Protein spatiotemporal dynamics in gene regulation and disease pathology

by

Ming Zheng

B.S., Peking University (2016)

Submitted to the Department of Physics in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Physics

at the

Massachusetts Institute of Technology

September 2024

Authored by: Ming Zheng
Department of Physics
August 15th, 2024

Certified by: Richard A. Young
Professor of Biology, Thesis supervisor

Certified by: Mehran Kardar
Professor of Physics, Thesis co-advisor

Accepted by: Lindley Winslow
Associate Department Head
Professor of Physics

Protein spatiotemporal dynamics in gene regulation and disease pathology

by

Ming Zheng

Submitted to the Department of Physics on August 15, 2024, in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Physics

ABSTRACT

A cell orchestrates billions of proteins to the right place at the right time to perform diverse cellular processes. Over the decades, this field has been evolving by integrating advances in microscopy, biochemistry, and molecular biology to unravel the intricate mechanisms governing protein spatiotemporal dynamics as well as the functional consequences. This thesis focuses on the physical motions of proteins at a length scale of tens of nanometers to several microns, where the apparent diffusion and the condensate dynamics of assembly and disassembly are specifically studied. In the studies presented in this thesis, the functional relevance of protein motion is exemplified in the context of gene regulation and disease pathology. We find that the apparent diffusion of transcription factors (TFs) is preferentially partitioned into slowly diffusing states by interacting with RNA, leading to enhanced chromatin occupancy and gene expression (Oksuz et al., 2023). The assembly and disassembly dynamics of transcriptional condensates are coupled to the active RNA synthesis, linking gene expression and the spatiotemporal organization of transcriptional proteins in a feedback loop (Henninger et al., 2021). In addition to transcriptional proteins, we find insulin receptors (IRs) are incorporated in dynamic condensates in normal cells to perform metabolic signaling transduction. In insulin-resistant cells which could occur in chronic diseases such as type 2 diabetes (T2D), IR signaling is dysregulated, associated with diminished IR condensate dynamics of assembly and disassembly (Dall'Agnese et al., 2022). Furthermore, pathogenic signaling reduces the mobility of key proteins–both inside and outside of condensates—that act in many cellular functions. Such reduced protein mobility under diverse pathogenic stimuli, termed proteolethargy, may account for diverse cellular dysregulation seen in chronic disease (Dall'Agnese, Zheng, Moreno et al., 2024).

Thesis supervisor: Richard A. Young
Title: Member, Whitehead Institute; Professor of Biology

Thesis co-advisor: Mehran Kardar
Title: Professor of Physics

ACKNOWLEDGEMENTS

I am grateful for having worked alongside the many talented members of the Young lab. Their collective guidance, support, and help have greatly shaped my development in the lab. A few lab members that I've worked closely with deserve special acknowledgment. I benefited from working closely with Alessandra Dall'Agnese, Shannon Moreno, and Jesse Platt as a team on the Proteolethargy project. Ale's positive attitude, willingness to embrace challenges, enthusiasm for science, and leadership were a driving force throughout the project. Shannon's critical thoughts in designing experiments and clarifying the logic and dedication to getting the job organized and done were essential in successfully completing the project. Jesse's proficient skills in the wet lab and unique insights as an MD-PhD were instrumental in completing this high-quality work with a broad vision. In addition, Ale deserves extra thanks for being an excellent mentor and helping me from multiple aspects. Shannon deserves extra thanks since it's always a pleasure to work with her and we are also good friends. I'd also like to thank Salman Banani and Jon Henninger as I enjoyed working with them as multidisciplinary scientists with clear logic and organized mind. Tony Lee has always been a great source of guidance and I appreciate him being very patient in coaching me towards the direction of becoming an all-round scientist. Many current and former lab members (Giuseppe, Hailey, Henry, Kalon, Alejandra, Ido, Daniel, Adrienne, Catherine, Hao-Che, Alfonso, Susana, Ozgur, Lena, Max, Mike, and others) have made the lab a fun and productive environment. Nancy Hannett deserves immense thanks for keeping the lab running.

I am extremely grateful to Rick Young for recruiting me to his lab. Rick is such an instructive mentor, and he has been helping every trainee to grow toward being a leading scientist in multidimensions. Rick's insistence on not only doing cutting-edge science but also communicating advanced ideas to general audiences and realizing scientific concepts in industry for actual therapeutics that improve human life, has been a positive source of personal and professional growth in my life. I'm sure that the many lessons I've learned from Rick will continue to be a guiding force into the future.

I would like to thank my thesis committee, Mehran Kardar, Jeff Gore, and Nikta Fakhri, for the time and effort they have dedicated to guiding my growth as a scientist. Mehran has been so supportive and thoughtful about my PhD training and graduation. Jeff and Nikta, as world-leading biophysicists, gave me important advice on how to bridge the physics and biology communities. Their achievements in biophysics also set a high bar for me to pursue. In addition, I thank Professor Arup Chakraborty for inspiring discussions.

I'd also like to thank my previous scientific mentor, Ibrahim Cissé. My time as a research assistant in the Cissé lab allowed me to explore and develop my scientific interests in single-molecule techniques. Working closely with Choongman, Manyu, Fatma, Owen, Wonki, Jan, Takuma, Mohammed, Micca, and Arjun taught me the fun of doing science as a team and helped inspire me to pursue a PhD.

Next, I want to thank my friends at Whitehead Institute and Biology and Biophysics at MIT: Pu, Yi, Jingwei, Liguo, Xiaohang, Deepti, Pradeep, Andriy, Tenzin, Baboucarr, Jinghui, Weishun, Jiliang, Junang, Tzer-Han, and Yuki.

None of this would have been possible without my family. I'm so lucky to be their son and I grew up in a family full of love and an open mind. My parents have always encouraged me to pursue science and provided ample opportunities for me to do so.

Lastly, I want to express my heartfelt gratitude to my wife, Yutong. I'm so lucky to meet her in my life, as finding a fountain in a desert. Her company and support have been crucial to my career and life. I cannot count how many days she came to the Whitehead with me at night or during weekends, waiting for me to do late experiments and cell culturing. I have also been learning a lot from her and I look forward to an even brighter future that we shall create together.

STATEMENT OF WORK

Chapter 1

I wrote Chapter 1 myself with suggestions from Rick Young and Tony Lee.

Chapter 2

I wrote the manuscript presented in Chapter 2 with my co-authors Alessandra Dall'Agnese, Shannon Moreno, Tony Lee, as well as Rick Young. In collaboration with Alessandra Dall'Agnese, Shannon Moreno, Jesse Platt, I conceptualized the study, constructed biological materials, designed and executed experiments, and performed analyses and physical modeling pertaining to all figures in Chapter 2.

Chapter 3

Ozgur Oksuz, Jon Henninger, and Rick Young wrote the manuscript presented in Chapter 3. I designed and executed experiments and analyses in Figure 5, Figure S7A, S7B, and S7D-F.

Chapter 4

Jon Henninger, Ozgur Oksuz, Krishna Shrinivas, Arup Chakraborty and Rick Young wrote the manuscript presented in Chapter 4. I designed and executed experiments in Figures 5G-I.

Chapter 5

Alessandra Dall'Agnese, Jesse Platt, Tony Lee, and Rick Young wrote the manuscript presented in Chapter 4. I designed and executed experiments and analyses in Figure 3b-e, Figure 4a,b, Figure 5d,f, Figure S13, Figure S14, Figure S16, and Figure S17.

Chapter 6

I wrote Chapter 6 myself.

TABLE OF CONTENTS

Chapter 1:
Introduction

Proteins perform a vast array of biological processes necessary for the survival, growth, and reproduction of the cell. For instance, many proteins function as enzymes, playing a central role in catalyzing virtually all biochemical reactions (1). The enormous scale and complexity of biological processes require meticulous coordination of both the spatial and temporal aspects of molecular interactions within the cell. A key consideration in all biological processes is the concept of protein spatiotemporal dynamics.

Protein spatiotemporal dynamics—encompassing the pathways, timing, and speeds—have predominantly been studied through the lens of diffusion and subsequent effects on collision-limited reactions (2-18). Recent studies have also increasingly focused on emergent properties, such as local protein concentration, that arise from the collective behaviors of proteins and their roles in cellular biology (19-24). Building on these insights, my Ph.D. research explores protein spatiotemporal dynamics, specifically looking at apparent diffusion at the single-molecule level and condensate formation and disassembly at the ensemble level. These investigations are applied to understand mechanisms in gene regulation and disease pathology, illustrating the practical implications of these dynamics.

In this introduction, the structure of the sections is outlined as follows. First, I will discuss several milestones in the history of studying protein motion in cells, providing a historical context for my research. Subsequently, I will summarize key insights into protein spatiotemporal dynamics that have shaped current understanding in the field. The background on gene regulation at the transcription level, which is pertinent to my thesis, will be covered next. I will then delve into the two levels of protein motion studied in my research: apparent diffusion in intracellular space as single molecules and collective formation and disassembly as dynamic condensates. Lastly, I will introduce the three key techniques I implemented to study protein spatiotemporal dynamics, focusing on a length scale of tens of nanometers to several microns.

Chapter 1.1: History of studying protein motion in cells

“Biology is wet and dynamic. Molecules […] immersed in an aqueous environment, are in continuous riotous motion” (25). Through the 19th and early 20th centuries, scientists gradually speculated that each cell contain billions of protein molecules that are in continuous motion within an aqueous milieu densely packed with biomolecules. In 1838-1839, the development of cell theory by Schleiden and Schwann laid the groundwork for understanding cellular structure (26): they proposed that cells were the basic unit of all living things and that they had a basic similarity in structure and growth. This theory set the stage for thinking about the cell as a container of fluid and a site of biochemical reactions. During the 1880s, observations of cytoplasmic streaming in plant cells, first described in detail by scientists like Eduard Strasburger, demonstrated the dynamic nature of the intracellular environment of plants (27): not only were the interior of a cell fluid, but they also exhibited organized flow, facilitating transport and mixing of organelles and molecules within the cell. By the end of the 19th century, the improvement of microscopes enabled clearer visualization of animal cellular components,

confirming that all the cells contained fluid as the medium incorporating biomolecules, known as cytoplasm (28). Einstein's theoretical explanation of Brownian motion in the early 20th century provided a scientific basis for the random, continuous movement observed in particles suspended in a fluid (29). This theory indirectly supported the idea that cellular fluids exhibit dynamic and continuous motion similar to other aqueous environments where Brownian motion occurs. However, the technology to directly observe protein physical motion in cells was limited.

The advent of electron microscopy in the 1930s allowed scientists to visualize the cell at unprecedented resolution, allowing for the examination of protein movement through submicron subcellular structures along the secretory pathway—such as the endoplasmic reticulum (ER). Electron microscopy autoradiography experiments published in the 1960s revealed for the first time that secretory proteins migrate in cells from ER to Golgi complexes, and then to zymogen granules (30, 31): radioactive labeling of secretory proteins at the initial synthesis step allowed the tracing of proteins through those submicron structures. However, this type of experiment lacks specificity as all proteins synthesized at ER are radioactively labeled.

Shortly after, the applications of the immunofluorescence (IF) technique allowed the tracing of non-secretory proteins in living cells, where only specific proteins were labeled by fluorophore based on the principle of antibody-antigen binding. With IF, the observation of protein migration was expanded to non-secretory proteins thanks to the classic heterokaryon surface-protein-mixing experiment by Frye and Edidin (32), and the complementary surface-protein mobility study by Edidin and Fambrough (2). However, the IF-based studies were limited to observing the migration of proteins on cell surface.

In the 1960s-1990s, the discovery of green fluorescent protein (GFP) with the subsequent development of genetic encoding and imaging techniques revolutionized the study of the field. Discovered from the jellyfish *Aequorea Victoria*, GFP is a β-barrel-shaped protein that undergoes a chemical rearrangement to form a fluorophore (33). With the genetic encoding of the GFP sequence adjacent to the protein sequence, virtually any protein of interest can be permanently tagged with GFP when being produced as a chimaera by cells. The resulting chimaera often retains parent-protein targeting and function (33). More fluorescent proteins with diverse excitation-emission spectrums were also developed later on. Therefore, for the first time, it was possible to visualize the live movement of virtually any specific proteins in any subcellular structures: imaging techniques and consequent analyzing pipelines based on fluorescent protein labeling such as fluorescence recovery after photobleaching (FRAP) and fluorescence correlation spectroscopy (FCS) enabled inferring protein transit dynamics and interactions within the cellular milieu (3, 4, 34-36).

Since then, the field has evolved by integrating advances in fluorophore and camera technologies to unveil protein transverse at nanoscale details. With the invention of bright, photostable, live-cell permeable fluorophores and cameras with single-photon sensitivity and fast frame rate, biophysicists can continuously track individual protein molecules' positions while they are in riotous move, namely single-molecule tracking or single-particle tracking (SPT) (37). With the development of photoconvertible fluorophores, super-resolution strategies such as Photo-Activation Localization Microscopy (PALM) (38) and STochastic Optical Reconstruction Microscopy (STORM) (39) localize fluorescence-labeled proteins with nanoscale spatial resolution. Time-correlated PALM (tcPALM) analysis maintains the spatial resolution of PALM and in the meantime captures the transient,

collective, nanoscale clustering dynamics of proteins (40). These methods have significantly refined our understanding of how proteins organize in space and time.

Chapter 1.2: Key insights into protein spatiotemporal dynamics

Up till now, cumulative investigations at the intersection of physics and biology have established a general agreement that protein spatiotemporal dynamics are not just passive processes but are actively regulated and integral parts of cellular life. Several key insights are summarized to represent the overview of the field and to recapitulate the concepts of my thesis research.

Significance of protein movement at various scales. The study of protein movement encompasses a wide range of length and time scales. Although my thesis research focuses on a subset of these scales—a length scale of tens of nanometers to several microns associated with a timescale of several milliseconds to tens of seconds, the concept that protein spatiotemporal dynamics are essential for cellular function and are disrupted in disease can be expanded to other variations of this theme. Consider, for instance, intra-protein dynamics of conformational change (<1 nm), inter-protein dynamics of complex assembly (1-10 nm), and directed movement via kinesin/dynein motors ($\gtrsim$10 μm). Under this broader theme, there are two levels of topics pivotal in expanding our awareness of protein movement: (1) the relationship between protein movement and the relevant functional implication at each length/time scale, and (2) the influence of protein movements and functional implications across different length/time scales.

Diffusion in a complex milieu. Diffusion is a baseline physical process affecting all protein movements in the cellular milieu. While the principles of diffusion in dilute solutions are well understood, the intracellular environment presents a complex scenario. The cell is a crowded space filled with a myriad of macromolecules, which creates a viscoelastic medium far from the ideal conditions often described in physical and chemical textbooks. Various interactions in intracellular space may alter protein diffusion rates and patterns, implying potential challenges of protein transverse cellular milieu and the consequent limit in the rate for diverse biological processes.

Collective dynamics and biomolecular condensates. In addition to movements as single molecules, proteins engage in collective behaviors when viewed from an ensemble perspective, such as forming biomolecular condensates. These condensates are dynamic entities whose size, composition, and lifetime can rapidly change in response to cellular signals or stress conditions. The dynamics of these condensates, such as the formation and disassembly, are tightly coupled to active biochemical processes. Knowing how the dynamics of condensates can influence cellular processes and vice versa will deepen our understanding of the roles of those emergent entities in a sophisticated level of regulation and responsiveness.

Functional implications in health and disease. Proteins' spatiotemporal dynamics—encompassing their pathways, timing, and speeds—are key parameters of their functionality within cells. Lines of evidence suggest that precise control over where proteins localize, when they arrive, and how fast they travel is critical to ensuring correct cellular function (41-45). For instance, the tumor suppressor protein p53 must be transported to the nucleus to activate genes responsible for cell cycle arrest and

apoptosis; incorrect localization due to mutations can lead to cancer. An understanding of the mechanisms underlying the control of protein spatiotemporal dynamics is crucial; it reveals not only how proteins fulfill proper roles within the cellular environment but also why disruptions in these processes can lead to disease.

In addition, interaction with active processes is also a critical feature of protein motion in living cells. Those insights underscore the complexity and critical nature of protein spatiotemporal dynamics in cellular function and disease, highlighting the cutting-edge interface of physics and biology in understanding life at the molecular level.

Chapter 1.3: Gene regulation at the transcription level

Transcription is a fundamental biological process by which genetic information encoded in DNA is transcribed into RNA, serving as the first step in gene expression (1). This process involves spatiotemporal orchestration of diverse regulatory proteins in a compact nuclear milieu, highlighting protein mobility as a key issue in transcription. Nowadays, transcription is well recognized to be carried out by condensates where hundreds of transcription apparatus assemble at the sites of active transcription as membraneless organelles. Cutting-edge works found pervasive syntheses of functional RNA molecules from non-coding regions, meaning those RNAs don't translate into proteins yet play vital roles in cellular processes.

The synthesis of a single transcript includes three major steps—initiation, elongation, and termination: an RNA polymerase is recruited to a specific DNA region at the starting site of a gene, synthesizing RNA as it traverses the DNA template, and stopping transcribing followed by disassociating from DNA (46). This basic description doesn't exclude the importance of other detailed steps such as promoter escape and pause-release after initiation and splicing during elongation. Each step of transcription incorporates delicate regulation from a number of proteins (46-51). Most of my thesis research focuses on proteins associated with initiation, where the synthesis of RNA from DNA is set in motion. Initiation happens at the promoter, a specific DNA sequence that marks the spot where transcription is to start and serves as a binding site for various proteins (52). Transcription factors (TFs) are particular proteins that can recognize and bind to the promoter (48, 49, 53-55), setting the stage for the assembly of the rest of the transcription machinery. Among the transcription machinery is RNA polymerase, a key enzyme responsible for synthesizing RNA from the DNA template (56, 57). TFs and the RNA polymerase form the pre-initiation complex (PIC), a crucial assembly that positions the polymerase correctly at the start site of transcription (51, 58). Initiation is significantly enhanced by the presence of enhancers (53, 59-67), which are distant DNA elements that allow TFs to bind and interact with promoters to increase transcription efficiency by looping closer and interacting with the PIC. Cofactors, which include various proteins and other molecules, also play essential roles in initiation by influencing the efficiency and specificity of transcription (68, 69). For instance, the Mediator complex is a cofactor essential for enhancer-promoter communication and the proper assembly of the PIC (70). In addition, initiation also involves dynamic interactions and modifications of chromatin, as chromatin must be remodeled to allow access to the DNA, a process mediated by chromatin remodeling complexes and histone-modifying enzymes (71-75). Clearly, diverse proteins need to be coordinated in space and time in order to synthesize even a single transcript, making

protein mobility a key issue in transcription regulation at the single-molecule (stoichiometric) level (58, 76-81).

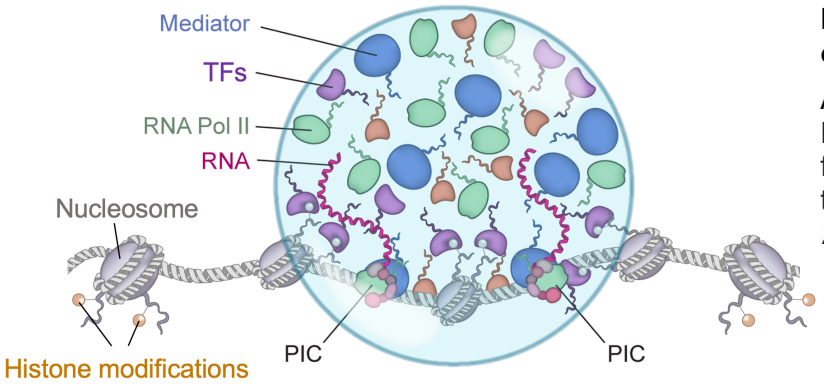


Figure 1.3 | Graphical illustration of a transcriptional condensate.

Adapted from Sharp, Chakraborty, Henninger, & Young (2022). RNA in formation and regulation of transcriptional condensates. *Rna*, *28*(1), 52-57.

Recent works from our lab and collaborators described transcriptional condensates in gene control (Figure 1.3) (21, 23, 82-90). Transcriptional condensates are membraneless organelles—several hundreds of nanometers in diameter—that concentrate hundreds of transcriptional machinery and regulatory factors at active gene loci (83). Proteins assemble these organelles through weak, multivalent interactions with other proteins (90). One important class of such interactions are mediated by proteins' intrinsically disordered regions (IDRs), which lack specific three-dimensional structures and interact with multiple other biomolecules in a non-stoichiometric fashion (91). *In vitro*, transcriptional proteins form droplets through liquid-liquid phase separation (LLPS) (82). In cells, LLPS is hypothesized to be an underlying mechanism driving condensate formation: a number of TFs bind to enhancers, leading to the local protein concentration beyond C_{sat} (84) thus switching to a state that favors incorporating further transcriptional proteins such as TFs, coactivators, positive elongation factors, and RNA polymerases. The grown condensate "wet" promoter(s), embedding distal gene(s) in a concentrated pool of transcriptional proteins that favors transcription; In other words, condensate formation establishes the long-range enhancer-promoter communication of transcription activation, revolutionized our understanding of how enhancers regulate distal gene expression (66, 82, 83). Transcriptional condensates are more than just static droplets. The protein composition concentrated within transcriptional condensates varies as the transcription step proceeds: at the initiation step, unphosphorylated RNA polymerases are enriched in transcriptional condensates; at the elongation step, hyperphosphorylated RNA polymerases leave the transcriptional condensates and incorporated into condensates that are formed by splicing factors (85). Transcriptional condensates also have a finite lifetime, lasting for seconds to minutes followed by disassembly (40, 83, 92). Mechanisms that regulate collective motion of transcriptional proteins underlying condensate formation and disassembly remain not fully understood.

Transcription is not confined to coding regions of the genome. Non-coding regions, including enhancers and promoters, can also be transcribed into non-coding RNAs (ncRNAs). Many individual case studies have revealed the regulatory roles of certain ncRNAs in gene expression, influencing the structure of chromatin and the recruitment of transcriptional machinery (93-95). One of the most well-known ncRNAs is the long non-coding RNA (lncRNA), which can modulate gene expression by acting as scaffolds that bring together multiple proteins to form ribonucleoprotein complexes, influencing chromatin architecture and transcriptional activity (94). For example, the lncRNA *Xist* is essential for

X-chromosome inactivation in females, highlighting its role in genetic regulation (96). Another significant group is the enhancer RNA (eRNA), which was thought to be nonspecific byproduct from active DNA enhancer regions. More recent works suggest that some eRNAs can mediate enhancer functions such as the loading of the Mediator complex and recruitment of transcription machinery (73, 97, 98). However, questions remain if these individual cases can be generalized to the whole genome and what might be the shared mechanisms. Furthermore, the roles of RNAs in altering the motion of individual proteins and the proteins' collective behaviors remain underexplored.

In summary, transcription is a multifaceted process with layers of regulation and complexity. The spatiotemporal coordination of transcriptional proteins, RNA, and chromatin is central to the precise control of gene expression. Advances in our understanding of condensate formation and non-coding RNA synthesis shed light on the sophisticated nature of transcriptional regulation.

Chapter 1.4: Apparent diffusion of proteins in intracellular space

Protein diffusion dynamics broadly influence cellular function. Diffusivity is a fundamental property of biomolecules—including proteins—as nearly all biomolecules experience diffusive motion at certain periods. Diffusion is utilized for most biochemical reactions in cells as it allows proteins to sample different spaces and find their reaction partners without the need of work (12). In cells, billions of proteins are densely packed in a crowded, viscoelastic environment, leading to an apparent diffusion rate at least tenfold lower in cells than in water: it takes several seconds for a protein to diffuse from one side of a cell to another, and it takes several hours for a protein to scan a whole cellular volume. In contrast, for a typical biochemical reaction in cells, the catalytic conversion of substrate into product occurs at subsecond scale. In line with these facts, diverse cellular processes are reported to be diffusion-limited, which implies two consequences: (1) protein diffusion dynamics as a rate-limiting parameter defining normal cellular function, and (2) the disruption of protein diffusion might be associated with disease.

Protein diffusion dynamics are based on Brownian motion yet the apparent rates are largely influenced by various "forces" in cells. In physics, the basic diffusion process—i.e., Brownian motion—is animated by the constant collisions with smaller particles based on thermal energy (25). Brownian motion is characterized by a linear dependence between the mean-square displacement (MSD) and the timelag τ: $\mathrm{MSD}(\tau) = 2nD\tau$, with *D* being the diffusion coefficient and *n* being the dimension of the diffusion process. In biological cells, the conventional diffusion and active transport intertwine within a crowded, porous intracellular matrix where both thermal energy and non-equilibrium activities contribute. Therefore, "apparent" diffusion might be more stringent when describing the nonspecific migration of proteins in cells. Accordingly, the apparent diffusion rate $D_{\mathrm{app}}(\tau) = (\mathrm{MSD}_{\mathrm{obs}}(\tau) - \sigma^2)/2n'\tau$ can deviate from the diffusion coefficient of pure Brownian motion, with $\mathrm{MSD}_{\mathrm{obs}}(\tau)$ being the observed MSD of timelag τ, σ being the localization error, and n' being the effective dimension of observation.

Four basic types of diffusion could be associated with protein motion in cells, namely Brownian motion, subdiffusion, superdiffusion, and confined fluctuation (15). The exact type of diffusion can be inferred from the MSD-timelag relation. The first three types obey the power law $\mathrm{MSD}(t) \sim t^{\alpha}$, where α=1, <1, and >1 are referred to as Brownian motion, subdiffusion, and superdiffusion, respectively

(Figure 1.4A). In the rest of the section, I will discuss these four types of diffusion, focusing on the relevant length/time scales, the underlying “forces”, and the relevance to my thesis research. Finally, I will discuss variations of a protein’s diffusivity due to the heterogeneity of intracellular space.

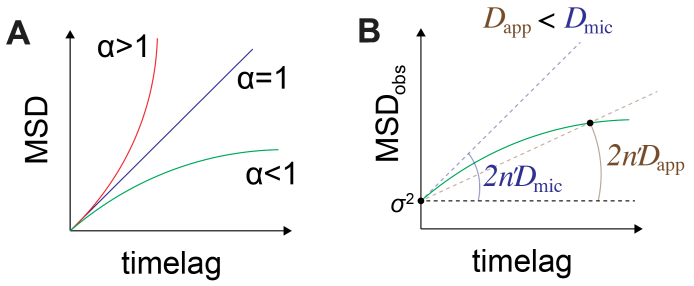


Figure 1.4 | MSD-timelag relations.

A. Brownian motion (α=1), subdiffusion (α<1), and superdiffusion (α>1) in theory.

B. Apparent diffusion rate (D_{app}) vs microscale diffusion coefficient (D_{mic}) in observed subdiffusion.

The Brownian motion applies to proteins at a short length/time scale when their random motion is dominated by constant collisions with smaller particles (e.g., water molecules). For example, the diffusion of a nuclear protein in the nucleoplasmic fluid in interchromatin space before hitting the chromatin network is roughly the Brownian motion (99); the diffusion of a cytoplasmic protein in the cytosol before colliding with large crowders such as ribosomes is roughly the Brownian motion (100). The Brownian motion is governed by Stokes–Einstein–Sutherland (SES) equation $D_{\mathrm{mic}} = \frac{k_{\mathrm{B}}T}{6\pi\eta r}$, where the microscale diffusion coefficient (D_{mic}) is defined by the temperature (T), microscopic viscosity (a.k.a. solvent viscosity) (η), or protein radius (r). Diverse factors can change temperature, viscosity, or effective protein size in cells. For example, two proteins bound together can increase the effective protein size. In Chapter 2, I will introduce a pervasive pathogenic mechanism among oxidative-stress-associated human diseases where diverse proteins show reduced mobility. We found evidence that pathogenic conditions associated with elevated oxidative stress upshift the chance of protein crosslinking. Such a correlation between upshifted protein crosslinking and reduced protein mobility is consistent with the SES dynamics: protein crosslinking leads to an effective increase in r and thus a decrease in D_{mic}.

Subdiffusion influences proteins’ random motion above length/time scales when hindering effects such as collisions with crowders and sieving by intracellular networks come into play. The degree of hindering depends on the relative dimension between the protein size and the characteristic length of the hindering: if a protein size is closer or goes beyond the spacing between crowders or the pore size of the intracellular network, a strong hindering effect will be sensed by the protein. The degree of hindering also depends on the interactions with macromolecules in the milieu: for example, nuclear proteins with a high chance of interacting with RNAs will display strong hindering. The hindering effect causes the MSD-timelag curve to grow with a slope smaller than $2n'D_{\mathrm{mic}}$, leading to a decreased apparent diffusion rate of any given timelag compared to the microscale diffusion coefficient (Figure 1.4B). In Chapter 2, in addition to the SES dynamics, this size-dependent hindering may contribute to the pathogenic mechanism of reduced protein mobility since protein crosslinking will increase the protein size. In Chapter 3, we reason that RNA exerts hindering effect on RNA-binding proteins, thus the term “subdiffusion” is used as an effective description of the observed slowdown of apparent diffusion based on MSD evaluated over a finite timelag.

Superdiffusion happens when protein diffusion is coupled to complex active processes in living cells. For example, proteins carried by a molecular motor with a timescale $\gtrsim D/v^2$ will reveal a significant

superdiffusion behavior, where v is the directed velocity of the molecular motor. In addition, superdiffusion of proteins can happen due to nonspecific fluctuations caused by mechanical agitation (101) and transient convective flow, which can result from cellular organelle transport, cell shape changes, cytoskeletal arrangements, and transient chemical/temperature gradients. If the timelag of measurement is much larger than the timescale of nonspecific fluctuations, such superdiffusion will appear as Brownian motion with an enhanced apparent diffusion rate. Superdiffusion is not explicitly studied in any Chapters of this thesis.

Confined fluctuation (a.k.a. confined diffusion) is the random motion restricted to a finite region. Proteins perform confined fluctuation when trapped in subdomains such as cytoskeleton “cages”. The actual MSD-timelag relation of confined fluctuation depends on the geometry (37). For confined fluctuation in a cellular matrix consisting of permeable subdomains, proteins can penetrate the confinement with certain possibility, displaying subdiffusion when being examined at long timescale. Confined diffusion is not explicitly studied in any Chapters of this thesis.

In addition to diffusion, proteins could be immobilized, such as when binding to chromatin. In Chapter 2, in addition to decreased rates of Brownian motion and subdiffusion when crosslinking happens between mobile proteins, the crosslinking of proteins with stationary structures might also account for reduced protein mobility in pathogenic conditions.

A protein’s diffusivity may vary at different locations due to the heterogeneity of intracellular space: polymers, organelles, and membraneless compartments occupying different cellular spaces with differential permeability, viscosity, chemical potential, and interaction characteristics establish distinct local environments where apparent diffusion rates differ. For example, transcription factors (TFs) are a class of proteins functioning by recruiting transcriptional proteins to active genes after locating the DNA-binding sites, doing so by making frequent switches among fast diffusion in the nucleoplasm, slow diffusion around chromatin, and immobilization on DNA-binding sites. In Chapter 3, I will demonstrate how the heterogeneous distribution of RNA can diversify the diffusion dynamics of TFs: nascent RNAs are synthesized at the active sites of transcription, forming submicron RNA-rich hubs; TFs can incorporate into those hubs and interact with RNA, leading to partition of diffusivity into multiple states.

Chapter 1.5: Protein condensate dynamics of formation and disassembly

In addition to the apparent diffusion of proteins at the single-molecule level, many recent works have focused on another type of protein spatiotemporal organization called biomolecular condensates. In this section, I will give a brief intro about biomolecular condensates, followed by discussing condensate dynamics of formation and disassembly, where the transcription process is exemplified. Finally, I will discuss the possibility that the dysregulation of condensate dynamics of formation and disassembly may be linked to disease pathology.

Biomolecular condensates are droplet-like compartments in cells that function to concentrate proteins and nucleic acids. The seminal discovery by Brangwynne et al. in 2009 recognized P granules as organelles in germline with characteristics of protein droplets, including the lack of membrane, rapid

molecule-rearrangement kinetics, unique surface tension and viscosity, and the capability of fission and fusion (102). Following studies identified more and more droplet-like organelles in cells that are recognized as biomolecular condensates. At this point, it is well appreciated that most fundamental cellular processes appear to be carried out by biomolecular condensates, including transcription, signal transduction, RNA metabolism, ribosome biogenesis, and DNA damage response (20, 23, 103-105). Biomolecular condensates function to concentrate specific proteins and nucleic acids and exclude others, which is speculated to establish unique local chemistry optimized for specific biochemical reactions. In addition, some biomolecular condensates may function as depots that sequester biochemical reactions.

Biomolecular condensates are dynamic, which means they have a finite lifetime, and their size, composition, and material properties can change over time. This fact suggests biomolecular condensates as integral parts of a dynamic, responsive system beyond just being static reactors or sequesters of biochemical processes. In turn, parameters associated with condensate dynamics may play key roles in defining condensates' overall functionality. In this thesis, I focus on studying condensates' lifetime, the regulation of which acts as a timer of biochemical processes as well as a controller of substrate-product turnover. For example, in some cases, condensates must be present for certain biochemical processes to continue; in other cases, condensates must dissolve followed by reforming to replace the product with fresh substrate. Understanding the regulation of condensate lifetime equals understanding the condensate dynamics of formation and disassembly.

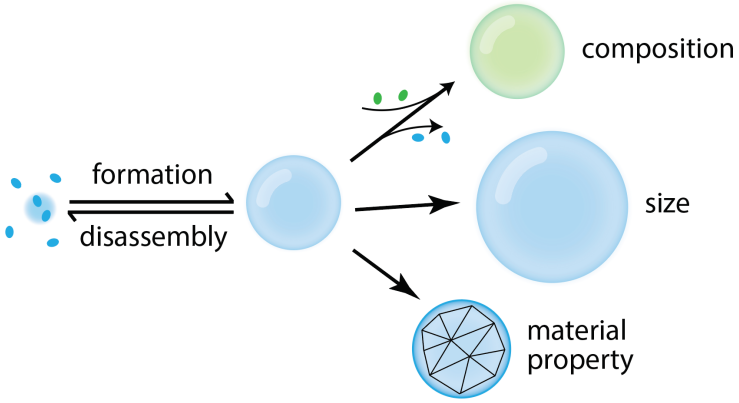


Figure 1.5 | Graphical illustration of condensate dynamics.

A condensate experiences formation and disassembly dynamics, resulting in a finite lifetime. Its size, composition, and material properties can change over time.

Mechanisms driving condensate formation are well understood in physical theories, biological regulations, and molecular features. Several physical theories can explain condensate formation, including liquid-liquid phase separation (LLPS), first-order phase transition, percolation, and surface condensation. LLPS is a prevailing model to explain how condensates form. The theory for biomolecular LLPS comes from polymer physics (e.g., Flory-Huggins Theory) (106), where a polymer solution can go from one well-mixed state to a de-mixed state, separating into a dilute and dense phase. The state of this system is concentration-dependent, meaning that the system will favor a de-mixed state over a well-mixed state above a critical concentration threshold (C_{sat}). Accordingly, specific biological regulations are required to drive phase transition. For example, recent works proposed that transcription factors (TFs) initiate condensate formation through concentration-dependent nucleation: multiple TFs bind to the gene to be activated, causing the local TF concentration to go above C_{sat} thus leading to the recruitment of additional proteins and the formation of a transcriptional condensate (82, 83, 90, 107). Molecular features have evolved to drive LLPS, with one key feature being multivalency. Multivalency can be achieved through a series of folded protein

domains, such as SH3 domains and proline-rich motifs (108, 109). Furthermore, intrinsically disordered regions (IDRs) – regions of proteins that lack a defined three-dimensional structure – establish the multivalency needed for condensate formation through low-affinity interactions. For example, virtually all TFs harbor IDR domains (well-recognized as “activation domains”) that mediate protein-protein interactions (82, 90). In addition, the system also prefers a de-mixed state with enhanced interaction affinity. The interaction affinity depends on the specific chemical composition of peptides and nucleic acids and can be altered by parameters such as pH and salt concentration (110-113).

In contrast, biological regulations that reverse the formation process and lead to condensate disassembly are much less explored. For example, regulations of transcriptional condensate formation have been studied extensively (82, 84, 90), but those involved in transcriptional condensate disassembly remain poorly understood. One appealing hypothesis involves negative feedback mechanisms, where the consequence of transcriptional condensate formation leads to the dissolution of the condensate. In Chapter 4, we proposed a charge balance model that couples the condensate formation-disassembly dynamics of positively charged transcriptional proteins and bursting of negatively charged RNA in a feedback fashion: low levels of RNA synthesis at the initiation step of gene expression promote condensate formation, which leads to high levels of RNA bursting that dissolves the condensate. This charge-balance-dependent condensate stability has been demonstrated in the field of colloid physics as reentrant phase behavior, which is driven by favorable opposite-charge interactions at low RNA concentrations and repulsive like-charge interactions at high RNA concentrations (114). Therefore, the regulation of transcription may be understood by a non-equilibrium RNA feedback control of transcriptional condensates, where the charge balance of electrostatic interactions can account for RNA-mediated feedback control of transcriptional condensates.

The rationale of regular condensate dynamics being critical for normal cellular functions implies that the dysregulation of condensate dynamics may be linked to disease pathology. Insulin receptor (IR) signaling is central to normal metabolic control and is dysregulated in metabolic diseases such as type 2 diabetes (T2D) (115, 116). In Chapter 5, we found that IRs are incorporated in dynamic condensates in normal cells to perform metabolic control. In insulin-resistant cells which could occur in T2D, IR signaling is dysregulated characterized by blunted downstream response. This dysregulated signaling correlates with diminished IR condensate dynamics of formation and disassembly: fewer and smaller IR condensates form in insulin-resistant cells, and the lifetime of those condensates increases. The correlation between blunted downstream signaling and diminished condensate dynamics might be explained as the suppressed substrate-product turnover rate in insulin resistance. Treating insulin-resistant cells with metformin, a first-line drug used to treat T2D, can rescue the IR condensate dynamics of formation and disassembly. Therefore, the dynamic formation and disassembly of signaling condensates may be an important means of metabolic control, whose disruption is linked to disease pathology.

Chapter 1.6: Techniques for studying protein spatiotemporal dynamics

Knowing proteins' detailed diffusion and condensation dynamics is crucial for understanding cellular function, where imaging-based techniques are helpful. Fluorescence microscopy integrating fluorophore technology, genetic encoding, and modern microscopy allows the versatile observation of the live motion of proteins in cells. Depending on the visible protein density, there are bulk-intensity-based methods and single-molecule-localization-based methods. In this thesis, I implemented one bulk-intensity-based method and two single-molecule-localization-based methods to investigate protein spatiotemporal dynamics in disease pathology and gene regulation.

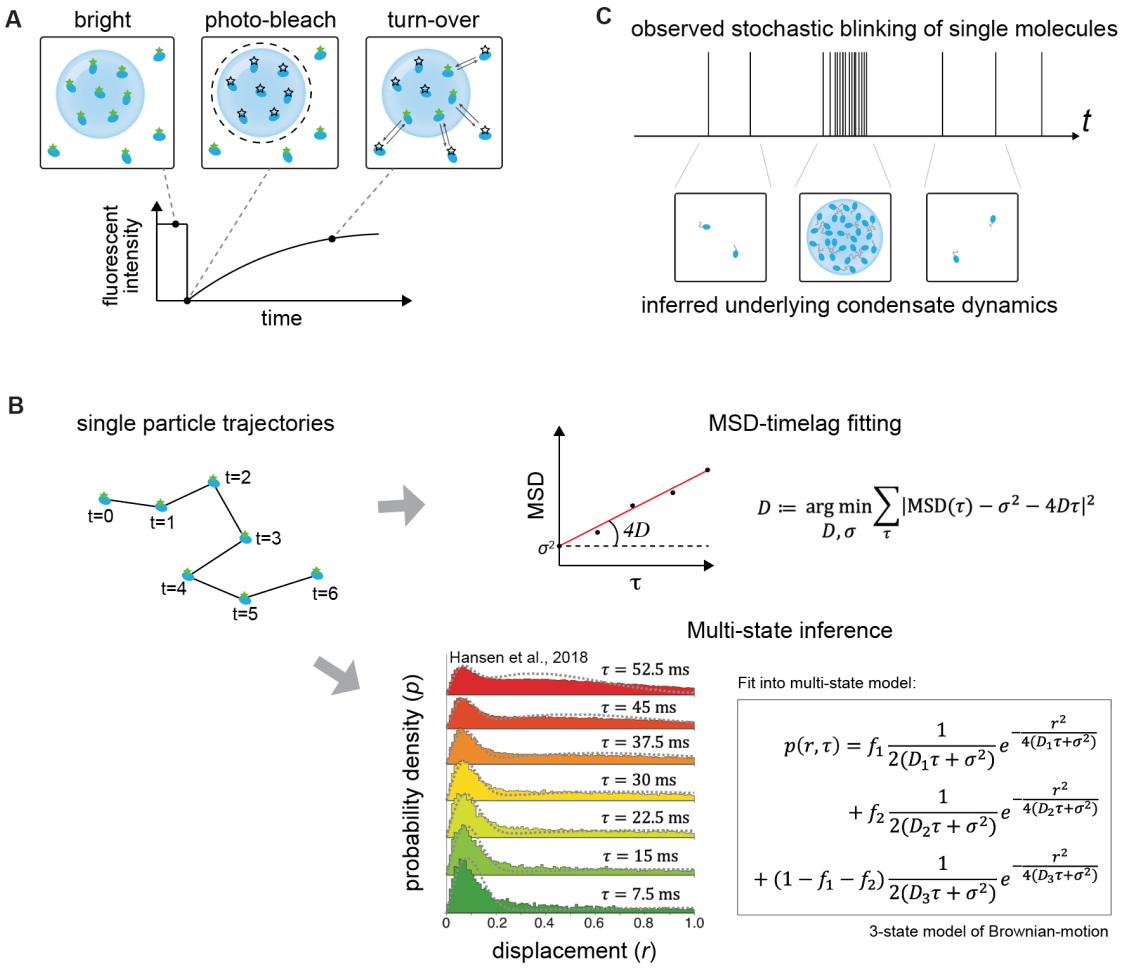


Figure 1.6 | Techniques for studying protein spatiotemporal dynamics.

A. Illustration of fluorescence recovery after photobleaching (FRAP). Fluorescence recovery is caused by turn-over of photobleached fluorophores inside a condensate with intact fluorophores outside the condensate.

B. Illustration of single-particle tracking (SPT). Single particle trajectories can be converted to different representations (e.g., MSD-τ relation or probability density distributions of *r* and τ) for different analyses.

C. Illustration of time-correlated photo-activation localization microscopy (tcPALM). The bursting of single molecule signals can be used to infer the condensate dynamics of formation and disassembly.

Fluorescence recovery after photobleaching (FRAP) infers bulk transit dynamics by photobleaching the fluorophores within a certain region followed by monitoring the recovery of fluorescence intensity of that region (34, 36, 117). The maximum degree of recovery indicates the mobile fraction within the imaging time window, and the rate of recovery indicates the mobility of the mobile molecules. In Chapter 2, FRAP was demonstrated as a robust method for investigating bulk transit dynamics of diverse proteins in normal versus pathogenic conditions, highlighting the decreased condensate liquidity in pathogenic conditions presumably due to intermolecular crosslinking under oxidative stress.

Single-particle tracking (SPT) traces the localizations of single molecules over time (Figure 1.6B) (37, 118). This method captures transit dynamics of individual proteins with unprecedented details based on single-molecule trajectories: the apparent diffusion rate can be calculated from each trajectory through MSD-timelag fitting, and the partition of molecules with different diffusivity can be estimated from a collection of trajectories via multi-state inference (Figure 1.6B). In Chapter 2, SPT empowered the discovery of a new molecular mechanism for chronic disease, where pathogenic stimuli decrease the apparent diffusion rate for proteins enriched in various subcellular regions. In Chapter 3, SPT revealed a differential partition of diffusivity of transcription factors (TFs) after losing the RNA-binding ability, implying a shift of TF localization from active sites of transcription to nucleoplasm and revealing a regulatory role of TF-RNA binding in gene activation.

Time-correlated photo-activation localization microscopy (tcPALM) detects transient condensation dynamics of proteins with single-molecule sensitivity (Figure 1.6C) (40). By stochastic photo-activation of fluorophores within a condensate when it's present, tcPALM can resolve the condensate beyond the diffraction limit and infer the underlying condensate dynamics (Figure 1.6C). In Chapter 4, I used tcPALM to monitor the persistency of transcriptional condensates, providing direct evidence of RNA-mediated feedback control of transcriptional condensate disassembly dynamics. In Chapter 5, I used tcPALM to resolve the lifetime of insulin-receptor condensates under insulin-sensitive vs insulin-resistant conditions and reveal reduced receptor condensate dynamics of formation and disassembly in insulin resistance, providing novel insights into the pathology of type 2 diabetes.

In addition to imaging-based techniques, appropriate probe design facilitates the research about protein spatiotemporal dynamics. In Chapter 2, I designed a protein mobility biosensor compatible with SPT and FREP. For endogenous proteins, the mobility response to a perturbation virtually always results from a series of entangled mechanisms and/or pathways, making it challenging to have a clear understanding of any molecular mechanisms that may lead to proteolethargy. This biosensor can detect small changes in protein diffusivity solely caused by inter-protein oxidative modifications and is insensitive to other cellular effects. Therefore, this biosensor and its variations are an ideal set of probes to measure protein mobility response to any desired biochemical modifications specifically and are convenient to be implemented in multiple cell lines and disease conditions.

References

1. Alberts B, Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. Molecular biology of the cell. New York: Garland Science. 2003(4th edition).

2. Edidin M, Fambrough D. Fluidity of the surface of cultured muscle fibers. Rapid lateral diffusion of marked surface antigens. J Cell Biol. 1973;57(1):27-37.
3. Liebman PA, Entine G. Lateral diffusion of visual pigment in photorecptor disk membranes. Science. 1974;185(4149):457-9.
4. Poo M, Cone RA. Lateral diffusion of rhodopsin in the photoreceptor membrane. Nature. 1974;247(5441):438-41.
5. Richards PM. Diffusion to nonoverlapping or spatially correlated traps. Phys Rev B Condens Matter. 1987;35(1):248-56.
6. McGuffee SR, Elcock AH. Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. PLoS Comput Biol. 2010;6(3):e1000694.
7. Jaqaman K, Kuwata H, Touret N, Collins R, Trimble WS, Danuser G, et al. Cytoskeletal control of CD36 diffusion promotes its receptor and signaling function. Cell. 2011;146(4):593-606.
8. Persson F, Linden M, Unoson C, Elf J. Extracting intracellular diffusive states and transition rates from single-molecule tracking data. Nat Methods. 2013;10(3):265-9.
9. Guo M, Gelman H, Gruebele M. Coupled protein diffusion and folding in the cell. PLoS One. 2014;9(12):e113040.
10. Krapf D. Mechanisms underlying anomalous diffusion in the plasma membrane. Curr Top Membr. 2015;75:167-207.
11. Esadze A, Stivers JT. Facilitated Diffusion Mechanisms in DNA Base Excision Repair and Transcriptional Activation. Chem Rev. 2018;118(23):11298-323.
12. Schavemaker PE, Boersma AJ, Poolman B. How Important Is Protein Diffusion in Prokaryotes? Front Mol Biosci. 2018;5:93.
13. Woringer M, Darzacq X. Protein motion in the nucleus: from anomalous diffusion to weak interactions. Biochem Soc Trans. 2018;46(4):945-56.
14. von Bulow S, Siggel M, Linke M, Hummer G. Dynamic cluster formation determines viscosity and diffusion in dense protein solutions. Proc Natl Acad Sci U S A. 2019;116(20):9843-52.
15. Briane V, Vimond M, Kervrann C. An overview of diffusion models for intracellular dynamics analysis. Brief Bioinform. 2020;21(4):1136-50.
16. Alejo JL, Kempes CP, Adamala KP. Diffusion control in biochemical specificity. Biophys J. 2022;121(8):1541-8.
17. Bigman LS, Levy Y. Protein Diffusion Along Protein and DNA Lattices: Role of Electrostatics and Disordered Regions. Annu Rev Biophys. 2023;52:463-86.
18. Di Bari D, Timr S, Guiral M, Giudici-Orticoni MT, Seydel T, Beck C, et al. Diffusive Dynamics of Bacterial Proteome as a Proxy of Cell Death. ACS Cent Sci. 2023;9(1):93-102.
19. Hyman AA, Weber CA, Julicher F. Liquid-liquid phase separation in biology. Annu Rev Cell Dev Biol. 2014;30:39-58.
20. Banani SF, Lee HO, Hyman AA, Rosen MK. Biomolecular condensates: organizers of cellular biochemistry. Nat Rev Mol Cell Biol. 2017;18(5):285-98.
21. Hnisz D, Shrinivas K, Young RA, Chakraborty AK, Sharp PA. A Phase Separation Model for Transcriptional Control. Cell. 2017;169(1):13-23.
22. Shin Y, Brangwynne CP. Liquid phase condensation in cell physiology and disease. Science. 2017;357(6357).
23. Sabari BR, Dall'Agnese A, Young RA. Biomolecular Condensates in the Nucleus. Trends Biochem Sci. 2020;45(11):961-77.
24. Feric M, Misteli T. Function moves biomolecular condensates in phase space. Bioessays. 2022;44(5):e2200001.

25. Berg H. Random Walks in Biology Expanded Edition. Princeton University Press. 1993(1993).
26. Cooper G AK. The cell: a molecular approach2022 Oct 26.
27. Volkmann D, Baluska F, Menzel D. Eduard Strasburger (1844-1912): founder of modern plant cell biology. Protoplasma. 2012;249(4):1163-72.
28. Wilson EB. The Structure of Protoplasm. Science. 1899;10(237):33-45.
29. Newburgh R, Joseph Peidle, and Wolfgang Rueckner. Einstein, Perrin, and the reality of atoms: 1905 revisited. American journal of physics. 2006;74(6):478-81.
30. Caro LG, Palade GE. Protein Synthesis, Storage, and Discharge in the Pancreatic Exocrine Cell. An Autoradiographic Study. J Cell Biol. 1964;20(3):473-95.
31. Palade GES, P.; Caro L.G. Structure, Chemistry and Function of the Pancreatic Exocrine Cell. Novartis Foundation Symposia Ciba Foundation Symposium - The Exocrine Pancreas: Normal and Abnormal Functions. 1962:23 - 55.
32. Frye LD, Edidin M. The rapid intermixing of cell surface antigens after formation of mouse-human heterokaryons. J Cell Sci. 1970;7(2):319-35.
33. Lippincott-Schwartz J, Snapp E, Kenworthy A. Studying protein dynamics in living cells. Nat Rev Mol Cell Biol. 2001;2(6):444-56.
34. Axelrod D, Koppel DE, Schlessinger J, Elson E, Webb WW. Mobility measurement by analysis of fluorescence photobleaching recovery kinetics. Biophys J. 1976;16(9):1055-69.
35. Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC. Green fluorescent protein as a marker for gene expression. Science. 1994;263(5148):802-5.
36. Lippincott-Schwartz J, Snapp EL, Phair RD. The Development and Enhancement of FRAP as a Key Tool for Investigating Protein Dynamics. Biophys J. 2018;115(7):1146-55.
37. Wieser S, Schutz GJ. Tracking single molecules in the live cell plasma membrane-Do's and Don't's. Methods. 2008;46(2):131-40.
38. Betzig E, Patterson GH, Sougrat R, Lindwasser OW, Olenych S, Bonifacino JS, et al. Imaging intracellular fluorescent proteins at nanometer resolution. Science. 2006;313(5793):1642-5.
39. Rust MJ, Bates M, Zhuang X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). Nat Methods. 2006;3(10):793-5.
40. Cisse, II, Izeddin I, Causse SZ, Boudarene L, Senecal A, Muresan L, et al. Real-time dynamics of RNA polymerase II clustering in live human cells. Science. 2013;341(6146):664-7.
41. Lu J, Wu T, Zhang B, Liu S, Song W, Qiao J, et al. Types of nuclear localization signals and mechanisms of protein import into the nucleus. Cell Commun Signal. 2021;19(1):60.
42. Wang H, Fang B, Peng B, Wang L, Xue Y, Bai H, et al. Recent Advances in Chemical Biology of Mitochondria Targeting. Front Chem. 2021;9:683220.
43. Fisher RP. The CDK Network: Linking Cycles of Cell Division and Gene Expression. Genes Cancer. 2012;3(11-12):731-8.
44. Guillaud L, El-Agamy SE, Otsuki M, Terenzio M. Anterograde Axonal Transport in Neuronal Homeostasis and Disease. Front Mol Neurosci. 2020;13:556175.
45. Guedes-Dias P, Holzbaur ELF. Axonal transport: Driving synaptic function. Science. 2019;366(6462).
46. Cramer P. Organization and regulation of gene transcription. Nature. 2019;573(7772):45-54.
47. Cramer P. Eukaryotic Transcription Turns 50. Cell. 2019;179(4):808-12.
48. Iwafuchi-Doi M, Zaret KS. Pioneer transcription factors in cell reprogramming. Genes Dev. 2014;28(24):2679-92.
49. Lee TI, Young RA. Transcriptional regulation and its misregulation in disease. Cell. 2013;152(6):1237-51.

50. Thomas MC, Chiang CM. The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol. 2006;41(3):105-78.
51. Sikorski TW, Buratowski S. The basal initiation machinery: beyond the general transcription factors. Curr Opin Cell Biol. 2009;21(3):344-51.
52. Muller F, Tora L. Chromatin and DNA sequences in defining promoters for transcription initiation. Biochim Biophys Acta. 2014;1839(3):118-28.
53. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell. 2013;153(2):307-19.
54. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The Human Transcription Factors. Cell. 2018;175(2):598-9.
55. Stadhouders R, Filion GJ, Graf T. Transcription factors and 3D genome conformation in cell-fate decisions. Nature. 2019;569(7756):345-54.
56. Dundr M, Hoffmann-Rohrer U, Hu Q, Grummt I, Rothblum LI, Phair RD, et al. A kinetic framework for a mammalian RNA polymerase in vivo. Science. 2002;298(5598):1623-6.
57. Liu X, Bushnell DA, Kornberg RD. RNA polymerase II transcription: structure and mechanism. Biochim Biophys Acta. 2013;1829(1):2-8.
58. Nguyen VQ, Ranjan A, Liu S, Tang X, Ling YH, Wisniewski J, et al. Spatiotemporal coordination of transcription preinitiation complex assembly in live cells. Mol Cell. 2021;81(17):3560-75 e6.
59. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. Cell. 2013;155(4):934-47.
60. Thomas HF, Buecker C. What is an enhancer? Bioessays. 2023;45(10):e2300044.
61. Pennacchio LA, Bickmore W, Dean A, Nobrega MA, Bejerano G. Enhancers: five essential questions. Nat Rev Genet. 2013;14(4):288-95.
62. Schoenfelder S, Fraser P. Long-range enhancer-promoter contacts in gene expression control. Nat Rev Genet. 2019;20(8):437-55.
63. Furlong EEM, Levine M. Developmental enhancers and chromosome topology. Science. 2018;361(6409):1341-5.
64. Krivega I, Dean A. Enhancer and promoter interactions-long distance calls. Curr Opin Genet Dev. 2012;22(2):79-85.
65. Barral A, Dejardin J. The chromatin signatures of enhancers and their dynamic regulation. Nucleus. 2023;14(1):2160551.
66. Yang JH, Hansen AS. Enhancer selectivity in space and time: from enhancer-promoter interactions to promoter activation. Nat Rev Mol Cell Biol. 2024.
67. Fukaya T, Lim B, Levine M. Enhancer Control of Transcriptional Bursting. Cell. 2016;166(2):358-68.
68. Fong YW, Cattoglio C, Yamaguchi T, Tjian R. Transcriptional regulation by coactivators in embryonic stem cells. Trends Cell Biol. 2012;22(6):292-8.
69. Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, et al. Mediator and cohesin connect gene expression and chromatin architecture. Nature. 2010;467(7314):430-5.
70. Kelleher RJ, 3rd, Flanagan PM, Kornberg RD. A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. Cell. 1990;61(7):1209-15.
71. Lorch Y, Kornberg RD. Chromatin-remodeling for transcription. Q Rev Biophys. 2017;50:e5.
72. Lorch Y, LaPointe JW, Kornberg RD. Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. Cell. 1987;49(2):203-10.
73. Kang Y, Kim YW, Kang J, Kim A. Histone H3K4me1 and H3K27ac play roles in nucleosome eviction and eRNA transcription, respectively, at enhancers. FASEB J. 2021;35(8):e21781.

74. Peng Y, Song W, Teif VB, Ovcharenko I, Landsman D, Panchenko AR. Detection of new pioneer transcription factors as cell-type specific nucleosome binders. bioRxiv. 2023.
75. Zhu F, Farnung L, Kaasinen E, Sahu B, Yin Y, Wei B, et al. The interaction landscape between transcription factors and the nucleosome. Nature. 2018;562(7725):76-81.
76. Izeddin I, Recamier V, Bosanac L, Cisse, II, Boudarene L, Dugast-Darzacq C, et al. Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus. Elife. 2014;3.
77. Wang Z, Deng W. Dynamic transcription regulation at the single-molecule level. Dev Biol. 2022;482:67-81.
78. Mueller F, Stasevich TJ, Mazza D, McNally JG. Quantifying transcription factor kinetics: at work or at play? Crit Rev Biochem Mol Biol. 2013;48(5):492-514.
79. Hager GL, McNally JG, Misteli T. Transcription dynamics. Mol Cell. 2009;35(6):741-53.
80. Elf J, Li GW, Xie XS. Probing transcription factor dynamics at the single-molecule level in a living cell. Science. 2007;316(5828):1191-4.
81. Wagh K, Stavreva DA, Upadhyaya A, Hager GL. Transcription Factor Dynamics: One Molecule at a Time. Annu Rev Cell Dev Biol. 2023;39:277-305.
82. Sabari BR, Dall'Agnese A, Boija A, Klein IA, Coffey EL, Shrinivas K, et al. Coactivator condensation at super-enhancers links phase separation and gene control. Science. 2018;361(6400).
83. Cho WK, Spille JH, Hecht M, Lee C, Li C, Grube V, et al. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science. 2018;361(6400):412-5.
84. Shrinivas K, Sabari BR, Coffey EL, Klein IA, Boija A, Zamudio AV, et al. Enhancer Features that Drive Formation of Transcriptional Condensates. Mol Cell. 2019;75(3):549-61 e7.
85. Guo YE, Manteiga JC, Henninger JE, Sabari BR, Dall'Agnese A, Hannett NM, et al. Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. Nature. 2019;572(7770):543-8.
86. Li CH, Coffey EL, Dall'Agnese A, Hannett NM, Tang X, Henninger JE, et al. MeCP2 links heterochromatin condensates and neurodevelopmental disease. Nature. 2020;586(7829):440-4.
87. Du M, Stitzinger SH, Spille JH, Cho WK, Lee C, Hijaz M, et al. Direct observation of a condensate effect on super-enhancer controlled gene bursting. Cell. 2024;187(2):331-44 e17.
88. Boehning M, Dugast-Darzacq C, Rankovic M, Hansen AS, Yu T, Marie-Nelly H, et al. RNA polymerase II clustering through carboxy-terminal domain phase separation. Nat Struct Mol Biol. 2018;25(9):833-40.
89. Nair SJ, Yang L, Meluzzi D, Oh S, Yang F, Friedman MJ, et al. Phase separation of ligand-activated enhancers licenses cooperative chromosomal enhancer assembly. Nat Struct Mol Biol. 2019;26(3):193-203.
90. Boija A, Klein IA, Sabari BR, Dall'Agnese A, Coffey EL, Zamudio AV, et al. Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell. 2018;175(7):1842-55 e16.
91. Chong S, Dugast-Darzacq C, Liu Z, Dong P, Dailey GM, Cattoglio C, et al. Imaging dynamic and selective low-complexity domain interactions that control gene transcription. Science. 2018;361(6400).
92. Cho WK, Jayanth N, English BP, Inoue T, Andrews JO, Conway W, et al. RNA Polymerase II cluster dynamics predict mRNA output in living cells. Elife. 2016;5.
93. Good DJ. Non-Coding RNAs in Human Health and Diseases. Genes (Basel). 2023;14(7).
94. Mattick JS, Amaral PP, Carninci P, Carpenter S, Chang HY, Chen LL, et al. Long non-coding RNAs: definitions, functions, challenges and recommendations. Nat Rev Mol Cell Biol. 2023;24(6):430-47.

