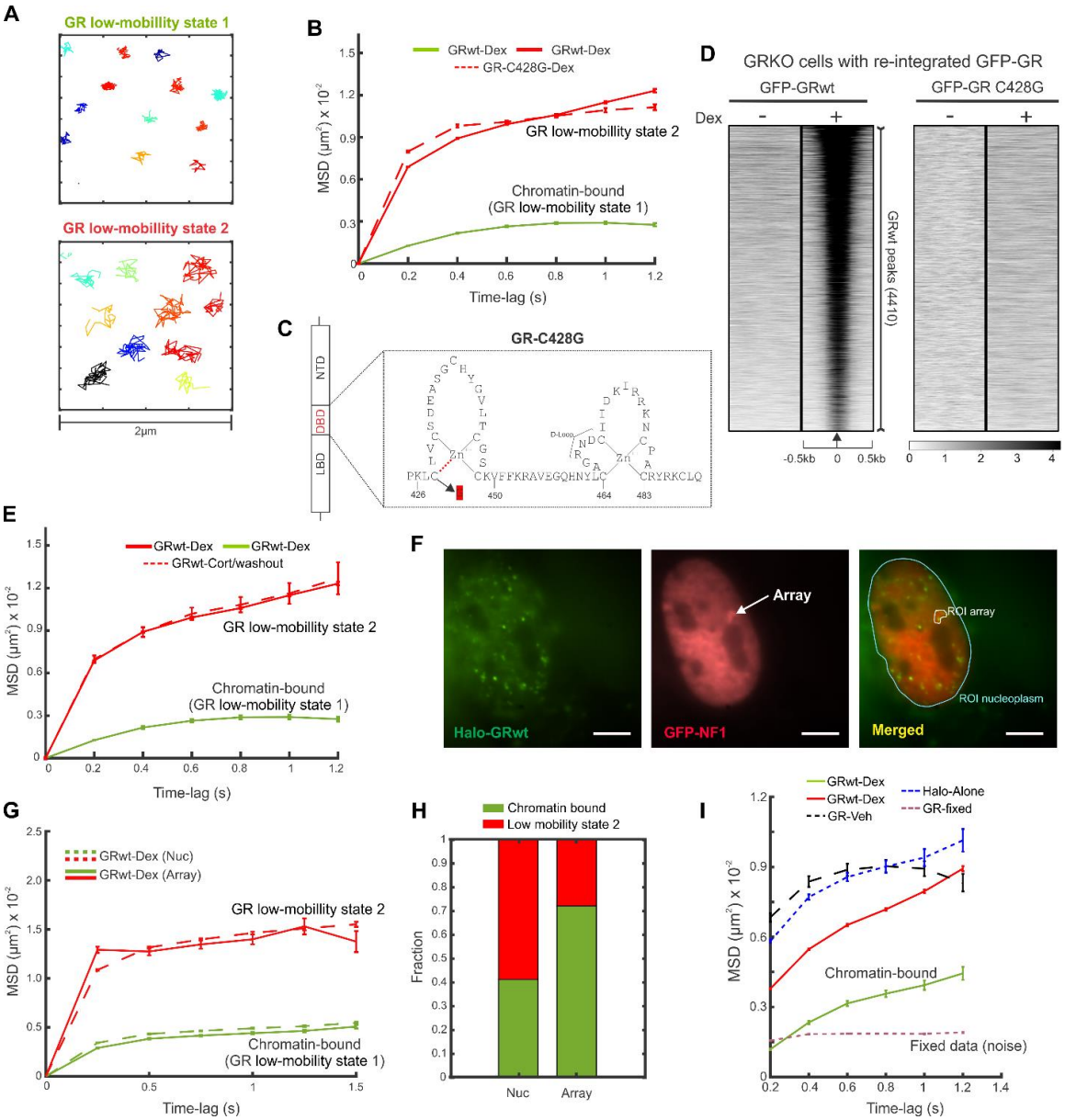


Figure 4-4 Chromatin binding accounts for one of the GR's low-mobility states.
(A) Randomly selected particle trajectories of the two low-mobility states of GRwt-Halo conjugated with JF549 and treated with 100 nM Dex (15-120 minutes prior to imaging) found by pEM analysis of 7-frame track segments, with 200 ms acquisition interval, 10 ms exposure. (B) MSD versus lag time of GRwt-Halo (solid lines, #cells/#tracks are 70/21,535) and GR-C428G (dashed lines, #cells/#tracks are 52/20,354). MSDs are calculated from 7-frame track segments, with 200 ms acquisition interval, 10 ms exposure. (C) Schematic of GR structural domains and location of the C428G mutation (arrow). (D) Heat map representation of ChIP-seq from the indicated cell lines, +/- 100 nM Dex for 1h. Binding intensity is noted below on a linear scale. Heat maps are sorted based on GRwt binding intensity and normalized for read depth and local tag density. (E) MSD versus lag time as described in B with Dex-treated GRwt-Halo (solid lines, #cells/#tracks are 70/21,535) and 4h Cort washout (dashed lines, #cells/#tracks are 60/32593). (F) Representative projection image of Halo-GRwt. GFP-NF1 serves as a marker for the tandem array. ROI, region of interest. Scale bar 5 μm. (G) MSD versus lag time as described in B for the nucleoplasm (#cells/#tracks are 82/7689) or the array (#cells/#tracks are 82/1866), with 252 ms acquisition interval, 10 ms exposure. (H) Proportions of two low-mobility states from panel G showing the relative fractions of

tracks obtained from the nucleoplasm vs the array. (I) Weighted MSD versus lag time for GRwt-Dex (solid lines, #cells/#tracks are 70/21,535), HaloTag-alone (blue dashed line, #cells/#tracks are 64/16,819), GR-vehicle (black dashed line, #cells/#tracks are 47/6236). The noise floor was calculated as in Figure 4-2C. In all cases error bars denote SEM.

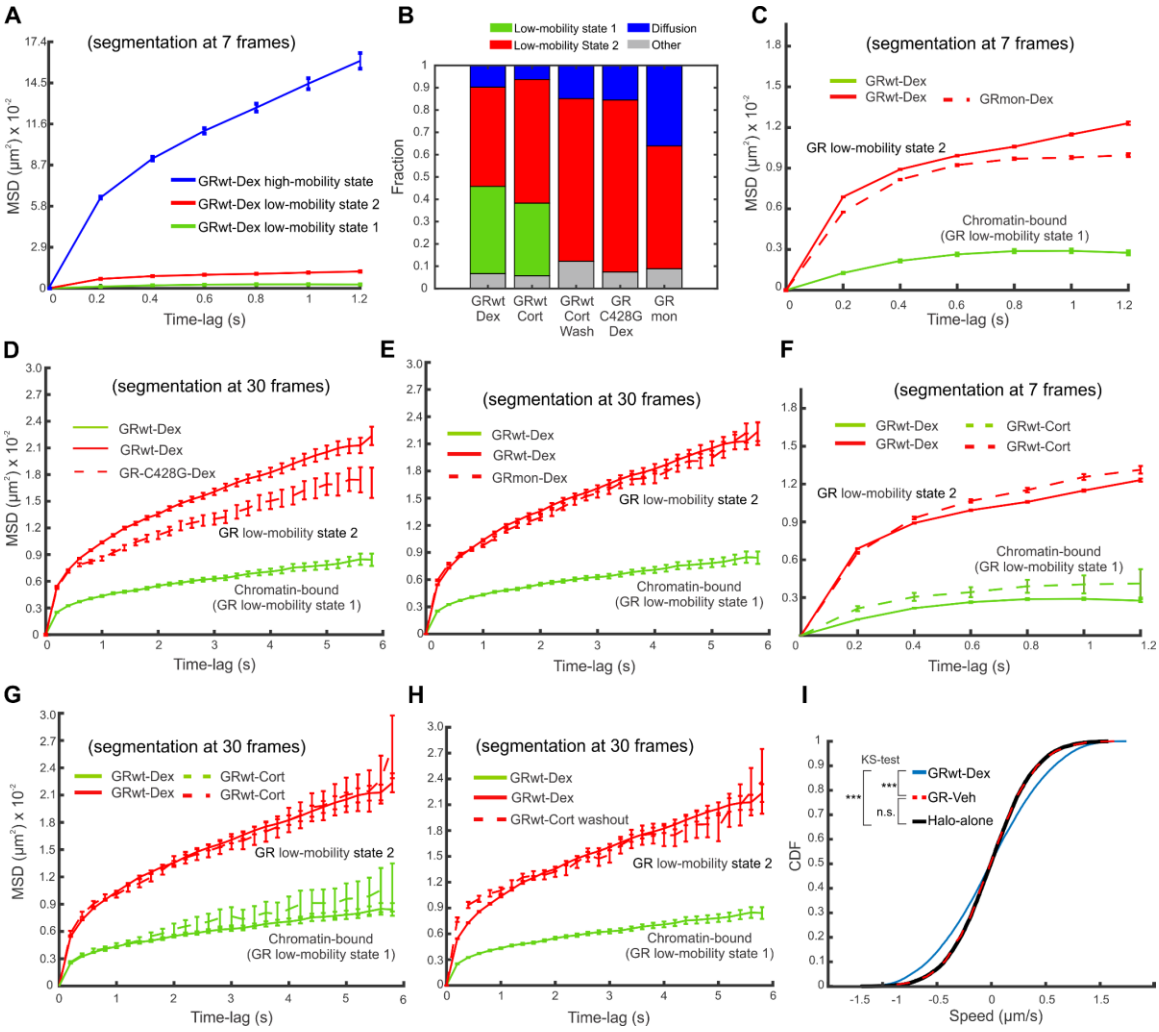


Figure 4-5 Chromatin binding accounts for one of the GR's low-mobility states.
(A) MSD versus time of GRwt-Halo (solid lines, #cells/#tracks are 70/21,535) conjugated with JF549 and treated with 100nM Dex (15-120 minutes prior to imaging). Plot is obtained from the MSD of 7-frame tracks, 200 ms acquisition interval, 10 ms exposure. (B) Proportions of diffusive and low-mobility states on the indicated GR isoforms from 200 ms acquisition interval data. Tracks that exhibit transitions between multiple diffusive states and a small fraction of anomalous tracks are classified in the "other" category (see methods for details). (C-H) MSD versus time of GRwt-Halo compared to the indicated mutants/treatments for 7-frame or 30-frame segmented tracks, as indicated; 200 ms acquisition interval, 10 ms exposure. (I) Cumulative distribution function (CDF) of the instantaneous velocity for the sub-diffusive state in HaloTag-alone or GR-vehicle compared to GRwt(Dex) low-mobility state 2 (see methods for details). A two-sample Komogorov-Smirnov (KS) test was performed: GRwt(Dex) vs GR-Veh p-value=1.8e-19; HaloTag-alone vs GRwt-Dex p-value: 1.8e-40; GR-Veh vs HaloTag-alone p-value: 0.79 (non-significant, n.s.). In all cases error bars denote SEM.

We reasoned that the apparent sub-diffusive (low mobility) populations may arise from the binding of GR to its cognate binding sites on the DNA. To test this, we first took advantage of the mouse C428G (C440G in rat) GR mutation, in which the first zinc finger is disrupted by replacing one of the four key cysteine residues. This mutation produces a drastic conformational change in the receptor's DBD (**Figure** 4-4**C**), and this mutant is unable to bind DNA *in vitro* (Hollenberg & Evans, 1988). We confirmed this result *in vivo* by performing global chromatin immunoprecipitation (ChIP-seq) in a GRKO cell line (Paakinaho, Johnson, Presman, & Hager, 2019a) stably expressing the GR-C428G mutant (see methods). While we were able to detect thousands of binding sites in GRwt (+Dex), peak calling methods found no ligand dependent peaks in GR-C428G expressing cells (**Figure** 4-4**D**).

MSD analysis of liganded GR-C428G revealed that the low mobility state 1 found in GRwt was lost, suggesting that this state corresponds to specific chromatin binding (**Figure** 4-4**B**). Consistent with this observation, the mouse GR-A465T/I634A monomeric mutant (GRmon) (Presman et al., 2014), which has almost no chromatin binding genome-wide (unpublished results), also lacks this "bound" sub-diffusive population (**Figure** 4-5**C).** Due to statistical limitations, if a small number of tracks (less than 5%) exhibit bound dynamics they will not be observed with the implemented classification. To rule out any artifacts of track segmentation, we analyzed tracks segmented at 30 frame intervals and found similar results (**Figure** 4-5**D-E**).

To further confirm the identity of GR low-mobility state 1, we analyzed the effect of inactivating GRwt by hormone withdrawal. We first activated the receptor with its natural ligand, corticosterone (Cort), and then inactivated it by culturing the cells with hormone-free media for several hours before imaging. Under these conditions, the receptor remains in the nucleus, but is mostly inactive (Stavreva et al., 2015; Stavreva et al., 2009). While

GR activated with Cort shows a similar MSD profile as the GR-Dex complex (**Figure** 4-5**F-G**), washing out the hormone from the GRwt removes the "bound" population (**Figure** 4-4**E, Figure** 4-5**H**).

The cell line used in these experiments harbors a tandem gene array that contains ~200 copies of a GR-responsive promoter structure (Stavreva et al., 2009), thus providing many specific sites for GR to bind at a discrete region of the nucleus. This array can be visualized using GFP-NF1 (Nuclear Factor 1) in live cells while GR dynamics are tracked at the single-molecule level (**Figure** 4-4**F**). We defined a region of interest around the array using the GFP-NF1 fluorescence. We then assigned Halo-GRwt tracks that either originated in or visited this region as "array tracks". Tracks were terminated when the particle left the array. We applied our MSD analysis to the array tracks and compared these with tracks of Halo-GRwt particles throughout the rest of the nucleoplasm. We found that while tracks from both the nucleoplasm and the array exhibited similar mobility states, the low-mobility state 1 (**Figure** 4-4**G-H**) was significantly enriched at the latter region. This observation supports our hypothesis that the low mobility state 1 represents chromatin binding.

To further link GR low-mobility state 1 to specific chromatin binding, we performed two additional controls. As unliganded GR constantly shuttles between nucleus and cytoplasm (Vandevyver et al., 2012), the (small) nuclear population of unliganded GR molecules serves as the vehicle control (GR-Veh). Analysis of the posterior-weighted MSD (see methods) shows that, as expected from its lack of DNA binding ability (GR-Veh in **Figure** 4-4**I**), unliganded GRwt lacks the GR "bound" state as does a control HaloTag (**Figure** 4-4**I**). In any SMT experiment, there will be a proportion of freely diffusive molecules that will appear as sub-diffusive because of the broad distribution of single molecule mobility and the anisotropy of the nucleoplasm (Banaz, Mäkelä, & Uphoff, 2019;

Mazza et al., 2012a). Accordingly, for both the HaloTag-alone and the nuclear population of untreated GR, we detected a sub-diffusive population, but with a larger effective mobility compared to the low mobility state 2 for activated GR (**Figure** 4-4**I**, *c.f.* red line vs blue and black). In fact, by calculating the instantaneous velocity distribution for this state across the three different conditions (**Figure** 4-5**I**, see methods), we find significant differences between GRwt-Dex and HaloTag alone and between GRwt-Dex and GR-Veh, but no differences between HaloTag alone and GR-Veh.

Thus far, our data indicates that GR exhibits two distinct sub-diffusive states. The most restricted state (termed the chromatin-binding state henceforth) is associated with specific binding of GR to chromatin. Accordingly, cellular conditions that preclude specific binding of GR exhibit only the sub-diffusive state with the larger mobility.

4.3.3. Intrinsic disordered regions account for the second GR low-mobility state

Recent studies have shown that protein-protein interactions mediated by IDRs form high density aggregates, likely phase separated liquid-liquid droplets, in living cells (Cho et al., 2018; Chong et al., 2018; Sabari et al., 2018; Shin et al., 2017). In the nucleus, these aggregates (also referred to as nuclear foci) might form through interactions with chromatin and have been linked to transcription (Boija et al., 2018; Y. Lu et al., 2020). GRwt, upon ligand binding, distributes nonhomogeneously throughout the nucleus, forming regions with a higher concentration of receptor molecules that are compatible with liquid condensates (Stortz et al., 2020). Using live cell confocal microscopy, we also observed fusion events between these structures, suggestive of an LLPS process (**Figure** 4-6**A-B**). We thus hypothesized that IDR interactions between GR molecules and other interacting proteins within the nucleus can create a local region of constrained motion or

"confinement" where GR kinetics will be different from the rest of nucleus and may explain our observations of the low-mobility state 2 for GR.

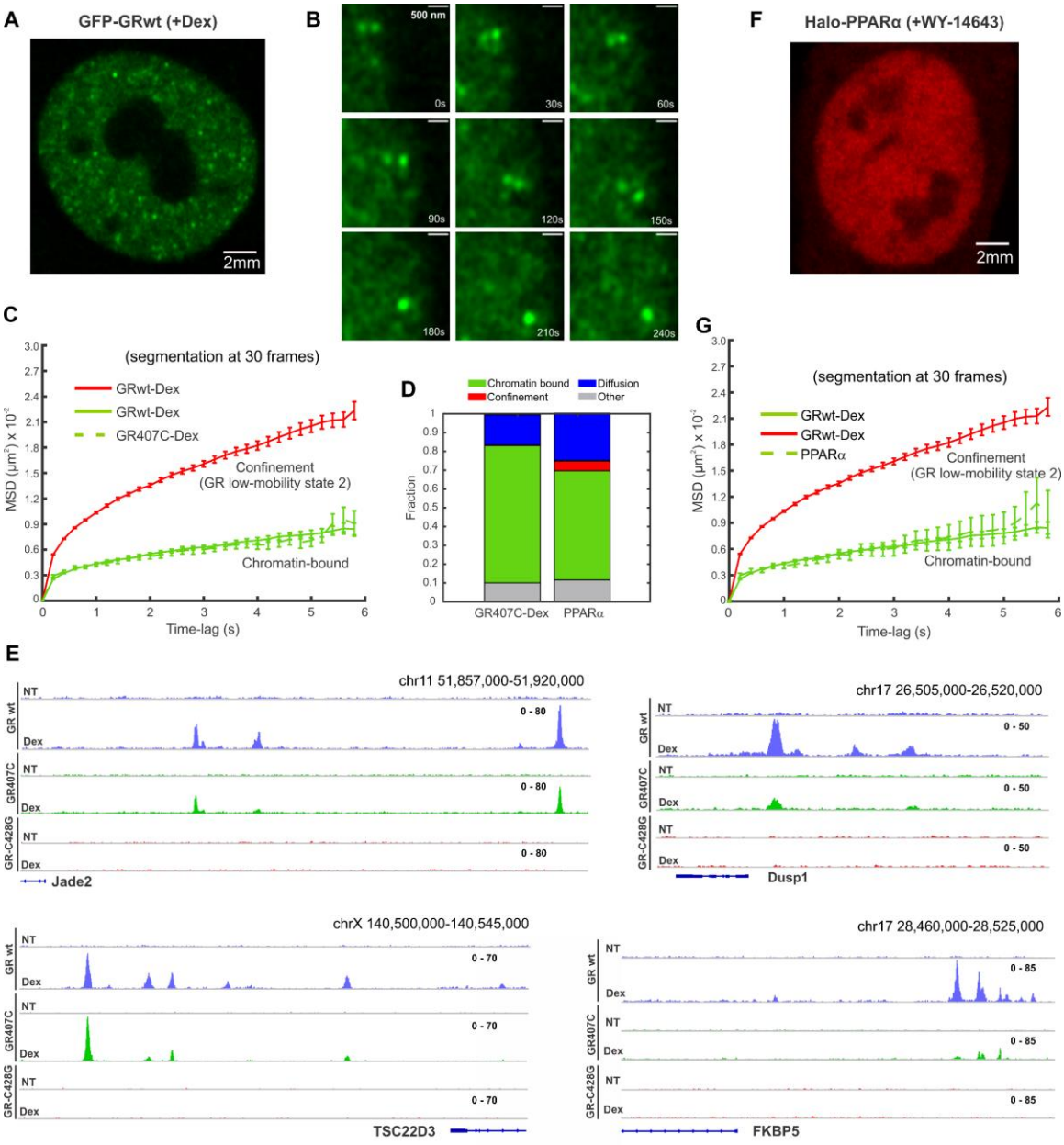


Figure 4-6 IDRs mediate confinement.

(A) Representative confocal microscopy (Airyscan) live cell image of liganded GFP-GRwt. Scale bar 2 µm. (B) Time-course of live cell imaging of two GFP-GR foci merging into one (30s intervals). Scale bar 500nm. (C) MSD versus time of GRwt-Halo (#cells/#tracks are 70/21,535) and GR407C-Halo (#cells/#tracks are 60/37,662) treated with 100nM Dex (15-120 minutes prior to imaging). Tracks segmented at 30-frames; 200 ms acquisition interval, 10 ms exposure time. Error bars denote SEM. (D) Proportions of diffusive and low-mobility states. Tracks that exhibit transitions between multiple diffusive states and a small fraction of anomalous tracks are classified in the

“other” category. (E) Example genome browser tracks of ChIP-seq data for GRwt (blue), GR407C (green) and GR-C428G (red). The closest RefSeq gene is noted below each track and the relative peak height scale of tracks is noted (right side of +Dex tracks). All genome browser tracks are normalized by reads per genomic content. (F) Representative confocal microscopy (Airyscan) live cell image of liganded HaloTag-PPARα (JF549). (G) MSD versus time of PPARα (#cells/#tracks are 60/12,237) treated with 10 μM WY-14643 (15-120 minutes prior to imaging). Tracks segmented at 30-frames; 200 ms acquisition interval, 10 ms exposure time. Error bars denote SEM.

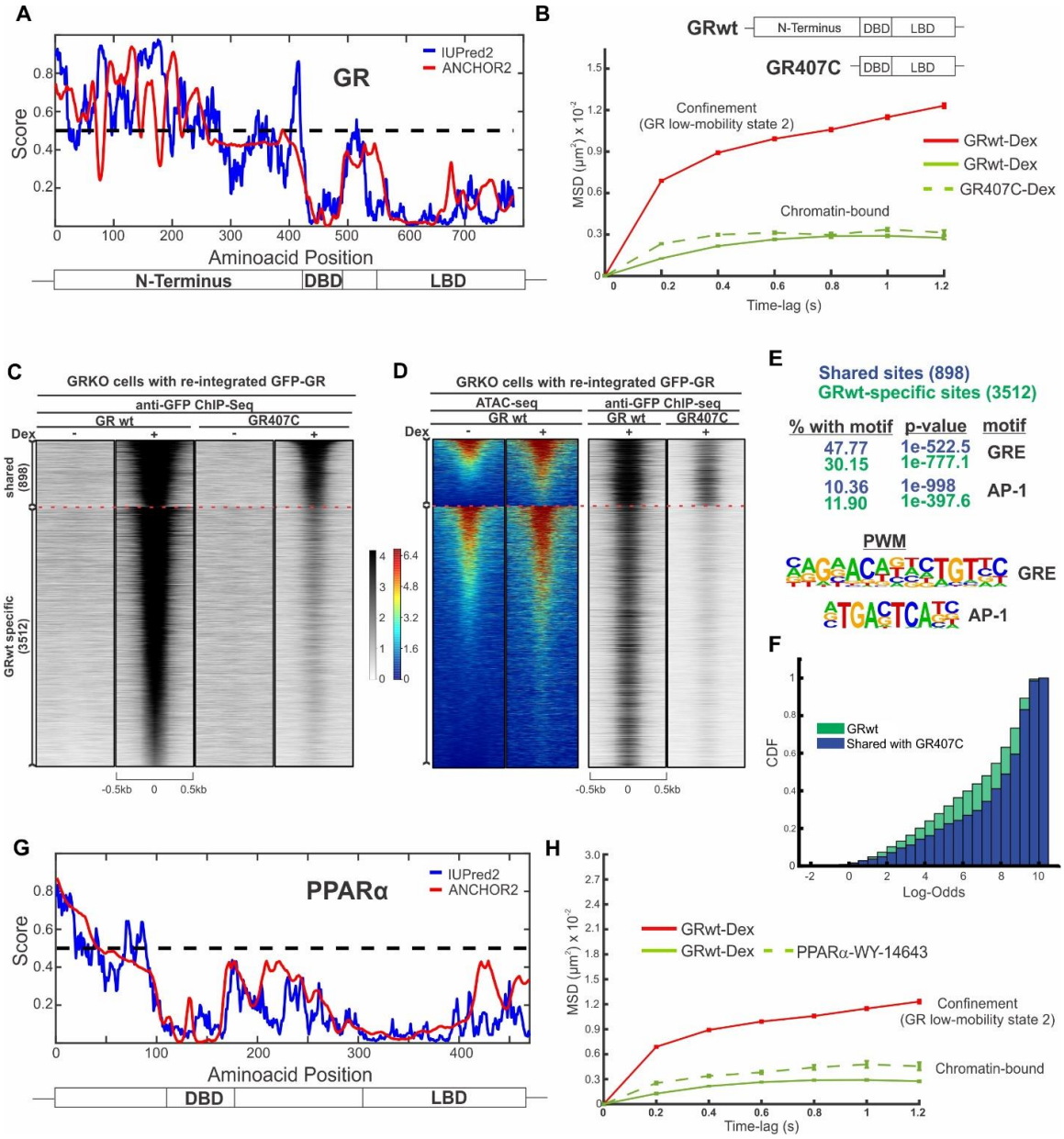


Figure 4-7 Interactions mediated by IDRs lead to confined diffusion of single TF molecules.

(A,G) Plot of inherent protein disorder probability due to a lack of intrachain interactions as predicted by IUPred2A (blue) and ANCHOR (red) models for GR (A) and PPARα (G). y-axis denotes probability (0-1) and x-axis denotes amino acid position. Regions that have a score exceeding 0.5 (dashed line) are classified as disordered regions. (B,H) MSD versus lag times of GRwt-Halo (solid lines, #cells/#tracks are 70/21,535), GR-407C (B, #cells/#tracks are 60/37,662) or PPARα (H, #cells/#tracks are 60/12,237, respectively) treated with 100 nM Dex or 10 μM WY-14643 (15-120 minutes prior to imaging). Plot shows MSD of 7-frame tracks; 200-ms acquisition interval, 10 ms exposure. Error bars denote SEM. (C) ChIP-seq heat maps (top) and aggregate plots (bottom) of GFP-tagged GRwt and GR407C stably expressed in GRKO cells, +/- 100 nM Dex for 1h. Heat maps are sorted by GRwt binding intensity and clustered by GRwt-specific peaks and GRwt/GR407C-shared peaks, noted on the left. Heat map binding intensity is noted to the right on a linear scale. (D) ATAC-seq heat map (left) and the same +Dex ChIP-Seq data as shown in C (right) re-sorted within each cluster by No-Dex ATAC signal intensity. ATAC signal intensity is noted

at left of heat map on a linear scale. (E) Motif analyses of each GR binding cluster at shared sites (blue) as compared to GRwt-specific sites (green). Position weight matrix (PWM) of the motifs are shown below. (F) Distribution of Log-odds of a GRE motif at shared sites (blue) and GRwt-specific sites (green). CDF, cumulative distribution function. X-axis represent bins of Log-odds. Comparisons using the two sample KS test ($p<0.01$).

To test this hypothesis, we removed the entire N-terminal domain (NTD) of GR, which is enriched in IDRs (**Figure** 4-7**A**). This deletion mutant is referred to hereafter as GR407C, as it only has the last 407 C-terminal amino acids (Meijsing, Elbi, Luecke, Hager, & Yamamoto, 2007). MSD analyses on the SMT data collected from this mutant revealed a complete loss of the second low-mobility state (henceforth termed "confinement") while retaining low-mobility state 1, associated with chromatin binding (**Figure** 4-7**B, Figure** 4-6**C-D**). This observation suggests that "confinement" may be a result of the protein-protein interactions associated with IDRs.

ChIP-seq analyses of the GR407C mutant in the presence of Dex shows a significant reduction in binding compared to GRwt (**Figure** 4-7**C,** 898 vs. 4410 peaks, respectively). Representative genome browser track examples are shown in **Figure** 4-6**E**. This loss of chromatin occupancy together with the loss of the confined sub-diffusive state suggests that IDR-mediated confinement may facilitate GR binding and contribute to its activity. Interestingly, the GR407C mutant is able to bind both closed and pre-accessible chromatin based on ATAC-seq data from GRwt (**Figure** 4-7**D**), suggesting that the mutant receptor can still recruit remodeling factors (Fan, Trotter, Archer, & Kingston, 2005; John et al., 2008), possibly through its LBD. Moreover, *de novo* motif analysis shows that the IDR-less GR mutant binds preferentially to more restrictive Glucocorticoid Response Elements (GREs) motif sequences compared to the wild type receptor (**Figure** 4-7**E**). In fact, the overall distribution of log-odds motif scores for GR407C is narrower compared to GRwt (**Figure** 4-7**F**), suggesting that GR407C binds a more homogenous population of GREs in relation to its consensus sequence.

To further test the proposed relationship between IDRs and the “confinement” population, we analyzed peroxisome proliferator-activated receptor alpha (PPARα), another member of the nuclear receptor superfamily (Zhao, Zhou, & Gustafsson, 2019). This protein naturally has a short IDR region (**Figure** 4-7**G**) and does not form foci (**Figure** 4-6**F**). Consistent with our IDR-mediated confinement hypothesis, diffusive state analysis on the SMT data collected with HaloTag-PPARα stimulated with the agonist WY-14643, only shows the chromatin-bound population and an absence of the confinement population (**Figure** 4-7**H, Figure** 4-6**G)**.

Taken together, our data suggest that IDRs, which have been implicated in the formation of condensates that further compartmentalize the nucleoplasm (Shin & Brangwynne, 2017), also lead to sub-diffusive behavior of GR in the nucleus (confinement), which is not directly related to specific chromatin binding.

4.3.4. Histone H2B also exhibits both confinement and chromatin-bound populations

It has been proposed that chromatin itself exhibits LLPS (Gibson et al., 2019; Sanulli et al., 2019). In fact, core histones are disordered proteins and histone tails that are exposed in nucleosomes are classified as IDRs (Peng, Mizianty, Xue, Kurgan, & Uversky, 2012), suggesting that histones may also exhibit similar diffusive dynamics. Hence, we examined the dynamics of H2B using SMT. The MSD analysis of H2B (**Figure** 4-8**A**) shows two low mobility states which are almost indistinguishable from the GR confinement and chromatin-bound states.

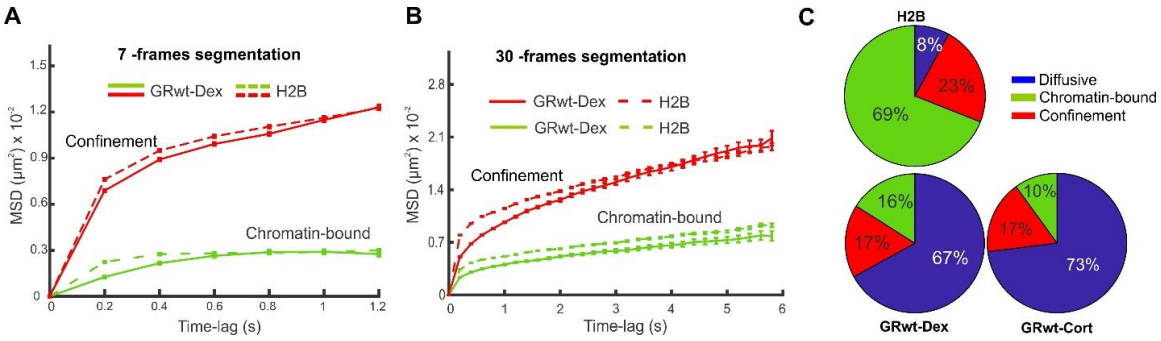


Figure 4-8 Histones exhibit both confined and chromatin-bound populations.
(A-B) MSD versus lag time of GRwt-Halo treated with 100 nM Dex (15-120 minutes prior to imaging) (solid lines, #cells/#tracks are 70/21,535) and untreated HaloTag-H2B (dashed lines, #cells/#tracks are 70/27,218, respectively). Plots obtained from 7-frame (A) or 30-frame (B) track segmentation, 200 ms acquisition interval, 10 ms exposure. Error bars denote SEM (C) Pie charts showing percentage of the different diffusive states for GRwt (Dex), GRwt (Cort) and H2B. For 12ms acquisition, #cells/#tracks are 100/20,000 for H2B; 109/33,377 for GRwt-Dex, and 101/22,182 for GRwt-Cort. For 200 ms acquisition, #cells/#tracks are 70/27,218 for H2B; 70/21,535 for GRwt-Dex, and 65/35,103 for GRwt-Cort.

The lowest sub-diffusive state of H2B is remarkably similar to the state of the GR subpopulation arising from chromatin binding (**Figure** 4-8**A**). This is to be expected, as the dynamics of H2B incorporated into nucleosomes (i.e. chromatin) should mirror the dynamics of proteins directly bound to them. Furthermore, the second low-mobility state of H2B (**Figure** 4-8**A, red line**) is compatible with the sub-population of GR that showed IDR mediated confinement. We obtained similar results with tracks segmented at 30 frame intervals (**Figure** 4-8**B**).

Using the population fraction estimates of the different diffusive states, we can calculate the relative proportions of tracks corresponding to confinement, chromatin binding, and diffusion (**Figure** 4-9 and methods). Fast acquisition SMT data (12 ms) can be used to estimate the relative proportions between the high mobility (diffusive and likely unbound) and the low mobility or “all bound” trajectories (Koo & Mochrie, 2016). We note that these estimates will be influenced by the diffusion coefficient of the particle and the fraction of faster diffusing proteins will likely be underestimated, especially due to tracking only in two dimensions (z-plane ~300nm). Furthermore, most diffusive particles will remain

in the focal plane for less than 7 frames and therefore the diffusion population is significantly underestimated. Despite these intrinsic limitations, for H2B, the immobile state dominates the dynamics as expected, in stark contrast with GRwt, which exhibits a larger diffusive population (**Figure** 4-8**C**). Longer acquisition interval SMT data (200 ms) provides the most reliable way to calculate the relative proportions between chromatin binding and confinement (**Figure** 4-9**B-C)**. We therefore used the 200 ms data to estimate the relative proportions of confined and chromatin-bound population for the "all bound" fractions identified in the fast acquisition data (**Figure** 4-9**D**). Using the proportions between bound/unbound and confinement/chromatin binding, we calculated the proportion estimates for the three different states – diffusive, confined and chromatin-bound (**Figure** 4-9**C-D**). As expected, H2B and GRwt-Dex have different fractions of chromatin-bound populations (**Figure** 4-8**C**, 69% and 16%, respectively), but surprisingly similar confined proportions (23% and 17%, respectively). GRwt-Cort shows a reduction in the chromatin-bound population compared to GRwt-Dex (**Figure** 4-8**C,** 10% vs. 16%, respectively), consistent with previous results (Stavreva et al., 2019).

Overall, our data support the idea that histones can exhibit constrained motion due to confinement in addition to their incorporation into chromatin itself.

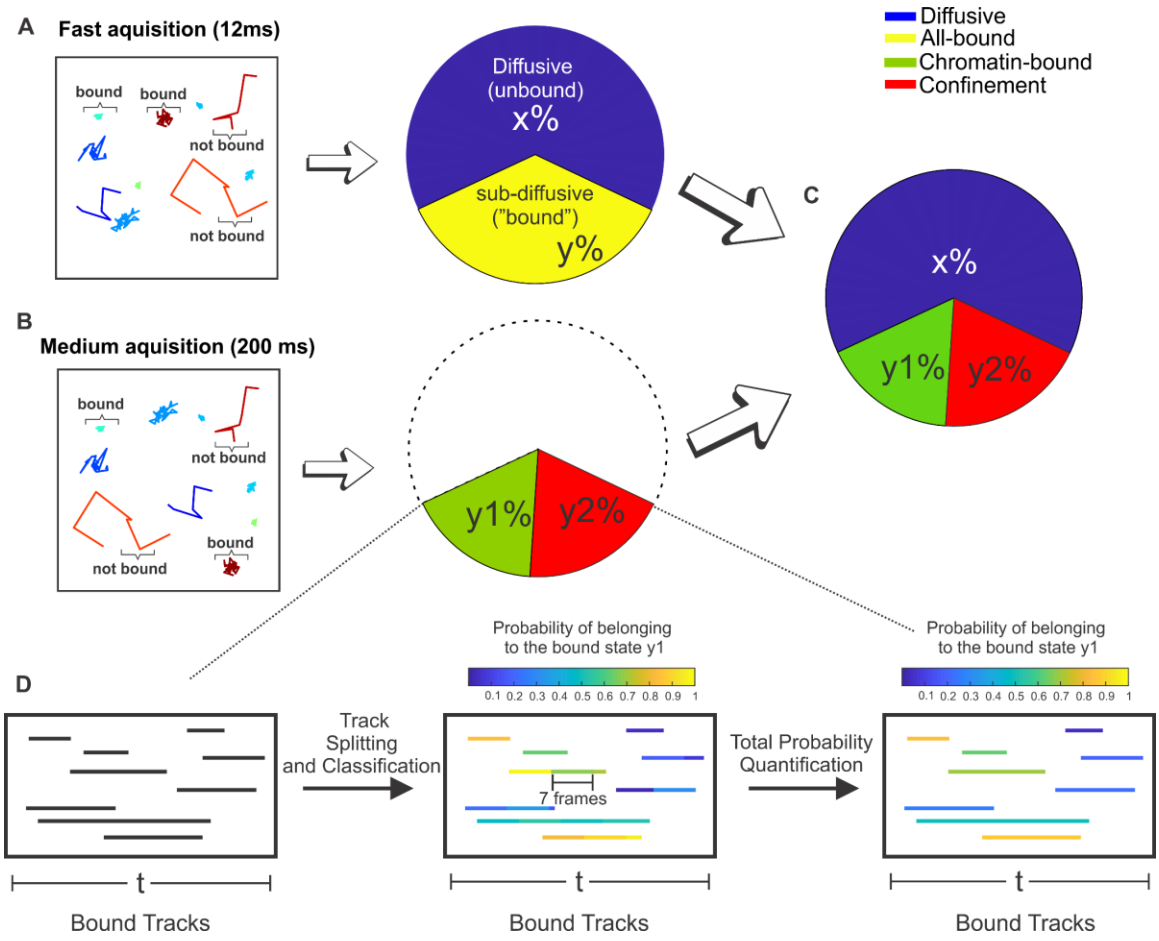


Figure 4-9 Schematic pipeline for calculation of proportions.
(A) The 12ms trajectories are classified into diffusive (high mobility) or bound (low mobility) states using pEM and the classification is used to estimate the proportion of low and high mobility trajectories (pie chart). (B) The 200 ms trajectories are classified into different diffusive states using pEM and the proportion of the two low mobility states (confinement and chromatin binding) is obtained. These proportions are then used in conjunction with the fraction of diffusive tracks identified from the 12 ms trajectories to calculate the final proportions (C). (D) Every track is segmented into 7-frame subtracks and classified into different diffusive states. After classification (middle panel), a posterior probability is assigned to each subtrack and a total posterior probability is constructed for the whole track (right panel). The posterior probability is used to estimate the proportions of the different states.

4.3.5. Dissecting the origins of the power-law distribution in SMT dynamics

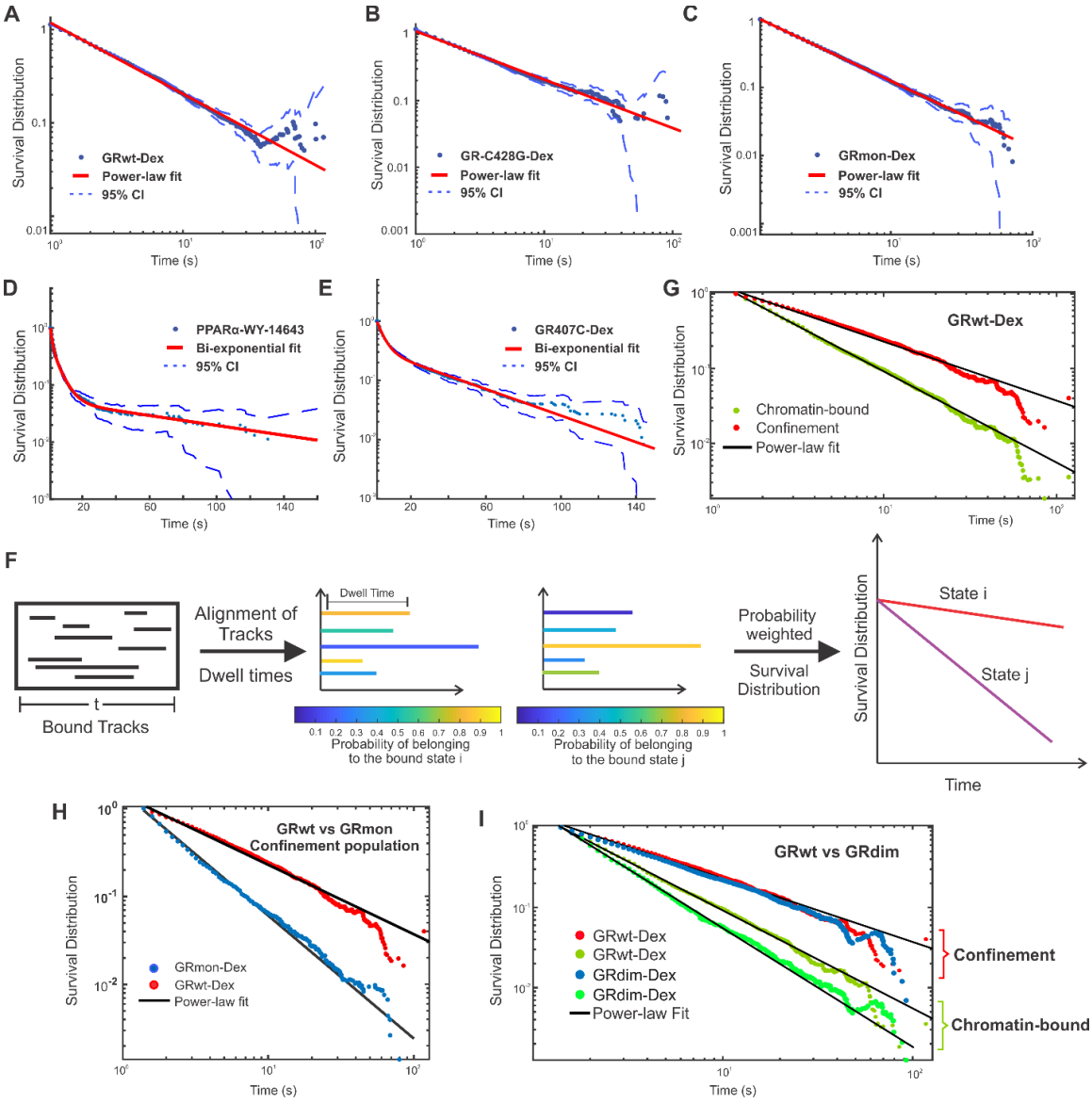


Figure 4-10 IDR-mediated interactions affect the distribution of dwell times.
(A-E) Survival distribution fit to a power-law for GRwt (Dex) (A, #cells/#tracks are 70/21,535), GR-C428G (B, #cells/#tracks are 52/20,354) and GRmon (Dex) (C, #cells/#tracks are 87/19,822). Survival distribution for PPARα (D, #cells/#tracks are 60/12,237) and GR-407C (E, #cells/#tracks are 60/37,662) fit to a bi-exponential. Fits are shown in red, 95% CI of the empirical survival distributions are indicated with dashed lines and data points are shown as solid circles. (F) Schematic pipeline for splitting chromatin-bound and confined tracks. Tracks are classified based on the posterior probability to belong to a particular state, which is then used to calculate the weighted dwell time distribution for each binding state. (G) GRwt (Dex) survival distribution for trajectories belonging to confinement (red) and chromatin binding (green) states fit to power laws (solid lines, #cells/#tracks are 70/21,535). (H) Survival distributions of the confined population for GRwt (Dex) (red) and GRmon (Dex) (blue) fit to power laws (solid lines, #cells/#tracks are 87/19,822). (I) Survival distributions of the confined (red) and chromatin-bound (green) population

of GRwt (Dex) and confined (blue) and chromatin-bound population (bright green) of GRdim (Dex, #cells/#tracks are 80/30,794). Solid lines show power law fits. In all cases error bars denote SEM.