95. Ferrer J, Dimitrova N. Transcription regulation by long non-coding RNAs: mechanisms and disease relevance. Nat Rev Mol Cell Biol. 2024.
96. Li J, Ming Z, Yang L, Wang T, Liu G, Ma Q. Long noncoding RNA XIST: Mechanisms for X chromosome inactivation, roles in sex-biased diseases, and therapeutic opportunities. Genes Dis. 2022;9(6):1478-92.
97. Mousavi K, Zare H, Dell'orso S, Grontved L, Gutierrez-Cruz G, Derfoul A, et al. eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. Mol Cell. 2013;51(5):606-17.
98. Tsai PF, Dell'Orso S, Rodriguez J, Vivanco KO, Ko KD, Jiang K, et al. A Muscle-Specific Enhancer RNA Mediates Cohesin Recruitment and Regulates Transcription In trans. Mol Cell. 2018;71(1):129-41 e8.
99. Hertzog M, Erdel F. The Material Properties of the Cell Nucleus: A Matter of Scale. Cells. 2023;12(15).
100. Delarue M, Brittingham GP, Pfeffer S, Surovtsev IV, Pinglay S, Kennedy KJ, et al. mTORC1 Controls Phase Separation and the Biophysical Properties of the Cytoplasm by Tuning Crowding. Cell. 2018;174(2):338-49 e20.
101. Fakhri N, Wessel AD, Willms C, Pasquali M, Klopfenstein DR, MacKintosh FC, et al. High-resolution mapping of intracellular fluctuations using carbon nanotubes. Science. 2014;344(6187):1031-5.
102. Brangwynne CP, Eckmann CR, Courson DS, Rybarska A, Hoege C, Gharakhani J, et al. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. Science. 2009;324(5935):1729-32.
103. Zamudio AV, Dall'Agnese A, Henninger JE, Manteiga JC, Afeyan LK, Hannett NM, et al. Mediator Condensates Localize Signaling Factors to Key Cell Identity Genes. Mol Cell. 2019;76(5):753-66 e6.
104. Jaqaman K, Ditlev JA. Biomolecular condensates in membrane receptor signaling. Curr Opin Cell Biol. 2021;69:48-54.
105. Su Q, Mehta S, Zhang J. Liquid-liquid phase separation: Orchestrating cell signaling through time and space. Mol Cell. 2021;81(20):4137-46.
106. Huggins ML. Some properties of solutions of long-chain compounds. J Phys Chem. 1942;46:151–8.
107. Lyons H, Veettil RT, Pradhan P, Fornero C, De La Cruz N, Ito K, et al. Functional partitioning of transcriptional regulators by patterned charge blocks. Cell. 2023;186(2):327-45 e28.
108. Li P, Banjade S, Cheng HC, Kim S, Chen B, Guo L, et al. Phase transitions in the assembly of multivalent signalling proteins. Nature. 2012;483(7389):336-40.
109. Banjade S, Rosen MK. Phase transitions of multivalent proteins can promote clustering of membrane receptors. Elife. 2014;3.
110. Morishita K, Watanabe K, Naguro I, Ichijo H. Sodium ion influx regulates liquidity of biomolecular condensates in hyperosmotic stress response. Cell Rep. 2023;42(4):112315.
111. Wang J, Choi JM, Holehouse AS, Lee HO, Zhang X, Jahnel M, et al. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. Cell. 2018;174(3):688-99 e16.
112. Martin EW, Holehouse AS, Peran I, Farag M, Incicco JJ, Bremer A, et al. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. Science. 2020;367(6478):694-9.
113. Jin X, Zhou M, Chen S, Li D, Cao X, Liu B. Effects of pH alterations on stress- and aging-induced protein phase separation. Cell Mol Life Sci. 2022;79(7):380.

114. Banerjee PR, Milin AN, Moosa MM, Onuchic PL, Deniz AA. Reentrant Phase Transition Drives Dynamic Substructure Formation in Ribonucleoprotein Droplets. Angew Chem Int Ed Engl. 2017;56(38):11354-9.
115. Batista TM, Haider N, Kahn CR. Defining the underlying defect in insulin action in type 2 diabetes. Diabetologia. 2021;64(5):994-1006.
116. James DE, Stockli J, Birnbaum MJ. The aetiology and molecular landscape of insulin resistance. Nat Rev Mol Cell Biol. 2021;22(11):751-71.
117. Wustner D, Solanko LM, Lund FW, Sage D, Schroll HJ, Lomholt MA. Quantitative fluorescence loss in photobleaching for analysis of protein transport and aggregation. BMC Bioinformatics. 2012;13:296.
118. Qian H, Sheetz MP, Elson EL. Single particle tracking. Analysis of diffusion and flow in two-dimensional systems. Biophys J. 1991;60(4):910-21.

Chapter 2:

Proteolethargy is a pathogenic mechanism in chronic disease

Alessandra Dall'Agnese[1†], Ming M. Zheng[1,2†], Shannon Moreno[1,3†], Jesse M. Platt[1,4], An T. Hoang[1], Deepti Kannan[2], Giuseppe Dall'Agnese[1], Kalon J. Overholt[1,5], Ido Sagi[1], Nancy M. Hannett[1], Hailey Erb[1], Olivia Corradin[1,3], Arup K. Chakraborty[2,6,7,8,9], Tong Ihn Lee[1*], Richard A. Young[1,3*]

[1] Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA
[2] Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
[3] Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
[4] Division of Gastroenterology, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts 02114, USA
[5] Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
[6] Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
[7] Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
[8] Institute of Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
[9] Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts 02139, USA.

† These authors contributed equally to this work
* Corresponding authors. Email: young@wi.mit.edu and tlee@wi.mit.edu

Abstract

The pathogenic mechanisms of many diseases are well understood at the molecular level, but there are prevalent syndromes associated with pathogenic signaling, such as diabetes and chronic inflammation, where our understanding is more limited. Here we report that pathogenic signaling suppresses the mobility of a spectrum of proteins that play essential roles in cellular functions known to be dysregulated in these chronic diseases. The reduced protein mobility, which we call proteolethargy, was linked to cysteine residues in the affected proteins and signaling-related increases in excess reactive oxygen species. Diverse pathogenic stimuli, including hyperglycemia, dyslipidemia and inflammation, produce similar reduced protein mobility phenotypes. We propose that proteolethargy is an overlooked cellular mechanism that may account for various pathogenic features of diverse chronic diseases.

Main Text

Diseases associated with chronic or pathogenic signaling are a leading cause of morbidity and mortality[1]. For prevalent syndromes such as diabetes and inflammatory disorders, the pathology typically involves a continuous and/or high-level stimulus but not necessarily a known mutation in a specific gene[2-7]. In contrast with monogenic diseases, where the causal link between gene mutation and disease pathology is evident and the cellular pathways directly impacted are thus defined, in chronic syndromes causal gene mutations are uncommon and diverse cellular processes such as gene regulation, ribosome biosynthesis and metabolic activity are dysregulated[8-18]. Thus, how to define hypotheses that will inform therapeutic development on the basis of such a breadth of cellular dysfunction has long vexed clinicians and research scientists.

The billions of protein molecules produced in cells must leave their site of synthesis and arrive at cellular locations where they carry out their specialized functions[19-26]. In so doing, they will transit through a milieu that is densely packed with biomolecules[19-24,27]. Recently, pathogenic signaling in certain chronic diseases was reported to cause reduced movement of receptor molecules into functional protein assemblies[28-30]. These findings led us to consider the possibility that dysregulated signaling might cause a more general defect in protein mobility in cells, and that reduced protein mobility in and of itself might be a pathogenic mechanism shared across these diseases. Biochemical reactions are often collision-limited[31,32], and reduced rates of protein diffusion would therefore be expected to reduce functional outputs.

Here, we show that pathogenic signaling reduces the mobility of key proteins involved in diverse cellular processes, and that this reduction in protein mobility, which we call proteolethargy, is associated with a dysregulated redox environment that consequently impacts oxidation-sensitive cysteines. Reduced protein mobility may account for the diversity of dysregulated cellular processes that are evident in chronic disease. We discuss a therapeutic hypothesis that emerges from these findings that might prove to be applicable in patients with diseases associated with proteolethargy.

Protein mobility in cells

We set out to develop a theoretical and experimental framework that would allow us to measure the mobility of multiple proteins with diverse functions in cells subjected to normal and pathogenic signaling. Single particle tracking (SPT) and fluorescence recovery after photobleaching (FRAP) allow measurement of the kinetics of protein mobility in living cells, and proteins studied with these methods have been reported to have average apparent diffusion coefficients that vary between 0.01 $\mu m^2/s$ and 30 $\mu m^2/s$[25,26,33-37]. This variation is thought to reflect that protein mobility is influenced by diverse factors, ranging from protein size to interaction with various biomolecules[33]. For our study, we selected proteins whose functions are key to cellular processes known to be dysregulated in prevalent syndromes[8-18]: a plasma membrane receptor (insulin receptor, IR), a transcriptional cofactor present at actively transcribed genes (Mediator subunit MED1), a regulator of silent genes in heterochromatin (heterochromatin protein

HP1α), a component of the nucleolus involved in ribosome biosynthesis (fibrillarin, FIB1) and a subunit of the mRNA splicing apparatus (serine and arginine-rich splicing factor 2, SRSF2) (Figure 1A; Methods). As a cell model, we chose HepG2 cells as they provide a well-established model system representative of human liver cells in healthy and disease states[28,38-40]. To monitor the mobility of each of these proteins, we engineered HepG2 cells to encode the endogenous protein fused with HaloTag or monomeric enhanced green fluorescence protein (GFP) and validated that each fusion protein was produced at normal levels and migrated to the compartment where it is known to function (Figure 1B,C and S1).

We determined the apparent diffusion coefficients of IR, MED1, HP1α, FIB1 and SRSF2 by SPT based on the mean-squared displacement of each individual protein molecule's trajectory. For each protein, we measured at least 200 protein trajectories and plotted the distribution of apparent diffusion coefficients (Figure 1D,E). As expected, most SPT protein trajectories for IR were contained within the plasma membrane and most SPT protein trajectories for MED1, HP1α, FIB1 and SRSF2 were contained within the nucleus (Figure 1D). The apparent diffusion coefficients ranged from 0.01 $\mu m^2/s$ to 28 $\mu m^2/s$ (Figure 1E), consistent with diffusion coefficients determined for other human proteins (Table S1), with SRSF2 having the highest average mobility and FIB1 the lowest (Figure 1E, S2A).

We also used FRAP to measure the mobility of proteins in HepG2 cells engineered to express the endogenous protein fused to GFP. Specifically, we bleached a selected region in the cell with a focused laser beam and measured the rate at which the fluorescence intensity recovered at the photobleached region. This fluorescence recovery reflects the average mobility of the bulk population of fluorescent proteins[41]. For all proteins under study, the fluorescence intensity recovered on a time scale of seconds (Figure 1F,G) and the relative mobilities of the proteins were in line with those determined using SPT (Figure S2). The mobility of all the proteins measured in these studies was within the range determined previously for other proteins in living cells[33-36,42-45].

Reduced protein mobility with pathogenic signaling

As an initial test of the hypothesis that protein mobility might be affected in a chronic disease (Figure 2A), we selected insulin signaling, since it is dysregulated in prevalent syndromes such as diabetes, known to be characterized by a range of affected cellular processes, including dysregulated intracellular signaling, gene activity, RNA splicing and ribosome biosynthesis, among others[5-12]. In fasting healthy individuals, liver cells are normally exposed to low concentrations of insulin (~ 0.1 nM), whereas after a meal, insulin transiently increases and activates the insulin signaling pathway[28,46,47]. In fasting patients with insulin resistance, liver cells are subject to continuous high concentrations of insulin (~3 nM), and this chronic high level of insulin no longer fully activates the signaling response[28,46,47]. Thus, normal and pathogenic insulin signaling can be

modeled in cell culture by treating liver-derived cells with normal or elevated (pathogenic) concentrations of insulin for prolonged periods of time (Figure 2B)[28,48].

To test the possibility that pathogenic insulin signaling may alter protein mobility, we treated HepG2 cells with normal or pathogenic concentrations of insulin (Figure 2C). SPT analysis revealed that the mobility of IR, MED1, HP1α, and FIB1 was reduced in cells that were treated with pathogenic levels of insulin, whereas that of SRSF2 was unaffected (Figure 2D,E, Table S2). For example, 50% of IR molecules had an apparent diffusion coefficient greater or equal to 0.1 $\mu m^2/s$ when cells were treated with normal concentrations of insulin, and this percentage decreased by ~20% when cells were treated with pathogenic insulin concentrations (Figure 2D,E). FRAP analysis of these proteins indicated a similar effect on this set of proteins; there was a reduction in the recovery of all proteins except SRSF2 (Figure 2F,G, Table S3). Taken together, these results suggest that pathogenic insulin signaling leads to a reduction in the mobility of many proteins in cells.

The proteins studied here have been reported to be associated with biomolecular condensates[28,44,49-52], which are non-stoichiometric assemblies of proteins that share cellular functions[25,44,53-55]. We thus tested whether suppressed protein mobility occurs when proteins are resident within the dense phase of condensates or when they are outside these bodies using FRAP with the GFP-tagged proteins. For the proteins that could be reliably assigned to be within or outside of condensates during image acquisition (MED1, HP1α, FIB1 and SRSF2), pathogenic signaling was found to produce a similar reduction in mobility for MED1, HP1α and FIB1, while SRSF2 mobility was unaffected (Figure 2G and Figure S3A, Table S3). Rapid movement of IR condensates prevented reliable assignments. Pathogenic signaling had little effect on condensate number, size or partition ratio for these proteins, with the exception of a slight decrease on condensate number for IR, as observed previously[28] (Figure S3B). Although there are reports that ROS can influence the properties of some condensates[56-58], these results suggest that the effects of pathogenic signaling can produce changes in protein mobility while having limited impact on condensate properties under the conditions studied here.

Oxidative environment affects protein mobility

Given the broad range of proteins whose mobility was affected by pathogenic insulin signaling, we asked whether changes in cellular viscosity or in the chemical environment might be responsible for the observed changes in protein mobility. To test the effect of pathogenic signaling on cellular viscosity, we monitored the mobility of GFP (not fused to any other protein) by FRAP, which is an established method for such studies[31,59-61], and the mobility of HaloTag (not fused to any other protein) by SPT. We detected a change in cytoplasmic viscosity but no change in nuclear viscosity (Figure S4A,B). These results suggest that altered viscosity could contribute to the mobility phenotype for IR in the plasma membrane but is unlikely to significantly impact the diverse nuclear proteins studied here.

Substantial changes in the chemical environment are known features of chronic diseases such as insulin resistance due to high levels of reactive oxygen species (ROS) (Figure 3A, B)[28,62]. Here, we hypothesized that if an oxidative environment leads to changes in protein mobility, then treating cells with pathologically-relevant concentrations of the oxidizing agent H_2O_2 should phenocopy the effects observed in cells treated with pathogenic insulin signaling (Figure 3B,C). Indeed, FRAP analysis showed that treatment of cells with H_2O_2 caused reduced mobility of IR, MED1, HP1α, and FIB1 but not SRSF2 or nuclear GFP (Figure 3D-F, Figure S4C,D, Table S3).

If high levels of ROS lead to reductions in protein mobility, then treatment with the antioxidant N-acetyl cysteine (NAC) should restore some degree of protein mobility in cells exposed to pathogenic levels of insulin. As expected, FRAP revealed that treating insulin-resistant cells with 1 mM NAC partially rescued the mobility of IR, MED1, HP1α and FIB1, but had little effect on the mobility of SRSF2 and nuclear GFP (Figure 3G-I, Figure S4D,E, Table S3). These results are consistent with the possibility that elevated levels of ROS cause a decrease in the mobility of certain proteins and suggest that the change in protein behavior is caused by an alteration in the oxidative environment.

Mobility of proteins with exposed cysteines

The sensitivity of proteins to the oxidative environment suggests that oxidation-sensitive amino acids might influence protein mobility. When we analyzed amino acid content, we found that the proteins whose mobility was affected by pathogenic insulin signaling and H_2O_2 have cysteines with surface-exposed side chains, whereas this was not the case for the proteins whose mobility was not affected by those pathogenic factors (Figure 4A, Table S4). Surface cysteines create the potential for cross-linking through disulfide bonds, which might reduce the rate of diffusion by diverse mechanisms, including increasing effective protein mass, altering protein conformation, promoting binding to immobile proteins, altering interaction with transporters, and increasing cellular viscosity (Figure 4B)[63-67].

To explore how different oxidative states of the cellular environment might be expected to influence diffusion of proteins with and without cysteines, we developed a physics-based model (Figure 4C, S5 see methods). In this model, proteins are simulated as spherical particles, half of which have sticky patches on their surface, representing surface-exposed cysteine residues, and half of which do not have sticky patches. As the oxidative state of the cellular environment increases, the propensity of interaction between the patches increases, leading to protein crosslinking and formation of protein dimers and multimers (Figure S5). Proteins without surface-exposed cysteines remain in a monomeric state even at higher levels of ROS. As a result, the average diffusion coefficient of proteins containing surface-exposed cysteine decreased more than that of proteins lacking surface-exposed cysteines, because of dimer and multimer formation (Figure 4C, S5). The mobility of proteins lacking cysteines slightly decreased at higher levels of ROS, due to the increase in effective viscosity caused by the crosslinking of

the proteins containing cysteines present in the environment (Figure 4C). This model predicts that increased ROS-driven intermolecular disulfide bond formation will reduce protein mobility due to the increased frequency and lifetime of these bonds. As an initial test of this model, we investigated whether treatment of cells with H_2O_2 promotes crosslinking of IR proteins using western blotting (Figure S6A). These results suggest enhanced formation of intermolecular crosslinking through disulfide bond formation and are consistent with the predictions from the theoretical work.

To further test the model that surface cysteines contribute to reduced protein mobility in an oxidative environment, we asked whether treatment with a thiol-protective agents might preserve protein mobility in a high ROS environment. To prevent cysteine disulfide bond formation, cells were treated with N-ethylmaleimide (NEM), a compound that forms stable, covalent bonds with the thiol group in cysteines. FRAP revealed that treating cells with NEM partially preserved the mobility of IR, MED1, HP1α and FIB1, but had little effect on the mobility of SRSF2 in high ROS condition generated by H_2O_2 (Figure 4D).

If surface cysteines contribute to reduced protein mobility, we might expect that addition of surface cysteines to SRSF2, which normally lacks these residues, would cause reduced mobility of the modified SRSF2 protein in a high ROS environment. We engineered HepG2 cells to express endogenous SRSF2 fused to a rigid linker (to ensure surface exposure) containing multiple cysteine residues (SRSF2-Cys) or, as a control, the same number of serine residues (SRSF2-Ser) (Figure 4E,F). Treating HepG2 cells with H_2O_2 or pathogenic insulin concentrations did not affect the mobility of the SRSF2-Ser protein, but decreased the mobility of SRSF2-Cys protein (Figure 4E,F). Taken together, these results indicate that surface-exposed cysteines can affect protein mobility when cells are exposed to oxidative stress and pathogenic signaling.

Next, we asked whether there are reports of any of the proteins studied here having missense mutations resulting in gaining a cysteine and, if so, whether these might affect protein mobility. A tyrosine to cysteine mutation (Y1361C) was reported in the insulin receptor. This mutation occurs outside of the structured domain and does not appear to decrease protein stability[68]. Modeling indicates that the cysteine gained through this mutation is surface exposed (Figure S6B). We introduced this mutation into the insulin receptor-GFP fusion protein (IR Y1361C-GFP) in both alleles in HepG2 cells (Figure 4G). By performing FRAP, we found that the gain-of-cysteine mutation caused a reduction in IR protein mobility in HepG2 cells under normal redox conditions (Figure 4H) and that treating cells expressing IR Y1361C-GFP with NAC enhanced IR Y1361C protein mobility (Figure 4I). Mutating the same amino acid to serine had little to no effect on insulin receptor protein mobility (Figure S6C). These results indicate that mutations that add surface cysteines sensitize the insulin receptor to physiological levels of ROS, reducing its mobility under normal redox conditions, and that addition of an antioxidant can enhance this receptor's mobility. It is possible that the Y1361C mutation confers this special sensitivity to normal redox conditions because it occurs in region known to interact with other proteins containing surface-exposed cysteines[69]. Gain-of-cysteine mutations are among the most pathogenic missense mutations (Figure S6D) and their

effect on protein mobility may not be limited to IR, but may extend to other disease-relevant proteins.

Diverse pathogenic factors decrease protein mobility

The pathogenic stimuli that are associated with diverse diseases are thought to commonly induce oxidative stress[70]. We developed a mobility biosensor assay to investigate relationships between surface-exposed cysteines and protein mobility under oxidative conditions, and to investigate whether diverse pathogenic stimuli produce similar mobility phenotypes in liver cells and in other disease-relevant cell types. We constructed the protein mobility sensor by adding a rigid linker containing 5 cysteine residues to the HaloTag protein (HaloTag-Cys) (Figure 5A) together with a control biosensor containing 5 serine residues (HaloTag-Ser). The HaloTag-Cys biosensor was appropriately sensitive to pathogenic levels of H_2O_2, as evidenced by the mobility of HaloTag-Cys decreasing upon H_2O_2 treatment in a dose-dependent fashion (Figure 5B). Similarly, treatment of cells containing the biosensor with pathogenic insulin concentrations led to reduced protein mobility (Figure 5C). Pathogenic levels of insulin had less of an effect on the mobility of a control HaloTag-Ser protein (Figure S6E).

Pathogenic stimuli that induce oxidative stress include hyperglycemia, high fat, inflammation, genotoxic stress, endotoxin and drug toxicity[71-76] (Figure 5D). These stimuli have been shown to increase ROS through diverse mechanisms, which include, but are not limited to, dysregulation of mitochondria, dysregulation of redox homeostasis proteins, ER stress and eNOS dysregulation[75,77-83]. Treating cells with these pathogenic stimuli led to elevated levels of ROS (Figure 5E), confirming previous results[28,71-75]. These treatments also reduced the mobility of the HaloTag-Cys protein (Figure 5F, S7A,B). In skeletal muscle cells, another disease-relevant cell type, pathogenic stimuli also decreased HaloTag-Cys mobility (Figure S7C). Taken together, these results are consistent with a model in which diverse pathogenic stimuli known to induce oxidative stress cause suppressed protein mobility in multiple disease-relevant cell types.

Protein mobility and functional activity

Biochemical reactions are typically collision-limited[31,32] and reduced rates of protein diffusion would be expected to reduce functional outputs (Figure 6A). We produced a mathematical model and conducted tests in vitro and in cells designed to confirm that reduced protein mobility confers reduced enzymatic activity with the insulin receptor (Figure 6B). Phosphorylation of substrates by protein kinases such as the insulin receptor (Figure 6C), which would be expected to be collision-limited, should be reduced when protein mobility is decreased. Mathematical modeling of phosphorylation of substrates by protein kinases showed reaction outputs are reduced when protein mobility is decreased (Figure 6D). IR and an IR substrate protein, IRS1, were purified and subjected to environments that would slow or accelerate the mobility of proteins *in vitro*. When the mobility of proteins was reduced *in vitro* by increasing glycerol

concentration, and thus viscosity, we observed reduced phosphorylation of IRS1 by IR (Figure 6E). Agitation of solutions can increase protein mobility and thus the collision rate of molecules[84], and agitation was found to partially rescue the reduction in phosphorylation with elevated viscosity (Figure 6F). These results support the expectation that reduced protein mobility reduces the kinase activity of IR.

To further probe the relationship between protein mobility and functional output in cells, we used the BirA/AviTag system, which was previously shown to exhibit collision-limited activity[31]. In this system, where the biotin ligase BirA biotinylates its substrate AviTag, fusion of BirA with SNAP-tag (BirA-SNAP) and fusion of our protein mobility biosensor HaloTag-Cys to the AviTag allowed us to monitor both protein mobility and BirA activity (Figure 6G,H) in HepG2 cells. Under conditions of pathogenic signaling in cells, the reduction in protein mobility correlated with reduced biotinylation (Figure 6I-K). These results support the concept that reduced protein mobility leads to reduced functional activity.

The cellular processes that have been reported to be dysregulated in chronic syndromes include reduced phosphorylation of substrates, altered gene regulation, repression of heterochromatic repeats, among others[8-18]. To confirm that these processes are indeed dysregulated in cells under conditions studied here, we conducted assays in cells that were treated with normal and with pathogenic insulin. The results showed evidence of dysregulated features noted previously in chronic syndromes (Figure 6L). Phosphorylation of IRS1 was reduced, genes occupied by the Mediator coactivator subunit MED1 were expressed at lower levels, and there was elevated expression of heterochromatic repeats. These results are consistent with a model where reduced protein mobility can contribute to the diversity of dysregulated processes that are evident in chronic disease.

DISCUSSION

Pathogenic signaling contributes to prevalent diseases characterized by dysregulation of remarkably diverse cellular processes[8-17]. Consequently, equally diverse pathogenic mechanisms are assumed to cause these phenotypes. However, the findings on protein mobility in healthy and dysregulated cells described here suggest an alternative explanation; namely, that a common mechanism - suppressed mobility, here referred to as proteolethargy - contributes to dysregulation of a range of cellular processes in the setting of diverse pathogenic stimuli.

Proteolethargy, the phenomenon of reduced protein mobility in the setting of pathogenic stimuli, might be caused by any number of mechanisms, but several lines of evidence converge on the effects of excess ROS on protein mobility as a common mechanism that can impact proteins throughout the cell in diverse chronic syndromes (Figure 7). Cells exposed to diverse pathogenic stimuli produce excess ROS through mechanisms that include dysregulation of mitochondria, dysregulation of redox homeostasis proteins, ER stress, eNOS dysregulation among others[75,77-83] (Figure 7A). Proteins exposed to

oxidative environments exhibit reduced mobility if they have surface-exposed cysteines or are engineered to have surface cysteines. These effects can be remedied in part by treatment of cells with reducing agents or agents that are thiol-protective. Gain-of-cysteine mutations can cause reduced mobility of the mutated protein. We estimate that ~50% of human proteins contain at least one surface-exposed cysteine (see Methods), so there is potential for half of the proteome to be directly susceptible to proteolethargy in high ROS environments.

Our experimental and modeling data jointly support a model in which proteins with surface-exposed cysteines, upon transiting through a milieu that is densely packed with biomolecules, have the potential to form transient disulfide crosslinks with other proteins (Figure 7B). An elevated oxidative environment has the potential to increase the lifetime of the inter-molecular crosslinks[63,85,86], effectively increasing the hydrodynamic radius and thereby decreasing protein mobility[64]. Variations of this model are possible, where changes in disulfide bond rates or lifetimes have additional influences on protein mobility through improper complex formation, changing protein conformation, promoting binding to immobile proteins, disrupting associations with transport proteins or altering cytoplasmic viscosity (Figure 4B)[63-67]. It is also likely that the effects of elevated oxidative environments can impact protein mobility more indirectly; for example, changes in plasma membrane fluidity due to altered lipid oxidation and composition have the potential to influence protein mobility[87-93] and changes that affect cytoskeleton-associated proteins have been noted to impact cellular fluidity (Figure 7C)[94,95].

Proteolethargy would be expected to adversely impact diverse functions in cells. In healthy cells, proteins with prominent roles in diverse cellular processes are highly mobile, and thus able to transit a space equivalent to the diameter of a cell in 2-10 seconds. In cells subjected to pathogenic signaling, however, the mobility of most proteins studied here was reduced by 20-35%. Since many biological processes in cells are collision-limited, decreases in protein mobility are expected to reduce functional outputs (Figure 7D)[31,32,96]. Supporting this view, we found that reducing the mobility of IR reduces its rate of phosphorylation of the IR substrate IRS1 in vitro and in vivo, and that a synthetic system designed to report biotin ligase activity in cells showed reduced ligase activity when cells were subjected to pathogenic signaling. The cellular processes that have been reported to be dysregulated in chronic syndromes such as diabetes and inflammatory disorders are diverse and include signaling activity, gene regulation, heterochromatin repression and metabolic activity[8-17]. These cellular functions were found to be dysregulated in the cell system studied here. We thus suggest that proteolethargy may account for the diversity of dysregulated cellular functions noted for at least some chronic diseases.

Many proteins have been shown to assemble together with functionally related proteins into biomolecular condensates, cellular organelles that are not physically delimited by membranes[25,28,44,49-55,97,98]. We found that the mobility of proteins was reduced both inside and outside of these compartments. The mobility of the synthetic proteins such as the protein mobility biosensor, which does not appear to assemble into condensates, was reduced by high ROS environments. Our results thus indicate that proteolethargy

can occur across the cell, and occurs both inside and outside of condensate compartments.

There is limited information on the mobility of a range of proteins with diverse functions in human cells, and even less information on the effects of pathogenic stimuli on protein mobility. This paucity of knowledge may explain why proteolethargy has previously not been considered as a potential pathogenic mechanism for chronic diseases. Previous studies have investigated the diffusive behaviors of certain transcriptional regulators in mammalian cells[24,35,99,100], and one noted that the mobility of the insulin receptor is reduced in rat hippocampal neurons by low concentrations of tumor necrosis factor α and by cholesterol depletion[34].

The model described here for proteolethargy in disease has implications for the development of novel therapeutics for certain chronic diseases. Restoring protein mobility might be considered among the therapeutic hypotheses for these chronic diseases. Protein mobility biosensors, such as the one developed for this study, may prove to be valuable for high throughput screening for drugs that restore normal protein mobility under pathogenic signaling conditions. Redox homeostasis is regulated by many pathways and proteins that counteract transient increases in ROS that occur normally in diverse cellular processes[70,101,102], so it is possible that therapeutic targeting of these natural pathways will prove beneficial for treating or preventing proteolethargy. The rescue of protein mobility with NAC treatment, as described here, is a proof of principle for this concept.

LIMITATIONS OF THE STUDY

We propose that pathogenic signaling reduces the mobility of a large fraction of cellular proteins, that reduced protein mobility is due largely to a dysregulated redox environment that impacts oxidation-sensitive cysteines, and that this proteolethargy may account for the diversity of dysregulated cellular processes that are evident in chronic disease. Pathogenic signaling could potentially affect ~50% of the proteome based on estimates of surface-exposed cysteines in proteins, but a necessarily limited number of proteins are surveyed in this work. There are additional oxidation-sensitive amino acids and oxidation-related mechanisms that may contribute to decreased protein mobility that are not studied here. As this work focused on movement of individual molecules at specific time- and distance-scales, we did not explicitly examine all potential mechanisms that may affect protein movement. Proteolethargy in the setting of pathogenic stimuli was observed using cell lines and a defined set of experimentally tractable treatments, so further studies will be needed to learn how reduced protein mobility manifests as disease phenotypes in whole organisms. Diverse endogenous proteins were studied in cells subjected to pathogenic insulin signaling, but these proteins were not studied in response to diverse pathogenic stimuli, where a biosensor was deployed to report protein mobility. Excess ROS has been implicated in aging and diseases not studied here, so further studies are necessary to learn whether reduced

protein mobility is associated with, and perhaps contributes to, aging and other diseases.

ACKNOWLEDGMENTS

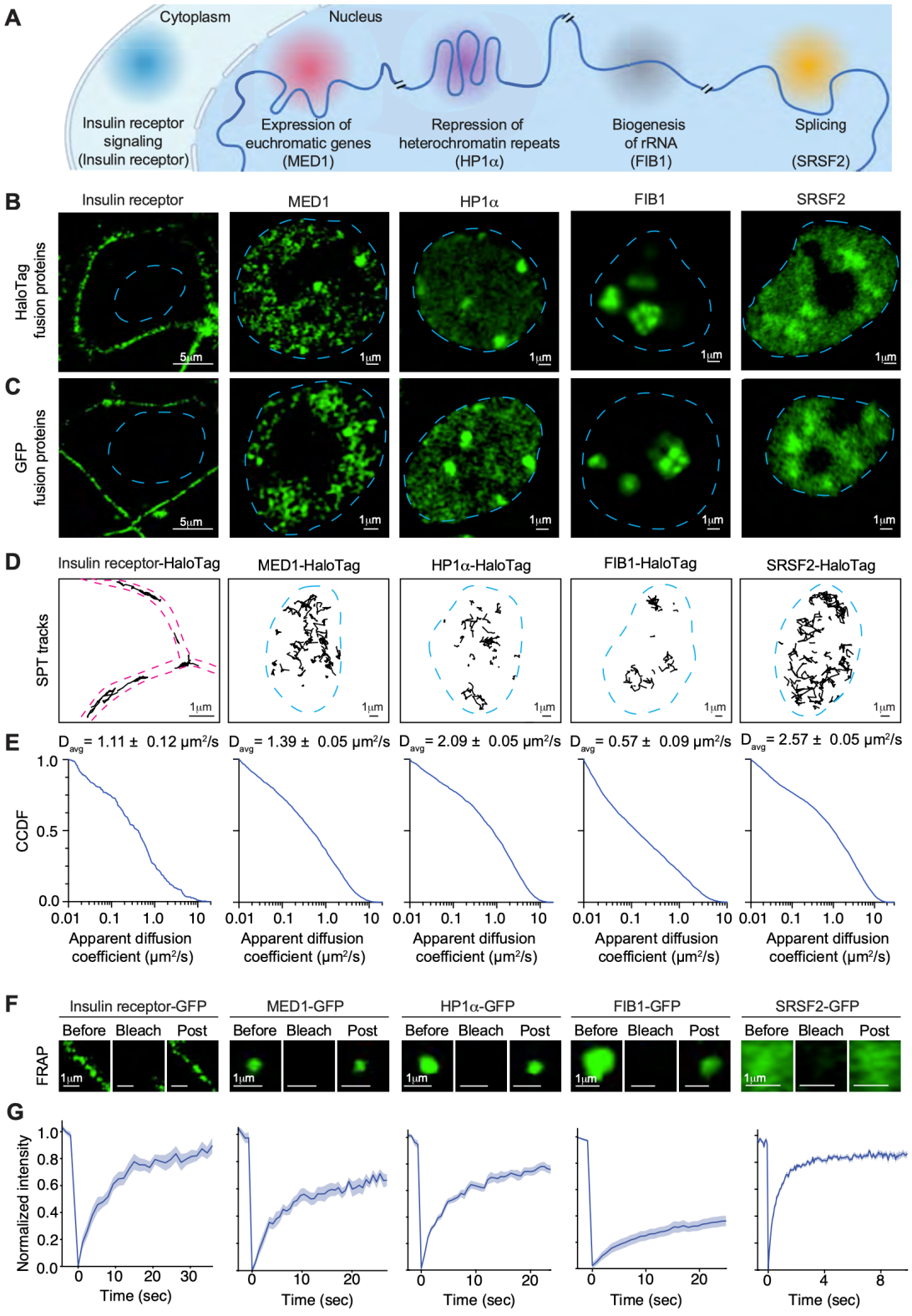
We thank Phillip Sharp, Eric Olson, Christine Ivashchenko, Max Friesen, and Mehran Kardar for enlightening discussions, Christina Lilliehook for feedback on the manuscript, Asier Vidal and Tenzin Kunchok of the W.M Keck Microscopy Facility and the Whitehead Institute Metabolomics Core for advice, members of the Harvard Center for Biological Imaging and the Whitehead Genome Technology Core for their assistance. Funding: National Institutes of Health grant GM144283 (RAY), National Institutes of Health grant CA155258 (RAY), National Science Foundation grant PHY2044895 (AC and RAY), St. Jude Children's Research Hospital Transcription Collaborative (RAY), National Institutes of Health grant KL2TR002542 (JMP).

DECLARATION OF INTERESTS

The Whitehead Institute has filed a patent application based on this paper. R.A.Y. is a founder or equity holder of Syros Pharmaceuticals, Camp4 Therapeutics, Omega Therapeutics, Dewpoint Therapeutics, Paratus Sciences and Precede Biosciences. A.D. is a consultant for Dewpoint Therapeutics. The other authors have no competing interests. AKC is a consultant (titled "Academic Partner") of Flagship Pioneering, a consultant and member of the Strategic Oversight Board of Apriori Bio (a Flagship company), and a consultant and SAB member of Metaphore Bio (a Flagship company).

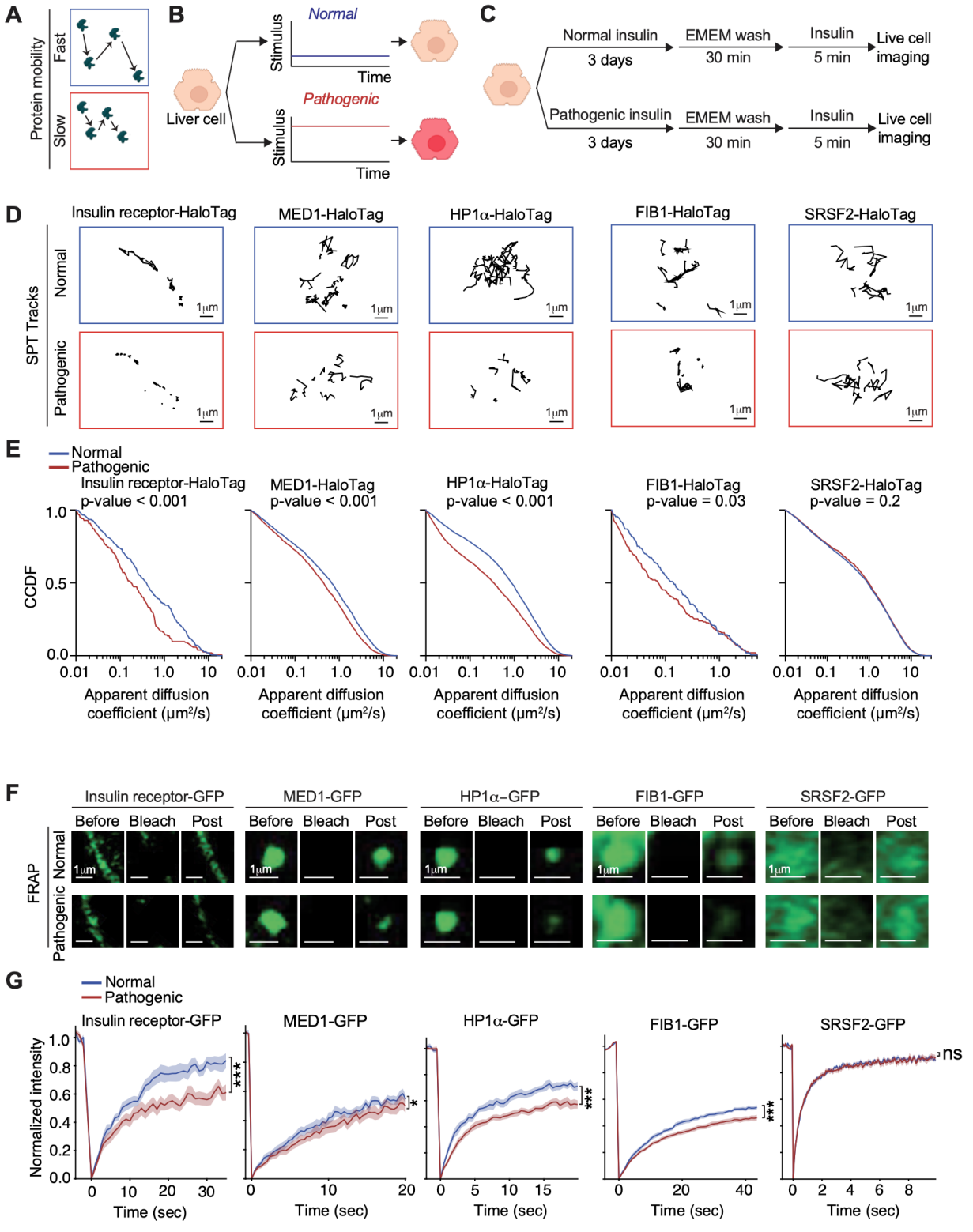
Figures and Tables

Figure 1. Mobility of diverse proteins in cells. See also Figures S1-S2, and Table S1.



(A) Cellular compartments, biological processes and proteins examined in this study.
(B, C) Live-cell imaging of HepG2 cells expressing HaloTag (B) or green fluorescent protein (GFP) -tagged (C) versions of the indicated proteins. Dashed lines show outline of nucleus. Scale bars are indicated.
(D) Representative tracks for movement of individual molecules as determined by single particle tracking (SPT) of HaloTag versions of the indicated proteins. Dashed magenta lines represent outline of the plasma membrane. Dashed blue lines represent outline of the nucleus.
(E) Complementary cumulative distribution function (CCDF) graphs of apparent diffusion coefficients as determined by SPT of the indicated proteins (n = 294, 1751, 2591, 2855, 5458 molecules for insulin receptor (IR), MED1, HP1α, FIB1, and SRSF2, respectively).
(F) Representative images of FRAP of HepG2 cells expressing GFP-tagged versions of the indicated proteins. Images before (Before), immediately following (Bleach), and after recovery (Post) are shown. Scale bars are indicated.
(G) Quantification of FRAP experiments of the indicated proteins (n=10, 11, 15, 15, 15 cells for IR, MED1, HP1α, FIB1, SRSF2, respectively). Data shown as mean (blue line) ± standard error of the mean (SEM) (light blue).

Figure 2. Protein mobility decreases in a model of pathogenic signaling. See also Figures S3 and S4 and Table S2 and S3.



(A, B) Model for protein mobility in pathogenic signaling: individual molecules move at fast or slow speeds (A), depending on exposure to normal or pathogenic signaling (B). **(C)** Schematic representation of cell treatments.

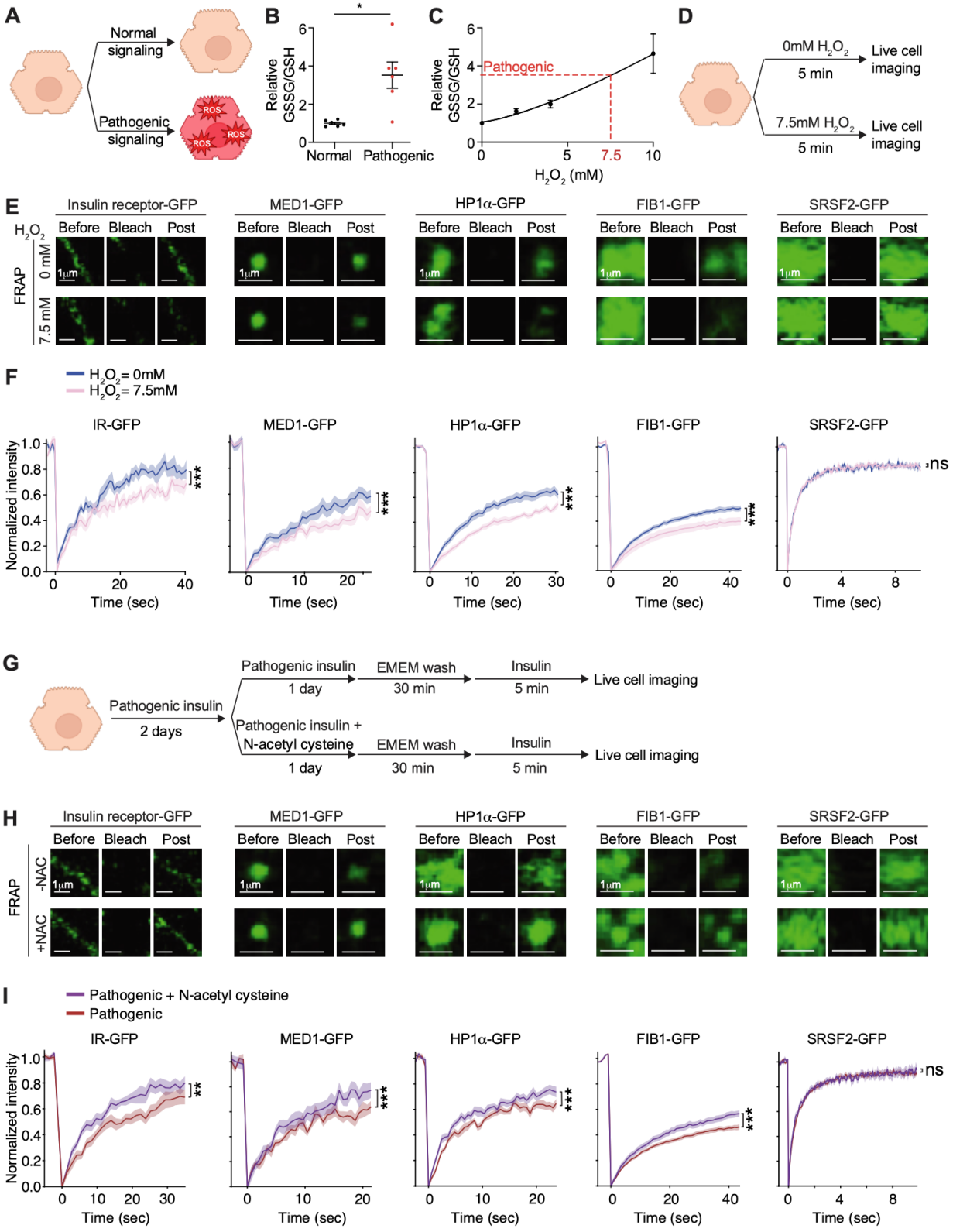
(D) Representative individual protein tracks as determined by SPT for the indicated proteins and experimental treatments. Scale bars are indicated.

(E) CCDF graphs of apparent diffusion coefficients as determined by SPT for the indicated proteins and experimental treatments (Normal, n = 357, 5719, 5199, 153, 3399 molecules for IR, MED1, HP1α, FIB1, and SRSF2, respectively; Pathogenic, n = 154, 2227, 3529, 146, 2872 molecules for IR MED1, HP1α, FIB1, and SRSF2, respectively). Mann-Whitney test was used for statistical analysis.

(F) Representative FRAP images for the indicated proteins and experimental treatments. Images before (Before), immediately following (Bleach), and after recovery (Post) are shown. Scale bars are indicated.

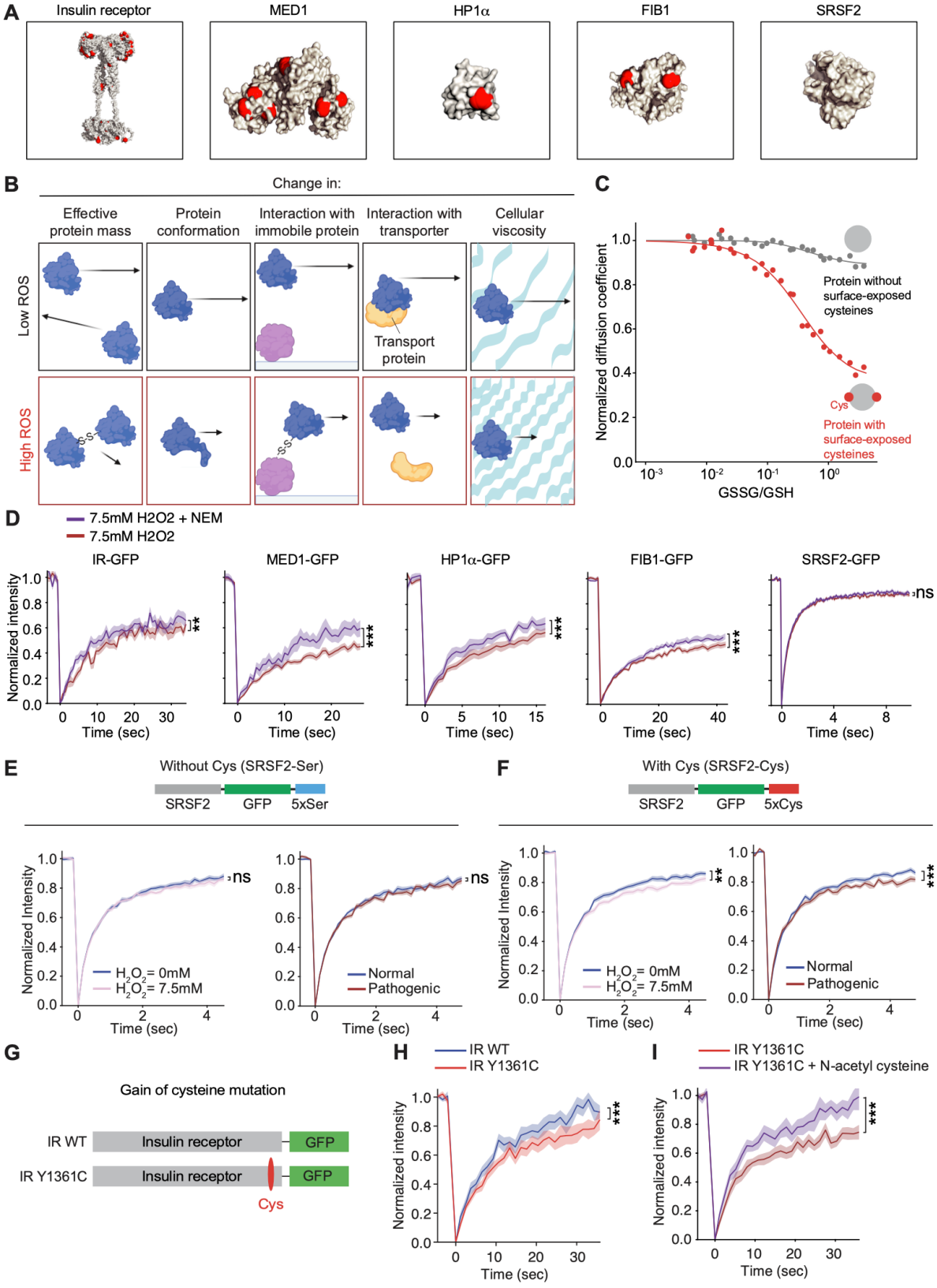
(G) Quantification of FRAP experiments for the indicated proteins and experimental conditions (Normal and Pathogenic, n = 16, 10, 14, 10, 20 cells each condition for IR, MED1, HP1α, FIB1, SRSF2, respectively). Data shown as mean (Normal, blue line; Pathogenic, red line) ± SEM (Normal, light blue; Pathogenic, light red). t-test was used for statistical analysis. Cohen's d = 0.9, 0.4, 1.2, 0.9, and 0.0 for IR, MED1, HP1α, FIB1, SRSF2, respectively.

Figure 3. Oxidative environment affects protein mobility. See also Figures S4 and Table S3.



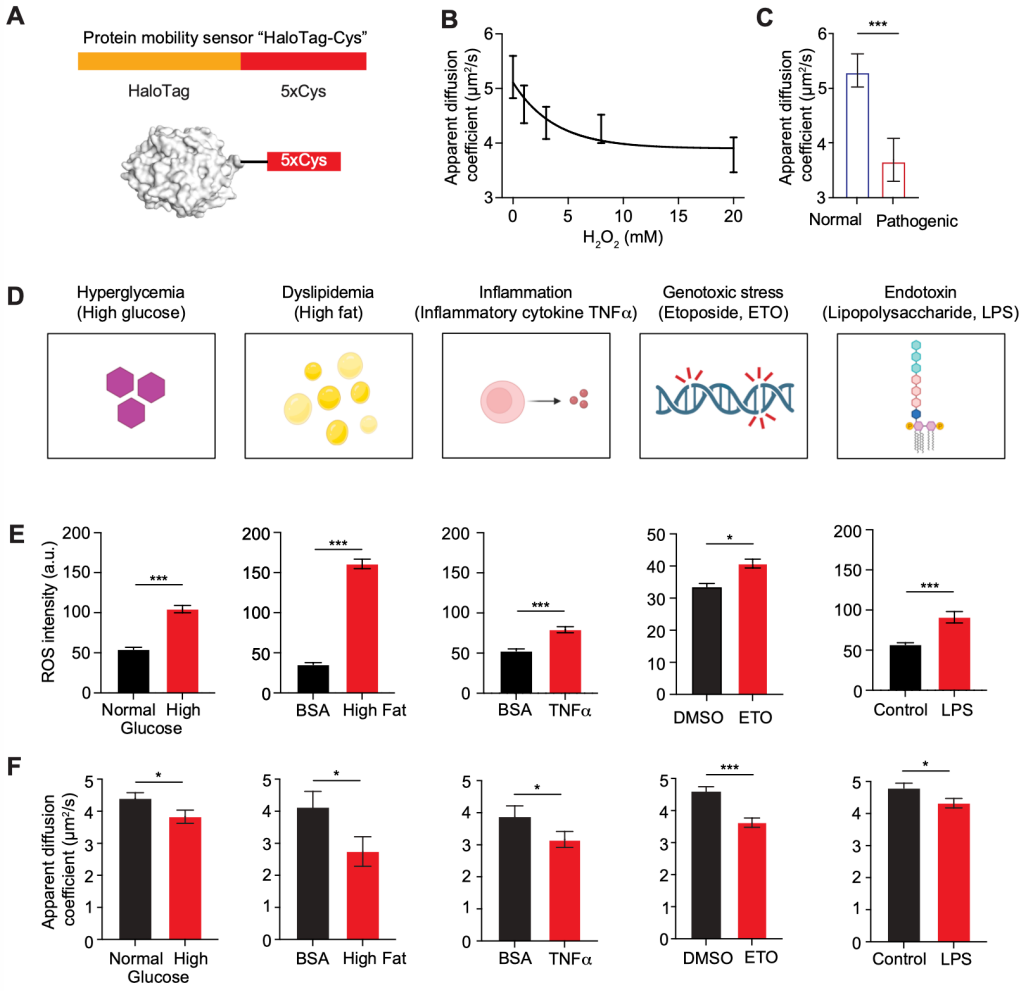
(A) Increased reactive oxygen species (ROS) in pathogenic signaling.
(B) Relative ratio of oxidized to reduced glutathione (GSSG/GSH) in cells treated as indicated. Data shown as mean ± SEM. T-test was used for statistical analysis.
(C) Relative GSSG/GSH ratio in cells treated with different hydrogen peroxide (H_2O_2) concentrations. Data shown as mean ± SEM. H_2O_2 concentration expected to phenocopy pathogenic signaling is indicated.
(D, G) Schematic representation of cell treatments.
(E, H) Representative FRAP images for the indicated proteins and experimental treatments. Images before (Before), immediately following (Bleach), and after recovery (Post) are shown.
(F, I) Quantification of FRAP experiments for the indicated proteins and experimental conditions. For (F), 0mM and 7.5mM, n = 10 cells each condition for each protein. Data shown as mean (0mM, blue line; 7.5mM, red line) ± SEM (0mM, light blue; 7.5mM, light red). For (I), (Pathogenic, n = 16, 10, 15, 10, 20 for IR, MED1, HP1α, FIB1, and SRSF2, respectively; Pathogenic + NAC, n = 16, 10, 15, 20, 20 for IR, MED1, HP1α, FIB1, and SRSF2. Data shown as mean (Pathogenic, red line; Pathogenic + NAC, purple line) ± SEM (Pathogenic, light red; Pathogenic + NAC, light purple). t-test was used for statistical analysis (F,I). For (F), Cohen's d = 0.7, 0.7, 1.2, 1.0, and 0.0 for IR, MED1, HP1α, FIB1, SRSF2, respectively. For (I), Cohen's d = 0.5, 0.8, 0.9, 0.6, and 0.2 for IR, MED1, HP1α, FIB1, SRSF2, respectively.

Figure 4. Surface-exposed cysteines sensitize proteins to oxidation-driven decrease in protein mobility. See also Figures S5 and S6 and Table S4.



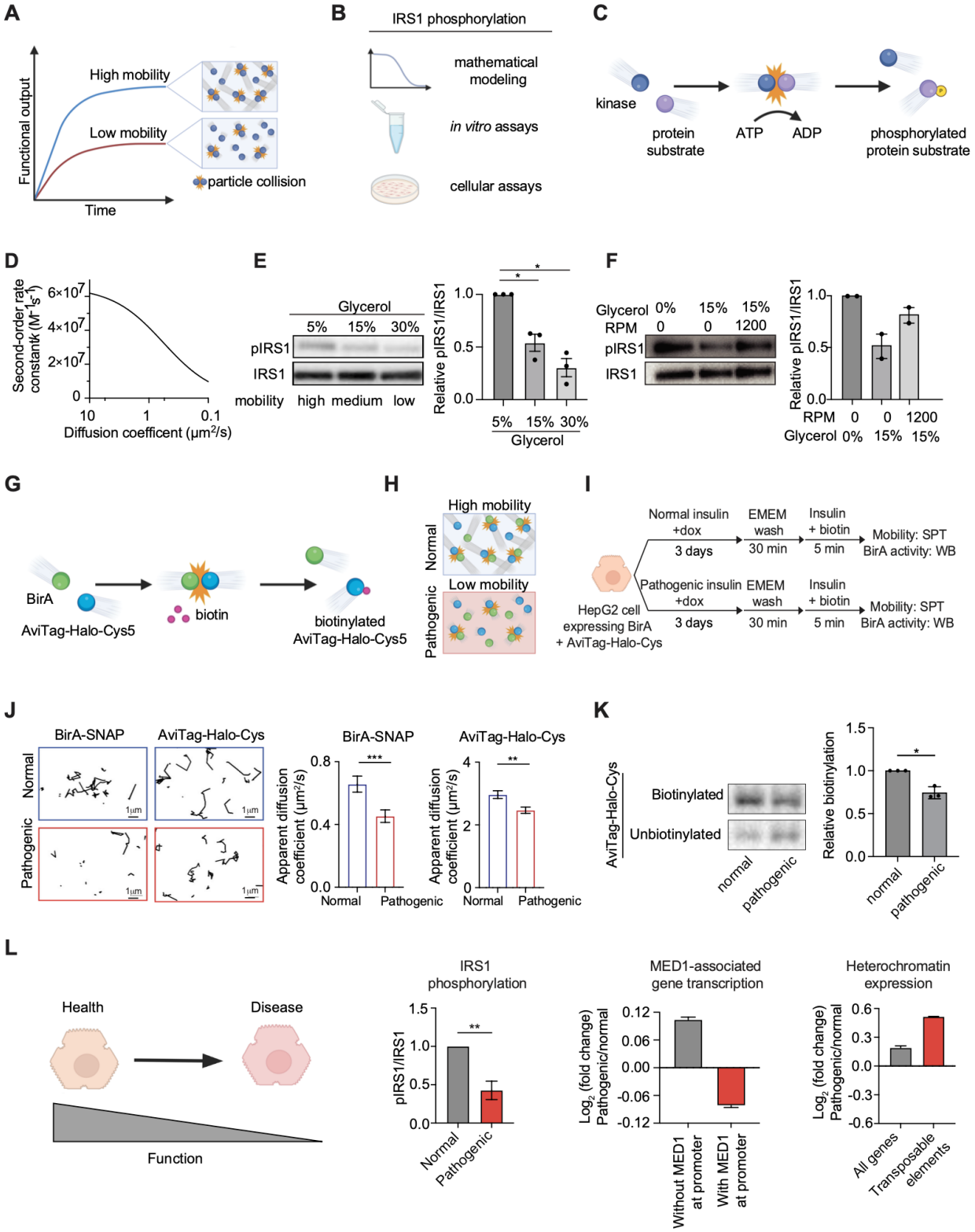
(A) Rendering of the crystal structure of indicated proteins showing cysteines in red.
(B) Diverse models for decreased protein mobility, including change in effective protein mass, protein conformation, interaction with immobile protein, interaction with a protein that facilitates transport, cellular viscosity increasing resistance to movement.
(C) Predicted normalized diffusion coefficient from simulations of a mixture of proteins with (red) and without (gray) surface-exposed cysteines as a function of the ratio of oxidized (GSSG) to reduced (GSH) glutathione. The diffusion coefficient was normalized to the mean of all simulated data points for GSSG/GSH<10^{-3} (see Methods).
(D) Quantification of FRAP data for insulin receptor (7.5 mM H_2O_2 n=16 cells, 7.5mM H_2O_2 + NEM n=16 cells), MED1 (7.5 mM H_2O_2 n=29 cells, 7.5mM H_2O_2 + NEM n=15 cells), HP1α (7.5 mM H_2O_2 n=14 cells, 7.5mM H_2O_2 + NEM n=13 cells), FIB1 (7.5 mM H_2O_2 n=24 cells, 7.5mM H_2O_2 + NEM n=24 cells) and SRSF2 (7.5 mM H_2O_2 n=12 cells, 7.5mM H_2O_2 + NEM n=12 cells) in HepG2 cells treated with 0mM or 7.5mM of H_2O_2 after pre-treatment with 10 μM N-ethyl maleimide. Data are plotted as means ± SEM.
(E, F) Top: representation of SRSF2 fusion proteins with an added serine or cysteine-containing rigid linker. Bottom: quantification of FRAP data for SRSF2 fusion proteins in cells treated with the indicated experimental conditions (SRSF2-Ser, 0mM H_2O_2, n=13 cells, 7.5mM H_2O_2, n=12, Normal, n=10 cells, Pathogenic, n=10 cells; SRSF2-Cys, 0mM H_2O_2, n=13 cells, 7.5mM H_2O_2, n=13, Normal, n=10 cells, Pathogenic, n=10 cells). Data are plotted as mean ± SEM.
(G) Representation of wildtype and mutant IR fusion proteins.
(H) Quantification of FRAP data for wildtype (IR WT, n=15 cells) or Y1361C mutant IR (IR Y1361C, n=15 cells). Data are plotted as mean ± SEM.
(I) Quantification of FRAP data for Y1361C mutant IR in cells treated with (n=15 cells) or without (n=15 cells) N-acetyl cysteine. Data are plotted as mean ± SEM. T-test was used for statistical analysis (D-I).

Figure 5. Diverse pathogenic factors decrease protein mobility. See also Figure S6.