In addition to providing information on diffusive properties, SMT experiments allow us to calculate dwell time distributions, which represent the time that a protein stays "bound" (Paakinaho et al., 2017; D. M. Presman et al., 2017). Most TFs have been described to exhibit a bi-exponential survival distribution, with the longer time constants representing specific binding [reviewed in (Goldstein & Hager, 2018b)]. However, alternative multi-exponential models (Hipp et al., 2019; Reisser et al., 2020) or power-law may better describe survival distributions (Garcia et al., 2021b; Normanno et al., 2015; Stavreva et al., 2019).

To determine how the different types of trajectories from TFs contribute to the dwell time distribution, we performed long exposure SMT experiments. We imaged HaloTag-GRwt, -GR-C428G, -GR407C, and -PPARα with 500ms exposure/acquisition time (see methods for details), using HaloTag-H2B as a reference for photobleaching correction and Bayesian inference criteria (BIC) for model fitting (Garcia et al., 2021b) (see methods). Upon either Dex (**Figure** 4-10**A)** or Cort (**Figure** 4-11**A**) stimulation, the dwell time distribution of GRwt is better explained by power-law behavior compared to bi-exponential or multi-exponential models (see **Table** 4-1 for statistics). Interestingly, the survival distributions of GR-C428G (**Figure** 4-10**B),** GRmon (**Figure** 4-10**C)** or GRwt after washing out the hormone (**Figure** 4-11**B)** also show a power-law, indicating that specific chromatin binding is not fully responsible for this type of distribution.

Strikingly, the absence of IDRs, either from PPARα (**Figure** 4-10**D)** or the IDR deletion mutant GR407C (**Figure** 4-10**E, Figure** 4-11**C**), results in a change in survival distributions to a bi-exponential behavior. The average residence times for the slowest component (**Figure** 4-11**D**) are similar to those previously reported by 3D orbital tracking

(Stavreva et al., 2019). Hence, it appears that IDRs might be responsible for the power-law distribution. To test this hypothesis, we took advantage of our MSD analysis to split tracks into chromatin-bound and confined tracks (**Figure** 4-10**F,** see methods), allowing us to independently analyze the dwell-time distribution of each low-mobility state.

If confinement were solely responsible for power-law behavior, then the chromatin-bound population should exhibit exponentially distributed dwell times consistent with chromatin binding (Garcia et al., 2021b). However, contrary to our original hypothesis, both confinement and chromatin binding remain power-law distributed (**Figure** 4-10**G**), with the confined population exhibiting overall longer dwell times than chromatin binding for GRwt. We also found that the confined population of GRmon and GRwt have different survival distributions (**Figure** 4-10**H**), with GRwt exhibiting longer dwell times, likely reflecting the different interacting partners inside the confined region.

Overall, it appears that some properties of GRwt-chromatin interactions must account for the power-law behavior in contrast to the bi-exponential distribution seen for PPARα and GR407C. One plausible explanation could rely on the differences in heterogeneity of binding affinities between GRwt and PPARα/GR407C. The GR407C mutant binds to a narrower array of GRE motif sequences as illustrated by the distribution of log-odds motif scores as compared to GRwt (**Figure** 4-7**F**), suggesting that GR407C binds to less heterogenous GREs. In the case of PPARα, it has been reported that it binds to relatively fewer sites compared to GRwt (Ratman et al., 2016), which might be indicative of less heterogeneity in binding. Taken together, our data suggest that when the heterogeneity in binding affinity is low, chromatin binding will follow a bi-exponential distribution.

Finally, we wondered if we could correlate our findings with transcriptional activity. The GRA465T mutant (GRdim) (Presman et al., 2016) is an extremely poor transcriptional activator, even though it binds to ~85% of GRwt's cistrome (Lim et al., 2015). Like GRwt, the GRdim SMT data shows both confined and chromatin-bound populations (**Figure** 4-11**E-F**). By analyzing the dwell time distributions, we discovered that while confinement dynamics are almost identical for GRwt and GRdim, there is a difference in their chromatin-bound populations (**Figure** 4-10**I**). Specifically, our observations show that GRwt has longer binding events than GRdim, which could explain why the mutant receptor is not a good transcriptional activator.

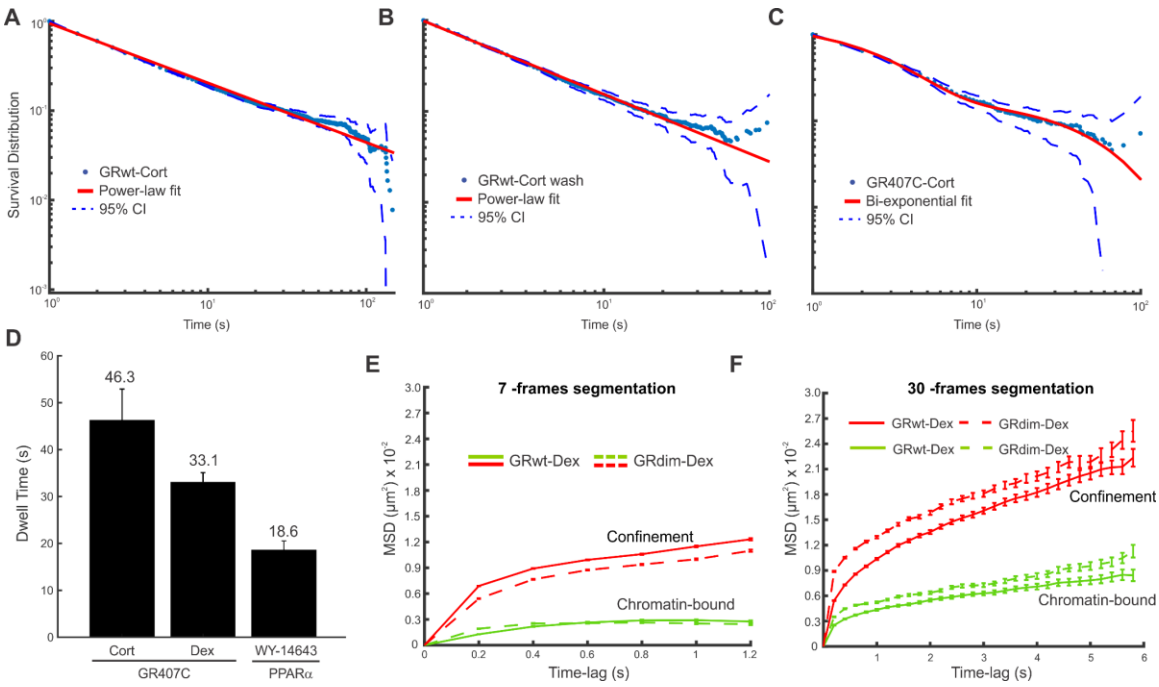


Figure 4-11 IDRs effect on survival distribution models
Survival distribution of dwell time fits to power-law for GRwt-Cort (A, #cells/#tracks are 65/35,103) and GRwt-Cort-wash (B, #cells/#tracks are 60/32,593). Power law fit in red (solid line), 95% CI in blue (dashed lines). Dwell time is exponentially distributed for GR407C (C, #cells/#tracks are 60/37,662, respectively). (D) Average residence time (slowest component) for proteins that exhibit exponentially distributed dwell times: GR407 Cort, Dex and PPARα. (E-F) GRdim (dashed lines) presents similar chromatin binding and confinement population based on MSD analysis compared to GRwt (Dex) (Solid Lines). 7-frames segmentation (E, #cells/#tracks for GRdim are 80/30,794) or 30 frames segmentation (F). Error bars denote SEM.

4.4. Discussion

Studying TF dynamics at the single-molecule level inside living cells is a sensitive approach that can unravel complex diffusion and binding kinetics of TFs as they locate and bind to their genomic targets. By analyzing single-molecule trajectories of TF using a machine learning (ML) based method and classifying their residence times based on their kinetic profiles, we show that TF "binding" is composed of at least two distinct subpopulations, one reflecting chromatin binding, and a newly identified subpopulation that arises from IDR-IDR interactions and appears transiently confined.

In support of the notion that the most restricted sub-diffusive population arises from chromatin binding of TFs we showed: 1) loss of this population in GR mutants that do not bind to chromatin (**Figure** 4-4**B,Figure** 4-5**C**), 2) its disappearance upon inactivation of the wild type receptor by hormone withdrawal (**Figure** 4-4**E**), 3) an increase of this population at the tandem array that is expected to show enhanced binding (**Figure** 4-4**H)**, and 4) the identical constrained behavior of the single H2B molecules likely incorporated into nucleosomes (**Figure** 4-8**A-B**).

The ML based analysis identified a second novel population of trajectories that exhibits constrained but higher mobility when compared to chromatin binding (**Figure** 4-2**, Figure** 4-4). Surprisingly, we observe that this apparent confinement is sustained for longer time intervals than previous observations of transient confinement (A. S. Hansen et al., 2020). We showed that this confinement behavior is mediated by interactions between intrinsically disordered regions on the TFs. In support of this, confinement is lost when the IDR is removed as in the GR407C mutant (**Figure** 4-7**B**). Additionally, TFs such as PPARα that naturally presents smaller IDR regions, but would otherwise have a full complement of motif preferences, do not show this confined state (**Figure** 4-7**H**).

Furthermore, TFs that are not capable of binding their specific recognition sequences (GR-C428G and GRmon), still display a confined state. Finally, tracks within the MMTV array have a significantly higher fraction of the lowest mobility (bound) state. Taken together, these observations suggest that IDR-IDR interactions between GR molecules and GR with other nuclear proteins can create local regions where the receptor exhibits altered diffusivity resulting in the detection of a less constrained sub-diffusive state. Our data suggest that IDR-mediated condensates [i.e. nuclear foci (Stortz et al., 2020)] are a good "microscopic" representation of this subpopulation. Whether IDR-IDR interactions are the only mechanism behind foci formation and/or confinement dynamics, or whether the foci are a functional homogenous entity, remains to be proven.

SMT also enables the measurement of TF dwell times which are indicative of their binding kinetics. We have recently reported that the dwell time distributions of GRwt and many other TFs exhibit power-law behavior rather than bi-exponential, suggesting that non-specific and specific binding cannot be simply classified according to their residence times (Garcia et al., 2021b). In fact, deviations from the bi-exponential model has been reported elsewhere (Hipp et al., 2019; Normanno et al., 2015; Reisser et al., 2020). Our ML-based techniques allowed us to classify the trajectories of bound proteins into chromatin-bound or confined sates and independently calculate their dwell time distributions. We discovered that the dwell times of both these populations exhibit power-law behavior (**Figure** 4-10**G**). Theoretical considerations suggest that such distributions can emerge due to a broad distribution of effective binding affinities (Garcia et al., 2021b), or from polymer models of chromatin with rapid rebinding of proteins (Amitai, 2018). Moreover, both models predict trapping of transcription factors in regions of the nucleus, consistent with the confinement population found in this study.

In the case of confinement, the broad distribution of effective binding affinities can originate from the heterogeneity in protein-protein IDR interactions. Equally plausible, TFs with IDRs could adopt a broad distribution of conformations due to the diversity of protein folding (Simons, Edwards, & Kumar, 2014), hence accounting for a broad distribution of binding affinities (Brodsky et al., 2020). Alternatively, but not mutually exclusive, the heavily tailed size distribution of foci (Berry, Weber, Vaidya, Haataja, & Brangwynne, 2015; Onuki, 2007; Ratke & Voorhees, 2011; Shakya, Park, Rana, & King, 2020) would produce effective binding energies consistent with power-law distributed dwell times (Garcia et al., 2021b).

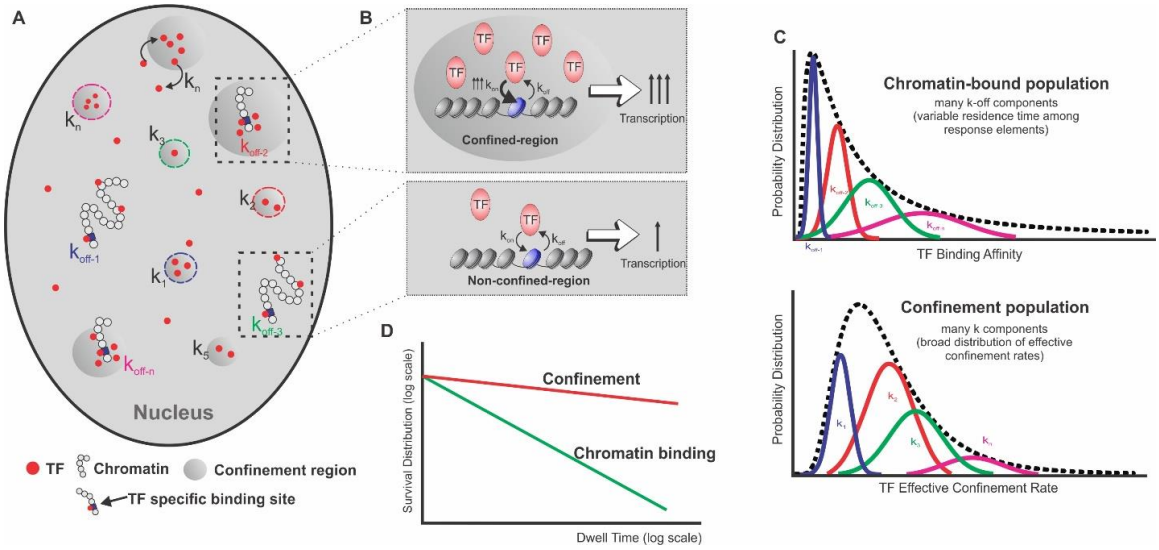


Figure 4-12 Proposed model for the modulation of gene expression by confinement and the emergence of power-law dwell-time distributions.

(A) Transcription factors (TF, red spots) navigate the nucleoplasm until they find their targets. They can be freely diffusing in the nucleoplasm (isolated red spots), confined in high density IDR-dependent hubs (shaded areas), or interacting with chromatin either specifically or non-specifically. koff, dissociation rate from chromatin; k, dissociation rate from the confined region. (B) Confined regions concentrate TFs, reducing the search time (i.e. greater kon, thicker arrow). Hence, transcriptional activity is potentiated compared to a gene whose enhancer element is not located in a confined region. (C) Broadly distributed binding affinities of a TF (dashed line) are composed of binding distributions arising from different chromatin environments and/or motifs (solid lines, top graph). Similarly, a confined transcription factor can exhibit a broad distribution of effective binding affinities related to the time that it takes to escape the confinement region, which depends on the size and physical properties of the hub (solid lines, bottom graph). (D) For a heavy tailed distribution of binding affinities and confinement, the dwell time distribution is expected to follow a power-law. In the case of GR, the confinement dwell times are longer than for chromatin binding

(as depicted). However, other TFs might present the opposite behavior if they have larger binding affinities.

In the case of chromatin binding, the power-law behavior likely emerges as a consequence of the broad affinity distribution of the k_{off} among the TF's cistrome due to the heterogeneity of binding to response elements (**Figure** 4-12). This could explain why heterologous expression of the tetracycline receptor (TetR) in mammalian cells, where it does not bind specifically anywhere, follows power-law throughout the genome but on an artificial array of single-response elements follows a single exponential (Normanno et al., 2015). In support of this model, GR407C, which does not have a population of confined trajectories and binds less variable response elements (**Figure** 4-7**F)**, behaves bi-exponentially (**Figure** 4-10**E**).

The dwell time distributions of the confined populations of GR mutants can either be similar to (GRdim, **Figure** 4-10**I**) or different (GRmon, **Figure** 4-10**H**) from the GRwt. More importantly, the survival distribution of the chromatin-bound population is significantly different between GRwt and GRdim (**Figure** 4-10**I**), which might explain the difference in their transcriptional activities (Jewell, Scoltock, Hamel, Yudt, & Cidlowski, 2012; Lim et al., 2015; Presman et al., 2014; Rogatsky et al., 2003). The approach used here thus provides a powerful means to correlate binding affinities of proteins to specific interactions within the cell nucleus.

We also found that H2B exhibits both confinement and chromatin-bound populations. Since free non-nucleosomal histones may electrostatically interact with different IDR mediated liquid-liquid phase separated aggregates (Peng et al., 2012), a free histone that has not yet been incorporated into nucleosomes can have two different types of kinetics, *normal diffusion*, or *confinement* as described above. The kinetics under confinement will be determined by the diffusive properties of the histone in these high

density/high viscosity regions, which in general will display larger MSD values than the incorporated histones due to the elastic properties of chromatin (Everaers & Schiessel, 2015; Koslover, Fuller, Straight, & Spakowitz, 2010). The kinetics of a histone incorporated into the nucleosome will be dominated by the physical properties of chromatin (e.g. elasticity and thermodynamic properties) and thus set the scale for the mobility of the chromatin-bound fraction.

Our observations suggest that the subpopulation of TFs directly bound to chromatin could serve as "nucleators" for IDR-mediated condensates, consistent with a recent report (Stortz et al., 2020). While these nucleator molecules will exhibit a typical chromatin-bound behavior, their IDR-interacting partners will exhibit confined behavior (**Figure** 4-12). Considering the growing evidence for a role of phase separated condensates in transcription (Boehning et al., 2018; Chong et al., 2018; Hnisz et al., 2017; H. Lu et al., 2018; Sabari et al., 2018), the confined population herein described might be critical for efficient transcription-initiation. Higher concentration of TF molecules found in these condensates should increase the on-rates (k_{on}) for the TF binding to chromatin (**Figure** 4-12**B**), therefore improving the chances of successful activation of RNA polymerase II resulting in a more frequent RNA bursting (Brouwer & Lenstra, 2019; Donovan et al., 2019; Stavreva et al., 2019). A prediction of this hypothesis is that if one could measure TF binding at a single site within a confined region, then only the k_{on}, and not the off-rate (k_{off}) (i.e. residence time) of the TF should change with respect to that in a non-confined region. In other words, confinement can only modulate the k_{on} of TFs, not their residence times, which may ultimately be a combination of the nature of the TF, the choice of ligand, the strength of the TF motif, and the chromatin landscape in which a response element is located (Coons, Burkholder, Hewitt, McDonnell, & Korach, 2019; T. A. Johnson et al., 2018; Syed, Greulich, Ansari, & Uhlenhaut, 2020).

Taken together, our data suggest that IDR-mediated confinement is a natural mechanism that many TFs can use to regulate gene expression more efficiently (**Figure** 4-12). This is achieved by providing a higher local concentration of TFs at specific genomic sites (i.e. increasing k_{on}) while also effectively decreasing the exploration area within the complex nuclear environment. Although not essential for all TFs, dysregulated phase separation has been implicated in a number of disease conditions (Basu et al., 2020; Darling & Uversky, 2017; Innis et al., 2004; Muragaki, Mundlos, Upton, & Olsen, 1996; Nakamura et al., 2001). Our results suggest that the modulation of TF mobility by IDR-mediated interactions and the formation of condensates likely contributes to regulation of transcriptional efficiency.

4.5. Limitations of this study

Even though we found four distinct populations of GR within the nucleus, our implementation of HiLO only allows us to recover the dynamics of the population of particles with lower mobility. This is because rapidly diffusing, i.e. free TF molecules can quickly move away from the focal plane precluding accurate tracking and recovery of MSD information. While these populations might have some useful information regarding the properties of the nuclear environment, our implementation of HiLO does not allow us to report those. 3D tracking of TFs at very high frame rates is needed to better understand the behavior of the two freely diffusing populations found in this study.

A second limitation is that we do not consider possible transitions between the diffusive states within a single trajectory. This may lead to some tracks not being classified into one of the diffusive states described here (**Figure** 4-5**B and Figure** 4-6**D**). Thus, it is possible that several biologically meaningful states with low probability of occurrence remain hidden to the classification implemented in this work. A more robust classification

algorithm using Bayesian non-parametrics might reveal these potential transitions and hidden states.

We have speculated that the confinement population described in this work is mediated by protein-protein interactions that could exhibit LLPS compartmentalization (Boehning et al., 2018; Chong et al., 2018; Hnisz et al., 2017; H. Lu et al., 2018; Sabari et al., 2018). However, a rigorous link between them has not been yet established and needs further investigation.

Finally, fluorophore stability is a limiting factor in all SMT experiments. To better quantify the dwell time distribution of TFs in different diffusive states, experiments must be performed at low exposure/high laser power for optimal localization precision. However, these conditions limit the fluorescent lifetime of fluorophores to the order of seconds, which is not enough to reliably quantify long binding times in the tail of power-law distributed dwell times.

4.6. Acknowledgments

We thank Tatiana Karpova and David Ball from the Optical Microscopy Core at the NCI, NIH for the assistance in imaging and data processing. We thank Luke Lavis (Janelia Research Campus) for providing HALO dyes. This research was supported (in part) by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. D.M.P was supported by CONICET. V.P. was supported by the Academy of Finland, the Cancer Foundation Finland, and the Sigrid Jusélius Foundation. R.A.M.J. was supported by the Vissing Foundation, the William Demant Foundation, the Knud Højgaard Foundation, the Frimodt-Heineke Foundation, the Director Ib Henriksen Foundation, and

the Ove and Edith Buhl Olesen Memorial Foundation. S.M. acknowledges support from the Danish Independent Research Council | Natural Sciences. A.U. acknowledges support from the NCI-UMD Cancer Technology Partnership, and the awards NSF PHY 1607645, NSF PHY 1806903.

4.7. Methods

4.7.1. Plasmids and cell lines

The pHaloTag-GR expresses rat GR with HaloTag (Promega, Madison, WI, USA) fused in the C-terminal domain under the CMVd1 promoter (Morisaki et al., 2014). The pHalo-GRA477T (GRdim), pHalo-GRA477T/I646A, and pHalo-GRC440G (C428G in mouse) were generated by using a QuikChange II XL Site Directed Mutagenesis Kit according to the manufacturer's instructions (Stratagene, La Jolla, CA, USA), as previously described (Paakinaho et al., 2017). The pHaloTag-H2B expresses histone H2B fused with the HaloTag through its N-terminal end (Mazza et al., 2012b). The pHalo-PPARα was purchased from Promega (Cat# pfn21ab9549) and expresses the human PPARα fused with the HaloTag through its N-terminal end. The pHalo-GR407C was generated by 1) PCR amplification of rGR's coding amino acids 407-794 from the pHaloTag-GR template using the Herculase II fusion DNA polymerase system (Agilent Technologies, Santa Clara, CA, USA), introducing XhoI and PvuI restriction sites; and 2) swapping the PCR product into the pHaloTag-GR using the same restriction enzymes. The pGFP-NF1 expresses Nuclear factor-1 fused to GFP (D. M. Presman et al., 2017). The mouse GFP-GR407C plasmid was generated by circular PCR with high-fidelity polymerase from the GFP-GRwt plasmid (Paakinaho et al., 2019a) using primers to exclude the N-terminus and sequence verified prior to use in making the cell line. The GFP-GRC428G plasmid was generated

with the same PCR/verification methods as above using primers to incorporate a point mutation at residue 428 of mouse GR.

The 3617 mouse mammary adenocarcinoma cell line and its derivatives were routinely cultured in high glucose DMEM supplemented with 10% fetal bovine serum and 2 mM L-glutamine at 37°C in a CO_2-controlled humidified incubator. The GRKO cell lines with stably integrated forms of GFP-GR has been described previously (Paakinaho et al., 2019a). Briefly, they were generated by integrating the designated form of GR into the GT(Rosa)26Sor locus. GFP-GR integrated cells were selected with puromycin and FACS sorted for similar levels of GFP expression and size uniformity. A GR knock-out subclone expressing Halo-GR was also used in this study and described elsewhere (Paakinaho et al., 2017).

For transfections, 5 million cells were electroporated using BTX T820 Electro Square Porator (Harvard Apparatus, Holliston, MA, USA) in 100 μl of DPBS with 2.5 μg of plasmid. 25 ms pulses of 120 V were used and cells were resuspended in fresh media. Single-molecule imaging experiments were set up as follows: 100,000 electroporated cells were seeded onto each well of a 2-well Lab-Tek chamber (1.5 German borosilicate coverglass, Thermo Fisher, Waltham, MA, USA) in high glucose DMEM supplemented with 10% FBS (Life Technologies), 2mM L- glutamine, 5 μg/ml tetracycline, and cultured overnight. The media was then replaced with high glucose DMEM supplemented with 10% charcoal stripped FBS (Life Technologies), 2mM L- glutamine, 5 μg/ml tetracycline, and incubated at 37°C O.N. before labeling.

4.7.2. Fluorescent labeling of Halo-tagged molecules and hormone treatments

Transfected cells were incubated with 5 nM JF_{549}-HaloTag (Grimm et al., 2015) or 50 nM photo-activatable PA-JF_{549}-HaloTag (Grimm et al., 2016) ligand for 20 min at 37°C. PA-JF_{549} was used for the fast acquisition SMT experiments (12ms) by stroboscopic activation of the fluorophore by a 473nm laser. Free ligand was depleted by washing three times with phenol red free DMEM media (supplemented with 10% charcoal stripped FBS and 5 µg/ml tetracycline) in 15 min intervals at 37°C. Next, cells were treated with 600 nM Corticosterone (Cort) (Sigma-Aldrich) or 100 nM Dexamethasone (Dex) (Sigma-Aldrich), and incubated for 20 min at 37°C before imaging. For "noise" control in **Figure** 4-2**C and Figure** 4-4**F**, after activating GR for 30 min with 100 nM dexamethasone, the cells were fixed with 4% (wt/vol) paraformaldehyde (Electron Microscopy Sciences) and 0.2% (wt/vol) glutaraldehyde (Electron Microscopy Sciences) for 35 min at room temperature. After fixation, cells were incubated with ProLong Gold Antifade Mountant for 24 hours. Cells treated with 10 µM WY-14643 (PPARα activation) were incubated for 1 hour at 37°C before imaging. For wash-out experiments, cells were washed with media three times for 4 different intervals (every 15 minutes for 1 hour or every hour for 4 hours) after 20 minutes of hormone treatment and finally imaged.

METHOD DETAILS

4.7.3. Image acquisition for single-molecule tracking

A custom HiLO microscope was used as previously described in detail elsewhere (D. M. Presman et al., 2017), with an objective heater to reduce drift.

Briefly, the custom-built microscope from the CCR, LRBGE Optical Microscopy Core facility is controlled by µManager software (Open Imaging, Inc., San Francisco, CA.), equipped with an Okolab state top incubator for CO_2 (5%) and temperature control (37°C), a 150X 1.45 numerical aperture objective (Olympus Scientific Solutions, Waltham, MA), a 561nm and 473nm lasers (iFLEX-Mustang, Excelitas Technologies Corp., Waltham, MA), and an acousto-optic tunable filter (AOTFnC- 400.650, AA Optoelectronic, Orsay, France). Images were collected on an EM-CCD camera (Evolve 512, Photometrics). The imaging conditions were as follow: 12ms exposure and 12ms acquisition time for fast acquisition SMT data with a laser power of 0.96 mW, 10ms exposure and 200ms acquisition time with a laser power of 0.96 mW for confinement and chromatin binding analysis and 500ms exposure and 500ms acquisition time with laser power of 0.16 mW for survival distribution analysis. For the array experiments (**Figure** 4-4**F-H**), the imaging conditions were 10ms exposure with sequential acquisition of GFP (for NF-1) and JF549 (for GR-halo) channels, giving a 252ms acquisition time due to limitations in the speed of the AOTF.

4.7.4. Image acquisition by Airyscan

Single plane images were acquired on a Zeiss LSM 880 point scanning confocal microscope using the Airyscan detector, a 100x Plan-Apochromat 1.4NA DIC oil immersion objective (Zeiss) and 488 nm and 561 nm laser lines. The Zeiss Zen 2.3 (black edition) software was used to control the microscope, adjust spectral detection for the emission of EGFP and JF_{549}-HaloTag fluorophores and for

processing of the Airyscan raw images. Zeiss Definite Focus was used at each time point for time-lapse imaging.

4.7.5. Single-molecule tracking analysis

We used the custom-made software TrackRecord (Mazza et al., 2013) in MATLAB (The MathWorks, Inc., Natick, MA). Briefly, to analyze each time series, data were filtered using top-hat, Wiener, and Gaussian filters. A region of interest (ROI) was defined to encompass the nucleus (and the array using the GFP-NF1 fluorescence when necessary), then particles were detected, fitted to two dimensional gaussian function for subpixel localization, and finally tracked using a nearest neighbor algorithm (D. M. Presman et al., 2017). The tracking parameters were as follows: window size for particle detection 7 pixels, maximum frame to frame displacement of 6 pixels, shortest track 2 frames, and gaps to close 1.

The average single molecule localization precision (SMLP) was estimated as $\sigma_{SMLP} = \langle \frac{\sigma_{PSF}}{\sqrt{N_{Photons}}} \rangle$, where σ_{PSF} corresponds to the standard deviation of the 2D Gaussian used for particle localization, $N_{Photons}$ the number of photons and $\langle\ \rangle$ indicates the ensemble average over all the detected particles in an experiment. The number of photons were estimated by converting the gray values (camera offset corrected) to electrons by using the conversion gain of the camera adjusted and quantum efficiency (QE) for digital gain as indicated by the manufacturer. The average localization precision for 12ms exposure – 12ms interval experiments is 23.8 nm and for 10ms exposure – 200ms interval experiments is 31.7 nm. The higher localization precision in the 12ms experiments is due to significantly higher laser power used.

Tracks were classified into distinct diffusive states using perturbation-Expectation Maximization [pEM v2, (Koo & Mochrie, 2016)]. Prior assumptions on the type of diffusion of the tracked particles are not needed with pEM and the number of diffusive states can be deduced from the analysis. pEM analysis requires all analyzed tracks to be of the same length. Tracks were split into 7 frame segments and the pEM classification analysis was performed on the set of all these track segments. For instance, a track of length 30 is segmented into 4 subtracks of length 7. Short tracks minimize the probability of transitions between diffusive states on the same tracks. To confirm the results, pEM analysis was also performed on tracks split into 15 or 30 frame segments. The minimum number of states for the system to converge to was set at 2 and the maximum at 7. If the optimal number of states that the analysis converged to was 7, the algorithm was rerun with a higher number of maximum states. The number of reinitializations was set to 30 with 50 perturbation trials. The maximum number of iterations was 10000 with a convergence criterion for the change of log-likelihood of 10^{-7}. The number of features for the covariance matrix was set to 5 for tracks of length 7 or 15 and 3 for tracks of length 30. Motion blur coefficient was calculated as $\frac{1}{6}\frac{\Delta e}{\Delta t}$, where Δe corresponds to the exposure time and Δt the acquisition interval.

4.7.6. Mean Squared Displacement calculation (MSD)

Ensemble MSD was calculated using custom scripts and routines previously published (Tarantino et al., 2014). All the calculations assume that the stochastic process of TFs diffusing in the nucleus or interacting with chromatin is

wide sense stationary. SMT experiments provide us with tracks that can be described as a series of positions $\{\vec{r}i(t)\}$, acquired every Δt sec. For any stationary stochastic process, the MSD of a trajectory *i* at time lag τ can be calculated as:

$$MSD_i(\tau) = \frac{1}{N}\sum_n [\vec{r}i(n\Delta t + \tau) - \vec{r}i(n\Delta t)]^2$$

After pEM classification, states with a population fraction less than 5% are removed. For a particular state *i*, let us define the set $I \equiv \{k|P_k(i) \geq 0.6\}$ where *k* is a track of length 7 and $P_k(i)$ corresponds to the probability of *k* to belong to the state *i*. If the population of a set I (i.e. the ratio of the number of tracks assigned to state I to the total number of tracks) is less than 0.05, the set I is removed. The ensemble average MSD for a particular state *i* is given by:

$$eMSD_i(\tau) = \frac{\sum_{j\in I} MSD_j(\tau)}{|I|}$$

where $|I|$ is the number of tracks assigned to state I. In other words, after tracks of length 7 (or 15) are classified into different diffusive states, the ensemble average MSD is calculated for each particular diffusive state from tracks of length 7 (or 15) that have a posterior probability higher than 0.6 of belonging to that particular state.

Alternately, we can also calculate the weighted ensemble averaged MSD for a particular state by using the posterior probability of a track to belong to a state I as the weight function:

$$WMSD_i(\tau) = \frac{\sum_{j=1}^{N} P_j(i).MSD_j\ (\tau)}{N}$$

Where $WMSD_i(\tau)$ is the weighted ensemble average MSD for state *i*, and N is the total number of tracks.

Standard error was calculated as $\frac{\sigma_w}{\sqrt{N_f}}$, where σ_w corresponds to the weighted standard deviation and N_f the number of degrees of freedom in the weighted mean. Weights correspond to the number of points averaged to generate the mean square displacement value at the given time lag.

Sample tracks for illustration were selected as follows: sets composed of all tracks from the different diffusive states with a posterior probability larger than 0.9 were generated. For low-mobility tracks, tracks were selected randomly in a 2 µm X 2 µm window for illustration. For high-mobility states, a 4.8 µm X 4.8 µm window was selected instead.

Diffusion coefficients were estimated from tracks with a posterior probability larger than 0.6 to belong to the particular diffusive state for the 12ms acquisition interval experiments. The estimation was done from the variance of the instantaneous velocity vector v by $\langle v^2 \rangle = \frac{4D}{\Delta t}$, D corresponds to the diffusion coefficient and Δt the acquisition interval (Qian, Sheetz, & Elson, 1991).

4.7.7. Survival distribution calculation

Survival distributions were calculated from particle tracks as in (Garcia et al., 2021b) using the Kaplan-Meier estimate. The 95% confidence interval was estimated using Greenwood's Formula. HaloTag-H2B data was acquired with identical imaging

conditions as the transcription factors for the different acquisition conditions. The survival distribution of this was fitted to a double and triple exponential model to extract the photobleaching rate, and model selection was used to determine the best predictive model (See Fitting and model selection). The survival distribution of the different transcription factors was corrected for photobleaching as follows ($S(t) = e^{\gamma t} S_E(t)$, where *S(t)* corresponds to the survival distribution after photobleaching correction, $S_E(t)$ the empirical survival distribution and γ the photobleaching rate).

4.7.8. Weighted survival distribution calculation

For the diffusivity analysis, we divide every track into subtracks of length 7 frames. These subtracks are then classified into different diffusive states using pEM and states that are not representative are removed as explained above. Suppose a track *k* is divided into *n* subtracks of length 7. After the pEM analysis, each subtrack *i* is assigned a posterior probability $P_k^i(j)$ to belong to a diffusive state *j*. Here, $i \in \{1,2,\dots,n\}, j \in \{1,2,\dots,m\}$, where *m* is the number of diffusive states to which pEM converges. From this state assignment, we would like to calculate the survival distribution of a particular diffusive state. To do so, we first calculate a posterior probability for the reconstructed track *k*:

$$P_k(j) = \frac{\sum_i P_k^i(j)}{\sum_{i,j} P_k^i(j)}$$

We will use these probabilities to calculate a weight function for each bin of the dwell time histogram. All the dwell times can be distributed into *N* bins such that each bin Ω_i contains n_i tracks. Mathematically, $\Omega_i \equiv \{t \in \mathbb{R} | (i-1)\Delta t \leq t < i\Delta t\}$, where Δt is the acquisition interval.

We can define a weight function W_i^j for each of the bins *i*, per state *j* as

$$W_i^j = \sum_{k=1}^{n_i} P_k(j)$$

Previously survival distributions were calculated by assigning a unit weight to each observed bound track. Instead, we can define the statistic

$$\hat{p}_i^j = \frac{W_i^j}{n}$$

as the unbiased estimator (i.e. $E[p_i^j] = \hat{p}_i^j$) of the weighted survival distribution.

Let $P_k(j)$ be the probability assigned to a track *k* as defined previously. The proportion ($M(i)$) of a particular diffusive state *i* is given by:

$$M(i) = \frac{\sum_k P_k(i)}{\sum_i \sum_k P_k(i)}$$

4.7.9. Fitting and model selection

All fits performed to the data were implemented with the nonlinear least square method using bisquare weights due to the noise on the tail of the survival distribution.

Graphical inspection was used to qualitatively determine if a straight line was observed for multiple decades in the case of a power-law fit in a log-log plot. Three different metrics were used to determine the difference between exponential models and power-law models. The first metric corresponds to Bayesian information criterion (BIC) using the probability distribution function (PDF) corrected for photobleaching as the likelihood function. The PDF was normalized between the minimum and maximum observation range of TFs dwell time (BIC1). BIC is a criterion for model selection that penalizes for

model complexity (number of free parameters in the model). BIC1 is given by (James et al., 2017):

$$BIC1(M) = kln(n) - 2ln\big(P(D|\hat{\theta}, M)\big)$$

where *M* corresponds to the model (Power Law, Double Exponential and Triple Exponential), *k* corresponds to the number of parameters of the model, $\hat{\theta}$ corresponds to the model parameters found by fitting, *D* the observed data and *n* the number of observations. $P(x|\hat{\theta}, M)$ corresponds to the realization probability of x given the model PDF with parameters $\hat{\theta}$. For SMT, *D* is the set of independent and identically distributed discrete experimental events and $P\big(D\big|\hat{\theta}, M\big)$ is calculated as follows:

$$P\big(D\big|\hat{\theta}, M\big) = \prod_{x \in D} \int_{x-\frac{\delta t}{2}}^{x+\frac{\delta t}{2}} p\big(x\big|\hat{\theta}, M\big)\, dx$$

Where $p\big(x\big|\hat{\theta}, M\big)$ corresponds to the PDF of the model after photobleaching correction. For instance, the double exponential PDF is given by:

$$p\big(x\big|\hat{\theta} = (\alpha, \beta), M_{DE}\big) = C\big(f_1 \gamma \alpha e^{-(\gamma+\alpha)t} + (1 - f_1)\gamma \alpha e^{-(\gamma+\beta)t}\big)$$

where α, β are the exponential parameters, γ the photobleaching rate and C a normalization constant.

The second metric, the evidence in decibels (Db) for a particular model given the observed data and priors, was calculated to compare the alternative models explored. The evidence measures the probability of a particular model being the best predicting model in comparison with all the other models. For instance, for the power law model (M_{PL}) the evidence versus the double exponential model (M_{DE}) is given by (Jaynes & Bretthorst, 2019):

$$E\left(M_{PL}\middle|D,\hat{\theta},A\right) = E(M_{PL}|A) + 10log_{10}\left[\frac{P(D|M_{PL},\hat{\theta})}{P(D|M_{DE},\hat{\theta})}\right]$$

$$E(M_{PL}|A) = \frac{P(M_{PL}|,A)}{P(M_{DE}|A)}$$

where A corresponds to the priors; P,D and $\hat{\theta}$ as defined for BIC1. Uniform priors were used throughout the model comparison and therefore $E(M_{PL}|A)$ is set to 0. For instance, an evidence of 30 Db corresponds to a probability higher than 0.999 that the power law model better describes the data in comparison with the alternative models tested. If the evidence was not high enough to reach a conclusion between the different models, more data was acquired until the evidence reached a satisfactory value.

A final metric using BIC and the residual sum of squares (RSS) as a likelihood function was used (BIC2). For each functional model *f*, BIC2 was calculated as:

$$BIC2(M) = klog(N) + Nlog\left(\frac{RSS}{N}\right)$$

where *k* corresponds to the number of parameters estimated and *N* the number of observations in the survival distribution (James et al., 2017).

Delta-BIC1 (Delta-BIC2) is defined as the difference between BIC1 (BIC2) calculated for double exponential and power-law (DeltaBIC1=BIC1(DE)-BIC1(PL)). A negative value indicates that the double exponential model is a better model for the data in comparison to the power-law model, while a positive value indicates that a power-law better describes the data. A negative value of evidence (as calculated above) corresponds to a higher probability for a double exponential model being the best model for the data compared to the power-law model. **Table** 4-1 shows the values of the evidence and Delta-BIC1 and

Delta-BIC2 for fitting comparison between the power law and double-exponential models for all the different experimental conditions.

4.7.10. GR chromatin immunoprecipitation (ChIP) and ChIP-seq.

GFP-GR wild type and mutant expressing cells were left untreated or treated with 100 nM of Dex (Sigma) for 1 h. For ChIP, after cross-linking with 1% paraformaldehyde (5 min) and cell collection the chromatin was sonicated (Bioruptor, Diagenode) to an average DNA length of 200–500 bp. For immunoprecipitation, 600 μg of chromatin was incubated with anti-GFP antibody (Abcam #ab290) coupled onto Protein A/G magnetic beads (Millipore) with rotation overnight at 4°C. After stringent washes, the antibody-bound chromatin fragments' cross-linking was reversed, and the remaining proteins digested. Immunoprecipitated DNA was extracted from the samples with phenol-chloroform-isoamyl alcohol and ethanol precipitation. ChIP-seq libraries were generated using Illumina TruSeq Chip Sample Prep Kit (Illumina # IP-202-1012) according to manufacturer's instructions.

4.7.11. ChIP-sequencing data analyses

Two biological duplicate ChIP samples each for GRC428G and GR407C cell lines were sequenced using Illumina NextSeq 500 with single-end reads. The data were aligned to the mouse reference mm10 genome using Bowtie2 (Langmead & Salzberg, 2012). Subsequent downstream analysis was performed using HOMER (Heinz et al., 2010). Replicate data sets were merged and peaks in each dataset were called using findPeaks with style factor for TF, FDR 0.001, 4-fold enrichment of normalized reads to the control and 4-fold enrichment over local background; however, individual replicates correlated well with each other for all called peaks. The mergepeaks command was used to determine the shared/unique peaks from the GRwt and GR407C. Data matrices for

heatmaps/aggregate plots for the ChIP-seq and the ATAC-seq (Paakinaho, Swinstead, Presman, Grontved, & Hager, 2019b) were generated using the annotatePeaks.pl command with a 20bp sampling window.

4.7.12. Quantification and Statistical Analysis

For statistical analysis, all measured quantities are reported as ensemble averages with standard error and number of observations. At least three biological replicates of SMT experiments were performed for each condition. Two sample K-S tests on the survival distribution were performed between replicates to confirm statistical reproducibility. At least 20 cells were imaged per SMT replicate for each condition for slow acquisition intervals (200 ms and 500 ms) and 60-100 cells were imaged for fast acquisition intervals (12 ms). The exact number of tracks and cells are specified in figure legends.

4.7.13. Data and Code Availability

The GRC428G and GR407C ChIP-seq datasets are submitted to the NCBI Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/) database, accession code: GSE154771. The previously published GFP-GRwt ChIP and ATAC datasets are archived under GEO accession number GSE108634.

Tracking was performed in MATLAB (version 2016a) with custom scripts. The code for pEM analysis is freely available at the following link: https://github.com/p-koo/pEMv2

Other relevant data and materials that are not explicitly included in this article will be made available by the lead contact upon reasonable request.

4.7.14. Tables

Protein	Evidence	Delta-BIC1	Delta-BIC2
PPARα	-22.73	-5.23	-54
GR407C-Dex	-23.39	-4376	-1.37
GR407C-Cort	-46.35	-20.12	-46.35
GRwt-Dex	2118	985	204
GRwt-Cort	3003	1677	834
GRwt-Cortwash	800	417	106
GR-C428G-Dex	127	65	53
GRmon-Dex	1596	744	174
GRwt-Dex confinement	NA	NA	107
GRwt-Dex chromatin binding	NA	NA	215

Table 4-1 Fitting parameters for model selection
Values of the evidence and Delta-BIC1 and Delta-BIC2 for fitting comparison between the power law and double-exponential models for all the different experimental conditions. NA, not applicable. Negative value indicates that the double exponential model is a better model for the data in comparison to the power-law model, while a positive value indicates the contrary.

Chapter 5. Estrogen Receptor Mobility and Action is Regulated by Substrate Stiffness

5.1. Summary

Tumors exhibit altered tissue-level and cell mechanics, including extracellular matrix (ECM) remodeling and stiffening, elevated interstitial pressure and altered mass transport. Experimental models demonstrate that enhancing ECM stiffness promotes malignancy and, conversely, inhibiting matrix stiffening reduces tumor incidence and improves treatment. Consequently, the molecular and biophysical mechanisms by which mechanics influence cell behavior to modulate malignancy are under sustained investigation. In the context of breast tumorigenesis, changes to the stiffness of the cellular microenvironment have been observed. Independently, reprogramming of the estrogen receptor has been reported during breast cancer progression.