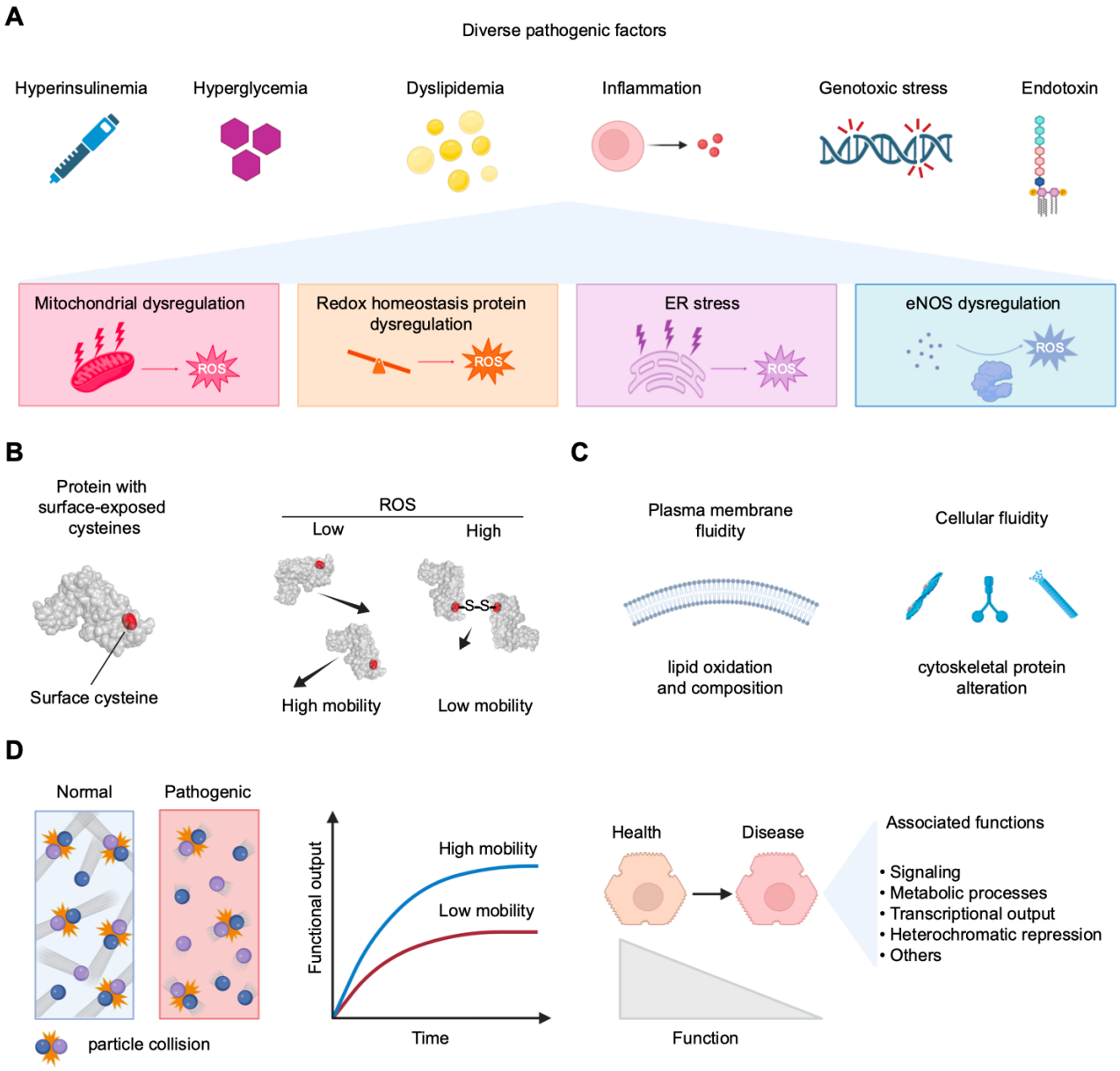
(A) Representations of HaloTag fusion protein (HaloTag-Cys).
(B) Apparent diffusion coefficient of HaloTag-Cys as determined by SPT in cells treated as indicated (n = 245, 316, 428, 560, 305 molecules for 0, 1, 3, 8 or 20mM H_2O_2, respectively).
(C) Apparent diffusion coefficient of HaloTag-Cys as determined by SPT in cells treated as indicated (n = 446, 173 molecules for normal and pathogenic, respectively).
(D) Cartoon depicting pathogenic stimuli.
(E) ROS quantification in cells treated as indicated. Data are plotted as mean ± SEM. Numbers of cells: normal glucose (77) vs. high glucose (67); BSA (115) vs. high fat (171); BSA (150) vs. TNFa (91); DMSO (152) vs. Etoposide (ETO, 83); control (82) vs. lipopolysaccharide (LPS, 78).
(F) Apparent diffusion coefficient of HaloTag-Cys as determined by SPT in cells treated as indicated. Numbers of molecules: normal glucose (1001) vs. high glucose (582); BSA (126) vs. high fat (101); BSA (265) vs. TNFa (363); DMSO (1718) vs. ETO (1804); control (1456) vs. LPS (1327). Cohen's d = 0.1, 0.2, 0.1, 0.2, and 0.1 from left to the right, respectively. Data are plotted as means ± SEM. Mann-Whitney test was used for statistical analysis (C,F). t-test was used for statistical analysis (E).

Figure 6. Protein mobility affects function. See also Figure S7.



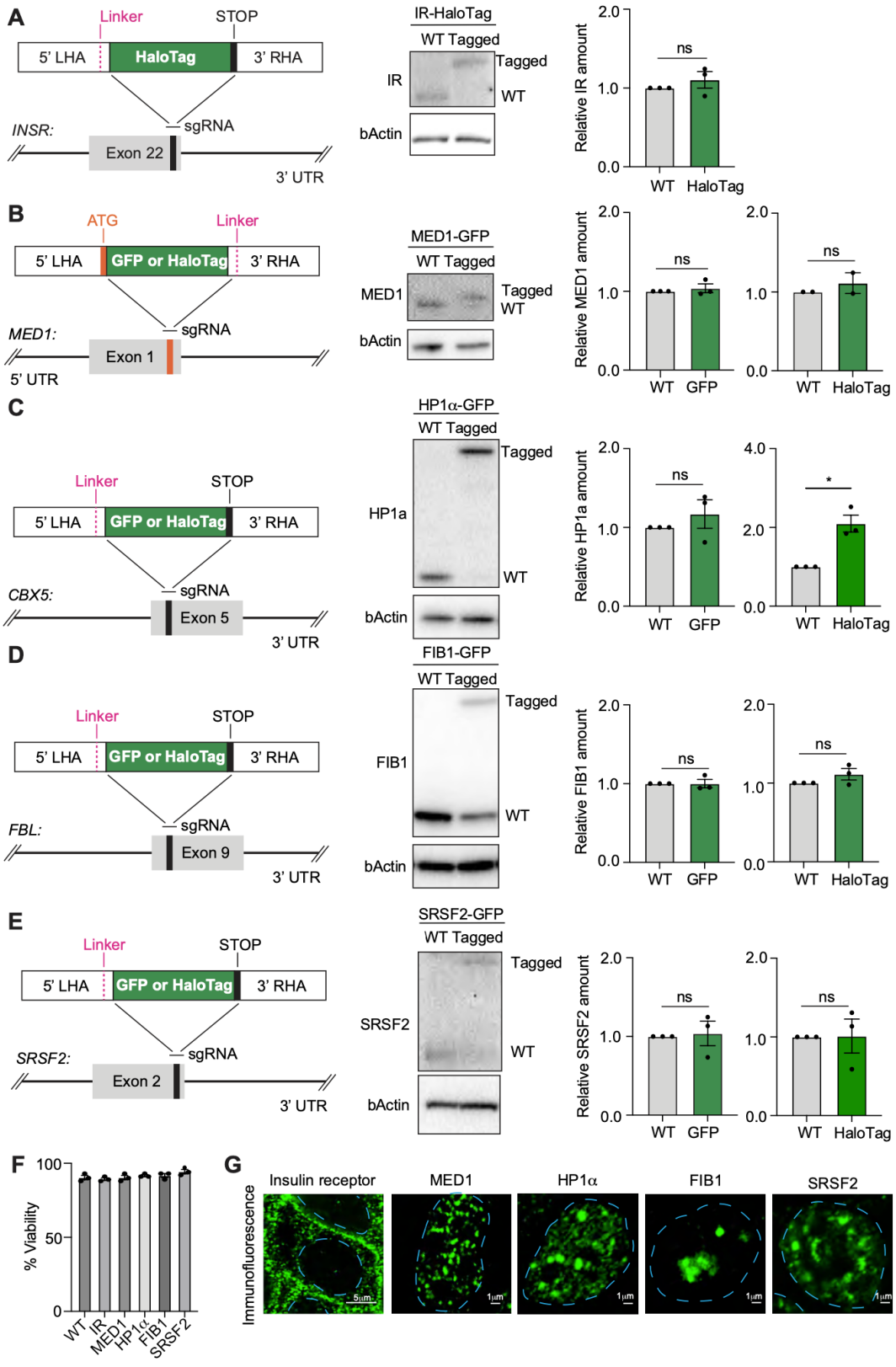
(A-C) Cartoons depicting relationship between protein mobility, functional output and collision frequency (A), models and assays used to study IRS phosphorylation **(B)** and the phosphorylation of IRS1 by a kinase.
(D) Second-order rate constant from simulations of IRS1 phosphorylation as a function of diffusion coefficient.
(E) Immunoblot for phosphorylated IRS1 (pIRS1) and IRS1 (left). IRS1 phosphorylation assay was performed in solutions containing 5%, 15% or 30% glycerol. Quantification of relative pIRS1 amount (right) (n = 3 biological replicates). T-test was used for statistical analysis.
(F) Immunoblot for phosphorylated IRS1 (pIRS1) and IRS1 (left). IRS1 phosphorylation assay was performed in solutions containing 0% or 15% glycerol with agitation (1200 RPM) or without agitation (0 RPM). Quantification of relative pIRS1 amount (right) (n = 2 biological replicates).
(G) Cartoon depicting biotinylation assay.
(H) Cartoon depicting high mobility in normal conditions and low mobility in pathogenic conditions.
(I) Schematic representation of cell treatments.
(J) Representative tracks for movement of individual molecules as determined by single particle tracking (SPT) of the indicated proteins (left). Apparent diffusion coefficient of the indicated proteins in cells treated with normal or pathogenic insulin (right). Numbers of molecules: BirA-SNAP normal (1003) vs. pathogenic (865); AviTag-Halo-Cys normal (1022) vs pathogenic (1067). Mann-Whitney test was used for statistical analysis.
(K) Immunoblot for biotinylated and unbiotinylated AviTag-Halo-Cys.
(L) Cartoon depicting function decreases in diseased cells (left). Quantification of relative pIRS1 determined by immunoblotting (t-test was used for statistical analysis), Log_2(fold change) of gene expression for genes whose promoter is occupied or not occupied by MED1, and Log_2(fold change) of expression of protein-coding genes or repetitive elements.

Figure 7. Proteolethargy is a pathogenic mechanism in chronic disease.



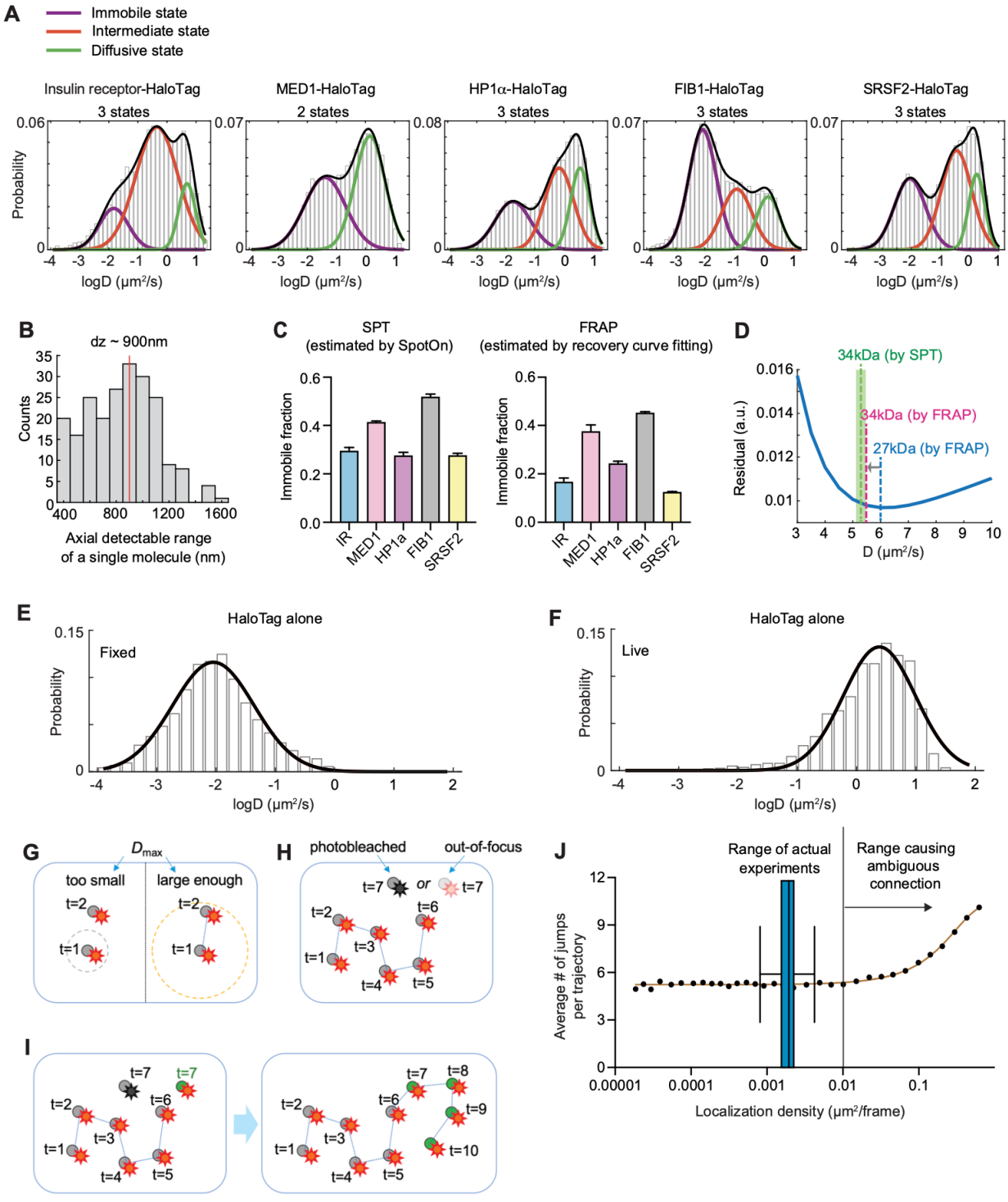
(A) Diverse pathogenic factors lead to oxidative stress via multiple cellular pathways and mechanisms.
(B) Proteins with surface-exposed cysteines suffer reduced mobility in high ROS environments due to their sensitivity to oxidation.
(C) Alterations in plasma membrane and cytoplasmic fluidity can also occur in high ROS environments.
(D) Mobility is decreased in pathogenic signaling, thereby reducing rates of particle collision and leading to reduced functional output for diverse cellular processes.

Figure S1. Cell line validation. Related to Figure 1.



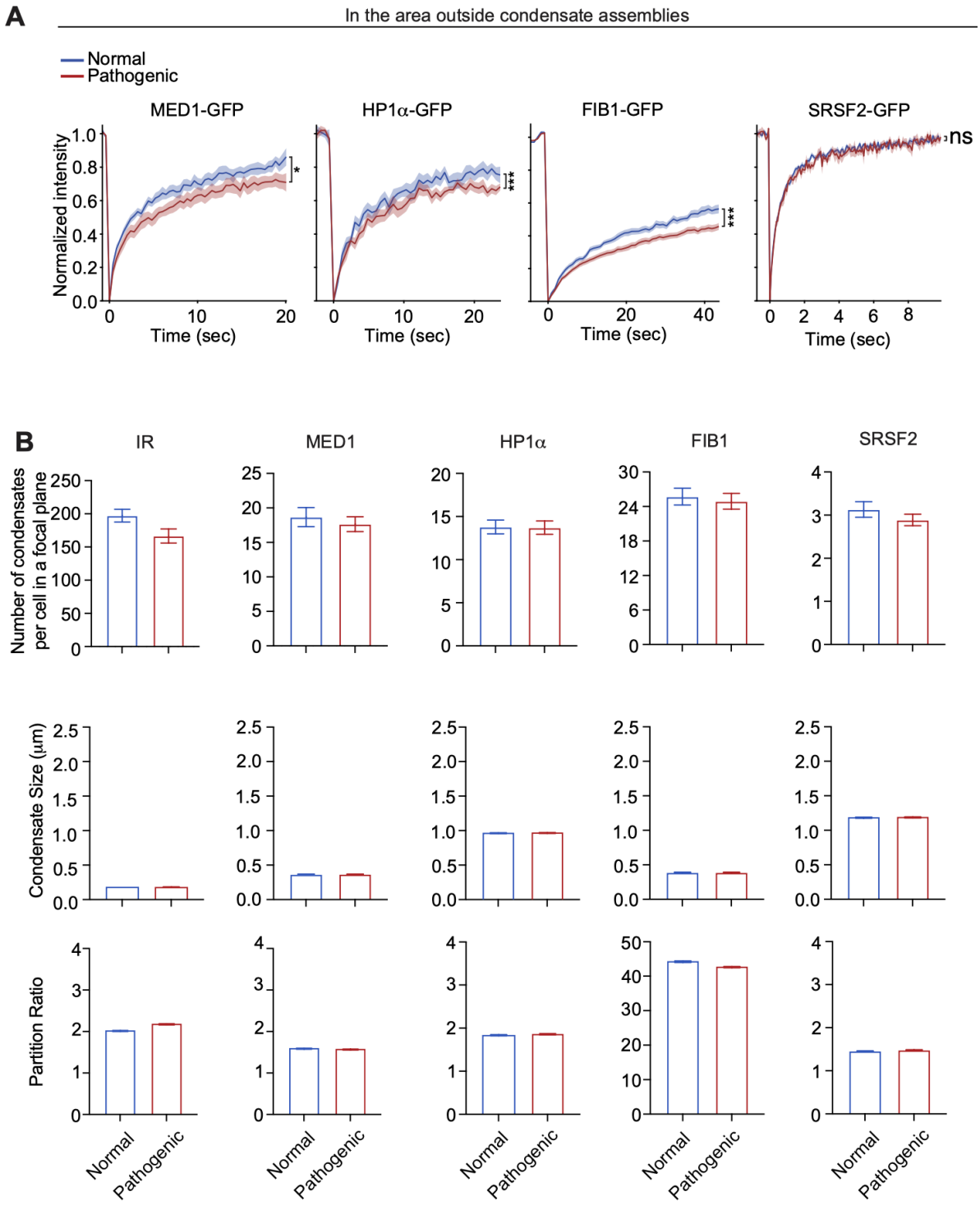
(A) Strategy to endogenously tag IR with HaloTag (left). Immunoblot for IR and beta actin (bActin, middle). Quantification of relative IR amount as compared to beta actin (right). Data are plotted as mean ± SEM (n= 3 biological replicates in each condition).
(B) Strategy to endogenously tag MED1 with GFP or HaloTag (left). Immunoblot for MED1 and beta actin (bActin, middle). Quantification of relative MED1 amounts as compared to beta actin (right). Data are plotted as mean ± SEM (n= 2 to 3 biological replicates in each condition).
(C) Same as (B), but for HP1α (n= 3 biological replicates in each condition).
(D) Same as (B), but for FIB1.
(E) Same as (B), but for SRSF2.
(F) Viability of WT cells or cells expressing endogenous IR, MED1, HP1α, FIB1 and SRSF2 tagged with HaloTag. Data are plotted as mean ± SEM (n= 3 biological replicates in each condition).
(G) Immunofluorescence images of IR, MED1, HP1α, FIB1 and SRSF2 (green) in WT HepG2 cells. Dashed blue lines represent nuclear outline. Scale bars are indicated.

Figure S2. Comparability between FRAP and SPT and SPT controls. Related to Figure 1.



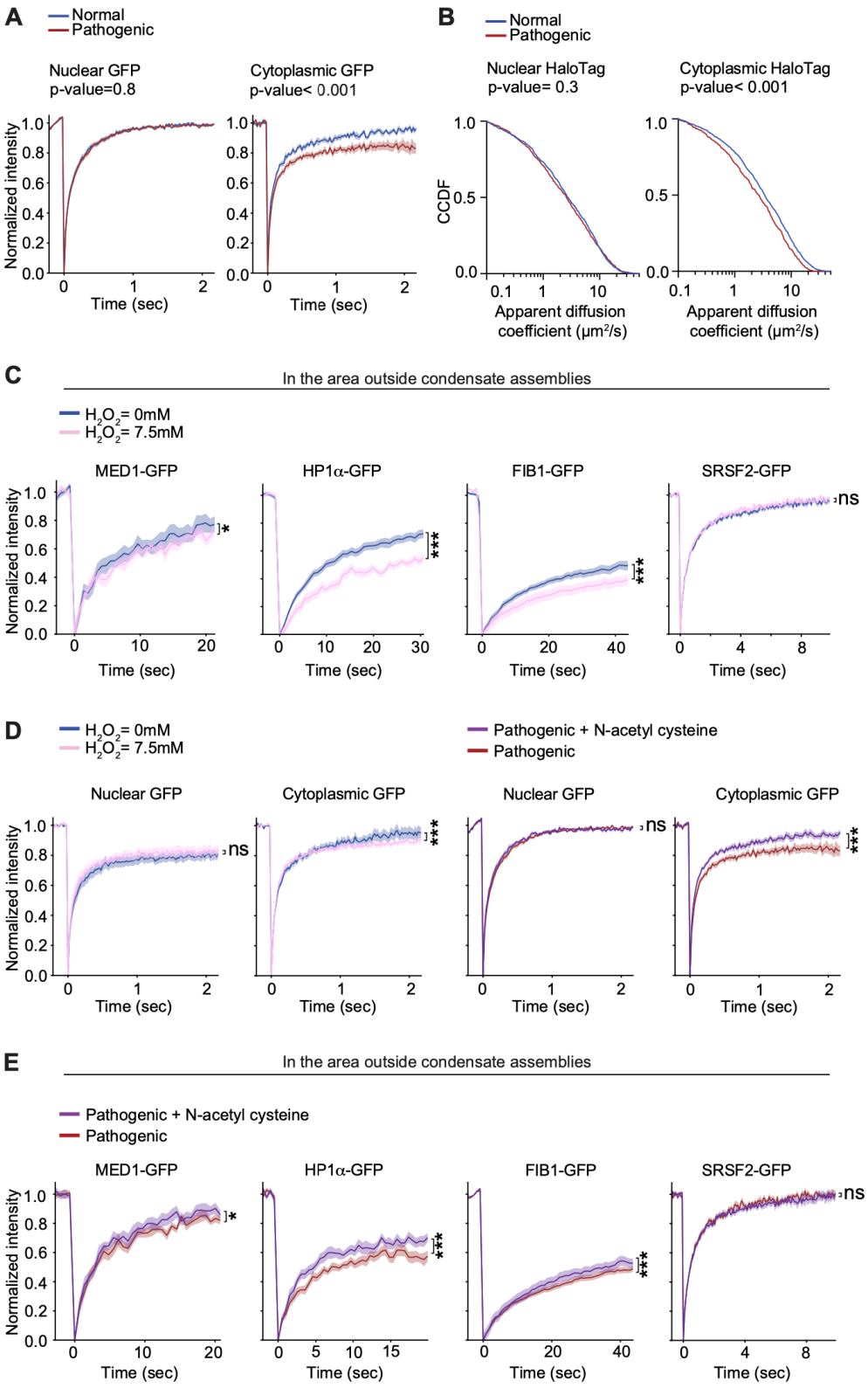
(A) log*D* distribution of individual molecules (histogram) fitted to either two or three Gaussian functions (colored curves indicate individual Gaussian functions, and the black curve is each graph is the sum of individual Gaussian functions).
(B) Distribution of the continuous axial detectable range of a single molecule. The distribution peaked at ~900 nm.
(C) Immobile fraction of endogenous proteins estimated from SPT dataset vs FRAP dataset. Data are plotted as mean + SEM.
(D) Plot of the residual of experimentally observed and theoretical models of FRAP recovery as a function of different diffusion coefficient *D* used for the model. The best fitted diffusion coefficient of GFP (27kDa) is indicated by a dashed blue line. The apparent diffusion coefficients of a HaloTag-JF646 (~34kDa) based on SPT are indicated as mean (dashed green line) and SEM (light green). The apparent diffusion coefficient of a HaloTag-JF646 (~34kDa) inferred from the relationship between FRAP-estimated diffusion of a protein of known molecular weight, GFP (27kDa), is indicated by a dashed magenta line.
(E) The log*D* distribution of individual HaloTag-alone molecules in fixed sample (histogram). The solid black line represents the best-fitted single Gaussian function.
(F) The log*D* distribution of individual HaloTag-alone molecules in live sample (histogram). The solid black line represents the best-fitted single Gaussian function.
(G) Graphical illustration of a premature stop during localization reconnection caused by assuming a too small maximumly-allowed prior apparent diffusion coefficient (D_{max}) (left) versus a successful reconnection because of assuming a large enough D_{max} (right).
(H) Graphical illustration of two reasons why the tracking of a protein may stop even if the next localization is within the 2-dimentional range defined by D_{max}: (i) photobleach of the dye molecule tagged to the protein and (ii) the protein moving out of focus.
(I) Graphical illustration of why the localization reconnection may continue by mistakenly joining the trajectories of two proteins together. In this case, the number of jumps per "trajectory" will go beyond normal.
(J) Average number of jumps per trajectory at different localization density. Beyond certain localization density threshold (vertical line at ~0.01 per μm^2 per frame), the number of jumps per "trajectory" will start to increase due to the reason shown in (I), which is associated with significant chance of ambiguous connection. The localization density range of the actual experiments for IR-HaloTag SPT is marked as the horizontal boxplot, which is safely below the threshold that will cause significant ambiguous connection.

Figure S3. The effect of pathogenic signaling on protein mobility outside of condensates and on condensate properties. Related to Figure 2.



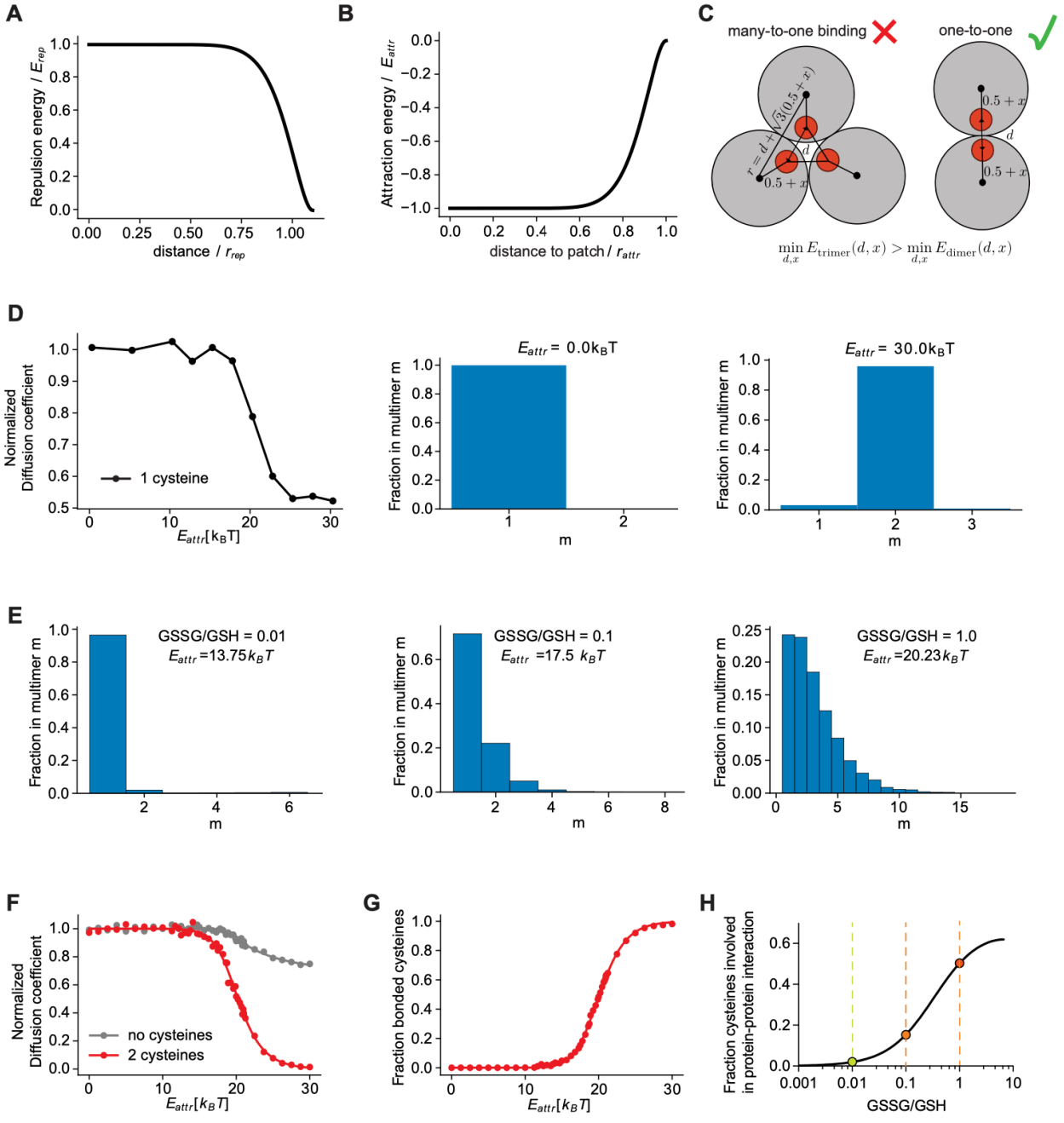
(A) Quantification of FRAP data for MED1, HP1α, FIB1 and SRSF2 in areas outside of condensate assemblies in cells that were treated with normal signaling (Normal) or pathogenic signaling (Pathogenic). MED1 normal n=10 cells, pathogenic n=10 cells; HP1α normal n=15 cells, pathogenic n=15 cells; FIB1 normal n=24 cells, pathogenic n=24 cells; SRSF2 normal n=14 cells, pathogenic n=14 cells. T-test was used for statistical analysis.
(B) Number, size and partition ratio of IR, MED1, HP1α, FIB1 and SRSF2 condensates in cells that were treated with normal signaling (Normal) or pathogenic signaling (Pathogenic). Number of condensates: IR normal n=22 cells, pathogenic n=24 cells; MED1 normal n=135 cells, pathogenic n=127 cells; HP1α normal n=44 cells, pathogenic n=56 cells; FIB1 normal n=150 cells, pathogenic n=214 cells; SRSF2 normal n=58 cells, pathogenic n=53 cells. Condensate size: IR normal n=3846 condensates, pathogenic n=3548 condensates; MED1 normal n=3522 condensates, pathogenic n=3426 condensates; HP1α normal n=1499 condensates, pathogenic n=1558 condensates; FIB1 normal n=541 condensates, pathogenic n=660 condensates; SRSF2 normal n=954 condensates, pathogenic n=699 condensates. Partition ratio: IR normal n=3846 condensates, pathogenic n=3548 condensates; MED1 normal n=3522 condensates, pathogenic n=3426 condensates; HP1α normal n=1499 condensates, pathogenic n=1558 condensates; FIB1 normal n=541 condensates, pathogenic n=660 condensates; SRSF2 normal n=954 condensates, pathogenic n=699 condensates.

Figure S4. The effect of oxidative environments on protein mobility. Related to Figure 2 and 3.



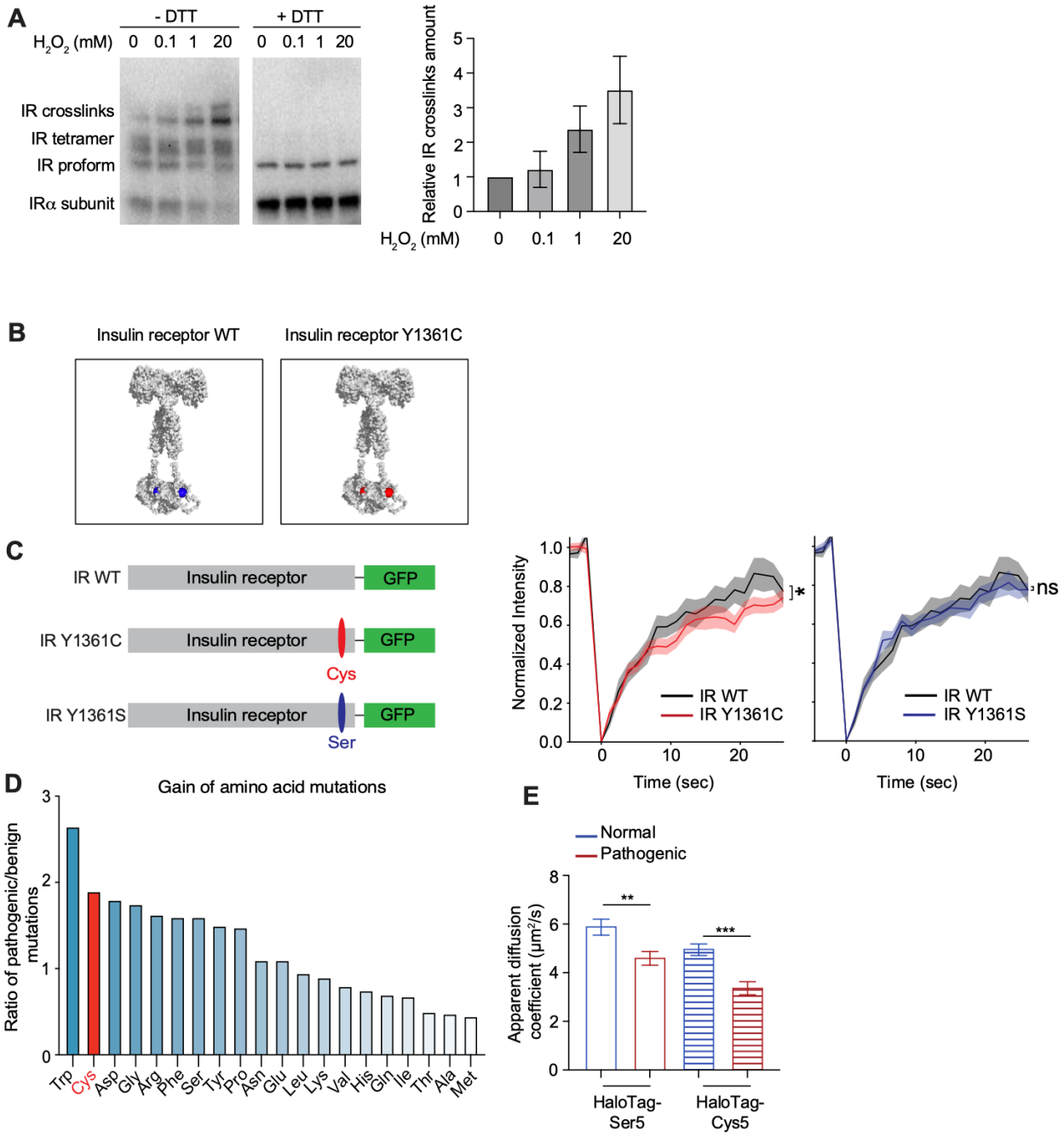
(A) Quantification of FRAP data for nuclear (left) and cytoplasmic (right) GFP in HepG2 cells that were treated with normal signaling (Normal) or pathogenic signaling (Pathogenic). Nuclear GFP: Normal n=16 cells, Pathogenic n=17 cells; Cytoplasmic GFP: n=15 cells for each condition. Data are plotted as mean (dark blue and dark red lines) ± SEM (light blue and light red regions).
(B) CCDF graphs of apparent diffusion coefficients as determined by SPT for nuclear (left) and cytoplasmic (right) HaloTag in HepG2 cells that were treated with normal signaling (Normal) or pathogenic signaling (Pathogenic). Nuclear HaloTag: Normal n=771 protein molecules, Pathogenic n=937 protein molecules; Cytoplasmic HaloTag: Normal n=1279 protein molecules, Pathogenic n=625 protein molecules. T-test was used for statistical analysis.
(C) Quantification of FRAP data for MED1, HP1α, FIB1, and SRSF2 in areas outside of condensates in HepG2 cells that were treated with 0mM H_2O_2 or 7.5mM H_2O_2. 0mM H_2O_2 n=11, 20, 15, 14 cells for MED1, HP1α, FIB1, SRSF2, respectively. 7.5mM H_2O_2 n=11, 20, 15, 14 cells for MED1, HP1α, FIB1, SRSF2, respectively.
(D) Quantification of FRAP data for nuclear and cytoplasmic GFP in HepG2 cells that were treated with 0mM H_2O_2 or 7.5mM H_2O_2 (n=7 cells per condition for nuclear FRAP and n=10 cells per condition for cytoplasmic FRAP). Data are plotted as mean (dark pink and dark blue lines) ± SEM (light pink and light blue regions). Quantification of FRAP data for nuclear GFP in HepG2 cells previously treated with pathogenic signaling with (Pathogenic + NAC, n=17 cells) or without (Pathogenic, n=11 cells) N-acetyl cysteine (left). Quantification of FRAP data for cytoplasmic GFP in HepG2 cells previously treated with pathogenic signaling with (Pathogenic + NAC, n=10 cells) or without (Pathogenic, n=10 cells) N-acetyl cysteine (left).
(E) Quantification of FRAP data for MED1, HP1α, FIB1, SRSF2 in areas outside of condensates in n HepG2 cells previously treated with pathogenic signaling with or without N-acetyl cysteine (NAC). Without NAC n=11, 20, 37, 15 cells for MED1, HP1α, FIB1, SRSF2, respectively. With NAC =11, 20, 28, 15 for MED1, HP1α, FIB1, SRSF2, respectively. Data are plotted as mean (dark purple and dark red lines) ± SEM (light purple and light red regions). T-test was used for statistical analysis.

Figure S5. Modeling protein diffusion with increasing ROS. Related to Figure 4.



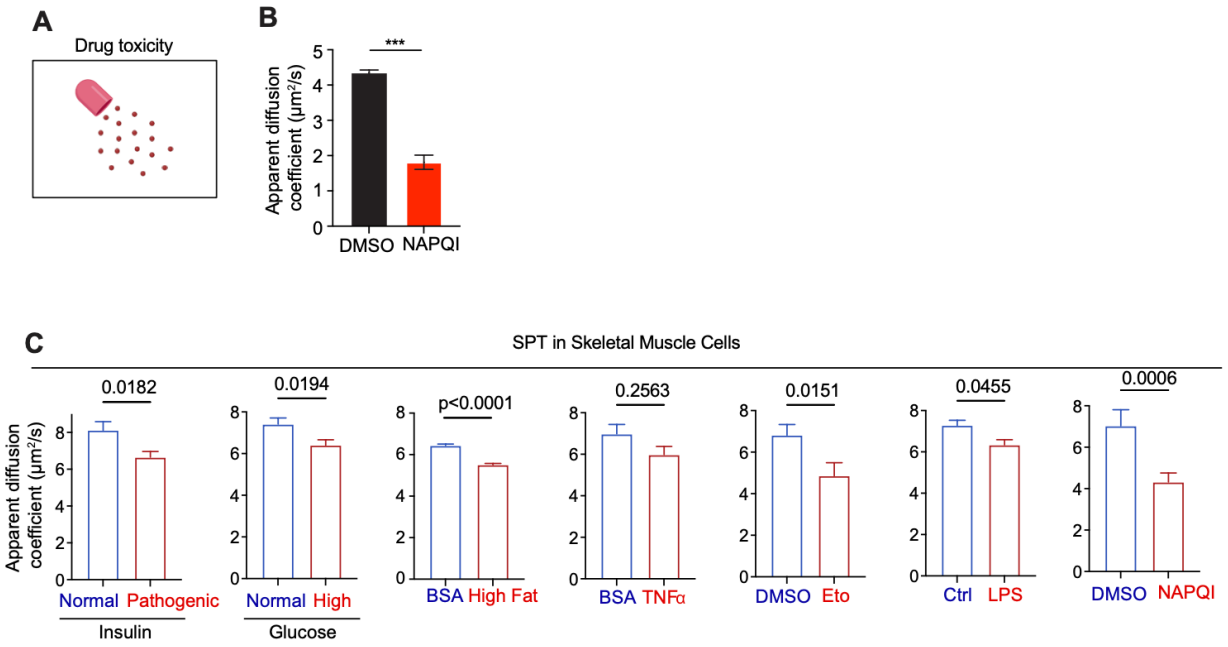
(A) Repulsive potential between protein spheres as a function of the inter-protein distance normalized to the protein diameter r_{rep}.
(B) Attractive potential between sticky patches (surface cysteines) as a function of the inter-patch distance normalized to the patch-patch attraction radius r_{attr}. The depth of the attractive potential, E_{attr}, controls the propensity for intermolecular disulfide bonding.
(C) Cartoon depicting minimum energy configurations of a trimer of proteins with a single surface cysteine, which represents an undesirable many-to-one bonding event, and a dimer, which represents a one-to-one bonding event. These configurations are determined by the equilibrium patch-patch distance d and protein-cysteine bond extension *x* which minimize the energy of the trimer or dimer. Simulation parameters are chosen such that trimers are energetically less favorable than dimers.
(D) Diffusion coefficient and cluster size distributions from simulations of 1000 proteins with one surface-exposed cysteine as a function of E_{attr}, normalized to the mean of the first five data points. At $E_{attr} = 0\ k_BT$, all proteins are in a monomeric state, but at $E_{attr} = 30\ k_BT$, nearly all proteins form dimers. Notably none form trimers, demonstrating that the choice of simulation hyperparameters minimizes many-to-one bonding.
(E) Fraction of simulated proteins that form a multimer of size m for simulations at three different values of E_{attr}, which correspond to three different GSSG/GSH ratios highlighted in (H).
(F) Normalized diffusion coefficient from simulations of a mixture of proteins with (red) and without (gray) surface-exposed sticky patches (cysteines) as a function of the patch-patch attraction energy E_{attr}.
(G) Fraction of surface cysteines which participate in intermolecular bonding as a function of E_{attr}, as calculated from protein dynamics simulations.
(H) Fraction of surface cysteines which participate in protein-protein disulfide bonding as a function of the steady state ratio of oxidized (GSSG) to reduced (GSH) glutathione, as calculated from a chemical reaction model. Representative GSSG/GSH ratios are highlighted with dashed lines and circles (yellow, orange and red).

Figure S6. Cysteines in protein mobility. Related to Figure 4.



(A) Immunoblot for IR (left). Cells were treated with indicated concentrations of H_2O_2 prior to protein isolation and western blotting in reducing (+ DTT) or non-reducing (- DTT) conditions. Quantification of relative IR crosslink amount as compared to IRα subunit (right).
(B) Renderings of dimers of wildtype (WT) or dimers of Y1361C IR. Cartoon design was based on both the previously published structure of IR (PDB 6PXV) and the AlphaFold structure of the unresolved region of IR. Tyrosines are represented as blue and cysteines are represented in red.
(C) Cartoon depicting wild type (IR WT) and mutant IR (IR Y1361C and IR Y1361S; left). Quantification of FRAP data for WT (IR WT, n= 15 cells) and mutant IR (IR Y1361C, n= 16 cells; IR Y1361S, n= 15 cells; right). These experiments were performed on the same day and as a result the IR WT FRAP curves are the same. Data are plotted as mean (dark black, dark red, and dark blue lines) ± SEM (light black, light red, and light blue regions). t-test was used for statistical analysis.
(D) Measurement of the pathogenicity of all twenty gain-of-amino acid mutations as determined by the ratio of the number of pathogenic mutations to the number of benign mutations for a specific amino acid throughout the proteome.
(E) Apparent diffusion coefficient for HaloTag-Ser5 and HaloTag-Cys5 in cells treated with normal or pathogenic insulin. HaloTag-Ser5 normal n=710 molecules, HaloTag-Ser5 pathogenic n=747 molecules; HaloTag-Cys5 normal n=1239 molecules, HaloTag-Cys5 pathogenic n= 569 molecules. Mann-Whitney was used for statistical analysis.

Figure S7. Pathogenic stimuli decrease protein mobility. Related to Figure 5.



(A) Cartoon depicting drug toxicity.
(B) Apparent diffusion coefficient as determined by SPT of the protein mobility biosensor expressed in HepG2 cells treated with (NAPQI, n=408 protein molecules) and without (DMSO, n=4921 protein molecules) *N*-acetyl-*p*-benzoquinone imine (NAPQI). NAPQI is a toxic intermediate in the breakdown of acetaminophen and is one of the main causes of acetaminophen-induced liver injury. Data are plotted as mean ± SEM. Mann-Whitney was used for statistical analysis.
(C) Apparent diffusion coefficient as determined by SPT of the protein mobility biosensor expressed in C2C12 skeletal muscle cells treated with the stimuli reported in the figure. Data are plotted as mean + SEM. Mann-Whitney test was used for statistical analysis. Numbers of molecules: normal insulin (327) vs. pathogenic insulin (510); normal glucose (706) vs. high glucose (673); BSA (42133) vs. high fat (38486); BSA (294) vs. TNFa (291); DMSO (186) vs. ETO (91); control (1015) vs. LPS (777); control (75) vs. NAPQI (156).

Table S1. Diffusion coefficients of human proteins. Related to Figure 1.

Protein	Cell Line	Diffusion Coefficient*	Reference
Tom20	Hela	0.02-0.49 $\mu m^2/s$	Bhagawati et al., 2020, *Angew Chem Int Ed Engl*[1]
Tom7	Hela	0.045 $\mu m^2/s$	Bhagawati et al., 2021, *Mol Biol Cell*[2]
Bace1	SH-SY5Y	0.1 $\mu m^2/s$ (median)	Capitini et al., 2023, *iScience*[3]
APP	SH-SY5Y	0.02 $\mu m^2/s$ (median)	Capitini et al., 2023, *iScience*[3]
Zap70	Jurkat	1.34 $\mu m^2/s$	Carr et al., 2017, *Biophys J*[4]
TCR	Jurkat	0.110 $\mu m^2/s$	Carr et al., 2017, *Biophys J*[4]
HIF-2α	786-O	5 $\mu m^2/s$	Chen et al., 2022, *Elife*[5]
HIF-1β	786-O	10 $\mu m^2/s$	Chen et al., 2022, *Elife*[5]
CD9	HeLa	0.24 $\mu m^2/s$	Dahmane et al., 2014 *Viruses*[6]
CD81	HepG2	0.17 $\mu m^2/s$	Dahmane et al., 2014 *Viruses*[6]
EGFR	A431	0.1 $\mu m^2/s$	Delcanale et al., 2020, *Angew Chem Int Ed Engl*[7]
pre-BCR	697 and Nalm6	0.09 - 0.13 $\mu m^2/s$	Erasmus et al., 2016, *Sci Signal*[8]
CD9	HB2	0.17 $\mu m^2/s$ (median)	Fernandez et al., 2021, *Int J Mol Sci*[9]
CD81	HB2	0.06 $\mu m^2/s$ (median)	Fernandez et al., 2021, *Int J Mol Sci*[9]
CD82	HB2	0.03 $\mu m^2/s$ (median)	Fernandez et al., 2021, *Int J Mol Sci*[9]
RON	A431	0.032 $\mu m^2/s$	Franco et al., 2021, *Elife*[10]
CXCR4	Jurkat	0.027 $\mu m^2/s$ (median)	García-Cuesta et al., 2022, *PNAS*[11]
CXCR4	HEK-293T	0.017 $\mu m^2/s$	Gardeta et al., 2022, *Front Immunol*[12]
CD81	HepG2	0.17 $\mu m^2/s$ (median)	Harris et al., 2013, *Cell Microbiol*[13]
MET	HeLa	0.126 $\mu m^2/s$	Harwardt et al., 2017 *FEBS Open Bio*[14]
Cav1	Hela	0.006-0.016 $\mu m^2/s$ (median)	Hirama et al., 2017, *J Biol Chem*[15]
CD3ε	Jurkat	0.095 $\mu m^2/s$	Ito et al. 2017, *Sci. Rep.*[16]
CD45	Jurkat	0.23 $\mu m^2/s$	Ito et al., 2017 *Sci. Rep.*[16]
c-Myc	U2OS	9.8 $\mu m^2/s$	Izeddin et al., 2014, *Elife*[17]

ADAM10	U2OS	0.067 $\mu m^2/s$	Jouannet et al., 2016, *Cell Mol Life Sci*[18]
Band3	human erythrocytes	0.0016-0.021 $\mu m^2/s$	Kodippili et al., 2020, *Biophys J*[19]
GLUT1	human erythrocytes	0.00043-0.0093 $\mu m^2/s$	Kodippili et al., 2020, *Biophys J*[19]
Duffy	human erythrocytes	0.00025-0.005 $\mu m^2/s$	Kodippili et al., 2020, *Biophys J*[19]
glycophorin A	human erythrocytes	0.00011-0.034 $\mu m^2/s$	Kodippili et al., 2020, *Biophys J*[19]
$KRas^{G12D}$	U2OS	0.08-0.81 $\mu m^2/s$	Lee et al. 2019, *Elife*[20]
CXCR4	Jurkat	0.0047 $\mu m^2/s$ (median)	Martínez-Muñoz et al., 2018, *Mol Cell*[21]
Scc2	HCT116	0.6 $\mu m^2/s$	Rhodes et al., 2017, *Elife*[22]
EGFR	A431	0.025 $\mu m^2/s$	Starok et al., 2015, *Biomacromolecules*[23]
hCD1d	THP-1	0.029 $\mu m^2/s$ (median)	Torreno-Pina et al., 2016, *PNAS*[24]
CD44v	MCF7-LR	0.016-0.213 $\mu m^2/s$	Tseng et al., 2019, *Life Sci Alliance*[25]
GRP78	MCF7-LR	0.00095-0.104 $\mu m^2/s$	Tseng et al., 2019, *Life Sci Alliance*[25]
CD56	293T	0.058 $\mu m^2/s$	Wäldchen et al., 2020, *Nat Commun*[26]
STIM1	HEK	0.116 $\mu m^2/s$	Wu et al., 2014, *Mol Biol Cell*[27]
Orai1	HEK	0.090 $\mu m^2/s$	Wu et al., 2014, *Mol Biol Cell*[27]
α6 integrin	MDA-MB-231	0.0004 $\mu m^2/s$	Yang et al., 2012, *J Cell Sci*[28]
IFNAR1 and 2	U5A	0.022-0.043 $\mu m^2/s$	You et al., 2016, *Sci Adv*[29]
EZH2	U2OS	2.09 $\mu m^2/s$	Youmans et al., 2018, *Genes Dev*[30]
SUZ12	U2OS	1.99 $\mu m^2/s$	Youmans et al., 2018, *Genes Dev*[30]

* Average diffusion coefficient unless otherwise noted.

Table S2. Information regarding SPT data collected for comparison of protein mobility in a pathogenic signaling condition. Related to Figure 2.

Protein	Condition	Total # of trajectories	# of cells examined	p-value *	Day
IR	Normal	294	46		1
	Normal	357	42		2
	Pathogenic	276	23	4.7E-05	1
	Pathogenic	154	34	2.0E-04	2
MED1	Normal	5719	214		1
	Normal	1751	281		2
	Pathogenic	2227	110	9.9E-09	1
	Pathogenic	1233	259	9.5E-03	2
HP1α	Normal	4568	180		1
	Normal	5199	157		2
	Normal	2591	180		3
	Pathogenic	3598	219	1.6E-13	1
	Pathogenic	3529	171	6.4E-62	2
	Pathogenic	1994	187	9.4E-12	3
FIB1	Normal	2855	205		1
	Normal	153	53		2
	Pathogenic	2561	162	9.8E-03	1
	Pathogenic	146	22	3.1E-02	2
SRSF2	Normal	5458	324		1
	Normal	3399	142		2
	Pathogenic	7267	323	5.1E-02	1
	Pathogenic	2872	162	2.3E-01	2

* p-values shown are for the decrease in protein mobility in pathogenic sample as compared to matching normal sample.

Table S3. Half-time of fluorescence recovery for IR, MED1, HP1α, FIB1 and SRSF2. Related to Figure 2, 3, S3 and S4.

Condensed Phase related to Figure 2G						
	Normal			**Pathogenic**		
	t_half_avg	t_half_sem	n of cells	t_half_avg	t_half_sem	n of cells
IR	7.3	0.8	20	13.9	2.1	13
MED1	5.3	0.7	10	7.6	1.2	10
HP1α	3.1	0.4	14	5.9	0.5	14
FIB1	9	0.7	24	13.9	1.2	24
SRSF2	0.6	0	14	0.6	0	14
In the area outside condensate assemblies related to Figure S3A						
	Normal			**Pathogenic**		
	t_half_avg	t_half_sem	n of cells	t_half_avg	t_half_sem	n of cells
MED1	2.9	0.4	10	4.7	0.6	10
HP1α	3.7	0.5	15	4.8	0.6	15
FIB1	10.8	1.1	24	16.8	1.3	24
SRSF2	0.6	0	14	0.7	0.1	14
Condensed Phase related to Figure 3F						
	H_2O_2 = 0 mM			**H_2O_2 = 7.5 mM**		
	t_half_avg	t_half_sem	n of cells	t_half_avg	t_half_sem	n of cells
IR	7.0	0.9	10	12.0	1.9	10
MED1	6.9	1.2	15	10.7	1.1	15
HP1α	6.7	0.9	14	12.1	0.9	14
FIB1	8.8	0.7	15	19.1	4.0	15
SRSF2	0.5	0.0	14	0.5	0.0	14
In the area outside condensate assemblies related to Figure S4C						
	H_2O_2 = 0 mM			**H_2O_2 = 7.5 mM**		
	t_half_avg	t_half_sem	n of cells	t_half_avg	t_half_sem	n of cells
MED1	4.6	1.0	11	5.5	0.8	11
HP1α	6.4	0.8	20	13.8	1.1	20
FIB1	10.3	1.5	15	24.0	5.4	15
SRSF2	0.7	0.0	14	0.7	0.0	14
Condensed Phase related to Figure 3I						
	Pathogenic			**Pathogenic+NAC**		
	t_half_avg	t_half_sem	n of cells	t_half_avg	t_half_sem	n of cells
IR	12.9	2.3	15	6.6	0.8	15
MED1	7.1	0.6	11	5.2	0.7	11
HP1α	6	0.5	14	4	0.4	14

FIB1	13.7	1.1	37	12.2	1.4	28
SRSF2	0.6	0	15	0.6	0.1	15
In the area outside condensate assemblies related to Figure S4E						
	Pathogenic			**Pathogenic+NAC**		
	t_half_avg	t_half_sem	n of cells	t_half_avg	t_half_sem	n of cells
MED1	3.3	0.3	11	2.8	0.5	11
HP1α	4.7	0.6	20	2.9	0.4	20
FIB1	15.2	2.1	37	13.9	1.2	28
SRSF2	0.7	0	15	0.7	0.1	15

Table S4. List of cysteines and predicted surface-exposure in IR, MED1, HP1α and FIB1. Related to Figure 4.

Protein	Amino acid #	Amino acid	Solvent accessible area (Å^2)	Surface-exposed cysteine	PDB
FIB 1	99	C	56.1	YES	7SE7
FIB 1	268	C	49	YES	7SE7
HP1α	59	C	in IDR	YES	
HP1α	133	C	1.8		AF-P45973-F1
HP1α	160	C	11.9		3I3C
IR	35	C	15.3		6PXV
IR	53	C	1.9		6PXV
IR	153	C	12		6PXV
IR	182	C	31.6	YES	6PXV
IR	186	C	33.8	YES	6PXV
IR	196	C	11.1		6PXV
IR	209	C	0		6PXV
IR	215	C	40		6PXV
IR	219	C	25.1	YES	6PXV
IR	223	C	8.8		6PXV
IR	228	C	5		6PXV
IR	234	C	41.3	YES	6PXV
IR	235	C	11.5		6PXV
IR	239	C	0.2		6PXV
IR	243	C	4.1		6PXV
IR	252	C	4.2		6PXV
IR	255	C	7.9		6PXV
IR	264	C	8.5		6PXV
IR	268	C	10.7		6PXV
IR	280	C	12.1		6PXV
IR	286	C	0.2		6PXV
IR	293	C	17.3		6PXV
IR	301	C	94.4	YES	6PXV
IR	311	C	4.1		6PXV
IR	315	C	17.6		6PXV
IR	328	C	8.2		6PXV
IR	331	C	45.1	YES	6PXV

IR	335	C	53.2	YES	6PXV
IR	339	C	26.1	YES	6PXV
IR	360	C	3.1		6PXV
IR	462	C	15.9		6PXV
IR	495	C	56	YES	6PXV
IR	551	C	12.2		6HN5
IR	674	C	31.5	YES	6PXV
IR	709	C	in IDR	YES	
IR	710	C	in IDR	YES	
IR	712	C	in IDR	YES	
IR	825	C	0		6PXV
IR	834	C	22.9	YES	6PXV
IR	899	C	29.9	YES	6PXV
IR	911	C	0		6PXV
IR	1008	C	114	YES	AF-P06213-F1
IR	1083	C	6.6		5HHW
IR	1165	C	0.2		5HHW
IR	1261	C	22.8	YES	5HHW
IR	1272	C	0		5HHW
IR	1335	C	in IDR	YES	
MED1	61	C	28.5	YES	7EMF
MED1	101	C	18.8		7EMF
MED1	121	C	62.6	YES	AF-Q15648-F1
MED1	135	C	12.9		7EMF
MED1	302	C	0.8		7EMF
MED1	324	C	28.1	YES	7EMF
MED1	373	C	0.6		7EMF
MED1	424	C	0.8		7EMF
MED1	443	C	14.8		7EMF
MED1	464	C	9		7EMF
MED1	477	C	8.1		7EMF
MED1	489	C	17.9		7EMF
MED1	501	C	7.8		7EMF
MED1	681	C	in IDR	YES	
MED1	745	C	in IDR	YES	
MED1	800	C	in IDR	YES	
MED1	1266	C	in IDR	YES	

Table S5. FRAP and SPT imaging settings. Related to Materials and Methods.

FRAP imaging settings						
POIs:	**Insulin receptor**	**HP1α**	**MED1**	**FIB1**	**SRSF2*#**	**GFP alone#**
Image Type	8-bit	8-bit	8-bit	8-bit	8-bit	8-bit
Acquisition mode	AiryScan SR	AiryScan SR	AiryScan SR	AiryScan SR	Multiplex SR-4Y	Multiplex SR-8Y
Laser power	2%	2%	2%	1%	1.5%	0.5%
Detector gain	900	900	900	800	800	700
Time interval	0.7 s	0.6 s	0.7 s	0.9 s	0.08 s	0.03 s
Total frames	30	44	44	54	142	110
Initial frames before photo bleaching	4	4	4	4	10	10
Bleaching cycles (100% 488 nm laser)	2	3	3	4	6	6
Remaining intensity fraction right after photobleaching (AVG±STD)%	(34±7)%	(20±7)%	(29±8)%	(27±5)%	(37±7)%	(59±12)%
Pixel size	66 nm	66 nm	66 nm	75 nm	43 nm	43 nm

Imaging area	512x512 pixels	512x512 pixels	512x512 pixels	512x512 pixels	632x632 pixels	528x528 pixels
Definite focus	On	On	On	On	Off	Off
Fluorescence bleached throughout a whole acquisition course (refer to an unbleached region) (AVG±STD)%	(16±8)%	(9±4)%	(7±4)%	(12±7)%	(11±3)%	(13±5)%
The number of final time points being binned for *t*-test	4	5	5	5	10	7
SPT imaging settings						
POIs:	**Insulin receptor**	**HP1α**	**MED1**	**FIB1**	**SRSF2**	**HaloTag alone†**
Dye used for tracking	Halo-(PA)JF549**	Halo-JF646	Halo-JF646	Halo-JF646	Halo-JF646	Halo-JF646
Tracking dye staining condition	10 nM, 2 hours	5 nM, 15 min	7.5 nM, 15 min	2.5 nM, 15 min	2.5 nM, 15 min	2.5 nM, 15 min
Image Type	16-bit	16-bit	16-bit	16-bit	16-bit	16-bit
Laser wavelength	561 nm	638 nm	638 nm	638 nm	638 nm	638 nm
Laser power	100%	90%	80%	100%	100%	100%

Collimator	1000	1378	1378	1781	1378	800
Time interval	5 ms	7 ms	7 ms	10 ms	6 ms	4 ms
Total frames	5,000	1,500	1,500	1,500	1,500	1,500
Pixel size	97 nm	97 nm	97 nm	97 nm	97 nm	97 nm
Imaging area	600x600 pixels	1024x1024 pixels	1024x1024 pixels	1024x1024 pixels	1024x1024 pixels	512x512 pixels
Prior D_{max} allowed when reconnecting sequential localizations	6 $\mu m^2/s$	6 $\mu m^2/s$	6 $\mu m^2/s$	3 $\mu m^2/s$	6 $\mu m^2/s$	9 $\mu m^2/s$
disappearance probability (exponential decay) for blinking	1 frame	1 frame	1 frame	1 frame	1 frame	1 frame
Maximum disappearance frames allowed for reconnection	2 frames	2 frames	2 frames	2 frames	2 frames	2 frames
Dye for visualizing the bulk distribution	Halo-JF646	Halo-JF549	Halo-JF549	Halo-JF549	Halo-JF549	Halo-JF549
Bulk dye staining condition	50 nM, 15 min	50 nM, 15 min	50 nM, 15 min	50 nM, 15 min	50 nM, 15 min	50 nM, 15 min
Region being analyzed	Plasma membrane	Nucleus	Nucleus	Nucleolus	Nucleus	Nucleus

*SRSF2-GFP tagged with five serines or cysteines adopt the same protocol.

#We switched to the Multiplexed mode when AiryScan mode could not provide high enough frame rate for capturing the fluorescence recovery dynamics.
†HaloTag alone tagged with five cysteines or serines adopt the same protocol.
**For (PA)JF549 dye, a very low 405 nm laser (<0.1%) is also applied to activate dye molecules sparsely.

Table S6. Sources of false positive identifications of proteins in SPT. Related to Materials and Methods.

Source of false positiveness	Sample	Localization density (μm^2/frame)	Trajectory density (μm^2/s)
Pixel noise	Empty region	$(8.3\pm0.2)\times10^{-6}$	0.0±0.0
Auto-fluorescence	HaloTag-IR HepG2 without dye	$(6.4\pm0.1)\times10^{-4}$	$(1.6\pm0.3)\times10^{-3}$
Non-specific dye staining*	WT HepG2 + PAJF549 dye	$(4.0\pm4.1)\times10^{-4}$	$(3.6\pm3.2)\times10^{-3}$
Actual experiment	HaloTag-IR HepG2 + PAJF549 dye	$(2.1\pm0.7)\times10^{-3}$	$(3.1\pm1.8)\times10^{-2}$

*This issue is only significant with PAJF549 dye staining. The JF646 dye has negligible non-specific staining issue (data not shown).

Materials and Methods

RESOURCE AVAILABILITY

Materials Availability

All plasmids and cell lines generated in this study are available upon request.

Data and Code availability

Metabolomic datasets generated in this study have been deposited in Metabolights under study ID number MTBLS9535. RNA-seq datasets generated in this study have been deposited in GEO. All codes are available to researchers by the corresponding authors.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

HepG2 (male) cells were acquired from ATCC (ATCC, HB-8065TM) and cultured in 150 mm cell culture grade dishes with EMEM media (ATCC, 30-2003) supplemented with 10% fetal bovine serum (FBS; Sigma-Aldrich, F4135) without antibiotic and kept in a humidified incubator at 37°C with 5% CO_2. These cells were chosen because they are widely used to study diverse pathologies and because they can be genetically modified. To passage the cells, 20 ml of room-temperature phosphate buffered saline solution (Gibco, 10010-023) was added to the dish, aspirated off, then 3 ml of TrypLE Express Enzyme (Life Technologies, 12604021) was added to help dissociate cells. The dish was then incubated at 37°C with humidity and 5% CO_2 for 5 minutes. After 5 minutes, cells were mechanically dissociated by pipetting them up and down 7 times using with a 10ml serological pipette fitted with a p200 tip. To quench the TrypLE, 7 ml of EMEM-FBS was added to the dish. 1 ml of the cell suspension was left on the dish and 20 ml of EMEM-FBS was added on top. HepG2 cells were continuously cultured in a 150 mm dish and split 1:10 when the cells became confluent. The cells were subcultured on a new plate monthly, seeded using a 1:2 split.