The focus of this chapter is to study the role of the physical properties of the matrix microenvironment in the regulation of transcription factor dynamics in single cells and uncover the links between stiffness, chromatin accessibility, transcription factor mobility and transcriptional regulation in the MCF-7 breast cancer model. We observe significant changes to the cell's transcriptome due to changes in matrix stiffness, accompanied by pharmacological resistance to tamoxifen. Genomic studies of the Estrogen Receptor (ER) reveal optimal regulatory functions of the receptor under physiological stiffness, with enhanced chromatin binding times correlated with levels of induction and repression of ER dependent genes. We report a novel crosstalk between substrate stiffness and unliganded ER activity

wherein the unliganded receptor binds to chromatin and regulate gene expression in a stiffness depending manner. Here, we propose mechanical regulation of gene expression mediated by changes in transcription factor mobility independent to changes in chromatin accessibility as previously reported in the field.

5.2. Introduction

Organs throughout the human body exhibit a broad range of stiffness (**Figure** 5-1**A**). The compliance of tissue and matrix microenvironments encountered by cells within these organs modulates their function, and plays a crucial role in tissue differentiation and disease (Y.L. Han et al., 2020; J. Park et al., 2012). For instance, normal breast tissue has a stiffness of ~1 kPa while during tumorigenesis breast tumors reach stiffnesses on the order of ~12 kPa (Plodinec et al., 2012). More generally, tumors exhibit altered tissue-level and cell mechanics, including extracellular matrix (ECM) remodeling and stiffening, elevated interstitial pressure and altered mass transport (Heldin, Rubin, Pietras, & Östman, 2004; Huang, Wang, Bates, & Zhuang, 2008; Winkler, Abisoye-Ogunniyan, Metcalf, & Werb, 2020). Enhancing ECM stiffness promotes malignancy and, conversely, inhibiting matrix stiffening reduces tumor incidence and improves treatment (Paszek et al., 2005). Therefore, the molecular and biophysical mechanisms by which mechanics influence gene expression to modulate malignancy are under sustained investigation. A consensus view that has emerged is that forces generated by the cytoskeleton as the cell adjusts to external perturbations such as substrate stretching, fluid flow, or alterations in substrate rigidity or topography are transmitted to the nucleus and chromatin via

physical links on the nuclear envelope and the lamin meshwork (Shumaker, Kuczmarski, & Goldman, 2003). Aberration in nuclear mechanics and deformability have been linked to metastatic progression (Shumaker et al., 2003). Additionally, changes in chromatin structure have been closely associated with tumor growth and cancer susceptibility (Morgan & Shilatifard, 2015; Zane, Sharma, & Misteli, 2014) by promoting genomic instability, stimulating cell proliferation, and facilitating cellular transformation.

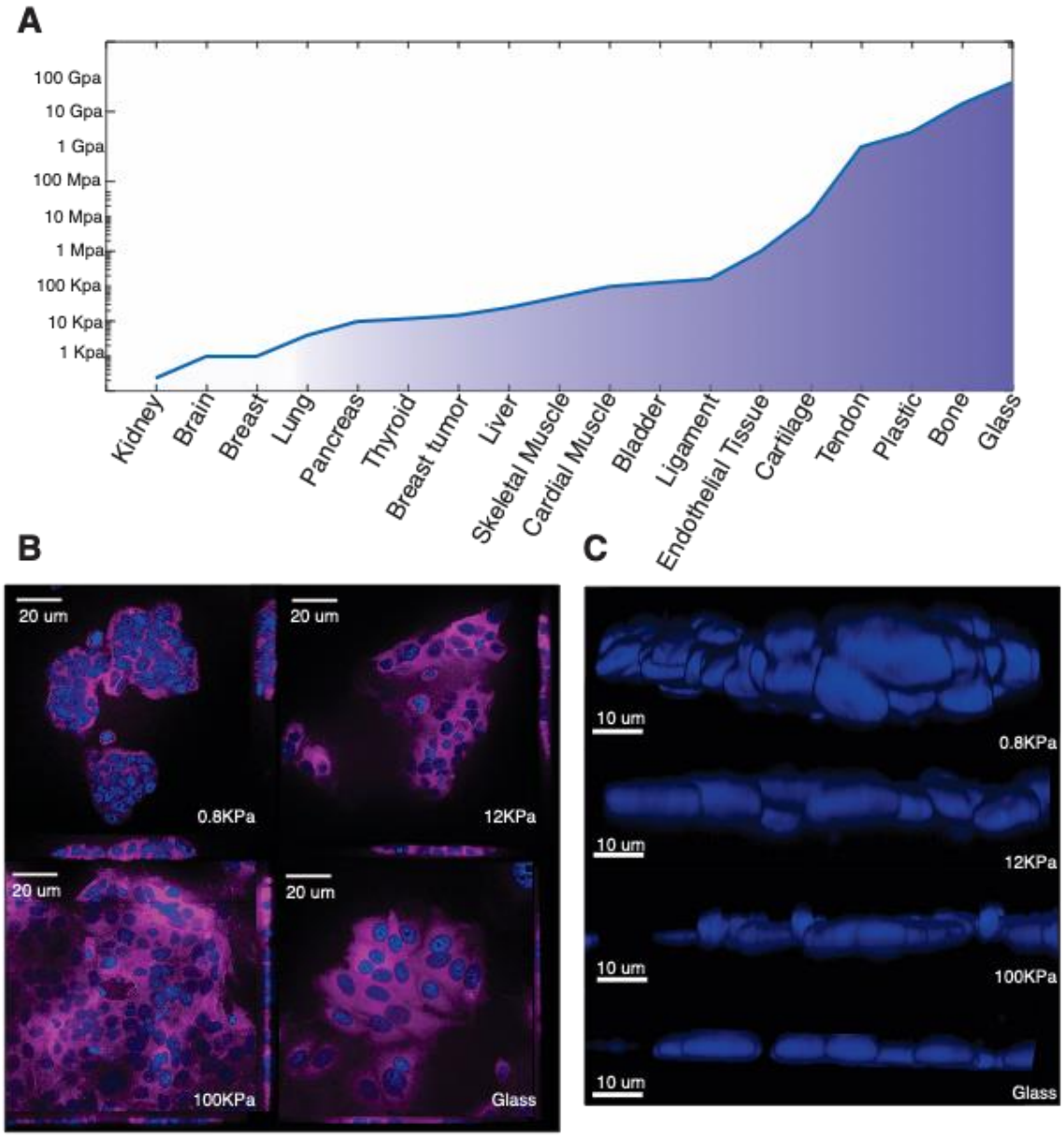


Figure 5-1 Stiffness Affects Cell Phenotype
(A) Schematic showing the broad range of stiffness of organs in the human body, ranging in the order of a few pascals in the Kidney and Brain to giga Pascals in bones and tendons. Young's modulus on the Y-axis in log scale, different organs on the X-axis. (B) Maximum intensity projection of fixed MCF-7 cells on polyacrylamide gels of different stiffness (0.8KPa, 12KPa, 100KPa and Glass) in a confocal microscope. F-actin (sir-actin) in magenta and nuclei staining in blue (DAPI). (C) Lateral view of 3D reconstruction of confocal imaging of fixed MCF-7 nuclei (DAPI) on substrates of different stiffness.

Understanding the role of mechanobiology in tumorigenesis and cell biology is pivotal in revealing its connection with metastasis, prognosis and therapeutic

treatments. Studies of mouse mammary glands revealed stiffening of breast tumors in comparison with normal tissue due to stiffening of the tumor stroma (Butcher et al., 2009; Sinkus et al., 2000). Recent nanomechanical measurements of human breast biopsies revealed unique mechanical markers that accompany different breast cancer stages with potential prognostic qualities (Plodinec & Lim, 2015). Besides changes in the mechanical properties of the tumor microenvironment due to increase in matrix deposition and crosslinking (Levental et al., 2009; Paszek et al., 2005), biophysical techniques have shown that alterations in cytoskeletal architecture are associated with malignant transformation (Guck et al., 2005).

The estrogen receptor (ER) is a key regulator of mammary development, differentiation and a dominant target in therapies for ER+ breast cancers (Howell, 2008; Jensen & Jordan, 2003). Moreover, ER has been shown to be an oncogenic driver of breast cancers but its role in tumorigenesis is still poorly understood. Understanding the role of ER at different stages of breast cancer progression is paramount for the development of specific targeting therapeutics. ER transcriptional activity has been shown to be altered between normal mammary epithelium and breast cancer cells driven by differences in the active ER cistrome (Chi et al., 2019), suggesting important changes in ER signaling during breast cancer progression. However, no clear connection between the changes in genomic landscape induced by mechanical stimuli and ER-dependent gene expression has been established.

Several recent studies have directly probed the effect of mechanical stimuli on chromatin modifications using next-generation sequencing approaches (Stowers et al., Ning wang 2020). A clear implication of these studies is that mechanical stimuli can alter the mobility and interaction dynamics of transcription factors. Our new analysis of single molecule dynamics of TFs (Chapter 3) has revealed that the residence time of TFs exhibits a power law distribution. Such power law distributions suggest that the chromatin environment encountered by transcription factors in the nucleus is complex, with binding sites having a broad distribution of affinities and local confinement of TFs due to protein-protein interactions (Chapter 4). However, the dependence of TF mobility on the physical properties of the local cellular microenvironment has not been studied and it is not known whether TF kinetics are dependent on substrate stiffness. The latter is even more important in the context of breast cancer where ER dynamics and chromatin interactions play a critical role in cell phenotype, prognosis and treatment.

Here, we examine how mechanical properties of the cell microenvironment regulate ER transcriptional activity, nuclear mobility, chromatin binding and its accessibility in MCF-7 cells. We show that ER binding kinetics are modulated by substrate stiffness with binding times significantly longer on physiological stiffness than previously reported on glass. These changes in ER kinetics induced by the mechanical properties of the cell microenvironment in turn modulate ER action and gene regulation. Using next-generation sequencing and high-throughput genomics, we demonstrate a novel modulation of unliganded ER by substrate stiffness that has antagonistic effects on liganded ER and significant gene

regulatory functions. Substrate stiffness appears to be a master regulator of transcription factor binding, mobility and gene regulation.

5.3. Results

5.3.1. Cell Phenotype and Nuclear Morphology are Determined by Mechanical Cues

To evaluate the role of substrate stiffness on cell phenotype and transcriptional regulation and dynamics in the context of breast cancer and TF dynamics, MCF-7 cells were grown on polyacrylamide gels of different stiffness (0.8 kPa, 12 kPa, 100 kPa) or glass, (see methods) for 48 hours. Confocal imaging revealed a clear dependence of cell phenotype on substrate stiffness. At physiological stiffness (0.8 kPa and 12 kPa), cells form acinar structures or spheroids similar to the *in vivo* phenotype (M.J. Bissell & Radisky, 2001). On stiffer substrates (glass and 100 kPa), cells formed monolayers with higher aggregation of cells in 100 kPa gels and decreased spread area as compared to glass (**Figure** 5-1**B**). In addition to morphological changes in cells, nuclear shape was strongly affected by substrate stiffness. We found that cells on physiological stiffness qualitatively display spherical nuclei in contrast to flat nuclei on stiffer substrates (**Figure** 5-1**C**). Several studies have shown that nuclear morphology has a profound impact on cellular function, gene expression and many pathologies (Biedzinski et al., 2020; Dahl, Ribeiro, & Lammerding, 2008; Lee, Stowers, & Chaudhuri, 2019; Skinner & Johnson, 2017). Changes in substrate stiffness are accompanied by changes in the cell transcriptome, phenotype and behavior (M. J. Bissell, Radisky, Rizki, Weaver, & Petersen, 2002). These remarkable changes in

cell phenotype and nuclear morphology we observed prompted us to characterize changes in transcriptional regulation in the context of the estrogen receptor and breast tumorigenesis.

5.3.2. MCF-7 Transcriptome is Significantly Altered at Physiological Stiffness

To characterize the differences in gene expression induced by changes in substrate stiffness, we performed RNA-seq on cells grown on surfaces of different stiffness in charcoal-stripped serum (CSS) medium to prevent confounding effects of hormones typically present in fetal bovine serum. Unsupervised clustering analysis revealed remarkable changes in the transcriptome as a function of stiffness (**Figure** 5-2**A**). Cells grown in gels show a remarkably different transcriptional profile than cells grown in glass. Noticeably, cells grown on substrates with physiological stiffness (0.8 and 12 kPa) exhibit similar transcriptional profiles (i.e. they clustered together) (**Figure** 5-2**A**). Polyacrylamide gels have been used extensively to grow cells under different physical environments without toxic effects (Caliari & Burdick, 2016). We found that cells on glass and 0.8 kPa gels exhibit significant differences at the transcriptional level with 1037 genes differentially expressed between them (**Figure** 5-2**B**, left panel) whereas 0.8 kPa and 12 kPa exhibit similar transcriptional profiles with only 80 genes differentially expressed between them (**Figure** 5-2**B**, right panel). In order to rule out changes to the transcriptome due to effects of polyacrylamide, we compared 100 kPa and glass transcriptional profiles. Differential expression analysis showed a smaller difference in expressed genes compared to 0.8KPa

(**Figure** 5-2**C**). This implies a strong dependence of transcriptional activity on substrate stiffness, with transcriptional responses naturally divided into stiff and soft (physiological) substrates (0.8 kPa and 12KPa).

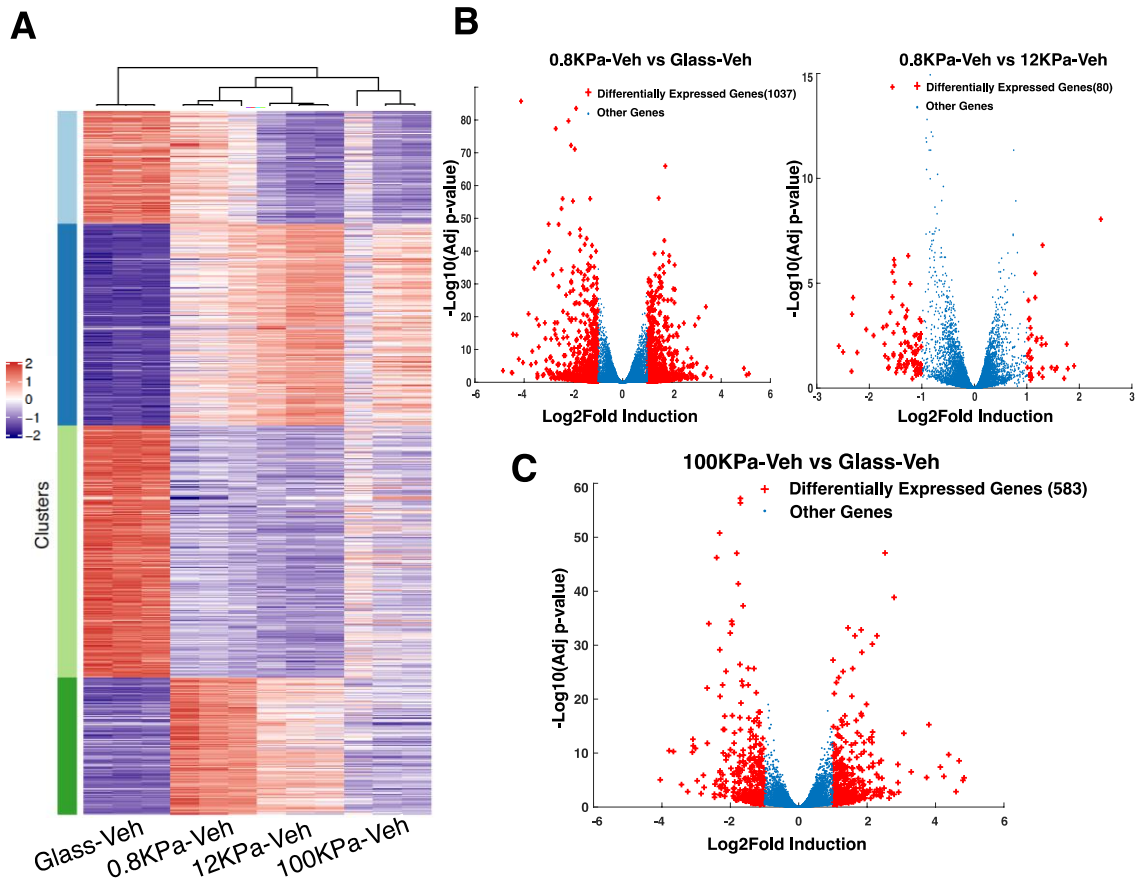


Figure 5-2 Transcriptome is determined by mechanical stimuli
(A) Hierarchical clustering of MCF-7 cells RNA-seq growth for 48 hours on substrates of different stiffness in CCS media. Transcriptome for soft substrates (0.8KPa and 12KPa) clustered together while Glass and 100KPa clustered in their own hierarchical group. **(B)** Log_2-fold change v/s adjusted P-value for differential gene expression analysis between Glass and 0.8KPa (Left Panel) and 12kPA and 0.8KPa (Right Panel). Significance level set to adjusted p-value of 0.01 and fold induction of 2. **(C)** Log_2-fold change and adjusted P-value for differential gene expression analysis between Glass and 100KPa. Significance level set to adjusted p-value of 0.01 and fold induction of 2.

Gene Set Enrichment Analysis (GSEA) of this dataset showed statistically significant enrichment of downregulated and upregulated genes detected in MCF-7 derived cells that are resistant to tamoxifen or inhibitors of aromatase (Masri et

al., 2008). Tamoxifen has been used in the management of ER+ breast cancer in the last 30 years with high degree of efficacy (Jordan, 2003; Shagufta & Ahmad, 2018). It is cytotoxic to cancer cells (Lippman & Bolan, 1975) and acts as an ER antagonist in breast tissue by inhibiting transcriptional activation of ER responsive genes (D. Y. Wang, Fulthorpe, Liss, & Edwards, 2004). Accordingly, the analysis predicts higher tamoxifen resistance in MCF-7 cells grown on softer substrates compared to glass (**Figure** 5-3**A**). To test this, we performed a cell survival assay (CellTiter-Glo 2.0) under different tamoxifen concentrations (see methods). We find that cells on stiffer substrates (Glass and 100 kPa) are more sensitive to tamoxifen treatment compare to cells on soft substrates (0.8 kPa). In particular, at a tamoxifen concentration of 25nM complete cellular death is observed on glass, 40% survival on 100 kPa and 80% survival on 0.8 kPa (**Figure** 5-3**B-C**).

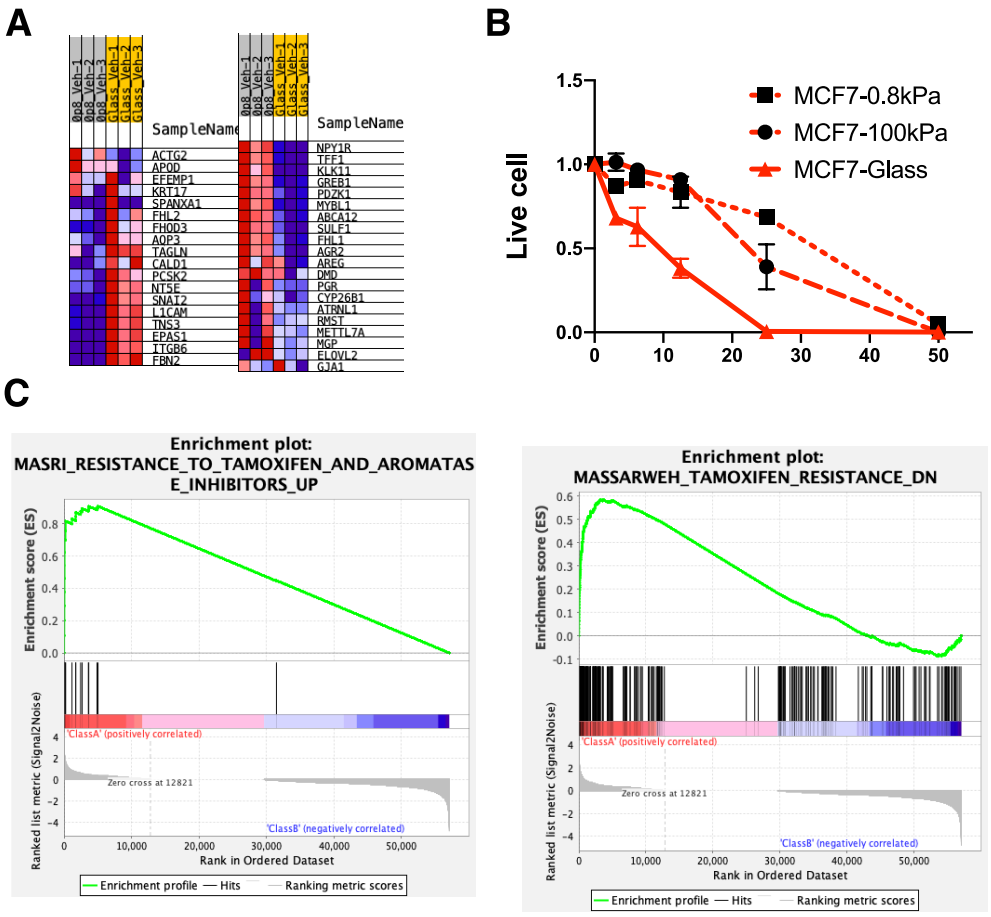


Figure 5-3 Substrate stiffness regulates tamoxifen resistance
(A) Downregulated (Left) and Upregulated (Right) genes in 0.8KPa compared to glass that are related to tamoxifen resistance in MCF-7 cells. Red represents upregulation, blue represents downregulation. Biological replicates are shown. (B) ATP survival assay for different concentrations of Tamoxifen (X-axis) of MCF-7 cells on substrates of different stiffness. Error bars represent standard error. (C) GSEA enrichment analysis results for MASRI resistance to tamoxifen and aromatase inhibitors. Upregulated genes involved in drug resistance are shown in the top panel and downregulated genes in the bottom panel. Class A corresponds to vehicle treatment on 0.8KPa gels and Class B corresponds to vehicle treatment on glass. Enrichment p-value of less than 0.01, normalized enrichment scores higher than 1.5 for upregulation and less than -1.5 for downregulation.

5.3.3. Liganded ER Gene Regulatory Functions are Optimized on Substrates of Physiological Stiffness

Given the remarkable changes in overall transcriptional behavior of MCF-7 cells induced by soft substrates, we next explored whether transcriptional activity of ER could be modulated by substrate stiffness. To this end, we performed RNA-seq following 2 hours of 100nM estradiol (E2) treatment on MCF-7 cells grown on

substrates with different stiffness. We found that liganded ER action was drastically affected by stiffness (**Figure** 5-4**A**), with a group of genes that are commonly upregulated in cells grown on physiological stiffness (0.8 kPa and 12 kPa), but are observed to be downregulated in cells grown on stiffer substrates (100 kPa and glass) and vice versa. Transcriptional profiles after E2 treatment were naturally clustered differentially in soft (0.8 and 12 kPa) and stiff (100 kPa and glass) substrates implying changes in transcriptional regulation of ER due to substrate stiffness. We found that while over 100 common genes undergo differential regulation on soft (0.8 and 12 kPa) substrates, only 13 common genes exhibited differential regulation on both soft and stiff substrates (**Figure** 5-4**B-C**). This implies that the expression of ER regulated genes is dependent on substrate stiffness, with unique sets of genes differentially expressed on soft versus stiff substrates. A number of induced/repressed genes are affected by stiffness with highest transcriptional regulation at 12 kPa and 0.8 kPa (**Figure** 5-4**B-D**), which indicates optimality of ER activity under physiological stiffness, since the receptor induces higher transcriptional activity on soft substrates seen by an increased number of differentially regulated genes (size of Venn diagram) and by the overall induction of the union of upregulated genes (**Figure** 5-4**E**).

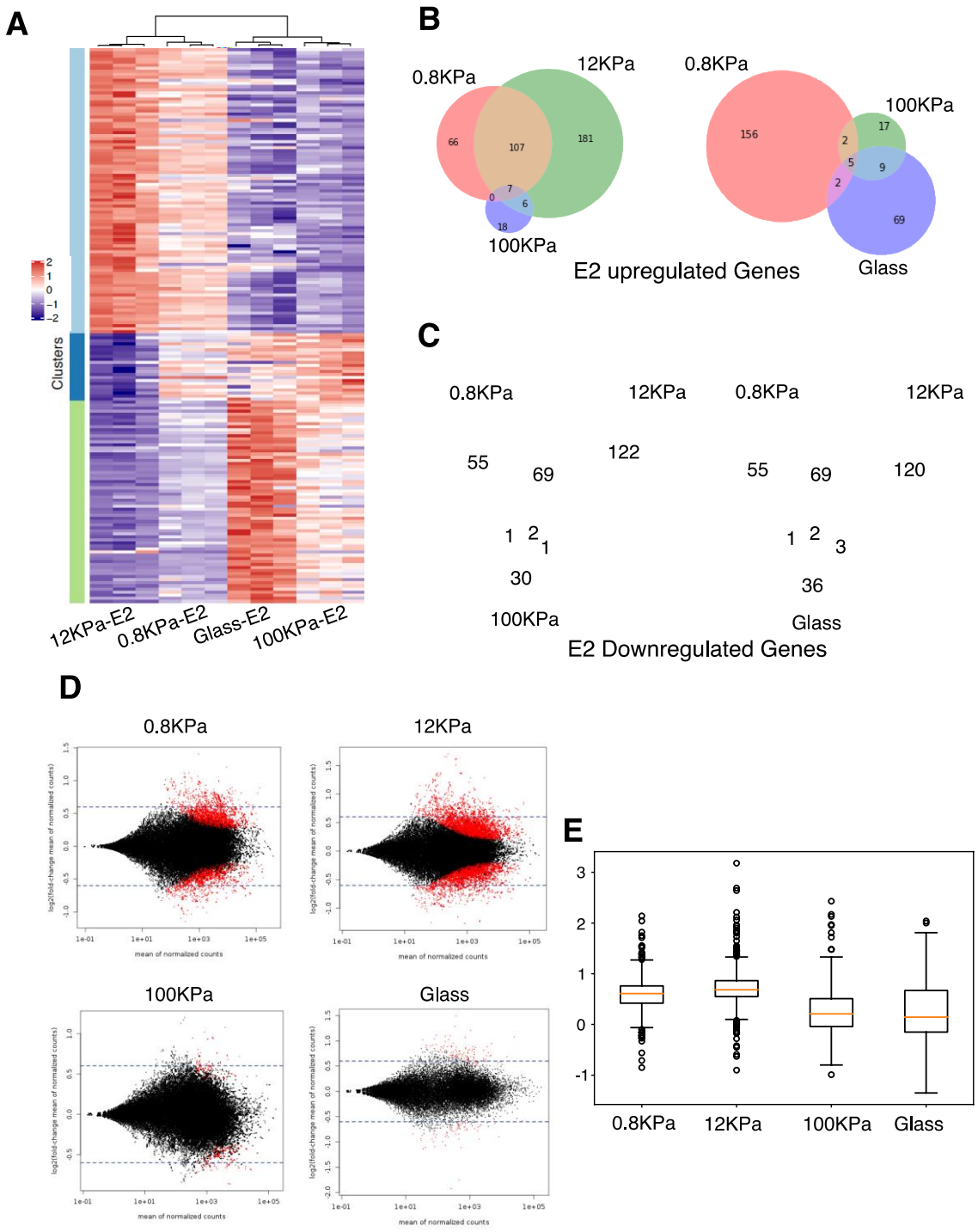


Figure 5-4 Estradiol response is potentiated at physiological stiffness

(A) Hierarchical clustering of MCF-7 cells RNA-seq grown for 48 hours on substrates of different stiffness in CCS media and 2 hours 100nM E2 treatment. Soft substrates (0.8KPa and 12KPa) transcriptome clustered together while the transcriptomes for cells on stiff substrates (100KPa and Glass) clustered in another hierarchical group. Two main gene sets with different response to E2 and stiffness is observed. Cluster 1 (light blue) corresponds to genes with higher expression on soft substrates while Cluster 3 (green) corresponds to genes with higher expression on stiff

substrates. **(B)** Venn diagram of differentially expressed genes (upregulated) in MCF-7 cells on substrates with different stiffness upon E2 treatment. Common upregulated genes correspond to the intersections in the Venn Diagrams. Differentially expressed genes corresponds to genes with adjusted p-value < 0.01 and Log2Fold induction of 0.5. **(C)** Venn diagram of differentially expressed genes (downregulated) of MCF-7 cells upon E2 treatment on substrates of different stiffness. Common downregulated genes correspond to the intersections in the Venn Diagrams. Differentially expressed genes correspond to genes with adjusted p-value < 0.01 and Log_2-fold induction of 0.5. **(D)** Volcano plot for RNA-seq of MCF-7 cells after 2 hours 100nM E2 treatment on substrates with different stiffness. Black dots correspond to genes with a p-value higher than 0.01 while red dots correspond to genes with a p-value lower than 0.01. Adjusted p-value calculated with biological triplicates. X-axis corresponds to the mean normalized count of a particular gene and y-axis the Log_2-fold induction with respect to vehicle. **(E)** Boxplot of the log_2-fold induction for the union of upregulated genes after E2 treatments on surface of different stiffness. The level of upregulation is similar between 0.8KPa and 12KPa (p-value higher than 0.01) and similar between Glass and 100 KPa (p-value higher than 0.01). Statistically significant difference between 0.8KP and 100KPa, Glass (p-value lower than 0.01) in the level of upregulation.

To further understand the dependence of ER activity on stiffness, differentially expressed genes between Glass-veh and 0.8 kPa-veh were compared to differentially expressed genes between 0.8 kPa-E2 and 0.8 kPa-veh. If ER activity is independent of substrate stiffness, no correlation would be expected between differentially expressed genes in the vehicle conditions and E2 treatment. However, this analysis revealed a strong correlation between ER and stiffness regulated genes (**Figure** 5-5**A** and **Figure** 5-4**E**) indicating ER transcriptional activity and stiffness-related transcriptional regulation. Upregulated (downregulated) genes due to stiffening of the substrates (yellow circles, **Figure** 5-5**A**) corresponds to upregulated (downregulated) genes by estradiol treatment (Blue circles, **Figure** 5-5**A**). Anticorrelation of differentially expressed genes between Veh 0.8 kPa/Veh Glass and 0.8KPa E2/0.8 kPa Veh (**Figure** 5-5**B-C**) implies antagonistic effect of E2 treatment and substrate stiffness, wherein E2 treatment works as an opposite force in terms of transcriptional changes due to stiffness (**Figure** 5-5**B-C**). This observation suggests the existence of a novel crosstalk between mechanical regulation of gene expression and ER activity.

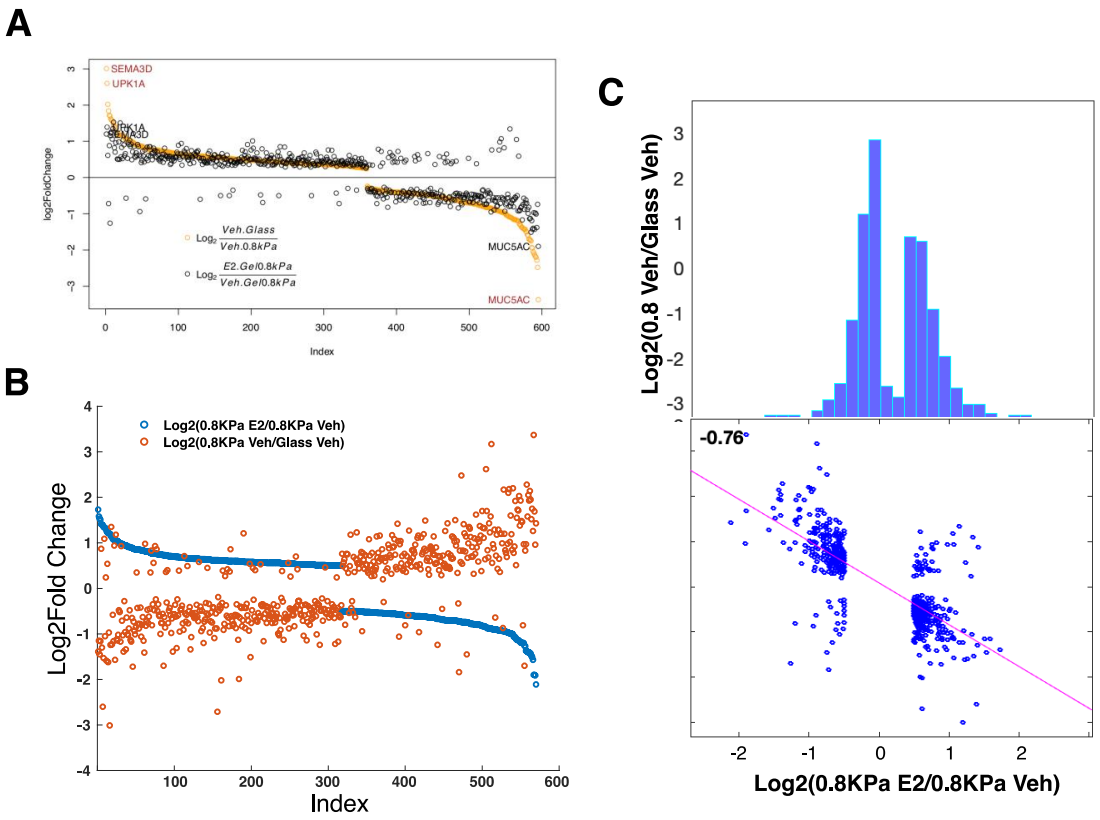


Figure 5-5 Antagonistic effects of ER gene regulation and stiffness.
(A) Log_2-Fold change of the intersection of the set of differentially expressed genes between Glass-Veh/0.8KPa-Veh (yellow) and 0.8KPa E2/0.8KPa Veh (black) ordered from highest upregulated gene in 0.8KPa-E2 to lowest. Differentially expressed genes correspond to genes with adjusted p-value < 0.01 and Log2Fold induction of 0.5. Example genes are listed. (B) Log_2-fold change of the intersection of the set of differentially expressed genes between 0.8KPa-Veh/Glass-Veh (Orange) and 0.8KPa E2/0.8KPa Veh (Blue) ordered from highest upregulated gene in 0.8KPa-E2 to lowest. (C) Histogram (Upper) and scatter plot (Lower) of the correlation in Log2 fold change between the genes in (B). Correlation coefficient of -0.76.

5.3.4. Unliganded ER Chromatin Binding and Transcriptional Activity is Modulated by the Mechanical Properties of the Cell Microenvironment

Changes in transcriptional activity of ER due to stiffness might be caused by modulation of ER dynamics, binding and/or chromatin accessibility. Moreover, it has been suggested that the receptor might have regulatory functions in the absence of hormone, but this is still under debate (Caizzi et al., 2014b). We thus performed ER CHIP-seq to characterize changes in binding of ER as a function of

changes in substrate stiffness. ER binding to genomic regions upon hormone treatment is comparable across different substrates but unliganded ER binding is strongly dependent on stiffness (**Figure** 5-6**A**) with minimal unliganded ER binding on glass (upper peaks). Surprisingly, in the absence of hormone, 30671 peaks were detected on 0.8 kPa stiffness gels, in contrast to 5180 peaks observed on glass. Unliganded ER binding and transcriptional activity is clearly modulated by stiffness and these observations are consistent with increased unliganded ER binding during tumorigenesis in breast cancer patients (Chi et al., 2019; T. W. Miller et al., 2011). Strikingly, 100 kPa shows a significant amount of unliganded ER binding but poor transcriptional response observed with RNA-seq, implying that binding is not a sole predictor of gene regulation and the dynamics of the receptor must be studied as a function of substrate stiffness. Moreover, unliganded ER binding is potentiated at lower stiffness (**Figure** 5-6**A**) further supporting the evidence for unliganded ER transcriptional activity and its antagonistic effect as observed in the RNA-seq data.

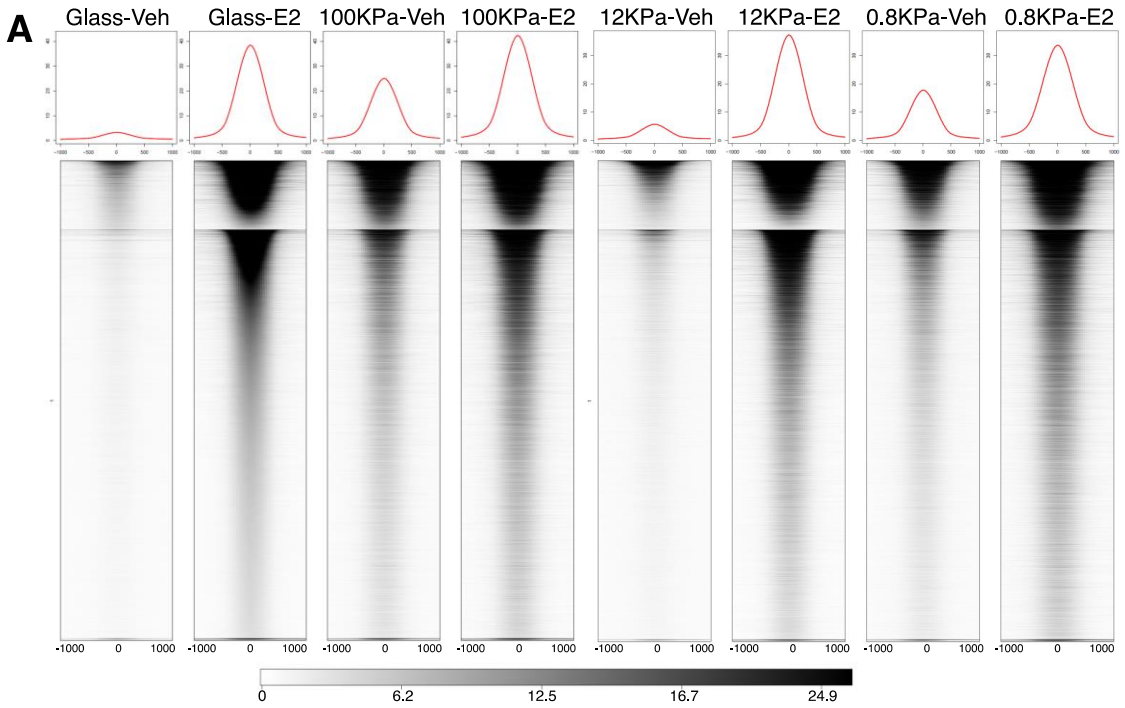


Figure 5-6 Unliganded ER binding is dependent on substrate stiffness
(A) Heatmap of the union of ER CHIP-seq of MCF-7 cells on different stiffness after 48 hours in CSS and 2 hours of 100nM E2 treatment. ER peaks are divided into group 1 with peaks observed in Glass-veh and group 2 corresponds to gained peaks. Aggregate plot for the heatmap is plot for each condition (Red).

To further explore the antagonistic action of liganded and unliganded ER, a new MCF-7 derived cell line was developed by depriving the cells from hormone for 6 months (see methods). Since MCF-7 cells are normally estrogen-dependent (Clarke et al., 1989), this long selection method allows us to isolate the small population of cells that are estrogen-independent, which we called MCF7-LTED (Long Term Estrogen Deprivation). As expected, LTED cells can grow in a hormone independent manner (**Figure** 5-7**A**) and ER levels are significantly reduced (**Figure** 5-7**B-C**). Thus, this cell line allows us to discriminate between unliganded ER action and ER deprivation. Using MCF7-LTED as a baseline, transcriptional effects of unliganded ER can be characterized by comparing RNA-seq datasets upon 48 hours hormone deprivation in wild type MCF-7 with MCF7-

LTED (**Figure** 5-7**D**). Most of the genes found are ER regulated genes (**Figure** 5-7**E-F**). The level of induction of these genes was then compared to the differentially expressed genes between E2 and CSS treatment on glass in MCF-7 (liganded ER gene regulation) confirming an antagonistic effect of liganded and unliganded ER transcriptional regulation (**Figure** 5-8**A**) with a negative correlation of 0.89 between the two groups (**Figure** 5-8**B**). These results are consistent with antagonistic effects of hormone treatment and hormone deprivation on transcription in breast tumors from patients (Chi et al., 2019) and suggest that the antagonistic effects of unliganded ER on MCF-7 are potentiated at physiological stiffness (Caizzi et al., 2014a).

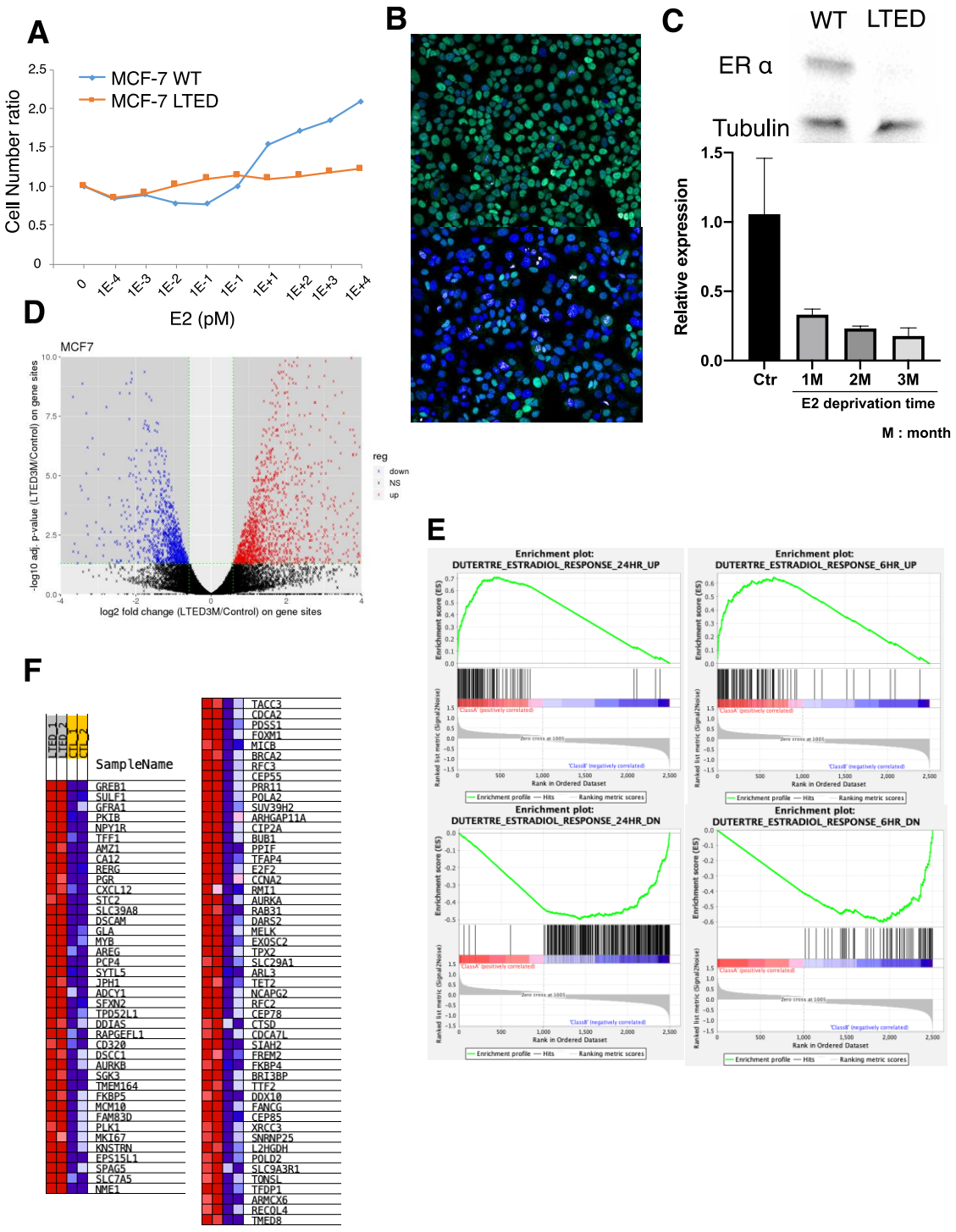


Figure 5-7 Unliganded ER exhibits transcriptional regulatory functions

(A) Cell survival assay for MCF-7 WT and LTED in terms of E2 concentration normalized to cell number in full media. MCF-7 LTED growth is independent of E2 concentration as expected from the downregulation of ESR1. (B) Immunostaining of ER (Green) in MCF7-LTED (Bottom) and MCF7-WT (Top). DAPI in Blue. (C) Western blot for ERα between MCF7-LTED after 3 months of hormone deprivation and MCF-7 LTED. Tubulin as control. Bottom panel, q-PCR relative expression of ESR1 with respect to estrogen deprivation time. (D) Log_2-fold change and adjusted P-value for differential gene expression analysis between MCF7-LTED Veh and MCF7-wt Veh.

Significance level set to adjusted p-value of 0.01 and fold induction of 2. (E) GSEA enrichment analysis results for DUTERTE Estradiol response 24 hours (Left panel) and 6 hours (Right panel). Class A corresponds to MCF-7-LTED and Class B corresponds to MCF-7 WT on glass. Enrichment p-value of less than 0.01, normalized enrichment scores higher than 1.5 for upregulation and less than -1.5 for downregulation. (F) Rightmost panel corresponds to the heatmap of ER upregulated genes in MCF7-LTED and MCF-7 wt.

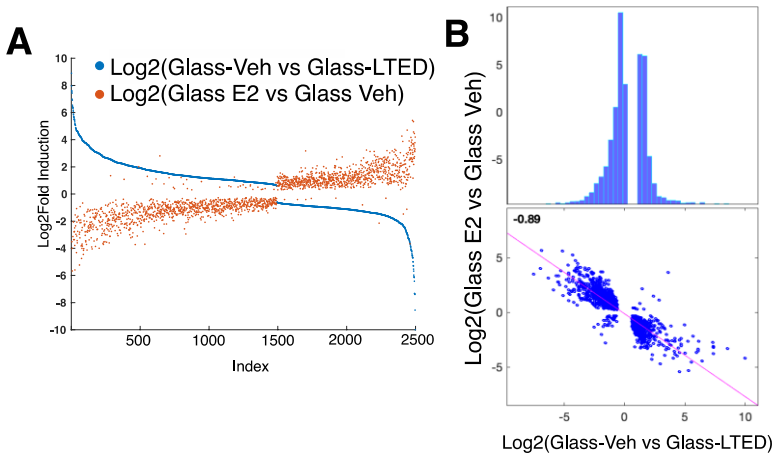


Figure 5-8 Uliganded ER presents antagonistic gene regulatory functions compared to liganded ER

(A) Log_2-fold change of the intersection of the set of differentially expressed genes between MCF7-Veh/MCF-7-LTED (blue) and MCF-7 E2/MCF-7 Veh (red) ordered from highest upregulated gene in MCF7-Veh to lowest. (B) Histogram (Upper) and scatter plot (Lower) of the correlation in Log2 fold change between the genes in (A). Correlation coefficient of -0.89.

5.3.5. Chromatin Accessibility is Weakly Affected by Substrate Stiffness

Several prior studies have suggested that chromatin accessibility is one of the main factors affected by the mechanical environment, with increased chromatin accessibility on softer substrates (Kirby & Lammerding, 2018; Stowers et al., 2019). To further study whether the effects of substrate stiffness on gene regulation and ER action similarly arise from changes in accessibility, we performed ATAC-seq on MCF-7 cells on different stiffness. In the absence of hormone, the differences in accessibility profiles across different substrates were small, with the vast majority of ATAC-seq peaks shared among conditions (**Figure**

5-9**A-B-C**). Comparison between extreme stiffness conditions (0.8 kPa and glass) reveals small differences in chromatin accessibility in the absence of estradiol (**Figure** 5-9**A-B-C**). Upon addition of estradiol, chromatin accessibility increased independent of stiffness, consistent with previous results (**Figure** 5-9**D-E**). However, analysis of ATAC signal at ER binding sites by ChIP revealed significant changes in chromatin accessibility induced by ER binding in a stiffness dependent manner (**Figure** 5-9**F**). In particular, E2 treatment increased chromatin accessibility in ~20% of ER binding sites with an ATAC-peak across different stiffness. On the other hand, chromatin accessibility was significantly decreased in the remaining 80% of ER binding sites at physiological stiffness (**Figure** 5-9**F**). Interestingly, this reduction in chromatin accessibility corresponds to higher levels of chromatin accessibility in the absence of hormone at 0.8 kPa compared to glass at unliganded ER binding sites. The latter implies a potential novel chromatin remodeling action of unliganded ER.

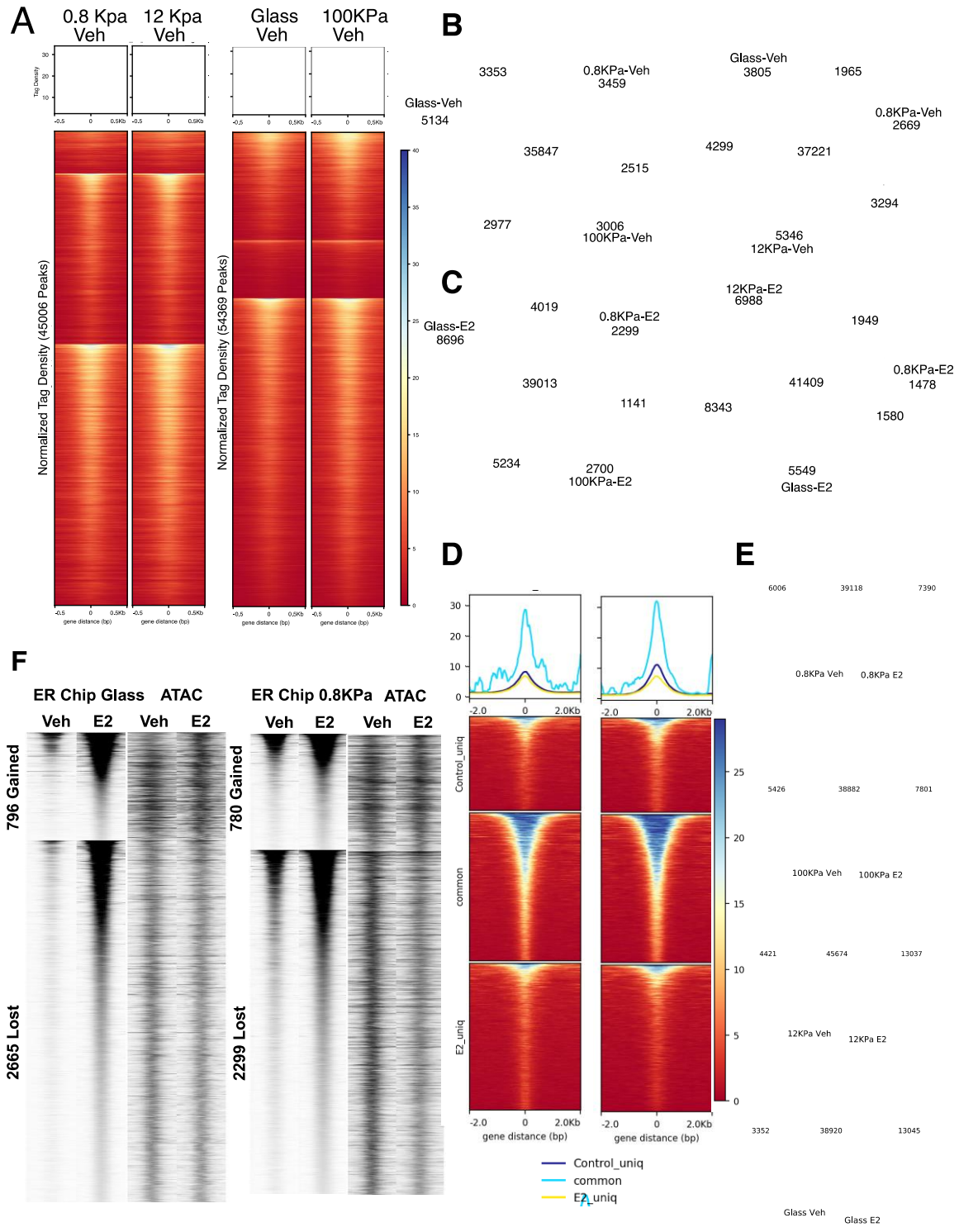


Figure 5-9 Chromatin accessibility is weakly affected by substrate stiffness
(A) Heatmap of the union ATAC-peaks between 0.8KPa-Veh, 12KPa-Veh and 100KPa-Veh, Glass-Veh around 500bp from the peak center. Aggregate plot for the ATAC-peaks is shown on top of the heatmap. **(B)** Venn Diagram for ATAC-seq between vehicle conditions and different substrate stiffness. Most ATAC peaks are shared across cells from different stiffness. **(C)** Venn Diagram for

ATAC-seq between 1-hour 100nM E2 treatment conditions and different substrate stiffness. Accessibility increases after E2 treatment for different stiffness conditions. (D) Heatmap of the union ATAC-peaks between Glass-Veh (Left) and Glass E2 (Right) divided into Vehicle (Control) unique peaks, common peaks between conditions and E2 treatment unique peaks. Heatmap around 2000bp from the peak center. Aggregate plot for the ATAC-peaks is shown on top of the heatmap. **(E)** Venn Diagram for ATAC-seq between 1-hour 100nM E2 treatment conditions and vehicle for different substrate stiffness. Accessibility increases after E2 treatment for different stiffness conditions. **(F)** Heatmap of gained and lost ATAC peaks after E2 treatment. ATAC peaks are aligned to ER CHIP peaks and sorted based on ER CHIP peak intensity. Stronger changes in chromatin accessibility are observed in 0.8KPa (right panel) compared to glass (left panel).

5.3.6. ER Binds in the Order of Minutes to Chromatin under Physiological Stiffness

Transcriptional regulation of ER as a function of stiffness is partially explained by changes in ER binding and its effect on chromatin accessibility but the higher transcriptional regulation on soft substrates and the increased ER binding at 100 kPa remains unexplained. In order to determine whether the observed alterations in transcriptional profiles induced by substrate stiffness are related to changes in ER mobility, we used single molecule tracking (SMT) experiments of Halo-tagged ER in MCF-7 cells under different stimulation conditions. We found that on substrates of physiological stiffness, fluorescent single molecule traces of ER persist in the imaging volume for several minutes, indicative of extremely long binding events (**Figure** 5-10**A**). This is contrary to previous reports of binding events on the order of seconds on glass (E.E. Swinstead et al., 2016). H2B-Halo serves as a good reporter for photobleaching since FRAP and other mobility experiments have shown significantly slower recovery of H2B compared to our experimental observation times (Kimura & Cook, 2001). Comparison of H2B-Halo and ER-Halo (after E2 treatment) reveals that ER dwell times are indistinguishable from H2B dwell times at physiological stiffness (**Figure** 5-10**B-C**) but are smaller than H2B dwell times on stiffer substrates

(**Figure** 5-10**B-D**). The photobleaching rate on 0.8 kPa substrates is 1.7 minutes (**Figure** 5-10**E**) and due to the exponential nature of photobleaching, events as long as 8 minutes can be observed, implying that ER binding times are much longer than previously thought at physiological stiffness. These longer binding times are consistent with the transcriptional efficiency of E2 treatment on physiological stiffness compared to stiffer substrates (**Figure** 5-4**B**).

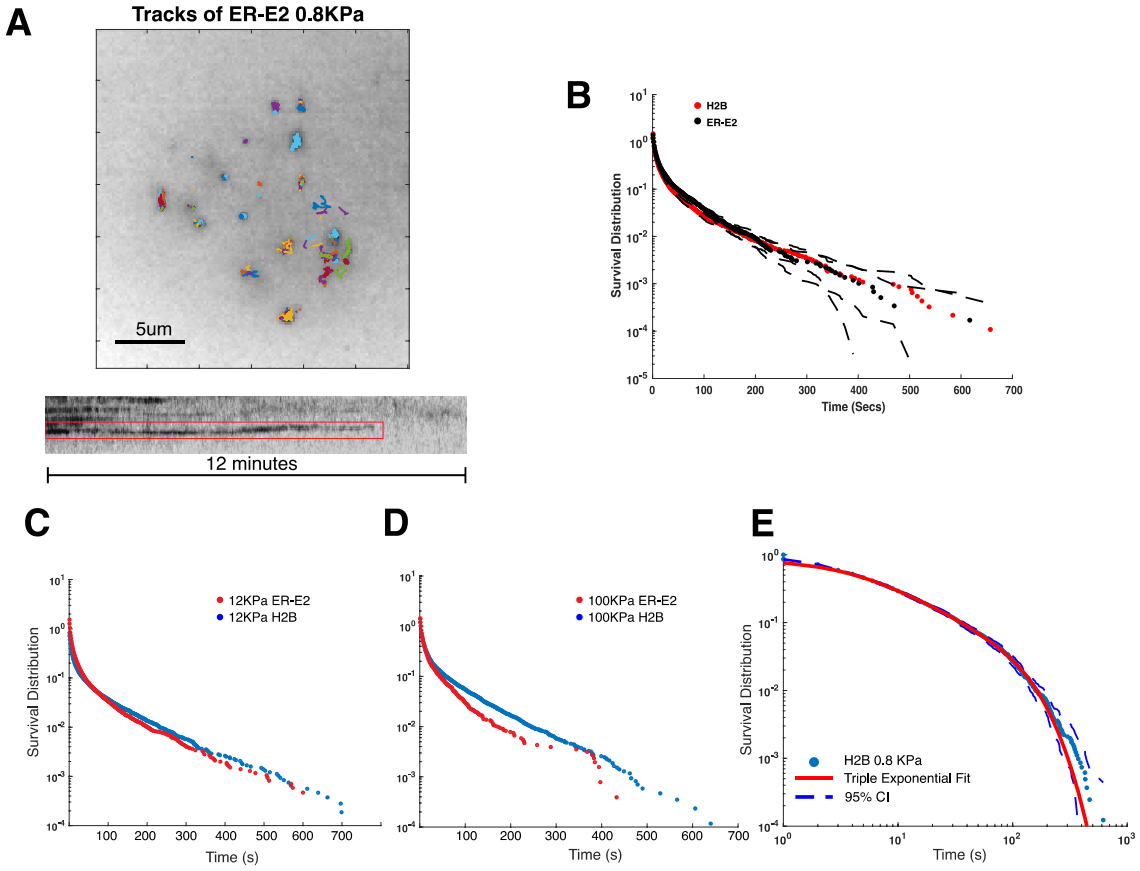


Figure 5-10 ER mobility is modulated by mechanical cues
(A) Sample ER tracks imaged by SMT after E2 treatment of MCF-7 cells on 0.8kPa substrate (top) after 12 minutes of imaging (500ms exposure and 1000ms acquisition interval). Sample kymograph showing a long ER binding event in the order of 10 minutes (bottom). (B) Semi-log plot of the dwell time distribution between ER (E2 treatment, Black) and H2B (Red) on 0.8KPa substrates. 95% Confidence Intervals (CI) are plotted in dashed lines. (C) Semi-log plot of the dwell time distribution of SMT tracks for ER (E2 treatment, red) and H2B (blue) on 12KPa substrates. 95% Confidence Intervals (CI) are plotted in dashed lines. ER binding events are as long as H2B photobleaching or disappearance from the focal plane. (D) Semi-log plot of the dwell time distribution of SMT tracks for ER (E2 treatment, red) and H2B (blue) on 100KPa substrates. ER binding events are shorter than H2B photobleaching or disappearance from the focal plane. (E) log-log plot of the dwell time

distribution of H2B (blue) in 0.8KPa and a triple exponential fit (red) characterizing photobleaching kinetics.

In Chapter 4, we showed that TF binding events result from a combination of direct interactions with chromatin as well as protein-protein interactions due to protein self-aggregation and LLPS (which we term as “confinement”) (Garcia et al., 2021a). To determine the relative contributions of chromatin binding and confinement to single molecule trajectories of ER, we used pEM (perturbation Expectation Maximization, (Koo & Mochrie, 2016)) to classify ER (E2 treatment) trajectories into these two states (**Figure** 5-11**A-B**). Mean Square Displacement (MSD) of the tracks classified as chromatin bound was independent of substrate stiffness, which is consistent with the motion expected due to the elastic properties of the chromatin polymer (Shukron & Holcman, 2017). On the other hand, we found that the MSDs for trajectories classified as confined were dependent on stiffness, with lowest MSDs on the softest substrates (**Figure** 5-11**A-B**). We next analyzed the dwell time distributions of the tracks classified as chromatin-bound or confined. The photobleaching corrected dwell time distributions showed significantly longer dwell times on 0.8 kPa gels for both confined and chromatin bound populations compared to those on stiffer substrates (**Figure** 5-11**C-D**) (Garcia et al., 2021b). Furthermore, trajectories classified as confined also showed significantly longer binding times across different stiffness compared to chromatin binding. The dwell times of these confined tracks were comparable to the timescales of H2B photobleaching on physiological stiffness (**Figure** 5-11**E**). These changes in the mobility and chromatin binding of liganded ER induced by substrate stiffness correlate with the observed changes in transcriptional activity, suggesting that

changes in ER mobility may be responsible for changes in the transcriptional response of hormones due to stiffness.

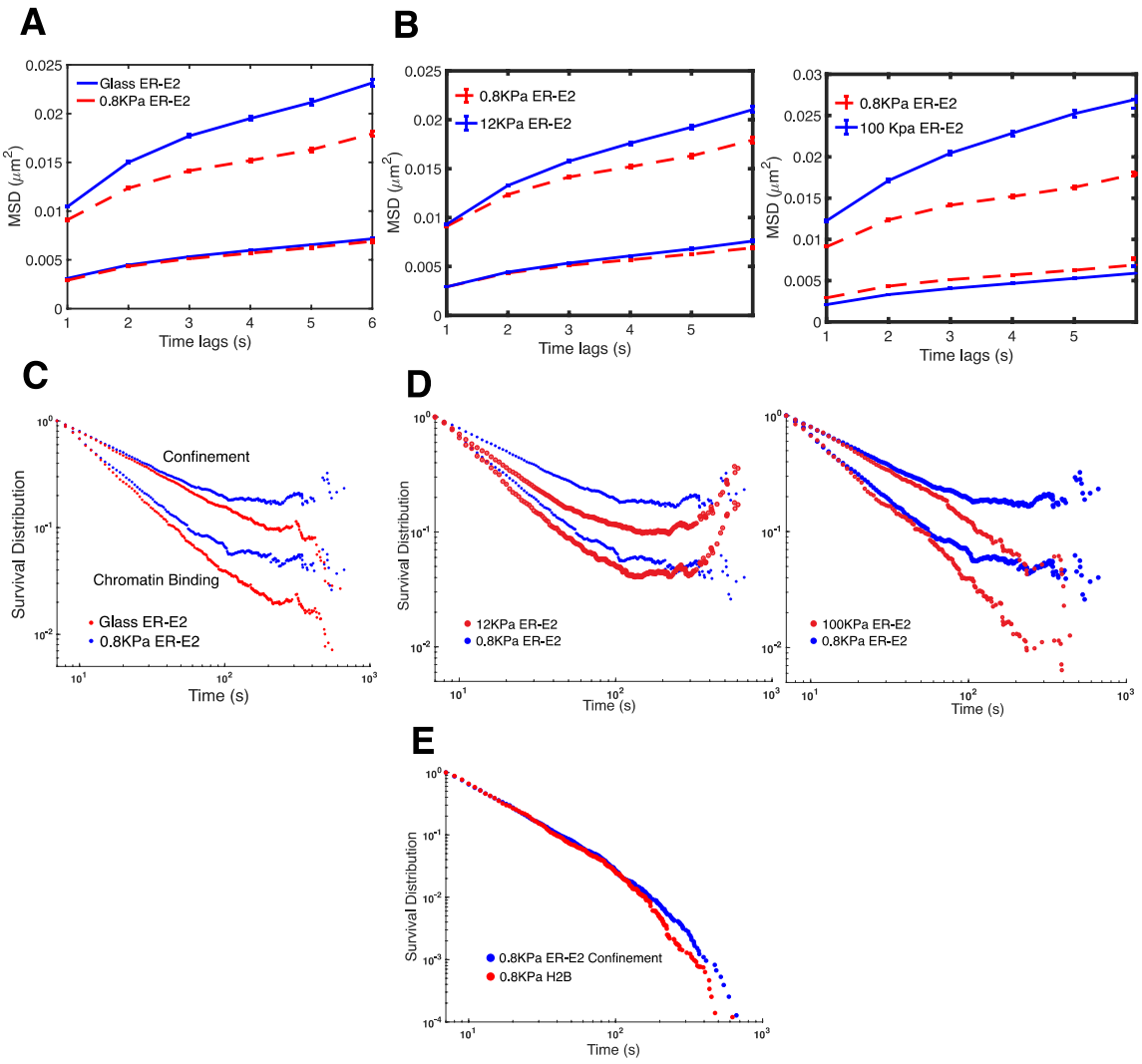


Figure 5-11 Liganded ER chromatin binding and confinement are modulated by mechanical stimuli.

(A) MSD for chromatin binding (bottom) and confined populations of SMT tracks (top) for 0.8KPa ER-E2 (red dashed lines) and Glass ER-E2 (solid blue lines). (B) MSD for chromatin binding (Bottom) and confined population (Top) of SMT tracks for 0.8KPa ER-E2 (red dashed lines) and 12KPa ER-E2 (solid blue lines), (left panel) and for 0.8KPa ER-E2 (red dashed lines) and 100KPa ER-E2 (solid blue lines), (right panel). Identical chromatin binding is observed between different stiffness while small changes in the confined population are observed. (C) log-log plot of the dwell time distribution for the confined and chromatin binding population between Glass ER-E2 (Red) and 0.8KPa ER-E2 (Blue). (D) log-log plot of the dwell time distribution for chromatin binding and confinement population of SMT tracks for 12KPa ER-E2 (red) and 0.8KPa ER-E2 (blue), (left panel) and for 100KPa ER-E2 (red) and 0.8KPa ER-E2 (blue) (right panel). Liganded ER binds longer on 12KPa than 0.8KPa as seen by the plateau in the tail of the distribution. Liganded ER binds longer on 0.8KPa than 100KPa. (E) log-log plot of the dwell time distribution for the confinement population

of SMT tracks for 0.8KPa ER-E2 (blue) and H2B (red). Liganded ER confinement dwell time are as long as photobleaching kinetics of the experiment.

Since ChIP-seq and RNA-seq analysis showed that substrate stiffness modulated unliganded ER binding and its transcriptional activation, we next performed SMT on unliganded ER to characterize the changes in unliganded ER mobility and its relation to gene regulation. We found that the mean-squared displacement of trajectories for the confined and chromatin bound states were similar between liganded and unliganded ER across substrates of different stiffness (**Figure** 5-12**A** and **Figure** 5-11**B**). However, analysis of the survival distributions showed that both the confined and chromatin bound states for E2 stimulation were longer lived than those for unliganded ER (**Figure** 5-12**B-C**). Since our whole genome and transcriptome analysis showed that even unliganded ER had significant transcriptional activity in cells on physiological substrates, we next compared unliganded ER binding dwell times at 0.8KPa and liganded ER binding dwell times on glass. Remarkably, we found that the distribution of binding times for the chromatin bound populations for unliganded ER on 0.8 kPa and liganded ER binding dwell times on glass were very similar (**Figure** 5-12**E**). Accordingly, we observed significantly shorter binding times of unliganded ER on 100 kPa substrates compared to physiological stiffness which correlates with the poor transcriptional activity of unliganded ER on 100KPa despite potentiated binding observed using CHIP-seq (**Figure** 5-12**D**). Further analysis showed that on softer substrates, the chromatin binding and confinement dwell times for vehicle treatment were not well separated (**Figure** 5-12**F**). In contrast, on stiffer substrates, the dwell time distributions were clearly separable. This separation in transcription

factor mobility is consistent with the natural transcriptome clustering between soft and stiff substrates by RNA-seq (**Figure** 5-12**G**).

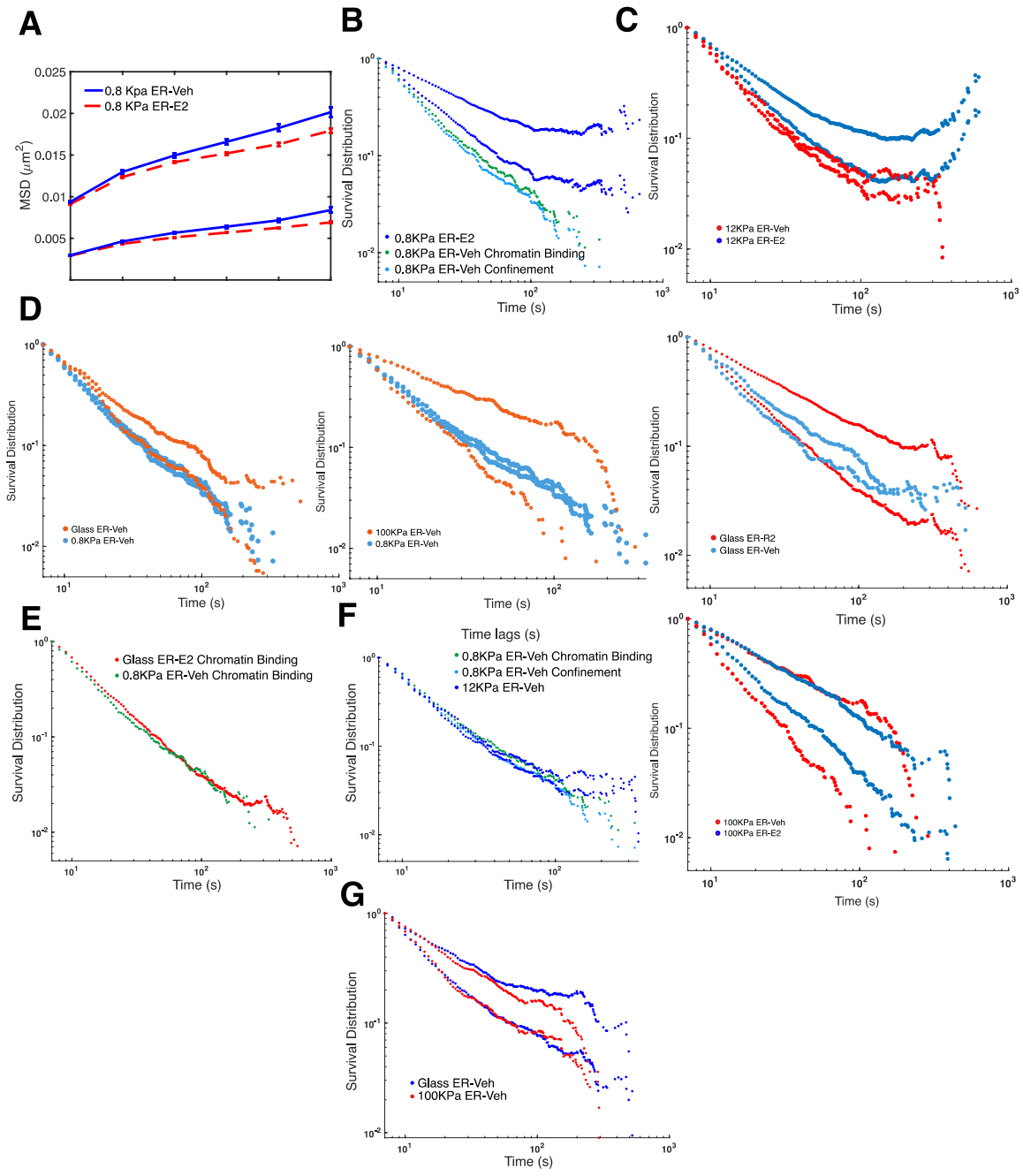


Figure 5-12 Unliganded ER chromatin binding and confinement are modulated by mechanical stimuli.

(A) MSD for chromatin binding (bottom) and confined populations of SMT tracks (top) for 0.8KPa ER-E2 (red dashed lines) and 0.8 ER-E2 (solid blue lines). MSD of the two population is unaltered by hormone binding to the receptor. (B) log-log plot of the dwell time distribution for the confined and chromatin binding populations for 0.8KPa ER-E2 (blue) and 0.8KPa ER-Veh (light blue and green). (C) log-log plot of the dwell time distribution for chromatin binding and confinement

population of SM tracks for unliganded ER (red) and liganded ER (blue) on substrates of different stiffness. Binding times are reduced in the unliganded version of the receptor. (D) log-log plot of the dwell time distribution for chromatin binding and confinement populations of SMT tracks for Glass ER-Veh (orange, top panel), 100KPa ER-Veh (orange, lower panel) and 0.8KPa ER-Veh (blue). Chromatin binding of the unliganded receptor is similar between Glass and 0.8KPa but faster in 100KPa. (E) log-log plot of the dwell time distribution for chromatin binding populations of SMT tracks for glass ER-E2 (red) and 0.8KPa ER-Veh (green). (F) log-log plot of the dwell time distribution for chromatin binding and confinement populations of SMT tracks for 12KPa ER-Veh (blue) and 0.8KPa ER-Veh (green and light blue). (G) log-log plot of the dwell time distribution for chromatin binding and confinement populations of SMT tracks for Glass ER-Veh (blue) and 100KPa ER-Veh (red).

The dramatic changes in nuclear morphologies (**Figure** 5-1**C**) and H2B dynamics (**Figure** 5-13**A**) in response to substrate stiffness suggest that these arise as a result of global changes in chromatin mobility (Zidovska, Weitz, & Mitchison, 2013). To investigate this possibility, we measured the dynamics of H2B-Halo fluorescence intensity and quantified chromatin displacements and speed (See methods). Single plane confocal images were acquired on substrates of different stiffness and the velocity vectors calculated using Particle Image Velocimetry (**Figure** 5-13**B**). We found that substrate stiffness had a strong effect on the speed distribution with chromatin mobility being largest at physiological stiffness compared to stiffer substrates (**Figure** 5-13**C**).

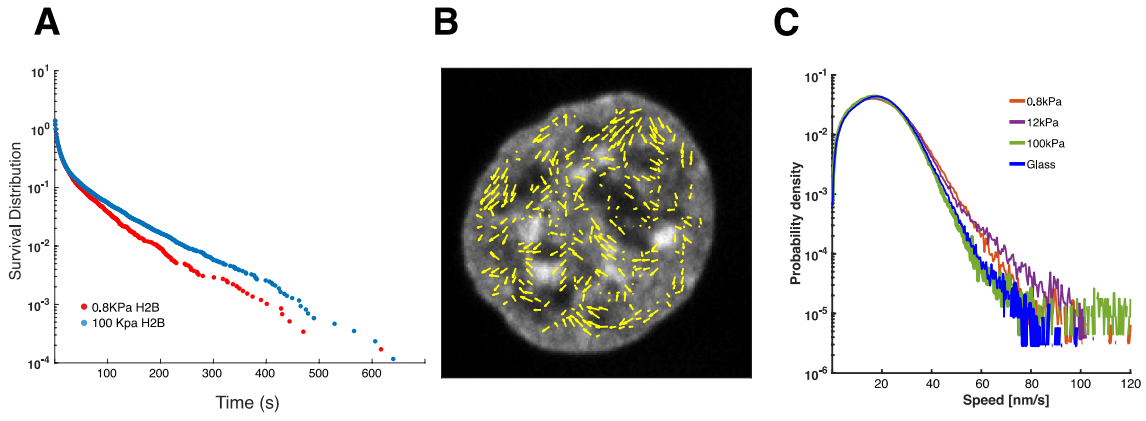


Figure 5-13 Chromatin mobility is potentiated on physiological stiffness
(A) log-log plot of the dwell time distribution of SMT tracks for H2B on 0.8KPa (red) and 100KPa (blue). Photobleaching is slower at 100kPa than 0.8KPa. (B) Example speed vector map of

chromatin motion on glass. (C) Chromatin speed probability density on substrates of different stiffness. Ensemble average from three biological replicates.

5.4. Discussion

This work demonstrates that alterations in ECM stiffness can alter chromatin state, global nuclear dynamics and transcription factor mobility and binding, which in turn regulates phenotypic changes. Prior work has shown the importance of the physical properties of the environment, in particular stiffness, to cell fate and behavior (Engler et al., 2006; Stowers et al., 2019) with profound changes in chromatin accessibility and gene expression. However, transcription factor action and cell response to external stimuli in substrates of different stiffness has not been carefully explored. Our findings highlight a previously undescribed effect of substrate stiffness on cellular response to external stimuli. Even though stiffness is accepted as an important stimulus that regulates cellular phenotype, most studies on the regulation of gene expression by hormonal stimulation performed on rigid surfaces like glass and. Here, we explore how substrate stiffness affects chromatin accessibility, transcription, transcription factor mobility, action and binding as well nuclear mobility. We find that changes in substrate stiffness induce significant phenotypic changes with clearly differentiated transcriptomes between rigid and physiological soft substrates. Moreover, breast cancer MCF-7 cells grown on physiological stiffness exhibit higher levels of tamoxifen resistance, further highlighting the importance of studying pharmacological processes in substrates of similar stiffness as in vivo.

Our RNA-seq and ATAC-seq analysis reveal that ER retains significant transcriptional activity and chromatin remodeling capacities on soft substrates

even in the absence of its ligand, while losing genomic activity on stiff substrates. In addition to the potentiation of its binding ability at physiological stiffness, chromatin binding times of unliganded ER on physiological stiffness are comparable to liganded ER on rigid substrates. Finally, we find that unliganded ER exhibits significant antagonistic transcriptional action at physiological stiffness to liganded ER which is consistent with clinical studies of estrogen deprivation, ER binding and transcriptional regulation (Chi et al., 2019; T. W. Miller et al., 2011) and potential unliganded ER activity on glass (Caizzi et al., 2014a; Welboren et al., 2009). Taken together, our results suggest a relevant role for unliganded ER in hormone resistance in breast cancer and the importance of studying breast cancer biology on physiologically relevant substrates.

We discovered a remarkable dependence of liganded ER dynamics and transcriptional regulation on substrate stiffness. We found that liganded ER activity is potentiated at physiological stiffness compared to rigid substrates with higher transcriptional activation, repression and longer binding times to chromatin. Single molecule tracking studies show that ER binding time to chromatin and protein-protein interaction is modulated by substrate mechanics with liganded ER binding on the order of minutes, much longer than previous binding times reported on glass (on the order of seconds) (E.E. Swinstead et al., 2016). Moreover, transcriptional regulation of liganded ER shows distinct differences in response to substrate stiffness with one group of genes upregulated on soft substrates (0.8KPa and 12KPa) while being downregulated on stiff substrates (100KPa and Glass) and vice versa for a second group of genes (**Figure** 5-4**A**). These results are consistent

with previous work on RNA-seq analysis of tumorigenesis where a group of E2-responsive genes are induced in normal breast were downregulated in breast tumor and vice-versa (Chi et al., 2019) and this serves as further evidence of the importance of microenvironment stiffening in tumorigenesis (J. Bauer et al., 2020; Broders-Bondon, Ho-Bouldoires, Fernandez-Sanchez, & Farge, 2018; Kalli & Stylianopoulos, 2018).

To the best of our knowledge, this is the first time that experiments on TF dynamics have been performed on physiological stiffness. SMT of liganded and unliganded ER reveals a remarkable dependence of ER binding and protein-protein interaction on substrate stiffness. Similar to changes in gene expression, ER dynamics in the nucleus show distinctly different binding times depending on substrate stiffness, with significantly longer binding times on physiological stiffness. In addition to changes in TF mobility, substrate stiffness significantly alters nuclear dynamics and in particular increased global chromatin mobility on soft substrates, emphasizing the importance of studying transcriptional regulation at the dynamical level due to the rapid changes in the nuclear environment.

By analyzing transcription, chromatin accessibility, chromatin mobility and ER dynamics and binding, we revealed the intricate connection between substrate stiffness and transcription factor action with important connections to the understanding of ER biology with far reaching consequences for breast cancer biology. We emphasize the importance of studying biological processes, as basic as transcriptional regulation, in substrates with similar mechanical properties *in vivo*. Further experiments on acquired drug resistance *in vitro* and tumorigenesis

will benefit tremendously from using compliant matrices such polyacrylamide gels. Overall, stiffness modulates transcriptional regulation in terms of chromatin accessibility, transcription factor binding, mobility and action with profound consequences in cancer biology such tumorigenesis and drug resistance.

5.5. Methods

5.5.1. Plasmids and cell lines

The pHaloTag-ER expresses human ER with HaloTag (Promega, Madison, WI, USA) fused in the C-terminal domain under the CMVd1 promoter. The pHaloTag-H2B expresses histone H2B fused with the HaloTag through its N-terminal end (Mazza et al., 2012a).

MCF7 cells were obtained from ATCC and cultured in 5% CO2 at 37°C. MCF7 cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) with 4.5g/l of D-glucose (Thermo Fisher Scientific), and supplemented with 10% FBS, 2 mM L-glutamine, 1 mM sodium pyruvate. MCF7 LTED cells were established by culturing MCF7 cells in phenol red-free DMEM with 4.5g/l of D-glucose (Thermo Fisher Scientific), and supplemented with 10% charcoal- stropped FBS, 2 mM L-glutamine, 1 mM sodium pyruvate for > 3 months.

For transfections, 5 million cells were electroporated using BTX T820 Electro Square Porator (Harvard Apparatus, Holliston, MA, USA) in 100 μl of DPBS with 2.5 μg of plasmid. 25 ms pulses of 120 V were used and cells were resuspended in fresh media. Single-molecule imaging experiments were set up as follows: 500,000 electroporated cells were seeded onto a 35 mm circular glass

bottom dish (1.5 German borosilicate coverglass, Cellvis, Mountain View, CA, USA) in high glucose DMEM supplemented with 10% FBS (Life Technologies), 2mM L- glutamine, 5 µg/ml tetracycline, and cultured overnight. The media was then replaced with high glucose DMEM supplemented with 10% charcoal stripped FBS (Life Technologies), 2mM L- glutamine, 5 µg/ml tetracycline, and incubated at 37°C for 24 hours before labeling.

5.5.2. Endogenous H2B gene tagging using CRISPR-Cas9

For C terminal tagging of H2B family member G gene (HIST1H2BG, NM_003518) with a Halo tag we used a donor plasmid consisting of an upstream homology arm (~800 bp long), 48 bp of TEV linker and a Halo sequence with a stop codon followed by a downstream homology arm (~800 bp long). MCF7 cells grown to ~70% confluence were trypsinized and $5x10^6$ cells were resuspended in 100 µl of growth media (DMEM) without serum or antibiotic and electroporated with the donor plasmid and two guide RNA plasmids to increase knock-in efficiency at a ratio of 2:1:1 (total 12 µg). The guide RNAs were cloned into the Addgene lentiCRISPR v2 plasmid (Plasmid#52961) that contains the inserts Cas9 and Puromycin resistance. As the cells transfected with the guide RNA plasmids acquire a temporal resistance to Puromycin, the electroporated cells were grown with 1 µg/ml Puromycin for 2 days and the surviving cells were washed 3X with PBS and regrown without antibiotic to ~$10x10^6$ cells. For the isolation of Halo-tagged clones, cells were incubated with 5nM of HaloTag® TMR Ligand for 20 minutes, at 37^{o}C washed 3x with warm media and left to recover for additional 20 minutes at 37oC. The cells were further trypsinized, pelleted in DMEM, washed

with PBS and resuspended in PBS before FACS sorting and plating of the individual positive cells in 96 well plate. The expression of the H2B-Halo was further verified by microscopy and 3 individual single cell clones were selected and expanded further.

5.5.3. Polyacrylamide Gel Fabrication

35 mm circular glass bottom dish (#1.5, Cellvis) bottom glass dishes and rectangular glass slides (178x127 mm) are incubated for 15 minutes in 5% 3-aminopropyltrimethoxysilane in ddH2O (v/v) and then washed three times for 10 minutes in ddH2O. Dishes and glass slides are incubated for 15 minutes in 5% glutaraldehyde in ddH2O(v/v) and then washed three times for 10 minutes in ddH2O, dishes and glass slides are dried overnight at RT. An aliquot of the acrylamide and bis-acrylamide working solution (Table 5-1) is degassed for 20 minutes, 5ul of 10% APS and 1.5ul of TEMED (Biorad) are added per 994.5ul of working solution; 10 ul of solution are added to the circular glass bottom dish and a Rainx treated 15mm coverslip is added on top (2 ml of solution is added to the rectangular glass slides and a rainx treated glass slide of same size is added on top). The solution is left to polymerize at RT for the time in Table 5-1.

Polymerized gels are incubated in hydrazine hydrate (78-82% iodometric, Sigma-Aldrich) over night at RT. Gels are washed several times with ddH2O and incubated in 5% glacial acetic acid in ddH2O (w/w) for one hour. Gels are coated with Collage IV from Human Placenta (Sigma) through EDC carbodiimide crosslinking. Gels are incubated for 24 hours with 4 exchanges of MES buffer (pH

5). After 24 hours, gels are incubated in 200ul (2 ml for rectangular glass slide gels) of coupling solution (1mg/ml Collagen IV from Human Placenta, 1mM NHS and 1mM EDC in MES buffer) O.N at RT. Gels are washed extensively with DPBS and they are ready for cell culture.

5.5.4. Fluorescent labeling of Halo-tagged molecules and hormone treatments

Transfected cells were incubated with 5 nM JF_{646}-HaloTag or 250nM JF_{549}-HaloTag (For chromatin mobility experiments) (Grimm et al., 2015) ligand for 20 min at 37°C. Free ligand was depleted by washing three times with phenol red free DMEM media (supplemented with 10% charcoal stripped FBS and 5 µg/ml tetracycline) in 15 min intervals at 37°C. Next, cells were treated with 100 nM Estradiol (E2) (Sigma-Aldrich) or left untreated and incubated for 20 min at 37°C before imaging. For confocal imaging in **Figure 1B-C**, Cells were fixed with 4% (wt/vol) paraformaldehyde (Electron Microscopy Sciences) and 0.2% (wt/vol) glutaraldehyde (Electron Microscopy Sciences) for 35 min at room temperature. After fixation, cells were incubated with ProLong Gold Antifade Mountant for 24 hours.

5.5.5. Image acquisition for single-molecule tracking

A custom HiLO microscope was used as previously described in detail elsewhere (D.M. Presman et al., 2017), with an objective heater to reduce drift. Briefly, the custom-built microscope from the CCR, LRBGE Optical Microscopy Core facility is controlled by µManager software (Open Imaging, Inc., San

Francisco, CA.), equipped with an Okolab state top incubator for CO_2 (5%) and temperature control (37°C), a 100X 1.35 numerical aperture silicone immersion objective (Olympus Scientific Solutions, Waltham, MA), a 561nm and 473nm lasers (iFLEX-Mustang, Excelitas Technologies Corp., Waltham, MA), and an acousto-optic tunable filter (AOTFnC- 400.650, AA Optoelectronic, Orsay, France). Images were collected on an EM-CCD camera (Evolve 512, Photometrics). The imaging conditions were as follow: 500ms exposure and 1000ms acquisition time with a laser power of 0.16 mW.

5.5.6. Image acquisition by Airyscan

Single plane images were acquired on a Zeiss LSM 880 point scanning confocal microscope using the Airyscan detector, a 40X/1.2 NA water immersion objective (Zeiss) and 561 nm laser line. The Zeiss Zen 2.3 (black edition) software was used to control the microscope, adjust spectral detection for the emission of JF_{549}-HaloTag fluorophores and for processing of the Airyscan raw images. Zeiss Definite Focus was used at each time point for time-lapse imaging. The microscope was set to line scanning mode with the pinhole set at 1 Airy Unit. The detector gain was set to 600. 120 images were acquired every 5 seconds for a total acquisition time of 5 minutes.

Images were converted to 16-bit TIFF stacks using Fiji. Drift correction was performed using the Correct 3D drift plugin (Parslow, Cardona, & Bryson-Richardson, 2014) within Fiji. After drift correction, flows were quantified using Particle Image Velocimetry (PIV) using the MatPIV implementation in MATLAB

(https://www.mn.uio.no/math/english/people/aca/jks/matpiv/). 16x16 pixel interrogation windows with 75% overlap were used to calculate the PIV flow fields. The raw velocity fields were filtered using snrfilt, pkhfilt, and globfilt. Jitter and remaining drift were corrected using the displacements of the centroid of the nucleus to account for measurement noise.