C2C12 (female) cells were acquired from ATCC (ATCC, CRL-1772) and cultured in 150 mm cell culture grade dishes with DMEM media (Gibco, 11965-092) supplemented with 10% fetal bovine serum (FBS; Sigma-Aldrich, F4135) without antibiotic and kept in a humidified incubator at 37°C with 5% CO_2. These cells were chosen because they are widely used to study diverse pathologies and because they can be genetically modified. Cells were passaged at 30-50% confluence to prevent differentiation. To passage the cells, 20 ml of room-temperature phosphate buffered saline solution (Gibco, 10010-023) was added to the dish, aspirated off, then 3 ml of TrypLE Express Enzyme (Life Technologies, 12604021) was added to help dissociate cells from the dish and one another. The dish was then incubated at 37°C with humidity and 5% CO_2 for 5 minutes. After 5 minutes, the cells were dissociated by tapping the sides of the plate. To quench the TrypLE, the cells were resuspended in 7 ml of TrypLE, and these resuspended cells were used to seed a new dish at a 1:20 dilution.

Method details

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Constructs and construct generation

For tagging endogenous proteins (IR, MED1, HP1α, FIB1, and SRSF2) with GFP, HaloTag, or GFP-5xSer/Cys, the homology directed repair (HDR) strategy of CRISPR was adopted. For this strategy, three components are needed: 1) Cas9 protein to cut the DNA, 2) sgRNA to guide Cas9 to the desired target, and 3) a DNA repair template that contains the desired edit as well as 800 bp of homologous sequence immediately upstream and downstream of the target. The sgRNA sequence and Cas9 coding sequence for transient expression of both in cells were integrated in the same plasmid (which was refer to as "sgRNA-Cas9 plasmid"), while the repair templates were integrated into a second plasmid.

sgRNA-Cas9 plasmids. 20 bp of target sequences were cloned into a plasmid containing sgRNA backbone, a codon-optimized version of Cas9, and mCherry. The mCherry was used during FACS sorting to select for Cas9-mCherry$^+$ cells. Constructs for the generation of MED1-GFP, HP1α-GFP, SRSF2-GFP, FIB1-GFP, and insulin receptor-GFP (IR-GFP) cell lines were described in previous publications[28,103]. To generate the IR-Y1361C-GFP cell line, the following sgRNA sequences with PAM sequence in parentheses were used for CRISPR/Cas9 editing:
sgRNA_IR_C-term_1: CACGGTAGGCACTGTTAGGA(AGG)
sgRNA_IR_C-term_2: TAGGCACTGTTAGGAAGGAT(TGG)
sgRNA_IR_C-term_3: CCTCCGTTCATGTGTGTGTA(AGG)
The other sgRNA sequences are reported in previous publications [28,103].
Cloning was performed using NEBuilder HiFi DNA Assembly Master Mix (NEB, E2621S) according to manufacturer's specifications.

Repair templates for GFP tagging. Approximately 800bp of Homo sapiens genomic DNA sequences flanking the Cas9 cutting sites were cloned into the pUC19 vector using NEBuilder HiFi DNA Assembly Master Mix (NEB, E2621S), with in-frame monomeric enhanced fluorescent protein (GFP) sequence being inserted together with a flexible 10-animo acid linker sequence (GGSGGGGSGG) to space the fluorophore and the protein of interest. Constructs for MED1-GFP, HP1α-GFP, SRSF2-GFP, FIB1-GFP, and IR-GFP cell line generations are described in previous publications[28,103]. For the IR-Y1361C-GFP cell line generation, the homology repair template consists of INSR exon 22 containing the Y1361C missense mutation in frame with GFP flanked on either side by 800-bp homology arms amplified from HepG2 genomic DNA using PCR. For SRSF2-GFP-Ser/Cys cell line generation, the SRSF2-GFP repair template was modified to fuse SRSF2-GFP to a flexible linker followed by either a 5xSer array or a 5xCys array. The 5xSer array contains 5 serines spaced by a rigid linker (AEAAAKEAAAKA)[110], while the 5xCys array contains 5 cysteines spaced by the same

rigid linker[110]. These constructs were cloned using NEBuilder HiFi DNA Assembly Master Mix.

Repair templates for HaloTag tagging. Constructs for MED1-HaloTag, HP1α-HaloTag, SRSF2-HaloTag, FIB1-HaloTag, and IR-HaloTag were generated by replacing the mEGFP with HaloTag in the repair templates using NEBuilder HiFi DNA Assembly Master Mix.

To generate cells for doxycycline-inducible expression of GFP, HaloTag, HaloTag-Ser/Cys, or AviTag-HaloTag-Cys/BirA-SNAP, a PiggyBac vector[28] was used to make the GFP, HaloTag or SNAP-tag containing construct.

PiggyBac vectors for doxycycline-inducible expression of GFP, HaloTag or SNAP. Sequences of SiriusGFP, HaloTag, or SNAP-tag were cloned using NEBuilder HiFi DNA Assembly Master Mix into a doxycycline-inducible, PiggyBac vector, which was described in our previous publication[28].

PiggyBac vectors for doxycycline-inducible expression of HaloTag-Ser/Cys arrays. Constructs for doxycycline-inducible HaloTag-Ser and HaloTag-Cys were generated by inserting the coding sequence for HaloTag protein whose C-terminal is fused to a flexible linker (GAPGSAGSAAGGSGA)[111] and to an array containing either 5 serines or 5 cysteines which are separated by a rigid linker (AEAAAKEAAAKA)[110] into a PiggyBac vector. Constructs were made using NEBuilder HiFi DNA Assembly Master Mix.

PiggyBac vectors for doxycycline inducible expression of AviTag-HaloTag-Cys and BirA. Constructs for doxycycline-inducible HaloTag-Ser and HaloTag-Cys were generated by inserting coding sequences for the constructs into PiggyBac vectors. The AviTag-Halo-Cys construct encodes the coding sequence for HaloTag-Cys construct described above with the AviTag peptide (GLNDIFEAQKIEWHE) with FLAG and HA tags all separated by flexible linkers (PGGSG) fused to the N-terminus. The BirA construct encodes a coding sequence for a human codon-optimized version of BirA with a C-terminal flexible linker (GAPGSAGSAAGGSGA) followed by a SNAP-tag and HA-tag. Constructs were made using NEBuilder HiFi DNA Assembly Master Mix.

Cell editing

Selection criteria for the studied endogenous proteins. We chose for study a plasma membrane receptor (insulin receptor, IR), a transcriptional cofactor (Mediator subunit MED1), a regulator of heterochromatin (heterochromatin protein HP1α), a component of the nucleolus (fibrillarin, FIB1) and a subunit of the mRNA splicing apparatus (serine and arginine-rich splicing factor 2, SRSF2) for multiple reasons. These proteins are well-studied and important regulators of diverse processes in cells (signaling, gene expression, gene silencing, rRNA biogenesis and splicing, respectively). The biological processes associated with these proteins have been shown to be dysregulated in prevalent syndromes. The expression level of these proteins in the liver is similar between healthy donors and patients with type 2 diabetes[112]. Previous studies have

shown that the endogenous proteins can be successfully tagged with fluorescent probes[28,44,103,111,113]. When labeled with florescent probes, they retained their ability to concentrate in the proper locations in cells[28,44,103,111,113]. All these proteins have been reported to assemble into condensate compartments together with other biomolecules with shared functions[28,44,49,53,113].

Endogenously tagged cell line generation. The IR-GFP cell line used here was generated in our previous study[28]. A CRISPR/Cas9 system is used to generate genetically modified HepG2 cell lines as previously performed [28]. 1×10^6 cells were transfected with 1.5 µg sgRNA-Cas9 plasmid and 1.5 µg of homology repair template using Lipofectamine 3000 (Invitrogen, L3000). 24 hours post-transfection, transfection media was replaced with fresh cell culture media (EMEM-FBS). To enrich for transfected cells, cells were sorted 72 hours after transfection based on the expression of mCherry fluorescent protein encoded from the sgRNA-Cas9 plasmid. mCherry positive cells were expanded for 1.5 to 2 weeks before a second sorting for the expression of GFP or HaloTag. To sort based on HaloTag expression, cells were cultured for 15 minutes with Janelia Fluor 585 (a gift of the Lavis Laboratory) prior to sorting. Cells were then expanded and the cell lines were validated by Western blot, PCR genotyping using Phusion polymerase (Thermo Fisher Scientific, F531S) and imaging experiments.

To generate the clonal cell line used in Figure 4H,I, after the second sort, single cells were plated into individual wells of a 96-well plate. The single cells were cultured for 1–1.5 months in conditioned media. To make conditioned media, HepG2 cells were first cultured in fresh media (EMEM-FBS) for 3 days and this media was subsequently harvested. Conditioned media was then made by mixing the harvested media 1:1 with fresh media and filter-sterilizing prior to use.

Genotyping PCR was performed according to the manufacturer's specifications, using the following primers:

IR_fwd: GGAGAATGTGCCCCTGGAC
IR_rev: TTGGTAACCAAACGAGTCCACCT

Doxycycline inducible expression cell line generation. A PiggyBac transposon system (Systems Biosciences) was used for stable integration. 1×10^6 wildtype HepG2 cells were plated in a 6-well plate and simultaneously transfected with 0.5 µg of the PiggyBac expression vector and 0.2 µg of a plasmid encoding PiggyBac transposase (gift of Jaenisch lab) using Lipofectamine 3000 (Invitrogen, L3000). 24 hours post-transfection, transfection media was replaced with fresh media, EMEM with 10% FBS. 72 hours post-transfection, the cells were treated with media containing 150 µg/mL hygromycin (Thermo Fisher Scientific, 10687-010) (for cells edited to express HaloTag-Cys, HaloTag-Ser, and AvTag-Halo-Cys) or 2 µg/mL puromycin (Millipore, P4512-1MLX10) (for cells edited to express BirA-SNAP). Selection media was refreshed every 3 days and un-transfected cells were also treated with hygromycin as a positive control, confirming the efficiency of selection. Typically, 7-10 days were required for the

hygromycin to kill all the non-transfected HepG2 cells. For cells with doxycycline-inducible co-expression of AviTag-Halo-Cys and BirA-SNAP, additional sorting was applied to get cells with low expression of both proteins and minimized cell-to-cell expression variability when performing doxycycline induction: cells were treated with 10 ng/mL doxycycline overnight, followed by co-staining with 50 nM of Halo-JF549 and 50 nM SNAP-JF646 for 20 minutes and proceeding to FACS sorting.

Cell viability

Cell viability was measured by mixing 1:1 TrypanBlue (Invitrogen, T10282) with single cell suspension, then 10 µl of the TrypanBlue/cell mixture was loaded into Countess cell counting chamber slides (Invitrogen, 100078809) and viability was measured using the Countess 3 FL (Invitrogen). All samples were prepared in biological triplicate.

Cell treatments for HepG2

Insulin treatment. The cell plating and insulin treatment regime used in this study is the same as the one previously published [28]. Cells were seeded at a density of 32,000 cells/cm^2 onto 35-mm glass bottom dishes (MatTek Corporation, P35G-1.5-20-C). Starting the day after plating, cells were serum-starved for two days by washing the plates twice with EMEM media without FBS (EMEM) and maintaining the cells in EMEM for 48 hours. Then cells were treated with EMEM supplemented with 0.125% fatty acid-free BSA (Sigma-Aldrich, A8806-5G) ("EMEM-BSA") that contained either 1) 0.1 nM insulin (Sigma-Aldrich, I9278-5ML) or 2) 3nM insulin, which are the concentrations of insulin in the portal vein of healthy and insulin resistant patients [46]. The media was refreshed twice per day (every ~12 hours) for 3 days. This treatment regime produced either a baseline "normal" signaling state or a "pathogenic" elevated signaling state [28]. To ensure that the protein mobility was due to the cellular state and not due to differences in the concentration of insulin, insulin wash-outs were performed. Insulin wash-outs were performed by extensively washing cells with EMEM: the cells were washed six times each with 2mL of EMEM, including three quick washes, two 5 min washes, and a 15 min wash at 37 °C. Cells were then acutely treated with 3 nM insulin for 5 minutes in EMEM-BSA at 37 °C with 5% CO2 in a humidified incubator and then subjected to the desired assay.

H_2O_2 treatment. For experiments in Figure 3E,F and S2C,D, cells were seeded at a density of 32,000 cells/cm^2, serum-starved for two days as described above. Then cells were treated with EMEM-BSA containing 0.1 nM insulin for three days refreshing the media containing insulin twice per day (every ~12 hours). Following a quick wash in EMEM, cells were treated with 0 mM or 7.5 mM H_2O_2 (Sigma-Aldrich, H1009) in EMEM for 5 minutes. This treatment regime was selected because it lead to a similar degree of oxidative stress as pathogenic insulin (see Figure 3B,C and "Metabolomics for quantification of GSSG and GSH ratio" below) and minimizes potential indirect effects of extended H_2O_2 treatment. Cells were then subjected to the desired assay. For the H_2O_2

titration experiment in Figure 5, cells were seeded at a density of 100,000 cells/cm^2, serum-starved for two days, followed by treating with 0, 1, 3, 8, or 20 mM H_2O_2 in EMEM for 5 minutes before proceeding to imaging. For Figure S6A, cells were plated at a density of 56,000 cells/cm^2. When cells reached 80-90% confluency, cells were washed with EMEM once for 30 minutes before treating the cells with 0, 0.1, 1 or 20 mM H_2O_2 for 5-10 minutes. Cells were then collected for western blot.

N-acetyl cysteine (NAC) treatment. Cells were seeded at a density of 32,000 cells/cm^2, serum-starved for two days as described above. Following serum starvation, the cells were treated with 1) EMEM-BSA containing 3 nM insulin for two days and then with 2) EMEM-BSA containing 3 nM insulin and 1 mM NAC (Sigma-Aldrich, A9165-25G) for one day, refreshing the media twice per day (every ~12 hours). We treated the cells with 1mM NAC for 24 hours, because it is reported as a clinically relevant concentration[114] and treating HepG2 cells with 1mM NAC for 24 hours partially restores insulin signaling[28]. Insulin washouts and final stimulation was performed as described above. For NAC treatments of cells expressing IR-Y1361C mutant protein, cells were seeded at a density of 32,000 cells/cm^2 in 35mm glass-bottom dishes, serum-starved for 16 hours and treated with EMEM-BSA containing 0.1 nM insulin and 1 mM NAC for two days refreshing the media twice per day (every ~12 hours). Insulin washouts and final stimulation was performed as described above and cells were then subjected to the desired assay.

NEM with H_2O_2 treatment. Cells were seeded at a density of 55,000 cells/cm^2 in 35 mm glass-bottom imaging dishes (Mattek, P35G-1.5-20-C) and the following day were washed once with EMEM and then serum-starved in 2 ml of EMEM for 24 hours as described above. 100 mM N-ethyl maleimide (NEM) (Thermo Fisher Scientific, 156100500) stock solution was freshly prepared in sterile water prior to experiments. A final concentration of 0 or 10 μM NEM in EMEM was added to cells and incubated at 37°C, 5% CO_2 for 10 minutes, then the cells were treated with H_2O_2 (Sigma-Aldrich, H1009) to a final concentration of 7.5 mM and imaged immediately. Imaging did not proceed longer than 10 minutes to limit secondary effects from extended hydrogen peroxide treatment.

High glucose treatment. Cells were seeded at a density of 100,000 cells/cm^2, and then serum-starved for 16 hours. Cells were then cultured in media containing high glucose concentrations (EMEM supplemented with 33 mM of glucose, Sigma-Aldrich, G8270) or in media containing physiological concentrations of glucose (EMEM supplemented with 33 mM of mannitol, Sigma-Aldrich, M1902) for 12 hours. EMEM media provides physiological concentrations of glucose, and mannitol is used to ensure cells are under similar osmolarity conditions as the high glucose condition. Cells were then subjected to the desired assay.

High fat treatment. Cells were seeded at a density of 32,000 cells/cm^2, and then serum-starved for two days as described above. Cells were then cultured for two days with either EMEM supplemented with fatty acids and BSA (EMEM-HF) or with EMEM-BSA as a control. For EMEM-HF, a 50x stock solution is first made by supplementing EMEM

with the following components to the indicated concentrations: 6.25% fatty acid-free BSA (Sigma-Aldrich, A8806-5G), 2.25 mM oleic acid (Sigma-Aldrich, O7501), and 1.5 mM palmitic acid (Sigma-Aldrich, P9767). This mixture was then incubated at 37°C for 1 hour with constant shaking in a thermomixer. The stock solution was then diluted 1:49 for use in experiments resulting in a final concentration of 0.125% BSA, 45 mM oleic acid and 30 mM palmitic acid for cell treatments. Media was refreshed twice a day (every ~12 hours). Cells were then subjected to the desired assay.

Tumor necrosis factor alpha (TNFα) treatments. Cells were seeded at a density of 32,000 cells/cm^2, and then serum-starved for two days as described above. Cells were then treated with EMEM-BSA media with/without 10 ng/ml Human TNF-α Recombinant Protein (Thermo Fisher Scientific, PHC3016) for two days, refreshing the media twice per day (every ~12 hours). The cells were then subjected to the desired assay.

Etoposide treatment. Cells were seeded at a density of 100,000 cells/cm^2 for etoposide treatment and 10,000 cells/cm^2 for DMSO control. The differences in seeding densities were required to ensure both sets of cells eventually reached similar levels of confluency, as etoposide blocks cell proliferation. Cells were treated with EMEM-FBS media containing 1) 1.5μM etoposide (Thermo Scientific Chemicals, J63651, reconstituted in DMSO) or 2) the same volume of DMSO (Sigma-Aldrich, D2438) as a DMSO control. After 3 days, cells were treated again with 1.5μM etoposide or DMSO for 3 more days. Cells were then subjected to the desired assay.

Lipopolysaccharide (LPS) treatment. Cells were seeded at a density of 100,000 cells/cm^2, and then serum-starved for 16 hours. Cells were then cultured in EMEM containing 1 μg/ml of LPS (Sigma-Aldrich, L2630) for 24 hours. Cells were then subjected to the desired assay.

N-acetyl-p-benzoquinone imine (NAPQI) treatment. Cells were seeded at a density of 32,000 cells/cm^2, and then serum-starved for 2 days. Cells were then treated with EMEM media containing 150 mM NAPQI (Sigma-Aldrich, A7300-1mg) in DMSO or with EMEM media containing DMSO as a control for 15 minutes. Cells were then subjected to the desired assay.

BirA/Avi Assay. Cells were treated with the insulin treatment described above. For each treatment with insulin besides the acute stimulation, 1ng/ml doxycycline was added. 10 μM biotin (Millipore, B4501) was added to the acute insulin treatment and cells were incubated at 37°C 5% CO_2 for 5 minutes and subjected to the desired assay.

Cell treatments for C2C12

Insulin treatment. Cells were seeded at a density of 100,000 cells/cm^2 onto 35-mm glass bottom dishes (MatTek Corporation, P35G-1.5-20-C) for imaging. Starting the day after plating, cells were serum-starved for two days by washing the plates twice with DMEM media without FBS (DMEM) and maintaining the cells in DMEM for 48 hours.

Then cells were treated with DMEM supplemented with 0.125% fatty acid-free BSA (Sigma-Aldrich, A8806-5G) ("DMEM-BSA") that contained either 1) 0.1 nM insulin (Sigma-Aldrich, I9278-5ML) or 2) 3nM insulin. The media containing insulin was refreshed twice per day (every ~12 hours) for two days. This treatment regime produced either a baseline "normal" signaling state or a "pathogenic" elevated signaling state [28]. Right before imaging, insulin wash-outs were performed as follows: in total six washes with 2mL of DMEM each, including three quick washes, two 5 min washes, and a 15 min wash at 37 °C. Cells were then acutely treated with 3 nM insulin for 5 minutes in DMEM-BSA at 37 °C with 5% CO2 in a humidified incubator and then subjected to the desired assay.

High glucose treatment. Cells were seeded at a density of 100,000 cells/cm^2, and then serum-starved for 16 hours. Cells were then cultured in media containing high glucose concentrations (DMEM supplemented with 33 mM of glucose, Sigma-Aldrich, G8270) or in media containing physiological concentrations of glucose (DMEM supplemented with 33mM of mannitol, Sigma-Aldrich, M1902) for 12 hours. DMEM media provides physiological concentrations of glucose, and mannitol is used to ensure cells are under similar osmolarity conditions as the high glucose condition. Cells were then subjected to the desired assay.

High fat treatment. Cells were seeded at a density of 100,000 cells/cm^2. After one day, cells were then cultured for one day with either DMEM-FBS supplemented with 200 μM palmitic acid (Sigma-Aldrich, P9767) (DMEM-HF) or with DMEM-FBS supplemented with 200 μM BSA as a control. For DMEM-HF, a 50x stock solution was first made by supplementing DMEM with the following components to the indicated concentrations: 6.25% fatty acid-free BSA (Sigma-Aldrich, A8806-5G) and 10 μM palmitic acid (Sigma-Aldrich, P9767). This mixture was then incubated at 37°C for 1 hour with constant shaking in a thermomixer. The stock solution was then diluted 1:49 for use in experiments resulting in a final concentration of 0.125% BSA, 200 μM palmitic acid for cell treatments. Cells were then subjected to the desired assay.

Tumor necrosis factor alpha (TNFα) treatment. Cells were seeded at a density of 100,000 cells/cm^2, and then serum-starved for two days as described above. Cells were then treated with DMEM-BSA media with/without 10 ng/ml Human TNF-α Recombinant Protein (Thermo Fisher Scientific, PHC3016) for two days, refreshing the media twice per day (every ~12 hours). The cells were then subjected to the desired assay.

Etoposide treatment. Cells were seeded at a density of 100,000 cells/cm^2 for etoposide treatment and 1,000 cells/cm^2 for DMSO control. The differences in seeding densities were required to ensure both sets of cells eventually reached similar levels of confluency, as etoposide blocks cell proliferation. Cells were treated with DMEM-FBS media containing 1) 1.5μM etoposide (Thermo Scientific Chemicals, J63651, reconstituted as 10mM in DMSO) or 2) the same volume of DMSO (Sigma-Aldrich, D2438) as a DMSO control. After 3 days, cells were treated again with 1.5μM etoposide or DMSO control for 3 more days. Cells were then subjected to the desired assay.

Lipopolysaccharide (LPS) treatment. Cells were seeded at a density of 100,000 cells/cm^2, and then serum-starved for 16 hours. Cells were then cultured in DMEM containing 1 μg/ml of LPS (Sigma-Aldrich, L2630) for 24 hours. Cells were then subjected to the desired assay.

N-acetyl-p-benzoquinone imine (NAPQI) treatment. Cells were seeded at a density of 100,000 cells/cm^2, and then serum-starved for 2 days. Cells were then treated with DMEM media containing 150 μM NAPQI (Sigma-Aldrich, A7300-1mg) in DMSO or DMSO as a control for 15 minutes. Cells were then subjected to the desired assay.

Live-cell imaging experiments: general setup

General imaging conditions. Cells were plated on 35-mm glass bottom dishes (MatTek Corporation, P35G-1.5-20-C). For imaging doxycycline inducible proteins, 0.1 μg/ml of doxycycline was added to the media 8-12 hours prior to imaging. Cells were imaged for no longer than 10 minutes inside an incubation chamber supplemented with warmed (37°C) humidified air and with 5% CO_2. For exogenous GFP or HaloTag visualization, the cells were also stained with 1:20,000 of Hoechst 33342 (Thermo Scientific, 62249) for 10 minutes to stain the nuclei before imaging.

Live-cell super-resolution microscopy. ZEISS LSM 980 with Airyscan 2 was used to acquire the super-resolution images of GFP-/HaloTag-tagged proteins. Cells were imaged with a ZEISS incubation system that stably maintained the samples at 37°C with 5% CO_2 and humidified air. A 63X objective with oil immersion was used. For GFP-tagged proteins, 488 nm laser was used for excitation. For Halo-tagged proteins, cells were first incubated with media containing 100 nM Janelia Fluor 549 (Halo-JF549, Promega, GA1110) for 15 minutes. The cells were then washed with fresh media and then cultured at 37°C with 5% CO_2 for 10 minutes before imaging with 561 nm laser excitation and the mCherry filter for emission. Due to the various expression levels of different proteins, the laser power was adjusted for each protein, such that the brightest pixels remained below the saturated levels of the detection range (maximum brightness = 255 for 8-bit images). Following raw image acquisition, Airyscan super-resolution processing was performed via ZEN Blue.

Fluorescence recovery after photobleaching (FRAP). ZEISS LSM 980 with 63X objective, oil immersion was used to perform FRAP experiments on GFP-tagged proteins in live cells. The acquisition mode, laser power, time interval between frames, total number of frames, and other FRAP-specific settings were customized for each protein of interest (POI) such that each experiment would satisfy four criteria: (1) have sufficient signal, (2) have sufficient duration to capture recovery, (3) have sufficient temporal resolution, (4) endure minimal photobleaching throughout the time course. Detailed configurations for different protein targets are summarized in Table S5. For each single FRAP acquisition course, several frames were first recorded to establish pre-bleach levels of signal, followed by photobleaching with 100% laser power of a 2μm-by-2μm square region. For insulin receptor, this square region contained a portion of the plasma membrane. For MED1, HP1α, and SRSF2, this square region contained

an area of the relevant punctate high signal (a condensate). For FIB1, this square region overlapped, but did not completely cover the nucleolus. For GFP, this square region was either randomly sampled within the nucleus or randomly sampled within the cytoplasm. The number of bleaching cycles is reported in Table S5. After photobleaching, fluorescence recovery was recorded over time. Raw image series were processed via ZEN Blue (2D Airyscan processing), followed by drift correction using a cross-correlation algorithm. Averaged intensity measurements from an unbleached region were further used to correct for the photobleaching occurring during the image acquisition.

Statistical analysis was performed with the Statistics and Machine Learning Toolbox of MATLAB. A two-tailed student's *t*-test was used to generate p-values comparing timepoints at the later end of recovery curves, at which point recovery intensities had stabilized. The number of timepoints (n) for each comparison was ~6-13% of the total number of timepoints collected in the recovery curve (the number of time points considered is specified in Table S5). All FRAP experiments were performed twice using a total of 4 biologically independent samples. Each dish was imaged for no more than ten minutes to minimize secondary effects of extended treatment.

Single particle tracking (SPT). ZEISS Elyra 7 with 63X objective, oil immersion was used to perform SPT experiments on Halo-tagged proteins in live cells. Cells were co-stained with two Halo dyes: one used for tracking individual molecules of a protein and the other for visualizing the bulk distribution of the protein. After staining, cells were washed by incubating with fresh media without dyes for at least 10 minutes. HILO illumination was used during the tracking. The detailed sample preparations and configurations of SPT for different proteins are summarized in Table S5. For the tracking of AviTag-Halo-Cys and BirA-SNAP, cells were first incubated with media containing 0.1 nM Janelia Fluor 549-Halo and 0.1 nM Janelia Fluor 646-SNAP (a generous gift from Luke Lavis Lab at Janelia Farm Research Campus) for 20 minutes and molecules were tracked at 100 Hz acquisition rate. There are four major steps in the SPT analysis to obtain single molecule trajectories: pixelwise peak detection, subpixel localization of the peaks, reconnection of the peaks (to construct trajectories) and validation of trajectories. For the first three steps, point spread functions (PSFs) from single molecules were detected, subpixel-localized, and reconnected with custom code in MATLAB based on the published multiple-target tracing (MTT) tool [115]. During the pixelwise peak detection step, for each pixel, two hypotheses H_0 and H_1 were compared based on a generalized likelihood ratio test, where H_0 defines the non-presence of particles and H_1 the presence of a particle at the center of the pixel. Valid peaks were identified with a constant false alarm rate ($\approx 1.5\times10^{-6}$). Additional peaks were identified with a B-spline wavelet filter [116]. The subpixel localization of the peaks was performed by maximizing the likelihood of the PSF to match the local intensity distribution of a 7x7 pixel area using Gauss-Newton regression. The reconnection of the peaks to construct trajectories was performed based on the multiple-target reconnection as described in MTT [115], with the prior maximum diffusion coefficients, the disappearance probability for blinking and maximum number of disappearance frames summarized in Table S5. A set of reconnected peaks comprises a trajectory. Validation

of trajectories for those occurring in the relevant biological compartments was performed using bulk distribution or nuclear stain as a reference. For insulin receptors, the plasma membranes were manually selected by drawing polygons via MATLAB; for other proteins of interest, nuclei or nucleoli regions were labeled by a deep learning based algorithm Cellpose[104].

Live-cell imaging experiments: analysis and additional validation

Inferring mobility from FRAP courses. With FRAP, we investigated protein mobility in cell models of disease at the length and time scales of 1-2 microns and (tens of) seconds, with the additional question of whether protein mobility was altered both inside and outside of condensates. For the proteins that could be reliably assigned to be within or outside of condensates during image acquisition (MED1, HP1α, FIB1 and SRSF2), a square region containing both the condensate(s) and the surrounding dilute phase was imaged, allowing separate FRAP analyses both inside and outside of the condensates. The areas inside the condensates (condensed region) and outside the condensates (dilute regions) were differentiated by custom code in MATLAB by fitting the cumulative distribution of pixel-wise intensities of each FRAP region to a two-step function, with the first step identifying the low-intensity pixels in the nucleus (the dilute region) and the second step by identifying the high-intensity pixels with enriched fluorescent signal in the nucleus (the condensed region). For FRAP outside the condensates, the dilute region was selected for analysis, and for FRAP inside the condensates, the condensed phase was selected for analysis. For insulin receptor (IR), due to the rapid movement of IR condensates, we could not perform reliable assignments of the condensate-occupied pixels throughout a FRAP acquisition course. Instead, the fluorescence signal analysis was limited to the plasma membrane (where most signals are located), which was manually selected in ZEN Blue. For GFP alone, the pixels inside the entire 2μm-by-2μm photobleached region were selected for fluorescence signal analysis as GFP alone does not form condensates.

For each frame of each FRAP course, the average intensity of selected pixels was calculated. The average intensities from different frames were further normalized through the following linear transformation: the averaged intensity of pre-photobleaching frames was set to 1, while the intensity right after photobleaching was set to 0. This resulted in a FRAP curve for each independent photobleaching experiment. A single average FRAP curve from all replicate samples was obtained by plotting the mean normalized pixel intensity and SEM for each timepoint. To evaluate the difference of protein mobility between any two conditions, the maximum extent of recovery at the end of the recorded time window was compared. For this comparison, a two-tailed student's t-test was used to calculate statistical significance. Data points used for the comparison were chosen by selecting time points where signal recovery was approaching an apparent plateau (listed in Table S5) and aggregating those signal intensities.

Statistical analysis was performed with the Statistics and Machine Learning Toolbox of MATLAB. A two-tailed student's *t*-test was used to generate p-values comparing timepoints at the later end of recovery curves, at which point recovery intensities had stabilized. The number of timepoints (n) for each comparison was ~6-13% of the total number of timepoints collected in the recovery curve (the number of

time points considered is specified in Table S5). All FRAP experiments were performed twice using a total of 4 biologically independent samples. Each dish was imaged for no more than ten minutes to minimize secondary effects of extended treatment.

Inferring mobility from 2D SPT trajectories. We investigated protein mobility in cell models of disease at the length and time scales of (sub)micron and (sub)second, and focused on the question of whether apparent diffusion rates are reduced under pathogenic conditions. We used a wide-field microscope to measure planar (XY) movement at a desired, feasible length/time scale. The use of 2D projections to measure apparent diffusion coefficient (*D*) has been a widely used approach[117-120], given the assumption that the molecules diffuse isotropically along the three-dimensional axes X, Y and Z. This assumption is supported by the observation that a similar apparent diffusion coefficient was observed when Sox2–a nuclear protein–was tracked in 3D (D≈2.5μm^2/s[121]) and 2D projection in XY (D≈2.8μm^2/s[118]). To provide a scalar measurement of the molecular movement we infer from a given SPT trajectory, we calculate an apparent diffusion coefficient for each trajectory. This value *D* is derived from the relationship between the mean square displacement (MSD) versus timelag (τ). Only trajectories with at least 5 reconnected peaks were selected. For trajectories with more than 20 reconnected peaks, only the first 20 peaks were used for estimating the *D*. A linear regression between MSD and timelag with an additional zero-order term (localization error due to limited spatial resolution) was used to fit the apparent diffusion coefficient in two dimensions: $\mathrm{MSD}(\tau) = 4D\tau + \sigma^2$. At this point, we obtained the fitted *D* for each trajectory, which was a reliable estimation only if 1) the *D* was above the effective magnitude caused by localization error, and 2) the fitting noise was relatively low. To filter for reliable *D* above the localization uncertainty, molecules with D≥0.01 μm^2/s for endogenous proteins and D≥0.1 μm^2/s for exogenous HaloTag were selected for the analysis (drop-off rate ≃ 10-30%), To filter for reliable *D* with low fitting noise, molecules whose fitting residual was below the MSD of one-frame timelag were selected for the analysis (drop-off rate ≃ 5%-10%). These two filters were sequentially applied to obtain the final well-fitted trajectories. Mann-Whitney test was used to evaluate the statistical significance between the diffusion coefficients in different conditions. For HepG2 cells, all SPT experiments were performed in 4 biologically independent samples, for C2C12 cells, all SPT experiments were performed in 2 biologically independent samples.

In this work, we are using mobility to describe the transit of individual molecules or ensembles of molecules through space over a given unit time. Such movement is likely to be the net effect of diverse forces within the cell. In this work, the duration of tracking mostly ranges from 0.02s-0.1s, with a temporal resolution of 4ms-10ms, and spatial resolution of 30nm-80nm.

Evaluation of the comparability between FRAP and SPT – general setup. There are two biophysical parameters inferred from both FRAP and SPT that can be used to address how comparable the values generated by these two methods are: (1) fraction of immobile molecules; (2) apparent diffusion coefficient of mobile molecules. For (1), we used five endogenously tagged proteins (IR, MED1, HP1α, FIB1, and SRSF2) which are known to have a “bound state” and thus a putative immobile fraction of molecules;

For (2), we used exogenously expressed GFP vs HaloTag with the assumption that these proteins demonstrate homogeneous apparent diffusion coefficients throughout the nucleoplasm.

Evaluation via the fraction of immobile molecules. We used SpotOn[118] to evaluate the fraction of immobile molecules by SPT, which requires two hyperparameters: 1) the total number of diffusivity states and 2) the axial detection range. An estimate of the total number of diffusivity states is achieved by identifying the minimum number of Gaussian functions needed to sufficiently fit the log*D* distribution of individual molecules (Figure S2A)[120]. The fitting residual was plotted as a function of the number of Gaussian functions (*N*) tested to perform the fitting, and the minimum number of Gaussian functions needed was the inflection point of the residual-N relation. We found MED1 can be well fitted by 2 states, and other targets are well fitted by three states. An estimate of the axial detection range is achieved by examining a z-stack scan of fixed cells with sparsely labeled PAJF549 molecules to establish a limit of expectations for focal depth within which a single molecule can be consecutively tracked. We concluded that the focal depth (dz) peaked at ~900nm for our specific setup (Figure S2B). These values were used in SpotOn to estimate of the fraction of immobile molecules by SPT (Figure S2C).

For FRAP datasets, we fitted the normalized recovery curve to the following equation:

$$I(t) = f \cdot (1 - 2^{-t/\tau}),$$

Where $1 - f$ would be the immobile fraction. The fraction of immobile molecules estimated via two methods are indeed comparable (Figure S2C): FIB1 shows the highest immobile fraction among the five proteins tested; IR, HP1α, and SRSF2 showed relatively lower immobile fractions in both SPT and FRAP. The immobile fractions estimated by FRAP were slightly lower compared to those estimated by SPT. Given that SPT can capture stable immobile events, and transient immobile events in a timescale as short as ~10^{-2} s, while FRAP is only sensitive to intermediate/long-term immobile events in a timescale of ~10^{1} s, this could explain why the immobile fractions estimated by FRAP are consistently lower.

Evaluation via the apparent diffusion coefficient of mobile molecules. The HaloTag tagged with a JF646 molecule (~34kDa in total) and GFP (~27kDa) have comparable protein size; we thus expect that the intrinsic diffusion coefficients of these two proteins should be similar. For SPT, we estimated the apparent diffusion coefficient of HaloTag by calculating the average apparent diffusion coefficient of mobile molecules. For FRAP, we estimated the apparent diffusion coefficient of GFP by matching the experimental FRAP data to a theoretical model of the diffusion process within a photobleached area. We modeled the theoretical diffusion process of a photobleached region ($l \times l$) as the following partial derivative equation (PDE) problem:

$$\frac{\partial u}{\partial t} = D\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right), \qquad 0 < x < l, 0 < y < l, t > 0,$$
$$u\rfloor_{x=0} = 0,\ u\rfloor_{x=l} = 0, \qquad 0 \le y \le l, t \ge 0,$$
$$u\rfloor_{y=0} = 0,\ u\rfloor_{y=l} = 0, \qquad 0 \le x \le l, t \ge 0,$$
$$u\rfloor_{t=0} = 1 - \phi(x, y), \quad 0 \le x \le l, 0 \le y \le l.$$

In this PDE problem, $u(x,y,t)$ is the normalized density of photobleached molecules of a certain pixel (x,y) at certain time t. $x,y=0\ or\ l$ are boundaries. $\phi(x,y)$ is the normalized pixelwise intensity (i.e., normalized density of intact molecules) right after photobleach ($t=0$), thus $1-\phi$ is the density of photobleached molecules at $t=0$ (the total normalized intact molecules plus photobleached molecules always equals 1 within each pixel). In the spirit of separation of variables, one general analytical solution that satisfied the boundary conditions was derived as:

$$u(x,y,t)=\sum_{n,m=1}^{\infty} A_{nm}\times\sin\frac{n\pi}{l}x\times\sin\frac{m\pi}{l}y\times\exp\left\{-\left[\left(\frac{n\pi}{l}\right)^2+\left(\frac{m\pi}{l}\right)^2\right]Dt\right\},$$

Where the coefficients of different modes were computed as:

$$A_{nm}=\frac{4}{l^2}\iint_0^l\left(1-\phi(x,y)\right)\times\sin\frac{n\pi}{l}x\times\sin\frac{m\pi}{l}y\,\mathrm{d}x\mathrm{d}y.$$

The D is determined when the 2-norm residual of the $u(x,y,t)$ throughout the whole FRAP process between the experimental measurement and the theoretical prediction are minimized. The calculations were done by custom MATLAB code. The diffusion coefficient of HaloTag estimated from SPT is 5.3±0.2 μm^2/s, and the diffusion coefficient of GFP estimated from FRAP is ~6 μm^2/s (Figure S2D). Given that the relation between the molecular weight (M) and diffusion coefficient is $D\sim \mathrm{M}^{-0.33}$ [122], the diffusion coefficient of a "weighted GFP" (~34kDa) estimated from FRAP would have been ~5.5 μm^2/s. Therefore, the apparent diffusion coefficient estimated via two methods are comparable indeed.

Estimation of the localization uncertainty. To experimentally derive a lower bound for apparent diffusion constants that indicate molecules that are moving, we examined a fixed sample of cells with sparsely labeled HaloTag-PAJF549 molecules. As these molecules are fixed, diffusion constants derived would theoretically represent "no movement". The distribution of apparent diffusion coefficient fitted from individual molecules in fixed sample centered around D~0.01 μm^2/s (Figure S2E), and this pseudo diffusivity is due to the localization uncertainty of single molecules at each frame. Therefore, we used 0.01 μm^2/s as the lower bound cutoff when filtering for truly mobile molecules for endogenous protein targets based on the limitation of the localization uncertainty. As for exogenously expressed HaloTag alone, because most molecules are diffusive, we can apply a higher cutoff (D>0.1 μm^2/s) to select mobile molecules in order to eliminate any false positive mobile molecules without increasing the chance of false negative elimination (Figure S2F).

Validation of reconnecting during trajectory reconstruction. There are two main sources of error when reconstructing a trajectory from localizations: too stringent prior maximum allowed diffusion coefficient (D_{max}), or too great of localization density (Fig S2G-I). If D_{max} is smaller than the typical diffusivity of the protein of interest, it will result in an early stop of reconnecting peaks of signal of the same molecule (Figure S2G), and the estimated apparent diffusion coefficient will hit a ceiling set by D_{max} (the estimated apparent diffusion coefficient is artificially low). To prevent this issue, the D_{max} we chose when reconnecting sequential peaks (Table S5) is much larger than the apparent diffusion coefficients estimated from the final trajectories. It should be noted that for all

D_{max}, the tracking process will still stop at some point, mainly due to either the photobleach of the dye molecule, or the molecule moving out of the focus (Figure S2H). If the localization density per frame is too great, there will be an increased likelihood that the trajectories of two different molecules form an “ambiguous connection”. In this situation, we expect to see an increase in the average number of jumps (connection between two consecutive peaks) per trajectory (Figure S2I). To determine a threshold density that would minimize ambiguous connections in our experiments, we generated IR-HaloTag SPT data with different peak density per frame, followed by reconnecting the peaks with the prior D_{max}=6 μm^2/s. We found that the average number of jumps per trajectory starts to increase with peak density when the peak density is above 0.01 μm^2/frame, and we ensured peak densities of our experiments were always below this threshold density (Figure S2J). Therefore, we concluded that the prior D_{max} chosen for reconnecting is large enough to capture consecutive jumps for the same molecule, and the peak densities of our actual experiments are low enough to avoid significant “ambiguous connection” given the prior D_{max}.

Estimation of the false positive identifications of trajectories in SPT. The sources of false positive identifications of proteins in SPT include pixel noise, auto-fluorescence, and non-specific dye staining (restricted to PAJF549). The control experiments are summarized in Table S6. The overall rate of false positive identification of trajectories is either ~5% (for JF646 staining) and ~12% (for PAJF549 staining) at maximum. Therefore, we concluded that the SPT dataset of the actual experiments are dominated by trajectories from real proteins of interest.

Quantification of condensate properties

Three condensate properties were evaluated in both normal and pathogenic conditions: (1) number of condensates per cell in the focal plane, (2) condensate size in diameter, and (3) the partition ratio, which can be defined as the relative enrichment of the intensity inside the condensate versus outside the condensate. Live-cell super-resolution images taken by ZEISS LSM 980 with Airyscan 2 were used for such quantifications of GFP-tagged proteins.

The first step is to identify/segment puncta. We implemented two approaches to identify/segment puncta depending on the size, morphology, and distribution of the protein condensates. Insulin receptor (IR), MED1 and FIB1 have relatively small condensates (less than ten pixels in diameter), with a round shape and compact distribution, thus the Laplacian of Gaussian (LoG) Blob Detection (sigma = 200nm, 500nm, and 450nm for IR, MED1 and FIB1, respectively) was applied to the images (MATLAB code source: Jason Klebes, 2024. LoG Blob, GitHub), and puncta were identified with the quality filter set to 0.2. Additional intensity filters and partition ratio filters were applied to call puncta, such that there was high agreement between auto-identified puncta and puncta called by a trained eye. HP1α and SRSF2 have varying condensate sizes (ranging from several pixels to more than ten pixels in diameter), with an ellipse or irregular shape and dispersed distribution, thus the images were

background-subtracted with a median filter (filter size = 2μm), followed by feature segmentation with Cellpose[104]—an AI-based segmentation tool—with the "cytoplasm" model (feature dimension to be recognized = 1μm) to obtain the punctum features.

The second step is to quantify the three condensate properties of the identified puncta. To quantify the number of puncta in each cell, the regions where puncta could be detected per cell were defined based on the GFP signal and the detected puncta in each cell were counted. To measure the condensate size, the full-width half-maximum was used as an estimator of IR, MED1 and FIB1 punctum diameters. To measure the condensate size of HP1α and SRSF2, the area of each punctum feature segmented by Cellpose was covered to an effective diameter with the formula: $d = 2*(area/\pi)^{0.5}$. To quantify the partition ratio[123], the intensity inside the puncta was divided by the intensity of the local dilute phase. The intensity of the local dilute phase for the plasma membrane was used for IR, and the intensity of the local dilute phase of the nucleoplasm was used for the rest of the proteins.

Western Blotting

Protein preparation. HepG2 cells were treated according to the specified treatment protocol, then the media was aspirated off and cells were washed once with ice-cold PBS (Gibco, 10010-023) on ice. The PBS was then removed and Cell Lytic M (Sigma-Aldrich, C2978) supplemented with protease and phosphatase inhibitors (Sigma-Aldrich, 11873580001 and 4906837001) was added to each well to lyse the cells. The cells were scraped with a plastic cell scraper, and the lysates were transferred to a 1.5 ml eppendorf tube and allowed to rotate on a rotator for 15 minutes at 4°C. For proteins that required sheering of DNA to be accurately measured by Western blot (MED1, HP1α, FIB1, SRSF2), the lysates were sonicated in 1.5 ml Eppendorf tubes on ice water (15 seconds on, 20 seconds off, 30% amplitude, for 3 cycles, Fisher Scientific, FB120 Model CL-18) and then centrifuged at 12,000 x g for 15 minutes. The supernatant was transferred to a fresh 1.5 ml tube and the protein concentration was quantified using a BCA Protein Assay Kit (Life Technologies, 23250) according to the manufacturer's instructions.

Preparations of western blot samples. For samples prepared in Figures 6E,F, S1, and S6A (blot with DTT, right side), dithiothreitol (DTT) and XT Sample Buffer 4x (BioRad, 1610791) were added to the purified proteins in reaction buffer or protein lysates to final concentrations of 100mM and 1x, respectively and boiled at 95°C for 5 minutes. For western blot in non-reducing conditions, DTT was not added. For samples in Figure 6K, 2-mercaptoethanol and Native Sample Buffer (BioRad, 1610738) were added to the protein lysate to final concentrations of 2.5% and 1x, respectively, boiled for 5 minutes at 95°C and allowed to cool completely before addition of streptavidin (Invitrogen, 43-430-2) to a final concentration of 10μM to cause a shift in molecular weight of proteins that were biotinylated by BirA.

Running western blot samples. 5-35 µg of proteins were separated on 10% or 4-12% Criterion™ XT Bis-Tris Protein Gel (BioRad, 3450112, 3450125) in XT MOPS running buffer (Bio-Rad Laboratories, 1610788) at 100 V. Proteins were transferred to a 0.45-µm PVDF membrane (Millipore, IPVH00010) in ice-cold transfer buffer (25 mM Tris, 192 mM glycine, 20% methanol) at 300 mA for 2 hours at 4 °C. Membranes were blocked in either 5% nonfat milk (LabScientific, M0842) dissolved in TBST (2% Tris-HCl pH 8.0, 1.3% 5 M NaCl, 0.05% Tween 20) or 5% BSA (VWR, 102643-516) in TBST for 1 hours at room temperature. Membranes were then incubated overnight at 4°C with primary antibodies (list below) diluted in 5% nonfat milk in TBST or 5% BSA in TBST. Membranes were then washed three times in TBST for 5 minutes at room temperature and then incubated with donkey anti-rabbit IgG (Cytiva Life Sciences, NA934-1ML, 1:10,000 dilution) or sheep anti-mouse IgG (Cytiva Life Sciences, NXA931V, 1:10,000 dilution) diluted in 5% nonfat milk in TBST for 1 hours at room temperature. Membranes were washed three times for 10 minutes in TBST. Membranes were developed with ECL substrate (Millipore, WBKL20500) and imaged using a CCD camera (BIO RAD, 1708265). The “analyze gel” tool on Fiji/ImageJ v2.1.0/153c was used to quantify immunoblot signal. A two-tailed student's t-test was used to generate p-values. Statistical analysis was performed using Prism Version 9.4.0 (GraphPad, La Jolla, CA).

Primary antibodies for Western blotting:
Anti-insulin receptor (Cell Signaling, 3025, dilution 1:1000)
Anti-MED1 (Bethyl, A300-793A, dilution 1:1000)
Anti-HP1α (Abcam, ab109028, dilution 1:1000)
Anti-FIB1 (Abcam, ab5821, dilution 1:1000)
Anti-SRSF2 (Thermo Fisher, PA5-12402, dilution 1:1000)
Anti-b-actin (Sigma-Aldrich, A5441, dilution 1:10,000)
Anti-H3 (Cell Signaling, 4499, dilution 1:1000)
Anti-HA (Cell Signaling, 3724, dilution 1:1000)
Anti-pIRS1 (Cell Signaling, 3070, dilution 1:1000)
Anti-IRS1 (Cell Signaling, 2382, dilution 1:1000)
Anti-IRα (Cell Signaling, 74118, dilution 1:1000)

Immunofluorescence

Wildtype HepG2 cells were fixed with 4% paraformaldehyde in PBS for 10 minutes at room temperature (RT), washed three times with PBS for 5 minutes at RT, permeabilized with 0.5% TX100 for 10 minutes at RT, washed with PBS for 5 minutes at RT, blocked with 4% BSA (Jackson Immunoresearch Laboratories - 001-000-162) in PBS for 1 hour at RT. Cells were incubated with primary antibodies diluted 1:500 in 4% BSA in PBS overnight at 4°C. Cells were washed three times with PBS for 5 minutes at RT and incubated with secondary antibodies Goat anti-Rabbit IgG Alexa Fluor 488 (Life Technologies, A11008) diluted 1:500 in 4% BSA in PBS for 1 hour at RT. Cells were then washed in PBS three times for 5 minutes. Nuclei were stained with Hoechst (Thermo Fischer Scientific, 3258) diluted 1:5000 in PBS for 5 minutes at RT and excess Hoechst was removed by washing cells 3 times for 5 minutes with PBS. Cells were stored at 4°C in PBS and imaged using the ZEISS LSM 980 with Airyscan detector

using 63x objective. Raw image series were processed via ZEN Blue (2D Airyscan processing). Images were converted in JPEG format using Fiji/ImageJ v2.1.0/153c.

Primary antibodies for immunofluorescence:
Anti-insulin receptor (Cell Signaling, 23413)
Anti-MED1 (Abcam, ab64965)
Anti-HP1α (Abcam, ab109028)
Anti-FIB1 (Abcam, ab582)
Anti-SRSF2 (Abcam, ab11826)

Metabolomics for quantification of GSSG and GSH ratio

All solvents, including water, were purchased from Fisher and were Optima LC/MS grade.

HepG2 cells were treated according to the specified treatment protocol in 6-well culture plates, then the media was removed, and cells were washed twice with ice-cold PBS (Gibco, 10010-023) on ice. The PBS was then removed and 500 µl of ice-cold 80% methanol (Thermo Fisher Scientific, A456-4)/ 20% LC-MS grade water (Thermo Fisher Scientific, W6-4) solution with isotope-labeled amino acid mass-spec internal standards (Cambridge Isotope Labs, MSK-CAA-1) was added to each well on dry ice. The plate was chilled at -80°C for a minimum of 15 minutes, then the cells were scraped for 30 seconds with a plastic cell scraper (Corning, 3008). The methanol-cell mixture was transferred to a 1.5 ml eppendorf tube (Eppendorf, 0223641). The well was washed again with 300 µl of the ice-cold methanol solution to extract most of the remaining cells from the well, which was added to the same 1.5 ml eppendorf tube. The mixture was vortexed on high for 10 minutes at 4°C, then centrifuged on a table top centrifuge on max speed for 10 minutes at 4°C. 600 µl of supernatant were removed from the tube and transferred to a fresh tube on dry ice. The supernatant was dried for 5 hours at 4°C using a speed vac (Labconco 7310020), then resuspended in 1/10th of the volume of the original supernatant in LC-MS grade water on ice. The resuspended metabolites were vortexed on high for 10 minutes at 4°C, then centrifuged on a table top centrifuge on max speed for 10 minutes at 4°C. The supernatant containing the endogenous metabolites and internal standards were transferred to LC-MS vials and liquid chromatography and mass spec was carried out by the Whitehead Institute Metabolomics Core.

Metabolite profiling was conducted on a QExactive bench top orbitrap mass spectrometer equipped with an Ion Max source and a HESI II probe, which was coupled to a Dionex UltiMate 3000 HPLC system (Thermo Fisher Scientific, San Jose, CA). External mass calibration was performed using the standard calibration mixture every 7 days and an additional custom mass calibration was performed weekly alongside standard mass calibrations to calibrate the lower end of the spectrum (m/z 70-1050 positive mode and m/z 60-900 negative mode) using the standard calibration mixtures spiked with glycine (positive mode) and aspartate (negative mode). Typically, samples were reconstituted in 50 uL water and 2 uL were injected onto a SeQuant® ZIC®-

pHILIC 150 x 2.1 mm analytical column equipped with a 2.1 x 20 mm guard column (both 5 mm particle size; Millipore-Sigma). Buffer A was 20 mM ammonium carbonate, 0.1% ammonium hydroxide; Buffer B was acetonitrile (Thermo Fisher Scientific, A955-4). The column oven and autosampler tray were held at 25°C and 4°C, respectively. The chromatographic gradient was run at a flow rate of 0.150 mL/min as follows: 0-20 min: linear gradient from 80-20% B; 20-20.5 min: linear gradient form 20-80% B; 20.5-28 min: hold at 80% B. The mass spectrometer was operated in full-scan, polarity-switching mode, with the spray voltage set to 3.0 kV, the heated capillary held at 275°C, and the HESI probe held at 350◦C. The sheath gas flow was set to 40 units, the auxiliary gas flow was set to 15 units, and the sweep gas flow was set to 1 unit. MS data acquisition was performed in a range of m/z = 70-1000, with the resolution set at 70,000, the AGC target at 1x106, and the maximum injection time at 20 msec. Relative quantitation of polar metabolites was performed with TraceFinder™ 4.1 (Thermo Fisher Scientific) using a 5 ppm mass tolerance and referencing an in-house library of chemical standards. Data were filtered according to predetermined QC metrics: CV of pools <25%; R of linear dilution series <0.975.

Metabolomics Analysis for quantification of GSSG and GSH ratio

FreeStyle (Thermo Scientific, Version 1.3) was used to check quality, mass shift, and retention time drift for each metabolite. The MS2 spectra for each metabolite was also verified in FreeStyle. TraceFinder (Thermo Fisher Scientific, Version 4.1) was used to call metabolite peaks and determine raw peak areas. The peak detection defaults were as follows - Mass tolerance: 5 ppm, Retention time window: 30 sec, Ion Ratio Window type: relative +/- 20%, Ion coelution (min) 0.100, Detection algorithm: ICIS (Area noise factor: 5, Peak noise factor: 10, Baseline window: 40, Noise method: repetitive, Min peak width: 3, Multiplet resolution: 10, Area tail extension: 5). Each peak was manually verified to have the correct shape, retention time, and m/z.

Peak area ratios were determined by normalizing the raw peak area for each metabolite by the raw peak area of the appropriate internal standard.

During the mass spec run, a pooled sample made from pooling 5 µl of each sample was run 4-6 times as technical replicates to measure the reliability o

f detection for each metabolite—a coefficient of variation < 0.30 was used as a cutoff for metabolites to be measured reliably. Similarly, a dilution series of the pool was also run to determine whether each metabolite was in the linear range of detection—A correlation coefficient R < 0.95 was used as a cutoff.

Two-tailed student's t-test with Welch’s correction was used to generate p-values. Statistical analysis was performed using Prism Version 9.4.0 (GraphPad, La Jolla, CA).

Identification of surface-exposed cysteines of individual proteins

Surface-exposed cysteines were identified for specific, individual proteins based on two criteria: 1) the cysteine residue is located within an intrinsically disordered region (IDR) of the protein[124], as determined by the Predictors of Natural Disordered Regions (PONDR) VSL2 algorithm, or 2) the solvent-accessible area of the cysteine exceeds 20 Å² [125], as measured by STRIDE[126].

Variant annotation

Variants and their genomic coordinates (hg38) were obtained from ClinVar (https://ftp.ncbi.nlm.nih.gov/pub/clinvar/vcf_GRCh38/clinvar_20230903.vcf.gz, https://ftp.ncbi.nlm.nih.gov/pub/clinvar/tab_delimited/variant_summary.txt.gz). Only germline missense variants were considered. We only considered variants with at least one clinical significance annotation as Pathogenic or Benign. The number of missense variants considered in ClinVar is 52,188.

When needed, variants were annotated with impact on protein sequence and other measures of computationally predicted pathogenicity (SIFT, PolyPhen, CADD etc) using Ensembl VEP 110. Gene-level and 1kb-window constraint metrics were obtained from gnomAD v4 and v3, respectively.

For all downstream analyses, variants were counted as protein variants—i.e., DNA variants resulting in the same protein-coding alteration, regardless of their similarity or differences at the DNA level, were counted as the same variant. Variants were mapped to gene, then mapped to proteins using mapping from Uniprot Swiss-Prot (https://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/idmapping/by_organism/HUMAN_9606_idmapping_selected.tab.gz, https://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/idmapping/by_organism/HUMAN_9606_idmapping.dat.gz), using the gene's HGNC ID to Uniprot-KB Accession ID of the canonical isoform.

Physics-based model for how cellular environment influences diffusion of proteins with and without cysteines

Estimation of the average number of surface-exposed cysteine per protein and the concentration of surface-exposed cysteine. We used iCysMod[127] to estimate the number of proteins with surface-exposed cysteines and the average number of surface-exposed cysteines per protein across the whole proteome. There are 18,350 proteins in the proteome, among which we tried two commonly used relative solvent accessibility (RSA) cutoffs to evaluate the cysteine surface exposure. With RSA > 40, the estimated number of proteins with at least one surface-exposed cysteine is 10,333 (56.3% of total proteins) with an estimated average number of surface-exposed cysteines of 2.8. With RSA > 50, the estimated number of proteins with at least one surface-exposed cysteine is 6,754 (36.8% of total proteins) with an estimated number of surface-exposed cysteines of 1.9. The mean estimated number of proteins with at least one surface-exposed cysteine using these two cutoffs is 8,544 (46.6% of total proteins, rounded to 50%) with an estimated number of surface-exposed cysteines of 2.35, rounded to 2.

The total number of protein molecules per cubic micron in the cell is 2-4 million[128], which yields 3.3-6.6 mM of proteins. Even though some measured values of cellular protein concentration can be 3 times lower[128], the protein concentration should still be at least on the order of 1 mM. Given that around 50% of the proteins have at least one surface cysteine, among which each protein has 2 surface cysteines on average, the final surface cysteine concentration should be at least 1mM*50%*2 = 1mM.

Simulations of proteins with surface-exposed cysteines. Brownian dynamics simulations of proteins with surface cysteines (available at https://github.com/kannandeepti/protein_mobility) were performed by adapting the polychrom software package (https://doi.org/10.5281/zenodo.3579473), a thin wrapper around OpenMM [129]. We model proteins as self-avoiding, spherical particles of diameter $r_{rep} = 1.2$ which interact through a repulsive potential, $U(r_{ij}) = U_0\{1 + (\widetilde{r_{ij}})^{12}[(\widetilde{r_{ij}})^2 - 1]\}$ [Eq1], where $\widetilde{r_{ij}} = \frac{r_{ij}}{r_{rep}}\sqrt{6/7}$ and $U_0 = 50\ k_BT$ represent a finite energy barrier to allow particle overlaps when $r_{ij} < 0.6r_{rep}$ (Figure S5A). Based on our estimates, the average number of surface-exposed cysteines on proteins that have surface-exposed cysteines is two (see "Estimation of the average number of surface cysteine per protein and the concentration of surface cysteine"). Thus, simulated protein spheres are bonded to two surface "cysteines" via the harmonic potential $0.5k(r_{ij} - 0.5)^2$, where k is chosen such that the average extension of the bond is 0.01 when the bond energy is equal to k_BT. A harmonic angle potential of the form $0.5\kappa(\theta - 180º)^2$ with $\kappa = 30k_BT$ enforces that the two cysteines are on opposite sides of the protein sphere. Cysteines on separate proteins can form intermolecular disulfide bridges, which are modeled via a short-ranged attractive potential of the same form as Eq. 1 (Figure S5B). In this case, $U_0 = -E_{attr}$ is the depth of the attractive potential and $\widetilde{r_{ij}} = \frac{r_{ij}}{r_{attr}}\sqrt{6/7}$, where the cysteine-cysteine attraction radius is set to $r_{attr} = 0.2$. These parameters were chosen to minimize many-to-one bonding of cysteines, i.e. such that proteins with only 1 surface cysteine predominantly form dimers instead of higher order multimers (Figure S5C,D). We simulate 1000 proteins with periodic boundary conditions in a cube whose side length is chosen such that the proteins occupy 30% of the cube volume. Consistent with estimates of the fraction of proteins with surface-exposed cysteines in the cell (see "Estimation of the average number of surface cysteine per protein and the concentration of surface cysteine"), 50% of the simulated proteins have two surface cysteines which can form disulfide bonds according to the value of E_{attr}, and the surface patches of the other 50% do not participate in disulfide bonding. As seen in Figure 4C, the proteins without surface-exposed cysteines diffuse more slowly at high E_attr since they are diffusing through a mesh of crosslinked proteins. However, this mobility reduction is far less pronounced than that of the proteins with surface exposed cysteines, which form dimers and multimers at high E_attr (Figure S5E).

For each value of E_{attr}, the diffusion coefficient is quantified as the slope of the protein's mean squared displacement over time. We normalize the diffusion coefficient to the mean of all data points for $E_{attr} \leq 11.25k_BT$ and fit the resulting data to a decreasing S-curve of the form $D(E_{attr}) = min + (1.0 - min)\ [e^{-k(E_{attr}-E_0)}/(1 + e^{-k(E_{attr}-E_0)})]^a$

(solid line in Figure S5F). For each simulation, we also calculate the fraction of cysteines that participate in intermolecular disulfide bonding, i.e. the fraction of sticky patches which are within r_{attr} of a neighboring patch (Figure S5G). This data is fit to an increasing S-curve of the form $f(E_{attr}) = [1/(1 + e^{-k(E_{attr}-E_0)}]^a$ to obtain the solid line in Figure S5G. The fraction of bonded cysteines can be mapped to the oxidative state of a cell as measured by the ratio of oxidized to reduced glutathione (GSSG/GSH) using a chemical reaction model (see "*Chemical reaction model for coupling protein-protein disulfide bonding to redox state"* section). For a given value of E_{attr}, we use $f(E_{attr})$ to compute the fraction of bonded cysteines as obtained from the simulations. Figure S5H then allows us to read off the corresponding value of GSSG/GSH. Thus, we relate E_{attr} to the oxidative state of the cell. This in turn allows us to graph the diffusion coefficient as a function of the oxidative state in Figure 4C, where the solid lines represent the fit relationships in Figure S5F-H and points show raw simulation data.