5.5.7. ATP Survivability Assay

MCF7 was plated on acrylamide-based hydrogel covered with human placenta derived collagen type IV (Sigma Aldrich) in a 24-well black plate with clear bottom at $1x10^5$ per well. The cell was cultured overnight in the regular growth media then treated with Tamoxifen (Sigma Aldrich) at final concentration of 50, 25, 12.5, 6.25, and 3.125 uM for 24 hours in the regular growth media. The normalized number of viable cells were estimated by ATP-based cell viability test CellTiter-Glo 2.0 (Promega) according to the manufacturer's protocol. The resulting luminescent signal was measured by VictorX plate reader (PerkinElmer) and normalized by the tamoxifen free vehicle control group. The experiment was done in biological triplicates.

5.5.8. RNA seq for (MCF7 and MCF7LTED)

Cells were trypsinized with phenol red free trypsin and total RNA was extracted from cultured cells with Nucleospin RNA kit (Macherey Nagel) in biological triplicates. Samples were sequenced on HISeq3000/4000 using Illumina TruSeq Stranded Total RNA Library Prep and paired-end sequencing. The reads of the samples were trimmed for adapters and low-quality bases using

Trimmomatic software before alignment with the reference genome (hg19) and the annotated transcripts using STAR. Differential expression analysis on RNA-seq was performed by using DESeq2.

5.5.9. ATAC seq

ATAC was performed according to OPTI-ATAC (PMID:30679562). Briefly cells were trypsinized from cell culture. 50,000 cells were harvested and lysed in lysis buffer (10 mM PIPES pH 6.8, 100 mM NaCl, 300 mM sucrose, 3 mM $MgCl_2$, 0.1% TritonX-100) for 5 min incubation on ice and 5 min spin. Nuclei were resuspended in 15ul water, 25ul TD buffer and 10ul TDE1 enzyme from the Nextera DNA Prep kit (Illumina) and incubated for 30min at 37°C in a shaker at 1000rpm. Fragmented DNA was extracted with MinElute PCR Purification kit (Qiagen) and amplified by using NEBNext High-Fidelity 2X PCR Master Mix (New England BioLabs) and primers published in (Buenrostro et al., 2013). Size selection was performed using SPRIselect (Beckman Coulter) to remove <150-bp and >1000-bp fragments according to the manufacturer's instructions and was verified by Agilent TapeStation system (Agilent Technologies).

The samples were subjected to paired-end sequencing using 2x75-bp reads by the Illumina NextSeq High V2 at the National Cancer Institute Sequencing Facility, (Frederick, MD) or the Illumina HiSeq X at the Psomagen Inc. (Rockville, MD). The reads were trimmed in silico to remove adapter sequences, low-quality reads, and 50 bp length using Trimmomatic 0.30 software. The reads were aligned

to human reference genome hg38 using Bowtie2 alignment tool. Mitochondrial reads were filtered for the subsequent analyses.

5.5.10. ChIP and ATAC peak calling and heatmaps

Peak calling was performed using HOMER (Heinz et al., 2010). Peaks in each data set were called using findPeaks with style factor for transcription factors. The mergepeaks command was used to find common/specific peaks between conditions. Heatmaps and aggregate plots were generated with the deepTools computeMatrix (reference-point, +/- 500 bp around the peak center), and plotHeatmap using the normalized bigWig files. The bigWig files were generated by deepTools using the above described bam files and normalized for reads per genomic content per bin (5kb) (1x normalization scaled to the average coverage across the genome).

Venn diagrams were generated to demonstrate the numbers of peaks shared by 2- or 3- conditions using the venneuler package of the statistical software R. For the replicates, we obtained the merged peak sets from pooled data as well as the sets of peaks from each individual replicate. We retained those peaks from pooled data that have at least 50% overlap in each replicate. The number of overlapped or unique peaks was determined using the software BEDtools suite.

5.5.11. Immunofluorescence

Cells were fixed with 4% paraformaldehyde in PBS for 10 mins at room temperature, washed three times with PBS, then permeabilized with PBS

containing 0.5% TritonX-100 for 10 min at room temperature. Cells were washed three times with PBS, then incubated at room temperature with PBS with 3% BSA for 20 min, followed by incubation with rabbit polyclonal anti-human ER α (Santa Cruz Biotechnology, sc-543,1: 300) in 3% BSA-0.05% TritonX-100 in PBS for 60min at room temperature. Cells were incubated with Alexa Fluor 488 goat anti-rabbit IgG (Invitrogen, 1:500) for 30 min at room temperature. DNA was counterstained with DAPI, then washed with PBS three times. Images were captured in CellVoyager CV7000 (Yokogawa Electric Corp).

5.5.12. qRT-PCR

Total RNA was extracted from cultured cells with RNeasy Plus Mini kit (Qiagen) and reverse transcription was carried out with iScript cDNA Synthesis Kit (BioRad). PCR was performed with iQ SYBR Green Supermix (BioRad) by BioRad CFX96 Real Time PCR Thermal Cycler (BioRad). Values were normalized to glyceraldehyde-3-phosphate dehydrogenase (GAPDH) mRNA expression before calculating relative fold changes. Primer sequences are listed in Table 5-2.

5.5.13. ER chromatin immunoprecipitation (ChIP) and ChIP-seq.

MCF-7 cells were left untreated or treated with 100 nM of E2 (Sigma) for 1 h. Cells were trypsinized with phenol red free trypsin and cross-linked in suspension with 1% paraformaldehyde (5 min), the chromatin was sonicated (Bioruptor, Diagenode) to an average DNA length of 200–500 bp. For immunoprecipitation, 600 μg of chromatin was incubated with anti-ER antibody (Abcam #ab108398) coupled onto Protein A/G magnetic beads (Millipore) with

rotation overnight at 4°C. After stringent washes, the antibody-bound chromatin fragments' cross-linking was reversed, and the remaining proteins digested. Immunoprecipitated DNA was extracted from the samples with phenol-chloroform-isoamyl alcohol and ethanol precipitation. ChIP-seq libraries were generated using Illumina TruSeq Chip Sample Prep Kit (Illumina # IP-202-1012) according to manufacturer's instructions.

5.5.14. Single-molecule tracking analysis

We used the custom-made software TrackRecord (Mazza et al., 2013) in MATLAB (The MathWorks, Inc., Natick, MA). Briefly, to analyze each time series, data were filtered using top-hat, Wiener, and Gaussian filters. A region of interest (ROI) was defined to encompass the nucleus (and the array using the GFP-NF1 fluorescence when necessary), then particles were detected, fitted to two dimensional gaussian function for subpixel localization, and finally tracked using a nearest neighbor algorithm (D.M. Presman et al., 2017). The tracking parameters were as follows: window size for particle detection 7 pixels, maximum frame to frame displacement of 2 pixels, shortest track 2 frames, and gaps to close 1.

Tracks were classified into distinct diffusive states using perturbation-Expectation Maximization [pEM v2, (Koo & Mochrie, 2016)]. Tracks were split into 7 frame segments and the pEM classification analysis was performed on the set of all these track segments. The minimum number of states for the system to converge to was set at 2 and the maximum at 7. If the optimal number of states that the analysis converged to was 7, the algorithm was rerun with a higher number

of maximum states. The number of reinitializations was set to 30 with 50 perturbation trials. The maximum number of iterations was 10000 with a convergence criterion for the change of log-likelihood of 10^{-7}. The number of features for the covariance matrix was set to 5. Motion blur coefficient was calculated as $\frac{1}{6}\frac{\Delta e}{\Delta t}$, where Δe corresponds to the exposure time and Δt the acquisition interval.

5.5.15. Mean Squared Displacement calculation (MSD)

We calculate the weighted ensemble averaged MSD for a particular state by using the posterior probability of a track to belong to a state I as the weight function:

$$WMSD_i(\tau) = \frac{\sum_{j=1}^{N} P_j(i).MSD_j\,(\tau)}{N}$$

Where $WMSD_i(\tau)$ is the weighted ensemble average MSD for state i, and N is the total number of tracks.

Standard error was calculated as $\frac{\sigma_w}{\sqrt{N_f}}$, where σ_w corresponds to the weighted standard deviation and N_f the number of degrees of freedom in the weighted mean. Weights correspond to the number of points averaged to generate the mean square displacement value at the given time lag.

Survival distribution calculation

Survival distributions were calculated from particle tracks as in (Garcia et al., 2021b) using the Kaplan-Meier estimate. The 95% confidence interval was

estimated using Greenwood's Formula. HaloTag-H2B data was acquired with identical imaging conditions as the transcription factors for the different acquisition conditions. The survival distribution of this was fitted to a double and triple exponential model to extract the photobleaching rate, and model selection was used to determine the best predictive model (See Fitting and model selection). The survival distribution of the different transcription factors was corrected for photobleaching as follows ($S(t) = e^{\gamma t} S_E(t)$, where S(t) corresponds to the survival distribution after photobleaching correction, $S_E(t)$ the empirical survival distribution and γ the photobleaching rate).

5.5.16. Weighted survival distribution calculation

After pEMV analysis, subtracks are classified into different diffusive states using pEM and states that are not representative are removed as explained above. Suppose a track k is divided into n subtracks of length 7. After the pEM analysis, each subtrack i is assigned a posterior probability $P_k^i(j)$ to belong to a diffusive state j. Here, $i \in \{1,2,\dots,n\}, j \in \{1,2,\dots,m\}$, where m is the number of diffusive states to which pEM converges. From this state assignment, we would like to calculate the survival distribution of a particular diffusive state. To do so, we first calculate a posterior probability for the reconstructed track k:

$$P_k(j) = \frac{\sum_i P_k^i(j)}{\sum_{i,j} P_k^i(j)}$$

We will use these probabilities to calculate a weight function for each bin of the dwell time histogram. All the dwell times can be distributed into N bins such

that each bin Ω_i contains n_i tracks. Mathematically, $\Omega_i \equiv \{t \in \mathbb{R} | (i-1)\Delta t \leq t < i\Delta t\}$, where Δt is the acquisition interval.

We can define a weight function W_i^j for each of the bins i, per state j as

$$W_i^j = \sum_{k=1}^{n_i} P_k(j)$$

Previously survival distributions were calculated by assigning a unit weight to each observed bound track. Instead, we can define the statistic

$$\hat{p}_i^j = \frac{W_i^j}{n}$$

as the unbiased estimator (i.e. $E\left[p_i^j\right] = \hat{p}_i^j$) of the weighted survival distribution.

Let $P_k(j)$ be the probability assigned to a track k as defined previously. The proportion ($M(i)$) of a particular diffusive state i is given by:

$$M(i) = \frac{\sum_k P_k(i)}{\sum_i \sum_k P_k(i)}$$

All fits performed to the data were implemented with the nonlinear least square method using bisquare weights due to the noise on the tail of the survival distribution.

5.5.17. Quantification and Statistical Analysis

For statistical analysis, all measured quantities are reported as ensemble averages with standard error and number of observations. At least three biological replicates of SMT

experiments were performed for each condition. Two sample K-S tests on the survival distribution were performed between replicates to confirm statistical reproducibility. At least 20 cells were imaged per SMT replicate for each condition.

5.6. Tables

Stiffness (KPA)	Acrylamide 40% (ul)	BIS-Acrylamide 2% (ul)	10% APS	TEMED	HEPES(ul)	Polymerization Time (min)
0.8	56.25	50	5	1.5	887.25	50
12	187	100	5	1.5	707	20
100	300	300	5	1.5	393.5	10

Table 5-1 Polyacrylamide Fabrication
Chemical concentration for polyacrylamide gel fabrication of three different stiffness (0.8KPa, 12KPa and 100KPa).

Name		Sequence
GAPDH	forward	ACACCCACTCCTCCACCTTT
GAPDH	reverse	TAGCCAAATTCGTTGTCATACC
ESR1	forward	TATGTGTCCAGCCACCAACC
ESR1	reverse	TCGGTCTTTTCGTATCCCACC

Table 5-2 Primers Sequence
Primers used for RT-qPCR.

Chapter 6. Conclusions and Future Directions

6.1. Conclusions

Transcriptional regulation is fundamental to all cellular processes, and dysregulation of transcriptional control underlies developmental disorders and diseases such as cancer and Alzheimer's. Owing to its importance, transcription is exquisitely controlled by the collective action of multiple proteins that work together in the complex nuclear microenvironment. The key initial step is the binding of sequence-specific TFs to the promoter of a gene or enhancer sequences thereby tuning transcriptional output, chromatin topology and phenotype. Understanding how TFs interact with chromatin, other proteins and among themselves is essential our basic understanding of biology and the development of targeted therapeutics. In this thesis, we were able to advance the knowledge of TF dynamics and transcriptional regulation with far reaching implications for our understanding of molecular biology and physiology.

The current paradigm of TF binding to chromatin assumes discrete and distinguishable binding times for interactions with specific and non-specific sites. In Chapter 3, we developed phenomenological models of TF binding, in which the interactions of TFs with chromatin have a broad distribution of binding affinities. An implication of this model is that the resulting dwell times (i.e. measurements of TF-chromatin interactions) will be power-law distributed. This model is compatible with current genomic and *in-vitr*o data of TF binding to the rich repertoire of binding sequences and the complex properties of the nuclear microenvironment. We

imaged multiple families of TFs (nuclear receptors, pioneer factors, chromatin remodelers and architectural proteins) at the single molecule level as they moved in the nucleus and bound to chromatin. We developed a new pipeline to analyze SMT binding times which reveals a power-law distribution of binding times of TFs. Using this pipeline, we showed that a large number of TFs show power law distributed interaction times with chromatin, implying that the current view of biexponential binding times is incomplete

Motivated by our observations of power law distributed TF-chromatin interactions, in Chapter 4 we explored TF mobility and how this is modulated by their interactions with chromatin and the local nuclear microenvironment. Characterization of single molecule trajectories revealed that TFs can either freely diffuse within the nucleus, bind to chromatin or are confined within nuclear hubs due to protein-protein interactions. In agreement with the models introduced in Chapter 3, mutants and native TFs with a narrower distribution of binding motifs, as assessed by CHIP-seq experiments, exhibit exponentially distributed chromatin binding times. Intriguingly, the dwell times of confined TFs were also power-law distributed with confinement times significantly longer than chromatin binding times. We proposed a model in which TF-TF interactions mediated by intrinsically disordered regions locally increase protein concentrations in the vicinity of chromatin, thereby decreasing the “search time” for TFs to find their target binding sites and promoting transcription.

With our robust analysis for TF dynamics in Chapter 3 and 4, we explored TF mobility and transcriptional regulation in the context of breast cancer

progression in Chapter 5. We discovered a novel unliganded ER regulatory function dependent on microenvironment stiffness. On substrates of physiological stiffness (0.8KPa and 12KPa) liganded estrogen receptor has significantly stronger transcriptional regulatory activity in comparison with stiffer substrates, which is correlated with tamoxifen resistance. Surprisingly, in soft substrates, unliganded ER binds broadly to chromatin and acts antagonistic to transcriptional regulation by liganded ER. Our findings put to rest the debate on whether unliganded ER has transcriptional activity. The relatively modest changes in chromatin accessibility in response to substrate stiffness argues for other regulatory mechanisms of mechanosensing in the context of breast cancer.

In addition to controlling the transcriptional activity of ER, substrate stiffness also affects its mobility and binding to chromatin. SMT experiments revealed that ER has binding times on the order of minutes in cells grown on substrates of physiological stiffness compared to the shorter binding times observed in cells grown on stiffer substrates (100 KPa and Glass). This finding implies that current models of rapid TF-chromatin interactions must be revised. We also showed that changes in ER binding and mobility are accompanied by changes in chromatin mobility with highest mobility at lower stiffness which correlates with increased cell stemness. The studies detailed in Chapter 5 revealed novel ER reprogramming due to substrate stiffness with potential applications in targeted treatments of ER-positive breast cancer.

6.2. Future Directions

6.2.1. 1. In search for specific response elements

In Chapter 4, we showed that GR exhibits two fast diffusing states in the nucleus, with one state characterized by a different diffusion coefficient than observed for GR in vitro. We hypothesize this diffusion state is related to one dimensional diffusion and hoping on chromatin, which has been postulated to serve as a search mechanism for TFs to find and bind to specific response elements. Theoretical studies have shown that 3D diffusion alone is incompatible with rates of transcription found *in vivo* and different mechanisms of facilitated diffusion have been suggested, but experimental evidence for these is lacking (Benichou et al., 2011; Kampmann, 2005). SMT experiments with higher temporal and spatial resolution coupled with improved photostable dyes will allow us to probe the search mechanism of TFs in eukaryotic cells.

One technique that could be useful in this regard is minflux nanoscopy, which allows for nm resolution microscopy with significant reduction in photons needed for localization (Gwosch et al., 2020). Moreover, the complexity of 3D tracking and tracking errors are reduced since tracking is built in the design of the microscope. This type of microscopy coupled with Bayesian non-parametric methods (Hjort, 2010; Sgouralis & Presse, 2017) to discern different types of TF mobility could be capable of revealing the search mechanisms utilized by TFs.

6.2.2. 2. Confinement, binding and transcriptional bursting

We showed in Chapter 4 that TFs are confined in nuclear hubs through protein-protein interactions. We proposed a model in which these hubs increase the local concentration of TFs thereby facilitating binding and promoting transcription, but experiments are needed to validate this. Recent experiments have shown that different complexes involved in transcription are in hubs formed by LLPS but the mechanisms that dictate their formation and their role in transcriptional regulation remain unknown (Alberti, Gladfelter, & Mittag, 2019; Hyman et al., 2014). Our proposed model predicts an increase in transcriptional bursting frequency of genes associated with transcriptional hubs due to the decreased searching times of transcription factors and transcriptional complexes. Other outstanding question correspond to the formation and dynamics of these protein hubs - How and when are they nucleated? What is their lifetime? Do they coalesce with other proteins involved in transcription and chromatin remodeling?

Different experimental techniques could be used to probe the formation and role of protein hubs in transcriptional control. A recent study used orbital tracking to correlate transcriptional bursting of a reporter gene and showed that longer TF binding times were correlated with longer ON times (Stavreva et al., 2019). Using a similar approach but labeling endogenous genes and real time PALM of TFs (Cho et al., 2016) might allow us to characterize the role of protein hubs in transcriptional regulation. Another approach is to use pair correlation PALM to explore the formation of GR hubs and PolII recruitment followed by transcriptional

activation. This will provide information about the dynamic signature of hub formation and their role in transcriptional regulation (Cho et al., 2018).

6.2.3. 3. Dynamic Assisted Loading

The dogma of TFs binding to chromatin has evolved rapidly in the past several years. The static view of TFs binding proposes continuous binding on the order of minutes to hours from activation to the completion of transcription (Spitz & Furlong, 2012). Recent advances in microscopy have allowed the observation of real time dynamics of transcriptional bursting and TF binding, shifting the dogma to a dynamic view where TFs bind transiently to chromatin and transcription happens in bursts on the order of minutes (Tunnacliffe & Chubb, 2020). TFs binding time was accepted to be on the order of seconds (Mazza et al., 2013) leading to the concept of “Dynamic Assisted Loading” (Voss et al., 2011) where TFs rapidly exchange at enhancers with other proteins that promote binding instead of competition even though they bind to the same genomic site.

However, in Chapters 3, 4 and 5 we showed that for different TFs and different experimental conditions, TF binding times are broadly distributed in the order of seconds to minutes. The latter implies an urgent need to characterize the dynamics of TFs interactions to reassess assisted loading or to define a new model of TFs interaction that includes the complexity of the nuclear microenvironment, chromatin topology and the broad distribution of TFs dwell times. In the dynamic assisted loading model, TFs interact transiently with closed chromatin recruiting remodeling factors which facilitate transient open chromatin states. Studying the

thermodynamic properties between these states will provide insightful information in the properties of the dynamics. In particular, for a system in thermal equilibrium, the transition rates of realizable microstates are balanced. In other words, they obey detailed balance where the rate out of and into pairwise microstates is zero (Van Kampen, 1992). Searching for signatures of broken detailed balance in TF binding adjusted for power-law distributed dwell times (Biddle et al., 2019) allows for identification of the asymmetry in protein binding that potentiates pairwise TFs binding. Another approach is to use orbital tracking or pair correlation PALM between pairs of TFs and Pol II loading to characterize the kinetics of interactions and transcriptional initiation. These experiments need to be done at different endogenous sites with diverse chromatin microenvironments to characterize the role of the nuclear microenvironment and TF interactions in gene regulation.

4. LINC-king mechanical information

There is mounting evidence that transcriptional regulation of cell fate in many contexts is influenced by the mechanical properties of the cell microenvironment (Engler et al., 2006; Handorf et al., 2015). In Chapter 5, we found that TF mobility, binding and action is altered by the stiffness of the substrate. This alteration of TF dynamics is accompanied by changes in cellular and nuclear shape, chromatin mobility and to a lesser extent, chromatin accessibility. However, how mechanical stimuli are transmitted to the nucleus and affect transcriptional regulation is not fully understood. Another outstanding question in the field is the amount of information in bits that cells can process about

the stiffness of the environment (Gregor, Tank, Wieschaus, & Bialek, 2007; Walczak, Tkacik, & Bialek, 2010).

In Chapter 5, we identified sets of genes that are influenced by the stiffness of the environment and dependent on ER signaling. Using smRNA-FISH or Hybridization Chain Reaction (HCR) (Dirks & Pierce, 2004) as a readout of gene expression of cells on different stiffness, the upper bound of the mutual information between gene expression and stiffness can be calculated to quantify information transmission of mechanical stimuli. The latter, coupled with degron systems (S. Li, Prasanna, Salo, Vattulainen, & Ikonen, 2019) targeting members of the LINC complex and transcriptional coactivators such YAP and TAZ that are involved in mechanotransduction (Bouzid et al., 2019), might provide direct evidence of the mechanism involved in the mechanical regulation of gene expression.

6.2.4. 5. Antagonizing liganded Estrogen Receptor

There has long been a debate about the transcriptional activity and clinical relevance of unliganded Estrogen Receptor and its mode of action (Stellato et al., 2016). Different experiments have pointed to changes in ER phosphorylation and cofactor interactions as regulators of unliganded ER activity (Caizzi et al., 2014b; de Leeuw, Neefjes, & Michalides, 2011), but these experiments have been carried out with conditional knockdowns or pharmacological treatments with unknown off-target effects. In Chapter 5, we showed antagonistic action of unliganded ER with respect to liganded ER and its dependence on substrate stiffness. This provides a

physiologically relevant framework to study unliganded ER biology and its role in breast cancer progression.

Follow up experiments to quantify changes of phosphorylation in ER AF-1 domain (ligand independent activity) on substrates of different stiffness to correlate unliganded ER activity with the changes in the transcriptome will shed light on these questions. Immunoprecipitation of ER followed by shotgun spectrometry (Zhang, Fonslow, Shan, Baek, & Yates, 2013) of cells on substrates with different stiffness will allow the characterization of changes in interacting partners of unliganded ER that are involved in potentiating binding and mechanical response. Identification of such partners followed by genomic and cell biological studies to quantify assisted loading will allow for the development of specific targeted strategies in the treatment of ER+ breast cancer.

6.2.5. 6. Mechanical regulation of GR activity in chromatin accessibility

In Chapter 5, we carried out extensive studies of the changes in ER activity due to substrate stiffness. ER reprogramming is coupled to the ability of the receptor to bind to chromatin in its unliganded form. Other hormone receptors such as GR might have a different response to stiffness since GR is bound to chaperones in the cytoplasm in the absence of ligand and after hormone treatment, the receptor translocates into the nucleus. Similar techniques used in Chapter 5 can be applied to understand the biology of GR in the context of different cellular microenvironments. This can potentially help to reveal mechanisms of TF regulation in physiological settings.

We performed preliminary studies on GR action on accessibility by performing ATAC-seq after 100 nM Dex treatment on substrates of different stiffness (0.8KP, 12KPa, 100KPa and Glass). Surprisingly, we found dramatic changes in GR's ability to remodel chromatin in a stiffness dependent manner (**Figure** 6-1). GR plays a repressive role on soft substrates (decreased chromatin accessibility) compared to an activating role on stiffer substrates (increased chromatin accessibility). Interestingly, GR activity on glass has been widely reported to be activated by increasing chromatin accessibility genome-wide after hormone treatment. Further experiments are needed to assess these changes in GR remodeling action on chromatin. If confirmed, this will have important clinical consequences due to the importance of GR in inflammation and disease (Baschant, Culemann, & Tuckermann, 2013; Baschant et al., 2011).

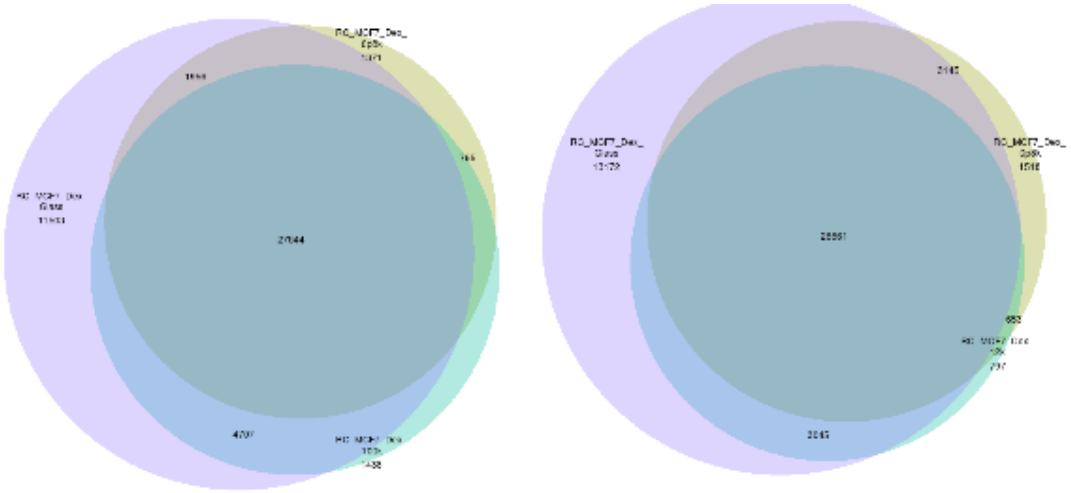


Figure 6-1 GR Chromatin Accessibility Effects are Modulated by Substrate Stiffness

Venn Diagram for ATAC-seq between 1-hour 100nM Dex treatment conditions and vehicle for different substrate stiffness. Accessibility increases after Dex treatment in a stiffness dependent manner.

I expect the work presented in this thesis to serve as a cornerstone in future studies of the biophysics properties of transcription factor mobility, their interactions with chromatin and the role of molecular dynamics in gene expression. The different analysis developed and presented here can be applied to study biophysical properties of single molecules in different systems. In this thesis, different models of TF mobility and gene regulation have emerged with far reaching consequences in our basic understanding of biology and with important therapeutic implications.

Appendix A.

Theoretical Models for TF Survival Distribution

1 Theoretical models for TF survival distribution

In an SMT experiment, the protein of interest is tagged with a fluorescent probe and imaged. Binding events are then associated with stationary particles in the focal plane. The final experimental information that can be recovered is the time that a protein can be detected in the imaging volume before it bleaches, or moves out of the focal plane. From these observations, one can calculate a dwell time for transcription factors (TFs) which is defined as the time interval between a single molecule transitioning from a diffusive state to a bound state and its subsequent unbinding from DNA and return to the diffusive state. The dwell time distribution is obtained by calculating the ensemble distribution of bound times for a specific TF in different cells in the experiment after photobleaching correction (see Methods).The survival distribution is then calculated as 1-CDF, where CDF is the empirical cumulative distribution function of dwell times.

1.1 Absorbing Boundary State Method

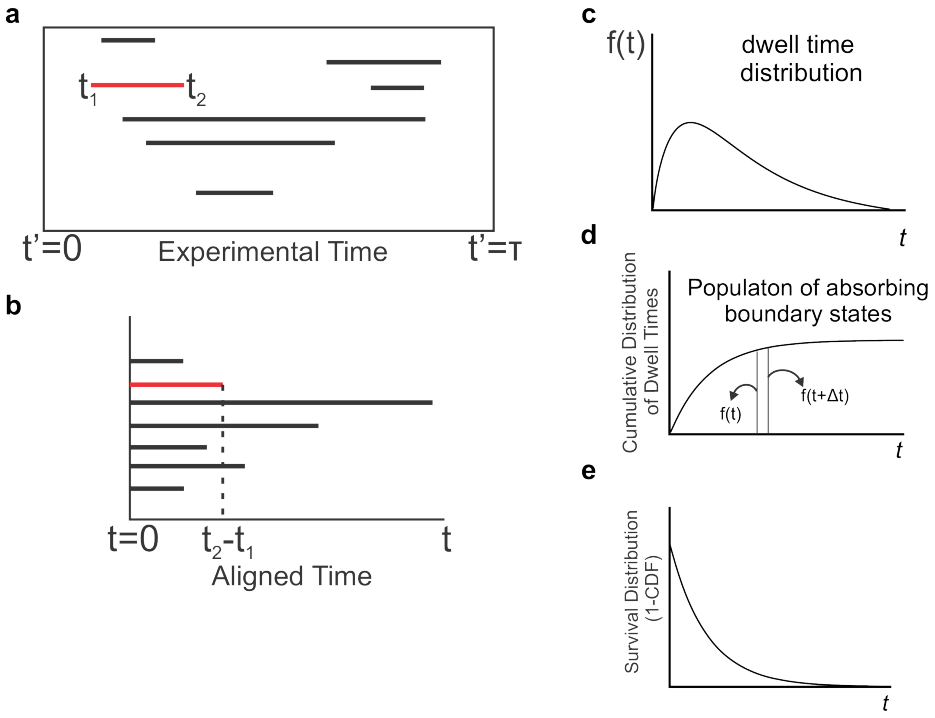
Calculation of dwell time distributions is a first-passage time problem in stochastic analysis and these distributions have been widely used to characterize kinetic properties of molecular motors and ion channels (1). In cases involving simple kinetic schemes, the dwell time distributions can be calculated analytically but for more complex schemes, a number of methods have been utilized. One particularly powerful approach is to assign one or more states to act as an absorbing boundary and then solve the associated first-order kinetic equations to obtain dwell time distributions. We assume that the diffusive (unbound) state corresponds to an absorbing boundary state, since the measurement ends with such transitions, because the particle either photobleaches, disappears from the focal plane or begins diffusing; any rebinding of the TF is considered as an independent event. This assumption implies that the population of particles in the absorbing boundary state increases with time. At the end of every experimental measurement, all the observed TFs transition to the absorbing state since the experiments are continued until most particles are bleached.

For a general process, a TF can be found in any state i such as diffusing around the nucleus, confined in a microenvironment, bound to a particular specific or non specific site of the DNA. When a TF transitions to a diffusive state, it cannot be observed experimentally and this state plays the role of an absorbing boundary state. We observe the system over a time interval from $t' = 0$ to $t' = \tau$, during which individual TFs may undergo transitions between different states $i \in \{1, ..., n\}$. When a transcription factor is in a "bound" state, it can be observed experimentally as a trace as displayed in Supplementary Note Fig. 1a. Each TF in any of these bound states will be experimentally recorded from a certain time interval t_1 to t_2 with $t_2 - t_1 < \tau$. t_1 corresponds to the time when the transcription factor transitions to a "bound" state and t_2 corresponds to the time when the TF enters an absorbing state (diffusion in the nucleus). All the traces will be shifted by a time t_1 to a new aligned time $t = t' - t_1$ so that all the TFs begin in a bound state at $t = 0$ (Supplementary Note Fig. 1b). During the experimental time τ, a finite number of TFs (N, equal to the number of traces) will be observed. When a TF transitions to a diffusive state, it cannot longer be observed experimentally and this state plays the role of an absorbing boundary state.

Let $p_i(t)$ correspond to the probability of being in state i at time t.$\sum_{i \in boundaries} p_i(t)$ corresponds to the population of all absorbing states. To calculate the number of unbinding events over a certain time interval ($f(t)$ dwell-time distribution, Supplementary Note Fig. 1c - adapted from (1)), we take the time derivative of this population,

$$f(t) = \frac{d}{dt} \sum_{i \in boundaries} p_i(t) \tag{1}$$

$f(t)$ can be visualized as the probability distribution of experimental track lengths of TFs entering a bound state and evolving independently from a registered time $t = 0$, until they transition to an absorbing boundary state, at which time they leave the bound state (2). The cumulative distribution



Supplementary Note Fig. 1: Survival distribution calculation. (**a**) Experimentally, slow events are seeing as traces in a kymograph. (**b**) These traces are aligned on a new time t and the distribution of lengths corresponds to the dwell time distribution $f(t)$ (**c**) . Shown in red a sample trace before and after time alignment. (**d**) The CDF of $f(t)$ corresponds to the normalized population of an absorbing boundary state. (**e**) (1-CDF) corresponds to the survival distribution $\hat{D}$.

of $f(t)$ is calculated (Supplementary Note Fig. 1d) and $1 - CDF$ corresponds to the survival distribution ($\hat{D}$, Supplementary Note Fig. 1e) .

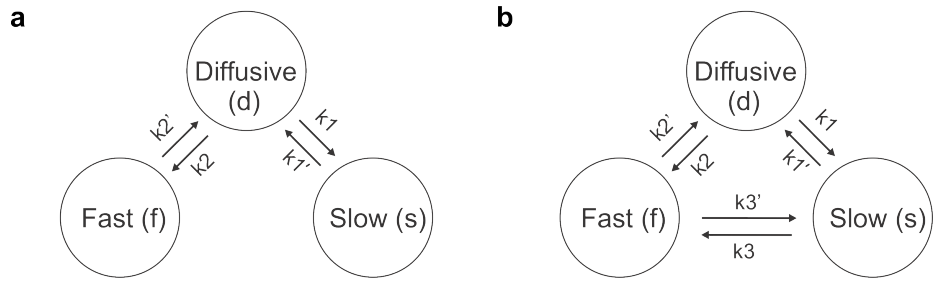
1.2 Revised Bi-Exponential Model

We consider an idealized system containing a fixed total number of TFs, each of which can be found in one of three states: Slow (s), Fast (f) and Diffusive (d) (Supplementary Note Fig. 2a) In this model, transitions between s and f states are forbidden. When a transcription factor is either in state s or f, it can be observed experimentally as a trace. When a TF transitions to the state d, its experimental observation will stop and this state plays the role of an absorbing boundary state.

Let $S(t)$ and $F(t)$ denote the number of TFs in the state s and f at the aligned time t, respectively. These functions will decrease monotonically to 0 from the initial values S_0 and F_0, $S_0 + F_0 = N$, which correspond to the total number of TFs in the s and f state during the experiment, respectively. To obtain the survival distribution, the probability that a particle stayed bound (either in state s or f) for a time t or longer, the population of absorbing boundary states needs to be calculated. The latter (population of absorbing boundary states), corresponds to transitions from s to d or from f to d, which denotes the following events:

$$s \xrightarrow{k_1'} d \quad f \xrightarrow{k_2'} d \tag{2}$$

Let $D(t)$ denote the number of experimentally observed transcription factors that transitioned from any bound state (s or f) to the diffusive state up to an aligned time t, with $D(0) = 0$. This



Supplementary Note Fig. 2: Double Exponential and Kinetic Models. (**a**) Schematic representation of the double exponential model. Transitions between specific (s) and non-specific binding (f) are forbidden. (**b**) Kinetic model allows for transitions between non-specific and specific binding.

population accumulates over time and the dwell time distribution corresponds to the time derivative of this population. Thus,

$$\frac{dS}{dt} = -k_1' S \quad \frac{dF}{dt} = -k_2' F \tag{3}$$

$$\frac{dD}{dt} = k_1' S + k_2' F \tag{4}$$

Then,

$$S = S_0 e^{-k_1' t}, \quad F = F_0 e^{-k_2' t} \tag{5}$$

$$\frac{dD}{dt} = S_0 k_1' e^{-k_1' t} + F_0 k_2' e^{-k_2' t} \tag{6}$$

where $\frac{1}{N}\frac{dD}{dt}$ corresponds to the dwell time distribution ($f(t)$). The dwell time distribution depends on the initial populations of S and F. These correspond to the relative populations entering a cycle in the dwell time ($D \to S$ or $D \to F$), the proportion of traces either in state s or f. Traces in state s or f are born randomly with proportions $\frac{k_1}{k_1+k_2}$ and $\frac{k_2}{k_1+k_2}$ respectively (Supplementary Note Fig 2A).

Then, $S_0 = \frac{Nk_1}{k_1+k_2}$ and $F_0 = \frac{Nk_2}{k_1+k_2}$ which corresponds to the initial values of $S(t)$ and $F(t)$.

Finally, the dwell time distribution is given by (Supplementary Note Fig 1c):

$$\boxed{f(t) = \frac{1}{N}\frac{dD}{dt} = \frac{1}{k_1+k_2}\left(k_1 k_1' e^{-k_1' t} + k_2 k_2' e^{-k_2' t}\right)} \tag{7}$$

The Survival distribution ($\hat{D} = \int_t^\infty \frac{dD}{du} du$) is given by (Supplementary Note Fig 1e):

$$\boxed{\hat{D} = \frac{1}{k_1+k_2}\left(k_1 e^{-k_1' t} + k_2 e^{-k_2' t}\right)} \tag{8}$$

$\frac{1}{k_1'}$ and $\frac{1}{k_2'}$ correspond to the expected residence time of the slow and fast component respectively. After calculating the survival distribution, we are interested in finding the proportions of TFs in state s and f in steady state.

At steady state, the slow and fast component populations ($\bar{S}$ and $\bar{F}$) are calculated as follows (Supplementary Note Fig 2a):

$$\frac{d\bar{S}}{dt'} = k_1\bar{D} - k_1'\bar{S} = 0 \rightarrow \bar{S} = \frac{k_1}{k_1'}\bar{D} \tag{9}$$

$$\frac{d\bar{F}}{dt'} = k_2\bar{D} - k_2'\bar{F} = 0 \rightarrow \bar{F} = \frac{k_2}{k_{2'}}\bar{D} \tag{10}$$

The slow and fast component proportions ($\hat{S}$ and $\hat{F}$) are given by:

$$\boxed{\hat{S} = \frac{k_1k_2'}{k_1k_2' + k_1'k_2};\ \ \hat{F} = \frac{k_1'k_2}{k_1k_2' + k_1'k_2}} \tag{11}$$

Previously (3–7), the survival distribution was phenomenologically fitted to $\hat{D} = \left(f_1e^{-k_1't} + (1-f_1)e^{-k_2't}\right)$ and f_1, $(1-f_1)$ were interpreted as the slow and fast component proportions at steady state contrary as the values found in equation 11. The derivation of the survival distribution shows that $f_1 = \frac{k_1}{k_1+k_2}$ represents the proportion of traces in the s state during the experimental observation and not their steady state populations.

1.3 Kinetic Model

We next extended the bi-exponential model so that transitions between the slow and fast components are allowed but indistinguishable. S, F and D defined as in 1.2 (Supplementary Note Fig. 2b). If a TF is in a non-specific bound state (f, diffusing or hopping on the DNA), and it binds to a specific site (s), this transition cannot be observed due to the spatial resolution limit (∼30nm, ∼3 nucleosomes). The particle will appear bound regardless of the number of transitions between the fast and the slow component inside the resolution limited volume. The diffusive state corresponds to an absorbing boundary state (Supplementary Note Fig. 2b), as before. Then, the dwell time distribution $\frac{dD}{dt}$ can be calculated as follows:

$$d \xleftarrow{k_1'} s \underset{k_3'}{\overset{k_3}{\rightleftharpoons}} f \xrightarrow{k_2'} d \tag{12}$$

$$\frac{dS}{dt} = -(k_1' + k_3)S + k_3'F \quad \frac{dF}{dt} = -(k_3' + k_2')F + k_3S \tag{13}$$

$$\frac{dD}{dt} = k_1'S + k_2'F \tag{14}$$

This can be solved in matrix form as:

$$\begin{pmatrix} \frac{dS}{dt} \\ \frac{dF}{dt} \end{pmatrix} = \begin{pmatrix} -(k_1' + k_3) & k_3' \\ k_3 & -(k_3' + k_2') \end{pmatrix} \begin{pmatrix} S \\ F \end{pmatrix} \tag{15}$$

The solution will be given by:

$$\begin{pmatrix} S \\ F \end{pmatrix} = C_1 \begin{pmatrix} \alpha_1 \\ \beta_1 \end{pmatrix} e^{\lambda_1 t} + C_2 \begin{pmatrix} \alpha_2 \\ \beta_2 \end{pmatrix} e^{\lambda_2 t} \tag{16}$$

Where λ_i are the eigenvalues and $\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix}$ the corresponding eigenvectors of the matrix in equation 15. C_1 and C_2 are calculated from the initial populations of the state s and f. These populations are given by $\frac{Nk_2}{k_1+k_2}$ and $\frac{Nk_1}{k_1+k_2}$ for states f and s respectively which corresponds to the number of TFs that entered a bound state through state s or f respectively. Then, the dwell time distribution is given by:

$$\boxed{f(t) = k_2' F(t) + k_1' S(t)}. \tag{17}$$

The survival distribution is calculated as:

$$\hat{D} = \int_t^\infty f(u)du$$

with S, F as defined in equation 16. The survival distribution corresponds then to a sum of two exponentials similar to the double exponential model but with exponential parameters (λ_1 and λ_2) that depend on the rate constants of the process. Therefore, double exponential fits to the experimental data cannot be directly used to extract the kinetic rates of the underlying process.

1.4 Diffusion and binding on a chain

A number of theoretical studies have posited that the process of TF binding to its cognate site on DNA involves a combination of bulk diffusion in the nucleus, 1-d sliding along the DNA, hopping and translocation, and have derived the search time subject to various conditions. In this extension of our basic model and to account for a multiplicity of fast bound states, we can model TF searching for a specific site on the DNA by assuming the DNA to be a circular chain composed of specific sites and non-specific sites (Supplementary Note Fig. 3a). The assumption in the following derivation is that the number of non-specific sites on the DNA is much larger than specific sites. This is a biologically reasonable assumption as only a few tens of thousands specific sites are bound by any TF according to genome wide studies (8, 9), while the entire genome contains millions of "other" potential chromatin binding sites. A TF binds stochastically to any site on the DNA and diffuses around the chain with a certain probability of unbinding from any state. The hopping rate from a non-specific site to another site (specific or non-specific), is given by k_1, the rate of hopping from a specific site to a non-specific or dissociation to the bulk is given by k_2 and the rate of dissociation from a non-specific site into the bulk is given by k_3.

Thus the empirically observed transitions of TF dissociations can result from two possibilities: (i) a TF begins at a random location in the chain and diffuses along the DNA on non-specific sites and unbinds to the bulk diffusive state (d) before finding the specific site; (ii) a TF diffuses along the DNA until reaches a specific site, and it unbinds from the DNA from either a specific site or a non-specific site.

1.4.1 A TF does not find a target

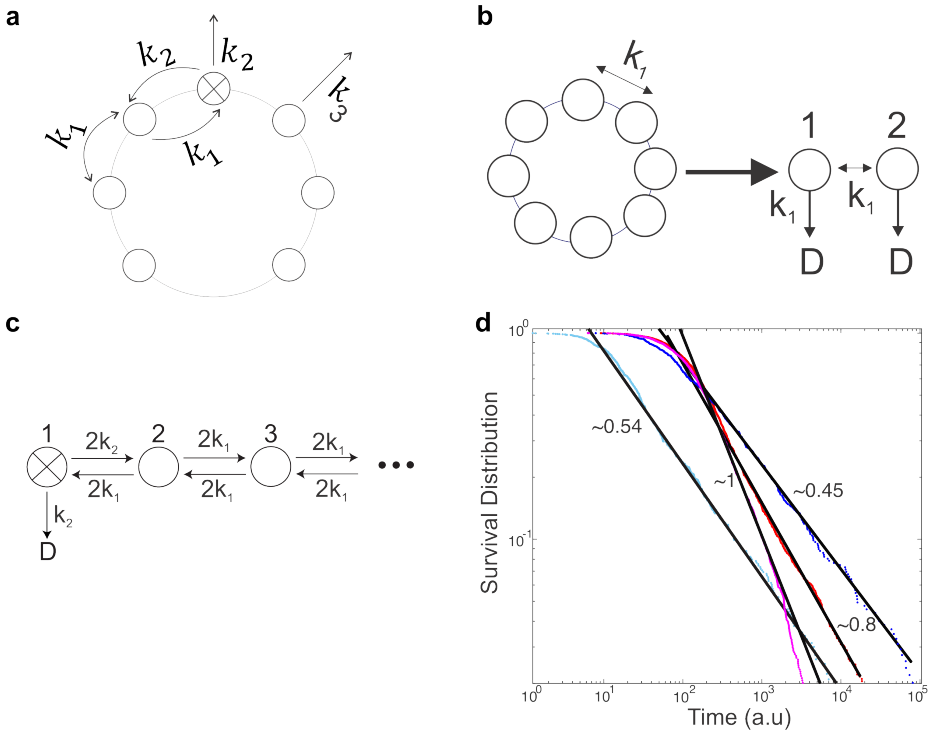
The case when a TF does not find a target is dynamically equivalent to a TF binding to a non-specific site with the possibility of diffusing to yet another non-specific site (Supplementary Note Fig. 3b). Due to the self-similarity of the chain, all non-specific sites are indistinguishable. Here, we assume that it is equally probable for a TF to jump between non-specific sites and unbind from a non-specific site, i.e $k_1 = k_3$. Applying the absorbing boundary method, the dwell time distribution can be easily found:

$$\frac{dD}{dt} = k_1 \left(P_1 + P_2\right) \tag{18}$$

where $P_1(t)$ and $P_2(t)$ correspond to the number of traces at time t that started in state 1 or 2 respectively.

$$\frac{dP_1}{dt} = -2k_1 P_1 + k_1 P_2, \;\; \frac{dP_2}{dt} = -2k_1 P_2 + k_1 P_1 \tag{19}$$

$$\rightarrow \begin{pmatrix} \frac{dP_1}{dt} \\ \frac{dP_2}{dt} \end{pmatrix} = k_1 \begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} P_1 \\ P_2 \end{pmatrix} \tag{20}$$



Supplementary Note Fig. 3: Diffusion and binding along a chain leads to power law behaviour of dwell times. (**a**) Schematic of TF diffusion on DNA represented as a circular chain that contains either non-specific sites (empty circles) or specific sites (crossed circle). (**b**) Equivalent state diagram for the condition when a TF does not find a specific site. The rates for diffusing along the chain and dissociating from the chain are set to k_1. (**c**) Equivalent state diagram for the case when the TF finds a specific site and unbinds from this site to the bulk. (**d**) Survival distribution of TF calculated using stochastic simulations. TFs bind uniformly around 20 sites away from the specific site. The simulations show different power law decays depending on different choice of parameters. For power-law of 0.45, $k_1 = 1$, $k_2 = 0.1$ and $k3 = 1$; for power-law of 0.54, $k_1 = 10$, $k_2 = 0.1$ and $k3 = 1$; for power-law of 0.8, $k_1 = 10$ $k_2 = 0.01$ and $k_3 = 0.01$; for power-law of 1, $k_1 = 10$ $k_2 = 0.01$ and $k_3 = 0.0001$.; for power-law 1.5, $k_1 = 1$ $k_2 = 0.01$ and $k_3 = 0.001$

$$\rightarrow \begin{pmatrix} P_1 \\ P_2 \end{pmatrix} = C_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} e^{-k_1 t} + C_2 \begin{pmatrix} -1 \\ 1 \end{pmatrix} e^{-3k_1 t} \tag{21}$$

After applying initial conditions for P_1 and P_2 (each non specific site is equally probable, $C_1 = \frac{1}{2}$ and $C_2 = 0$), we find that the normalized dwell time distribution is given by:

$$\boxed{\frac{d\tilde{D}}{dt} = k_1 e^{-k_1 t}}, \tag{22}$$

as expected for a Poisson process. The exponential term due to the dissociation of the TF from non-specific sites along the DNA dominates the internal diffusion of the TF along the chain.

1.4.2 A TF does find its target

Let $P_n(t)$ denote the probability of being in state n at time t. We will consider (without loss of generality) that $k_1 = 0.5$ and $\frac{k_2}{k_1} = \alpha$; k_1 will impose the time scale of the processes shown in Supplementary Note Fig. 3c. We are considering an infinite linear chain since a TF factor after finding its specific site will either move to a non-specific site to the right or left with equal probability and since we are considering a long chain, the right-side chain will be the reflection of the left-side chain. To solve this case analytically we are considering the special case where unbinding is only possible from a specific site. The master equation associated with this process is given by:

$$\frac{\partial P_1(t)}{\partial t} = P_2(t) - 1.5\alpha P_1(t) \quad \text{for } n = 1 \tag{23}$$

$$\frac{\partial P_2(t)}{\partial t} = \alpha P_1(t) + P_3(t) - 2P_2(t) \quad \text{for } n = 2 \tag{24}$$

$$\frac{\partial P_n(t)}{\partial t} = P_{n-1}(t) + P_{n+1}(t) - 2P_n(t) \quad \text{for } n > 2 \tag{25}$$

Site 1 corresponds to the specific site and the initial condition is such that $P_n(0) = \delta_{n,1}$.

Using the property of the Modified Bessel functions of first kind, $\frac{\partial I_n(t)}{\partial t} = \frac{1}{2}\left(I_{n-1}(t) + I_{n+1}(t)\right)$. It is easy to see that $P_n(t) = e^{-2t} I_{n+m}(2t)$ are solutions for 25, with $m \in \mathbb{N}$. Let's consider the linear combination of these particular solutions:

$$\boxed{P_n(t) = \sum_{m=0}^{\infty} A_m^n e^{-2t} I_m(2t)} \tag{26}$$

Applying the general differential equation for $n > 2$ (25), we get the following conditions on the coefficients:

$$\boxed{A_2^n + 2A_0^n = A_1^{n+1} + A_1^{n-1}} \tag{27}$$

$$\boxed{A_{m+1}^n + A_{m-1}^n = A_m^{n-1} + A_m^{n+1}} \quad m \neq 1 \tag{28}$$

Without loss of generality, we can set $A_1^4 = 0$. This implies that:

$$\boxed{A_{m+1}^n = A_m^{n-1};\ \ A_{m-1}^n = A_m^{n+1}} \tag{29}$$

Applying the initial condition that $P_n(0) = \delta_{n1}$, the solution is given by:

$$\boxed{P_n(t) = e^{-2t}\left\{A^n_{n-1}I_{n-1}(2t) + \sum_{m\neq n-1} A^n_m I_m(2t)\right\}} \tag{30}$$

with $A^1_0 = 1,\ A^n_0 = 0$, $n > 1$. Using this solution as ansatz for 24:

$$P_1(t) = e^{-2t}\left\{I_0(2t) + \sum_{m=1} A^1_m I_m(2t)\right\} \tag{31}$$

$$P_2(t) = e^{-2t}\left\{A^2_1 I_1(2t) + \sum_{m=1} A^2_{m+1} I_{m+1}(2t)\right\} \tag{32}$$

$$\frac{\partial P_2(t)}{\partial t} = \alpha P_1(t) + P_3(t) - 2P_2(t) \tag{33}$$

for simplicity the notation $I_m(2t) \equiv I_m$ is used from now on. Then,

$$I_0\left(A^2_1 - \alpha\right) - I_1\left(\alpha A^1_1 + A^3_1 - A^2_2\right) \tag{34}$$

$$= \sum_{m=3} I_m\left(A^3_m + \alpha A^1_m - A^2_{m-1} - A^2_{m+1}\right) + I_2\left(A^3_2 + \alpha A^1_2 - A^2_1 - A^2_3\right)$$

This implies:

$$\boxed{A^2_1 = \alpha; \quad A^3_1 = A^2_2 - \alpha A^1_1} \tag{35}$$

Applying $A^n_{m+1} = A^{n-1}_m$:

$$A^2_3 = \alpha A^1_2 \tag{36}$$

$$\boxed{A^2_{m+1} = \alpha A^1_m \quad m > 2} \tag{37}$$

Applying the ansatz to 24, we find:

$$(1.5\alpha - 2 + A^1_1)I_0 + (2 - A^2_1)I_1 = \sum_{m=1} I_m\left(A^2_m - 1.5\alpha A^1_m - A^1_{m+1} - A^1_{m-1} + 2A^1_m\right) \tag{38}$$

applying $A^2_{m+1} = \alpha A^1_m$, $A^2_1 = \alpha$, $A^2_2 = A^3_1 + \alpha A^1_1$, $A^2_3 = \alpha A^1_2$ and setting without loss of generality $A^3_1 = 0$,

$$\Rightarrow (1.5\alpha - 2 + A^1_1)I_0 + I_1(2 - \alpha - A^1_1(2 - 1.5\alpha) + A^1_2) = \sum_{m=4} I_m\left((2 - 1.5\alpha)A^1_m - A^1_{m+1} + (\alpha - 1)A^1_{m-1}\right) \tag{39}$$

$$+ I_2\left(A^2_2 - 1.5\alpha A^1_2 - A^1_3 - A^1_1 + 2A^1_2\right) + I_3\left(A^2_3 - 1.5\alpha A^1_3 - A^1_4 - A^1_2 + 2A^1_3\right)$$

Then,

$$\boxed{A^1_4 = (\alpha - 1)A^1_2 + (2 - 1.5\alpha)A^1_3} \tag{40}$$

$$\boxed{A^1_3 = A^3_1 + A^1_1(\alpha - 1) + A^1_2(2 - 1.5\alpha)} \tag{41}$$

$$\boxed{A_1^1 = 2 - 1.5\alpha;\ A_2^1 = \alpha - 2 + A_1^1(2 - 1.5\alpha);\ A_{m+2}^1 = (2 - 1.5\alpha)A_{m+1}^1 + (\alpha - 1)A_m^1} \tag{42}$$

Finally,

$$\boxed{P_1(t) = e^{-2t}\left\{I_0(2t) + (2 - 1.5\alpha)I_1(2t) + \left[\alpha - 2 + A_1^1(2 - 1.5\alpha)\right] I_2(2t) + \sum_{m=1}^{\infty} A_{m+2}^1 I_{m+2}(2t)\right\}} \tag{43}$$

In the limit when $t \gg 1$, we can apply the following approximation (10):

$$I_n(2t) \approx (4\pi t)^{-1/2} e^{2t}\left[1 - \frac{(4n^2 - 1)}{16t} + \frac{\left(4n^2 - 1\right)\left(4n^2 - 9\right)}{2!\,(16t)^2} + ...\right] \tag{44}$$

We want to derive the asymptotic behaviour of the solution for $P_1(t)$. For this, we need to calculate the leading terms of the expansion for $I_n(2t)$. First, we are going to calculate the behavior of the leading order of $P_1(t)$, terms of the form $(4\pi t)^{-1/2}$. Assuming that $\alpha \neq 1$:

$$P_1(t) \approx (4\pi t)^{-1/2} \sum_{m=0}^{\infty} A_m + \mathcal{O}(t^{-3/2})$$

$$\sum_{m=0}^{\infty} A_m = A_0 + A_1 + A_2 + A_3 + A_4 + \sum_{m=3}^{\infty} A_{m+2} \tag{45}$$

Using 42,

$$\sum_{m=0}^{\infty} A_m = A_0 + A_1 + A_2 + A_3 + A_4 + \sum_{m=4}^{\infty}(2 - 1.5\alpha)A_m^1 + \sum_{m=3}^{\infty}(\alpha - 1)A_m^1 \tag{46}$$

$$\Rightarrow 0.5\alpha \sum_{m=0}^{\infty} A_m = 0.5\alpha(A_0 + A_1 + A_2) + A_3(1.5\alpha - 1) + A_4 \tag{47}$$

Replacing the values for A_0, A_1, A_2, A_3 and A_4; it is easy to see that:

$$\Rightarrow \boxed{\sum_{m=0}^{\infty} A_m = 0} \tag{48}$$

Which implies that terms of order $t^{-1/2}$ vanish for the solution of $P_1(t)$.We need to calculate the behaviour of the second leading term of the form $\frac{(4\pi t)^{-1/2}}{16t}$ (terms of the order $t^{-3/2}$) which will dominate the asymptotic behaviour of $P_1(t)$. It is easy to see in a similar way that for $\sum_{m=0}^{\infty} A_m$, that

$$\boxed{0 < \sum_{m=0}^{\infty} A_m \frac{(4m^2 - 1)}{16} < \infty} \tag{49}$$

This implies that the asymptotic behavior of $P_1(t)$ will be dominated by a term of the form $\frac{(4\pi t)^{-1/2}}{16t}$:

$$\boxed{P_1(t) \propto t^{-3/2}} \tag{50}$$

Finally, the dwell time distribution $\tilde{D}$ is given by:

$$\frac{d\tilde{D}}{dt} = 0.5\alpha P_1(t) \propto 0.5\alpha t^{-3/2} \tag{51}$$

The survival distribution $\hat{D}$ in the limit $t \gg 1$:

$$\boxed{\hat{D}(t) \propto 0.5\alpha t^{-1/2}}, \tag{52}$$

The survival distribution for such a chain will be the sum of an exponential term due to non-specific binding with a fast dwell time and a term of the form $t^{-1/2}$ as described above ($t \gg 1$). The asymptotic behavior for longer times compared to diffusion along the DNA (random walk along non-specific sites) will therefore show power-law behavior. This suggests that the phenomenological fits of survival time distributions by a mixture of exponentials may have to be modified depending on TF affinity and dynamics.

This analytical result can be applied to transcription factors that undergo multiple conformational changes or interact with multiple protein complexes before unbinding from a specific site. Such a process will produce asymptotically power-law distributed dwell times for specific binding.

Conceptually, our calculation shows that diffusion on the DNA as illustrated in Section 1.4 produces asymptotic power-law behavior for the survival distribution of TFs. This diffusion on the DNA may be seen as a broad distribution of effective affinities depending on how many non-specific targets the TF visited prior to binding to a specific site with each specific binding event having a distinct effective binding affinity.

The next section will explore computationally a more general process where unbinding is allowed from non-specific sites.

1.4.3 Simulating a Complete Model

We used the Gillespie algorithm (11) to simulate the residence time of a TF binding to a specific target in the background of multiple non-specific sites.We modeled the chromatin environment as a circular chain of 2000 sites, with 1999 non-specific and a single specific site (Supplementary Note Fig. 3). TFs were initially allowed to bind at a randomly chosen non-specific site at most 20 sites away from the specific site, since TFs bound farther away contribute negligibly to the final residence time distribution due to the low probability of finding the specific site before unbinding. TFs were allowed to diffuse to a neighboring site with the rate constants as shown in Supplementary Note Fig. 3a or dissociate from the specific site at a rate k_2 or from a nonspecific site at a rate k_3. The dwell time for an iteration corresponds to the time interval between the initial condition of binding at a random site and unbinding of the TF. The simulation was terminated when 1000 specific binding events occurred (multiple specific binding for the same TF in a single iteration were counted as a single binding event) or after 1×10^7 time points (a.u.). The survival distribution was calculated using the Kaplan-Meier estimate for the empirical cumulative distribution function. We find that the survival time distribution clearly shows a power law dependence for different combinations of k_1, k_2 and k_3 when $k_2 < k_1$ (Supplementary Note Fig. 3d). Unbinding from a non-specific site was allowed after a TF binds to a specific site. The power-law exponent is largely determined by the search time or the effective diffusion on the DNA, before finding the specific site based on exploration of different parameters through simulations. Conceptually, our simulation implies that since a TF can visit multiple non-specific sites before binding to the specific site and subsequently unbinding, the effective survival distribution resembles the situation in which the TF encounters sites with a broad distribution of affinities. This suggests that the phenomenological fits of survival time distributions by a mixture of exponentials may have to be modified depending on TF affinity and dynamics.

1.5 Broad distribution of Binding Affinities

Independent of diffusion along the DNA, the affinity landscape in the nucleus may be highly heterogeneous resulting in a broad distribution of binding affinities contrary to the assumption that TF dynamics on chromatin results from well-separated and narrow distributions of specific and non-specific binding (Fig. 5). Given the heterogeneity in local organization and nuclear structure, TF binding sites on chromatin can be viewed as a collection of traps with a distribution of trap depths (analogous to binding affinities, Figure 4H). Each trap can be viewed as a potential energy well of depth ΔE, which is related to the affinity as $k(T) = k_0 e^{-\Delta E/k_B T}$, where k_0 is the bare transition rate (Arrhenius equation). The dwell time distribution of a TF in a particular well of depth ΔE is given by $f_{\Delta E}(t) = k(T)e^{-k(T)t}$. If the energy landscape across the nucleus is described by a distribution $P(\Delta E)$, then the overall dwell time distribution observed in an experiment across all different sites ($f(t)$) will take the form:

$$\boxed{f(t) = \int k_0 e^{-\Delta E/k_B T} e^{-k_0 e^{-\Delta E/k_B T} t} P(\Delta E) d\Delta E} \tag{53}$$

Under very general conditions, Bouchaud et al. have shown that for such finite disordered systems, there exist a family of energy distributions $P(\Delta E)$, for which the distribution of dwell times asymptotically approaches a power law (12, 13). As an example, the well-known Random Energy model and the Sherrington-Kirkpatrick model of spin glasses from physics have energy landscape in which the distribution of deep traps (large ΔE) is given by:

$$P(\Delta E) = \frac{f_0}{k_B T} e^{-x\Delta E/k_B T} \tag{54}$$

where x is a temperature dependent parameter between 0 to 1. For the case of TF binding, this parameter x can be seen as the interaction strength of different types of TFs to a specific site described by a well with energy depth ΔE. Using this distribution for the energy landscape, the dwell time distribution can be calculated:

$$f(t) = \frac{f_0 k_0}{k_B T} \int_0^\infty e^{-\Delta E/k_B T(1+x)} e^{-k_0 e^{-\Delta E/k_B T} t} d\Delta E \tag{55}$$

let's $u = e^{-\Delta E/k_B T}$ and $t' = k_0 t$:

$$f(t) = f_0 k_0 \int_0^{t'} u^x e^{-e^u t'} du \tag{56}$$

setting $ut' = v$,

$$f(t) = f_0 k_0 \int_0^{t'} \left(\frac{v}{t'}\right)^x e^{-v} \frac{dv}{t'} \tag{57}$$

$$\Rightarrow f(t) = \frac{f_0 k_0}{t'^{(1+x)}} \int_0^{t'} v^x e^{-v} dv \tag{58}$$

Using the definition for the incomplete gamma function ($\Gamma(a,x) \equiv \int_x^\infty t^{a-1} e^{-t} dt$)and using $t' = k_0 t$

$$f(t) = t^{-(1+x)} k_0^{-x} f_0 \left[\Gamma(x+1, k_0 t) - \Gamma(x+1, 0)\right] \tag{59}$$

We are interested in the asymptotic behavior of $f(t)$. For $t \gg 1$:

$$f(t) \approx -t^{-(1+x)} k_0^{-x} f_0 \Gamma(x+1, 0) \tag{60}$$

Let $C(x) = k_0^{-x} f_0 \Gamma(x+1, 0)$ which is a time independent quantity ($C(x) < 0$),

$$f(t) \approx -C(x) t^{-(1+x)} \tag{61}$$

Finally, the asymptotic behavior of the survival distribution $\hat{D}(t)$ ($t \gg 1$:) is given by:

$$\boxed{\hat{D}(t) \propto t^{-x}} \tag{62}$$

Thus, the survival distribution exhibits asymptotic power-law behavior over a range of parameters that are experimentally plausible.

1.5.1 Simulation

To simulate a broad distribution of binding affinities, the chromatin was modeled as a chain of 10000 sites. The depth of potential well for each site was randomly chosen from an exponential distribution with a mean of $\frac{x}{k_B T}$ (x was set up to 0.5). The affinity distribution for the site then becomes $k = k_0 e^{-\Delta E / k_B T}$ and the resulting dwell time distribution for any particular well was randomly generated from an exponential distribution with mean $1/k$. The Survival distribution was then calculated using the Kaplan-Meier estimator.

References

1. Liao, J.-C., Spudich, J. A., Parker, D., and Delp, S. L. (2007). Extending the absorbing boundary method to fit dwell-time distributions of molecular motors with complex kinetic pathways. *Proceedings of the National Academy of Sciences*, **104**(9), 3171–3176.

2. Kampen, N. G. V. (1992). *Stochastic processes in physics and chemistry*. Elsevier.

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

4. Paakinaho, V., Presman, D. M., Ball, D. A., Johnson, T. A., Schiltz, R. L., Levitt, P., Mazza, D., Morisaki, T., Karpova, T. S., and Hager, G. L. (2017). Single-molecule analysis of steroid receptor and cofactor action in living cells. *Nature communications*, **8**, 15896.

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

6. Morisaki, T., Muller, W. G., Golob, N., Mazza, D., and McNally, J. G. (2014). Single-molecule analysis of transcription factor binding at transcription sites in live cells. *Nature communications*, **5**, 4456.

7. 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**.

8. John, S., Sabo, P. J., Thurman, R. E., Sung, M.-H., Biddie, S. C., Johnson, T. A., Hager, G. L., and Stamatoyannopoulos, J. A. (2011). Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nature genetics*, **43**, 264–268.

9. Wang, J., Zhuang, J., Iyer, S., Lin, X., Whitfield, T. W., Greven, M. C., Pierce, B. G., Dong, X., Kundaje, A., Cheng, Y., et al. (2012). Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome research*, **22**(9), 1798–1812.

10. Abromowitz, M. and Stegun, I. A. (1965). Handbook of mathematical functions. *NBS (now NIST)*.

11. Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry, 81(25), 2340-2361.*

12. Bouchaud, J.-P. and Georges, A. (1990). Anomalous diffusion in disordered media: statistical mechanisms, models and physical applications. *Physics reports*, **195**(4-5), 127–293.

13. Bouchaud, J.-P. (1992). Weak ergodicity breaking and aging in disordered systems. *Journal de Physique I*, **2**(9), 1705–1713.

Appendix B.

Instructions for Using Single Molecule Tracking Routines Implemented in MATLAB

To follow this guide, you will need:

- MATLAB R2015b or newer, with the *following tools installed*
    - Curve Fitting Toolbox
    - Image Processing Toolbox
    - Optimization Toolbox
- Statistics and Machine Learning Toolbox
- Global Optimization Toolbox
- Basic knowledge of MATLAB (no coding expertise necessary)
- A spreadsheet software such as Excel

Installation:

1 - Unpack the zipped folders “MatlabTrack_v6”and “Analysis_scripts” from https://github.com/davidalejogarcia/PL_HagerLab into your MATLAB files folder (example (C:\Matlab Files\))

2 - Open MATLAB

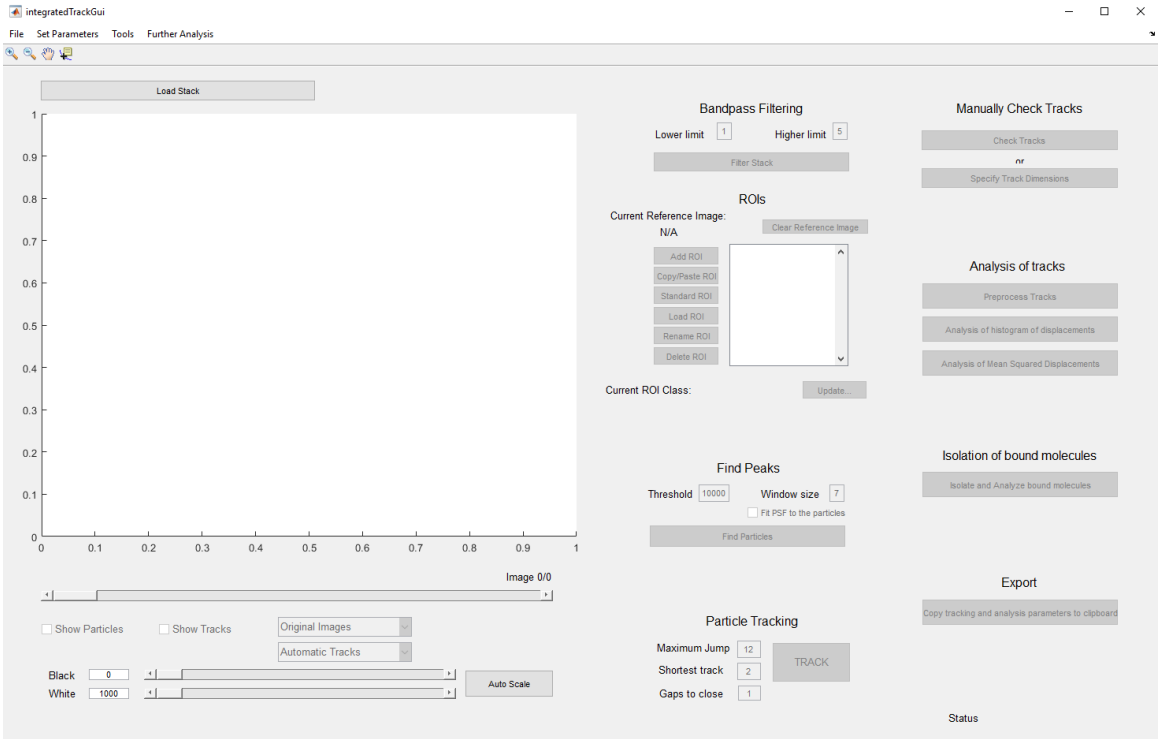
3 - Add the directory to the MATLAB path. To do this, select “Set Path” from the MATLAB “File” menu, click on “Add with Subfolders” and select the folder

“MatlabTrack_v6” and “Analysis_scripts”). Select “Save” in the Set Path panel to have MATLAB memorize the location of the newly installed routines.

Running the MatlabTrack routines:

TRACKING

1 – From MATLAB, in the command window type **integratedTrackGui** and press “Enter”. The main graphical user interface (GUI) for the track analysis should open up (it should look like below).



You will notice a series of buttons and some menus at the top. The menus are used to perform some basic actions (load previously tracked data, save the tracked data, set the parameters of data acquisition and analysis) and some more sophisticated analysis (single molecule colocalization, and global analysis of

multiple tracked movies). The main window enables performing the whole framework of the single molecule analysis.

2 – The first step in analyzing a single molecule movie is to load the movie by pressing the big button Load Stack at the top of the GUI. The movie needs to be in ".tif" format and it is preferable to have the files stored locally (not on some remote drive) for better performance. Once you select the movie to analyze it should appear in the GUI (in the big white area on the left). Make sure to inspect the movie to ensure it is appropriate for analysis. If not (too many particles, blurred, lost focus etc.) proceed to the next movie by clicking on the "Load Stack" button again.

3 – Set the acquisition parameters by clicking on the menu "Set parameters > Set Acquisition Parameters" at the top. A dialog box will appear, requiring as input the time interval between images (Frame Time which is the inverse of the acquisition rate) and the pixel size in micrometers. These parameters can be stored as default so that they are saved for the next time the software is opened. However, it is good practice to double check that the right parameters are being used.

4 – The next step in particle detection is to filter out some of the noise with a bandpass filter. The bandpass filter smoothens the image by "discarding" features smaller than a certain threshold and larger than another threshold. You can change these thresholds by entering lower and higher limits in the corresponding boxes (top of the center column of the GUI). For our system, we find that a "lower limit" equal to 1 and a "higher limit" equal to 5 works well. Then,

press the button “Filter Stack”. This applies a top-hat, Wiener, and Gaussian filter. You can switch between the original stack and the filtered stack by selecting the proper entry in the pop-up menu below the image window. NOTE: changing this entry will only affect what is visualized but NOT how the data analysis is performed. Tracking will always be done on the “Original images”.

5 – Once the stack has been processed, you can choose to analyze only the molecules found in a particular region of interest (ROI) by clicking on the “Define ROI” button in the center column. At this point, if desired, it is possible to open a reference image to define a particular area of the cells [for example a brightfield image of the same field of view as the ROI (such as the nucleus)]. Otherwise the GUI will prompt you to generate a “Sum” or “Maximum” projection image from the stack. Just draw the region of interest around the feature you are interested in. You will also have a choice to adjust the brightness of the generated Sum/Maximum image to help you select the ROI. Additionally, a new pop-up window will ask whether you want to add a second ROI that could be analyzed independently.

6 – The next step in tracking molecules is to identify the peaks corresponding to single molecules. To do so, we need to set two parameters. First, select a “Threshold” that discards all the dim peaks that are likely to be noise. This threshold will depend on the brightness of the particles and the background intensity. Select the threshold so that most of the peaks which are likely to correspond to particles are detected, while keeping the number of “false positive” particles low. Next, select a “window size” (in pixels). This sets the minimum

distance between two particles to be recognized as separated particles (it is set by default to 7, which is our recommended value). Then, click on the “Find Particles” button. The identified particles will be circled in blue. Visually inspect the movie to check if you agree with the results of the particle identification procedure and change the threshold value accordingly (too many particles: increase the threshold; too few particles: decrease the threshold). Before proceeding to the next step, select the “Fit PSF to the particles” option1 and hit “Find Particles again” 2. This procedure is much slower but identifies the position of the particles much more precisely.

7 – Once the particles have been identified, the different positions obtained at different frames of the tracks need to be linked. This is a very critical step, as incorrect tracking might result in nonsensical results. There are a number of options for tracking, as outlined below.

First, set the maximum jump (in pixels) allowed between contiguous frames. The value of this parameter depends on the mobility of the molecule. To track bound molecules, set the “maximum jump” relatively low (4 pixels, depending on the acquisition interval). Otherwise, set it higher to track all particles. Keep in mind that if a large number of particles are identified, the routine might link the particle

[1] Point spread function (PSF) is fitted with a 2D gaussian with a width of 3 pixels independent of numerical aperture (NA) and wavelength

[2] Another way to know if the threshold is ok, you can do a “mock” tracking, without the “fit PSF….” Function. Just track the particles and press the “Preprocess Tracks” button (read and understand steps 7-9). A new window will pop-up. You can now see a qualitative estimation of the signal-to-noise (SNR) ratio in the form of a histogram. If the SNR is low (blue and red histograms overlaps too much) you should increase the threshold and try again. If the SNR is still poor no matter how high the threshold has been set up, then you should consider eliminating this cell. A more quantitative approach is taken on step 10.

positions between subsequent time frames incorrectly, especially if the “maximum jump” setting is too high. Therefore, in order to track molecules moving over long distances, it is optimal to acquire movies with a low number of particles (less expression level, less fluorescent labeling ratio, etc.).

Next, in order to discard tracks that are too short, set the minimal track length. It is recommended to keep this value low (2-4 frames).

Finally, you can select the “close gap” option. This is useful when particles blink, so that if “gaps” are found in the tracks they will be “filled”. A good starting value is gap = 2.

Once these parameters have been wisely selected, click on the “Track” button. The algorithm might be challenged by two opposing situations (in both cases an error message will appear): (i) Too many particles, you need to either increase the threshold to find the particles or set a lower “maximum jump” allowed; ii) Too few particles, here the solution is to decrease the shortest track value.

8 – Before proceeding with the analysis it is important to pre-process the tracks. Click on the button “Preprocess Tracks” in the main GUI[3]. A new pop-up window will give you information on the background and signal intensity levels, as well as the decay in the number of particles over time. Press “DONE”. This process converts the measured tracks to microns, calculates the number of particles

[3] Make sure that the acquisition parameters are set correctly (in the menu “Set acquisition parameters”) before preprocessing the tracks

detected per frame and quantifies the intensity of the particles over the background. After preprocessing, remember to save the MATLAB track file.

9 – After processing all movies (repeat steps 1-9 for all your data), run the script “SNRParticlesCells.mat” located in the “Analysis_scripts” folder by opening it from the HOME tab > OPEN. This script is used to calculate the SNR (Signal to Noise Ratio) of the tracks. It will produce a vector file “SNR” that corresponds to the SNR for each track. This can be used to measure the quality of your data.

9a- Open the script and manually change the required values in the editor window for your set up (W1, wavelength used; NA, numerical aperture used; pixel size used)

9b- It is important that all the .mat files to be analyzed are in the same folder and the current folder in the MATLAB console corresponds to it. Run the script. Select all the .mat files you wish to analyze.

9c - Calculate the fraction of tracks that have an SNR smaller than 2 (you can use the command “length(find(SNR<=2))/length(SNR)”. We put our quality standard at <0.05. If the fraction of tracks with low SNR is higher, consider increasing the thresholding before tracking.[4]

9d – For visualization purposes, you can generate a plot using the command “histogram(SNR,'Normalization','pdf')”

[4] There is a possibility to change the threshold of all files at the same time, however, it will apply the same thresholding value to all. Under Tools, select “Retrack Batch” and change the corresponding parameter. The system will re-track, pre-process and automatically saved a copy of all files. This step may take a while depending on the number of files and processing power.

CALCULATION OF R_{MIN} AND R_{MAX} USING HISTONES

R_{min} (the minimum frame to frame displacement that a molecule may exhibit to be considered bound) and R_{max} (maximum displacement in a shortest track number of frames that a molecule may exhibit to be considered bound) are essential to define when a molecule is bound and when it is not. We calculate R_{min} and R_{max} using histone data. This should be done for each acquisition condition and cell line used.

1- Under the “Further Analysis” menu, select “Merge and analyze jump histograms”. Then, select the preprocessed MATLAB files generated during tracking histones from different experiments. Select the properties of the histogram in the pop-up window (you can use the default ones). Make sure the right frame time is selected. Then, enter ‘yes’ for photobleaching correction and ‘yes’ for normalization of the jump histogram to 1 to make the displacement a normalized distribution.

2- Click on the “Copy 2D hist of displacements” button and paste the results on a spreadsheet such as Excel. The table is the probability for a given displacement (in micrometers) at different time intervals. An example is shown below.

Frame displacement

Rmin

Rmax

Time interval	0	0.01	0.03	0.05	0.07	0.09	0.11	0.13	0.15	0.17	0.19	0.21	0.23	0.25	0.27	0.29	0.31
0.2	0.17131	0.2822	0.21586	0.13654	0.079971	0.045142	0.026207	0.014634	0.008468	0.0052	0.003581	0.002357	0.001689	0.001453	0.001143	0.000981	
0.4	0.1049	0.21813	0.20619	0.15093	0.10166	0.064388	0.039847	0.025066	0.016042	0.010145	0.006493	0.004514	0.003389	0.002527	0.001991	0.001513	
0.6	0.085631	0.18837	0.18656	0.1466	0.10418	0.070566	0.046336	0.03045	0.020217	0.012887	0.009064	0.006061	0.00435	0.003162	0.002487	0.001975	
0.8	0.07549	0.16906	0.17144	0.13862	0.10311	0.071733	0.049109	0.033124	0.022598	0.015248	0.010374	0.007369	0.00515	0.003653	0.002789	0.002167	
1	0.069527	0.15448	0.16115	0.1323	0.099391	0.071308	0.050689	0.035198	0.024612	0.016895	0.011229	0.007968	0.00545	0.004072	0.003033	0.002495	
1.2	0.064146	0.14431	0.15207	0.12688	0.096051	0.070689	0.050956	0.036656	0.02583	0.017815	0.011995	0.008117	0.006299	0.004452	0.003455	0.002638	
1.4	0.059986	0.13547	0.14411	0.12124	0.093624	0.070249	0.051257	0.037378	0.026931	0.018229	0.01283	0.009135	0.00638	0.004703	0.003538	0.002917	
1.6	0.056041	0.12921	0.13806	0.11613	0.091641	0.069044	0.051526	0.036581	0.027228	0.018826	0.013524	0.0097	0.006966	0.005238	0.003862	0.003017	
1.8	0.053476	0.12214	0.1321	0.11349	0.088816	0.068254	0.05037	0.037821	0.027092	0.019419	0.014028	0.009882	0.007317	0.005229	0.004252	0.003329	
2	0.050637	0.11736	0.12694	0.10948	0.086753	0.067082	0.050854	0.037407	0.027516	0.019588	0.014195	0.010096	0.007815	0.00575	0.004281	0.003089	

R_{min} and R_{max} calculation example. The figure shows an example of the output table generated from the "Copy 2D hist of displacements" function. Please note the actual table will be longer. The top row (in grey) displays the displacements in microns. The left column (in blue) displays the time intervals (in seconds). Each cell displays the probability of the particle's displacements. In this example, the interval time is 200ms (you can deduce it from the time increments 0.2, 0.4, 0.6, etc) and the Shortest Track (ST) was set to 4 during tracking. To calculate R_{min}, you must choose the row corresponding to the interval time used (in this case 0.2s, yellow) and SUM all the probabilities until you reach 0.99. In this case, the displacement is 0.23. This is R_{min}. To calculate R_{max}, you must choose the row that corresponds to the ST time interval (In this case ST=4, so 4*0.2s = 0.8s; highlighted in green). Because this row is not normalized, first you must SUM all the probabilities in the row (in this case this is 0.8873) and second, calculate 99% of this value (in this case 0.874). Finally, you must SUM all the probabilities until you reach 0.8873; in this case R_{max} is 0.29.

2a- For **R_{min}**, calculate from the row corresponding to your frame time, the displacement which corresponds to the cumulative probability less than or equal to 0.99 (i.e. sum all probabilities until you reach 0.99). The corresponding distance is R_{min}.

2b – For **R_{max}**, use the row corresponding to the value of the Shortest Track (in seconds). Calculate the displacement that corresponds to the cumulative probability less or equal to 0.99. Care must be taken for determining R_{max}. Because R_{max} is not normalized, you need to use the 99% value from the sum of all rows (i.e. the command: sum(row)*0.99).

ANALYSIS OF THE RESIDENCE TIME HISTOGRAM OF BOUND MOLECULES

1 – In the "Further Analysis" menu select "Merge and analyze residence time distribution". Then, select the individual preprocessed MATLAB files generated during tracking that you want to analyze. A pop-up window will ask to

select either “Residence time histogram” or “Survival Time Distribution”. Choose the latter. Enter the “maximum jump between consecutive frames” (Rmin) and “Maximum end to end distance” (Rmax) as calculated in the previous step. “Minimum length of bound tracks” (or Nmin) corresponds to the least number of frames of a track to be considered bound. It should mirror the shortest tracks (ST) previously selected. For long acquisition intervals (e.g. 500 ms) we suggest ST=2 frames and for short acquisition intervals (e.g. 200 ms) we suggest ST =4 frames. The frame time and the maximum number of frames to be analyzed need to be entered as well in the pop-up window.

2 – Once these parameters are set, run the analysis of the bound molecules by clicking on the “ok” button. Upon completion of the computation (it may take a few minutes), a pop-up window will emerge with a bi-exponential analysis (old methodology). Please close this window and ignore it. The MATLAB command window will display the execution time and a MATLAB file will be saved in the current folder called “DwellTimehistogram.mat”, containing the histogram of track lengths.

ESTIMATION OF PHOTOBLEACHING RATE USING HISTONES (H2B, H3 or H4)

This can only be done after generating the “DwellTimehistogram.mat” file for the bound tracks. However, if you wish to use another tracking software to generate the tracks, you can continue with the analysis if and only if the generated tracks are saved with the current format: an excel file with two columns, column A with the experimental time points in seconds and column B with the number of

particles with a lived time of column A (see sample file "Sample_ResTimeHist.xlsx"). Run the script "ImportData" and select the excel file. The script will generate a file with the name of said excel file that can be further analyzed.

NOTE. If you use uTrack to generate your tracks, we have generated a script to make your dataset compatible with the following analysis. Create a folder with the tracking data from utrack for all the experiments of interest (files in "TrackingPackage/tracks"). Run the script "Utrack_ResTimeHist" modifying the lane 3 with the acquisition time of the experiment in seconds. The script will generate a file called "UtrackMergedData" that can be used for further analysis.

1 – Making sure that the pathname of the file to be analyzed is current in the MATLAB console, access the "PhotobleachingFit" file (located in the Analysis scripts folder) by opening it from the HOME tab > OPEN. Modify line 6 ("period") with the acquisition rate in seconds and line 7 ("st") with the shortest track value in frames. Load the "DwellTimehistogram.mat" file and run the script. Two plots and two fits will be generated. The figures correspond to fitting the curve with a triple exponential (**Figure1**) or double exponential (**Figure2**) with dashed lines showing the 99% CI in the survival distribution. In the console, a message will display the evidence in decibels (Db) for the triple exponential model in comparison with the double exponential mode. **If the evidence is larger than 30, then the triple exponential model is a better predictor of the data**. The program also prints BIC1, corresponding to the Bayesian Inference Criterion. The minimum value corresponds to the best predictive model. Alternatively, visual inspection of the

graphs should also give you the best fit (the one that closely resembles the raw data inside the CI). Typing “Parameter1” will display the exponential parameters for the triple exponential fit[5] and “CI_Parameter1” the corresponding 95% Confidence Interval. The photobleaching rate corresponds to **the smallest exponential parameter**. Typing “Parameter2” will display the fitting parameters for the double exponential fit[6] and “CI_Parameter2” the corresponding 95% Confidence Interval. **Make note** of the smallest exponential parameter for the best fit. **Type in command “clear all”** before uploading/analyzing new files.

2 – Multiple biological replicates taken on different days can be combined to obtain a better estimate of the photobleaching rate[7]. Open the “MergeExperiments” script. Put all files to be merged in the same folder and make sure that the pathname is current in the MATLAB console. Run the script to generate a MATLAB file “Merged.mat”. The photobleaching analysis can be run using the new “Merged.mat” file as in the above step. Make note of the **smallest exponential parameter from the best fit (i.e. photobleaching rate).**

SURVIVAL DISTRIBUTION FITS

[5] For the triple exponential case, the equation to fit is: $f(t|f_1, f_2, f_3, \beta, \gamma, \eta) = f_1 e^{-\beta t} + f_2 e^{-\gamma t} + f_3 e^{-\eta t}$. The parameters are displayed in the following order: $f_1, f_2, f_3, \beta, \gamma, \eta$

[6] For the double exponential case, the equation to fit is: $f(t|f_1, f_2, f_3, \beta, \gamma, \eta) = f_1 e^{-\gamma t} + f_2 e^{-\eta t}$. The parameters are displayed in the following order: $f_1, f_2, f_3, \gamma, \eta$

[7] While you should finally use this “merged” value, it is recommended to run all individual experiments (step1) in case one of them deviates too much from the rest and therefore should be discarded

After estimating the photobleaching rate, we are ready to obtain the fits to the survival distribution of our protein of interest. Once again, if you wish to use another tracking software to generate the tracks, please refer to the previous section.

1 – Open the “SurvivalFit.m” file and modify line 6 (“period”) with the time interval of acquisition in seconds, line 7 (“st”) with the shortest track value in frames and line 8 with the photobleaching rate (PB) calculated from histone data. Select the part of data that you want to include in the analysis (sometimes the tail is really noisy, and it ought to be removed). Change the value of line 11 (“final”) for the minimum number of cumulative tracks in the tail to be analyzed (Default 30, the portion of the Survival distribution with less than 30 tracks will be censored in the analysis) or 1 otherwise which will include the tail of the distribution. Load the “DwellTimehistogram.mat” file of your protein of interest and run the script. This will generate three plots with 99% CI as dashed lines and three fits:

The first plot corresponds to a double exponential fit; type “Parameter1” in the MATLAB console to display the fitting parameters[8] and “CI_Parameter1” for their respective 95 CI.

[8] For the double exponential model, the equation to fit is: $f(t|f_1,\gamma,\eta,A) = A(f_1 e^{-\gamma t} + (1-f_1)e^{-\eta t})$. The parameters are displayed in the following order: f_1,γ,η,A

The second plot corresponds to a Power Law fit; type “Parameter 2” in the MATLAB console to display the fitting parameters[9] and “CI_Parameter2” for their respective 95 CI.

The third plot corresponds to the triple exponential model[10]; type “Parameter3” in the MATLAB console to display the fitting parameters.