Simulation hyperparameter tuning and validation. Since disulfide bonding is modeled via a pairwise attractive potential between surface-exposed cysteines, it is possible for one cysteine to attract more than one binding partner on neighboring proteins. Such many-to-one bonding events can be minimized by tuning the following simulation hyperparameters: the protein-protein repulsion radius, r_{rep} , the protein-protein repulsion energy E_{rep} (Figure S5A), the cysteine-cysteine attraction radius, r_{attr} (Figure S5B), , and the spring constant k for the harmonic bonds connecting each protein to each of the cysteines on its surface.

For a given set of parameters ($r_{attr}, r_{rep}, E_{rep}, k$), we calculate the equilibrium cysteine-cysteine distance ${d_3}^*$ and bond extension ${x_3}^*$ that minimizes the energy of a trimer of proteins with just one surface cysteine, $E_{trimer}(d, x) = 3U_{attr}(d) + 3U_{rep}(d + \sqrt{3}(0.5 + x)) + 3/2kx^2$ (Figure S5C). Analogously, we compute the values of ${d_2}^*$ and ${x_2}^*$ that minimize the energy of a dimer, $E_{dimer}(d, x) = U_{attr}(d) + U_{rep}(d + 2(0.5 + x)) + kx^2$.
We then choose a parameter set where $E_{trimer}({d_3}^*, {x_3}^*) > 0 > E_{dimer}({d_2}^*, {x_2}^*)$ for all values of $E_{attr} \in [0, 30]\ k_BT$. From this approach, we identified that the choice of $r_{attr} = 0.2, r_{rep} = 1.2, E_{rep} = 50k_BT, k = 2k_BT/(0.01)^2$ ensured that trimers are always less energetically favorable than dimers. In Figure S5D, we confirm that using these parameters, proteins with one surface cysteine only form dimers even at high values of E_{attr}.

Chemical reaction model for coupling protein-protein disulfide bonding to redox state. To investigate the coupling between cellular redox and the propensity for proteins with surface exposed cysteines to form disulfide bonds, we developed a minimal chemical reaction model. In this model, we assume the level of oxidative stress is represented by a fixed concentration of hydrogen peroxide $[H_2O_2]$, which is the predominant reactive oxygen species (ROS) in the cell[130]. We also assume that glutathione is the primary species responsible for regulating ROS, given that it is the most abundant non-protein

antioxidant in the cell[131]. Glutathione (GSH) and proteins with surface exposed thiol groups (PSH) are oxidized in the presence of H_2O_2 and can form disulfide bridges via the following set of reactions:

$$GSH + H_2O_2 \xrightarrow{k_{\mathrm{on}}} GSOH + H_2O$$
$$PSH + H_2O_2 \xrightarrow{k'_{\mathrm{on}}} PSOH + H_2O$$

$$GSOH + GSH \xrightarrow{k_{\mathrm{SS}}} GSSG + H_2O$$
$$GSOH + PSH \xrightarrow{k_{\mathrm{SS}}} PSSG + H_2O$$
$$PSOH + GSH \xrightarrow{k_{\mathrm{SS}}} PSSG + H_2O$$
$$PSOH + PSH \xrightarrow{k_{\mathrm{SS}}} PSSP + H_2O$$

The second-order rate constant of the oxidation of the thiol of a glutathione is k_{on}= 0.42 $M^{-1}s^{-1}$ [132] and the rate constant for the oxidation of the thiol of a protein is $k_{on}{}'$= 2.3 $M^{-1}s^{-1}$ (based on the oxidation of Cys-34 in BSA)[132]. GSOH and PSOH are highly reactive intermediate products which then undergo a much faster disulfide bridging process ($k_{SS} \gg k_{on}, k_{on}{}'$) with another thiol[133]. We set k_{SS} = 50 $k_{on}{}'$. We assume that proteins can form disulfide bonds with other proteins (PSSP) or with glutathione (PSSG) at equal rates.

Intramolecular disulfide bonds of proteins are removed by a thiol group interchange reaction with GSH[134]. We reason that removal of intermolecular disulfide bonds could be achieved by a similar mechanism,

$$PSSP + GSH \xrightarrow{k_{\mathrm{ex}}} PSSG + PSH$$
$$PSSG + GSH \xrightarrow{k_{\mathrm{ex}}} GSSG + PSH$$
$$GSSG + PSH \xrightarrow{k_{\mathrm{ex}}} PSSG + GSH$$
$$PSSG + PSH \xrightarrow{k_{\mathrm{ex}}} PSSP + GSH$$

where we take $k_{ex} \approx$ 0.15 $M^{-1}s^{-1}$ (estimated from the disulfide interchange between GSSG and 2-Mercaptoethanol at pH=7)[135]. Ultimately, GSSG is reduced by NADPH,

$$GSSG \xrightarrow{k_{\mathrm{NADPH}}} 2\,GSH,$$

while the total concentration of glutathione[136] and surface-exposed protein cysteines (see “Estimation of the average number of surface cysteine per protein and the concentration of surface cysteine”) are conserved at 1mM:

$$G_{tot} = [GSH] + [GSOH] + [PSSG] + 2[GSSG] = 1\text{mM}$$
$$P_{tot} = [PSH] + [PSOH] + [PSSG] + 2[PSSP] = 1\text{mM}.$$

We simulate the ordinary differential equations (see "ODEs of the chemical reaction model for coupling protein-protein disulfide bonding to redox state") associated with the above chemical reactions for the dynamics of [GSOH], [PSOH], [GSSG], [PSSG], and [PSSP] using MATLAB with ode45. The rate of glutathione reduction $k_{\mathrm{NADPH}} = 2.72 \times 10^{-6}$s^{-1} is chosen such that when the steady state [GSSG]/[GSH] ratio is 0.01 (physiological redox ratio[137]), $[H_2O_2]$ is 10nM (physiological hydrogen peroxide concentration[130]).

We then determine the fraction of surface-exposed cysteines that participate in protein-protein disulfide bonding, $2[PSSP]/P_{tot}$, as a function of the steady state ratio [GSSG]/[GSH] (Figure S5H). Note that even at high ROS, this fraction is capped at 0.62 since surface-exposed cysteines are equally likely to bind to a thiol on a neighboring protein or the thiol of glutathione. Our protein simulations do not include glutathione. Thus, the fraction of bonded cysteines as computed from simulations, which can go up to 1.0 (Figure S5G), corresponds to the fraction of cysteines participating in inter-protein disulfide bridges in our chemical reaction model. To map E_{attr} to GSSG/GSH, we only consider simulation data for which the fraction of bonded cysteines is less than 0.62.

ODEs of the chemical reaction model for coupling protein-protein disulfide bonding to redox state

$$\frac{d[GSOH]}{dt} = k_{\mathrm{on}}[GSH][H_2O_2] - k_{\mathrm{SS}}[GSH][GSOH] - k_{\mathrm{SS}}[PSH][GSOH]$$

$$\frac{d[PSOH]}{dt} = k'_{\mathrm{on}}[PSH][H_2O_2] - k_{\mathrm{SS}}[GSH][PSOH] - k_{\mathrm{SS}}[PSH][PSOH]$$

$$\frac{d[GSSG]}{dt} = k_{\mathrm{SS}}[GSH][GSOH] + k_{\mathrm{ex}}[PSSG][GSH] - k_{\mathrm{ex}}[GSSG][PSH] - k_{\mathrm{NADPH}}[GSSG]$$

$$\frac{d[PSSG]}{dt} = k_{\mathrm{SS}}[PSH][GSOH] + k_{\mathrm{SS}}[GSH][PSOH] + k_{\mathrm{ex}}[PSSP][GSH] + k_{\mathrm{ex}}[GSSG][PSH] - k_{\mathrm{ex}}[PSSG][GSH] - k_{\mathrm{ex}}[PSSG][PSH]$$

$$\frac{d[PSSP]}{dt} = k_{\mathrm{SS}}[PSH][PSOH] + k_{\mathrm{ex}}[PSSG][PSH] - k_{\mathrm{ex}}[PSSP][GSH]$$

$$[GSH] = G_{tot} - [GSOH] - [PSSG] - 2[GSSG]$$

$$[PSH] = P_{tot} - [PSOH] - [PSSG] - 2[PSSP]$$

where $[GSH] = G_{tot}$ and $[PSH] = P_{tot}$ at t = 0. $[H_2O_2]$ is a constant value for each simulation ranging from 10^{-3} uM to 10 uM.

ROS stain and imaging

Following cell treatment, media was removed and cells were incubated with 5 mM CellROX Deep Red Reagent (Thermo Fisher Scientific, C10422) diluted in EMEM for 30 minutes. Cells were then fixed with 4% paraformaldehyde in PBS (BTC Beantown Chemical, 140770-10x10ML) for 10 minutes. Cells were washed with PBS three times and imaged using the RPI Spinning disk confocal microscope, 60x objective. ROS signal intensity was measured using the "measure tool" on Fiji/ImageJ v2.1.0/153c. A two-tailed student's t-test was used to generate p-values. Statistical analysis was

performed using Prism Version 9.4.0 (GraphPad, La Jolla, CA). All ROS imaging experiments were performed twice using 2-4 biologically independent samples.

Modeling of the diffusion-limited tyrosine kinase receptor phosphorylation

For a generic reaction $A + B \rightarrow C^*$ at the molecular level, two basic steps are needed to accomplish this reaction: (i) molecules A and B need to "find" each other, and (ii) they transform into an activated complex C^*. The first basic step is called "collision" and the second basic step is called "activation". Accordingly, there are two fundamental rate constants that defines the overall reaction rate: the diffusion limited rate constant k_D describes the rate of collision through the diffusion process, and the inherent reaction constant k_r describes the rate of activation. While k_r is determined by the intrinsic chemical property, $k_D \coloneqq 4\pi(D_A + D_B)\beta$ largely depends on the diffusion coefficients of molecule A and B, which are D_A and D_B, respectively. β is the characteristic length-scale, defined below.

The exact relation between the overall reaction rate and the two fundamental rate constants is[138]:

$$k = \frac{k_D k_r}{k_r + k_D \exp\left(\frac{U(R_{\mathrm{AB}})}{k_B T}\right)}$$

In this equation, R_{AB} is the center-to-center distance when spherical molecules A and B touch, $U(r)$ is the potential between molecules A and B when spaced by a center-to-center distance r, $k_B T$ has the dimension of energy as the product of Boltzmann constant and temperature.

We then use the receptor tyrosine kinase phosphorylation reaction in the collision-limited realm as an example to quantify the relation between protein mobility (D) and protein functional activity (k) in cell:

$$k(D) = \frac{8\pi D\beta k_r}{k_r + 8\pi D\beta \exp\left(\frac{U(2R)}{k_B T}\right)}$$

Here we have assumed equal diffusion coefficients (D) and protein sizes (R) for substrate and enzyme for simplicity. To draw the *k-D* relation, additional parameters/functions need to be determined, including R, $U(r)$, $k_B T$, β, and k_r.

- The radius of receptor tyrosine kinases is estimated to be R=3 nm.
- For $U(r)$, we adapted a 10-5 Lennard-Jones potential in the colloid-type spherical model to describe the interactions between substrate and enzyme: $U(r) = 4\varepsilon\left(\left(\frac{\sigma}{r}\right)^{10} - \left(\frac{\sigma}{r}\right)^{5}\right)$, and common value of ε = 4.0 kJ/mol was used[139]. Lennard-Jones parameters σ were obtained from the protein radius: $\sigma \coloneqq 2^{5/6}R$ = 5.35 nm.
- $k_B T$ is set to the value that represents 37°C: $k_B T$ = 2.5 kJ/mol.
- β is a characteristic length-scale defined as $\beta^{-1} = \int_{R_{\mathrm{AB}}}^{\infty} dr \cdot \exp\left(U(r)/k_B T\right)/r^2$, which is computed as 7.60 nm in this case (note that R_{AB}=2R).

- To estimate the inherent reaction constant k_r, the apparent receptor tyrosine phosphorylation rate of EGFR *in vitro* is adopted: k = 5.5×10^7 M^{-1}s^{-1} [140]. Therefore, the k_r can be reversely solved in a dilute solution scenario (i.e., $\exp\left(\frac{U(R_{\mathrm{AB}})}{k_B T}\right) \approx 1$) as:

$$k_r = \left(\frac{1}{k} - \frac{1}{8\pi D_{vitro}\,\beta}\right)^{-1}$$

To back-calculate k_r, we also need an estimation of the *in vitro* diffusion coefficient of receptor tyrosine kinases. The diffusion coefficient of insulin receptor measured in our paper in live cells (D_{cell} ~ 1 μm^2/s) is adopted given the comparable molecular weights among insulin receptor, IRS1, and EGFR. It is also known that the diffusion coefficient *in vitro* is around 3 times higher than in cell[141], thus the effective diffusion coefficient of EGFR in the referred *in vitro* work[140] is estimated to be D_{vitro} ~ 3 μm^2/s. Hence, $k_r = 6.5 \times 10^7$ M^{-1}s^{-1}.

With those parameters/functions in hand, *k-D* relation is generated as plotted in Figure 6D.

In vitro IRS1 phosphorylation

Purified active recombinant human insulin receptor (IR) (Millipore, 14-466) and purified recombinant insulin responsive substrate 1 (IRS1) (Abcam, ab70538) were incubated in freshly prepared reaction buffer consisting of 50 mM Tris pH 7.5, 0.1 mM EGTA, 0.1 mM Na_3VO_4, 0.1 mM 2-mercaptoethanol, 10 mM $MnCl_2$ and 0.01 mg/ml bovine serum albumin with the indicated concentrations of glycerol (Invitrogen, 15514011) for 5 minutes at 30°C immediately after the addition of 50 μM ATP in 5 mM magnesium acetate. All samples were prepared with 75 ng of IR and 240 ng of IRS1. These amounts and ratios were chosen because they fell within the linear range of IRS1 phosphorylation by IR and they provided equivalent moles of IR and IRS1. For samples that were agitated, tubes were subjected to orbital mixing at 1200 rpm using a Thermomixer (Eppendorf, ThermoMixer C, EP5382000023) during the entire incubation. After 5 minutes, reactions were immediately quenched with dithiothreitol (DTT) and XT Sample Buffer 4x (BioRad, 1610791) to a final concentration of 100mM and 1x, respectively and incubated at 95°C for 5 minutes, then ran on Western blot or frozen at -80°C until subjected to western blot.

RNA-seq

Cells were treated with normal or pathogenic insulin concentrations for three days and washed with EMEM as described above. RNA was then purified using TRIzol™ reagent (Thermo Fisher Scientific, 15596026) following manufacturer's instructions. RNA-seq libraries were prepared using KAPAHyperRiboErase (Roche, KK8561) and were sequenced on Illumina NovaSeq 6000, generating at least 200 million paired-ended 150-bp reads per sample. Reads were mapped to the human genome GRCh38 using STAR aligner[105] (v2.7.1a), allowing up to 100 multiple alignments and up to 200 loci

anchors (--outFilterMultimapNmax 100 --winAnchorMultimapNmax 200). Differential expression analysis of genes and transposable elements comparing triplicates of samples treated with normal or pathogenic concentrations of insulin was performed using TEtranscripts[106] (v2.2.3). The list of protein coding genes was downloaded from ENSEMBL BioMart (http://www.ensembl.org/biomart/martview/6e82036bfd2b9ca0c5044d2c7449824d).

ChIP-seq

Published MED1 ChIP-seq data (GSM2040029) and input (GSM2864933) were used in this study. ChIP-seq bioinformatics analysis was performed on the Whitehead High-Performance Computing Facility using the nf-core ChIP-seq pipeline v1.2.1112 with Nextflow v20.04.1. Quality control of fastq files was performed with FastQC v0.11.9. Trim Galore! v0.6.4_dev was used to trim low-quality reads. Alignment was performed against the hg19 genome assembly using BWA v0.7.17-r1188[107]. Peak calling was performed using MACS2[108] v2.2.7.1 with q value of 0.01. For the identification of genes whose promoter (transcription start site, TSS, +/- 1kb) were occupied by MED1, the same hg38 gene list used for the RNA-seq analysis was used. The coordinates of the promoters were converted from hg38 to hg19 using LiftOver (https://genome.ucsc.edu/cgi-bin/hgLiftOver). Bedtools[109] v2.29.2 was used to measure the distance between MED1 peaks and gene promoters (bedtools closest -d). A gene was considered occupied by MED1 the distance between MED1 peak and the promoter was 0. The changes in gene expression measured by RNA-seq were matched to each MED1-occupied and non-occupied gene using the VLOOKUP tool in Excel v16.78.3.

Illustrations

PyMOL [142] was used for protein illustrations in Figure 4A, 6A, 7B, S6B. PDB ID: IR 6PXV, MED1 7EMF, HP1α 3I3C, FIB1 7SE7, SRSF2 2LEC. Cartoon illustrations were created with BioRender.com. Figures were generated using Adobe Illustrator v27.0.1.

Quantification and statistical analysis

Statistical analyses for FRAP were performed using the Statistics and Machine Learning Toolbox of MATLAB R2021b or R2024a (The MathWorks, Inc., Natick, MA). All other statistical analyses were performed using Prism Version 9.4.0 (GraphPad, La Jolla, CA). All statistical tests used, the exact value of n, and what n represents can all be found in the figure legends. All data are reported as mean $\pm$ SEM or mean + SEM. For Figures 2E, 5B, 5C, 5F, 6J, S4B, S6E, and S7, a two-tailed Mann-Whitney U test was applied. For Figures 2G, 3F, 3I, 4D-F, 4H, 4I, 5E, S3A, S4A, S4C-E, and S6C, an unpaired two-tailed student's *t*-test was applied. For Figures 3B, 6E, 6K, 6L, and S1A-E, an unpaired two-tailed student's *t*-test with Welch's correction was applied. All statistical results were done without randomization or stratification. The notation for statistical significance is as follows: * represents $p < 0.05$, ** represents $p < 0.01$ and *** represents $p < 0.001$.

References

1. Hajat, C., and Stein, E. (2018). The global burden of multiple chronic conditions: A narrative review. Prev Med Rep *12*, 284-293. 10.1016/j.pmedr.2018.10.008.
2. Collins, F.S.D., J.A.; Lander, E.S.; Rotimi, C.N. (2021). Human Molecular Genetics and Genomics — Important Advances and Exciting Possibilities. The New England Journal of Medicine, 1-4.
3. Roden, M., and Shulman, G.I. (2019). The integrative biology of type 2 diabetes. Nature *576*, 51-60. 10.1038/s41586-019-1797-8.
4. Langley, M.R., Rangaraju, S., Dey, A., and Sarkar, S. (2022). Editorial: Environmental Effect on Neuroinflammation and Neurodegeneration. Front Cell Neurosci *16*, 935190. 10.3389/fncel.2022.935190.
5. Janssen, J. (2021). Hyperinsulinemia and Its Pivotal Role in Aging, Obesity, Type 2 Diabetes, Cardiovascular Disease and Cancer. Int J Mol Sci *22*. 10.3390/ijms22157797.
6. Shanik, M.H., Xu, Y., Skrha, J., Dankner, R., Zick, Y., and Roth, J. (2008). Insulin resistance and hyperinsulinemia: is hyperinsulinemia the cart or the horse? Diabetes Care *31 Suppl 2*, S262-268. 10.2337/dc08-s264.
7. Crofts, C.A.P.Z., C.; Wheldon, M.C.; Schofield, G.M. (2015). Hyperinsulinemia: A unifying theory of chronic disease? Diabesity *1*, 34-43.
8. DeFronzo, R.A., Ferrannini, E., Groop, L., Henry, R.R., Herman, W.H., Holst, J.J., Hu, F.B., Kahn, C.R., Raz, I., Shulman, G.I., et al. (2015). Type 2 diabetes mellitus. Nat Rev Dis Primers *1*, 15019. 10.1038/nrdp.2015.19.
9. Liu, J., Liu, S., Yu, Z., Qiu, X., Jiang, R., and Li, W. (2022). Uncovering the gene regulatory network of type 2 diabetes through multi-omic data integration. J Transl Med *20*, 604. 10.1186/s12967-022-03826-5.
10. Ahmed, S.A.H., Ansari, S.A., Mensah-Brown, E.P.K., and Emerald, B.S. (2020). The role of DNA methylation in the pathogenesis of type 2 diabetes mellitus. Clin Epigenetics *12*, 104. 10.1186/s13148-020-00896-4.
11. Dlamini, Z., Mokoena, F., and Hull, R. (2017). Abnormalities in alternative splicing in diabetes: therapeutic targets. J Mol Endocrinol *59*, R93-R107. 10.1530/JME-17-0049.
12. Kobiita, A., Godbersen, S., Araldi, E., Ghoshdastider, U., Schmid, M.W., Spinas, G., Moch, H., and Stoffel, M. (2020). The Diabetes Gene JAZF1 Is Essential for the Homeostatic Control of Ribosome Biogenesis and Function in Metabolic Stress. Cell Rep *32*, 107846. 10.1016/j.celrep.2020.107846.
13. Yuan, Q., Tang, B., and Zhang, C. (2022). Signaling pathways of chronic kidney diseases, implications for therapeutics. Signal Transduct Target Ther *7*, 182. 10.1038/s41392-022-01036-5.
14. Bradner, J.E., Hnisz, D., and Young, R.A. (2017). Transcriptional Addiction in Cancer. Cell *168*, 629-643. 10.1016/j.cell.2016.12.013.
15. Meng, G., and Mei, H. (2019). Transcriptional Dysregulation Study Reveals a Core Network Involving the Progression of Alzheimer's Disease. Front Aging Neurosci *11*, 101. 10.3389/fnagi.2019.00101.
16. Gandal, M.J., Haney, J.R., Wamsley, B., Yap, C.X., Parhami, S., Emani, P.S., Chang, N., Chen, G.T., Hoftman, G.D., de Alba, D., et al. (2022). Broad transcriptomic dysregulation occurs across the cerebral cortex in ASD. Nature *611*, 532-539. 10.1038/s41586-022-05377-7.

17. Fuller, H., Zhu, Y., Nicholas, J., Chatelaine, H.A., Drzymalla, E.M., Sarvestani, A.K., Julian-Serrano, S., Tahir, U.A., Sinnott-Armstrong, N., Raffield, L.M., et al. (2023). Metabolomic epidemiology offers insights into disease aetiology. Nat Metab *5*, 1656-1672. 10.1038/s42255-023-00903-x.
18. Jiao, L., Liu, Y., Yu, X.Y., Pan, X., Zhang, Y., Tu, J., Song, Y.H., and Li, Y. (2023). Ribosome biogenesis in disease: new players and therapeutic targets. Signal Transduct Target Ther *8*, 15. 10.1038/s41392-022-01285-4.
19. Misteli, T. (2008). Physiological importance of RNA and protein mobility in the cell nucleus. Histochem Cell Biol *129*, 5-11. 10.1007/s00418-007-0355-x.
20. Misteli, T. (2001). Protein dynamics: implications for nuclear architecture and gene expression. Science *291*, 843-847. 10.1126/science.291.5505.843.
21. Ellis, R.J. (2001). Macromolecular crowding: an important but neglected aspect of the intracellular environment. Curr Opin Struct Biol *11*, 114-119. 10.1016/s0959-440x(00)00172-x.
22. Axelrod, D., Koppel, D.E., Schlessinger, J., Elson, E., and Webb, W.W. (1976). Mobility measurement by analysis of fluorescence photobleaching recovery kinetics. Biophys J *16*, 1055-1069. 10.1016/S0006-3495(76)85755-4.
23. Berg, H. (1993). Random Walks in Biology Expanded Edition. Princeton University Press.
24. Lakadamyali, M. (2022). Single nucleosome tracking to study chromatin plasticity. Curr Opin Cell Biol *74*, 23-28. 10.1016/j.ceb.2021.12.005.
25. Su, X., Ditlev, J.A., Hui, E., Xing, W., Banjade, S., Okrut, J., King, D.S., Taunton, J., Rosen, M.K., and Vale, R.D. (2016). Phase separation of signaling molecules promotes T cell receptor signal transduction. Science *352*, 595-599. 10.1126/science.aad9964.
26. Ditlev, J.A., Vega, A.R., Koster, D.V., Su, X., Tani, T., Lakoduk, A.M., Vale, R.D., Mayor, S., Jaqaman, K., and Rosen, M.K. (2019). A composition-dependent molecular clutch between T cell signaling condensates and actin. Elife *8*. 10.7554/eLife.42695.
27. Delarue, M., Brittingham, G.P., Pfeffer, S., Surovtsev, I.V., Pinglay, S., Kennedy, K.J., Schaffer, M., Gutierrez, J.I., Sang, D., Poterewicz, G., et al. (2018). mTORC1 Controls Phase Separation and the Biophysical Properties of the Cytoplasm by Tuning Crowding. Cell *174*, 338-349 e320. 10.1016/j.cell.2018.05.042.
28. Dall'Agnese, A., Platt, J.M., Zheng, M.M., Friesen, M., Dall'Agnese, G., Blaise, A.M., Spinelli, J.B., Henninger, J.E., Tevonian, E.N., Hannett, N.M., et al. (2022). The dynamic clustering of insulin receptor underlies its signaling and is disrupted in insulin resistance. Nat Commun *13*, 7522. 10.1038/s41467-022-35176-7.
29. Li, H., Zhang, J., Shi, Y., Zhao, G., Xu, H., Cai, M., Gao, J., and Wang, H. (2022). Mechanism of INSR clustering with insulin activation and resistance revealed by super-resolution imaging. Nanoscale *14*, 7747-7755. 10.1039/d2nr01051h.
30. Nair, S.J., Yang, L., Meluzzi, D., Oh, S., Yang, F., Friedman, M.J., Wang, S., Suter, T., Alshareedah, I., Gamliel, A., et al. (2019). Phase separation of ligand-activated enhancers licenses cooperative chromosomal enhancer assembly. Nat Struct Mol Biol *26*, 193-203. 10.1038/s41594-019-0190-5.
31. Persson, L.B., Ambati, V.S., and Brandman, O. (2020). Cellular Control of Viscosity Counters Changes in Temperature and Energy Availability. Cell *183*, 1572-1585 e1516. 10.1016/j.cell.2020.10.017.

32. Alejo, J.L., Kempes, C.P., and Adamala, K.P. (2022). Diffusion control in biochemical specificity. Biophys J *121*, 1541-1548. 10.1016/j.bpj.2022.03.005.
33. Schavemaker, P.E., Boersma, A.J., and Poolman, B. (2018). How Important Is Protein Diffusion in Prokaryotes? Front Mol Biosci *5*, 93. 10.3389/fmolb.2018.00093.
34. Gralle, M., Labrecque, S., Salesse, C., and De Koninck, P. (2021). Spatial dynamics of the insulin receptor in living neurons. J Neurochem *156*, 88-105. 10.1111/jnc.14950.
35. Oksuz, O., Henninger, J.E., Warneford-Thomson, R., Zheng, M.M., Erb, H., Vancura, A., Overholt, K.J., Hawken, S.W., Banani, S.F., Lauman, R., et al. (2023). Transcription factors interact with RNA to regulate genes. Mol Cell *83*, 2449-2463 e2413. 10.1016/j.molcel.2023.06.012.
36. Mir, M., Stadler, M.R., Ortiz, S.A., Hannon, C.E., Harrison, M.M., Darzacq, X., and Eisen, M.B. (2018). Dynamic multifactor hubs interact transiently with sites of active transcription in Drosophila embryos. Elife *7*. 10.7554/eLife.40497.
37. Gomez-Garcia, P.A., Portillo-Ledesma, S., Neguembor, M.V., Pesaresi, M., Oweis, W., Rohrlich, T., Wieser, S., Meshorer, E., Schlick, T., Cosma, M.P., and Lakadamyali, M. (2021). Mesoscale Modeling and Single-Nucleosome Tracking Reveal Remodeling of Clutch Folding and Dynamics in Stem Cell Differentiation. Cell Rep *34*, 108614. 10.1016/j.celrep.2020.108614.
38. Sefried, S., Haring, H.U., Weigert, C., and Eckstein, S.S. (2018). Suitability of hepatocyte cell lines HepG2, AML12 and THLE-2 for investigation of insulin signalling and hepatokine gene expression. Open Biol *8*. 10.1098/rsob.180147.
39. Irvine, K.M., Skoien, R., Bokil, N.J., Melino, M., Thomas, G.P., Loo, D., Gabrielli, B., Hill, M.M., Sweet, M.J., Clouston, A.D., and Powell, E.E. (2014). Senescent human hepatocytes express a unique secretory phenotype and promote macrophage migration. World J Gastroenterol *20*, 17851-17862. 10.3748/wjg.v20.i47.17851.
40. Müller, F.A., Sturla, S.J. (2019). Human in vitro models of nonalcoholic fatty liver disease. Current Opinion in Toxicology *16*, 9-16.
41. Wang, Z.D., W. (2022). Dynamic transcription regulation at the single-molecule level. Developmental Biology *482*, 67-81.
42. Chen, K.Y., Jenkins, E., Korbel, M., Ponjavic, A., Lippert, A.H., Santos, A.M., Ashman, N., O'Brien-Ball, C., McBride, J., Klenerman, D., and Davis, S.J. (2021). Trapping or slowing the diffusion of T cell receptors at close contacts initiates T cell signaling. Proc Natl Acad Sci U S A *118*. 10.1073/pnas.2024250118.
43. Bryan, L.C., Weilandt, D.R., Bachmann, A.L., Kilic, S., Lechner, C.C., Odermatt, P.D., Fantner, G.E., Georgeon, S., Hantschel, O., Hatzimanikatis, V., and Fierz, B. (2017). Single-molecule kinetic analysis of HP1-chromatin binding reveals a dynamic network of histone modification and DNA interactions. Nucleic Acids Res *45*, 10504-10517. 10.1093/nar/gkx697.
44. Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A.V., Manteiga, J.C., et al. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. Science *361*. 10.1126/science.aar3958.
45. Feric, M., Vaidya, N., Harmon, T.S., Mitrea, D.M., Zhu, L., Richardson, T.M., Kriwacki, R.W., Pappu, R.V., and Brangwynne, C.P. (2016). Coexisting Liquid Phases Underlie Nucleolar Subcompartments. Cell *165*, 1686-1697. 10.1016/j.cell.2016.04.047.

46. Song, S.H., McIntyre, S.S., Shah, H., Veldhuis, J.D., Hayes, P.C., and Butler, P.C. (2000). Direct measurement of pulsatile insulin secretion from the portal vein in human subjects. J Clin Endocrinol Metab *85*, 4491-4499. 10.1210/jcem.85.12.7043.
47. Porksen, N., Grofte, T., Greisen, J., Mengel, A., Juhl, C., Veldhuis, J.D., Schmitz, O., Rossle, M., and Vilstrup, H. (2002). Human insulin release processes measured by intraportal sampling. Am J Physiol Endocrinol Metab *282*, E695-702. 10.1152/ajpendo.00516.2000.
48. Alaaeldin, R., Abdel-Rahman, I.A.M., Hassan, H.A., Youssef, N., Allam, A.E., Abdelwahab, S.F., Zhao, Q.L., and Fathy, M. (2021). Carpachromene Ameliorates Insulin Resistance in HepG2 Cells via Modulating IR/IRS1/PI3k/Akt/GSK3/FoxO1 Pathway. Molecules *26*. 10.3390/molecules26247629.
49. Sabari, B.R., Dall'Agnese, A., and Young, R.A. (2020). Biomolecular Condensates in the Nucleus. Trends Biochem Sci *45*, 961-977. 10.1016/j.tibs.2020.06.007.
50. Strom, A.R., Emelyanov, A.V., Mir, M., Fyodorov, D.V., Darzacq, X., and Karpen, G.H. (2017). Phase separation drives heterochromatin domain formation. Nature *547*, 241-245. 10.1038/nature22989.
51. Larson, A.G., Elnatan, D., Keenen, M.M., Trnka, M.J., Johnston, J.B., Burlingame, A.L., Agard, D.A., Redding, S., and Narlikar, G.J. (2017). Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin. Nature *547*, 236-240. 10.1038/nature22822.
52. Spector, D.L., and Lamond, A.I. (2011). Nuclear speckles. Cold Spring Harb Perspect Biol *3*. 10.1101/cshperspect.a000646.
53. Banani, S.F., Lee, H.O., Hyman, A.A., and Rosen, M.K. (2017). Biomolecular condensates: organizers of cellular biochemistry. Nat Rev Mol Cell Biol *18*, 285-298. 10.1038/nrm.2017.7.
54. Brangwynne, C.P., Eckmann, C.R., Courson, D.S., Rybarska, A., Hoege, C., Gharakhani, J., Julicher, F., and Hyman, A.A. (2009). Germline P granules are liquid droplets that localize by controlled dissolution/condensation. Science *324*, 1729-1732. 10.1126/science.1172046.
55. Alberti, S. (2017). The wisdom of crowds: regulating cell function through condensed states of living matter. J Cell Sci *130*, 2789-2796. 10.1242/jcs.200295.
56. Huang, X., Chen, S., Li, W., Tang, L., Zhang, Y., Yang, N., Zou, Y., Zhai, X., Xiao, N., Liu, W., et al. (2021). ROS regulated reversible protein phase separation synchronizes plant flowering. Nat Chem Biol *17*, 549-557. 10.1038/s41589-021-00739-0.
57. Chong, P.A., and Forman-Kay, J.D. (2019). Oxidative Inhibition of Pbp1 Phase Separation. Biochemistry *58*, 3057-3059. 10.1021/acs.biochem.9b00511.
58. Kato, M., Yang, Y.S., Sutter, B.M., Wang, Y., McKnight, S.L., and Tu, B.P. (2019). Redox State Controls Phase Separation of the Yeast Ataxin-2 Protein via Reversible Oxidation of Its Methionine-Rich Low-Complexity Domain. Cell *177*, 711-721 e718. 10.1016/j.cell.2019.02.044.
59. Visser, A.J., Westphal, A.H., Skakun, V.V., and Borst, J.W. (2016). GFP as potential cellular viscosimeter. Methods Appl Fluoresc *4*, 035002. 10.1088/2050-6120/4/3/035002.
60. Adrien, V., Rayan, G., Astafyeva, K., Broutin, I., Picard, M., Fuchs, P., Urbach, W., and Taulier, N. (2022). How to best estimate the viscosity of lipid bilayers. Biophys Chem *281*, 106732. 10.1016/j.bpc.2021.106732.

61. Pucadyil, T.J., and Chattopadhyay, A. (2006). Confocal fluorescence recovery after photobleaching of green fluorescent protein in solution. J Fluoresc *16*, 87-94. 10.1007/s10895-005-0019-y.
62. Ge, X., Yu, Q., Qi, W., Shi, X., and Zhai, Q. (2008). Chronic insulin treatment causes insulin resistance in 3T3-L1 adipocytes through oxidative stress. Free Radic Res *42*, 582-591. 10.1080/10715760802158448.
63. Cumming, R.C., Andon, N.L., Haynes, P.A., Park, M., Fischer, W.H., and Schubert, D. (2004). Protein disulfide bond formation in the cytoplasm during oxidative stress. J Biol Chem *279*, 21749-21758. 10.1074/jbc.M312267200.
64. Wilkins, D.K., Grimshaw, S.B., Receveur, V., Dobson, C.M., Jones, J.A., and Smith, L.J. (1999). Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. Biochemistry *38*, 16424-16431. 10.1021/bi991765q.
65. Winter, J., Ilbert, M., Graf, P.C., Ozcelik, D., and Jakob, U. (2008). Bleach activates a redox-regulated chaperone by oxidative protein unfolding. Cell *135*, 691-701. 10.1016/j.cell.2008.09.024.
66. Ilbert, M., Horst, J., Ahrens, S., Winter, J., Graf, P.C., Lilie, H., and Jakob, U. (2007). The redox-switch domain of Hsp33 functions as dual stress sensor. Nat Struct Mol Biol *14*, 556-563. 10.1038/nsmb1244.
67. Knudsen, J.R., Persson, K.W., Henriquez-Olguin, C., Li, Z., Di Leo, N., Hesselager, S.A., Raun, S.H., Hingst, J.R., Trouillon, R., Wohlwend, M., et al. (2023). Microtubule-mediated GLUT4 trafficking is disrupted in insulin-resistant skeletal muscle. Elife *12*. 10.7554/eLife.83338.
68. Mahmud, Z., Malik, S.U., Ahmed, J., and Azad, A.K. (2016). Computational Analysis of Damaging Single-Nucleotide Polymorphisms and Their Structural and Functional Impact on the Insulin Receptor. Biomed Res Int *2016*, 2023803. 10.1155/2016/2023803.
69. Borsari, C., Keles, E., McPhail, J.A., Schaefer, A., Sriramaratnam, R., Goch, W., Schaefer, T., De Pascale, M., Bal, W., Gstaiger, M., et al. (2022). Covalent Proximity Scanning of a Distal Cysteine to Target PI3Kalpha. J Am Chem Soc *144*, 6326-6342. 10.1021/jacs.1c13568.
70. Sharifi-Rad, M., Anil Kumar, N.V., Zucca, P., Varoni, E.M., Dini, L., Panzarini, E., Rajkovic, J., Tsouh Fokou, P.V., Azzini, E., Peluso, I., et al. (2020). Lifestyle, Oxidative Stress, and Antioxidants: Back and Forth in the Pathophysiology of Chronic Diseases. Front Physiol *11*, 694. 10.3389/fphys.2020.00694.
71. Panahi, G., Pasalar, P., Zare, M., Rizzuto, R., and Meshkani, R. (2018). High glucose induces inflammatory responses in HepG2 cells via the oxidative stress-mediated activation of NF-kappaB, and MAPK pathways in HepG2 cells. Arch Physiol Biochem *124*, 468-474. 10.1080/13813455.2018.1427764.
72. Fischer, R., and Maier, O. (2015). Interrelation of oxidative stress and inflammation in neurodegenerative disease: role of TNF. Oxid Med Cell Longev *2015*, 610813. 10.1155/2015/610813.
73. Ly, L.D., Xu, S., Choi, S.K., Ha, C.M., Thoudam, T., Cha, S.K., Wiederkehr, A., Wollheim, C.B., Lee, I.K., and Park, K.S. (2017). Oxidative stress and calcium dysregulation by palmitate in type 2 diabetes. Exp Mol Med *49*, e291. 10.1038/emm.2016.157.
74. Raza, H., John, A., and Shafarin, J. (2016). Potentiation of LPS-Induced Apoptotic Cell Death in Human Hepatoma HepG2 Cells by Aspirin via ROS and Mitochondrial

Dysfunction: Protection by N-Acetyl Cysteine. PLoS One *11*, e0159750. 10.1371/journal.pone.0159750.
75. Shin, H.J., Kwon, H.K., Lee, J.H., Anwar, M.A., and Choi, S. (2016). Etoposide induced cytotoxicity mediated by ROS and ERK in human kidney proximal tubule cells. Sci Rep *6*, 34064. 10.1038/srep34064.
76. Jackson, S.P., and Bartek, J. (2009). The DNA-damage response in human biology and disease. Nature *461*, 1071-1078. 10.1038/nature08467.
77. Singh, A., Koduru, B., Carlisle, C., Akhter, H., Liu, R.M., Schroder, K., Brandes, R.P., and Ojcius, D.M. (2017). NADPH oxidase 4 modulates hepatic responses to lipopolysaccharide mediated by Toll-like receptor-4. Sci Rep *7*, 14346. 10.1038/s41598-017-14574-8.
78. Gonzalez, P., Lozano, P., Ros, G., and Solano, F. (2023). Hyperglycemia and Oxidative Stress: An Integral, Updated and Critical Overview of Their Metabolic Interconnections. Int J Mol Sci *24*. 10.3390/ijms24119352.
79. Inoguchi, T., Li, P., Umeda, F., Yu, H.Y., Kakimoto, M., Imamura, M., Aoki, T., Etoh, T., Hashimoto, T., Naruse, M., et al. (2000). High glucose level and free fatty acid stimulate reactive oxygen species production through protein kinase C--dependent activation of NAD(P)H oxidase in cultured vascular cells. Diabetes *49*, 1939-1945. 10.2337/diabetes.49.11.1939.
80. Chen, X., Andresen, B.T., Hill, M., Zhang, J., Booth, F., and Zhang, C. (2008). Role of Reactive Oxygen Species in Tumor Necrosis Factor-alpha Induced Endothelial Dysfunction. Curr Hypertens Rev *4*, 245-255. 10.2174/157340208786241336.
81. Griffin, M.E., Marcucci, M.J., Cline, G.W., Bell, K., Barucci, N., Lee, D., Goodyear, L.J., Kraegen, E.W., White, M.F., and Shulman, G.I. (1999). Free fatty acid-induced insulin resistance is associated with activation of protein kinase C theta and alterations in the insulin signaling cascade. Diabetes *48*, 1270-1274. 10.2337/diabetes.48.6.1270.
82. Schulze-Osthoff, K., Bakker, A.C., Vanhaesebroeck, B., Beyaert, R., Jacob, W.A., and Fiers, W. (1992). Cytotoxic activity of tumor necrosis factor is mediated by early damage of mitochondrial functions. Evidence for the involvement of mitochondrial radical generation. J Biol Chem *267*, 5317-5323.
83. Rosca, M.G., Vazquez, E.J., Chen, Q., Kerner, J., Kern, T.S., and Hoppel, C.L. (2012). Oxidation of fatty acids is the source of increased mitochondrial reactive oxygen species production in kidney cortical tubules in early diabetes. Diabetes *61*, 2074-2083. 10.2337/db11-1437.
84. Bałdyga, J., and Bourne, J.R. (1999). Turbulent mixing and chemical reactions (Wiley).
85. Yoshimura, S.H., Otsuka, S., Kumeta, M., Taga, M., and Takeyasu, K. (2013). Intermolecular disulfide bonds between nucleoporins regulate karyopherin-dependent nuclear transport. J Cell Sci *126*, 3141-3150. 10.1242/jcs.124172.
86. Higo, S., Asano, Y., Kato, H., Yamazaki, S., Nakano, A., Tsukamoto, O., Seguchi, O., Asai, M., Asakura, M., Asanuma, H., et al. (2010). Isoform-specific intermolecular disulfide bond formation of heterochromatin protein 1 (HP1). J Biol Chem *285*, 31337-31347. 10.1074/jbc.M110.155788.
87. Portoles, M.T., Pagani, R., Diaz-Laviada, I., and Municio, A.M. (1987). Effect of Escherichia coli lipopolysaccharide on the microviscosity of liver plasma membranes and hepatocyte suspensions and monolayers. Cell Biochem Funct *5*, 55-61. 10.1002/cbf.290050107.

88. de la Haba, C., Palacio, J.R., Martinez, P., and Morros, A. (2013). Effect of oxidative stress on plasma membrane fluidity of THP-1 induced macrophages. Biochim Biophys Acta *1828*, 357-364. 10.1016/j.bbamem.2012.08.013.
89. Perona, J.S. (2017). Membrane lipid alterations in the metabolic syndrome and the role of dietary oils. Biochim Biophys Acta Biomembr *1859*, 1690-1703. 10.1016/j.bbamem.2017.04.015.
90. Tsuda, K., Kinoshita, Y., Nishio, I., and Masuyama, Y. (2001). Hyperinsulinemia is a determinant of membrane fluidity of erythrocytes in essential hypertension. Am J Hypertens *14*, 419-423. 10.1016/s0895-7061(00)01247-4.
91. Pilon, M. (2016). Revisiting the membrane-centric view of diabetes. Lipids Health Dis *15*, 167. 10.1186/s12944-016-0342-0.
92. Winocour, P.D., Watala, C., and Kinglough-Rathbone, R.L. (1992). Membrane fluidity is related to the extent of glycation of proteins, but not to alterations in the cholesterol to phospholipid molar ratio in isolated platelet membranes from diabetic and control subjects. Thromb Haemost *67*, 567-571.
93. Kamboj, S.S., Chopra, K., and Sandhir, R. (2009). Hyperglycemia-induced alterations in synaptosomal membrane fluidity and activity of membrane bound enzymes: beneficial effect of N-acetylcysteine supplementation. Neuroscience *162*, 349-358. 10.1016/j.neuroscience.2009.05.002.
94. Jaqaman, K., and Grinstein, S. (2012). Regulation from within: the cytoskeleton in transmembrane signaling. Trends Cell Biol *22*, 515-526. 10.1016/j.tcb.2012.07.006.
95. Carlini, L., Brittingham, G.P., Holt, L.J., and Kapoor, T.M. (2020). Microtubules Enhance Mesoscale Effective Diffusivity in the Crowded Metaphase Cytoplasm. Dev Cell *54*, 574-582 e574. 10.1016/j.devcel.2020.07.020.
96. Molines, A.T., Lemiere, J., Gazzola, M., Steinmark, I.E., Edrington, C.H., Hsu, C.T., Real-Calderon, P., Suhling, K., Goshima, G., Holt, L.J., et al. (2022). Physical properties of the cytoplasm modulate the rates of microtubule polymerization and depolymerization. Dev Cell *57*, 466-479 e466. 10.1016/j.devcel.2022.02.001.
97. Cho, W.K., Spille, J.H., Hecht, M., Lee, C., Li, C., Grube, V., and Cisse, II (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science *361*, 412-415. 10.1126/science.aar4199.
98. Keber, F.C., Nguyen, T., Mariossi, A., Brangwynne, C.P., and Wuhr, M. (2024). Evidence for widespread cytoplasmic structuring into mesoscale condensates. Nat Cell Biol *26*, 346-352. 10.1038/s41556-024-01363-5.
99. Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G.M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X., and Tjian, R. (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. Science *361*. 10.1126/science.aar2555.
100. Cho, W.K., Jayanth, N., English, B.P., Inoue, T., Andrews, J.O., Conway, W., Grimm, J.B., Spille, J.H., Lavis, L.D., Lionnet, T., and Cisse, II (2016). RNA Polymerase II cluster dynamics predict mRNA output in living cells. Elife *5*. 10.7554/eLife.13617.
101. Ray, P.D., Huang, B.W., and Tsuji, Y. (2012). Reactive oxygen species (ROS) homeostasis and redox regulation in cellular signaling. Cell Signal *24*, 981-990. 10.1016/j.cellsig.2012.01.008.
102. Schieber, M., and Chandel, N.S. (2014). ROS function in redox signaling and oxidative stress. Curr Biol *24*, R453-462. 10.1016/j.cub.2014.03.034.

103. Klein, I.A., Boija, A., Afeyan, L.K., Hawken, S.W., Fan, M., Dall'Agnese, A., Oksuz, O., Henninger, J.E., Shrinivas, K., Sabari, B.R., et al. (2020). Partitioning of cancer therapeutics in nuclear condensates. Science *368*, 1386-1392. 10.1126/science.aaz4427.
104. Stringer, C., Wang, T., Michaelos, M., and Pachitariu, M. (2021). Cellpose: a generalist algorithm for cellular segmentation. Nat Methods *18*, 100-106. 10.1038/s41592-020-01018-x.
105. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21. 10.1093/bioinformatics/bts635.
106. Jin, Y., Tam, O.H., Paniagua, E., and Hammell, M. (2015). TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. Bioinformatics *31*, 3593-3599. 10.1093/bioinformatics/btv422.
107. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760. 10.1093/bioinformatics/btp324.
108. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137. 10.1186/gb-2008-9-9-r137.
109. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842. 10.1093/bioinformatics/btq033.
110. Chen, X., Zaro, J.L., and Shen, W.C. (2013). Fusion protein linkers: property, design and functionality. Adv Drug Deliv Rev *65*, 1357-1369. 10.1016/j.addr.2012.09.039.
111. Li, C.H., Coffey, E.L., Dall'Agnese, A., Hannett, N.M., Tang, X., Henninger, J.E., Platt, J.M., Oksuz, O., Zamudio, A.V., Afeyan, L.K., et al. (2020). MeCP2 links heterochromatin condensates and neurodevelopmental disease. Nature *586*, 440-444. 10.1038/s41586-020-2574-4.
112. Li, L., Pan, Z., and Yang, X. (2019). Key genes and co-expression network analysis in the livers of type 2 diabetes patients. J Diabetes Investig *10*, 951-962. 10.1111/jdi.12998.
113. Guo, Y.E., Manteiga, J.C., Henninger, J.E., Sabari, B.R., Dall'Agnese, A., Hannett, N.M., Spille, J.H., Afeyan, L.K., Zamudio, A.V., Shrinivas, K., et al. (2019). Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. Nature *572*, 543-548. 10.1038/s41586-019-1464-0.
114. Radtke, K.K., Coles, L.D., Mishra, U., Orchard, P.J., Holmay, M., and Cloyd, J.C. (2012). Interaction of N-acetylcysteine and cysteine in human plasma. J Pharm Sci *101*, 4653-4659. 10.1002/jps.23325.
115. Serge, A., Bertaux, N., Rigneault, H., and Marguet, D. (2008). Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes. Nat Methods *5*, 687-694. 10.1038/nmeth.1233.
116. Ovesny, M., Krizek, P., Borkovec, J., Svindrych, Z., and Hagen, G.M. (2014). ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. Bioinformatics *30*, 2389-2390. 10.1093/bioinformatics/btu202.
117. Izeddin, I., Recamier, V., Bosanac, L., Cisse, II, Boudarene, L., Dugast-Darzacq, C., Proux, F., Benichou, O., Voituriez, R., Bensaude, O., et al. (2014). Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus. Elife *3*. 10.7554/eLife.02230.

118. Hansen, A.S., Woringer, M., Grimm, J.B., Lavis, L.D., Tjian, R., and Darzacq, X. (2018). Robust model-based analysis of single-particle tracking experiments with Spot-On. Elife *7*. 10.7554/eLife.33125.
119. Kent, S., Brown, K., Yang, C.H., Alsaihati, N., Tian, C., Wang, H., and Ren, X. (2020). Phase-Separated Transcriptional Condensates Accelerate Target-Search Process Revealed by Live-Cell Single-Molecule Imaging. Cell Rep *33*, 108248. 10.1016/j.celrep.2020.108248.
120. Nguyen, V.Q., Ranjan, A., Liu, S., Tang, X., Ling, Y.H., Wisniewski, J., Mizuguchi, G., Li, K.Y., Jou, V., Zheng, Q., et al. (2021). Spatiotemporal coordination of transcription preinitiation complex assembly in live cells. Mol Cell *81*, 3560-3575 e3566. 10.1016/j.molcel.2021.07.022.
121. Chen, J., Zhang, Z., Li, L., Chen, B.C., Revyakin, A., Hajj, B., Legant, W., Dahan, M., Lionnet, T., Betzig, E., et al. (2014). Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. Cell *156*, 1274-1285. 10.1016/j.cell.2014.01.062.
122. Krouglova, T., Vercammen, J., and Engelborghs, Y. (2004). Correct diffusion coefficients of proteins in fluorescence correlation spectroscopy. Application to tubulin oligomers induced by Mg2+ and Paclitaxel. Biophys J *87*, 2635-2646. 10.1529/biophysj.104.040717.
123. Banani, S.F., Afeyan, L.K., Hawken, S.W., Henninger, J.E., Dall'Agnese, A., Clark, V.E., Platt, J.M., Oksuz, O., Hannett, N.M., Sagi, I., et al. (2022). Genetic variation associated with condensate dysregulation in disease. Dev Cell *57*, 1776-1788 e1778. 10.1016/j.devcel.2022.06.010.
124. Moses, D., Yu, F., Ginell, G.M., Shamoon, N.M., Koenig, P.S., Holehouse, A.S., and Sukenik, S. (2020). Revealing the Hidden Sensitivity of Intrinsically Disordered Proteins to their Chemical Environment. J Phys Chem Lett *11*, 10131-10136. 10.1021/acs.jpclett.0c02822.
125. Janin, J. (1979). Surface and inside volumes in globular proteins. Nature *277*, 491-492. 10.1038/277491a0.
126. Heinig, M., and Frishman, D. (2004). STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. Nucleic Acids Res *32*, W500-502. 10.1093/nar/gkh429.
127. Wang, P., Zhang, Q., Li, S., Cheng, B., Xue, H., Wei, Z., Shao, T., Liu, Z.X., Cheng, H., and Wang, Z. (2021). iCysMod: an integrative database for protein cysteine modifications in eukaryotes. Brief Bioinform *22*. 10.1093/bib/bbaa400.
128. Milo, R. (2013). What is the total number of protein molecules per cell volume? A call to rethink some published values. Bioessays *35*, 1050-1055. 10.1002/bies.201300066.
129. Eastman, P., Swails, J., Chodera, J.D., McGibbon, R.T., Zhao, Y., Beauchamp, K.A., Wang, L.P., Simmonett, A.C., Harrigan, M.P., Stern, C.D., et al. (2017). OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. PLoS Comput Biol *13*, e1005659. 10.1371/journal.pcbi.1005659.
130. Sies, H. (2017). Hydrogen peroxide as a central redox signaling molecule in physiological oxidative stress: Oxidative eustress. Redox Biol *11*, 613-619. 10.1016/j.redox.2016.12.035.
131. Zhang, H., and Forman, H.J. (2012). Glutathione synthesis and its role in redox signaling. Semin Cell Dev Biol *23*, 722-728. 10.1016/j.semcdb.2012.03.017.

132. Davies, M.J. (2016). Protein oxidation and peroxidation. Biochem J *473*, 805-825. 10.1042/BJ20151227.
133. Winterbourn, C.C., and Metodiewa, D. (1999). Reactivity of biologically important thiol compounds with superoxide and hydrogen peroxide. Free Radic Biol Med *27*, 322-328. 10.1016/s0891-5849(99)00051-9.
134. Fra, A., Yoboue, E.D., and Sitia, R. (2017). Cysteines as Redox Molecular Switches and Targets of Disease. Front Mol Neurosci *10*, 167. 10.3389/fnmol.2017.00167.
135. Szajewski, R.P.W., G.W. (1980). Rate constants and equilibrium constants for thiol-disulfide interchange reactions involving oxidized glutathione. JCIM *102*, 2011-2026.
136. Forman, H.J., Zhang, H., and Rinna, A. (2009). Glutathione: overview of its protective roles, measurement, and biosynthesis. Mol Aspects Med *30*, 1-12. 10.1016/j.mam.2008.08.006.
137. Zitka, O., Skalickova, S., Gumulec, J., Masarik, M., Adam, V., Hubalek, J., Trnkova, L., Kruseova, J., Eckschlager, T., and Kizek, R. (2012). Redox status expressed as GSH:GSSG ratio as a marker for oxidative stress in paediatric tumour patients. Oncol Lett *4*, 1247-1253. 10.3892/ol.2012.931.
138. Roussel, M.R., and Institute of, P. (2023). Foundations of chemical kinetics : a hands-on approach (IOP Publishing).
139. Dutagaci, B., Nawrocki, G., Goodluck, J., Ashkarran, A.A., Hoogstraten, C.G., Lapidus, L.J., and Feig, M. (2021). Charge-driven condensation of RNA and proteins suggests broad role of phase separation in cytoplasmic environments. Elife *10*. 10.7554/eLife.64004.
140. Hsieh, M.Y., Yang, S., Raymond-Stinz, M.A., Edwards, J.S., and Wilson, B.S. (2010). Spatio-temporal modeling of signaling protein recruitment to EGFR. BMC Syst Biol *4*, 57. 10.1186/1752-0509-4-57.
141. Gura Sadovsky, R., Brielle, S., Kaganovich, D., and England, J.L. (2017). Measurement of Rapid Protein Diffusion in the Cytoplasm by Photo-Converted Intensity Profile Expansion. Cell Rep *18*, 2795-2806. 10.1016/j.celrep.2017.02.063.
142. Schrödinger, L.D., W. (2020). PyMOL. Available at: http://www.pymol.org/pymol.

Chapter 3:
Transcription factors interact with RNA to regulate genes

Originally published in *Molecular Cell* 83, 2449–2463, July 20, (2023)

Ozgur Oksuz[1,13], Jonathan E Henninger[1,13], Robert Warneford-Thomson[2,3], Ming M Zheng[1,4], Hailey Erb[1], Adrienne Vancura[1], Kalon J Overholt[1,5], Susana Wilson Hawken[1,6], Salman F Banani[1,7], Richard Lauman[2,3], Lauren N. Reich[2,3], Anne L Robertson[8,10], Nancy M Hannett[1], Tong I Lee[1], Leonard I. Zon[8,9,10,11], Roberto Bonasio[2,3], Richard A. Young[1,12,14,*]

[1] Whitehead Institute for Biomedical Research, Cambridge MA 02142
[2] Epigenetics Institute, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA 19104
[3] Department of Cell and Developmental Biology, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA 19104
[4] Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139
[5] Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139
[6] Program of Computational & Systems Biology, Massachusetts Institute of Technology, Cambridge, MA 02139
[7] Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115
[8] Stem Cell Program, Division of Hematology/Oncology, Boston Children's Hospital and Dana Farber Cancer Institute, Boston, MA 02115,
[9] Harvard Medical School, Boston, MA 02115
[10] Howard Hughes Medical Institute, Boston, MA 02115
[11] Stem Cell and Regenerative Biology Department, Harvard University, Cambridge, MA 02138
[12] Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139
[13] Authors contributed equally
[14] Lead contact
* Correspondence: young@wi.mit.edu (R.A.Y.)

https://doi.org/10.1016/j.molcel.2023.06.012

Abstract

Transcription factors (TFs) orchestrate the gene expression programs that define each cell's identity. The canonical TF accomplishes this with two domains, one that binds specific DNA sequences and the other that binds protein coactivators or corepressors. We find that at least half of TFs also bind RNA, doing so through a previously unrecognized domain with sequence and functional features analogous to the arginine-rich motif of the HIV transcriptional activator Tat. RNA binding contributes to TF function by promoting the dynamic association between DNA, RNA and TF on chromatin. TF-RNA interactions are a conserved feature important for vertebrate development and disrupted in disease. We propose that the ability to bind DNA, RNA and protein is a general property of many TFs and is fundamental to their gene regulatory function.

Main Text

Transcription factors (TFs), which are encoded by ~1,600 genes in the human genome, comprise the single largest protein family in mammals. Each cell type expresses approximately 150-400 TFs, which together control the gene expression program of the cell[1–5]. TFs typically contain DNA-binding domains that recognize specific sequences and multiple TFs collectively bind to enhancers and promoter-proximal regions of genes[6,7]. The DNA-binding domains form stable structures whose conserved features are reliably detected by homology and are therefore used to classify TFs (e.g. C2H2 zinc finger, homeodomain, bHLH, bZIP) (Figure 1A)[1,2]. TFs also contain effector domains that exhibit less sequence conservation and sample many transient structures that enable multivalent protein interactions[8–10]. These effector domains recruit coactivator or corepressor proteins, which contribute to gene regulation through mechanisms that include mobilizing nucleosomes, modifying chromatin-associated proteins, influencing genome architecture, recruiting transcription apparatus and controlling aspects of transcription initiation and elongation[11,12]. This canonical view of TFs that function with two domains, one binding DNA and the other protein, has been foundational for models of gene regulation[13,14].

RNA molecules are produced at loci where TFs are bound, but their roles in gene regulation are not well-understood[15,16]. A few TFs and cofactors have been reported to bind RNA[17–28], but TFs do not harbor domains characteristic of well-studied RNA binding proteins[29]. We wondered whether TFs might have evolved to interact with RNA molecules that are pervasively present at gene regulatory regions but harbor a heretofore unrecognized RNA-binding domain. Here we present evidence that a broad spectrum of TFs do bind RNA molecules, that TFs accomplish this with a domain analogous to the RNA-binding arginine-rich motif of the HIV Tat transactivator, and that this domain promotes TF occupancy at regulatory loci. These domains are a conserved feature important for vertebrate development, and they are disrupted in cancer and developmental disorders.

Transcription factor binding to RNA in cells

Using nuclei isolated from human K562 cells, we performed a high throughput RNA-protein crosslinking assay (RNA-binding region identification - RBR-ID), which uses UV crosslinking and mass spectrometry to detect angstrom-scale crosslinks, typically thought to reflect direct interactions[30], between protein and RNA molecules in cells[31] (Figure 1B). The results included the expected distribution of peptides from known RNA-binding proteins (RBPs) and revealed that a broad distribution of TFs had peptides crosslinked to RNA in this assay independent of their cellular abundance (Figures 1C, 1D, and S1A). Nearly half (48%) of TFs identified in the RBR-ID dataset showed evidence of RNA binding in K562 cells (Figure S1B) when the analysis was conducted using thresholds that retain RBPs verified by independent methods[31] (Table S1). These results prompted a re-examination of previously published RBR-ID data for murine embryonic stem cells (ESCs)[31] which confirmed that a substantial fraction of TFs (41%) in those cells also bind RNA (Figures S1C-1E, and Table S2). A meta-analysis of data from multiple studies using proteomics to identify RNA-binding proteins, including data collected in this study, provides an extensive list of RNA-binding TFs (Table S3).

Specific TFs are notable for their roles in control of cell identity and have been subjected to more extensive study than others. Many well-studied TFs that contribute to the control of cell identity were observed among the TFs that showed evidence of RNA binding. In K562 hematopoietic cells, these included GATA1, GATA2, and RUNX1, which play major roles in regulation of hematopoietic cell genes[32], as well as MYC and MAX, oncogenic regulators of

these tumor cells[33] (Figure 1C). In the ESCs, these included the master pluripotency regulators Oct4, Klf4, and Nanog, as well as the MYC family member that is key to proliferation of these cells, Mycn[34] (Figure S1D). The RNA-binding TFs also included those involved in other important cellular processes, including regulation of chromatin structure (CTCF, YY1) and response to signaling (CREB1, IRF2, ATF1) (Figure 1C). It was notable that RNA binding was a property of TFs that span many TF families (Figures S1F and S1G). These results suggest that RNA binding is a property shared by TFs that participate in diverse cellular processes and that possess diverse DNA-binding domains.