In the console, a message with the evidence for all models in decibels (Db) is displayed. If the evidence is larger than 30 for any model, then said model is a better predictor of the data. The program also prints BIC1, corresponding to the Bayesian Inference Criterion. The minimum value corresponds to the best predictive model. This is a simplified model selection. It is important to note that when you are performing model selection, you need to graphically inspect the plots, understand that there is biological and statistical variability in the data, and these are models that must be seen from a biological perspective. The results may not be conclusive due to low range in the experimental observations of dwell times. For instance, if the photobleaching rate is too high that events longer that 20 s are not observable, then a distinction between power law and exponential models become inaccurate. This may indicate that experiments with longer acquisition times and lower photobleaching are required.

[9] For the Power Law model, the equation to fit is: $f(t|A,b) = At^{-b}$. The parameters are displayed in the following order: A, b

[10] For the triple exponential model, the equation to fit is: $f(t|f_1,\gamma,\eta,A) = A(f_1 e^{-\gamma t} + f_2 e^{-\eta t} + f_3 e^{-\beta t})$. The parameters are displayed in the following order: $f_1, f_2, f_3, \gamma, \eta, \beta, A$

2 – Multiple biological replicates performed on different days may be combined to obtain a better estimate of the survival distribution[11]. Open the "MergeExperiments" script. Select the portion of data that you want to include in the survival distribution analysis. As mentioned before, the tail can be noisy. This script will generate a file called "Merged.mat".

3 – Load the file "Merged.mat" and run the script "SurvivalFit.mat" as above. The best fit and the corresponding parameters should be used to characterize your protein of interest.

[11] While you should finally use this "merged" value, it is recommended to run all individual experiments (step1) in case one of them deviates too much from the rest.

References

Abraham, A. V., Ram, S., Chao, J., Ward, E. S., & Ober, R. J. (2009). Quantitative Study of Single Molecule Location Estimation Techniques. *Opt Express*(17), 23352-23373.

Afek, A., Schipper, J. L., Horton, J., Gordân, R., & Lukatsky, D. B. (2014). DNA binding in the absence of specific base-pair recognition. *Proceedings of the National Academy of Sciences*, 17140-17145.

Agarwal, H., Reisser, M., Wortmann, C., & Gebhardt, J. C. M. (2017). Direct Observation of Cell-Cycle-Dependent Interactions between CTCF and Chromatin. *Biophys J, 112*(10), 2051-2055. doi:10.1016/j.bpj.2017.04.018

Agrawal, A., & Lele, T. P. (2019). Mechanics of nuclear membranes. *J Cell Sci, 132*(14). doi:10.1242/jcs.229245

Agresti, A., Scaffidi, P., Riva, A., Caiolfa, V. R., & Bianchi, M. E. (2005). GR and HMGB1 interact only within chromatin and influence each other's residence time. *Mol.Cell, 18*(1), 109-121.

Alam, S. G., Zhang, Q., Prasad, N., Li, Y., Chamala, S., Kuchibhotla, R., . . . Lele, T. P. (2016). The mammalian LINC complex regulates genome transcriptional responses to substrate rigidity. *Sci Rep, 6*, 38063. doi:10.1038/srep38063

Alberti, S. (2017). The wisdom of crowds: regulating cell function through condensed states of living matter. *J Cell Sci, 130*(17), 2789-2796. doi:10.1242/jcs.200295

Alberti, S., Gladfelter, A., & Mittag, T. (2019). Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell, 176*(3), 419-434. doi:10.1016/j.cell.2018.12.035

Allan, J., Cowling, G. J., Harborne, N., Cattini, P., Craigie, R., & Gould, H. (1981). Regulation of the higherorder structure of chromatin by histones H1 and H5. *The journal of cell biology*(90), 279-288.

Allen, B. L., & Taatjes, D. J. (2015). The Mediator complex: a central integrator of transcription. *Nat Rev Mol Cell Biol, 16*(3), 155-166. doi:10.1038/nrm3951

Amitai, A. (2018). Chromatin Configuration Affects the Dynamics and Distribution of a Transiently Interacting Protein. *Biophys J, 114*(4), 766-771. doi:10.1016/j.bpj.2017.12.037

Arpino, G., Wiechmann, L., Osborne, C. K., & Schiff, R. (2008). Crosstalk between the estrogen receptor and the HER tyrosine kinase receptor family: molecular mechanism and clinical implications for endocrine therapy resistance. *Endocrinology Review, 29*, 217-233.

Aumiller, W. M., & Keating, C. D. (2016). Phosphorylation-mediated RNA/peptide complex coacervation as a model for intracellular liquid organelles. *Nat Chem*(8), 129-137.

Axelrod, D. (2001). Total internal reflection fluorescence microscopy in cell biology. *Traffic, 2*(11), 764-774. doi:10.1034/j.1600-0854.2001.21104.x

Babu, M. M., Kriwacki, R. W., & Pappu, R. V. (2012). Structural biology. Versatility from protein disorder. *Science*(337), 1460-1461.

Babu, M. M., van der Lee, R., de Groot, N. S., & Gsponer, J. (2011). Intrinsically disordered proteins: regulation and disease. (21), 432-440.
Baldwin, A. K., Cain, S. A., Lennon, R., Godwin, A., Merry, C. L., & Kielty, C. M. (2014). Epithelial-mesenchymal status influences how cells deposit fibrillin microfibrils. *J Cell Sci, 127*(Pt 1), 158-171. doi:10.1242/jcs.134270
Ball, D. A., Mehta, G. D., Salomon-Kent, R., Mazza, D., Morisaki, T., Mueller, F., . . . Karpova, T. S. (2016). Single molecule tracking of Ace1p in Saccharomyces cerevisiae defines a characteristic residence time for non-specific interactions of transcription factors with chromatin. *Nucleic Acids Res, 44*(21), e160. doi:10.1093/nar/gkw744
Banani, S. F., Lee, H. O., Hyman, A. A., & Rosen, M. K. (2017). Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol, 18*(5), 285-298. doi:10.1038/nrm.2017.7
Banaz, N., Mäkelä, J., & Uphoff, S. (2019). Choosing the right label for single-molecule tracking in live bacteria: side-by-side comparison of photoactivatable fluorescent protein and Halo tag dyes. *J Phys D Appl Phys, 52*(6), 064002. doi:10.1088/1361-6463/aaf255
Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., . . . Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell, 129*(4), 823-837.
Baschant, U., Culemann, S., & Tuckermann, J. (2013). Molecular determinants of glucocorticoid actions in inflammatory joint diseases. *Mol Cell Endocrinol., 380*(1-2), 108-118.
Baschant, U., Frappart, L., Rauchhaus, U., Bruns, L., Reichardt, H. M., Kamradt, T., . . . Tuckermann, J. P. (2011). Glucocorticoid therapy of antigen-induced arthritis depends on the dimerized glucocorticoid receptor in T cells. *Proc.Natl.Acad.Sci U.S.A, 108*(48), 19317-19322.
Basu, S., Mackowiak, S. D., Niskanen, H., Knezevic, D., Asimi, V., Grosswendt, S., . . . Hnisz, D. (2020). Unblending of Transcriptional Condensates in Human Repeat Expansion Disease. *Cell, 181*(5), 1062-1079.e1030. doi:10.1016/j.cell.2020.04.018
Bauer, J., Emon, M. A. B., Staudacher, J. J., Thomas, A. L., Zessner-Spitzenberg, J., Mancinelli, G., . . . Jung, B. (2020). Increased stiffness of the tumor microenvironment in colon cancer stimulates cancer associated fibroblast-mediated prometastatic activin A signaling. *Sci Rep, 10*(1), 50. doi:10.1038/s41598-019-55687-6
Bauer, M., & Metzler, R. (2012). Generalized facilitated diffusion model for DNA-binding proteins with search and recognition states. *Biophys J, 102*(10), 2321-2330. doi:10.1016/j.bpj.2012.04.008
Becker, M., Baumann, C. T., John, S., Walker, D., Vigneron, M., McNally, J. G., & Hager, G. L. (2002). Dynamic behavior of transcription factors on a natural promoter in living cells. *EMBO Rep., 3*(12), 1188-1194.
Ben-Avraham, Z., & Havlin, S. (2000). *Diffusion and reactions in fractals and disordered systems*. Cambridge: Cambridge University Press.

Benichou, O., Chevalier, C., Meyer, B., & Voituriez, R. (2011). Facilitated diffusion of proteins on chromatin. *Phys Rev Lett, 106*(3), 038102. doi:10.1103/PhysRevLett.106.038102

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J.R.Statist.Soc.B, 57*, 289-300.

Berg, O. G., & Blomberg, C. (1976). Association kinetics with coupled diffusional flows. Special application to the lac repressor--operator system. *Biophys Chem, 4*(4), 367-381.

Berry, J., Weber, S. C., Vaidya, N., Haataja, M., & Brangwynne, C. P. (2015). RNA transcription modulates phase transition-driven nuclear body assembly. *Proc Natl Acad Sci U S A, 112*(38), E5237-5245. doi:10.1073/pnas.1509317112

Biddle, J. W., Nguyen, M., & Gunawardena, J. (2019). Negative reciprocity, not ordered assembly, underlies the interaction of Sox2 and Oct4 on DNA. *Elife, 8*. doi:10.7554/eLife.41017

Biedzinski, S., Agsu, G., Vianay, B., Delord, M., Blanchoin, L., Larghero, J., . . . Brunet, S. (2020). Microtubules control nuclear shape and gene expression during early stages of hematopoietic differentiation. *EMBO J, 39*(23), e103957. doi:10.15252/embj.2019103957

Bintu, B., Mateo, L. J., Su, J. H., Sinnott-Armstrong, N. A., Parker, M., Kinrot, S., . . . Zhuang, X. (2018). Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science, 362*(6413). doi:10.1126/science.aau1783

Bissell, M. J., & Radisky, D. (2001). Putting tumours in context. . *Nature REviews Cancer, 1*, 46-54.

Bissell, M. J., Radisky, D. C., Rizki, A., Weaver, V. M., & Petersen, O. W. (2002). The organizing principle: microenvironmental influences in the normal and malignant breast. *Differentiation, 70*(9-10), 537-546. doi:10.1046/j.1432-0436.2002.700907.x

Bledsoe, R. K., Montana, V. G., Stanley, T. B., Delves, C. J., Apolito, C. J., McKee, D. D., . . . Xu, H. E. (2002). Crystal structure of the glucocorticoid receptor ligand binding domain reveals a novel mode of receptor dimerization and coactivator recognition. *Cell, 110*(1), 93-105.

Boehning, M., Dugast-Darzacq, C., Rankovic, M., Hansen, A. S., Yu, T., Marie-Nelly, H., . . . Zweckstetter, M. (2018). RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat Struct Mol Biol, 25*(9), 833-840. doi:10.1038/s41594-018-0112-y

Boeynaems, S., Bogaert, E., Kovacs, D., Konijnenberg, A., Timmerman, E., Volkov, A., . . . Janke, A. M. (2017). Phase separation of C9orf72 dipeptide repeats perturbs stress granule dynamics. *Molecular Cell*(65), 1044-1055.

Boija, A., Klein, I. A., Sabari, B. R., Dall'Agnese, A., Coffey, E. L., Zamudio, A. V., . . . Young, R. A. (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell*. doi:10.1016/j.cell.2018.10.042

Remodelling the extracellular matrix in development and disease, 15, Nature Publishing Group 786-801 (2014).
Bouchaud, J.-P., & Georges, A. (1990). Anomalous diffusion in disordered media: Statistical mechanisms, models and physical applications. *Physics Reports, 195*(4), 127-293. doi:https://doi.org/10.1016/0370-1573(90)90099-N
Bouzid, T., Kim, E., Riehl, B. D., Esfahani, A. M., Rosenbohm, J., Yang, R., . . . Lim, J. Y. (2019). The LINC complex, mechanotransduction, and mesenchymal stem cell function and fate. *J Biol Eng, 13*, 68. doi:10.1186/s13036-019-0197-9
Boyle, A. P., Guinney, J., Crawford, G. E., & Furey, T. S. (2008). F-Seq: a feature density estimator for high-throughput sequence tags. *Bioinformatics, 24*(21), 2537-2538. doi:10.1093/bioinformatics/btn480
Brangwynne, C. P., Eckmann, C. R., Courson, D. S., Rybarska, A., Hoege, C., Gharakhani, J., . . . Hyman, A. A. (2009). Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science*(324), 1729-1732.
Mechanotransduction in tumor progression: The dark side of the force, 217, Rockefeller University Press 1571-1587 (2018).
Brodsky, S., Jana, T., Mittelman, K., Chapal, M., Kumar, D. K., Carmi, M., & Barkai, N. (2020). Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding Specificity. *Mol Cell*. doi:10.1016/j.molcel.2020.05.032
Brouwer, I., & Lenstra, T. L. (2019). Visualizing transcription: key to understanding gene expression dynamics. *Curr Opin Chem Biol, 51*, 122-129. doi:10.1016/j.cbpa.2019.05.031
Bryois, J., Garrett, M. E., Song, L., Safi, A., Giusti-Rodriguez, P., Johnson, G. D., . . . Sklar, P. (2018). Evaluation of chromatin accessibility in prefrontal cortex of individuals with schizophrenia. *Nature Communications, 9*, 1-15.
Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & 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*(12), 1213-1218.
Buenrostro, J. D., Wu, B., Chang, H. Y., & Greenleaf, W. J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol, 109*, 21 29 21-21 29 29. doi:10.1002/0471142727.mb2129s109
Burdall, S., Hanby, A., Lansdown, M. R., & Speirs, V. (2003). Breast cancer cell lines: friend or foe? *Breast Cancer Res*(5), 89-95.
Burris, T. P., Busby, S. A., & Griffin, P. R. (2012). Targeting orphan nuclear receptors for treatment of metabolic diseases and autoimmunity. *Chemistry & Biology, 19*(1), 8.
Butcher, D. T., Alliston, T., & Weaver, V. M. (2009). A tense situation:forcing tumour progression. *Nature REviews Cancer*(9), 108-122.
Caccianini, L., Normanno, D., Izeddin, I., & Dahan, M. (2015). Single molecule study of non-specific binding kinetics of LacI in mammalian cells. *Faraday Discuss, 184*, 393-400. doi:10.1039/c5fd00112a

Caizzi, L., Ferrero, G., Cutrupi, S., Cordero, F., Ballare, C., Miano, V., . . . De Bortoli, M. (2014a). Genome-wide activity of unliganded estrogen receptor-alpha in breast cancer cells. *Proc.Natl.Acad.Sci U.S.A, 111*(13), 4892-4897.

Caizzi, L., Ferrero, G., Cutrupi, S., Cordero, F., Ballare, C., Miano, V., . . . De Bortoli, M. (2014b). Genome-wide activity of unliganded estrogen receptor-alpha in breast cancer cells. *Proc Natl Acad Sci U S A, 111*(13), 4892-4897. doi:10.1073/pnas.1315445111

Caliari, S. R., & Burdick, J. A. (2016). A practical guide to hydrogels for cell culture. *Nat Methods, 13*(5), 405-414. doi:10.1038/nmeth.3839

Callegari, A., Sieben, C., Benke, A., Suter, D. M., Fierz, B., Mazza, D., & Manley, S. (2019). Single-molecule dynamics and genome-wide transcriptomics reveal that NF-kB (p65)-DNA binding times can be decoupled from transcriptional activation. *PLoS Genet, 15*(1), e1007891. doi:10.1371/journal.pgen.1007891

Carter, D., Chakalova, L., Osborne, C. S., Dai, Y. F., & Fraser, P. (2002). Long-range chromatin regulatory interactions in vivo. *Nat.Genet., 32*(4), 623-626.

Chen, J., Zhang, Z., Li, L., Chen, B. C., Revyakin, A., Hajj, B., . . . Liu, Z. (2014). Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell, 156*(6), 1274-1285.

Chi, D., Singhal, H., Li, L., XiaoT., Liu, W., Pun, M., . . . Rao, P. (2019). Estrogen receptor signaling is reprogrammed during breast tumorigenesis. *Proc Natl Acad Sci U S A*, 11437-11443.

Cho, W. K., Jayanth, N., English, B. P., Inoue, T., Andrews, J. O., Conway, W., . . . Cisse, II. (2016). RNA Polymerase II cluster dynamics predict mRNA output in living cells. *Elife, 5*. doi:10.7554/eLife.13617

Cho, W. K., Spille, J. H., Hecht, M., Lee, C., Li, C., Grube, V., & Cisse, II. (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science, 361*(6400), 412-415. doi:10.1126/science.aar4199

Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G. M., Cattoglio, C., . . . Tjian, R. (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science, 361*(6400). doi:10.1126/science.aar2555

Chubb, J. R., Bloomfield, G., Xu, Q., Kaller, M., Ivens, A., Skelton, J., . . . Singer, R. H. (2006). Developmental timing in Dictyostelium is regulated by the Set1 histone methyltransferase. *Dev.Biol., 292*(2), 519-532.

Clarke, R., Brunner, N., Katzenellenbogen, B. S., Thompson, E. W., Norman, M. J., Koppi, C., . . . Dickson, R. B. (1989). Progression of human breast cancer cells from hormone-dependent to hormone-independent growth both in vitro and in vivo. *Proc Natl Acad Sci U S A, 86*(10), 3649-3653. doi:10.1073/pnas.86.10.3649

Coons, L. A., Burkholder, A. B., Hewitt, S. C., McDonnell, D. P., & Korach, K. S. (2019). Decoding the Inversion Symmetry Underlying Transcription Factor

DNA-Binding Specificity and Functionality in the Genome. *iScience, 15*, 552-591. doi:10.1016/j.isci.2019.04.006

Corbett, A. H. (2018). Post-transcriptional regulation of gene expression and human disease. *Curr Opin Cell Biol*(52), 96-104.

Corces, M. R., Buenrostro, J. D., Wu, B., Greenside, P. G., Chan, S. M., Koenig, J. L., . . . Chang, H. Y. (2016). Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet, 48*(10), 1193-1203. doi:10.1038/ng.3646

Coulon, A., Chow, C. C., Singer, R. H., & Larson, D. R. (2013). Eukaryotic transcriptional dynamics: from single molecules to cell populations. *Nat Rev.Genet., 14*(8), 572-584.

Cox, T. R., & Erler, J. T. (2011). Remodeling and homeostasis of the extracellular matrix: implications for fibrotic diseases and cancer. *Dis Model Mech, 4*(2), 165-178. doi:10.1242/dmm.004077

Cramer, P. (2019). Organization and regulation of gene transcription. *Nature, 573*(7772), 45-54. doi:10.1038/s41586-019-1517-4

Cross, S. E., Jin, Y. S., Rao, J., & Gimzewski, J. K. (2007). Nanomechanical analysis of cells from cancer patients. *Nature Nanotechnology, 2*, 780-783.

Dahirel, V., Paillusson, F., Jardat, M., Barbi, M., & Victor, J. M. (2009). Nonspecific DNA-protein interaction: why proteins can diffuse along DNA. *Physical review letters*, 228101.

Dahl, K. N., Ribeiro, A. J., & Lammerding, J. (2008). Nuclear shape, mechanics, and mechanotransduction. *Circ Res, 102*(11), 1307-1318. doi:10.1161/CIRCRESAHA.108.173989

Darling, A. L., & Uversky, V. N. (2017). Intrinsic Disorder in Proteins with Pathogenic Repeat Expansions. *Molecules, 22*(12). doi:10.3390/molecules22122027

Darzacq, X., Yao, J., Larson, D. R., Causse, S. Z., Bosanac, L., de Turris, V., . . . Singer, R. H. (2009). Imaging transcription in living cells. *Annu.Rev.Biophys., 38*, 173-196.

Dasgupta, S., Lonard, D. M., & O'Malley, B. W. (2014). Nuclear receptor coactivators: master regulators of human health and disease. *Annu Rev Med, 65*, 279-292. doi:10.1146/annurev-med-051812-145316

De Bosscher, K., Vanden Berghe, W., & Haegeman, G. (2006). Cross-talk between nuclear receptors and nuclear factor kappaB. *Oncogene, 25*(51), 6868-6886.

de Leeuw, R., Neefjes, J., & Michalides, R. (2011). A role for estrogen receptor phosphorylation in the resistance to tamoxifen. *Int J Breast Cancer, 2011*, 232435. doi:10.4061/2011/232435

Desantis, C., Ma, J., Bryan, L., & Jemal, A. (2013). Breast cancer statistics. *CA Cancer J Clin*(64), 52-62.

Dirks, R. M., & Pierce, N. A. (2004). Triggered amplification by hybridization chain reaction. *Proc Natl Acad Sci U S A, 101*(43), 15275-15278. doi:10.1073/pnas.0407024101

Donovan, B. T., Huynh, A., Ball, D. A., Patel, H. P., Poirier, M. G., Larson, D. R., . . . Lenstra, T. L. (2019). Live-cell imaging reveals the interplay between transcription factors, nucleosomes, and bursting. *EMBO J, 38*(12). doi:10.15252/embj.2018100809

Dundr, M., Hoffmann-Rohrer, U., Hu, Q., Grummt, I., Rothblum, L. I., Phair, R. D., & Misteli, T. (2002). A kinetic framework for a mammalian RNA polymerase in vivo. *Science, 298*(5598), 1623-1626.

Elbi, C., Walker, D. A., Romero, G., Sullivan, W. P., Toft, D. O., Hager, G. L., & DeFranco, D. B. (2004). Molecular chaperones function as steroid receptor nuclear mobility factors. *Proc.Natl.Acad.Sci.USA, 101*(9), 2876-2881.

Engler, A. J., Sen, S., Sweeney, H. L., & Discher, D. E. (2006). Matrix elasticity directs stem cell lineage specification. *Cell, 126*(4), 677-689. doi:10.1016/j.cell.2006.06.044

Erdmann, R. S., Baguley, S. W., Richens, J. H., Wissner, R. F., Xi, Z., Allgeyer, E.S., Z., S., . . . R. and Bewersdorf, J. (2019). Labeling Strategies Matter for Super-Resolution Microscopy: A Comparison

between HaloTags and SNAP-tags. *Cell Chem Biol*(26), 584-592.

Evans, R. M. (1988). The steroid and thyroid hormone receptor superfamily. *Science, 240*, 6.

Everaers, R., & Schiessel, H. (2015). The physics of chromatin. *J Phys Condens Matter, 27*(6), 060301. doi:10.1088/0953-8984/27/6/060301

Ewen, M. E. (2000). Where the cell cycle and histones meet. *Genes Dev, 14*(18), 2265-2270.

Falk, M., Feodorova, Y., Naumova, N., Imakaev, M., Lajoie, B. R., Leonhardt, H., . . . Mirny, L. A. (2019). Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature*(570), 395-399.

Fan, H. Y., Trotter, K. W., Archer, T. K., & Kingston, R. E. (2005). Swapping function of two chromatin remodeling complexes. *Mol.Cell, 17*(6), 805-815.

Feric, M., Vaidya, N., Harmon, T. S., Mitrea, D. M., Zhu, L., Richardson, T. M., . . . Brangwynne, C. P. (2016). Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell, 165*(7), 1686-1697. doi:10.1016/j.cell.2016.04.047

Finn, E. H., & Misteli, T. (2019). Molecular basis and biological function of variability in spatial genome organization. *Science, 365*(6457). doi:10.1126/science.aaw9498

Tumors: Wounds That Do Not Heal, 315, N Engl J Med 1650-1659 (1986).

Fryer, C. J., & Archer, T. K. (1998). Chromatin remodeling by the glucocorticoid receptor requires the BRG1 complex. *Nature, 393*(6680), 88-91.

Galon, J., Franchimont, D., Hiroi, N., Frey, G., Boettner, A., Ehrhart-Bornstein, M., . . . Bornstein, S. R. (2002). Gene profiling reveals unknown enhancing and suppressive actions of glucocorticoids on immune cells. *FASEB J., 16*(1), 61-71.

Garcia, D. A., Johnson, T. A., Presman, D. M., Fettweis, G., Wagh, K., Rinaldi, L., . . . Upadhyaya, A. (2021a). An intrinsically disordered region-mediated

confinement state contributes to the dynamics and function of transcription factors. *Molecular Cell*.

Garcia, D. A., Presman, D. M., Fettweis, G., Paakinaho, V., Jarsinski, C., Upadhyaya, A., & Hager, G. L. (2021b). A new model for single-molecule tracking analysis of transcription factor dynamics. *Nucleic Acids Res*.

Gautier, A., Juillerat, A., Heinis, C., Correa, I. R., Jr., Kindermann, M., Beaufils, F., & Johnsson, K. (2008). An engineered protein tag for multiprotein labeling in living cells. *Chem.Biol., 15*(2), 128-136.

Gebhardt, J. C., Suter, D. M., Roy, R., Zhao, Z. W., Chapman, A. R., Basu, S., . . . Xie, X. S. (2013). Single-molecule imaging of transcription factor binding to DNA in live mammalian cells. *Nat Methods, 10*(5), 421-426.

Geertz, M., Shore, D., & Maerkl, S. J. (2012). Massively parallel measurements of molecular interaction kinetics on a microfluidic platform. *Proceedings of the National Academy of Sciences*, 16540-16545.

Gest, C., Joimel, U., Huang, L., Pritchard, L. L., Petit, A., Dulong, C., . . . Soria, C. (2013). Rac3 induces a molecular pathway triggering breast cancer cell aggressiveness: differences in MDA-MB-231 and MCF-7 breast cancer cell lines. *BMC Cancer*(13), 63.

Gibson, B. A., Doolittle, L. K., Schneider, M. W. G., Jensen, L. E., Gamarra, N., Henry, L., . . . Rosen, M. K. (2019). Organization of Chromatin by Intrinsic and Regulated Phase Separation. *Cell, 179*(2), 470-484.e421. doi:10.1016/j.cell.2019.08.037

Gillespie, D. (1977). Exact stochastic simulation of coupled chemical reactions. *J Phys Chem, 81*(25), 2340-2361.

Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R., & Lieb, J. D. (2007). FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res., 17*, 877-885.

Glass, C. K., & Rosenfeld, M. G. (2000). The coregulator exchange in transcriptional functions of nuclear receptors. *Genes Development, 14*, 20.

Goldberg, A. D., Allis, C. D., & Bernstein, E. (2007). Epigenetics: a landscape takes shape. *Cell, 128*(4), 635-638.

Goldstein, I., Baek, S., Presman, D. M., Paakinaho, V., Swinstead, E. E., & Hager, G. L. (2017). Transcription factor assisted loading and enhancer dynamics dictate the hepatic fasting response. *Genome Res., 27*(3), 427-439. doi:10.1101/gr.212175.116

Goldstein, I., & Hager, G. L. (2018a). Dynamic enhancer function in the chromatin context. *Wiley Interdiscip Rev Syst Biol Med, 10*(1), e1390. doi:10.1002/wsbm.1390

Goldstein, I., & Hager, G. L. (2018b). The Three Ds of Transcription Activation by Glucagon: Direct, Delayed, and Dynamic. *Endocrinology, 159*(1), 206-216. doi:10.1210/en.2017-00521

Göttlicher, M., Heck, S., & Herrlich, P. (1998). Transcriptional cross-talk, the second mode of steroid hormone receptor action. *Journal of Molecular Medicine 76*, 480-489.

Gregor, T., Tank, D. W., Wieschaus, E. F., & Bialek, W. (2007). Probing the limits to positional information. *Cell, 130*(1), 153-164. doi:10.1016/j.cell.2007.05.025

Grimm, J. B., English, B. P., Chen, J., Slaughter, J. P., Zhang, Z., Revyakin, A., . . . Lavis, L. D. (2015). A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nat Methods, 12*(3), 244-250.

Grimm, J. B., English, B. P., Choi, H., Muthusamy, A. K., Mehl, B. P., Dong, P., . . . Lavis, L. D. (2016). Bright photoactivatable fluorophores for single-molecule imaging. *Nat Methods, 13*(12), 985-988. doi:10.1038/nmeth.4034

Gruber, C. J., Tschugguel, W., Schneeberger, C., & Huber, J. C. (2002). Production and actions of estrogens. *New England Journal of Medicine, 346*, 12.

Grünwald, D., Spottke, B., Buschmann, V., & Kubitscheck, U. (2006). Intranuclear binding kinetics and mobility of single native U1 snRNP particles in living cells. *Mol Biol Cell, 17*(12), 5017-5027. doi:10.1091/mbc.e06-06-0559

Guck, J., Schinkinger, S., Lincoln, B., Wottawah, F., Ebert, S., Romeyke, M., . . . Käs, J. (2005). Optical deformability as an inherent cell marker for testingmalignant

transformation and metastatic competence. *Biophysical Journal*(88), 3689-3698.

Guilluy, C., Osborne, L. D., Van Landeghem, L., Sharek, L., Superfine, R., Garcia-Mata, R., & Burridge, K. (2014). Isolated nuclei adapt to force and reveal a mechanotransduction pathway in the nucleus. *Nat Cell Biol, 16*(4), 376-381. doi:10.1038/ncb2927

Guo, Y. E., Manteiga, J. C., Henninger, J. E., Sabari, B. R., Dall'Agnese, A., Hannett, N. M., . . . Abraham, B. J. (2019). Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature*(572), 543-548.

Gurdon, J. B., Javed, K., Vodnala, M., & Garrett, N. (2020). Long-term association of a transcription factor with its chromatin binding site can stabilize gene expression and cell fate commitment. *Proc Natl Acad Sci U S A*. doi:10.1073/pnas.2000467117

Gustafsson, J.-A. (2016). Historical overview of nuclear receptors. *The Journal of Steroid Biochemistry and Molecular Biology, 157*, 3.

Gwosch, K. C., Pape, J. K., Balzarotti, F., Hoess, P., Ellenberg, J., Ries, J., & Hell, S. W. (2020). MINFLUX nanoscopy delivers 3D multicolor nanometer resolution in cells. *Nat Methods*(17), 217-224.

Hager, G. L., Fletcher, T. M., Xiao, N., Baumann, C. T., M:uller, W. G., & McNally, J. G. (2000). Dynamics of gene targeting and chromatin remodeling by nuclear receptors. In: Steroid Receptor Coactivators and the Remodeling of Chromatin. *Biochemical Society Transactions, 28*(4), 405-410.

Hager, G. L., McNally, J. G., & Misteli, T. (2009). Transcription dynamics. *Mol.Cell, 35*(6), 741-753.

Hall, M. S., Alisafaei, F., Ban, E., Feng, X., Hui, C. Y., Shenoy, V. B., & Wu, M. (2016). Fibrous nonlinear elasticity enables positive Mechanical feedback between cells and ECMs. In *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 113, pp. 14043-14048): National Academy of Sciences.

Hamilton, K. J., Arao, Y., & Korach, K. S. (2014). Estrogen hormone physiology: reproductive findings from estrogen receptor mutant mice. *Reproductive Biology, 14*, 3-8.

Han, Y. L., Pegoraro, A. F., Li, H., Li, K., Yuan, Y., Xu, G., . . . Li, Y. (2020). Cell swelling, softening and invasion in a three-dimensional breast cancer model. . *Nature Physics*, 101-108.

Han, Y. L., Ronceray, P., Xu, G., Malandrino, A., Kamm, R. D., Lenz, M., . . . Guo, M. (2018). Cell contraction induces long-ranged stress stiffening in the extracellular matrix. In *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 115, pp. 4075-4080): National Academy of Sciences.

Handorf, A. M., Zhou, Y., Halanski, M. A., & Li, W. J. (2015). Tissue stiffness dictates development, homeostasis, and disease progression. *Organogenesis, 11*(1), 1-15. doi:10.1080/15476278.2015.1019687

Hansen, A. S., Amitai, A., Cattoglio, C., Tjian, R., & Darzacq, X. (2020). Guided nuclear exploration increases CTCF target search efficiency. *Nat Chem Biol, 16*(3), 257-266. doi:10.1038/s41589-019-0422-3

Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R., & Darzacq, X. (2017). CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife, 6*. doi:10.7554/eLife.25776

Hansen, J. C. (2002). Conformational dynamics of the chromatin fiber in solution: determinants, mechanisms, and functions. *Annu. Re. Biophys. Biomol. Struct*(31), 361-392.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., . . . Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol.Cell, 38*(4), 576-589.

Heldin, C. H., Rubin, K., Pietras, K., & Östman, A. (2004). High interstitial fluid pressure—an obstacle in cancer therapy. *Nature REviews Cancer*, 806-813.

Hipp, L., Beer, J., Kuchler, O., Reisser, M., Sinske, D., Michaelis, J., . . . Knöll, B. (2019). Single-molecule imaging of the transcription factor SRF reveals prolonged chromatin-binding kinetics upon cell stimulation. *Proc Natl Acad Sci U S A, 116*(3), 880-889. doi:10.1073/pnas.1812734116

Hjort, N. L. (2010). *Bayesian nonparametrics*. Cambridge, UK ; New York: Cambridge University Press.

Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K., & Sharp, P. A. (2017). A Phase Separation Model for Transcriptional Control. *Cell, 169*(1), 13-23. doi:10.1016/j.cell.2017.02.007

Hoffmann, J. M., & Partridge, L. (2015). Nuclear hormone receptors: Roles of xenobiotic detoxification and sterol homeostasis in healthy aging. *Critical Reviews in Biochemistry and Molecular Biology, 50*(5), 12.

Hollenberg, S. M., & Evans, R. M. (1988). Multiple and cooperative trans-activation domains of the human glucocorticoid receptor. *Cell, 55*, 899-906.

Howell, A. (2008). The endocrine prevention of breast cancer. *Best Pract. Res. Clin. Endocrinol. Metab.*, 615-623.

Huang, B., Wang, W., Bates, M., & Zhuang, X. (2008). Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science, 319*(5864), 810-813. doi:10.1126/science.1153529

Hug, C. B., & Vaquerizas, J. M. (2018). The Birth of the 3D Genome during Early Embryonic Development. *Trends Genet, 34*(12), 903-914. doi:10.1016/j.tig.2018.09.002

Hyman, A. A., Weber, C. A., & Jülicher, F. (2014). Liquid-liquid phase separation in biology. *Annu Rev Cell Dev Biol, 30*, 39-58. doi:10.1146/annurev-cellbio-100913-013325

Innis, J. W., Mortlock, D., Chen, Z., Ludwig, M., Williams, M. E., Williams, T. M., . . . Utsch, B. (2004). Polyalanine expansion in HOXA13: three new affected families and the molecular consequences in a mouse model. *Hum Mol Genet, 13*(22), 2841-2851. doi:10.1093/hmg/ddh306

Izeddin, I., Recamier, V., Bosanac, L., Cisse, II, Boudarene, L., Dugast-Darzacq, C., . . . Darzacq, X. (2014). Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus. *Elife, 3*, e02230. doi:10.7554/eLife.02230

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *An introduction to statistical learning : with applications in R*.

Jameson, J. L., Fauci, A. S., Kasper, D. L., Hauser, S. L., Longo, D. L., & Loscalzo, J. (2018). *Harrison's Principles of Internal Medicine*. New York: McGraw-Hill Education.

Janissen, R., Eslami-Mossallam, B., Artsimovitch, I., Depken, M., & Dekker, N. H. (2020). A unifying mechanistic model of bacterial transcription pith three interconnected pause states and non-diffusive backtrack recovery. *Biophysical Journal, 118*(3, Supplement 1), 543a. doi:https://doi.org/10.1016/j.bpj.2019.11.2974

Jaqaman, K., Loerke, D., Mettlen, M., Kuwata, H., Grinstein, S., Schmid, S. L., & Danuser, G. (2008). Robust single-particle tracking in live-cell time-lapse sequences. *Nat. Methods, 5*(8), 695-702. doi:10.1038/nmeth.1237

Jaynes, E. T., & Bretthorst, G. L. (2019). *Probability theory : the logic of science*.

Jensen, E. V., & Jordan, V. C. (2003). The estrogen receptor: A model for molecular medicine. *Clin. Cancer Res.*, 1980-1989.

Jewell, C. M., Scoltock, A. B., Hamel, B. L., Yudt, M. R., & Cidlowski, J. A. (2012). Complex human glucocorticoid receptor dim mutations define glucocorticoid induced apoptotic resistance in bone cells. *Mol Endocrinol, 26*(2), 244-256. doi:10.1210/me.2011-1116

John, S., Sabo, P. J., Johnson, T. A., Sung, M. H., Biddie, S. C., Lightman, S. L., . . . Hager, G. L. (2008). Interaction of the glucocorticoid receptor with the global chromatin landscape. *Mol.Cell, 29*, 611-624.

Johnson, D. S., Mortazavi, A., Myers, R. M., & Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science, 316*(5830), 1497-1502.

Johnson, T. A., Chereji, R. V., Stavreva, D. A., Morris, S. A., Hager, G. L., & Clark, D. J. (2018). Conventional and pioneer modes of glucocorticoid receptor interaction with enhancer chromatin in vivo. *Nucleic Acids Res, 46*, 203-214. doi:10.1093/nar/gkx1044

Johnston, S. R., & Dowsett, M. (2003). Aromatase inhibitors for breast cancer: lessons from the laboratory. *Nat Rev Cancer, 3*(11), 821-831. doi:10.1038/nrc1211

Jordan, V. C. (2003). Tamoxifen: a most unlikely pioneering medicine. *Nat Rev Drug Discov, 2*(3), 205-213. doi:10.1038/nrd1031

Jothi, R., Cuddapah, S., Barski, A., Cui, K., & Zhao, K. (2008). Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. *Nucleic Acids Res., 36*(16), 5221-5231.

Kadmiel, M., & Cidlowski, J. A. (2013). Glucocorticoid receptor signaling in health and disease. *Trends in Pharmacological Sciences, 34*, 12.

Kalli, M., & Stylianopoulos, T. (2018). Defining the Role of Solid Stress and Matrix Stiffness in Cancer Cell Proliferation and Metastasis. *Front Oncol, 8*, 55. doi:10.3389/fonc.2018.00055

Kampmann, M. (2005). Facilitated diffusion in chromatin lattices: mechanistic diversity and regulatory potential. *Mol Microbiol, 57*(4), 889-899. doi:10.1111/j.1365-2958.2005.04707.x

Kaplan, N., Moore, I. K., Fondufe-Mittendorf, Y., Gossett, A. J., Tillo, D., Field, Y., . . . Segal, E. (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature, 458*(7236), 362-366.

Kato, M., Han, T. W., Xie, S., Shi, K., Du, X., Wu, L. C., . . . Grishin, N. V. (2012). Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. (149), 753-767.

Kent, S., Brown, K., Yang, C. H., Alsaihati, N., Tian, C., Wang, H., & Ren, X. (2020). Phase-Separated Transcriptional Condensates Accelerate Target-Search Process Revealed by Live-Cell Single-Molecule Imaging. *Cell Rep, 33*(2), 108248. doi:10.1016/j.celrep.2020.108248

Khan, S. H., Awasthi, S., Guo, C., Goswami, D., Ling, J., Griffin, P. R., . . . Kumar, R. (2012). Binding of the N-terminal region of coactivator TIF2 to the intrinsically disordered AF1 domain of the glucocorticoid receptor is accompanied by conformational reorganizations. *J Biol Chem, 287*(53), 44546-44560.

Kharchenko, P. V., Woo, C. J., Tolstorukov, M. Y., Kingston, R. E., & Park, P. J. (2008). Nucleosome positioning in human HOX gene clusters. *Genome Res., 18*(10), 1554-1561.

Kilic, S., Bachmann, A. L., Bryan, L. C., & Fierz, B. (2015). Multivalency governs HP1alpha association dynamics with the silent chromatin state. *Nat Commun, 6*, 7313. doi:10.1038/ncomms8313
Kim, N. G., & Gumbiner, B. M. (2015). Adhesion to fibronectin regulates Hippo signaling via the FAK-Src-PI3K pathway. *J Cell Biol, 210*(3), 503-515. doi:10.1083/jcb.201501025
Kim, S., & Shendure, J. (2019). Mechanisms of Interplay between Transcription Factors and the 3D Genome. *Mol Cell, 76*(2), 306-319. doi:10.1016/j.molcel.2019.08.010
Kimura, H., & Cook, P. R. (2001). Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. *J.Cell Biol., 153*(7), 1341-1353.
Kirby, T. J., & Lammerding, J. (2018). Emerging views of the nucleus as a cellular mechanosensor. *Nat Cell Biol*. doi:10.1038/s41556-018-0038-y
Klemm, S. L., Shipony, Z., & Greenleaf, W. J. (2019). Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet, 20*(4), 207-220. doi:10.1038/s41576-018-0089-8
Klosin, A., & Hyman, A. A. (2017). A liquid reservoir for silent chromatin. *Nature, 547*(7662), 168-170. doi:10.1038/nature23089
Koo, P. K., & Mochrie, S. G. (2016). Systems-level approach to uncovering diffusive states and their transitions from single-particle trajectories. *Phys Rev E, 94*(5-1), 052412. doi:10.1103/PhysRevE.94.052412
Koo, P. K., Weitzman, M., Sabanaygam, C. R., van Golen, K. L., & Mochrie, S. G. (2015). Extracting Diffusive States of Rho GTPase in Live Cells: Towards In Vivo Biochemistry. *PLoS Comput Biol, 11*(10), e1004297. doi:10.1371/journal.pcbi.1004297
Kornberg, R. D. (1974). Chromatin structure: a repeating unit of histones and DNA. *Science, 184*(4139), 868-871.
Kornberg, R. D., & Lorch, Y. (1999). Chromatin-modifying and -remodeling complexes. *Curr.Opin.Genet.Dev., 9*(2), 148-151.
Kornberg, R. D., & Thomas, J. O. (1974). Chromatin structure; oligomers of the histones. *Science, 184*, 865-868.
Koslover, E. F., Fuller, C. J., Straight, A. F., & Spakowitz, A. J. (2010). Local geometry and elasticity in compact chromatin structure. *Biophys J, 99*(12), 3941-3950. doi:10.1016/j.bpj.2010.10.024
Kudalkar, E. M., Davis, T. N., & Asbury, C. L. (2016). Single-Molecule Total Internal Reflection Fluorescence Microscopy. *Cold Spring Harb Protoc, 2016*(5). doi:10.1101/pdb.top077800
Kumar, R., & Litwack, G. (2009). Structural and functional relationships of the steroid hormone receptors' N-terminal transactivation domain. *Steroids, 74*(12), 877-883. doi:10.1016/j.steroids.2009.07.012
Kumar, R., Zakharov, M. N., Khan, S. H., Miki, R., Jang, H., Toraldo, G., . . . Jasuja, R. (2011). The dynamic structure of the estrogen receptor. *Journal of Aminoacids*, 812540.
Kuser-Abali, G., Alptekin, A., Lewis, M., Garraway, I. P., & Cinar, B. (2015). YAP1 and AR interactions contribute to the switch from androgen-

dependent to castration-resistant growth in prostate cancer. *Nat Commun, 6*, 8126. doi:10.1038/ncomms9126
Lacroix, M., & Leclercq, G. (2004). Relevance of breast cancer cell lines as models for breast tumours: an update. *Breast Cancer Res.Treat., 83*(3), 249-289.
Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., . . . Weirauch, M. T. (2018). The Human Transcription Factors. *Cell, 172*(4), 650-665. doi:10.1016/j.cell.2018.01.029
Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods, 9*(4), 357-359. doi:10.1038/nmeth.1923
Lazar, M. A. (2017). Maturing of the nuclear receptor family. *J Clin Invest, 127*(4), 1123-1125. doi:10.1172/jci92949
Lee, H. P., Stowers, R., & Chaudhuri, O. (2019). Volume expansion and TRPV4 activation regulate stem cell fate in three-dimensional microenvironments. *Nat Commun, 10*(1), 529. doi:10.1038/s41467-019-08465-x
Lerner, J., Gomez-Garcia, P. A., McCarthy, R. L., Liu, Z., Lakadamyali, M., & Zaret, K. S. (2020). Two-Parameter Mobility Assessments Discriminate Diverse Regulatory Factor Behaviors in Chromatin. *Mol Cell*. doi:10.1016/j.molcel.2020.05.036
Lettice, L. A., Heaney, S. J., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., . . . de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet, 12*(14), 1725-1735.
Levental, K. R., Yu, H., Kass, L., Lakins, J. N., Egeblad, M., Erler, J. T., . . . Yamauchi, M. (2009). Matrix crosslinking forces tumor progression by enhancing integrin signaling. *Cell, 891-906*.
Levi, V., & Gratton, E. (2007). Exploring dynamics in living cells by tracking single particles. *Cell Biochem Biophys, 48*(1), 1-15. doi:10.1007/s12013-007-0010-0
Levine, M., Cattoglio, C., & Tjian, R. (2014). Looping back to leap forward: transcription enters a new era. *Cell, 157*(1), 13-25. doi:10.1016/j.cell.2014.02.009
Li, Q., Gao, H., Yang, H., Wei, W., & Jiang, Y. (2019). Estradiol promotes the progression of ER+ breast cancer through methylation-mediated RSK4 inactivation. *Onco Targets Therapy, 12*, 5907-5916.
Li, S., Prasanna, X., Salo, V. T., Vattulainen, I., & Ikonen, E. (2019). An efficient auxin-inducible degron system with low basal degradation in human cells. *Nat Methods, 16*(9), 866-869. doi:10.1038/s41592-019-0512-x
Liao, J. C., Spudich, J. A., Parker, D., & Delp, S. L. (2007). Extending the absorbing boundary method to fit dwell-time distributions of molecular motors with complex kinetic pathways. *Proc Natl Acad Sci U S A, 104*(9), 3171-3176. doi:10.1073/pnas.0611519104
Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., . . . Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science, 326*(5950), 289-293.