We next sought to identify the RNAs that interact with specific TFs. We conducted CLIP for the TF GATA2, a major regulator of hematopoietic genes in K562 cells that showed evidence of RNA binding in our RBR-ID data (Figure 1C). Immunoprecipitation of HA- and FLAG-tagged GATA2 in K562 cells subjected to UV cross-linking showed that GATA2 interacts with RNA in cells in a 4SU-dependent manner (Figure S2A). Interacting RNAs were then sequenced and cross-linked sites were identified with nucleotide resolution (Figure S2B, Table S4, STAR Methods). A diversity of RNA species were bound by GATA2, including many enhancer- and promoter-derived RNAs (Table S4). We reasoned that GATA2 may interact with RNAs transcribed in proximity to regions where GATA2 binds chromatin to regulate genes. Indeed, as illustrated for a specific locus, GATA2 binds chromatin at the *HINT1* gene measured by ChIP-seq, and GATA2 interacts with RNA transcribed from the *HINT1* gene measured by CLIP-seq (Figure 1E). A metagene analysis revealed that GATA2 CLIP signal was enriched at GATA2 ChIP-seq peaks (Figure 1F). Enrichment of GATA2 CLIP signal was not evident at ChIP-seq peaks of RUNX1, another major regulator of hematopoietic genes (Figure 1F). These results prompted a re-examination of previously published CLIP/ChIP data for RBR-ID$^+$ YY1 and CTCF[21,35,36], which also showed that these TFs interact with RNAs transcribed from loci near their chromatin-binding sites (Figures S2C and S2D). These results suggest that TFs bind to RNAs produced in the vicinity of their DNA-binding sites.

Transcription factor binding to RNA in vitro

To corroborate evidence that TFs can bind RNA molecules in cells, we sought to confirm that purified TFs bind RNA molecules in vitro using a fluorescence polarization assay (Figure 2A, STAR Methods). The assay was validated with multiple control proteins with an RNA of random sequence, including three well-studied RNA-binding proteins (U2AF2, HNRNPA1, and SRSF2) and proteins that were not expected to have substantial affinity for RNA (GFP and the DNA-binding restriction enzyme BamHI). The RBPs bound RNA with nanomolar affinities, consistent with previous studies[37–40], whereas GFP and BamHI showed little affinity for RNA (K_d > 4 μM) (Figure 2B). We then selected 13 TFs that showed evidence of crosslinking to RNA in cells, are well-studied for their diverse cellular functions and are members of different TF families, purified them from human cells and measured their RNA-binding affinities. These TFs exhibited a range of binding affinities for the RNA, ranging from 41 to 505 nM, which is remarkably similar to the range of affinities measured for known RBPs (42 to 572 nM) (Figure 2C). Thus, a diverse set of TFs can bind RNA with affinities similar to proteins with known physiological roles in RNA processing. The thousands of enhancers and promoter-proximal regions where TFs bind have diverse sequences, and thus RNA molecules produced from these sites differ in sequence, so we investigated whether TFs bind diverse RNA sequences. Six TFs were investigated, and the results indicate that these TFs do bind various RNA sequences with similar affinities (Figures S2E and S2F).

An arginine-rich domain in transcription factors

We next sought to identify regions in TFs that contribute to RNA binding. TFs do not contain sequence motifs that resemble those of structured RNA-binding domains[29,38] (Figures S3A and S3B), so we searched for local amino acid features that might be common to TFs. Nearly 80% of TFs were found to have a cluster of basic residues (R/K) adjacent to their DNA-binding domain (Figure 3A). Derivation of a position-weight matrix from these “basic patches” revealed that they contain a sequence motif similar to the RNA-binding domain of the HIV Tat transactivator, which has been termed the arginine-rich motif (ARM)[41,42] (Figure 3B). These ARM-like domains were enriched in TFs compared to the remainder of the proteome (Figure 3C). Furthermore, the ARM-like domains have sequences that are evolutionarily conserved and appear adjacent to diverse types of DNA-binding domains, as illustrated for KLF4, SOX2, and GATA2 (Figures 3D, S3C, and S3D). This analysis suggests that TFs often contain conserved ARM-like domains, which we will refer to hereafter as TF-ARMs.

To investigate whether TF-ARMs are necessary for RNA binding, we purified wild-type and deletion mutant versions of KLF4, SOX2 and GATA2 and compared their RNA binding affinities. The 7SK RNA was used in this assay because it is one of a number of RNA species known to be bound by HIV Tat[43]. RNA binding by the ARM-deleted proteins was substantially reduced (Figure 3E). To determine if the TF-ARMs are sufficient for RNA binding, peptides containing the HIV Tat ARM and TF-ARMs were synthesized and their ability to bind 7SK RNA was investigated using an electrophoretic mobility shift assay (EMSA). The results showed that all the TF-ARM peptides can bind 7SK RNA, as did the control HIV Tat ARM peptide (Figure 3F). This binding was dependent on arginine and lysine residues within the TF-ARMs (Figure 3F), as has been previously demonstrated for the Tat ARM[41,43]. These results indicate that TF-ARMs are necessary and sufficient for RNA binding.

We considered the possibility that the TF-ARM also contributes to DNA-binding. Synthesized peptides of the SOX2 and KLF4 ARMs were tested for binding to either DNA or RNA. The results show that both ARMs bind RNA with greater affinity compared to DNA (Figures S4A and S4B). Full-length wildtype and ARM-deleted SOX2 and KLF4 were also tested for binding to motif-containing DNA. The results show that deletion of the SOX2 ARM did not affect DNA-binding (Figure S4C). Deletion of the KLF4 ARM did affect DNA-binding (Figure S4D), although not to the extent that it affected RNA binding (Figure 3E). It thus appears possible that some TF-ARMs can contribute to DNA-binding to some extent whereas others do not.

Having found that TF-ARMs bind to RNA in vitro in assays with purified components, we next asked whether TF-ARMs bind RNA in the more complex environment of the cell. To investigate this, we analyzed the RBR-ID data (Figures 1B-D), which can provide spatial information on the regions of proteins that bind RNA in cells. If TF-ARMs were binding to RNA in cells, then we would expect an enrichment of RBR-ID$^+$ peptides overlapping or adjacent to the TF-ARMs. Global analysis of RBR-ID$^+$ peptides in human K562 cells, as well as inspection of RBR-ID$^+$ peptides for individual TFs, confirmed that this was the case (Figure S5). These results provide evidence that ARM-like regions in TFs bind to RNA in cells.

To investigate if TF-ARMs could function similarly to the Tat ARM in cells, we tested whether TF-ARMs could replace the Tat ARM in a classical Tat transactivation assay[41]. In this assay, the HIV-1 5’ long terminal repeat (LTR) is placed upstream of a luciferase reporter gene. Transcription of the LTR generates an RNA stem loop structure called the Trans-activation Response (TAR), and HIV Tat binds to the TAR RNA to stimulate expression of the reporter gene[44] (Figure 3G). We confirmed that expression of full-length Tat stimulates luciferase

expression, and that mutation of the lysines and arginines in the Tat ARM reduces this activity (Figure 3H). Replacing the Tat ARM with the TF-ARMs of KLF4, SOX2, or GATA2 rescued the loss of the Tat ARM (Figure 3H). In all cases, activation was dependent on the TAR RNA bulge structure, which is required for Tat binding[44] (Figure 3H). These results indicate that the TF-ARMs can perform the functions described for the Tat ARM and activate gene expression in an RNA-dependent manner.

TF-ARMs enhance TF chromatin occupancy and gene expression

TFs bind enhancer and promoter elements in chromatin and regulate transcriptional output, so it is possible that RNA binding, enabled by TF-ARMs, contributes to chromatin occupancy and gene expression. We investigated whether TF-ARMs contributed to TF association with chromatin by measuring the relative levels of TFs in chromatin and nucleoplasmic fractions from ES cells containing HA-tagged TFs with wild-type and mutant ARMs. Genome-wide localization of KLF4 and SOX2 was globally reduced upon deletion of their ARMs (Figure 4A) as determined by CUT&Tag and illustrated for specific genes regulated by KLF4 or SOX2 (Figure 4B). Nuclear fractionation confirmed that deletion of the ARMs reduced the levels of KLF4 and SOX2 in chromatin (Figures S6A and S6B), and treatment of the extracts with RNase reduced TF enrichment in the chromatin fraction (Figures S6C and S6D). These results are consistent with a model whereby TF-RNA interactions enhance the association of TFs with chromatin.

We next sought to determine whether TF-ARMs contribute to gene output by using a transcriptional reporter assay that has been used extensively to investigate the functions of domains in TFs that contribute to transcriptional output[8]. KLF4 was selected for study because previous studies have used this assay to study KLF4 function in various cellular contexts[45–47], KLF4 has a single ARM-like domain (Figures 4C and 4D), it has contiguous effector and DNA-binding domains, and our assays show that deletion of the ARM has a strong effect on RNA binding (Figure 3E). In this assay, the KLF4 zinc fingers (DBD) were replaced with the yeast GAL4 DBD, and this fusion was tested for its ability to activate expression of a luciferase reporter downstream of GAL4-binding UAS sites (Figure 4E). GAL4-KLF4^{WT} activated reporter expression, while substitution of arginines and lysines for alanines in the ARM (GAL4-KLF4$^{R/K>A}$) significantly reduced reporter expression (Figure 4F). Importantly, this reduction was rescued by replacement of the ARM with the HIV Tat ARM (Figure 4F). Similar effects were observed with the replacement of KLF4 DBD with the bacterial TetR DBD, which recognizes TetO elements in the presence of doxycycline (Figures 4E and 4F). The mutation of the KLF4 ARM caused a reduction in reporter expression rather than complete ablation of expression. These results, taken together with previous studies[45–47], suggest that while the DNA and protein binding portions of the TF play major roles in gene activation, TF-RNA binding contributes to fine-tune transcriptional output.

A role for TF RNA-binding regions in TF nuclear dynamics

TFs are thought to engage their enhancer and promoter DNA-binding sites through search processes that involve dynamic interactions with diverse components of chromatin. Single molecule image analysis of TF dynamics in cells indicates that TFs conduct a highly dynamic search for their binding sites in chromatin[48,49]. The tracking data can be fit to a three-state model, where TFs are interpreted to be immobile (potentially DNA-bound), subdiffusive (potentially interacting with chromatin components) and freely diffusing[50,51]. If TFs interact with chromatin-associated RNA through their ARMs, then we might expect that mutation of their

ARMs would reduce the portion of TF molecules in the immobile and sub-diffusive states. To test this, we conducted single-molecule tracking experiments with murine embryonic stem cell (mESC) or human K562 leukemia lines that enable inducible expression of Halo-tagged wildtype or ARM-mutant TFs. For these experiments, we chose the TFs SOX2, KLF4, GATA2, and RUNX1 because of their prominent roles in mES or hematopoietic cells[32,34] and our earlier characterization of their RNA-binding regions (Figure 3). As a control, we included the deletion of an ARM-like region from CTCF that overlaps the previously described RNA-binding region (RBR)[36], which was shown to reduce both the immobile and subdiffusive fractions of CTCF[52]. Single-molecule imaging data was fit to a three-state model: immobile, subdiffusive, and freely diffusing (Figures 5A, S7A-S7C, Videos S1-S3, and STAR Methods). Inspection of single-molecule traces for wildtype and ARM-mutant TFs (Figures 5B and S7A), as well as global quantification across replicates (Figures 5C, S7D, and S7E), showed that deletion of the ARM-like domains in TFs reduces the fraction of molecules in the subdiffusive fraction for all factors and immobile fraction for all factors but one (GATA2), while increasing the fraction of freely diffusing molecules. Although diffusive fractions changed with expression level, the behavior of the mutant TF was consistent across expression regimes (Figures S7F). The observed changes in diffusivity upon ARM mutation could reflect changes in binding between TFs and RNA or DNA molecules. The observation that ARM peptides have a preference for RNA binding (Figure S4), and evidence that TF chromatin occupancy is reduced upon RNase treatment or ARM mutation (Figure S6), is consistent with a role for RNA interactions in TF nuclear dynamics. These results suggest that TF-ARMs enhance the timeframe in which TFs are associated with chromatin.

TF-ARMs are important for normal development and disrupted in disease

Transcription factors are fundamental controllers of cell-type specific gene expression programs during development, so we next asked whether the TF-ARMs contribute to the factor's role in normal development in vivo. For this purpose, we turned to the zebrafish, which has served as a valuable model system to study and perturb vertebrate development. Previous study showed that knockdown of zebrafish *sox2* by injection of antisense morpholinos at the one-cell stage led to growth defects and embryonic lethality, which could be rescued by co-injection with messenger RNA (mRNA) encoding human SOX2[53]. Using this system, we injected zebrafish with the *sox2* morpholino while co-injecting mRNA encoding either wildtype or ARM-mutant human SOX2 (Figures 6A and S7G), which reduced RNA but not DNA binding in vitro (Figures 3E and S4C). Embryos were scored at 48 hours post-fertilization for growth defects by the length of the anterior-posterior axis compared to embryos injected with a non-targeting control morpholino (Figure 6B). Whereas wildtype human SOX2 could partially rescue the growth defect induced by *sox2* knockdown, ARM-mutant SOX2 was unable to do so (Figures 6C). These results indicate that TF-ARMs contribute to proper development.

The presence of ARMs in most TFs, and evidence that they can contribute to TF function in a developmental system, prompted us to investigate whether pathological mutations occur in these sequences in human disease. Analysis of curated datasets of pathogenic mutations revealed hundreds of disease-associated missense mutations in TF-ARMs (Figure 6D, Table S5, STAR Methods). These mutations are associated with both germline and somatic disorders, including multiple cancers and developmental syndromes, that affect a range of tissue types (Figure 6E). Variants that mutate arginine residues were the most enriched compared to the other amino acid residues in ARMs (STAR Methods), which is consistent with their importance in RNA binding (Figure 6F) [42]. To confirm that such mutations could affect RNA binding, we selected for further study the estrogen receptor (ESR1) R269C mutation (Figure 6G), which is found in multiple cancers and is particularly enriched in a subset of patients with pancreatic

cancer[54]. An EMSA assay showed that RNA binding was reduced with an ESR1 ARM peptide containing the R269C mutation (Figure 6H). Furthermore, when the Tat ARM was replaced with wildtype and mutant versions of the ESR1 ARM in the Tat transactivation assay, the mutation caused reduced reporter expression compared to wildtype (Figure 6I). These results support the hypothesis that disease-associated mutations in TF-ARMs can disrupt TF RNA binding.

DISCUSSION

The canonical view of transcription factors is that they guide the transcription apparatus to genes and control transcriptional output through the concerted function of domains that bind DNA and protein molecules[1,3,55,56]. The evidence presented here suggests that many transcription factors also harbor RNA-binding domains that contribute to gene regulation (Figure 7A). Given the large portion of TFs that showed evidence of RNA interaction in cells and the presence of an ARM-like sequence in nearly 80% of TFs, it is possible that the majority of TFs engage in RNA binding.

RNA molecules are pervasive components of active transcriptional regulatory loci[15,16,57–59] and have been implicated in the formation and regulation of spatial compartments[60]. The noncoding RNAs produced from enhancers and promoters are known to affect gene expression[15], and plausible mechanisms by which these RNA species could influence gene regulation have been proposed to include binding to cofactors and chromatin regulators[61–64], and electrostatic regulation of condensate compartments[58]. The evidence that TFs bind RNA suggests additional functions for RNA molecules at enhancers and promoters (Figures 7B and 7C).

Transcription and processing of RNA is a highly localized and dynamic process, producing high local concentrations of RNA at active loci. RNA molecules transcribed by Pol II will typically undergo rapid capping and splicing while tethered to Pol II[65]. In some cases, RNA molecules accumulate in proximity to the loci where they are transcribed[60], but in others they are rapidly exported into the cytoplasm once fully processed. This high local concentration of RNA molecules at sites of transcription would be expected to provide a multivalent interaction network between TFs, DNA and RNA and thereby influence the recruitment and dynamics of TFs at these sites (Figures 7B and 7C). Indeed, previous studies have shown that tethering of RNA molecules to modestly active sites in the genome will enhance the concentration of certain TFs at those loci[21].

The observation that many TFs can bind DNA, RNA and protein molecules offers new opportunities to further advance our understanding of gene regulation and its dysregulation in disease. Knowledge that TFs can interact with both DNA and RNA molecules may help with efforts to decipher the “code” by which multiple TFs collectively bind to specific regulatory regions of the genome[66–68] and inspire novel hypotheses that may provide additional insight into gene regulatory mechanisms. It might also provide new clues to the pathogenic mechanisms that accompany GWAS variants in enhancers, where those variations occur in both DNA and RNA.

This study shows that many transcription factors bind RNA and harbor RNA-binding domains that resemble the HIV Tat ARM. Our results demonstrate for a few tested examples that these domains contribute to the dynamic association of TFs with chromatin, which may provide a mechanism by which TF-RNA interactions contribute to gene control. Although the observed changes in diffusivity of TFs upon ARM mutation was consistent across expression regimes, we cannot exclude the possibility that expression level itself affects TF diffusivity and could explain

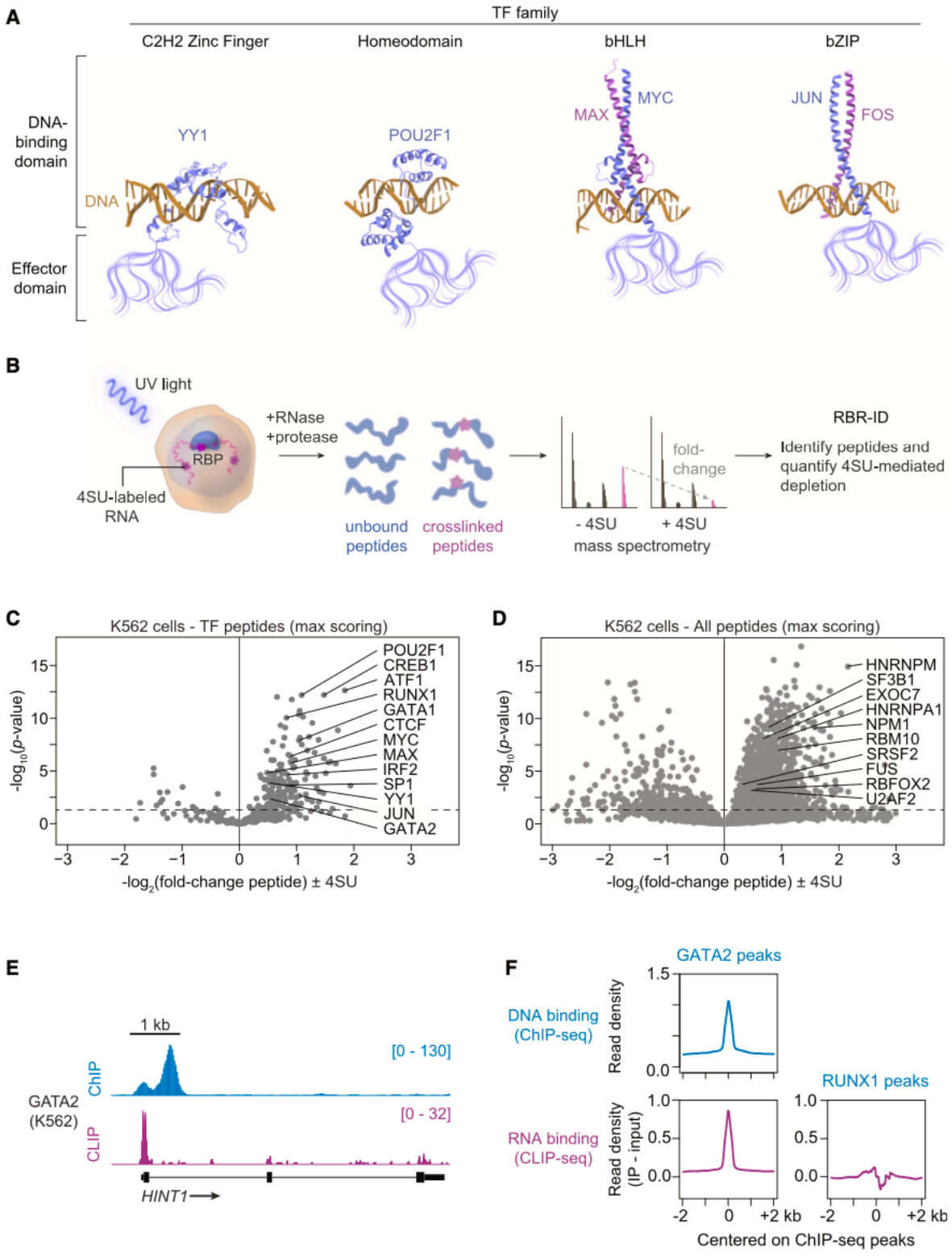
some of these changes. There are several ways in which the binding of TFs to RNA could affect their function (Figures 7B and 7C), and these mechanisms could result in positive or negative effects on transcriptional output. It is also possible that these domains have additional RNA-dependent functions, some of which may be general and some TF-specific[69]. Another limitation of the study is the extent to which cellular and organismal phenotypes observed upon deletion of ARM-like domains can be attributed to RNA binding. We believe that characterization of these domains in TFs, including systematic identification of the precise residues required for RNA binding and RNA sequence preferences, will inspire investigation of their roles in many aspects of TF function, including but not limited to locus-specific chromatin association, chromatin architecture, transcriptional output, splicing, translational control, and RNA polymerase II pausing. A key challenge will be to delineate these functions in cells and explore how these functions are related to cooperative or competitive interactions of these domains with RNA, DNA or proteins.

ACKNOWLEDGEMENTS

We are grateful to Phillip Sharp, Amy Gladfelter, Seychelle Vos, and Ibrahim Cissé for discussions regarding HIV Tat, RNA-binding proteins, RNA interactions, and single-molecule imaging. We thank L.D. Lavis (HHMI, Janelia) for the gift of the Halo Tag-(PA)-JF549 dyes. This work was supported by NIH grants GM144283 (R.A.Y), CA155258 (R.A.Y.), NSF PHY2044895 (R.A.Y.), F32CA254216 (J.E.H.), GM138788 (R.B.), GM127408 (R.B.), R01 HL144780-01 (L.I.Z.), R24 OD017870-01 (L.I.Z), CA251062 (S.F.B.) and NSF Graduate Research Fellowship 1745302 (K.J.O.).

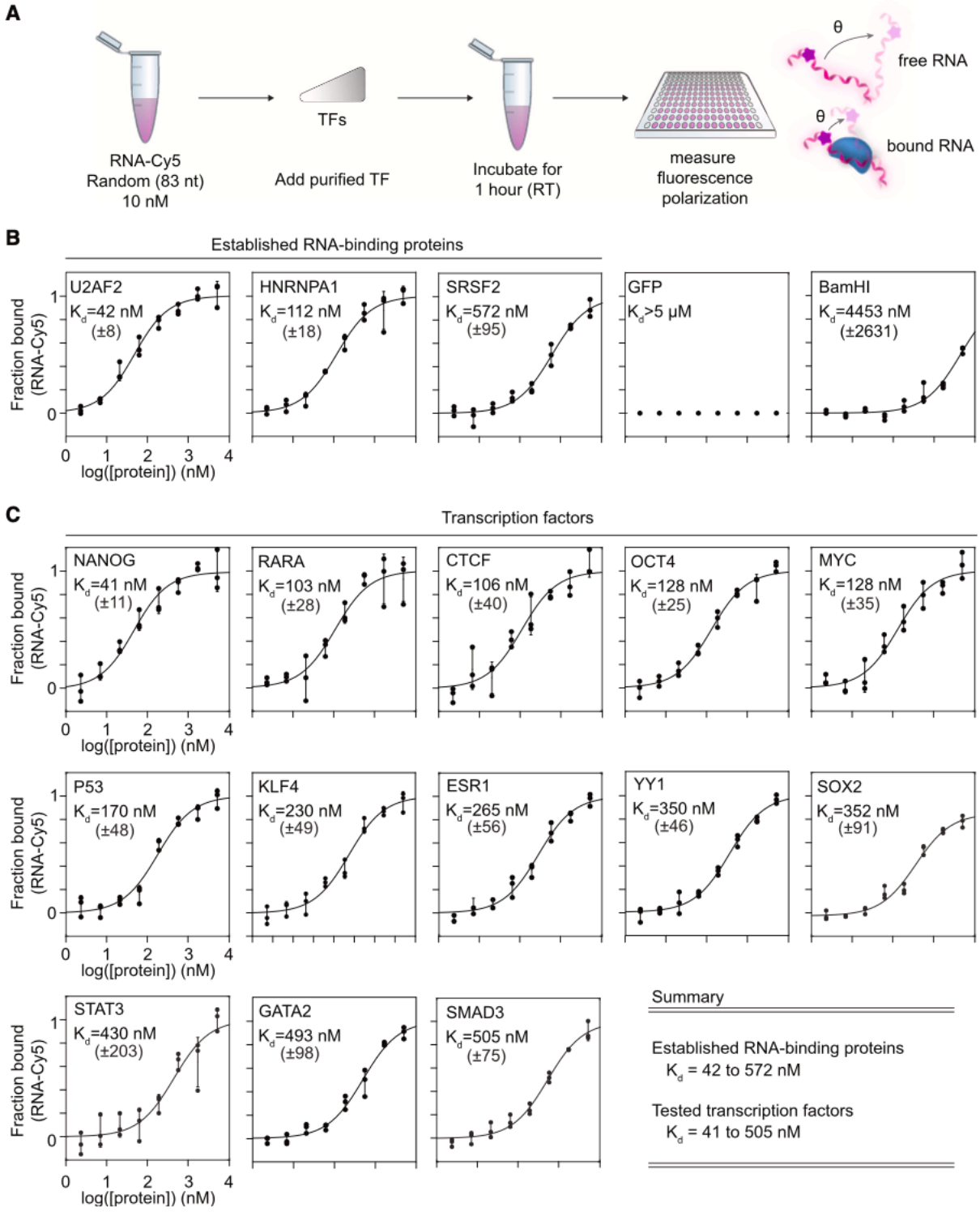
Figures and Tables

Figure 1. Transcription factor binding to RNA in cells



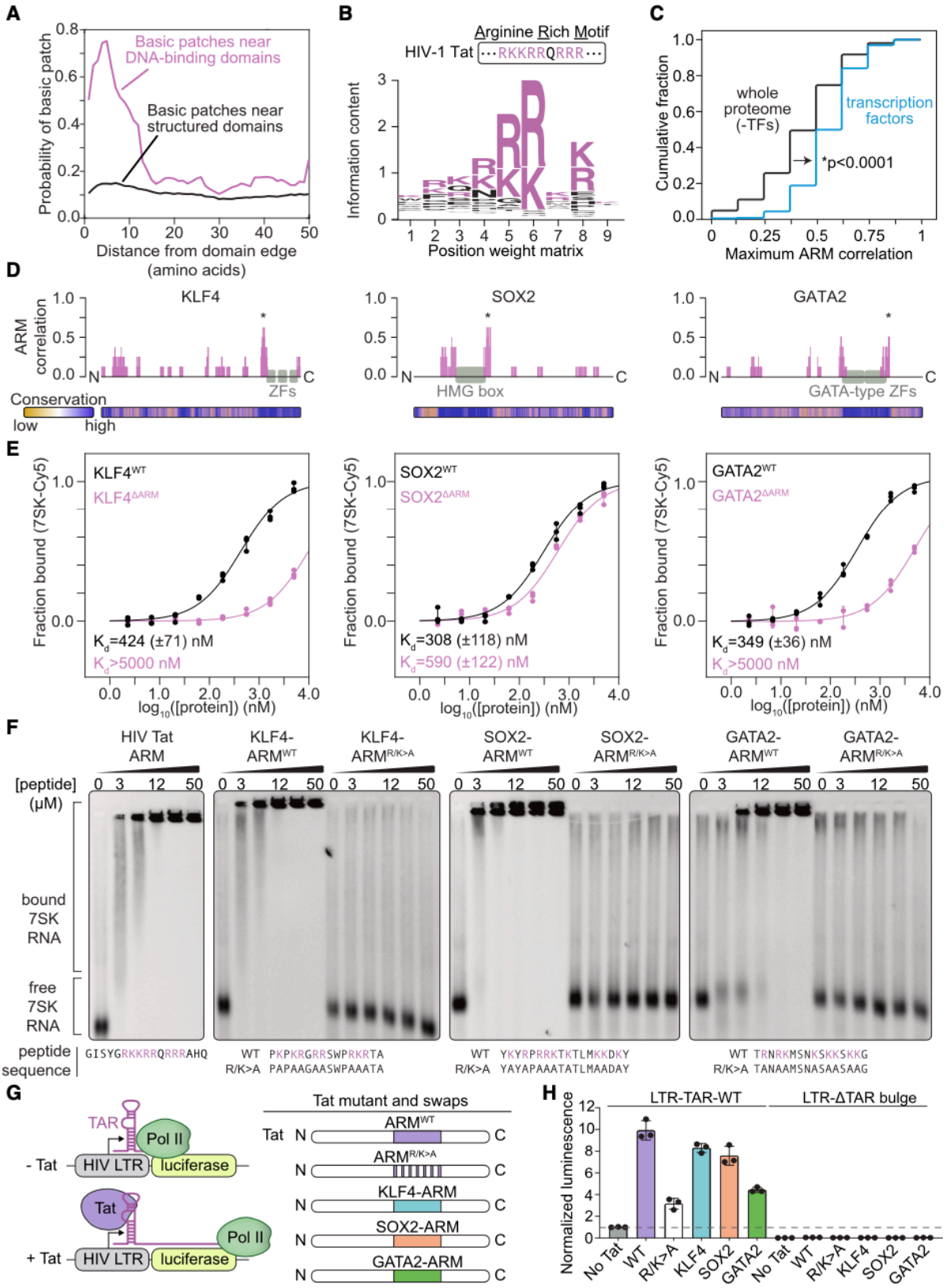
(A) Schematic of DNA-binding and effector domains in transcription factors from different families (PDB accession numbers in Methods).
(B) Experimental scheme for RBR-ID in human K562 cells. 4SU-labeled RNAs are crosslinked to proteins with UV light. RNA-binding peptides are identified by comparing the levels of crosslinked and unbound peptides by mass spectrometry.
(C) Volcano plot of TF peptides in RBR-ID for human K562 cells with select highlighted TFs (dotted line at p=0.05). Each marker represents the peptide with maximum RBR-ID score for each protein.
(D) Volcano plot of all detected peptides in RBR-ID for human K562 cells with select highlighted RBPs (dotted line at p=0.05). Each marker represents the peptide with maximum RBR-ID score for each protein.
(E) ChIP-seq and CLIP signal for GATA2 at the *HINT1* locus in K562 cells.
(F) Meta-gene analysis of input-subtracted CLIP signal centered on GATA2 or RUNX1 ChIP-seq peaks in K562 cells.

Figure 2. Transcription factor binding to RNA in vitro.



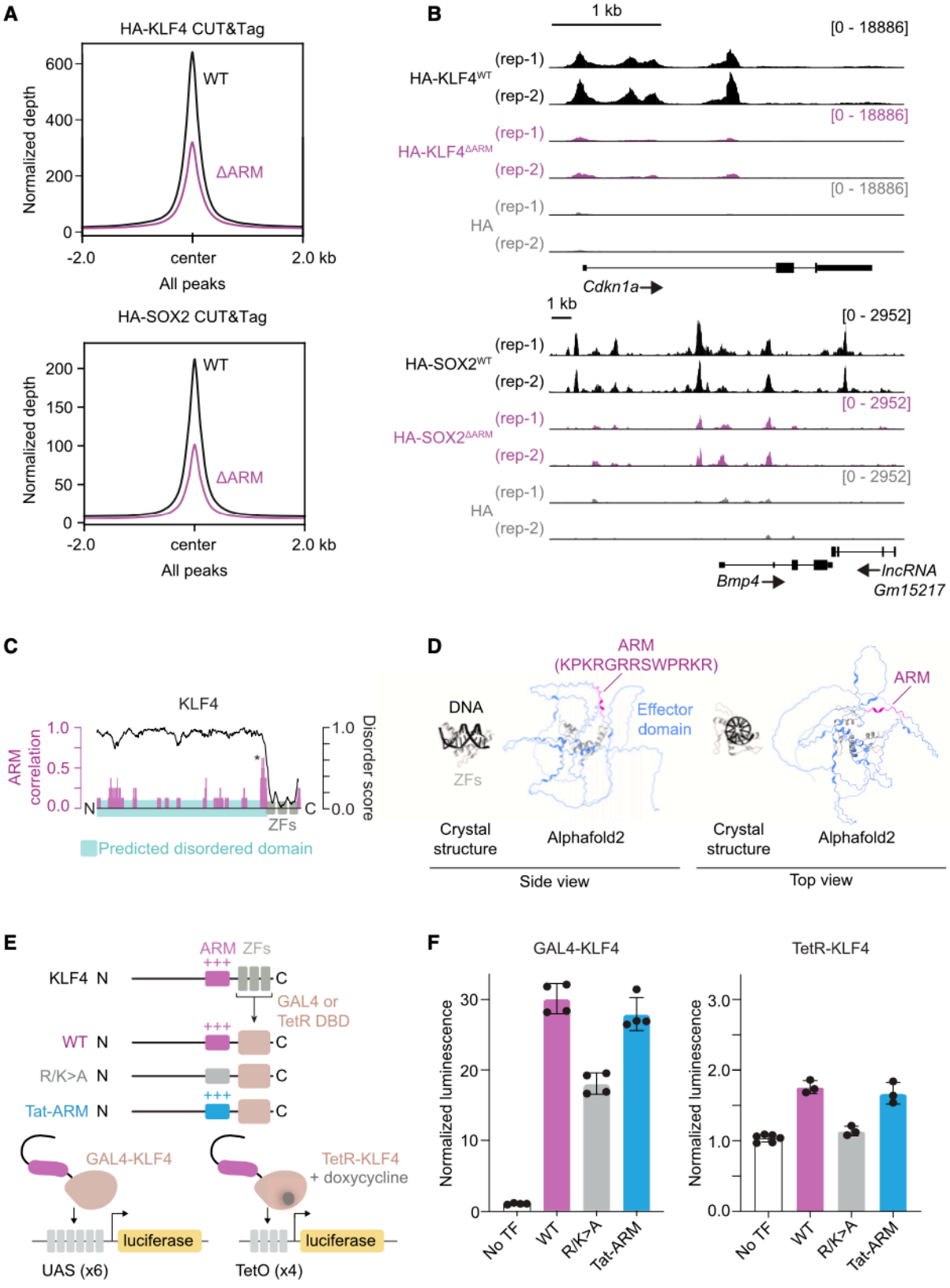
(A) Experimental scheme for measuring the equilibrium dissociation constant (K_d) for protein-RNA binding. Cy5-labeled RNA and increasing concentrations of purified proteins are incubated and protein-RNA interactions is measured by fluorescence polarization assay.
(B) Fraction bound RNA with increasing protein concentration for established RNA-binding proteins, GFP, and the restriction enzyme BamHI (error bars depict s.d.).
(C) Fraction bound RNA with increasing protein concentration for select transcription factors (error bars depict s.d.). A summary of K_d values for established RNA-binding proteins and TFs are indicated.

Figure 3. An arginine-rich domain in transcription factors.



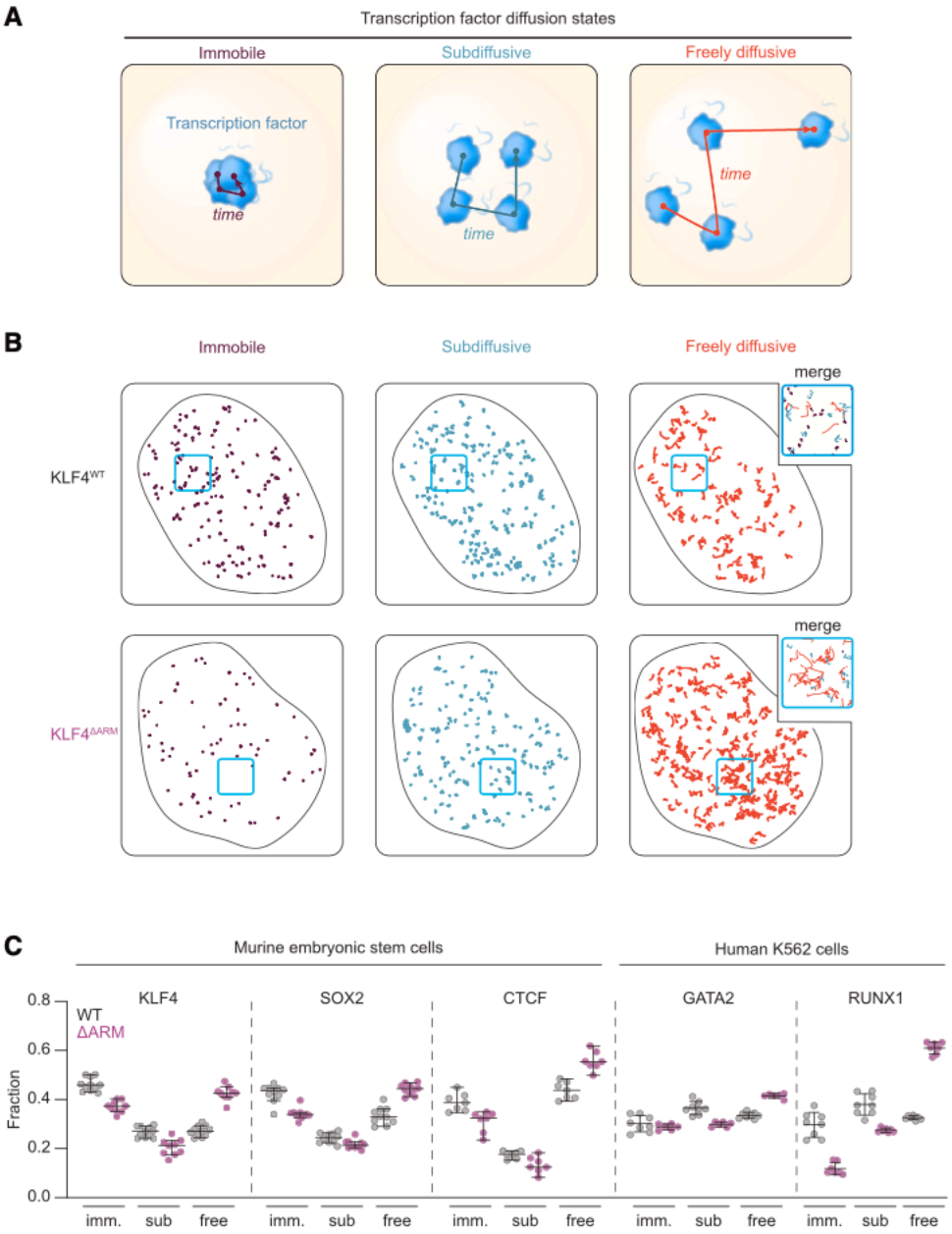
(A) Plot depicting the probability of a basic patch as a function of the distance from either DNA-binding domains (magenta) or all other annotated structured domains (black).
(B) Sequence logo derived from a position-weight matrix generated from the basic patches of TFs.
(C) Cumulative distribution plot of maximum cross-correlation scores between proteins and the Tat ARM (*p < 0.0001, Mann Whitney U test) for the whole proteome excluding TFs (black line) or TFs alone (blue line).
(D) Diagram of select TFs and their cross-correlation to the Tat ARM across a sliding window (*maximum scoring ARM-like region). Evolutionary conservation as calculated by ConSurf (Methods) is provided as a heatmap below the protein diagram.
(E) Fraction bound RNA with increasing protein concentration for wildtype (WT) or deletion (ΔARM) TFs (KLF4 WT vs ΔARM: p=0.017; SOX2 WT vs ΔARM: p=0.0012; GATA2 WT vs ΔARM: p=0.018).
(F) Gel shift assay for 7SK RNA with synthesized peptides encoding wildtype or R/K>A mutations of TF-ARMs.
(G) Experimental scheme for Tat transactivation assay. RNA Pol II transcribes the luciferase gene in the presence of Tat protein and bulge-containing TAR RNA. Indicated TF-ARMs are tested for their ability to replace Tat ARM.
(H) Bar plots depicting the normalized luminescence values for the Tat transactivation assay with or without the TAR RNA bulge with the indicated TF-ARM replacements. Values are normalized to the control condition (p_{adj}<0.0001 for Tat RK>A compared to No Tat, WT Tat, KLF4, SOX2, and all conditions with TAR deletion; p_{adj} = 0.0086 for Tat RK>A compared to GATA2, Sidak multiple comparison test).

Figure 4. TF-ARMs enhance chromatin occupancy and gene expression



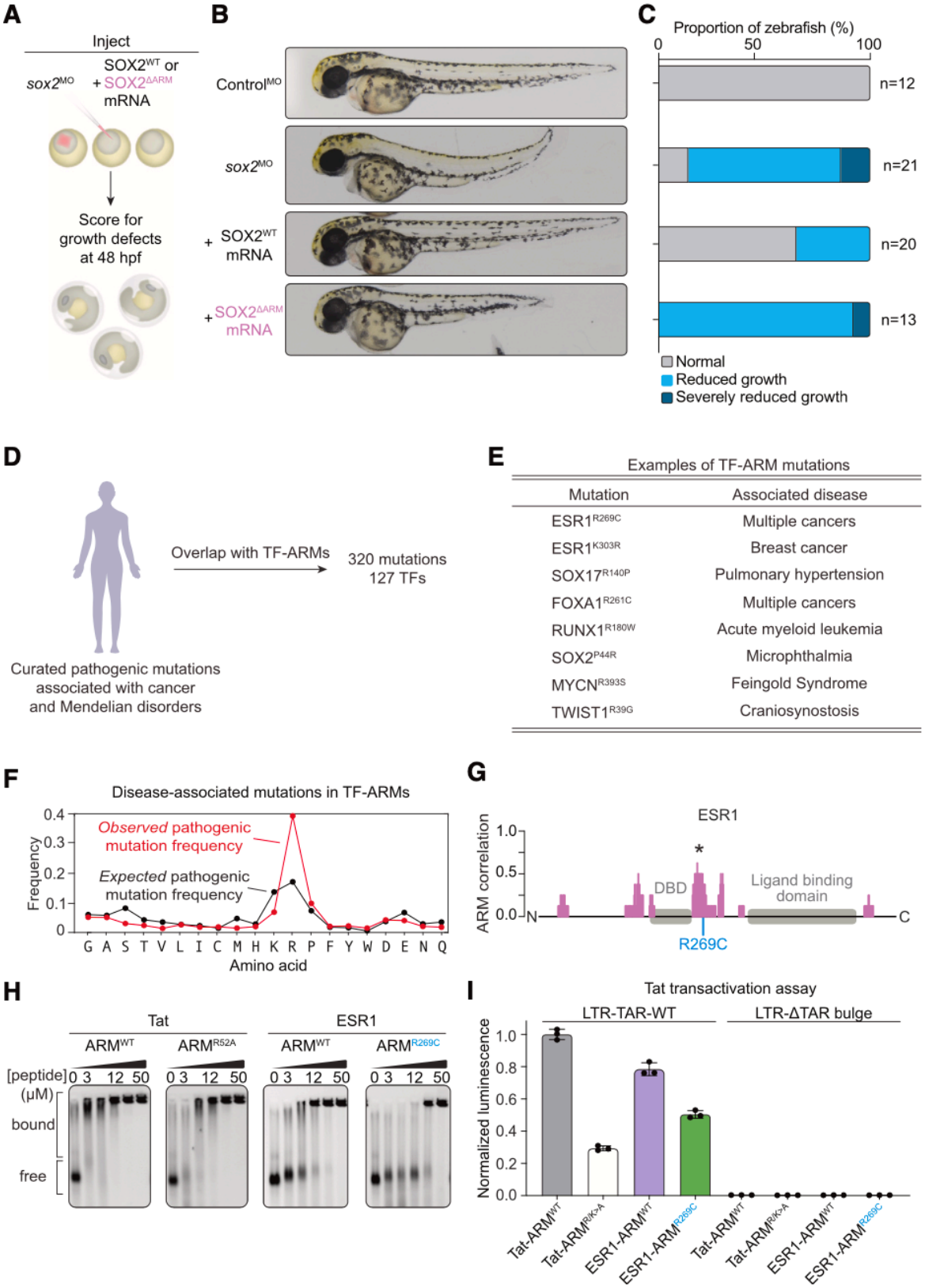
(A) Meta-gene analysis of CUT&Tag for WT or ΔARM HA-tagged KLF4 or SOX2, centered on called WT peaks in mESCs
(B) Example tracks of CUT&Tag (spike-in normalized) at specific genomic loci.
(C) Diagram of KLF4 and its cross-correlation to the Tat ARM (magenta), predicted disorder (black line), DNA-binding domain (grey boxes) and predicted disordered domain (cyan).
(D) Side and top views of the crystal structure of KLF4 with DNA (PDB: 6VTX) or AlphaFold predicted structure (ID: O43474)
(E) Experimental scheme for TF gene activation assays. KLF4 ZFs are replaced either by GAL4 or TetR DBD. The effect of KLF4-ARM mutation or replacement of KLF4-ARM with Tat-ARM on gene activation is tested by UAS or TetO containing reporter system.
(F) Normalized luminescence of gene activation assays, normalized to the “No TF” condition (error bars depict s.d., GAL4: $p<0.0001$ for all pairwise comparisons except WT vs. Tat-ARM, $p=0.3363$; TetR: NoTF vs. WT, $p<0.0001$, NoTF vs. R/K>A, $p=0.5668$, NoTF vs. Tat-ARM, $p=0.0002$, WT vs. R/K>A, $p=0.0003$, WT vs. Tat-ARM, $p=0.7126$, Tat-ARM vs. R/K>A, $p=0.0008$, one-way ANOVA)

Figure 5. A role for TF RNA-binding regions in TF nuclear dynamics.



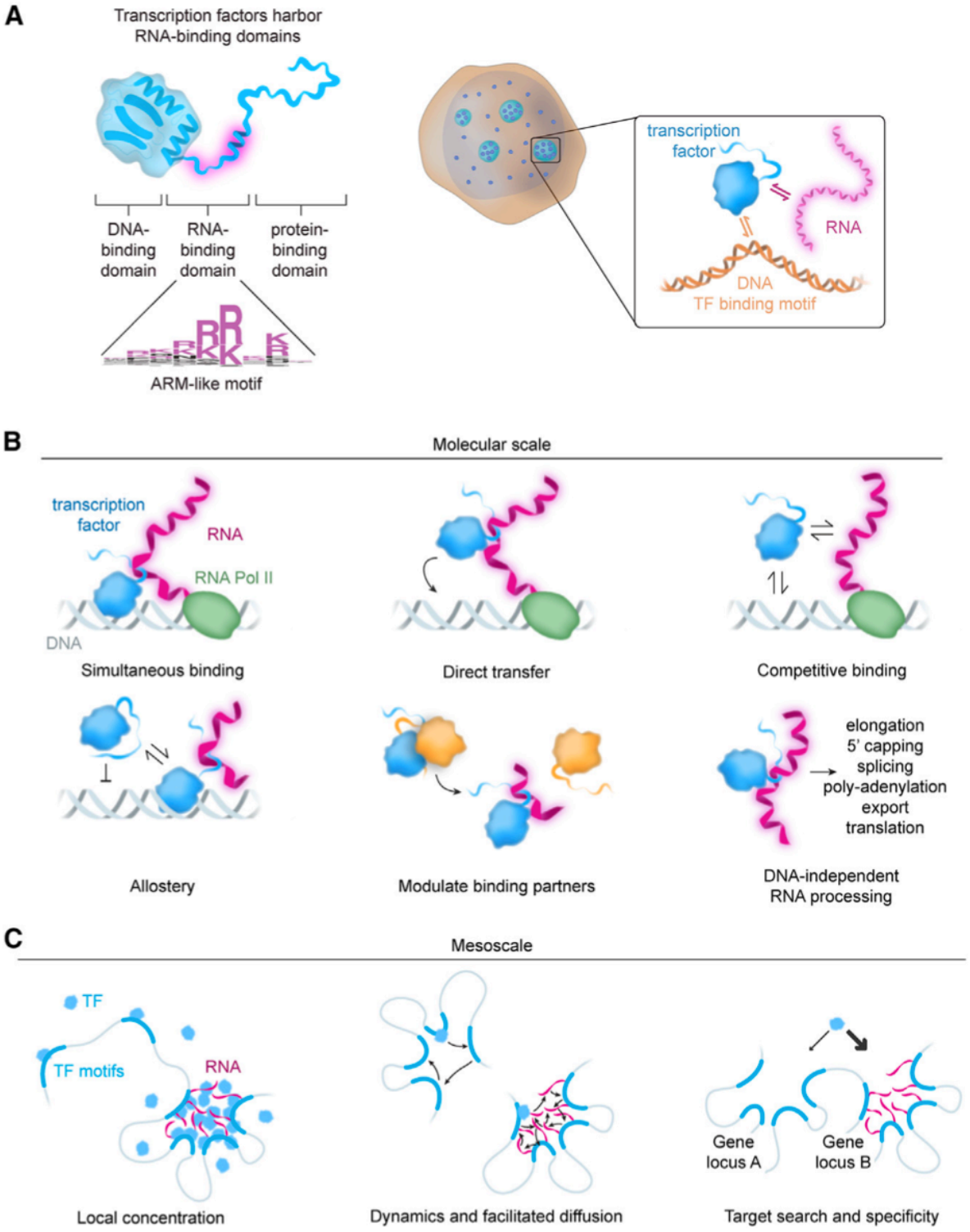
(A) Cartoon depicting a 3-state model of TF diffusion.
(B) Example of single nuclei single-molecule tracking traces for KLF4-WT and KLF4-ARM deletion. The traces are separated by their associated diffusion coefficient (D_{imm}: <0.04 $\mu m^2 s^{-1}$; D_{sub}: 0.04-0.2 $\mu m^2 s^{-1}$; D_{free}: >0.2 $\mu m^2 s^{-1}$). For each nucleus, 500 randomly sampled traces are shown.
(C) Dot plot depicting the fraction of traces in the immobile, subdiffusive, or freely diffusing states. Each marker represents an independent imaging field (comparing WT and ARM-deletion, p<0.0001 for $KLF4^{free}$, $SOX2^{free}$, $CTCF^{free}$, $GATA2^{free}$, $RUNX1^{free}$, $KLF4^{sub}$, $GATA2^{sub}$, $RUNX1^{sub}$, $KLF4^{imm}$, $SOX2^{imm}$, $RUNX1^{imm}$; p=0.0094 for $SOX2^{sub}$; p=0.0101 for $CTCF^{sub}$, p=0.0034 for $CTCF^{imm}$, p=0.38 for $GATA2^{imm}$, two-tailed Student's t-test; error bars depict 95% C.I.).

Figure 6. TF-ARMs are important for normal development and disrupted in disease.



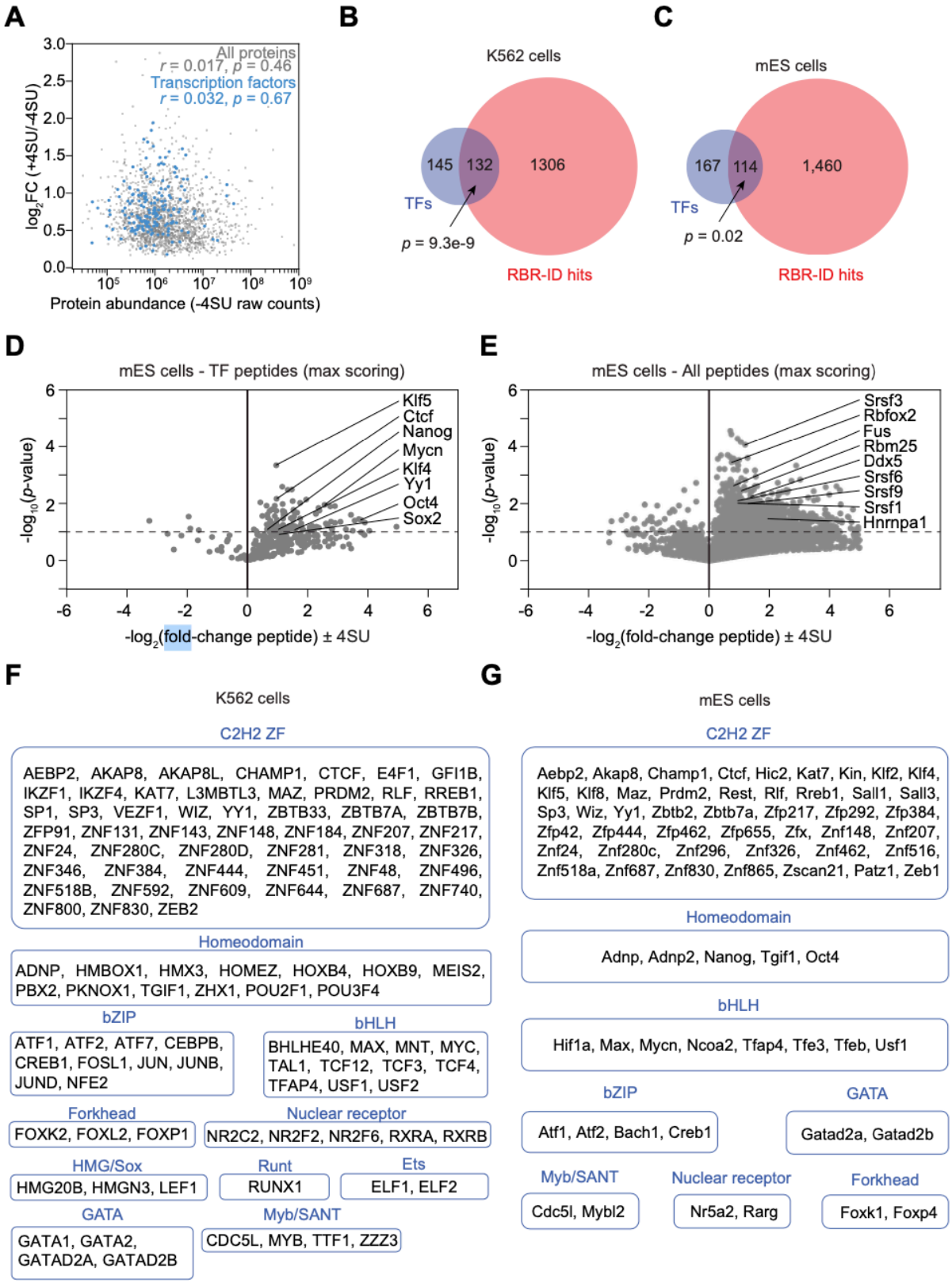
(A) Experimental scheme for injection of zebrafish embryos with morpholinos and rescue by co-injection with the indicated mRNAs (hpf = hours post-fertilization).
(B) Representative images of injected zebrafish embryos at 48 hpf.
(C) Scoring of zebrafish anterior-posterior axis growth.
(D) The landscape of mutations in TF-ARMs associated with human disease
(E) Examples of disease-associated mutations in TF-ARMs.
(F) Line plot of the observed frequency (red) or expected frequency (black) of mutations for amino acids in TF-ARMs ($p = 2.7 \times 10^{-74}$ for enrichment of mutations in arginine, one-side binomial test with Benjamini-Hochberg correction).
(G) Representation of the ESR1 protein and its correlation to the Tat ARM (*Maximum scoring ARM-like region). The selected mutation is provided in blue.
(H) Gel shift assay with 7SK RNA and synthesized peptides for Tat-ARM-WT, Tat-ARM-R52A, ESR1-ARM-WT, and ESR1-ARM-R269C.
(I) Tat transactivation reporter assay with wildtype or mutant versions of Tat and ESR1 ARMs and a version of the reporter without the Tat-binding TAR bulge. Values are normalized to the Tat-ARM-WT condition.

Figure 7. Transcription factors harbor functional RNA-binding domains.



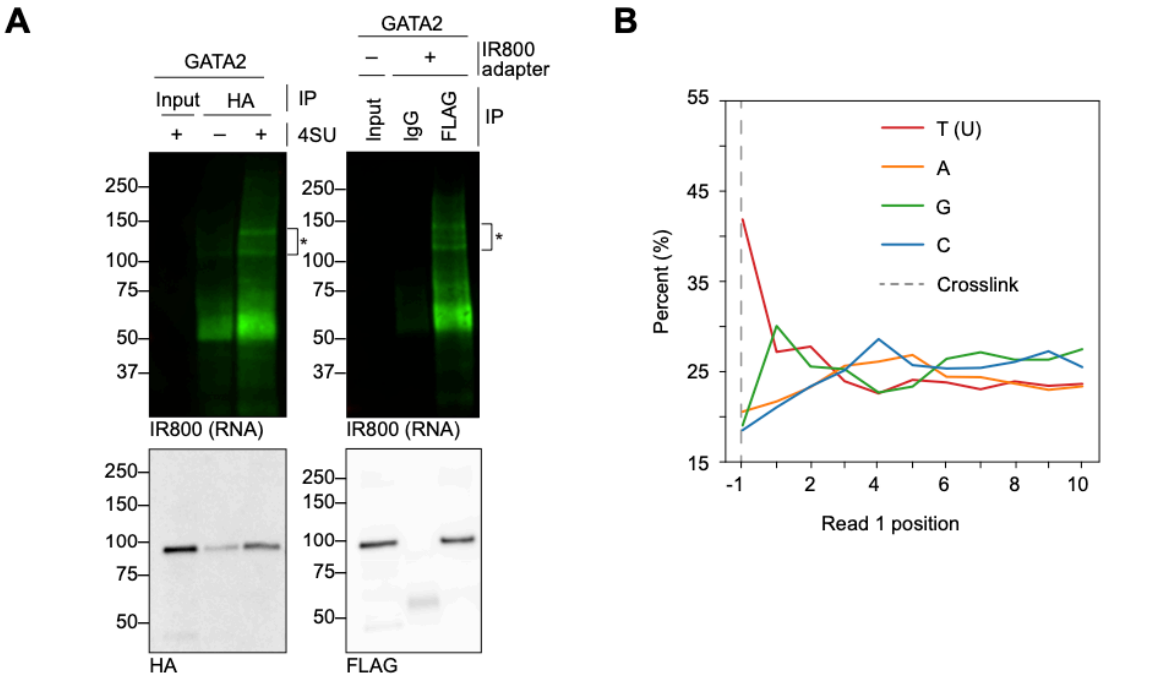
(A) A model depiction of a previously unrecognized RNA-binding domain in a large fraction of transcription factors and its role in TF function.
(B) Various ways by which RNA interactions could impact TF function at the molecular scale
(C) Various ways by which RNA interactions could impact TF function at the mesoscale

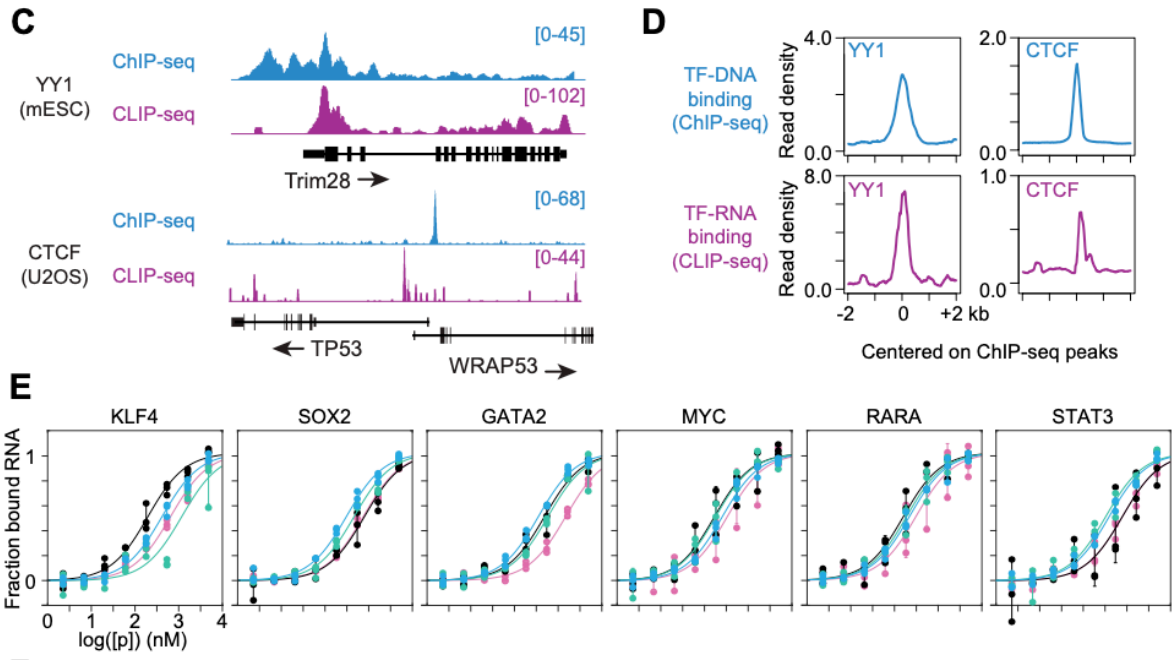
Figure S1. RNA-binding TFs in mammalian cells (Related to Figure 1).



(A) Scatter plot of 4SU-mediated fold change vs. protein abundance (raw peptide counts of -4SU condition) for the K562 RBR-ID (transcription factors in blue)
(B) Venn diagram depicting overlap of RBR+ protein hits and TFs for K562 cells (p=9.3e-9, Fisher's exact test).
(C) Venn diagram depicting overlap of RBR+ protein hits and TFs for mES cells (p=0.02, Fisher's exact test).
(D) Volcano plot of TF peptides in RBR-ID for murine embryonic stem cells with select highlighted TFs (dotted line at p=0.10). Each marker represents the peptide with maximum RBR-ID score for each protein.
(E) Volcano plot of all detected peptides in RBR-ID for murine embryonic stem cells with select highlighted RBPs (dotted line at p=0.10). Each marker represents the peptide with maximum RBR-ID score for each protein.
(F) List of RBR-ID+ TFs ($p<0.05$, $log2FC>0$) for K562 RBR-ID categorized by DBD family
(G) List of RBR-ID+ TFs ($p<0.10$, $log2FC>0$) for mESC RBR-ID categorized by DBD family.

Figure S2. Transcription factor binding to various RNAs (Related to Figure 1 and Figure 2).