Lim, H. W., Uhlenhaut, N. H., Rauch, A., Weiner, J., Hubner, S., Hubner, N., . . . Steger, D. J. (2015). Genomic redistribution of GR monomers and dimers mediates transcriptional response to exogenous glucocorticoid in vivo. *Genome Res., 25*(6), 836-844. doi:10.1101/gr.188581.114

Lippman, M. E., & Bolan, G. (1975). Oestrogen-responsive human breast cancer in long term tissue culture. *Nature, 256*(5518), 592-593. doi:10.1038/256592a0

Lit, L. C., Scott, S., Zhang, H., Stebbing, J., Photiou, A., & Giamas, G. (2013). LATS2 is a modulator of estrogen receptor alpha. *Anticancer Res, 33*(1), 53-63. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/23267128

Liu, Z., Merkurjev, D., Yang, F., Li, W., Oh, S., Friedman, M. J., . . . Rosenfeld, M. G. (2014). Enhancer Activation Requires trans-Recruitment of a Mega Transcription Factor Complex. *Cell, 159*(2), 358-373.

Liu, Z., & Tjian, R. (2018). Visualizing transcription factor dynamics in living cells. *J Cell Biol, 217*(4), 1181-1191. doi:10.1083/jcb.201710038

Loffreda, A., Jacchetti, E., Antunes, S., Rainone, P., Daniele, T., Morisaki, T., . . . Mazza, D. (2017). Live-cell p53 single-molecule binding is modulated by C-terminal acetylation and correlates with transcriptional activity. *Nat Commun, 8*(1), 313. doi:10.1038/s41467-017-00398-7

Lonai'd, D. M., & O'Malley, B. W. (2012). Nuclear receptor coregulators: Modulators of pathology and therapeutic targets. *Nature Reviews Endocrinology, 8*, 6.

Long, H. K., Prescott, S. L., & Wysocka, J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell, 167*(5), 1170-1187. doi:10.1016/j.cell.2016.09.018

Lu, H., Yu, D., Hansen, A. S., Ganguly, S., Liu, R., Heckert, A., . . . Zhou, Q. (2018). Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. *Nature*. doi:10.1038/s41586-018-0174-3

Lu, Y., Wu, T., Gutman, O., Lu, H., Zhou, Q., Henis, Y. I., & Luo, K. (2020). Phase separation of TAZ compartmentalizes the transcription machinery to promote gene expression. *Nat Cell Biol, 22*(4), 453-464. doi:10.1038/s41556-020-0485-0

Luecke, H. F., & Yamamoto, K. R. (2005). The glucocorticoid receptor blocks P-TEFb recruitment by NFkappaB to effect promoter-specific transcriptional repression. *Genes Dev., 19*(9), 1116-1127.

Luger, K., Rechsteiner, T. J., Flaus, A. J., Waye, M. M., & Richmond, T. J. (1997). Characterization of nucleosome core particles containing histone proteins made in bacteria. *J.Mol.Biol., 272*, 301-311.

Luisi, B. F., Xu, W. X., Otwinowski, Z., Freedman, L. P., Yamamoto, K. R., & Sigler, P. B. (1991). Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA [see comments]. *Nature, 352*, 497-505.

Lupianez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., . . . Mundlos, S. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell, 161*(5), 1012-1025. doi:10.1016/j.cell.2015.04.004

Maeshima, K., Ide, S., Hibino, K., & Sasai, M. (2016). Liquid-like behavior of chromatin. *Current opinion in genetics and development*(37), 36-45.
Makhija, E., Jokhun, D. S., & Shivashankar, G. V. (2016). Nuclear deformability and telomere dynamics are regulated by cell geometric constraints. *Proc Natl Acad Sci U S A, 113*(1), E32-40. doi:10.1073/pnas.1513189113
Malek, G., & Lad, E. M. (2014). Emerging roles for nuclear receptors in the pathogenesis of age-related macular degeneration. *Cellular and Molecular Life Sciences, 71*, 19.
Mangelsdorf, D. J., Thummel, C., Beato, M., Herrlich, P., Schutz, G., Umesono, K., . . . Evans, R. M. (1995). The nuclear receptor superfamily: the second decade. *Cell, 83*, 4.
Manzo, C., & Garcia-Parajo, M. F. (2015). A review of progress in single particle tracking: from methods to biophysical insights. *Rep Prog Phys, 78*(12), 124601. doi:10.1088/0034-4885/78/12/124601
Marklund, E. G., Mahmutovic, A., Berg, O. G., Hammar, P., van der Spoel, D., Fange, D., & Elf, J. (2013). Transcription-factor binding and sliding on DNA studied using micro- and macroscopic models. *Proc Natl Acad Sci U S A, 110*(49), 19796-19801. doi:10.1073/pnas.1307905110
Masri, S., Phung, S., Wang, X., Wu, X., Yuan, Y. C., Wagman, L., & Chen, S. (2008). Genome-wide analysis of aromatase inhibitor-resistant, tamoxifen-resistant, and long-term estrogen-deprived cells reveals a role for estrogen receptor. *Cancer Res, 68*(12), 4910-4918. doi:10.1158/0008-5472.CAN-08-0303
Mazza, D., Abernathy, A., Golob, N., Morisaki, T., & McNally, J. G. (2012a). A benchmark for chromatin binding measurements in live cells. *Nucleic Acids Res, 40*(15), e119.
Mazza, D., Ganguly, S., & McNally, J. G. (2013). Monitoring dynamic binding of chromatin proteins in vivo by single-molecule tracking. *Methods Mol Biol, 1042*, 117-137.
Mazza, D., Stasevich, T. J., Karpova, T. S., & McNally, J. G. (2012b). Monitoring dynamic binding of chromatin proteins in vivo by fluorescence correlation spectroscopy and temporal image correlation spectroscopy. *Methods Mol Biol, 833*, 177-200.
McNally, J. G., Mueller, W. G., Walker, D., Wolford, R. G., & Hager, G. L. (2000). The glucocorticoid receptor: Rapid exchange with regulatory sites in living cells. *Science, 287*(5456), 1262-1265.
Meaburn, K. J., & Misteli, T. (2008). Locus-specific and activity-independent gene repositioning during early tumorigenesis. *J Cell Biol., 180*(1), 39-50.
Mehta, G. D., Ball, D. A., Eriksson, P. R., Chereji, R. V., Clark, D. J., McNally, J. G., & Karpova, T. S. (2018). Single-Molecule Analysis Reveals Linked Cycles of RSC Chromatin Remodeling and Ace1p Transcription Factor Binding in Yeast. *Mol Cell, 72*(5), 875-887.e879. doi:10.1016/j.molcel.2018.09.009
Meijsing, S. H., Elbi, C., Luecke, H. F., Hager, G. L., & Yamamoto, K. R. (2007). The ligand binding domain controls glucocorticoid receptor dynamics independent of ligand release. *Mol.Cell Biol., 27*(7), 2442-2451.

Metivier, R., Penot, G., Hubner, M. R., Reid, G., Brand, H., Kos, M., & Gannon, F. (2003). Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell, 115*(6), 751-763.

Metzler, R., Jeon, J. H., Cherstvy, A. G., & Barkai, E. (2014). Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking. *Phys Chem Chem Phys, 16*(44), 24128-24164. doi:10.1039/c4cp03465a

Michieletto, D., & Gilbert, N. (2019). Role of nuclear RNA in regulating chromatin structure and transcription. *Curr Opin Cell Biol, 58*, 120-125. doi:10.1016/j.ceb.2019.03.007

Mikuni, S., Tamura, M., & Kinjo, M. (2007). Analysis of intranuclear binding process of glucocorticoid receptor using fluorescence correlation spectroscopy. *FEBS Lett., 581*(3), 389-393.

Mikuni, S., Yamamoto, J., Horio, T., & Kinjo, M. (2017). Negative Correlation between the Diffusion Coefficient and Transcriptional Activity of the Glucocorticoid Receptor. *Int J Mol Sci, 18*(9). doi:10.3390/ijms18091855

Miller, J. H., & Reznikoff, W. S. (1980). The Operon. *Cold Spring Harbor Laboratory*.

Miller, T. W., Balko, J. M., Fox, E. M., Ghazoui, Z., Dunbier, A., Anderson, H., . . . Arteaga, C. L. (2011). ERalpha-dependent E2F transcription can mediate resistance to estrogen deprivation in human breast cancer. *Cancer Discov, 1*(4), 338-351. doi:10.1158/2159-8290.Cd-11-0101

Mitrea, D. M., & Kriwacki, R. W. (2016). Phase separation in biology; functional organization of a higher order. *Cell Commun. Signal*(14), 1.

Morgan, M. A., & Shilatifard, A. (2015). Chromatin signatures of cancer. . *Genes Dev*, 238-249.

Morisaki, T., Muller, W. G., Golob, N., Mazza, D., & McNally, J. G. (2014). Single-molecule analysis of transcription factor binding at transcription sites in live cells. *Nat Commun., 5*, 4456.

Muchardt, C., & Yaniv, M. (1993). A human homologue of Saccharomyces cerevisiae SNF2/SWI2 and Drosophila brm genes potentiates transcriptional activation by the glucocorticoid receptor. *EMBO J., 12*(11), 4279-4290.

Mueller, F., Stasevich, T. J., Mazza, D., & McNally, J. G. (2013). Quantifying transcription factor kinetics: at work or at play? *Crit Rev Biochem Mol Biol, 48*(5), 492-514. doi:10.3109/10409238.2013.833891

Muragaki, Y., Mundlos, S., Upton, J., & Olsen, B. R. (1996). Altered growth and branching patterns in synpolydactyly caused by mutations in HOXD13. *Science, 272*(5261), 548-551. doi:10.1126/science.272.5261.548

Nadji, M., Gomez-Fernandez, C., Ganjei-Azar, P., & Morales, A. R. (2005). Immunohistochemistry of es- trogen and progesterone receptors reconsid-ered: experience with 5,993 breast cancers. *Am J Clin Pathol*(123), 21-27.

Nakahashi, H., Kwon, K. R., Resch, W., Vian, L., Dose, M., Stavreva, D., . . . Casellas, R. (2013). A genome-wide map of CTCF multivalency redefines the CTCF code. *Cell Rep., 3*(5), 1678-1689.

Nakamura, K., Jeong, S. Y., Uchihara, T., Anno, M., Nagashima, K., Nagashima, T., . . . Kanazawa, I. (2001). SCA17, a novel autosomal dominant cerebellar ataxia caused by an expanded polyglutamine in TATA-binding protein. *Hum Mol Genet, 10*(14), 1441-1448. doi:10.1093/hmg/10.14.1441

Nelson, L. R., & Bulun, S. E. (2001). Estrogen production and action. *Journal of American Academy of Dermatology, 45*, 8.

Newby, J. M., Schaefer, A. M., Lee, P. T., Forest, M. G., & Lai, S. K. (2018). Convolutional neural networks automate detection for tracking of submicron-scale particles in 2D and 3D. *Proc Natl Acad Sci U S A*(115), 9026-9031.

Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics, 46*, 323-351. Retrieved from https://ui.adsabs.harvard.edu/\#abs/2005ConPh..46..323N

Nissen, R. M., & Yamamoto, K. R. (2000). The glucocorticoid receptor inhibits NFkappaB by interfering with serine-2 phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev., 14*(18), 2314-2329.

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., . . . Heard, E. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature, 485*(7398), 381-385.

Normanno, D., Boudarene, L., Dugast-Darzacq, C., Chen, J., Richter, C., Proux, F., . . . Dahan, M. (2015). Probing the target search of DNA-binding proteins in mammalian cells using TetR as model searcher. *Nat Commun, 6*, 7357. doi:10.1038/ncomms8357

O'Neill, L. P., & Turner, B. M. (2003). Immunoprecipitation of native chromatin: NChIP. *Methods, 31*(1), 76-82.

Onuki, A. (2007). *Phase transition dynamics*. Cambridge: Cambridge University Press.

Orlando, V. (2000). Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. *Trends Biochem.Sci., 25*(3), 99-104.

Paakinaho, V., Johnson, T. A., Presman, D. M., & Hager, G. L. (2019a). Glucocorticoid receptor quaternary structure drives chromatin occupancy and transcriptional outcome. *Genome Res., 29*(8), 1223-1234. doi:10.1101/gr.244814.118

Paakinaho, V., Presman, D. M., Ball, D. A., Johnson, T. A., Schiltz, R. L., Levitt, P., . . . Hager, G. L. (2017). Single-molecule analysis of steroid receptor and cofactor action in living cells. *Nat Commun, 8*, 15896. doi:10.1038/ncomms15896

Paakinaho, V., Swinstead, E. E., Presman, D. M., Grontved, L., & Hager, G. L. (2019b). Meta-analysis of Chromatin Programming by Steroid Receptors. *Cell Rep, 28*(13), 3523-3534.e3522. doi:10.1016/j.celrep.2019.08.039

Panciera, T., Azzolin, L., Cordenonsi, M., & Piccolo, S. (2017). Mechanobiology of YAP and TAZ in physiology and disease. *Nat Rev Mol Cell Biol, 18*(12), 758-770. doi:10.1038/nrm.2017.87

Park, J., Kim, P., Helen, W., Engler, A. J., Levchenko, A., & Kim, D. H. (2012). Control of stem cell fate and function by engineering physical microenvironments. *Integrative biology*, 1008-1018.

Park, P. J. (2009). ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet, 10*(10), 669-680. doi:10.1038/nrg2641

Parslow, A., Cardona, A., & Bryson-Richardson, R. J. (2014). Sample drift correction following 4D confocal time-lapse imaging. *J Vis Exp*(86). doi:10.3791/51086

Paszek, M. J., Zahir, N., Johnson, K. R., Lakins, J. N., Rozenberg, G. I., Gefen, A., . . . Weaver, V. M. (2005). Tensional homeostasis and the malignant phenotype. *Cancer Cell, 8*(3), 241-254. doi:10.1016/j.ccr.2005.08.010

Peng, Z., Mizianty, M. J., Xue, B., Kurgan, L., & Uversky, V. N. (2012). More than just tails: intrinsic disorder in histone proteins. *Mol Biosyst, 8*(7), 1886-1901. doi:10.1039/c2mb25102g

Picard, D., Khursheed, B., Garabedian, M. J., Fortin, M. G., Lindquist, S., & Yamamoto, K. R. (1990). Reduced levels of hsp90 compromise steroid receptor action in vivo *Nature, 348*, 166-168.

Piperigkou, Z., & Karamanos, N. K. (2020). Estrogen receptor-mediated targeting of the extracellular matrix network in cancer. *Semin Cancer Biol, 62*, 116-124.

Plank, J. L., & Dean, A. (2014). Enhancer Function: Mechanistic and Genome-Wide Insights Come Together. *Mol.Cell, 55*(1), 5-14.

Plodinec, M., & Lim, R. Y. (2015). Nanomechanical characterization of living mammary tissues by atomic force microscopy. *Methods Mol Biol, 1293*, 231-246. doi:10.1007/978-1-4939-2519-3_14

Plodinec, M., Loparic, M., Monnier, C. A., Obermann, E. C., Zanetti-Dallenbach, R., Oertle, P., . . . Schoenenberger, C. A. (2012). The nanomechanical signature of breast cancer. *Nat Nanotechnol, 7*(11), 757-765. doi:10.1038/nnano.2012.167

Poirier, M. G., Bussiek, M., Langowski, J., & Widom, J. (2008). Spontaneous access to DNA target sites in folded chromatin fibers. *J Mol Biol, 379*, 772-786.

The role of the extracellular matrix and its molecular and cellular regulators in cancer cell plasticity, 8, Frontiers Media S.A. 431 (2018).

Pombo, A., & Dillon, N. (2015). Three-dimensional genome architecture: players and mechanisms. *Nat Rev.Mol.Cell Biol., 16*(4), 245-257.

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

Presman, D. M., Ball, D. A., Paakinaho, V., Grimm, J. B., Lavis, L. D., Karpova, T. S., & Hager, G. L. (2017). Quantifying transcription factor binding dynamics at the single-molecule level in live cells. *Methods, 123*, 76-88. doi:10.1016/j.ymeth.2017.03.014

Presman, D. M., Ganguly, S., Schiltz, R. L., Johnson, T. A., Karpova, T. S., & Hager, G. L. (2016). DNA binding triggers tetramerization of the

glucocorticoid receptor in live cells. *Proc Natl Acad Sci U S A, 113*(29), 8236-8241. doi:10.1073/pnas.1606774113
Presman, D. M., & Hager, G. L. (2017). More than meets the dimer: What is the quaternary structure of the glucocorticoid receptor? *Transcription, 8*(1), 32-39. doi:10.1080/21541264.2016.1249045
Presman, D. M., Ogara, M. F., Stortz, M., Alvarez, L. D., Pooley, J. R., Schiltz, R. L., . . . Pecci, A. (2014). Live cell imaging unveils multiple domain requirements for in vivo dimerization of the Glucocorticoid Receptor. *PLoS Biol, 12*(3), e1001813.
Qian, H., Sheetz, M. P., & Elson, E. L. (1991). Single particle tracking. Analysis of diffusion and flow in two-dimensional systems. *Biophys J, 60*(4), 910-921. doi:10.1016/s0006-3495(91)82125-7
Ranhotra, H. S. (2013). The orphan nuclear receptors in cancer and diabetes. *Journal of Receptors and Signal Transduction, 33*, 5.
Rao, S. S. P., Huang, S. C., Glenn St Hilaire, B., Engreitz, J. M., Perez, E. M., Kieffer-Kwon, K. R., . . . Aiden, E. L. (2017). Cohesin Loss Eliminates All Loop Domains. *Cell, 171*(2), 305-320.e324. doi:10.1016/j.cell.2017.09.026
Rastinejad, F., Huang, P., Chandra, V., & Khorasanizadeh, S. (2013). Understanding nuclear receptor form and function using structural biology. *Molecular Endocrinology, 51*, 20.
Rastogi, C., Rube, H. T., Kribelbauer, J. F., Crocker, J., Loker, R. E., Martini, G. D., . . . Mann, R. S. (2018). Accurate and sensitive quantification of protein-DNA binding affinity. *Proceedings of the National Academy of Sciences*, E3692-E3701.
Ratke, L., & Voorhees, P. W. (2011). *Growth and coarsening : Ostwald ripening in material processing*. Berlin; London: Springer.
Ratman, D., Mylka, V., Bougarne, N., Pawlak, M., Caron, S., Hennuyer, N., . . . De Bosscher, K. (2016). Chromatin recruitment of activated AMPK drives fasting response genes co-controlled by GR and PPARα. *Nucleic Acids Res, 44*(22), 10539-10553. doi:10.1093/nar/gkw742
Reingruber, J., & Holcman, D. (2011). Transcription factor search for a DNA promoter in a three-state model. *Phys Rev E Stat Nonlin Soft Matter Phys, 84*(2 Pt 1), 020901. doi:10.1103/PhysRevE.84.020901
Reisser, M., Hettich, J., Kuhn, T., Popp, A. P., Große-Berkenbusch, A., & Gebhardt, J. C. M. (2020). Inferring quantity and qualities of superimposed reaction rates from single molecule survival time distributions. *Sci Rep, 10*(1), 1758. doi:10.1038/s41598-020-58634-y
Reiter, F., Wienerroither, S., & Stark, A. (2017). Combinatorial function of transcription factors and cofactors. *Curr Opin Genet Dev, 43*, 73-81. doi:10.1016/j.gde.2016.12.007
Reits, E. A. J., & Neefjes, J. J. (2001). From Fixed to Frap: Measuring Protein Mobility and Activity in Living Cells. *Nat Cell Biol*(3), E145-E147.
Reznikoff, W. S. (2003). Tn5 as a model for understanding DNA transposition. *Mol Microbiol., 47*(5), 1199-1206.
Riaz, M., van Jaarsveld, M. T., Hollestelle, A., Prager-van der Smissen, W. J., Heine, A. A., Boersma, A. W., . . . Wiemer, E. A. (2013). miRNA

expression profiling of 51 human breast cancer cell lines reveals subtype and driver mutation-specific miRNAs. *Breast Cancer Res*(15), R33.

Robb, N. C., Cordes, T., Hwang, L. C., Gryte, K., Duchi, D., Craggs, T. D., . . . Kapanidis, A. N. (2013). The transcription bubble of the RNA polymerase–promoter open complex exhibits conformational heterogeneity and millisecond-scale dynamics: implications for transcription start-site selection. *J Mol Biol*, 875-885.

Rodriguez, J., Ren, G., Day, C. R., Zhao, K., Chow, C. C., & Larson, D. R. (2019). Intrinsic Dynamics of a Human Gene Reveal the Basis of Expression Heterogeneity. *Cell, 176*(1-2), 213-226.e218. doi:10.1016/j.cell.2018.11.026

Rogatsky, I., Wang, J. C., Derynck, M. K., Nonaka, D. F., Khodabakhsh, D. B., Haqq, C. M., . . . Yamamoto, K. R. (2003). Target-specific utilization of transcriptional regulatory surfaces by the glucocorticoid receptor. *Proc.Natl.Acad.Sci.USA, 100*(24), 13845-13850.

Rosenbluth, J. M., Schackmann, R. C. J., Gray, G. K., Selfors, L. M., Li, C. M. C., Boedicker, M., . . . Brugge, J. S. (2020). Organoid cultures from normal and cancer-prone human breast tissues preserve complex epithelial lineages. In *Nature Communications* (Vol. 11, pp. 1-14): Nature Research.

Rosowski, K. A., Sai, T., Vidal-Henriquez, E., Zwicker, D., Style, R. W., & Dufresne, E. R. (2020). Elastic ripening and inhibition of liquid–liquid phase separation. *Nature Physics*(16), 422-425.

Roux, K. J., Crisp, M. L., Liu, Q., Kim, D., Kozlov, S., Stewart, C. L., & Burke, B. (2009). Nesprin 4 is an outer nuclear membrane protein that can induce kinesin-mediated cell polarization. *Proc Natl Acad Sci U S A, 106*(7), 2194-2199. doi:10.1073/pnas.0808602106

Sabari, B. R., Dall'Agnese, A., Boija, A., Klein, I. A., Coffey, E. L., Shrinivas, K., . . . Young, R. A. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. *Science, 361*(6400). doi:10.1126/science.aar3958

A framework for advancing our understanding of cancer-associated fibroblasts, 20, Nature Research 174-186 (2020).

Sam, J., Peter, J. S., Thomas, A. J., Myong-Hee, S., Simon, C. B., Stafford, L. L., . . . Gordon, L. H. (2008). Interaction of the Glucocorticoid Receptor with the Chromatin Landscape. *Molecular Cell, 29*(5), 13.

Sanulli, S., Trnka, M. J., Dharmarajan, V., Tibble, R. W., Pascal, B. D., Burlingame, A. L., . . . Narlikar, G. J. (2019). HP1 reshapes nucleosome core to promote phase separation of heterochromatin. *Nature, 575*(7782), 390-394. doi:10.1038/s41586-019-1669-2

Sanyal, A., Lajoie, B. R., Jain, G., & Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature, 489*(7414), 109-113.

Saxton, M. J. (2008). Single-Particle Tracking: Connecting the Dots. *Nat Methods*(5), 671-772.

Saxton, M. J. (2020). Diffusion of DNA-binding species in the nucleus: A transient anomalous subdiffusion model. *Biophysical Journal.*

Schöne, S., Jurk, M., Helabad, M. B., Dror, I., Lebars, I., Kieffer, B., . . . Meijsing, S. H. (2016). Corrigendum: Sequences flanking the core-binding site modulate glucocorticoid receptor structure and activity. *Nat Commun, 7*, 13784. doi:10.1038/ncomms13784

Schones, D. E., Cui, K., Cuddapah, S., Roh, T. Y., Barski, A., Wang, Z., . . . Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. *Cell, 132*(5), 887-898.

Schurch, N. J., Schofield, P., Gierlinski, M., Cole, C., Sherstnev, A., Singh, V., . . . Barton, G. J. (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA, 22*(6), 839-851. doi:10.1261/rna.053959.115

Schwarz, G. (1978). Estimating the dimension of a model. *Ann Stat., 6*(2), 461-464.

Sengupta, P., van Engelenburg, S. B., & Lippincott-Schwartz, J. (2014). Superresolution Imaging of Biological Systems Using Photoactivated Localization Microscopy. *Chem. Rev.*(114), 3189-3202.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., . . . Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell, 148*(3), 458-472.

Sgouralis, I., & Presse, S. (2017). An Introduction to Infinite HMMs for Single-Molecule Data Analysis. *Biophys J, 112*(10), 2021-2029. doi:10.1016/j.bpj.2017.04.027

Shagufta, & Ahmad, I. (2018). Tamoxifen a pioneering drug: An update on the therapeutic potential of tamoxifen derivatives. *Eur J Med Chem, 143*, 515-531. doi:10.1016/j.ejmech.2017.11.056

Shakya, A., Park, S., Rana, N., & King, J. T. (2020). Liquid-Liquid Phase Separation of Histone Proteins in Cells: Role in Chromatin Organization. *Biophys J, 118*(3), 753-764. doi:10.1016/j.bpj.2019.12.022

Shin, Y., Berry, J., Pannucci, N., Haataja, M. P., Toettcher, J. E., & Brangwynne, C. P., 2017. (2017). Spatiotemporal control of intracellular phase transitions using light-activated optoDroplets. *Cell*(168), 158-171.

Shin, Y., & Brangwynne, C. P. (2017). Liquid phase condensation in cell physiology and disease. *Science, 357*(6357). doi:10.1126/science.aaf4382

Shirazi, F. H. (2011). Remarks in Successful Cellular Investigations for Fighting Breast Cancer Using Novel Synthetic Compounds. In: Breast Cancer – Focusing Tumor Microenvironment. *Stem Cells and Metastasis*, 85-102.

Shlyueva, D., Stampfel, G., & Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev.Genet., 15*(4), 272-286.

Shukron, O., & Holcman, D. (2017). Transient chromatin properties revealed by polymer models and stochastic simulations constructed from Chromosomal Capture data. *PLoS Comput Biol, 13*(4), e1005469. doi:10.1371/journal.pcbi.1005469

Shumaker, D. K., Kuczmarski, E. R., & Goldman, R. D. (2003). The nucleoskeleton: lamins and actin are major players in essential nuclear functions. *Curr Opin Cell Biol, 15*(3), 358-366.

Simons, S. S., Jr., Edwards, D. P., & Kumar, R. (2014). Minireview: dynamic structures of nuclear hormone receptors: new promises and challenges. *Mol Endocrinol., 28*(2), 173-182.
Simons, S. S., Jr., & Kumar, R. (2013). Variable steroid receptor responses: Intrinsically disordered AF1 is the key. *Mol Cell Endocrinol, 376*(1-2), 81-84. doi:10.1016/j.mce.2013.06.007
Sinkus, R., Lorenzen, J., Schrader, D., Lorenzen, M., Dargatz, M., & Holz, D. (2000). High-resolution tensor MR elastography for breast tumour detection. *Physics in Medicine & Biology, 45*, 1649.
Skinner, B. M., & Johnson, E. E. (2017). Nuclear morphologies: their diversity and functional relevance. *Chromosoma, 126*(2), 195-212. doi:10.1007/s00412-016-0614-5
Small, A., & Stahlheber, S. (2014). Fluorophore Localization Algorithms for Super-Resolution Microscopy. *Nat. Methods*(11), 267-279.
Song, L., & Crawford, G. E. (2010). DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb Protoc., 2010*(2), db.
Sorrentino, G., Ruggeri, N., Zannini, A., Ingallina, E., Bertolio, R., Marotta, C., . . . Del Sal, G. (2017). Glucocorticoid receptor signalling activates YAP in breast cancer. *Nat Commun, 8*, 14073. doi:10.1038/ncomms14073
Spencer, T. E., Jenster, G., Burcin, M. M., Allis, C. D., Zhou, J., Mizzen, C. A., . . . O'Malley, B. W. (1997). Steroid receptor coactivator-1 is a histone acetyltransferase. *Nature, 389*, 194-198.
Spitz, F., & Furlong, E. E. (2012). Transcription factors: from enhancer binding to developmental control. *Nat.Rev.Genet., 13*(9), 613-626.
Stadhouders, R., Filion, G. J., & Graf, T. (2019). Transcription factors and 3D genome conformation in cell-fate decisions. *Nature, 569*(7756), 345-354. doi:10.1038/s41586-019-1182-7
Stamatoyannopoulos, J. A. (2012). What does our genome encode? *Genome Res., 22*(9), 1602-1611.
Stasevich, T. J., & McNally, J. G. (2011). Assembly of the transcription machinery: ordered and stable, random and dynamic, or both? *Chromosoma, 120*(6), 533-545.
Stasevich, T. J., Mueller, F., Michelman-Ribeiro, A., Rosales, T., Knutson, J. R., & McNally, J. G. (2010). Cross-validating FRAP and FCS to quantify the impact of photobleaching on in vivo binding estimates. *Biophys.J, 99*(9), 3093-3101.
Stavreva, D. A., Coulon, A., Sung, M. H., Baek, S., John, S., Stixova, L., . . . Hager, G. L. (2015). Dynamics of chromatin accessibility and long-range interactions in response to glucocorticoid pulsing. *Genome Res., 25*(6), 845-857. doi:10.1101/gr.184168.114
Stavreva, D. A., Garcia, D. A., Fettweis, G., Gudla, P. R., Zaki, G. F., Soni, V., . . . Hager, G. L. (2019). Transcriptional Bursting and Co-bursting Regulation by Steroid Hormone Release Pattern and Transcription Factor Mobility. *Mol Cell, 75*(6), 1161-1177 e1111. doi:10.1016/j.molcel.2019.06.042

Stavreva, D. A., Muller, W. G., Hager, G. L., Smith, C. L., & McNally, J. G. (2004). Rapid glucocorticoid receptor exchange at a promoter is coupled to transcription and regulated by chaperones and proteasomes. *Mol.Cell Biol., 24*(7), 2682-2697.

Stavreva, D. A., Wiench, M., John, S., Conway-Campbell, B. L., McKenna, M. A., Pooley, J. R., . . . Hager, G. L. (2009). Ultradian hormone stimulation induces glucocorticoid receptor-mediated pulses of gene transcription. *Nat.Cell Biol., 11*(9), 1093-1102.

Stellato, C., Porreca, I., Cuomo, D., Tarallo, R., Nassa, G., & Ambrosino, C. (2016). The "busy life" of unliganded estrogen receptors. *Proteomics, 16*(2), 288-300. doi:10.1002/pmic.201500261

Stortz, M., Pecci, A., Presman, D. M., & Levi, V. (2020). Unraveling the molecular interactions involved in phase separation of glucocorticoid receptor. *BMC Biol, 18*(1), 59. doi:10.1186/s12915-020-00788-2

Stortz, M., Presman, D. M., Bruno, L., Annibale, P., Dansey, M. V., Burton, G., . . . Levi, V. (2017). Mapping the Dynamics of the Glucocorticoid Receptor within the Nuclear Landscape. *Sci Rep, 7*(1), 6219. doi:10.1038/s41598-017-06676-0

Stowers, R. S., Shcherbina, A., Israeli, J., Gruber, J. J., Chang, J., Nam, S., . . . Chaudhuri, O. (2019). Matrix stiffness induces a tumorigenic phenotype in mammary epithelium through changes in chromatin accessibility. *Nat Biomed Eng, 3*(12), 1009-1019. doi:10.1038/s41551-019-0420-5

Stumpf, M. P., & Porter, M. A. (2012). Mathematics. Critical truths about power laws. *Science, 335*(6069), 665-666. doi:10.1126/science.1216142

Style, R. W., Sai, T., Fanelli, N., Ijavi, M., Smith-Mannschott, K., Xu, Q., . . . Dufresne, E. R. (2018). Liquid-liquid phase separation in an elastic network. *Physical Review X*(8).

Sugo, N., Morimatsu, M., Arai, Y., Kousoku, Y., Ohkuni, A., Nomura, T., . . . Yamamoto, N. (2015). Single-Molecule Imaging Reveals Dynamics of CREB Transcription Factor Bound to Its Target Sequence. *Sci Rep, 5*, 10662. doi:10.1038/srep10662

Sunyer, R., Conte, V., Escribano, J., Elosegui-Artola, A., Labernadie, A., Valon, L., . . . Trepat, X. (2016). Collective cell durotaxis emerges from long-range intercellular force transmission. *Science, 353*(6304), 1157-1161. doi:10.1126/science.aaf7119

Sunyer, R., & Trepat, X. (2020). Durotaxis. *Curr Biol, 30*(9), R383-R387. doi:10.1016/j.cub.2020.03.051

Sweeney, E. E., Mcdaniel, R. E., Maximov, P. Y., Fan, P., & Craig, V. (2013). Models and Mechanisms of Acquired Antihormone Resistance in Breast Cancer: Significant Clinical Progress Despite

Limitations. *Horm Mol Biol Clin Investig*(9), 143-163.

Swift, J., Ivanovska, I. L., Buxboim, A., Harada, T., Dingal, P. C., Pinter, J., . . . Discher, D. E. (2013). Nuclear lamin-A scales with tissue stiffness and enhances matrix-directed differentiation. *Science, 341*(6149), 1240104. doi:10.1126/science.1240104

Swinstead, E. E., Miranda, T. B., Paakinaho, V., Baek, S., Goldstein, I., Hawkins, M., . . . Hager, G. L. (2016). Steroid Receptors Reprogram FoxA1 Occupancy through Dynamic Chromatin Transitions. *Cell, 165*(3), 593-605. doi:10.1016/j.cell.2016.02.067

Swinstead, E. E., Miranda, T. B., Paakinaho, V., Baek, S., Goldstein, I., Hawkins, M., . . . Hager, G. L. (2016). Steroid Receptors Reprogram FoxA1 Occupancy through Dynamic Chromatin Transitions. *Cell, 165*(3), 593-605.

Syed, A. P., Greulich, F., Ansari, S. A., & Uhlenhaut, N. H. (2020). Anti-inflammatory glucocorticoid action: genomic insights and emerging concepts. *Curr Opin Pharmacol, 53*, 35-44. doi:10.1016/j.coph.2020.03.003

Tarantino, N., Tinevez, J. Y., Crowell, E. F., Boisson, B., Henriques, R., Mhlanga, M., . . . Laplantine, E. (2014). TNF and IL-1 exhibit distinct ubiquitin requirements for inducing NEMO-IKK supramolecular structures. *J Cell Biol, 204*(2), 231-245. doi:10.1083/jcb.201307172

Teves, S. S., An, L., Hansen, A. S., Xie, L., Darzacq, X., & Tjian, R. (2016). A dynamic mode of mitotic bookmarking by transcription factors. *Elife, 5*. doi:10.7554/eLife.22280

Thanos, D., & Maniatis, T. (1995). Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. *Cell, 83*(7), 1091-1100.

Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., . . . Stamatoyannopoulos, J. A. (2012). The accessible chromatin landscape of the human genome. *Nature, 489*(7414), 75-82.

Tian, Y., Martinez, M. M., & Pappas, D. (2011). Fluorescence Correlation Spectroscopy: A Review of Biochemical and Microfluidic Applications. *Appl. Spectrosc.*(65), 115-124.

Tokunaga, M., Imamoto, N., & Sakata-Sogawa, K. (2008). Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nat Methods, 5*(2), 159-161.

Totaro, A., Panciera, T., & Piccolo, S. (2018). YAP/TAZ upstream signals and downstream responses. *Nat Cell Biol, 20*(8), 888-899. doi:10.1038/s41556-018-0142-z

Tsompana, M., & Buck, M. J. (2014). Chromatin accessibility: a window into the genome. *Epigenetics.Chromatin., 7*(1), 33.

Tunnacliffe, E., & Chubb, J. R. (2020). What Is a Transcriptional Burst? *Trends Genet, 36*(4), 288-297. doi:10.1016/j.tig.2020.01.003

Turner, A. L., Watson, M., Wilkins, O. G., Cato, L., Travers, A., Thomas, J. O., & Stott, K. (2018). Highly disordered histone H1− DNA model complexes and their condensates. *Proc Natl Acad Sci U S A*(115), 11964-11969.

Uhler, C., & Shivashankar, G. V. (2017). Regulation of genome organization and gene expression by nuclear mechanotransduction. *Nat Rev Mol Cell Biol, 18*(12), 717-727. doi:10.1038/nrm.2017.101

Valouev, A., Johnson, D. S., Sundquist, A., Medina, C., Anton, E., Batzoglou, S., . . . Sidow, A. (2008). Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. *Nat.Methods, 5*, 829-834.

Strain Stiffening of Fibrillar Collagen during Individual and Collective Cell Migration Identified by AFM Nanoindentation, 8 21946-21955 (2016).

Van Kampen, N. G. (1992). *Stochastic processes in physics and chemistry* (Vol. 1): Elsevier.

Vandevyver, S., Dejager, L., Van, B. T., Kleyman, A., Liu, Y., Tuckermann, J., & Libert, C. (2012). Glucocorticoid receptor dimerization induces MKP1 to protect against TNF-induced inflammation. *J Clin.Invest, 122*(6), 2130-2140.

Voong, L. N., Xi, L., Wang, J. P., & Wang, X. (2017). Genome-wide Mapping of the Nucleosome Landscape by Micrococcal Nuclease and Chemical Mapping. *Trends Genet, 33*(8), 495-507. doi:10.1016/j.tig.2017.05.007

Voss, T. C., & Hager, G. L. (2014). Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nat.Rev.Genet., 15*(2), 69-81.

Voss, T. C., Schiltz, R. L., Sung, M. H., Yen, P. M., Stamatoyannopoulos, J. A., Biddie, S. C., . . . Hager, G. L. (2011). Dynamic exchange at regulatory elements during chromatin remodeling underlies assisted loading mechanism. *Cell, 146*, 544-554.

Wagh, K., Ishikawa, M., Garcia, D. A., Stavreva, D. A., Upadhyaya, A., & Hager, G. L. (2021). Mechanical Regulation of Transcription: Recent Advances. *Trends Cell Biol*. doi:10.1016/j.tcb.2021.02.008

Walczak, A. M., Tkacik, G., & Bialek, W. (2010). Optimizing information flow in small genetic networks. II. Feed-forward interactions. *Phys Rev E Stat Nonlin Soft Matter Phys, 81*(4 Pt 1), 041905. doi:10.1103/PhysRevE.81.041905

Wang, D. Y., Fulthorpe, R., Liss, S. N., & Edwards, E. A. (2004). Identification of estrogen-responsive genes by complementary deoxyribonucleic acid microarray and characterization of a novel early estrogen-induced gene: EEIG1. *Mol Endocrinol, 18*(2), 402-411. doi:10.1210/me.2003-0202

Wang, J., Zhuang, J., Iyer, S., Lin, X., Whitfield, T. W., Greven, M. C., . . . Weng, Z. (2012). Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res., 22*(9), 1798-1812.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet, 10*(1), 57-63. doi:10.1038/nrg2484

Wang, Z., Malone, M. H., Thomenius, M. J., Zhong, F., Xu, F., & Distelhorst, C. W. (2003). Dexamethasone-induced gene 2 (dig2) is a novel pro-survival stress gene induced rapidly by diverse apoptotic signals. *J.Biol Chem., 278*(29), 27053-27058.

Wedeking, T., Lochte, S., Richter, C. P., Bhagawati, M., Piehler, J., & You, C. (2015). SingleCellGfp-TrapRevealsStoichiometryandDynamics of Cytosolic Protein Complexes. *Nano Letters*(15), 3610-3615.

Wei, M. T., Elbaum-Garfinkle, S., Holehouse, A. S., Chen, C. C. H., Feric, M., Arnold, C. B., . . . Brangwynne, C. P. (2017). Phase behaviour of

disordered proteins underlying low density and high permeability of liquid organelles. *Nature Chemistry*(9), 1118.

Wei, S. C., & Yang, J. (2016). Forcing through Tumor Metastasis: The Interplay between Tissue Rigidity and Epithelial-Mesenchymal Transition. *Trends Cell Biol, 26*(2), 111-120. doi:10.1016/j.tcb.2015.09.009

Weikum, E. R., de Vera, I. M. S., Nwachukwu, J. C., Hudson, W. H., Nettles, K. W., Kojetin, D. J., & Ortlund, E. A. (2017). Tethering not required: the glucocorticoid receptor binds directly to activator protein-1 recognition motifs to repress inflammatory genes. *Nucleic Acids Res, 45*(14), 8596-8608. doi:10.1093/nar/gkx509

Welboren, W. J., van Driel, M. A., Janssen-Megens, E. M., van Heeringen, S. J., Sweep, F. C., Span, P. N., & Stunnenberg, H. G. (2009). ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands. *EMBO J., 28*(10), 1418-1428.

Winkler, J., Abisoye-Ogunniyan, A., Metcalf, K. J., & Werb, Z. (2020). Concepts of extracellular matrix remodelling in tumour progression and metastasis. *Nature Communications, 11*, 1-19.

Woringer, M., & Darzacq, X. (2018). Protein motion in the nucleus: from anomalous diffusion to weak interactions. *Biochem Soc Trans, 46*(4), 945-956. doi:10.1042/bst20170310

Wright, P. E., & Dyson, H. J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.*(293), 321-331.

Yanfang, L., Hong, M., & Jing, Y. (2020). ERα, A Key Target for Cancer Therapy: A Review. *Onco Targets Therapy, 13*, 2183–2191.

Yu, J. (2016). Single-molecule studies in live cells. *Annual review of physical chemistry*, 67.

Zane, L., Sharma, V., & Misteli, T. (2014). Common features of chromatin in aging and cancer: cause or coincidence? *Trends in cell biology*, 686-694.

Zhang, Y., Fonslow, B. R., Shan, B., Baek, M. C., & Yates, J. R., 3rd. (2013). Protein analysis by shotgun/bottom-up proteomics. *Chem Rev, 113*(4), 2343-2394. doi:10.1021/cr3003533

Zhao, L., Zhou, S., & Gustafsson, J. (2019). Nuclear Receptors: Recent Drug Discovery for Cancer Therapies. *Endocr Rev, 40*(5), 1207-1249. doi:10.1210/er.2018-00222

Zhen, C. Y., Tatavosian, R., Huynh, T. N., Duc, H. N., Das, R., Kokotovic, M., . . . Ren, X. (2016). Live-cell single-molecule tracking reveals co-recognition of H3K27me3 and DNA targets polycomb Cbx7-PRC1 to chromatin. *Elife, 5*. doi:10.7554/eLife.17667

Zhu, C., Chen, Y., & Ju, L. A. (2019). Dynamic bonds and their roles in mechanosensing. *Curr Opin Chem Biol, 53*, 88-97. doi:10.1016/j.cbpa.2019.08.005

Zhu, F., Farnung, L., Kaasinen, E., Sahu, B., Yin, Y., Wei, B., . . . Taipale, J. (2018). The interaction landscape between transcription factors and the nucleosome. *Nature, 562*(7725), 76-81. doi:10.1038/s41586-018-0549-5

Zidovska, A., Weitz, D. A., & Mitchison, T. J. (2013). Micron-scale coherence in interphase chromatin dynamics. *Proc Natl Acad Sci U S A, 110*(39), 15555-15560. doi:10.1073/pnas.1220313110