F

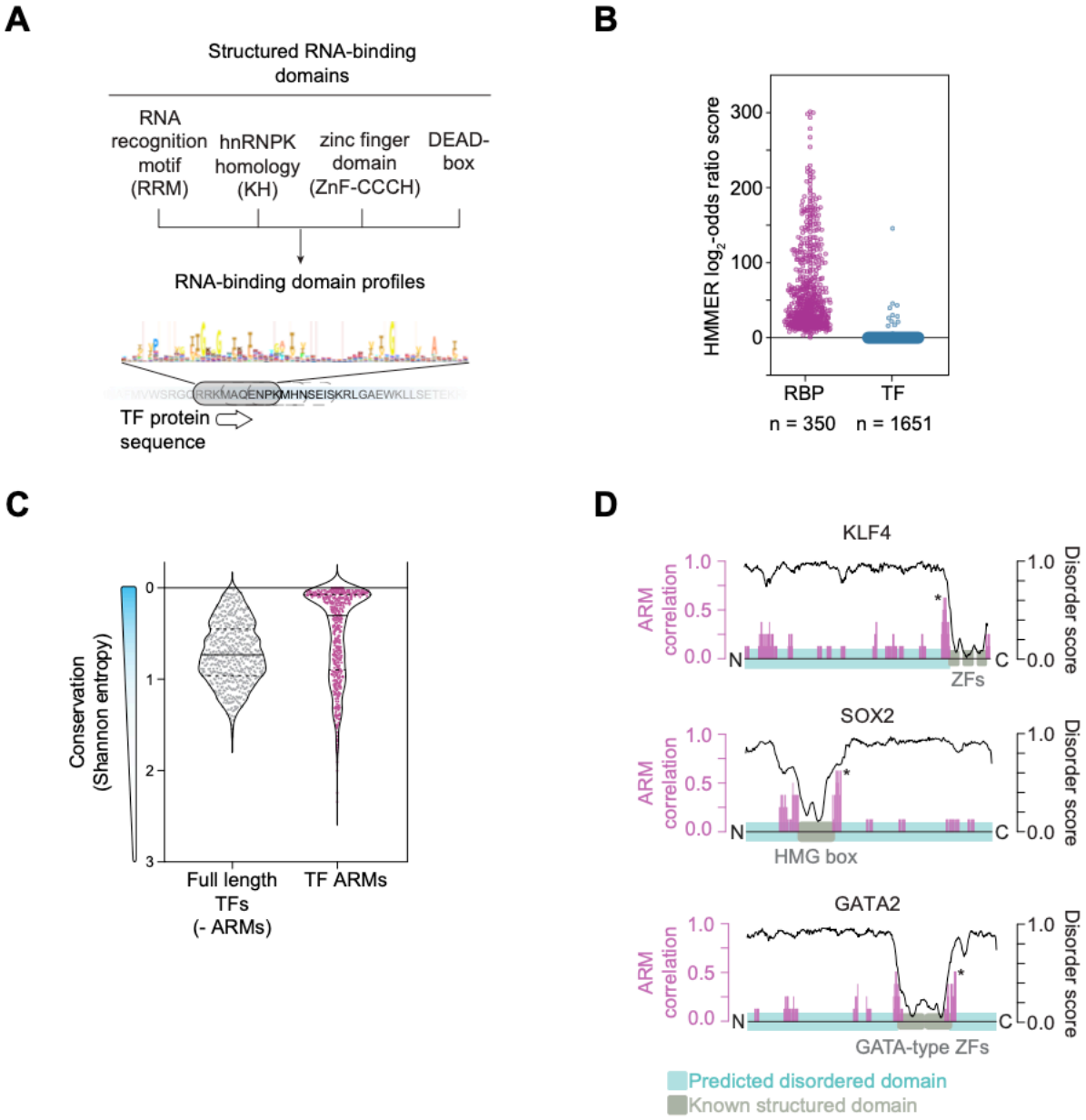
K_d (nM ± s.d.)

KLF4	SOX2	GATA2	MYC	RARA	STAT3	
198±28	810±191	493±83	165±45	158±36	795±408	random RNA
1210±525	467±56	591±109	176±37	188±30	327±70	*Pou5f1* pRNA
667±86*	761±65	1524±189*	329±105	311±105	794±144	*Pou5f1* eRNA
424±47*	308±33*	349±29	252±30	219±25	393±48	7SK RNA

* $p < 0.05$ compared to random RNA

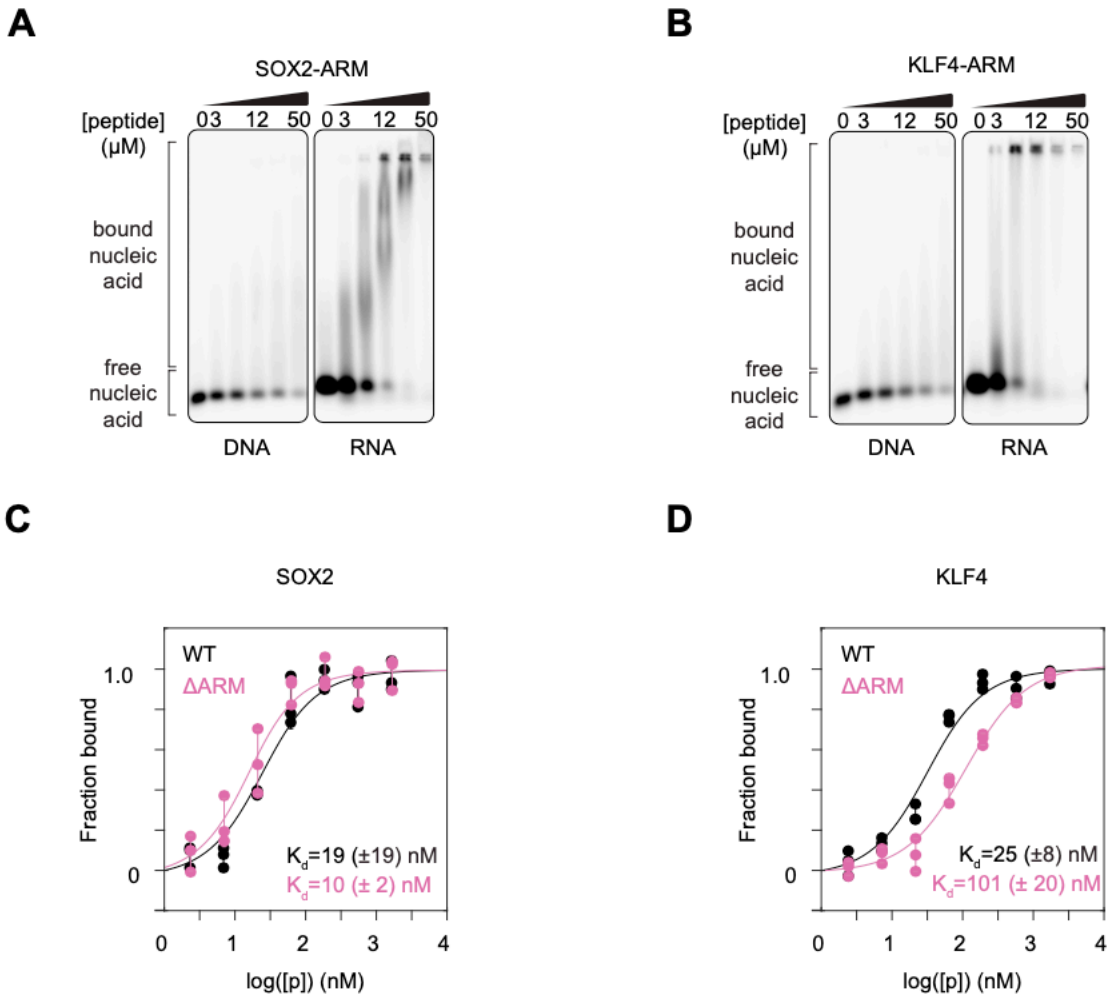
(A) Gel electrophoresis of UV-crosslinked HA-FLAG-GATA2 with visualization of RNA via IR800 adapter (top) and Western blot (bottom)
(B) Meta-analysis of nucleotide biases in CLIP reads
(C) ChIP-seq and CLIP signal for YY1 and CTCF at the Trim28 and TP53 genomic loci
(D) Meta-gene analysis of CLIP signal centered on YY1 or CTCF ChIP-seq peaks
(E) Fraction bound RNA with increasing protein concentration for 6 TFs and 4 RNA species per TF.
(F) Table of apparent Kd values for the binding assays in **(B)** (p-values comparing random RNA to pRNA, eRNA, and 7SK RNA respec- tively – KLF4: 0.06, 6.24e-6, 1.88e-4; SOX2: 0.09, 0.81, 0.013; GATA2: 0.47, 1.05e-5, 0.10; MYC: 0.84, 0.15, 0.11; RARA: 0.53, 0.17, 0.17; STAT3: 0.26, 0.99, 0.33).

Figure S3. Sequence analysis of RNA-binding regions in transcription factors (Related to Figure 3)



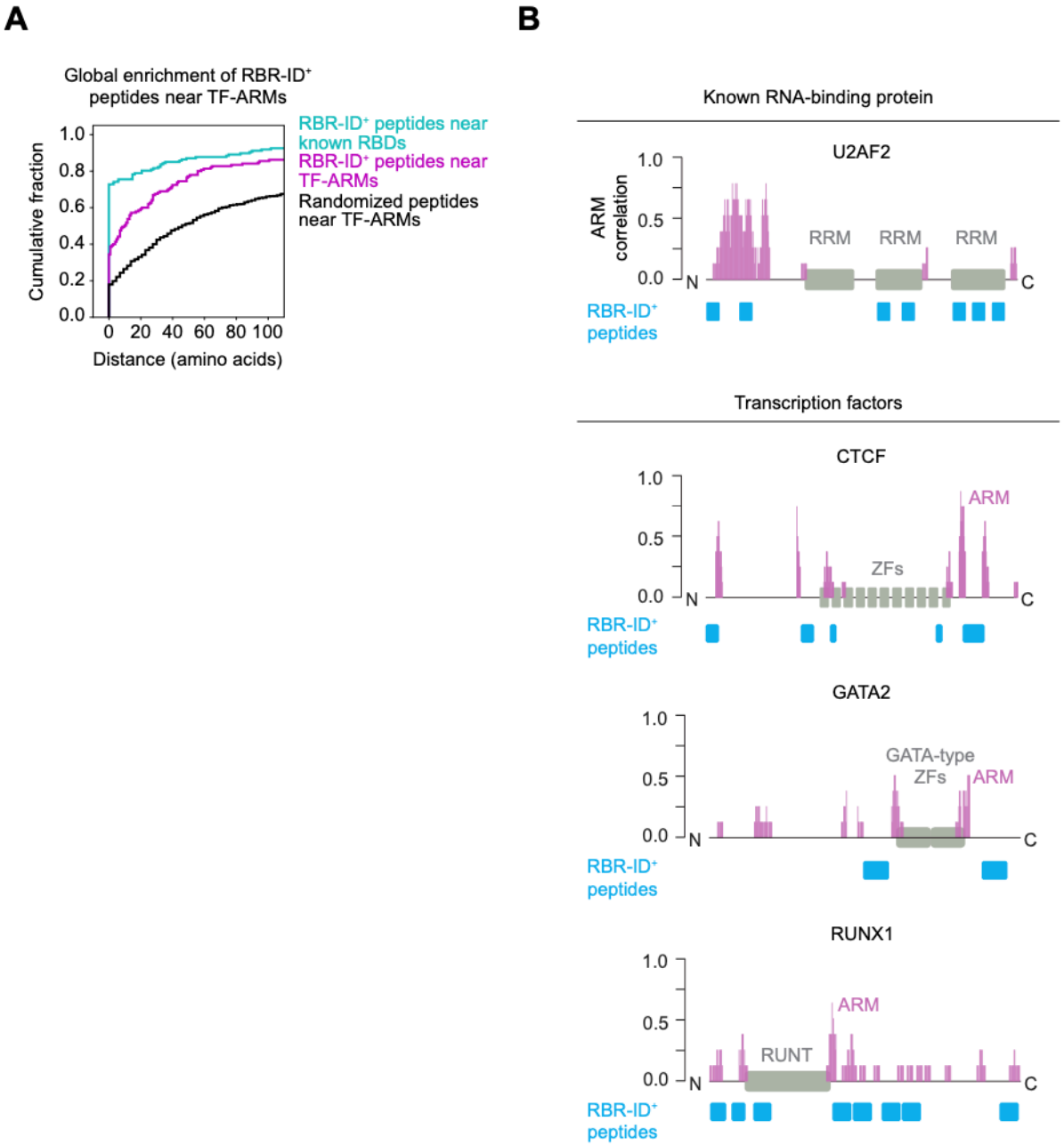
(A) Scheme to search for structured RNA-binding domain motifs in transcription factors.
(B) Scatter plot depicting the HMMER log2-odds ratio score for the 4 most abundant RNA-binding domains (RRM, KH, ZnF-CCCH, DEAD) for select RBPs and all human TFs.
(C) Evolutionary conservation analysis using Shannon entropy for TF-ARMs or TFs excluding the ARMs.
(D) Diagram of KLF4, SOX2, and GATA2 and their cross-correlation to the Tat ARM (magenta), predicted disorder (black line), DNA-binding domain (grey boxes) and predicted disordered domain (cyan).

Figure S4. Transcription factor binding to DNA in vitro (Related to Figure 3).



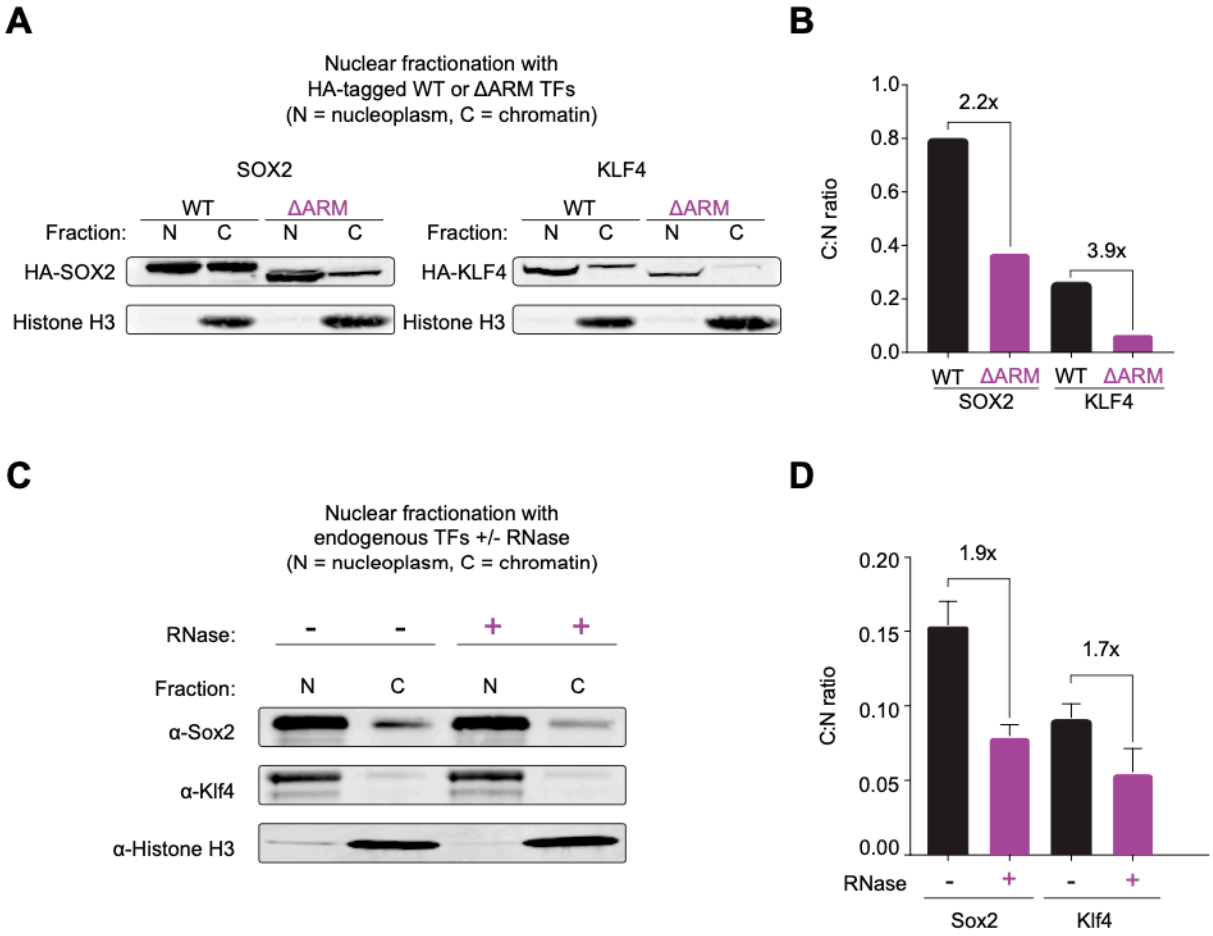
(A) Gel shift assay of the synthesized SOX2-ARM peptide with DNA or RNA
(B) Gel shift assay of the synthesized KLF4-ARM peptide with DNA or RNA
(C) Fraction bound motif-containing DNA with increasing protein concentration for SOX2 (SOX2WT vs SOX2ΔARM: p=0.11, error bars depict s.d.)
(D) Fraction bound motif-containing DNA with increasing protein concentration for KLF4 (KLF4WT vs KLF4ΔARM: p=8.75e-6; error bars depict s.d

Figure S5. Crosslinking of TF-ARMs to RNA in cells (Related to Figure 3).



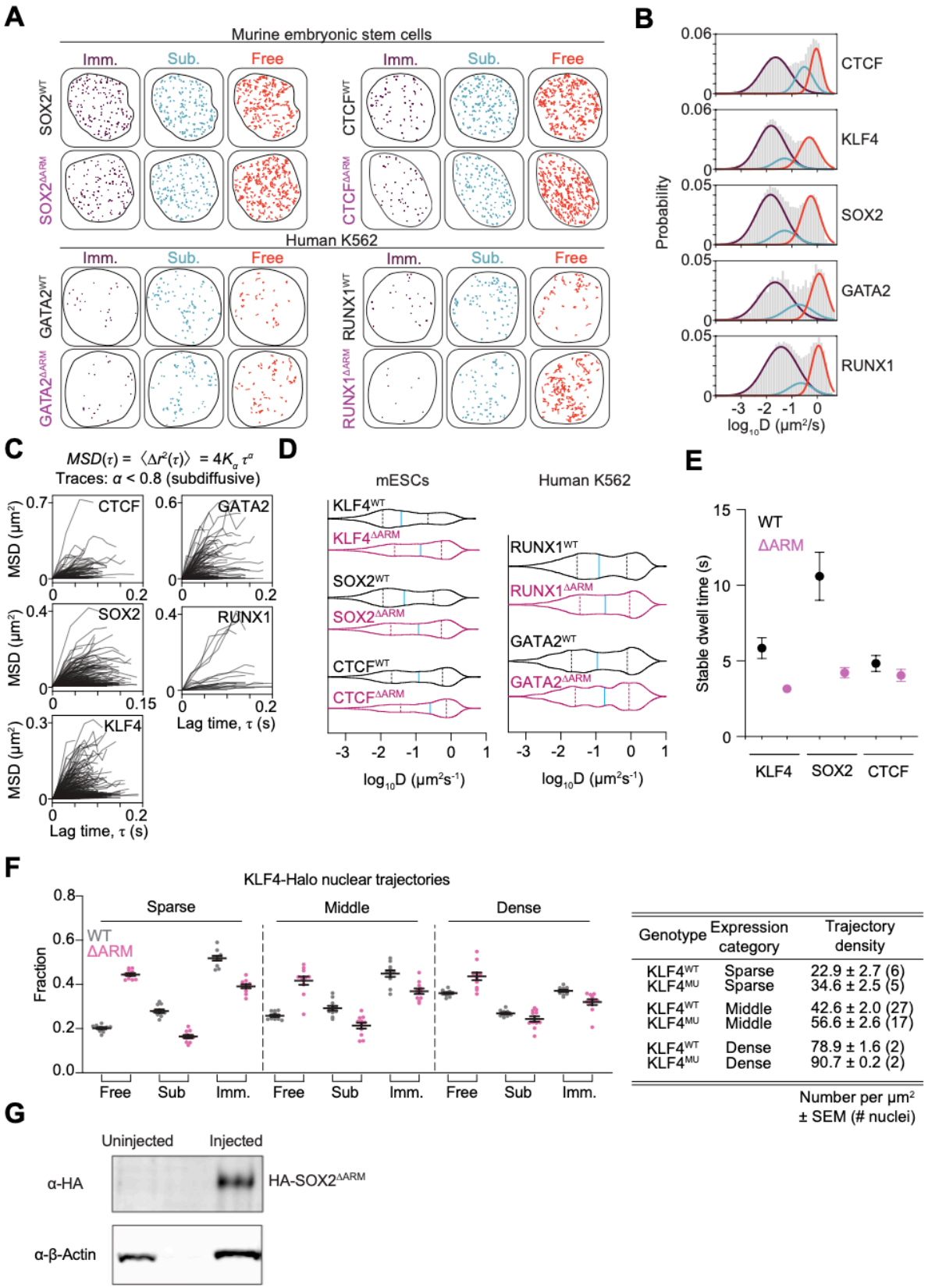
(A) Global analysis of RBR-ID+ peptide enrichment near known RNA-binding domains (cyan), TF-ARMs (magenta), or randomized peptides near ARMs (black).
(B) Examples of RBR-ID+ peptides for select TFs.

Figure S6. Transcription factor enrichment in sub-nuclear fractions (Related to Figure 4).



(A) Western blot of histone H3 and HA-tagged wildtype or ARM-mutant KLF4 and SOX2 in nucleoplasmic (N) or chromatin (C) fractions.
(B) Quantification of the relative intensity in N and C fractions of the samples in (A).
(C) Western blot of Sox2 or Klf4 and histone H3 in nucleoplasmic (N) or chromatin (C) fractions with or without RNase treatment.
(D) Quantification of the relative intensity in N and C fractions of the samples in (C).

Figure S7. Controls for in vivo experiments (Related to Figure 5 and Figure 6)



Genotype	Expression category	Trajectory density
KLF4^{WT}	Sparse	22.9 ± 2.7 (6)
KLF4^{MU}	Sparse	34.6 ± 2.5 (5)
KLF4^{WT}	Middle	42.6 ± 2.0 (27)
KLF4^{MU}	Middle	56.6 ± 2.6 (17)
KLF4^{WT}	Dense	78.9 ± 1.6 (2)
KLF4^{MU}	Dense	90.7 ± 0.2 (2)

Number per μm²
± SEM (# nuclei)

(A) Example of single nuclei single-molecule tracking traces for wildtype and ARM-mutant SOX2 and CTCF in mESCs, and GATA2 and RUNX1 in K562 cells. The traces are separated by their associated diffusion coefficient (Dimm: <0.04 μm2s-1; Dsub: 0.04-0.2 μm2s-1; Dfree: >0.2 μm2s-1). For each nucleus, up to 500 randomly sampled traces are shown.
(B) Histogram with curve fitting for a 3-state model of TF diffusion coefficients
(C) Individual traces of single molecules with subdiffusive (α<0.8) behavior
(D) Distribution of diffusion constants (D) for WT and ARM-mutant TFs
(E) Stable dwell times for KLF4, SOX2, and CTCF (error bars depict s.e.m.)
(F) Fraction of traces in 3-state model across different expression levels of KLF4 with table providing trajectory metrics across the different KLF4 expression levels.
(G) Western blot of lysates from zebrafish embryos injected with mRNA

Tables S1-S6 are available online:

https://doi.org/10.1016/j.molcel.2023.06.012

Materials and Methods

RESOURCE AVAILABILITY

Materials Availability
All unique/stable reagents generated in this study are available from the Lead Contact upon reasonable request with a completed Materials Transfer Agreement.

Data and Code Availability

- The RBR-ID mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD035484. CUT&Tag sequencing and CLIP sequencing data have been deposited to GEO with identifier GSE232181. Original images for EMSAs and Western blots as well as single molecule trace data are available through Mendeley Data (DOI: 10.17632/dkx9gsh42h.2). These data are publicly available as of the date of publication.
- Code generated during this study is available through Zenodo (DOI: 10.5281/zenodo.7974933; Link: https://zenodo.org/record/7974933)
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact by request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

The V6.5 murine embryonic stem cells were a gift from the Jaenisch laboratory of the Whitehead Institute, and these cells are derived from a cross of C57BL/6(F) x 129/sv(M). The human K562 and HEK293 cell lines were purchased from ATCC, and the HEK293F cells for protein purification were a gift from the Sabatini lab. Cell culture conditions are described below. Zebrafish experiments were conducted using male and female zebrafish from a wildtype Tübingen strain. Zebrafish embryos were scored at 48 hours post-fertilization, prior to sex determination, so the influence of sex on the results could not be determined. All animals were housed at Boston Children's Hospital following standard protocols (water temperature at 28.5 °C and a 14/10-hour light/dark cycle), and handled according to approved Institutional Animal Care and Use Committee (IACUC) of Boston Children's Hospital protocol 20-10-4254R.

METHOD DETAILS

Structures of known DNA-binding domains in TFs

TF-DNA X-ray structures were obtained from the RCSB Protein Data Bank (Accession numbers: YY1 = 1UBD, MYC/MAX = 1NKP, POU2F1 = 1CQT, JUN/FOS = 1FOS). These entries were modified using ChimeraX[70,71], and the effector domains, which are not included in the X-ray structures, are depicted as cartoons highlighting their dynamic and transient structure.

RNA binding region identification (RBR-ID)

K562 cells were cultured in suspension flasks containing culture medium [RPMI-1640 medium with GlutaMAX™ (ThermoFisher Cat. 72400047) supplemented with 10% FBS (ThermoFisher Cat. 10437028), 2 mM L-glutamine (Sigma-Aldrich Cat. G7513), 50 U/mL penicillin and 50 µg/mL streptomycin]. For each biological replicate of RBR-ID, 4 million K562 cells from actively

proliferating cultures were aliquoted into 2x T25 flasks. 4-thiouridine (4SU) was added to one of the two flasks for each replicate at a final concentration of 500 µM and incubated for 2 hrs at 37˚C with 5% CO_2. Cells from each flask were collected and resuspended in 600 µL 1x PBS [137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4] and transferred to 6-well plates. Plates were placed on ice with their lids removed and protein–RNA complexes were crosslinked with 1 J/cm^2 UVB (312 nm) light. Cells were lysed in Buffer A (10 mM Tris pH $7.9_{4˚C}$, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT, 0.2 mM PMSF) with 0.2% IGEPAL CA-630 for 5 min at 4˚C, then centrifuged at 2,500 g for 5 min at 4˚C to pellet nuclei. Nuclei were washed 3x with 1 mL cold Buffer A (without IGEPAL) and lysed at room temperature in 100 µL denaturing lysis buffer [9 M urea, 100 mM Tris pH 8_{RT}, 1x complete protease inhibitor, EDTA free (Roche Cat. 4693132001)]. Lysates were sonicated using a BioRuptor instrument (Diagenode) as follows: (energy: high, cycle: 15 sec ON, 15 sec OFF, duration: 5 min), centrifuged at 12,000 g for 10 min and supernatant was collected. Extracts were quantified using Pierce BCA assay kit (ThermoFisher Cat. 23225). 5 mM DTT was added to extracts and incubated at room temperature for one hr to reduce proteins, and then alkylated with 10 mM iodoacetamide in the dark for one hr. Samples were then diluted to 1.5 M urea with 50 mM ammonium bicarbonate and treated with 1 µL of 10,000U/µL molecular grade benzonase (Millipore Sigma Cat. E8263) and incubated at room temperature for 30 min. Sequencing grade trypsin (Promega Cat. V5117) was then added to samples at a ratio of 1:50 (trypsin:protein) by mass and incubated at room temperature for 16 hrs. The digested samples were loaded onto Hamilton C18 spin columns, washed twice with 0.1% formic acid, and eluted in 60% acetonitrile in 0.1% formic acid. Samples were dried using a speed vacuum apparatus and reconstituted in 0.1% formic acid, then measured via A_{205} quantification and diluted to 0.333 µg/µL.

For the proximity analysis in Figure S5, the nearest distance was calculated for each detected protein between RBR-ID+ peptides (p-val<0.05, log2FC<0) and either (1) TF-ARMs (cross-correlation to Tat ARM > 0.5, described below), (2) Known RNA-binding domains (RRM: IPR000504, KH: IPR004087, dsRBD: IPR014720). We required that at least 3 peptides were detected for each protein considered. As a control for the TF-ARM nearest distance analysis, the label (RBR-ID+ or RBR-ID-) of each peptide was randomly shuffled 100 times for all detected RBR-ID peptides for each protein, which provides the null distribution of the dataset.

The RBR-ID mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD035484.

LC-MS/MS

Peptide samples were batch randomized and separated using a Thermo Fisher Dionex 3000 nanoLC with a binary gradient consisting of 0.1% formic acid aqueous for mobile phase A and 80% acetonitrile with 0.1% formic acid for mobile phase B. 3 µL of each sample were injected onto a Pepmax C18 trap column and washed with a 0.05% trifluoroacetic acid 2% acetonitrile loading buffer. The linear gradient was 3 minutes until switching the valve at 2% mobile phase B and increasing to 25% by 90 minutes and 45% by 120 minutes at a flow rate of 300 nL/minute. Peptides were separated on a laser-pulled 75 µm ID and 30 cm length analytical column packed with 2.4 µm C18 resin. Peptides were analyzed on a Thermo Fisher QE HF using a DIA method.

The precursor scan range was a 385 to 1015 m/z window at a resolution of 60k with an automatic gain control (AGC) target of 10^6 and a maximum inject time (MIT) of 60 ms. The subsequent product ion scans were 25 windows of 24 m/z at 30k resolution with an AGC target of 10^6 and MIT of 60 ms and fragmentation of 27 normalized collision energy (NCE). All samples were acquired by LC-MS/MS in three technical replicates. Thermo .raw files were converted to indexed mzML format using ThermoRawFileParser utility (https://github.com/compomics/ThermoRawFileParser). To detect and quantify peptides, indexed mzML files from each set of technical replicates were searched together using Dia-NN v1.8.1[72] against a FASTA file of the *Homo sapiens* UniProtKB database (release 2022_02, containing Swiss-Prot + TrEMBL and alternative isoforms). Precursor and fragment m/z ranges of 300-1800 and 200-3000 were considered, respectively with peptides lengths from 6-40. Fixed and variable modifications included carbamidomethyl, N-term acetylation and methionine oxidation. A 0.01 q value cutoff was applied, and the options --peak-translation and --peak-center were enabled, while all other Dia-NN parameters were left as default.

Bioinformatic analysis of the RBR-ID data

After removal of suspected contaminants, identified peptides were re-mapped to an updated human proteome reference (UniProtKB release 2022_02, Swiss-Prot + TrEMBL + isoforms) to reannotate matching proteins. Where multiple protein matches were identified, peptides were assigned to a single protein annotation by first defaulting to Swiss-Prot accessions, where available, then by the accession with the most matching peptides in the dataset and therefore the most likely protein group[73]. Abundances of the different charge states of the same peptide were summed, and all abundances were normalized by the median peptide intensity in each run. To assess depletion mediated by RNA crosslinking, normalized abundances for each peptide in cells treated or not with 4SU were analyzed by unpaired, two-sided Student's *t* tests. For peptides that were missing across all 5 x 3 technical replicates in one of the treatments, Fisher's exact tests were used comparing the frequency of peptide detection between cells treated with or without 4SU. Statistical significance was determined by adjusting *p* values from both tests using the Benjamini-Hochberg method[74]. For mESC RBR-ID data from previous study[31], all peptides were re-mapped to an updated mouse reference proteome (UniProtKBrelease 2021_04) as described above while keeping original quantification and P-values. A relaxed p-value threshold (0.10) was used in the original study because it was validated to include additional RBPs[31]. Peptides were annotated using the InterPro database (release 87, accessed 28 Feb 2022) to identify functional domains. For volcano plots, outliers were removed and each marker represents the peptide with maximum RBR-ID score[31] for each protein. Transcription factors annotated in this dataset are from a previous census study[1].

Generating list of RNA-binding TFs

RNA-binding proteins identified in the current and previous studies using various methods were collected[18,23,31,75–81]. The list of RNA-binding proteins from these studies was overlapped with the list of transcription factors from a previous census study[1] using merge function in R. Transcription factors that are found at least in one dataset were reported in Table S3.

CLIP

CLIP experiments were performed as previously described[82] with minor modifications (see below for details). The protocol is a modified seCLIP protocol with the addition of 4SU incorporation (adapted from PAR-CLIP) and an IR800-conjugated 3' adapter. CLIP sequencing data have been deposited to GEO with identifier GSE232181.

Protein–RNA crosslinking
K562 cells stably expressing human GATA2 with N-terminal HA-FLAG-Halo tags under dox-inducible promoter were treated for 5 hours with 1 μM doxycycline (Sigma), and 24 hours with 100 μM of 4-Thiouridine (4SU) (Sigma-Aldrich T4509) prior to cell collection. Cells were resuspended in 1X PBS and transferred to a 6-well plate for crosslinking. Plates were placed on ice with lids removed and crosslinked at 365 nm at 0.3 J/cm^2. Cell suspension was transferred to microcentrifuge tubes and plates were washed with 1X PBS.

Lysate preparation
Cells were washed in 1X PBS and cell pellets were lysed in eCLIP lysis buffer [20 mM HEPES-NaOH pH 7.4, 1 mM EDTA, 100 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% sodium deoxycholate, 1x cOmplete™ EDTA-free protease inhibitor cocktail (Roche 4693132001)]. Samples were sonicated in a Diagenode Bioruptor (30 s ON/OFF) on medium for 5 minutes. RNase I (ThermoFisher AM2294) was added to lysates for a final concentration of 0.4 U/μL and incubated at 37 °C at 1200 rpm for 5 min. EDTA was immediately added at a final concentration of 21 mM. Lysates were clarified at 15,000g for 10 minutes at 4˚C and supernatant was transferred to fresh tubes. Protein concentration was measured using Protein Assay Dye Reagent (Bio-Rad 5000006).

Labeling of crosslinked protein–RNA complexes
Dynabeads™ were washed in eCLIP binding buffer (20 mM HEPES-NaOH pH 7.4, 20 mM EDTA, 100 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% sodium deoxycholate). Antibody was added to bead mixture and incubated, rotating at room temperature for 45 min. Antibody-bead mixture was washed in eCLIP binding buffer and mixed with calculated amount of lysate. Tubes were incubated overnight rotating at 4˚C. 2% of lysate-bead mixture was transferred to a new tube to serve as input sample. IP samples were washed with CLIP wash buffer (20 mM HEPES-NaOH pH 7.4, 20 mM EDTA, 5 mM NaCl, 0.2% Tween-20) and IP_{50} (20 mM Tris pH 7.3_{RT}, 0.2 mM EDTA, 50 mM KCl, 0.05% NP-40). Samples were treated with TURBO™ DNase (ThermoFisher AM2238) and 0.1 U/μL final concentration of RNase I (in some cases, 1 U/μL final concentration was used for better visualization of bands, e.g. **Fig. S2A**). IP samples were washed in CLIP wash buffer and FastAP buffer (10 mM Tris-Cl pH 7.5_{RT}, 5 mM $MgCl_2$, 100 mM KCl, 0.02% Triton X-100). IP RNA was dephosphorylated using FastAP phosphatase reaction FastAP Thermosensitive Alkaline Phosphotase (ThermoFisher EF0652), and T4 PNK (NEB M0201S).

IP samples were washed in CLIP wash buffer and 1X RNA Ligase buffer (50 mM Tris-Cl pH 7.5_{RT}, 10 mM $MgCl_2$]. A 3' IR-800 fluorescent adaptor was ligated using T4 RNA Ligase 1 high concentration (NEB M0437M). Samples were washed in eCLIP high-salt wash buffer (50 mM Tris-HCl pH 7.4_{RT}, 1M NaCl, 1 mM EDTA, 1% NP-40, 0.1% SDS, 0.5% sodium deoxycholate) and CLIP wash buffer. IP and input samples were eluted with 4X LDS Sample Buffer (ThermoFisher NP0007), run on an 8% bis-tris gel, and transferred overnight to a nitrocellulose membrane.

Library preparation and sequencing
The transferred membrane was cut ~0–50 kDa above protein size and incubated with Proteinase K (ThermoFisher AM2548) to isolate crosslinked RNA. Remaining steps were

performed as per the seCLIP protocol[83], with some modifications. RNA was purified and concentrated with phenol:chloroform:IAA (ThermoFisher AM9732) and ethanol precipitation. 3' and 5' adapters were designed to include an IR800 fluorophore and an 8-nt UMI for cDNA ligation, respectively. We did not include 5' deadenylase enzyme in our 5' ligation reactions and we used the AffinityScript RT (Agilent 600107) for crosslinking-induced truncation. Libraries were sequenced on an Illumina NextSeq 500 in paired-end mode for 47:8:8:29 cycles (read 1 : index 1 : index2 : read 2).

CLIP Analysis

Generating CLIP-seq peaks

Raw CLIP-seq reads were trimmed using Cutadapt[84]. The adapter sequence AGATCGGAAGAGCACACGTCTGAA was trimmed from the 5' end of the reads, AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT adapter sequence from the 3' end, and a universal four nucleotide UMI from the 3' end. Prior to mapping, UMIs were extracted from the 5' end of the reads using UMI-tools version 1.0.0 with the argument --bc-pattern=NNNNNNNN[85]. Bowtie2 was used to map all trimmed reads to the hg19 human genome using parameters -p 40 –end-to-end –no-discordant[86,87]. Trimmed, mapped, and unique reads were then sorted using the samtools sort function and indexed using the bedtools index function[88,89]. Lastly, reads were collapsed to account for PCR duplicates using the extracted UMIs with the UMI-tools dedup function. These trimmed, mapped, and collapsed reads were then used for downstream analysis. To call CLIP-seq peaks, .bed files were generated using MACS with parameters -g hs --keep-dup auto –-nomodel[90].

Identifying crosslinked nucleotides

As per the seCLIP protocol, during the reverse transcription step, polymerase terminates at the site of the cross-link[83]. This yields a cDNA product in which the 3' nucleotide of the cDNA is the nucleotide before the site of the cross-link on the pulled down RNA. During the paired end sequencing, the position 1 of the 5' end of read1 will therefore map to the site on the genome that is one nucleotide downstream the cross-linked nucleotide. To extract this site from the mapped CLIP-seq reads and generate Table S4 with sequences containing the site of the cross-link +/- 5bp, the genomic locations for the forward strand reads were first extracted. bedtools fasta was then used to extract the -1 position of the 5' end of the forward strand mapped reads (see CLIP methods) and +/- 5 bps around this site. This generated 11nt sequences in which the site of the cross-link is at the center of the sequence (nucleotide position 6).

To filter out any sequences in which the polymerase terminated early (i.e. prior to the cross-link) in the reverse transcription step, the sequences containing cross-linked nucleotides were filtered further for only the sequences containing a T (U) in the cross-link site (position 6). As expected, there was an enrichment of T (U) nucleotides as compared to G's, C's, and A's at this position within the sequences. The list was further filtered to only include sequences that overlap with called CLIP-seq peaks (see Generating CLIP-seq peaks)

To annotate the cross-link containing sequences with whether they fell within a gene, an enhancer, or a promoter in Table S4, the chromosomal locations of the cross-link containing sequences were overlapped with RefSeq genes, H3K27Ac ChIP-seq peaks (GSM733656), or RefSeq genes TSS +/- 200, respectively. H3K27Ac peaks were called using MACS with parameters -g hs --keep-dup auto –-nomodel.

Generating CLIP-seq metaplots

Fastq files from GATA2 ChIP-seq[91] (GSM467648) and RUNX1 ChIP-seq[92] (GSM2423457) experiments in K562 cells were downloaded from Gene Omnibus Expression database (GEO) and aligned to the hg19 human genome using Bowtie2. ChIP-seq peaks were called using MACS with parameters -g hs --keep-dup auto –-nomodel. Regions for metaplot analysis were generated using +/-2000 bases from the center of the called peaks. Normalized CLIP-seq densities within these regions were calculated using bamToGFF[93]. Input-corrected meta-gene plots were generated by subtracting the mean read density per bin of the input CLIP at ChIP peaks from the HA pull down CLIP at ChIP peaks. R matplot function was used to plot the density values across the 4Kb region.

Protein purification

To purify transcription factors (NANOG: pJH201, RARA: pJH203, CTCF: pJH205, OCT4: pJH199, MYC: pJH200, P53: pJH204, KLF4: pJH278, ESR1: pJH202, YY1: pJH087, SOX2: pJH198, STAT3: pJH227, GATA2: pJH247, SMAD3: pJH226, see Key Resource Table for plasmid information), a mammalian purification system using Freestyle HEK 293F cells (gift from Sabatini lab) were used. HEK cells were grown in FreeStyle 293 Expression Medium (Gibco) on an orbital shaker. Coding sequence of desired genes were synthesized by IDT as gBlock fragments (Table S6) containing proper Gibson overhangs. TF-ARM deletion mutants (pJH279, pJH245, pJH272, Key Resource Table) were generated by removal of a stretch of peptide adjacent to DNA binding domains that contain ARMs. The amino acid sequences that are removed in TF-ARM mutants are shown in parentheses as follows: hsKLF4_ΔARM (aa 355-386), hsSOX2_ΔARM (aa 118-178), hsGATA2_ΔARM (aa 360-395), and hsCTCF_ΔARM (576-611). To reduce sequence complexity for gBlock synthesis, codon optimization using the IDT codon optimization tool was applied when needed. The fragments are then cloned into a mammalian expression vector containing Flag and mEGFP (N- or C- terminal) (modified from Addgene #32104) using NEBuilder HiFi DNA Assembly kit (E2611). These vectors were transiently transfected into 293F cells at a concentration of 1 million/ml with 1 µg of DNA per million cells using branched polyethylenimine (PEI) (Polysciences). 60-72 hours post-transfection, cells were resuspended in 45 ml HMSD50 buffer (20 mM HEPES pH 7.5, 5 mM MgCl2, 250 mM sucrose, 1mM DTT, 50mM NaCl, supplemented with 0.2 mM PMSF and 5 mM sodium butyrate) and incubated for 30 min at 4° C with gentle agitation. After a spin down at 3500 rpm at 4°C for 10 min, the supernatant was discarded and the pellet containing nuclei were resuspended in 35 ml of BD450 buffer (10 mM HEPES pH 7.5, 5% Glycerol, 450 mM NaCl, and protease and phosphatase inhibitors) and incubated for 30 min at 4° C with agitation. The solution was spun down at 3500 rpm at 4°C for 10 min to clear the nuclear extract. The supernatant was transferred into fresh tube and the pellet containing chromatin was passed through 18G ½ syringe 5 times. The chromatin containing lysate was spun down at 8000 rpm at 4° C for 10 min and supernatant is combined with the previously collected supernatant. Then the combined supernatants were spun down again at 8000 rpm at 4°C for 10 min to clear the lysate. 500 ul of Flag-M2 beads (Sigma) were added to the cleared lysates and incubated overnight at 4° C. The Flag-M2 beads were washed 2 times with 45 ml BD450 buffer and they were transferred into a purification column (Biorad). The beads on the column were washed 2 more times with 10 ml BD450 buffer and 5 ml Elution buffer (20 mM HEPES pH 7.5, 10% Glycerol, 300 mM NaCl). Elutions were performed by incubating the beads overnight at 4° C with 800 elution buffer and 200 ul of 5mg/ml flag peptide (Sigma). The buffer exchange (into

elution buffer) and concentration of proteins were performed using spin columns (Milipore). Proteins were aliquoted and stored at -80°C. Canonical RNA-binding proteins (U2AF2, HNRNPA1, SRSF2) were purified in a previous study[94].

In vitro RNA synthesis and purification

To synthesize labeled RNA for fluorescence polarization measurements, in vitro transcription templates were generated from ssDNA oligos (for the random RNA template, Integrated DNA Technologies), gBlocks (for 7SK template, Integrated DNA Technologies), or PCR amplification of genomic DNA from V6.5 murine embryonic stem cells (for *Pou5f1* enhancer and promoter RNAs)[58] (Table S6). Templates were amplified by PCR with primers containing T7 (sense) or SP6 (antisense) promoters:

T7 (added to 5' of sense): 5' TAATACGACTCACTATAGGG 3'
SP6 (added to 5' of antisense): 5' ATTTAGGTGACACTATAGAA 3'

Templates were amplified using Phusion polymerase (NEB), and the products were gel-purified using the Monarch Gel Extraction Kit (NEB) following the manufacturer's instructions and eluted in 40 µL H2O. Each template was transcribed using the MEGAscript T7 kit using 200 ng total template according to the manufacturer's instructions. Reactions included a Cy5-labeled UTP (Enzo LifeSciences ENZ-42506) at a ratio of 1:10 labeled UTP:unlabeled UTP. The transcription reaction was incubated overnight at 37°C, and then it was incubated with 1 µL TURBO DNase (supplied in kit) for 15 minutes at 37°C. Transcribed RNA was purified by the MEGAclear Transcription Clean-Up Kit (Invitrogen) following the manufacturer's instructions and eluting in 40 µL H2O. The RNA was diluted to 2 µM and aliquoted to limit freeze/thaw cycles. Transcribed RNA was analyzed by gel electrophoresis to verify a single band of correct size.

Fluorescence polarization assay

To determine the binding affinity of a protein with RNA, we conducted the fluorescence polarization assay as previously described with some minor modifications[18] (Holmes et al 2020)., The concentration of protein is serially diluted from 5000 nM down to 2 nM by a 3-fold dilution factor. The series of protein concentrations is then mixed with a buffer containing 10 nM Cy5-labeled RNA, 10 mM Tris pH 7.5, 8% Ficoll PM70 (Sigma F2878), 0.05% NP-40 (Sigma), 150 mM NaCl, 1 mM DTT, 0.1 mg/mL non-acetylated BSA (Invitrogen AM2616), and 10 µM $ZnCl_2$. The reactions were performed in triplicates in a 20 µL reaction volume. After incubating the reactions 1 hr at room temperature, they are transferred into flat bottom black 384 well-plate (Corning 3575). Anisotropy was measured by a Tecan i-control infinite M1000 with the following parameters. Excitation Wavelength: 635 nm; Emission Wavelength: 665; Excitation/ Emission Bandwidth: 5 nm; Gain: Auto; Number of Flashes: 20; Settle Time: 200ms; G-Factor: 1. To account for instrument error, the plate was measured 3 times and the mean of the values are used in the affinity calculations. Reagents used for established RNA-binding proteins were generated previously[94] and BamHI was purchased from New England Biolabs.

To determine the binding affinity of a protein with DNA, the same buffer conditions and incubation times were used, as described above. The series of protein concentrations from

0.76-1666 nM (3-fold serial dilution) and 10 nM cy5-labeled DNA were used. The motif containing DNA sequences that have been shown to bind SOX2[18] and KLF4[95] were ordered from IDT. To prepare motif-containing DNA sequences, 50 μM of oligos with complementary sequences (one unlabeled and the other labeled with cy5) (Table S6) were annealed in TE+100 mM NaCl buffer by ramping down the temperature from 98°C to 4°C on a thermocycler. Then the annealed DNA fragments were diluted to appropriate concentrations with water for the assay.

Binding curves were fit to fluorescence anisotropy data via nonlinear regression with the Levenberg-Marquardt-based 'curve_fit' function in scipy (v. 1.7.3). Curve fitting was performed using a monovalent reversible equilibrium binding model accounting for ligand depletion, given by the equation below:

$$A = A_0 + (A_1 - A_0)\left[\frac{P_0 + L_0 + K_d - \sqrt{(P_0 + L_0 + K_d)^2 - 4P_0L_0}}{2L_0}\right]$$

where P_0 is the total protein concentration, L_0 is the total ligand (RNA) concentration, and A_0, A_1, and K_d are fit parameters. The measured anisotropy value A for each condition was determined by first averaging raw anisotropy measurements across three subsequent reads of the same well, then averaging these values across three technical replicates from separate wells. To calculate the bound fraction of RNA, A values were normalized to the range between the upper and lower anisotropy asymptotes A_0 and A_1. Error bars were computed from the standard deviation of RNA bound fraction across three technical replicates. The script used to calculate the affinities are available on Zenodo (https://zenodo.org/record/7974933).

Electrophoretic mobility shift assay

To determine the binding affinity of a TF-ARM peptides (synthesized by Genscript) (Table S6) with 7SK RNA, we conducted the electrophoretic mobility shift assay as previously described with some minor modifications[19,36]. The concentration of peptides was serially diluted from 50000 nM down to 3.125 nM by a 2-fold dilution factor in buffer containing 20 mM HEPES, 300 mM NaCl, and 10% Glycerol. The series of protein concentrations was then mixed 1:1 with a buffer containing an initial concentration of 20 nM Cy5-labeled RNA, 20 mM Tris pH 8.0, 5% glycerol, 0.1% NP40 (Sigma), 0.02 mM $ZnCl_2$, 1 mM $MgCl_2$, 2 mM DTT, and 0.2 mg/mL non-acetylated BSA (Invitrogen AM2616). For DNA-binding assays, 20 nM Cy5-labeled dsDNA or 20 nM Cy5-labeled ssRNA were used (Table S6). The reactions were performed in a 20 μL reaction volume. After incubating the reactions in the dark for 1 hr at room temperature, they were loaded into a 2.5% agarose gel that is pre-run for at least 30 min at 4°C. The samples then ran for 1.5 hr at 150V at 4°C. The gel is imaged using Typhoon FLA95 imager with a Cy5 fluorescence module.

Homology search for RNA-binding domains in TFs

We retrieved hidden Markov model based profiles (HMM-profiles) for RNA-binding domains corresponding to the following Pfam[96] entries using hmmfetch from the HMMER package (hmmer.org) – RRM_1, RRM_2, RRM_3, RRM_5, RRM_7, RRM_8, RRM_9, DEAD, zf-CCCH, zf-CCCH_2, zf-CCCH_3, zf-CCCH_4, zf-CCCH_6, zf-CCCH_7, zf-CCCH_8, KH_1, KH_2, KH_4, KH_5, KH_6, KH_7, KH_8, KH_9. These domains represent the largest families of RNA-

binding domains. We searched for these profiles using *hmmsearch* form the HMMER package with '-T 0' as a parameter in fasta files with sequences corresponding to TFs[1] or RNA-binding proteins[97]. The log2-odds ratio score from the *hmmsearch* output was plotted for RBPs with score > 0 (n=350, to provide scores that one would expect if these domains were in the protein) and for all 1651 TFs[1]. If a TF was not in the output, it was assigned a score of 0.

Analysis of ARM-like regions in TFs

We used an approach based on analogous functions in localCIDER[98] and on a previously applied procedure[99] used to map basic patches. For each TF, amino acid compositions of Lys and Arg in sliding 5-residue windows were computed. Basic patches were defined as regions of ≥ 5 consecutive residues that consisted of Lys and Arg occurring at a frequency of >0.5. This threshold was based on optimizing this approach against previously described basic patches in MECP2[99]. All identified basic patches were filtered for those that occurred within predicted IDRs (*metapredict*), determined as described above. For the adjacency analysis, DNA-binding domains were defined based on domains with annotations of *DNA-binding* in Interpro[100]. Probabilities of basic patch occurrence in all TFs were computed starting from the N-terminal edge of the first DNA-binding domain and moving N-terminally, or the C-terminal edge of the last DNA-binding domain and moving C-terminally. These probabilities were summed to arrive at the total probability as a function of distance from the bounds of the DNA-binding regions.

A consensus motif for bioinformatically identified basic patches (Figure 3B) was created using MEME (v. 4.11.4)[101]. Briefly, 963 basic patches found in TFs were padded by appending the 10 amino acid residues upstream and downstream of each the region. Next, a zero-order Markov model was created from 1,290 full sequences of annotated TFs using the 'fasta_get_markov' function to generate a background for the motif search. The TF basic patch sequences were input to the 'MEME' function using the TF background model, specifying a constraint to identify exactly one site per sequence, a minimum motif width of 5, a maximum motif width of 13, and defaults for the unspecified parameters.

A charge-based cross-correlation method was employed to identify ARMs in TF disordered regions similar to the HIV Tat ARM. Extensive in vitro and cellular analyses of the Tat ARM have mapped the critical residues responsible for Tat RNA-binding and HIV transactivation[41,42]. To properly function, the Tat ARM requires an arginine positioned near the motif center flanked by an enrichment of basic residues (R/K). The Tat ARM sequence "RKKRRQRRR" was digitized to the amino acid charge pattern "111110111" to create a 9-mer search kernel. A protein target sequence was created by first digitizing the sequence of the protein of interest to "1" for R/K amino acid residues and "0" otherwise, then refining the sequence by setting residues to "0" if they fell outside of disordered regions assessed through the *metapredict* package[102] (v. 2.2) with a disorder threshold of 0.2. The target sequence was further refined by setting all entries to "0" in 9-mer windows where no R's were originally present. The cross-correlation between the search kernel and the target sequence was then computed using the 'correlate' function in scipy using the "direct" method. Maximum cross-correlations were computed as the maximum of the returned array for each protein tested. This method was applied iteratively to all sequences from the UniProt database to generate distributions for TFs and the proteome.

Evolutionary conservation of TF-ARMs

Evolutionary conservation of specific human TFs was assessed using the ConSurf online server[103]. TF sequences were downloaded from UniProt and run without specifying a 3D structure or MSA, with automatic detection of homologs from the "NR_PROT_DB" database. Defaults were used for all other running parameters. Amino acid conservation scores from the ConSurf GRADES output were re-normalized between 0 and 1 for each protein, such that a score of 1 corresponded to the of the most conserved amino acid in a given protein.

To evaluate the extent of evolutionary conservation for a larger cohort of TF ARMs, the degree of conservation of TF ARMs was compared to non-ARM regions across vertebrates. The OrthoDB v10 database was used to identify the set of vertebrate orthologs for each protein in a list of annotated human TFs. For each TF, a multiple sequence alignment (MSA) of the retrieved vertebrate orthologs was generated using Clustal Omega (v. 1.2.4) with default parameters. The output ALN format MSA files were converted directly to FASTA format. TFs with an ARM maximum cross-correlation score of 5 or above were retained for further analysis. Each MSA file was parsed via the "prody" package (v. 2.3.1)[104] in Python using the 'parseMSA' command. Reference coordinates for the MSA were set with respect to the human TF of interest by using the 'refineMSA' command and specifying the ID of the human TF. The degree of conservation of each amino acid residue in the human TF was quantified by computing the Shannon entropy (H) for each residue via the 'calcShannonEntropy' function. Higher values of H represent more sequence variation at a specific residue position and therefore a lower degree of evolutionary conservation. To define ARM regions for the purpose of Shannon entropy analysis, the union of 9-mer regions with an ARM cross-correlation score of 5 or above was used. For each TF analyzed (N=580), the median value of H in the ARM region and the median value of H in the remainder of the sequence (non-ARM region) were calculated and plotted. Distributions of these paired data were compared via a Wilcoxon signed-rank test.

HIV Tat transactivation assay

To generate the HIV LTR luciferase reporter (pJH325, Key Resource Table), the HIV 5' LTR from the pNL4-3 isolate (Genbank AF324493) was cloned into pGL3-Basic (Promega) via Gibson assembly (NEB 2X HiFi) with a HindIII-digested pGL3-Basic and a gBlock (Integrated DNA Technologies) containing the HIV 5' LTR with compatible overhangs (Table S6). A mutant version of this reporter lacking the Tat activation site (TAR RNA bulge structure)[44] was also generated in a similar fashion (pJH326, Key Resource Table). Mammalian expression vectors encoding Tat, an R/K>A mutant of Tat, and replacements of the Tat ARM with TF-ARMs from KLF4, SOX2, GATA2, and ESR1 were generated by Gibson assembly with a NotI-XhoI-digested pcDNA3 (Invitrogen) and gBlocks encoding these variants with compatible overhangs (pJH327, pJH329, pJH330, pJH361, pJH371, pJH365, pJH366, Key Resource Table and Table S6).

For transfections, HEK293T cells were cultured in DMEM (Gibco) supplemented with 10% fetal bovine serum (Sigma F4135), 50 U/mL penicillin and 50 µg/mL streptomycin (Life Technologies 15140163). Transfections were conducted in triplicate. 24-well plastic plates were first coated with poly-L-lysine (Sigma) for 30 minutes at 37°C, washed once with 1X PBS, and then allowed to air dry. Cells were seeded in 500 µL of media in coated wells at a density of $2x10^5$ cells per well. The next day, each well was transfected using Lipofectamine 3000 (Life Technologies) (total reaction 50 µL Optimem, 1.5 µL Lipo-3000, 0.6 µL P3000, and the appropriate volume of

DNA) with 100 ng of the HIV 5' LTR reporter vector, 150 ng of the pcDNA3 expression vector (encoding Tat or the variants), and 50 ng of a renilla luciferase plasmid (pRL-SV40, Promega) to normalize transfection efficiency. As a control, we included a pcDNA3 vector expressing LacI-mCherry (labeled as "No Tat" in Figure 3). After 6 hours of incubation, luciferase activity was quantified by the Dual Luciferase Assay kit (Promega) following the manufacturer's instructions and a Safire II plate reader. The luminescence values were first normalized to the renilla luciferase luminescence for each well, and then all conditions were normalized to the average value of the "No Tat" control condition.

CUT&Tag experimental procedure

CUT&Tag sequencing was performed using the CUT&Tag-IT Assay Kit (Active Motif 53160) according to manufacturer's instructions. Stable mESC lines expressing HA-tagged versions of WT and ARM-mutant SOX2 and KLF4 were induced with doxycycline (1 µg/mL) for 6 hours, and $4x10^5$ mESCs were collected. The nuclei of the cells were extracted and incubated with 1µg of HA antibody (Abcam ab9110). After incubation with a rabbit secondary antibody and pA-Tn5 Transposomes, DNA was extracted and amplified with i7/i5 indexed primer combinations. SPRI Bead clean-up of the amplified DNA fragments were performed, and libraries were pooled, subjected to gel-based clean up and sequenced by Novaseq (50x50). CUT&Tag sequencing data have been deposited to GEO with identifier GSE232181.

CUT&Tag analysis

Reads were first trimmed by adapter sequence (CTGTCTCTTATACACATCT) in the forward and reverse directions using Cutadapt with default parameters. Subsequent analysis of the data was conducted according to a published protocol with no modification[105]. Reads were aligned to the mm10 mouse genome, and samples were spike-in normalized according to the protocol by calculating a scale factor from reads aligning to the E. coli genome. Peak calling for both WT and ARM-mutant samples was conducted using the Seacr algorithm using the "non" (non-normalized) and "stringent" parameters[106]. For meta-gene plots, raw read density was calculated by centering on called peaks for both WT and ARM-mutant TFs that were merged using bedTools merge with default parameters.

TF reporter assays

For KLF4 reporter assays, constructs were designed that replaced the 3 zinc fingers of KLF4 with either the yeast GAL4 DNA-binding domain or the bacterial TetR DNA-binding domain. Plasmids were cloned via Gibson assembly with gBlocks (IDT) encoding wildtype, mutant, or Tat-ARM-swap versions of KLF4, and expression of the KLF4 fusions were driven by the human UbiC promoter (pJH438, pJH439, pJH441, pJH375, pJH376, pJH377, Key Resource Table). Reporter constructs contained either 6X UAS sites (Key Resource Table pJH437) or 4X TetO sites (Key Resource Table, pJH175) upstream of a minimal CMV promoter driving firefly luciferase. For GAL4 experiments, HEK293 cells were plated at $2x10^5$ cells per well in a 24-well plate in triplicate. Cells were transfected with 100 ng reporter, 166 ng KLF4 expression construct, and 50 ng of a renilla luciferase transfection control (pRL-SV40, Promega) the following day using Lipofectamine 3000 following the manufacturer's instructions. As a control, we included a pcDNA3 vector expressing LacI-mCherry (labeled as "No TF"). After 4 hours of incubation, luciferase activity was quantified by the Dual Luciferase Assay Kit (Promega)

following the manufacturer's instructions and a Safire II plate reader. The luminescence values were first normalized to the renilla luciferase luminescence for each well, and then all conditions were normalized to the average value of the "No TF" control condition. For TetR assays, HEK293 cells were plated at $1x10^5$ cells per well in a 24-well plate in triplicate in media containing tetracycline-free serum. The following day, cells were transfected with 100 ng reporter, 100 ng KLF4 expression construct, and 50 ng of renilla luciferase. After 2 hours of incubation, the media was removed and replaced with a media containing 1 µg/mL doxycycline. After 4 hours in dox, the cells were processed for luminescence readings in an identical fashion to the GAL4 assays.

Single-molecule tracking

Cell line generation

Murine embryonic stem cells were cultured in 2i/LIF media on tissue culture plates coated with 0.2% gelatin (Sigma, G1890). The 2i/LIF media contained: 960 mL DMEM/F12 (Life Technologies, 11320082), 5 mL N2 supplement (Life Technologies, 17502048; stock 100X), 10 mL B27 supplement (Life Technologies, 17504044; stock 50X), 5 mL additional L-glutamine (GIBCO 25030-081; stock 200 mM), 10 mL MEM nonessential amino acids (GIBCO 11140076; stock 100X), 10 mL penicillin-streptomycin (Life Technologies, 15140163; stock 10^4 U/mL), 333 mL BSA fraction V (GIBCO 15260037; stock 7.50%), 7 mL b-mercaptoethanol (Sigma M6250; stock 14.3 M), 100 mL LIF (Chemico, ESG1107; stock 10^7 U/mL), 100 mL PD0325901 (Stemgent, 04-0006-10; stock 10 mM), and 300 mL CHIR99021 (Stemgent, 04-0004-10; stock 10 mM). Cells were passaged by washing once with 1X PBS (Life Technologies, AM9625) and incubating with TrypLE (Life Technologies, 12604021) for 3-5 minutes, then quenched with serum-containing media made by the following recipe: 500 mL DMEM KO (GIBCO 10829-018), MEM nonessential amino acids (GIBCO 11140076; stock 100X), penicillin-streptomycin (Life Technologies, 15140163; stock 10^4 U/mL), 5 mL L-glutamine (GIBCO 25030-081; stock 100X), 4 mL b-mercaptoethanol (Sigma M6250; stock 14.3 M), 50 mL LIF (Chemico, ESG1107; stock 10^7 U/mL), and 75 mL of fetal bovine serum (Sigma, F4135). Cells were passaged every 2 days.

K562 cells were cultured in suspension flasks containing culture medium [RPMI-1640 medium with GlutaMAX™ (ThermoFisher Cat. 72400047) supplemented with 10% FBS (ThermoFisher Cat. 10437028), 2 mM L-glutamine (Sigma-Aldrich Cat. G7513), 50 U/mL penicillin and 50 µg/mL streptomycin].

A piggyBac compatible base vector was assembled containing two tandem gene cassettes: (1) an insertion site downstream of a doxycycline-inducible promoter allowing for the expression of a Flag-HA-Halo-tagged ORF with SV40 NLS and bGH polyA termination sequence, and (2) the Tet-On 3G rtta element driven by the EF1a promoter that also produces hygromycin resistance via a 2A self-cleaving peptide. This base vector was generated by Gibson assembly. Plasmids encoding Halo-tagged versions of TFs (WT and ARM-deletion) were generated by Gibson assembly with BamHI-digested base vector and gBlocks (Integrated DNA Technologies) encoding the WT and ARM-deletion TFs. See Key Resource Table for plasmid information

(pJH294, pJH295, pJH290, pJH291, pJH357, pJH358, pJH337, pJH342, pJH308, pJH309; PBFH vectors stand for “PiggyBac Flag HA”).

To generate cell lines, $5x10^6$ mESCs or K562 cells per well were transfected in 6-well plates with 1 µg of the Halo-TF vector and 1 µg of the piggyBac transposase (Systems Biosciences) in serum-containing media (described above) using Lipofectamine-3000 for at least 4 hours. After transfection, the cells were passaged into 10 cm plates in 2i media or K562 media containing 500 µg/mL Hygromycin-B (Gibco 10687010). After 2-4 days of selection for mESC and 2 weeks of selection for K562, cells were maintained as described above.

Sample preparation
mESCs were plated on glass bottom dishes (Mattek Corporation P35G-1.5-20-C) coated with 5 µg/ml of poly-L-ornithine (Sigma-Aldrich P4957) for 2hrs min at 37°C and with 5µg/ml of Laminin (Corning® 354232) for 2hrs-24hrs at 37°C, growing from 20% confluency in 2i for one day. K562 cells were plated on poly-L-lysine coated glass bottom dishes and allowed to attach for at least 4 hours. Doxycycline=10ng/mL was added to dishes for 1hr, followed by adding 5nM of HaloTag-(PA) JF549 for another 3hrs. Cells were then rinsed once with PBS and washed in fresh 2i for 1hr. Dishes were refilled with 2mL prewarmed Leibovitz's L-15 Medium, no phenol red (ThermoFisher 21083027) and brought for imaging.

Imaging
Cells were imaged on an inverted, widefield setup with a Nikon Eclipse Ti microscope and a 100x oil immersion objective as previously described[58]. Images were acquired with an EMCCD camera (EM gain 1000, exposure time 10ms, conjugated pixel-size on sample 160nm). A 561nm laser beam of 150mW (attenuated with 50% AOTF) was 2x expanded for a uniform illumination across around 200x200 pixel region. 10,000 frames were recorded for each ROI (including 2-4 cells), and the 405nm activation was kept very low to guarantee the molecule sparsity needed for robust reconnection.

Analyses
Particle trajectories were detected and reconnected with customized MATLAB code from MTT[107]. Detection settings: false-positive threshold=24, window-size 7x7pixel, and Gaussian width fitting allowed. Reconnection settings: T_{off}=10ms, T_{cut}=20ms, and r_{max}=270nm. A collection of trajectories from each ROI were fitted to a 3-state model in Spot-on[108]. Spot-on settings: detection slice dZ=950nm, 8 delays to consider, and only first 10 jumps to consider for each trajectory. The final outputs include fractions and apparent diffusion coefficients of each state (immobile, sub-diffusive, and free, respectively). For expression dependence testing in Figure S7F, trajectories of the same genotype from different nuclei with similar trajectory density were gathered together first and resampled ten times (2,000 trajectories for each resampling) for ten independent Spot-on fittings, respectively. In this way, the accuracy of each fitting and the distributions across different conditions are comparable.

For dwell time analyses in Figure S7E, sparse detections from slow tracking mode were generated with the same MTT settings as for those in the fast tracking. The detections were

then grouped to different spatial clusters by running a Density-based spatial clustering of applications with noise (DBSCAN) with short radius. Within each spatial cluster, the time-correlated detections were further grouped into the same trajectory (two dark frames at maximum). In this manner, only immobile (i.e., bound) trajectories will be collected, whose duration (t_{last}-t_{first}) were the apparent dwelling time. The survival probabilities of apparent dwelling time distributions were fitted to a biexponential model for both fixed and live cell samples, where a short dwelling time scale and a long dwelling time scale were fitted. The stable dwell time of each live cell sample was based on the long dwelling time scale, which was calibrated by the long dwelling time scale of a fixed sample with the exact imaging condition as following:

$$\frac{1}{\hat{\tau}_{\mathrm{cali}}} = \frac{1}{\tau_{\mathrm{live}}} - \frac{1}{\tau_{\mathrm{fix}}},$$

where τ_{live} is the "apparent" long dwelling time scale of the live sample, τ_{fix} is the "apparent" long dwelling time scale of a fixed sample on the same date in the same imaging buffer, and $\hat{\tau}_{\mathrm{cali}}$ is the calibrated stable dwell time actually reported in final figures.

For curve fitting in Figure S7B, the sum of N constrained Gaussian functions is fitted to the probability distributions of logarithm of diffusion coefficients (logD). The total amplitude of N Gaussians is constrained to 1. The center of each Gaussian is constrained within an interval, where the N intervals are determined based on N equally assigned quantiles of the logD distribution. The sigma of each Gaussian is contained below the half width of the corresponding interval. The diffusion coefficients are fitted from 3,000-20,000 individual single-molecule trajectories with at least 5 jumps in assumption of 2-dimensional Brownian motion. Only the first 16 jumps are used to fit the diffusion coefficient if there are more than 16 jumps for a given trajectory.

For subdiffusion analysis in Figure S7C, single-molecule trajectories for KLF4, SOX2, CTCF, GATA2, and RUNX1 were analyzed by computing the mean squared displacement (MSD) of particles as a function of lag time τ according to a standard method[109]. MSD was computed for trajectories containing 5 or more time steps and the final lag time was trimmed from each trace prior to fitting. Traces were fit to the 2D anomalous diffusion equation:

$$MSD(\tau) = \langle \Delta r^2(\tau) \rangle = 4K_\alpha \tau^\alpha$$

where MSD is the mean squared displacement (μm^2) for each trajectory, r is the radial displacement of the particle (μm), Kα is the generalized diffusion coefficient ($\mu m^2/s^\alpha$), τ is the lag time (s), and α is the anomalous diffusion exponent. Two-parameter curve fitting of α and K_α was performed in Python using the 'curve_fit' function in scipy (v. 1.10.1). Traces with α between 10-5 and 0.8 and with $\alpha > 10*\sigma^2(\alpha)$, where σ(α) is the standard deviation of the α parameter estimate, were classified as subdiffusive. For visualization purposes, datasets for KLF4, CTCF, and SOX2 were randomly downsampled to display 5% of traces.

Sub-nuclear fractionation

mESCs with exogenous expression for SOX2 and KLF4 wild type and ARM deletion mutations expressing HA tag were used for nuclei sub fractionation. To extract nuclei, cells were resuspended in 10 ml HMSD50 buffer (20 mM HEPES pH 7.5, 5 mM MgCl2, 250 mM sucrose,

1mM DTT, 50mM NaCl, supplemented with 0.2 mM PMSF and 5 mM sodium butyrate) and incubated for 30 min at 4°C with gentle agitation. After a spin down at 3500 rpm at 4°C for 10 min, the supernatant was discarded and the pellet containing nuclei were subjected to subcellular protein fractionation for nucleoplasm and chromatin fractions using the Subcellular Protein Fractionation Kit for Cultured Cells (ThermoScientific, Ref 78840) according to manufacturer's instructions. For RNase treatment in wild type mESCs, nuclei were treated with RNase A (1:100, Thermo Fisher EN0531) and the initial 30-minute incubation at 4°C was adjusted to 20 minutes at 4°C and 10 minutes at 37°C. The pH of the buffer remained the same (~7.5) after RNase A treatment. SDS Page was run on 12% Bis-Tris gel (Criterion XT, BioRad) and western blotting was performed on the subfractions using anti Histone H3 antibody from Abcam (ab1791) and anti HA antibody from Abcam (ab9110) with secondary antibody against Rabbit (IRDye 800CW Goat anti-rabbit LI-COR 926-32211). For wild type transcription factor detection, antibody for Sox2 (R&D Systems, MAB2018) and Klf4 (R&D Systems, AF3158) with secondary antibody anti-mouse for Sox2 (IRDye 680CW goat anti-mouse LI-COR 926-32211) and anti-goat for Klf4 (IRDye 800CW donkey anti-goat LI-COR 926-32214), were used. Fluorescence was assessed using Odyssey CLX LiCOR and quantified using Fiji/ImageJ[111].

Zebrafish knockdown and rescue of *sox2*

Morpholinos (MO, GeneTools) were resuspended in nuclease free water, heated to 65°C for 5 minutes, and stored at room temperature. Wildtype AB zebrafish embryos were injected into the yolk at the 1-cell stage with 7ng of *sox2*-MO (TCTTGAAAGTCTACCCCACCAGCCG)[53], either alone or in combination with 25 pg of human wildtype or ARM-deletion SOX2 mRNA. Messenger RNA was synthesized using the T7 mMessage mMachine (Invitrogen) kit with templates generated from gBlocks (IDT). The mRNA was purified with the MEGAclear Clean-Up Kit (Invitrogen), run on a TBE agarose gel to confirm purity and size, aliquoted, and stored at -80°C. Embryos injected with 7ng of Standard Control MO (CCTCTTACCTCAGTTACAATTTATA) were used as controls. At 48 hours post fertilization (hpf), MO injected embryos were dechorionated using forceps, anaesthetized using 0.16 mg/ml Tricaine, then visually assessed for growth impairment using a Nikon SMZ18 stereoscope with DS-Ri2 camera and NIS-Elements software. Embryos were scored based on rescue of growth impairment in the presence of wildtype or mutant sox2 mRNA.

To assure that mutant SOX2 was expressed as protein, we conducted Western blots (Figure S7G). Protein extraction for zebrafish embryos (n = 20 per tube) that were uninjected or injected with mRNA encoding HA-tagged ARM-mutant SOX2 was performed with Urea Chaps lysis buffer. Cells were resuspended in Urea Chaps (1% Chaps, 8M Urea, 50mM Tris-Cl pH 7.5 containing protease inhibitors (Thermo Fisher)) and incubated for 30' at 4°C with gentle agitation. After a spin down at 14,000 rpm for 10' at 4°C, the supernatant was used for SDS-Page. SDS-Page was run on a 10% Bis-Tris (Criterion XT, BioRad) and western blotting was performed on uninjected and injected samples using anti HA antibody from Abcam (ab9110) and anti beta actin (Sigma A5441) with secondary antibody against Rabbit (IRDye 800CW Goat anti-rabbit LI-COR 926-32211 and IRDye 680RD Goat anti-mouse 926-68070). Fluorescence was assessed using Odyssey CLX LiCOR.

Overlap of pathogenic mutations in TF-ARMs

Pathogenic nonsynonmous substitution mutations were obtained from a prior dataset of pathogenic mutations that integrated multiple databases of somatic and germline variation associated with cancer and Mendelian disorders, including ClinVar (accessed January 29, 2021) and HGMD v2020.4 in hg38. Cancer variants were obtained from AACR Project GENIE v8.1 (AACR Project GENIE Consortium, 2017) and various TCGA and TARGET studies via cBioPortal[110]. Mutations were subsetted for those affecting TF-ARMs. For mutation frequency analysis, the expected mutation frequency for each amino acid type within TF-ARMs was estimated using the average nucleotide substitution rates within the entire mutation dataset and the frequency of nucleotide types encoding each amino acid type within TF-ARMs. It is important to note that this analysis does not take into account disease-specific mutational signatures, which could introduce potential biases. Enrichment was defined as a significantly higher pathogenic mutation frequency compared to the aforementioned expected amino acid mutation frequency. Statistical significance of the enrichment was determined using a one-sided binomial test, and p-values were corrected for the multiple tests across the twenty amino acids using the Benjamini-Hochberg method.

QUANTIFICATION AND STATISTICAL ANALYSIS

Details of quantification and statistical analysis for each experiment can be found in their respective section, and we provide additional details on sample sizes and statistical parameters here. Statistical tests were conducted using Prism software (GraphPad). Confidence intervals for K_d estimates from fluorescence polarization data were computed by multiplying the standard deviation of the K_d curve fit parameter with the Student's t-value corresponding to the 95% confidence interval with degrees of freedom equal to the number of data points in the concentration curve minus the number of fit parameters. Statistical comparisons between the K_d's of two fluorescence polarization curves (for Figure 3E, Figure S2E, and Figure S4) were assessed using a two-tailed Student's t-test based on the standard errors of the K_d parameters calculated from the diagonals of the covariance matrix returned by 'curve_fit' in scipy, with the degrees of freedom as specified above.

The distributions of ARM correlation scores (Figure 3C) for whole proteome (-TFs) vs TFs were compared using a two-tailed Mann Whitney U test, n1=1287, n2=20238.

The Tat reporter assays were conducted on 3 biological replicates per genotype, and luminescence readings were measured in technical duplicates. Each condition was compared to the Tat R/K>A condition using a Sidak multiple comparisons test (DF = 24, t statistics were as follow: TAR-WT - WT=20.15, KLF4=15.3, SOX2=13.17, GATA2=3.805, NoTat=6.419; ΔTAR-bulge – WT=9.263, KLF4=9.319, SOX2=9.329, GATA2=9.315, Tat R/K>A=9.302, No-Tat=9.364).

For comparison of the diffusive fractions reported in Figure 5C, multiple fields of cells were imaged per genotype (KLF4-WT n=11, KLF4-ΔARM n=9, SOX2-WT n=10, SOX2-ΔARM n=9, CTCF-WT n=7, CTCF-ΔARM n=7). The diffusive fractions were compared by 2-tailed Student t-test. The data was confirmed to have equal variance via F test, and the degrees of freedom and t statistics were as follows: KLF4-free (t=13.47, df=18), SOX2-free (t=8.297, df=18), CTCF-free

(t=6.044, df=12), KLF4-sub (t=5.152, df=18), SOX2-sub (2.908, df=18), CTCF-sub (t=3.051, df=12), KLF4-imm (t=7.824, df=18), SOX2-imm (t=6.203, df=18), CTCF-imm (t=3.639, df=12).

References

1. Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. Cell *172*, 650–665. 10.1016/j.cell.2018.01.029.
2. Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. Nat. Rev. Genet. *10*, 252–263. 10.1038/nrg2538.
3. Cramer, P. (2019). Organization and regulation of gene transcription. Nature *573*, 45–54. 10.1038/s41586-019-1517-4.
4. Lee, T.I., and Young, R.A. (2013). Transcriptional regulation and its misregulation in disease. Cell *152*, 1237–1251. 10.1016/j.cell.2013.02.014.
5. Stadhouders, R., Filion, G.J., and Graf, T. (2019). Transcription factors and 3D genome conformation in cell-fate decisions. Nature *569*, 345–354. 10.1038/s41586-019-1182-7.
6. Panne, D., Maniatis, T., and Harrison, S.C. (2007). An Atomic Model of the Interferon-β Enhanceosome. Cell *129*, 1111–1123. 10.1016/j.cell.2007.05.019.
7. Avsec, Ž., Weilert, M., Shrikumar, A., Krueger, S., Alexandari, A., Dalal, K., Fropf, R., McAnany, C., Gagneur, J., Kundaje, A., et al. (2021). Base-resolution models of transcription-factor binding reveal soft motif syntax. Nat. Genet. *53*, 354–366. 10.1038/s41588-021-00782-6.
8. Arnold, C.D., Nemčko, F., Woodfin, A.R., Wienerroither, S., Vlasova, A., Schleiffer, A., Pagani, M., Rath, M., and Stark, A. (2018). A high-throughput method to identify trans-activation domains within transcription factor sequences. EMBO J. *37*, e98896. 10.15252/embj.201798896.
9. Boija, A., Klein, I.A., Sabari, B.R., Dall'Agnese, A., Coffey, E.L., Zamudio, A.V., Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M., et al. (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell *175*, 1842-1855.e16. 10.1016/j.cell.2018.10.042.
10. Soto, L.F., Li, Z., Santoso, C.S., Berenson, A., Ho, I., Shen, V.X., Yuan, S., and Fuxman Bass, J.I. (2022). Compendium of human transcription factor effector domains. Mol. Cell *82*, 514–526. 10.1016/j.molcel.2021.11.007.
11. Richter, W.F., Nayak, S., Iwasa, J., and Taatjes, D.J. (2022). The Mediator complex as a master regulator of transcription by RNA polymerase II. Nat. Rev. Mol. Cell Biol., 1–18. 10.1038/s41580-022-00498-3.
12. Vos, S.M. (2021). Understanding transcription across scales: From base pairs to chromosomes. Mol. Cell *81*, 1601–1616. 10.1016/j.molcel.2021.03.002.
13. Lelli, K.M., Slattery, M., and Mann, R.S. (2012). Disentangling the many layers of eukaryotic transcriptional regulation. Annu. Rev. Genet. *46*, 43–68. 10.1146/annurev-genet-110711-155437.
14. Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. *13*, 613–626. 10.1038/nrg3207.
15. Kaikkonen, M.U., and Adelman, K. (2018). Emerging Roles of Non-Coding RNA Transcription. Trends Biochem. Sci. *43*, 654–667. 10.1016/j.tibs.2018.06.002.
16. Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent Transcription from Active Promoters. Science *322*, 1849–1851. 10.1126/science.1162253.
17. Cassiday, L.A., and Maher, L.J. (2002). Having it both ways: transcription factors that bind DNA and RNA. Nucleic Acids Res. *30*, 4118–4126. 10.1093/nar/gkf512.
18. Holmes, Z.E., Hamilton, D.J., Hwang, T., Parsonnet, N.V., Rinn, J.L., Wuttke, D.S., and Batey, R.T. (2020). The Sox2 transcription factor binds RNA. Nat. Commun. *11*, 1805. 10.1038/s41467-020-15571-8.
19. Hou, L., Wei, Y., Lin, Y., Wang, X., Lai, Y., Yin, M., Chen, Y., Guo, X., Wu, S., Zhu, Y., et al. (2020). Concurrent binding to DNA and RNA facilitates the pluripotency reprogramming activity of Sox2. Nucleic Acids Res. *48*, 3869–3887. 10.1093/nar/gkaa067.
20. Saldaña-Meyer, R., Rodriguez-Hernaez, J., Escobar, T., Nishana, M., Jácome-López, K., Nora, E.P., Bruneau, B.G., Tsirigos, A., Furlan-Magaril, M., Skok, J., et al. (2019). RNA Interactions Are Essential for CTCF-Mediated Genome Organization. Mol. Cell *76*, 412-422.e5. 10.1016/j.molcel.2019.08.015.

21. Sigova, A.A., Abraham, B.J., Ji, X., Molinie, B., Hannett, N.M., Guo, Y.E., Jangi, M., Giallourakis, C.C., Sharp, P.A., and Young, R.A. (2015). Transcription factor trapping by RNA in gene regulatory elements. Science *350*, 978–981. 10.1126/science.aad3346.
22. Theunissen, O., Rudt, F., Guddat, U., Mentzel, H., and Pieler, T. (1992). RNA and DNA binding zinc fingers in Xenopus TFIIIA. Cell *71*, 679–690. 10.1016/0092-8674(92)90601-8.
23. Xu, Y., Huangyang, P., Wang, Y., Xue, L., Devericks, E., Nguyen, H.G., Yu, X., Oses-Prieto, J.A., Burlingame, A.L., Miglani, S., et al. (2021). ERα is an RNA-binding protein sustaining tumor cell survival and drug resistance. Cell *0*. 10.1016/j.cell.2021.08.036.
24. Jeon, Y., and Lee, J.T. (2011). YY1 tethers Xist RNA to the inactive X nucleation center. Cell *146*, 119–133. 10.1016/j.cell.2011.06.026.
25. Yoshida, Y., Izumi, H., Torigoe, T., Ishiguchi, H., Yoshida, T., Itoh, H., and Kohno, K. (2004). Binding of RNA to p53 regulates its oligomerization and DNA-binding activity. Oncogene *23*, 4371–4379. 10.1038/sj.onc.1207583.
26. Steiner, H.R., Lammer, N.C., Batey, R.T., and Wuttke, D.S. (2022). An Extended DNA Binding Domain of the Estrogen Receptor Alpha Directly Interacts with RNAs in Vitro. Biochemistry *61*, 2490–2494. 10.1021/acs.biochem.2c00536.
27. Niessing, D., Driever, W., Sprenger, F., Taubert, H., Jäckle, H., and Rivera-Pomar, R. (2000). Homeodomain Position 54 Specifies Transcriptional versus Translational Control by Bicoid. Mol. Cell *5*, 395–401. 10.1016/S1097-2765(00)80434-7.
28. Dvir, S., Argoetti, A., Lesnik, C., Roytblat, M., Shriki, K., Amit, M., Hashimshony, T., and Mandel-Gutfreund, Y. (2021). Uncovering the RNA-binding protein landscape in the pluripotency network of human embryonic stem cells. Cell Rep. *35*. 10.1016/j.celrep.2021.109198.
29. Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. Nat. Rev. Mol. Cell Biol. *8*, 479–490. 10.1038/nrm2178.
30. Wheeler, E.C., Van Nostrand, E.L., and Yeo, G.W. (2018). Advances and challenges in the detection of transcriptome-wide protein-RNA interactions. Wiley Interdiscip. Rev. RNA *9*, e1436. 10.1002/wrna.1436.
31. He, C., Sidoli, S., Warneford-Thomson, R., Tatomer, D.C., Wilusz, J.E., Garcia, B.A., and Bonasio, R. (2016). High-Resolution Mapping of RNA-Binding Regions in the Nuclear Proteome of Embryonic Stem Cells. Mol. Cell *64*, 416–430. 10.1016/j.molcel.2016.09.034.
32. Orkin, S.H., and Zon, L.I. (2008). Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. Cell *132*, 631–644. 10.1016/j.cell.2008.01.025.
33. Delgado, M.D., Lerga, A., Cañelles, M., Gómez-Casares, M.T., and León, J. (1995). Differential regulation of Max and role of c-Myc during erythroid and myelomonocytic differentiation of K562 cells. Oncogene *10*, 1659–1665.
34. Young, R.A. (2011). Control of the embryonic stem cell state. Cell *144*, 940–954. 10.1016/j.cell.2011.01.032.
35. Ibarra, A., Benner, C., Tyagi, S., Cool, J., and Hetzer, M.W. (2016). Nucleoporin-mediated regulation of cell identity genes. Genes Dev. *30*, 2253–2258. 10.1101/gad.287417.116.
36. Saldaña-Meyer, R., González-Buendía, E., Guerrero, G., Narendra, V., Bonasio, R., Recillas-Targa, F., and Reinberg, D. (2014). CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. Genes Dev. *28*, 723–734. 10.1101/gad.236869.113.
37. Burd, C.G., and Dreyfuss, G. (1994). RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. EMBO J. *13*, 1197–1204.
38. Corley, M., Burns, M.C., and Yeo, G.W. (2020). How RNA-Binding Proteins Interact with RNA: Molecules and Mechanisms. Mol. Cell *78*, 9–29. 10.1016/j.molcel.2020.03.011.
39. Maji, D., Glasser, E., Henderson, S., Galardi, J., Pulvino, M.J., Jenkins, J.L., and Kielkopf, C.L. (2020). Representative cancer-associated U2AF2 mutations alter RNA interactions and splicing. J. Biol. Chem. *295*, 17148–17157. 10.1074/jbc.RA120.015339.
40. Zhang, J., Lieu, Y.K., Ali, A.M., Penson, A., Reggio, K.S., Rabadan, R., Raza, A., Mukherjee, S., and Manley, J.L. (2015). Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities. Proc. Natl. Acad. Sci. U. S. A. *112*, E4726–E4734. 10.1073/pnas.1514105112.
41. Calnan, B.J., Biancalana, S., Hudson, D., and Frankel, A.D. (1991). Analysis of arginine-rich peptides from the HIV Tat protein reveals unusual features of RNA-protein recognition. Genes Dev. *5*, 201–210. 10.1101/gad.5.2.201.

42. Calnan, B.J., Tidor, B., Biancalana, S., Hudson, D., and Frankel, A.D. (1991). Arginine-Mediated RNA Recognition: the Arginine Fork. Science *252*, 1167–1171. 10.1126/science.252.5009.1167.
43. Pham, V.V., Salguero, C., Khan, S.N., Meagher, J.L., Brown, W.C., Humbert, N., de Rocquigny, H., Smith, J.L., and D'Souza, V.M. (2018). HIV-1 Tat interactions with cellular 7SK and viral TAR RNAs identifies dual structural mimicry. Nat. Commun. *9*, 4266. 10.1038/s41467-018-06591-6.
44. Jakobovits, A., Smith, D.H., Jakobovits, E.B., and Capon, D.J. (1988). A discrete element 3' of human immunodeficiency virus 1 (HIV-1) and HIV-2 mRNA initiation sites mediates transcriptional activation by an HIV trans activator. Mol. Cell. Biol. *8*, 2555–2561. 10.1128/mcb.8.6.2555-2561.1988.
45. Ghaleb, A.M., and Yang, V.W. (2017). Krüppel-like factor 4 (KLF4): What we currently know. Gene *611*, 27–37. 10.1016/j.gene.2017.02.025.
46. Geiman, D.E., Ton-That, H., Johnson, J.M., and Yang, V.W. (2000). Transactivation and growth suppression by the gut-enriched Krüppel-like factor (Krüppel-like factor 4) are dependent on acidic amino acid residues and protein-protein interaction. Nucleic Acids Res. *28*, 1106–1113. 10.1093/nar/28.5.1106.
47. Yet, S.F., McA'Nulty, M.M., Folta, S.C., Yen, H.W., Yoshizumi, M., Hsieh, C.M., Layne, M.D., Chin, M.T., Wang, H., Perrella, M.A., et al. (1998). Human EZF, a Krüppel-like zinc finger protein, is expressed in vascular endothelial cells and contains transcriptional activation and repression domains. J. Biol. Chem. *273*, 1026–1031. 10.1074/jbc.273.2.1026.
48. Chen, J., Zhang, Z., Li, L., Chen, B.-C., Revyakin, A., Hajj, B., Legant, W., Dahan, M., Lionnet, T., Betzig, E., et al. (2014). Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. Cell *156*, 1274–1285. 10.1016/j.cell.2014.01.062.
49. Nguyen, V.Q., Ranjan, A., Liu, S., Tang, X., Ling, Y.H., Wisniewski, J., Mizuguchi, G., Li, K.Y., Jou, V., Zheng, Q., et al. (2021). Spatiotemporal coordination of transcription preinitiation complex assembly in live cells. Mol. Cell, S1097276521005918. 10.1016/j.molcel.2021.07.022.
50. Garcia, D.A., Johnson, T.A., Presman, D.M., Fettweis, G., Wagh, K., Rinaldi, L., Stavreva, D.A., Paakinaho, V., Jensen, R.A.M., Mandrup, S., et al. (2021). An intrinsically disordered region-mediated confinement state contributes to the dynamics and function of transcription factors. Mol. Cell *81*, 1484-1498.e6. 10.1016/j.molcel.2021.01.013.
51. Garcia, D.A., Fettweis, G., Presman, D.M., Paakinaho, V., Jarzynski, C., Upadhyaya, A., and Hager, G.L. (2021). Power-law behavior of transcription factor dynamics at the single-molecule level implies a continuum affinity model. Nucleic Acids Res. *49*, 6605–6620. 10.1093/nar/gkab072.
52. Hansen, A.S., Amitai, A., Cattoglio, C., Tjian, R., and Darzacq, X. (2020). Guided nuclear exploration increases CTCF target search efficiency. Nat. Chem. Biol. *16*, 257–266. 10.1038/s41589-019-0422-3.
53. Pavlou, S., Astell, K., Kasioulis, I., Gakovic, M., Baldock, R., Heyningen, V. van, and Coutinho, P. (2014). Pleiotropic Effects of Sox2 during the Development of the Zebrafish Epithalamus. PLOS ONE *9*, e87546. 10.1371/journal.pone.0087546.
54. Boldes, T., Merenbakh-Lamin, K., Journo, S., Shachar, E., Lipson, D., Yeheskel, A., Pasmanik-Chor, M., Rubinek, T., and Wolf, I. (2020). R269C variant of ESR1: high prevalence and differential function in a subset of pancreatic cancers. BMC Cancer *20*, 531. 10.1186/s12885-020-07005-x.
55. Keegan, L., Gill, G., and Ptashne, M. (1986). Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. Science *231*, 699–704. 10.1126/science.3080805.
56. Tjian, R., and Maniatis, T. (1994). Transcriptional activation: a complex puzzle with few easy pieces. Cell *77*, 5–8. 10.1016/0092-8674(94)90227-5.
57. Asimi, V., Sampath Kumar, A., Niskanen, H., Riemenschneider, C., Hetzel, S., Naderi, J., Fasching, N., Popitsch, N., Du, M., Kretzmer, H., et al. (2022). Hijacking of transcriptional condensates by endogenous retroviruses. Nat. Genet., 1–10. 10.1038/s41588-022-01132-w.
58. Henninger, J.E., Oksuz, O., Shrinivas, K., Sagi, I., LeRoy, G., Zheng, M.M., Andrews, J.O., Zamudio, A.V., Lazaris, C., Hannett, N.M., et al. (2021). RNA-Mediated Feedback Control of Transcriptional Condensates. Cell *184*, 207-225.e24. 10.1016/j.cell.2020.11.030.
59. Sharp, P.A., Chakraborty, A.K., Henninger, J.E., and Young, R.A. (2022). RNA in formation and regulation of transcriptional condensates. RNA N. Y. N *28*, 52–57. 10.1261/rna.078997.121.
60. Quinodoz, S.A., Jachowicz, J.W., Bhat, P., Ollikainen, N., Banerjee, A.K., Goronzy, I.N., Blanco, M.R., Chovanec, P., Chow, A., Markaki, Y., et al. (2021). RNA promotes the formation of spatial compartments in the nucleus. Cell *184*, 5775-5790.e30. 10.1016/j.cell.2021.10.014.

61. Bose, D.A., Donahue, G., Reinberg, D., Shiekhattar, R., Bonasio, R., and Berger, S.L. (2017). RNA Binding to CBP Stimulates Histone Acetylation and Transcription. Cell *168*, 135-149.e22. 10.1016/j.cell.2016.12.020.
62. Lai, F., Orom, U.A., Cesaroni, M., Beringer, M., Taatjes, D.J., Blobel, G.A., and Shiekhattar, R. (2013). Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature *494*, 497–501. 10.1038/nature11884.
63. Long, Y., Wang, X., Youmans, D.T., and Cech, T.R. (2017). How do lncRNAs regulate transcription? Sci. Adv. *3*, eaao2110. 10.1126/sciadv.aao2110.
64. Hemphill, W.O., Voong, C.K., Fenske, R., Goodrich, J.A., and Cech, T.R. (2022). RNA- and DNA-binding proteins generally exhibit direct transfer of polynucleotides: Implications for target site search. 2022.11.30.518605. 10.1101/2022.11.30.518605.
65. Reimer, K.A., Mimoso, C.A., Adelman, K., and Neugebauer, K.M. (2021). Co-transcriptional splicing regulates 3′ end cleavage during mammalian erythropoiesis. Mol. Cell *81*, 998-1012.e7. 10.1016/j.molcel.2020.12.018.
66. Brodsky, S., Jana, T., Mittelman, K., Chapal, M., Kumar, D.K., Carmi, M., and Barkai, N. (2020). Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding Specificity. Mol. Cell *79*, 459-471.e4. 10.1016/j.molcel.2020.05.032.
67. Inukai, S., Kock, K.H., and Bulyk, M.L. (2017). Transcription factor–DNA binding: beyond binding site motifs. Curr. Opin. Genet. Dev. *43*, 110–119. 10.1016/j.gde.2017.02.007.
68. Wasserman, W.W., and Sandelin, A. (2004). Applied bioinformatics for the identification of regulatory elements. Nat. Rev. Genet. *5*, 276–287. 10.1038/nrg1315.
69. Han, H., Braunschweig, U., Gonatopoulos-Pournatzis, T., Weatheritt, R.J., Hirsch, C.L., Ha, K.C.H., Radovani, E., Nabeel-Shah, S., Sterne-Weiler, T., Wang, J., et al. (2017). Multilayered Control of Alternative Splicing Regulatory Networks by Transcription Factors. Mol. Cell *65*, 539-553.e7. 10.1016/j.molcel.2017.01.011.
70. Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., and Ferrin, T.E. (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein Sci. Publ. Protein Soc. *27*, 14–25. 10.1002/pro.3235.
71. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. Publ. Protein Soc. *30*, 70–82. 10.1002/pro.3943.
72. Demichev, V., Messner, C.B., Vernardis, S.I., Lilley, K.S., and Ralser, M. (2020). DIA-NN: Neural networks and interference correction enable deep proteome coverage in high throughput. Nat. Methods *17*, 41–44. 10.1038/s41592-019-0638-x.
73. Nesvizhskii, A.I., Keller, A., Kolker, E., and Aebersold, R. (2003). A statistical model for identifying proteins by tandem mass spectrometry. Anal. Chem. *75*, 4646–4658. 10.1021/ac0341261.
74. Hochberg, Y., and Benjamini, Y. (1990). More powerful procedures for multiple significance testing. Stat. Med. *9*, 811–818. 10.1002/sim.4780090710.
75. Baltz, A.G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., et al. (2012). The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. Mol. Cell *46*, 674–690. 10.1016/j.molcel.2012.05.021.
76. Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012). Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. Cell *149*, 1393–1406. 10.1016/j.cell.2012.04.031.
77. Kwon, S.C., Yi, H., Eichelbaum, K., Föhr, S., Fischer, B., You, K.T., Castello, A., Krijgsveld, J., Hentze, M.W., and Kim, V.N. (2013). The RNA-binding protein repertoire of embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1122–1130. 10.1038/nsmb.2638.
78. Bao, X., Guo, X., Yin, M., Tariq, M., Lai, Y., Kanwal, S., Zhou, J., Li, N., Lv, Y., Pulido-Quetglas, C., et al. (2018). Capturing the interactome of newly transcribed RNA. Nat. Methods *15*, 213–220. 10.1038/nmeth.4595.
79. Huang, R., Han, M., Meng, L., and Chen, X. (2018). Transcriptome-wide discovery of coding and noncoding RNA-binding proteins. Proc. Natl. Acad. Sci. *115*, E3879–E3887. 10.1073/pnas.1718406115.
80. Trendel, J., Schwarzl, T., Horos, R., Prakash, A., Bateman, A., Hentze, M.W., and Krijgsveld, J. (2019). The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. Cell *176*, 391-403.e19. 10.1016/j.cell.2018.11.004.

81. Queiroz, R.M.L., Smith, T., Villanueva, E., Marti-Solano, M., Monti, M., Pizzinga, M., Mirea, D.-M., Ramakrishna, M., Harvey, R.F., Dezi, V., et al. (2019). Comprehensive identification of RNA-protein interactions in any organism using orthogonal organic phase separation (OOPS). Nat. Biotechnol. *37*, 169–178. 10.1038/s41587-018-0001-2.
82. He, C., Bozler, J., Janssen, K.A., Wilusz, J.E., Garcia, B.A., Schorn, A.J., and Bonasio, R. (2021). TET2 chemically modifies tRNAs and regulates tRNA fragment levels. Nat. Struct. Mol. Biol. *28*, 62–70. 10.1038/s41594-020-00526-w.
83. Blue, S.M., Yee, B.A., Pratt, G.A., Mueller, J.R., Park, S.S., Shishkin, A.A., Starner, A.C., Van Nostrand, E.L., and Yeo, G.W. (2022). Transcriptome-wide identification of RNA-binding protein binding sites using seCLIP-seq. Nat. Protoc. *17*, 1223–1265. 10.1038/s41596-022-00680-z.
84. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10–12. 10.14806/ej.17.1.200.
85. Smith, T., Heger, A., and Sudbery, I. (2017). UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res. *27*, 491–499. 10.1101/gr.209601.116.
86. Langmead, B., Wilks, C., Antonescu, V., and Charles, R. (2019). Scaling read aligners to hundreds of threads on general-purpose processors. Bioinformatics *35*, 421–432. 10.1093/bioinformatics/bty648.
87. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. 10.1038/nmeth.1923.
88. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinforma. Oxf. Engl. *25*, 2078–2079. 10.1093/bioinformatics/btp352.
89. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. 10.1093/bioinformatics/btq033.
90. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based Analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137. 10.1186/gb-2008-9-9-r137.
91. Fujiwara, T., O'Geen, H., Keles, S., Blahnik, K., Linnemann, A.K., Kang, Y.-A., Choi, K., Farnham, P.J., and Bresnick, E.H. (2009). Discovering Hematopoietic Mechanisms Through Genome-Wide Analysis of GATA Factor Chromatin Occupancy. Mol. Cell *36*, 667–681. 10.1016/j.molcel.2009.11.001.
92. Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74. 10.1038/nature11247.
93. Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. Cell *153*, 307–319. 10.1016/j.cell.2013.03.035.
94. Guo, Y.E., Manteiga, J.C., Henninger, J.E., Sabari, B.R., Dall'Agnese, A., Hannett, N.M., Spille, J.-H., Afeyan, L.K., Zamudio, A.V., Shrinivas, K., et al. (2019). Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. Nature, 1–6. 10.1038/s41586-019-1464-0.
95. Sharma, D., Zagore, L.L., Brister, M.M., Ye, X., Crespo-Hernández, C.E., Licatalosi, D.D., and Jankowsky, E. (2021). The kinetic landscape of an RNA-binding protein in cells. Nature *591*, 152–156. 10.1038/s41586-021-03222-x.
96. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., et al. (2021). Pfam: The protein families database in 2021. Nucleic Acids Res. *49*, D412–D419. 10.1093/nar/gkaa913.
97. Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. Nat. Rev. Genet. *15*, 829–845. 10.1038/nrg3813.
98. Holehouse, A.S., Das, R.K., Ahad, J.N., Richardson, M.O.G., and Pappu, R.V. (2017). CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. Biophys. J. *112*, 16–21. 10.1016/j.bpj.2016.11.3200.
99. Li, C.H., Coffey, E.L., Dall'Agnese, A., Hannett, N.M., Tang, X., Henninger, J.E., Platt, J.M., Oksuz, O., Zamudio, A.V., Afeyan, L.K., et al. (2020). MeCP2 links heterochromatin condensates and neurodevelopmental disease. Nature. 10.1038/s41586-020-2574-4.

100. Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., et al. (2021). The InterPro protein families and domains database: 20 years on. Nucleic Acids Res. *49*, D344–D354. 10.1093/nar/gkaa977.
101. Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME Suite: tools for motif discovery and searching. Nucleic Acids Res. *37*, W202–W208. 10.1093/nar/gkp335.
102. Emenecker, R.J., Griffith, D., and Holehouse, A.S. (2021). Metapredict: a fast, accurate, and easy-to-use predictor of consensus disorder and structure. Biophys. J. *120*, 4312–4319. 10.1016/j.bpj.2021.08.039.
103. Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. *44*, W344–W350. 10.1093/nar/gkw408.
104. Bakan, A., Meireles, L.M., and Bahar, I. (2011). ProDy: Protein Dynamics Inferred from Theory and Experiments. Bioinformatics *27*, 1575–1577. 10.1093/bioinformatics/btr168.
105. Henikoff, S., Henikoff, J.G., Kaya-Okur, H.S., and Ahmad, K. (2020). Efficient chromatin accessibility mapping in situ by nucleosome-tethered tagmentation. eLife *9*, e63274. 10.7554/eLife.63274.
106. Meers, M.P., Tenenbaum, D., and Henikoff, S. (2019). Peak calling by Sparse Enrichment Analysis for CUT&RUN chromatin profiling. Epigenetics Chromatin *12*, 42. 10.1186/s13072-019-0287-4.
107. Sergé, A., Bertaux, N., Rigneault, H., and Marguet, D. (2008). Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes. Nat. Methods *5*, 687–694. 10.1038/nmeth.1233.
108. Hansen, A.S., Woringer, M., Grimm, J.B., Lavis, L.D., Tjian, R., and Darzacq, X. (2018). Robust model-based analysis of single-particle tracking experiments with Spot-On. eLife *7*, e33125. 10.7554/eLife.33125.
109. Saxton, M.J. (1997). Single-particle tracking: the distribution of diffusion coefficients. Biophys. J. *72*, 1744–1753. 10.1016/S0006-3495(97)78820-9.
110. Banani, S.F., Afeyan, L.K., Hawken, S.W., Henninger, J.E., Dall'Agnese, A., Clark, V.E., Platt, J.M., Oksuz, O., Hannett, N.M., Sagi, I., et al. (2022). Genetic variation associated with condensate dysregulation in disease. Dev. Cell. 10.1016/j.devcel.2022.06.010.
111. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona (2012). Fiji: an open-source platform for biological-image analysis. Nat Methods. 9, 676–682. 10.1038/nmeth.2019

Chapter 4:
RNA-Mediated Feedback Control of Transcriptional Condensates

Originally published in *Cell* 184, 207–225, January 7, (2021)

Jonathan E. Henninger,[1,13] Ozgur Oksuz,[1,13] Krishna Shrinivas,[2,6,8,13] Ido Sagi,[1] Gary LeRoy,[9,10] Ming M. Zheng,[4] J. Owen Andrews,[4] Alicia V. Zamudio,[1,3] Charalampos Lazaris,[1,11] Nancy M. Hannett,[1] Tong Ihn Lee,[1] Phillip A. Sharp,[3,7,14] Ibrahim I. Cissé,[4,14] Arup K. Chakraborty,[2,4,5,6,12,14,*] and Richard A. Young[1,3,14,15,*]

[1]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA
[2]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[3]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[4]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[5]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[6]Institute of Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[7]Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[8]NSF-Simons Center for Mathematical & Statistical Analysis of Biology, Harvard University, Cambridge, MA 02138, USA
[9]Howard Hughes Medical Institute, New York University School of Medicine, New York, NY 10016, USA
[10]Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, NY 10016, USA [11]Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[12]Ragon Institute of Massachusetts General Hospital, MIT and Harvard University, Cambridge, MA 02139, USA
[13]These authors contributed equally
[14]Senior author
[15]Lead Contact
*Correspondence: arupc@mit.edu (A.K.C.), young@wi.mit.edu (R.A.Y.)

https://doi.org/10.1016/j.cell.2020.11.030

Abstract

Regulation of biological processes typically incorporates mechanisms that initiate and terminate the process and, where understood, these mechanisms often involve feedback control. Regulation of transcription is a fundamental cellular process where the mechanisms involved in initiation have been studied extensively, but those involved in arresting the process are poorly understood. Modeling of the potential roles of RNA in transcriptional control suggested a non-equilibrium feedback control mechanism where low levels of RNA promote condensates formed by electrostatic interactions whereas relatively high levels promote dissolution of these condensates. Evidence from in vitro and in vivo experiments support a model where RNAs produced during early steps in transcription initiation stimulate condensate formation, whereas the burst of RNAs produced during elongation stimulate condensate dissolution. We propose that transcriptional regulation incorporates a feedback mechanism whereby transcribed RNAs initially stimulate but then ultimately arrest the process.

Main Text

Diverse biological processes have evolved feedback mechanisms to enable positive and negative regulation. Examples of biological processes that are known to incorporate feedback regulation include signal transduction (Brandman and Meyer, 2008), production of RNA splicing factors (Jangi and Sharp, 2014), circadian rhythms (Dunlap, 1999), red blood cell production (Ebert and Bunn, 1999), and response to DNA damage (Lahav et al., 2004). In transcription, some factors that regulate amino acid biosynthetic pathway genes can be regulated allosterically by intermediates produced by those pathways (Bergot et al., 1992; Bruhat et al., 2000; Sellick and Reece, 2003), but a general feedback mechanism has not been described. Evidence that feedback control is often mediated by the product of the process (Brandman and Meyer, 2008; Elowitz and Leibler, 2000; Gardner et al., 2000; Monod and Jacob, 1961; Umbarger, 1956) is one of the factors that led us to postulate that RNA may regulate transcription by a feedback mechanism.

Mammalian transcription produces diverse RNA species from regulatory elements and genes (Smith et al., 2019), and transcription of genes occurs in bursts of RNA synthesis (Chubb et al., 2006; Raj and van Oudenaarden, 2008; Raj et al., 2006). Transcription factors and coactivators recruit RNA polymerase II (RNA Pol II) to enhancer and promoter elements, where short (20–400 bp) RNAs are bidirectionally transcribed before RNA Pol II pauses (Adelman and Lis, 2012; Core and Adelman, 2019; Jin et al., 2017; Kim et al., 2010; Seila et al., 2008). These RNA species are short-lived and are reported to have various regulatory roles, although there is not yet a consensus regarding their functions (Andersson et al., 2014; Catarino and Stark, 2018; Core et al., 2014; Gardini and Shiekhattar, 2015; Henriques et al., 2018; Lai et al., 2013; Li et al., 2016; Mikhaylichenko et al., 2018; Nair et al., 2019; Pefanis et al., 2015; Rahnamoun et al., 2018; Schaukowitch et al., 2014; Scruggs et al., 2015; Sigova et al., 2015; Smith et al., 2019; Struhl, 2007). RNA Pol II pause release leads to processive elongation, which occurs in periodic bursts (~1–10 min in duration), where multiple molecules of RNA Pol II can be released from promoters within a short time frame and produce multiple molecules of mRNA (~1–100 molecules per burst) (Cisse et al., 2013; Fukaya et al., 2016; Larsson et al., 2019). How and whether the diverse RNA species produced during transcription—which differ in length, half-life, and number—affect or regulate transcription is currently unclear.

Recent studies have shown that transcriptional condensates can compartmentalize and concentrate large numbers of transcription factors, cofactors, and RNA Pol II at super-enhancers, clusters of enhancers that regulate genes with prominent roles in cell identity (Boija et al., 2018; Cho et al., 2018; Cramer, 2019; Hnisz et al., 2017; Sabari et al., 2018). The component enhancer elements of such genes promote transcriptional condensate formation by crowding transcription factors and Mediator at densities above sharply defined thresholds for condensate formation (Shrinivas et al., 2019). Transcriptional condensates are highly dynamic and can be observed in live cells to form and dissolve at timescales ranging from seconds to minutes (Cho et al., 2018). The periodic formation and dissolution of dynamic transcriptional condensates, coupled with evidence that different species and levels of RNAs are produced at different stages of transcription, led us to wonder whether transcriptional condensates are regulated by a non-equilibrium feedback mechanism mediated by its RNA product.

RNA molecules are components of and play regulatory roles in diverse biomolecular condensates. These include the nucleolus, nuclear speckles, paraspeckles, and stress granules (Fay and Anderson, 2018; Roden and Gladfelter, 2020; Sabari et al., 2020; Strom and Brangwynne, 2019). RNA has a high negative charge density because of its phosphate backbone, and the effective charge of a given RNA molecule is directly proportional to its length

(Boeynaems et al., 2019). Condensates are thought to be formed by an ensemble of low-affinity molecular interactions, including electrostatic interactions, and RNA can be a powerful regulator of condensates that are formed and maintained by electrostatic forces (Banani et al., 2017; Maharana et al., 2018; Peran and Mittag, 2020; Shin and Brangwynne, 2017). Indeed, RNA has been shown to enter and modify the properties of simple condensates formed by polyelectrolyte- rich molecules (Drobot et al., 2018; Frankel et al., 2016; Mountain and Keating, 2020). In a phenomenon called complex coacervation, a type of liquid-liquid phase separation mediated by electrostatic interactions between oppositely charged polyelectrolytes, low levels of RNA can enhance condensate formation, whereas high levels can cause their dissolution (Lin et al., 2019; Overbeek and Voorn, 1957; Sing, 2017; Srivastava and Tirrell, 2016). Condensate formation and subsequent dissolution with increasing RNA concentration is an example of reentrant phase behavior, which is driven by favorable opposite-charge interactions at low RNA concentrations (formation) and repulsive like-charge interactions at high RNA concentrations (dissolution) (Banerjee et al., 2017; Milin and Deniz, 2018). We wondered whether such a reentrant equilibrium phase behavior coupled to the non-equilibrium processes that occur during transcription could regulate transcriptional output.

By combining physics-based modeling and experimental analysis, we propose and test a model where the products of transcription initiation stimulate condensate formation and those of a burst of elongation stimulate condensate dissolution. We provide experimental evidence that physiological RNA levels can enhance or dissolve transcriptional condensates. These results show a mechanism by which the products of transcription regulate condensate behaviors and, thus, transcription and suggest that this non-equilibrium process provides negative feedback to dissolve the transcriptional condensates that support initiation and thereby arrest transcription.

Low Levels of RNA Enhance and High Levels Dissolve Mediator Condensates

To explore the potential role of RNA in regulating transcriptional condensates, we sought to estimate the number and effective charge of RNA and protein molecules in a typical transcriptional condensate at different stages of transcription. In early stages of transcription, low levels of small noncoding RNAs are produced by RNA Pol II at enhancers and promoter-proximal regions (Figure S1A; Adelman and Lis, 2012; Core and Adelman, 2019; Kim et al., 2010; Seila et al., 2008). During pause release, RNA Pol II produces longer genic RNAs during bursts of transcription elongation (Figure S1A; Adelman and Lis, 2012; Core and Adelman, 2019). These protein- and RNA-rich states can be thought of as mixtures of poly-electrolytes that may undergo complex coacervation (Figure 1A; Lin et al., 2019; Overbeek and Voorn, 1957; Sing, 2017; Srivastava and Tirrell, 2016). We reasoned that this phenomenon is likely to be relevant to transcriptional condensates because electrostatic interactions contribute to formation of these condensates, even in the absence of RNA (Boija et al., 2018; Sabari et al., 2020). Complex coacervate formation through phase separation is promoted when polyelectrolytes are present at concentrations where their net charges are approximately balanced. When the concentration of a poly-electrolyte, such as RNA, becomes sufficiently high, domination of repulsive like-charge interactions can suppress phase separation (Banerjee et al., 2017; Lin et al., 2019; Milin and Deniz, 2018; Muthukumar, 2016; Overbeek and Voorn, 1957; Zhang et al., 2018). Thus, at constant protein concentration, titrating RNA levels results in reentrant phase behavior, whereby low RNA levels promote and high RNA levels suppress condensate formation (Figure 1A; Banerjee et al., 2017; Milin and Deniz, 2018; Zhang et al., 2018). We wondered whether the reentrant phase behavior might apply to regulation of transcriptional condensates during transcription. Because the quantities of the diverse RNA species and proteins present in transcriptional condensates in populations of cells can be

estimated (Figure S1; STAR Methods), it is possible to conduct experimental tests to determine whether reentrant phase behavior occurs under physiologically relevant conditions of these molecules.

As an initial test of whether low levels of RNA stimulate transcriptional condensate formation while high levels of RNA favor condensate dissolution, we used an *in vitro* droplet assay (Figure 1B). Using components at physiologically relevant conditions, we investigated whether an enhancer RNA transcribed from the *Trim28* super-enhancer, which has been shown previously to form a transcriptional condensate in living cells (Boija et al., 2018; Guo et al., 2019), influences condensate formation by purified Mediator complex. Measurement of enhancer RNA levels in cells indicated that ~0.2 molecules of this enhancer RNA exist at steady state in murine embryonic stem cells (mESCs) (Figure S1F). Given that multiple loci in a super-enhancer are transcribed into enhancer RNAs, this roughly corresponds to ~100–1,000 nM of RNA in a typical Mediator condensate in cells (STAR Methods). These condensates typically contain Mediator at a concentration of around 1–20 mM (STAR Methods). The results showed that addition of 6–400 nM *Trim28* enhancer RNA to 200 nM purified Mediator complex had a dose-dependent effect on the size of Mediator/RNA droplets (Figures 1C–1E). Droplet sizes peaked at 100 nM RNA (Figure 1D), and the relative enrichment of RNA in the droplets, as measured by the ratio of average intensity inside versus outside the droplet (partition ratio), followed a similar trend (Fig- ure 1E). Similar results were obtained using an enhancer RNA transcribed from the *Pou5f1* super-enhancer (Figures 1F–1H). Thus, within the range of physiological levels observed in cells, low levels of RNA can enhance condensate formation, and high levels of RNA can reduce condensate formation by Mediator *in vitro*.

Charge Balance Mediates Regulation of MED1-IDR Condensates by RNA

We next sought to determine whether the reentrant phase behavior of mixtures of RNA and transcriptional proteins is predominantly regulated by charge balance considerations, with other types of RNA-protein interactions playing a less significant role. We performed *in vitro* droplet assays (Figure 2A) using the MED1 C-terminal intrinsically disordered region (MED1-IDR), which has proven to be a useful surrogate for the multisubunit Mediator complex because it is not possible to purify sufficient amounts of this complex to test all parameters of interest (Boija et al., 2018; Guo et al., 2019; Klein et al., 2020; Li et al., 2020; Sabari et al., 2018; Shrinivas et al., 2019; Zamudio et al., 2019). Fusion of GFP to MED1-IDR allows quantification by fluorescence of a single species whose effective charge can be calculated to determine the charge ratio between protein and RNA. Addition of increasing levels of RNA to a constant protein concentration should have predictable effects on partitioning of either component according to its charge ratio (Figure 2B). Non-coding and coding RNAs produced from three different super-enhancer loci and their associated genes (*Trim28*, *Pou5f1*, and *Nanog*; Figure S1) were selected for this analysis based on prior studies of nascent RNA sequencing data in mESCs (Boija et al., 2018; Guo et al., 2019; Sabari et al., 2018; Sigova et al., 2015; Whyte et al., 2013). Addition of 6–400 nM of each of these RNAs to 1,000 nM MED1-IDR (protein:RNA ratios = 167:2.5) stimulated formation of MED1-IDR condensates at low RNA concentrations and dissolved MED1-IDR condensates at higher RNA concentrations (Figures 2C, 2D, S2A, and S2B). BRD4 is another key component of transcriptional condensates, and BRD4-IDR protein exhibits condensate behaviors very similar to those of MED1-IDR (Sabari et al., 2018); the effects of increasing RNA levels on formation and dissolution of BRD4- IDR condensates were very similar to those observed for MED1-IDR (Figures S2C and S2D). RNA did not stimulate formation of droplets with GFP alone or OCT4-GFP, both of which have a net negative charge (Figure S2E). Condensates exhibited internal dynamic reorganization (Figure S2F) with

apparent diffusion coefficients 3–5 ± 0.8 3 102 mm2/s (STAR Methods), consistent with liquid-like behavior (Nott et al., 2015; Sabari et al., 2018; Taylor et al., 2019). The incomplete recovery after photobleaching seen here has been observed previously with other condensates (Nott et al., 2015; Sabari et al., 2018; Taylor et al., 2019) and could result from a small portion of the droplet material being relatively immobile, rate-limiting material in the dilute phase or droplet aging (Taylor et al., 2019). These results show that diverse RNAs are capable of stimulating MED1-IDR condensate formation when present at relatively low levels and dissolving MED1-IDR condensates at high levels.

We sought to further test whether charge balance is the predominant phenomenon underlying the RNA-mediated effects on MED1-IDR condensates (STAR Methods). If so, then MED1-IDR/RNA condensate formation should be enhanced when the protein and RNA polymers are balanced in charge, and they should be sensitive to disruption of this balance. We quantified the relative charge of RNA and MED1-IDR and computed the correlation with the partition ratio of MED1-IDR (STAR Methods). RNA-mediated effects on MED1-IDR condensates correlated with their charge balance, as observed by the concordance of higher partition ratios near charge balance and lower partition ratios away from this balance (Figure 2D). We would expect an RNA length-dependent shift in the RNA level required for peak MED1-IDR partitioning when RNAs of different length are introduced into the droplet assay in equal numbers. This expectation, that a higher concentration of shorter RNAs is needed to disrupt condensate formation, was observed (Figures S3A and S3B). Another prediction from charge balance considerations is that these interactions should be largely independent of RNA sequence, so antisense versions of any one of the RNA species should exhibit the same quantitative effects as the sense strand, and this was also observed (Figures S3B and S3C). Consistent with charge balance considerations, MED1-IDR condensates formed with RNA were sensitive to increasing monovalent salt, which screens charged interactions (Figure S3D). The expectations from charge balance considerations also held when MED1-IDR and RNA concentrations were varied (Figures S4A–S4D) and when alternative polyanions (heparin and single-stranded DNA [ssDNA]) were employed (Figures S4E and S4F). RNA did not stimulate condensate formation by a MED1-IDR mutant lacking positively charged residues (MED1-IDR R/H/K > A) (Figures S4G and S4H). Although the experiments described above show a strong correlation between charge balance and partition ratios, the lack of complete correlation suggests that other features of RNA and MED1-IDR, such as RNA secondary structure (Roden and Gladfelter, 2020) or non-electrostatic interactions (Sabari et al., 2018), may influence the observed phase behavior. Nonetheless, these results further support the concept that RNA-mediated effects on equilibrium behavior of MED1-IDR condensates are predominantly regulated by electrostatic effects.

RNA-Mediated Effects on Condensates in Reconstituted In Vitro Transcription Assays

We sought to investigate the functional consequence of the RNA-mediated reentrant phase behavior on transcription. RNA Pol II-dependent transcription can be reconstituted *in vitro* with purified components (Roeder, 2019), so we investigated whether droplets containing transcriptional components are formed in these assays and whether conditions that alter droplet levels similarly alter transcriptional output. We used a classic reconstituted mammalian transcription system with purified components, including RNA Pol II, general transcription factors, Mediator, and a transcriptional activator (Gal4), where addition of nucleotides permits transcription of a linear DNA template (Figure 3A). We observed that component mixtures and buffer conditions that are optimal for transcriptional output (Carey et al., 2009; Flores et al., 1992; LeRoy et al., 2008; Orphanides et al., 1998) produced droplets containing the DNA template (Figure 3B). Quantification of the newly synthesized RNA in this system showed that

3.5 (±0.5) pM RNA was produced in the transcription reaction (STAR Methods). We were unable to demonstrate that RNA synthesis actually occurs in the droplets because we cannot eliminate the possibility that synthesis occurs in the bulk phase and the product subsequently partitions into the droplet, but the observation that protein and template DNA concentrate in droplets under conditions optimal for transcription (Figure 3B) and evidence that diverse condensate-altering treatments have similar effects on transcription, described below, are consistent with the notion that transcription occurs within condensates in this reconstituted system.

We reasoned that if transcription and droplet formation are mutually dependent in the reconstituted system, then treatments that alter transcription should similarly affect condensate formation and vice versa. Addition to the reaction of various chemicals that are known to inhibit transcription (elevated concentrations of nucleoside triphosphates (NTPs), NaCl, or heparin; Carey et al., 2009; Reinberg and Roeder, 1987) caused reductions in droplet area, DNA partitioning, and transcription (Figure S5). Spermine, a positively charged polyamine, enhances droplet formation when it contributes to charge balance in coacervate models (Aumiller et al., 2016). Addition of spermine at concentrations predicted to balance charge in the *in vitro* reactions simultaneously increased droplet area, partitioning of template DNA, and levels of RNA synthesis (Figures 3C–3F; Table S2; Blair, 1985; Moruzzi et al., 1975). These correlations suggest that optimal droplet formation and transcription are co-dependent.

An expectation of the RNA feedback model is that droplets in the reconstituted system might ultimately produce enough RNA to cause a reduction in droplet size and transcriptional output. However, the low concentrations of RNA produced in these systems (3.5 ± 0.5 pM; STAR Methods) are insufficient to dissolve the droplets. For this reason, we tested whether purified RNA, added to the reaction, would similarly affect droplets and transcription. Indeed, addition of exogenous RNA reduced the number and size of the droplets (Figures 3G and 3H) and reduced template-derived RNA synthesis, as measured by qRT-PCR (Figure 3I). Although these results do not rule out additional ways in which RNA may affect transcription (Pai et al., 2014), they are consistent with the expected behavior of transcriptional condensates if RNA contributes to negative feedback control.

A Model of RNA-Mediated Non-equilibrium Feedback Control of Transcriptional Condensates

The *in vitro* experiments, which provide evidence that key transcriptional proteins and RNA exhibit electrostatics-driven, RNA-protein ratio-dependent reentrant phase transition, were performed under equilibrium conditions (Figures 1 and 2). However, *in vivo*, RNA is synthesized and degraded at specific genomic loci by dynamic, ATP-dependent, non-equilibrium processes (Azofeifa et al., 2018; Li et al., 2016; Pefanis et al., 2015). To investigate how non-equilibrium processes underlying transcription may regulate transcriptional condensates, we built a physics-based model. The model consists of two inter-linked parts: (1) a free-energy function (Figure 4A) that depends on the concentrations of transcriptional proteins and RNA and recapitulates the equilibrium reentrant phase behavior of RNA-protein mixtures (Figures 1 and 2), and (2) a mathematical framework to study spatiotemporal evolution of condensates subject to non-equilibrium dynamical processes of RNA synthesis, degradation, and diffusion (Figure 4B).

The goal of the model described below is to explore non-equilibrium regulation of transcription by RNA output and obtain insights into the pertinent mechanistic principles. Quantitative descriptions of RNA-protein phase behavior even *in vitro* (Adhikari et al., 2018; Delaney and Fredrickson, 2017) and direct measurements of the dynamic parameters underlying transcription are largely unavailable (Rodriguez and Larson, 2020). Therefore, we sought to

develop a phenomenological model of non-equilibrium regulation of transcription by RNA output and use it to predict the qualitative effects of perturbing model parameters. Our experimental approaches allow such perturbations to be realized and can test whether the predicted effects are accurate. By coupling the predictions of effects of perturbing model parameters with experimental tests, we aimed to obtain mechanistic insights into how RNA synthesis may dynamically regulate transcription itself.

We first developed a free-energy function to recapitulate the experimentally observed reentrant phase behavior of RNA-protein mixtures (Figure 4A). The free-energy function depends on the concentrations of transcriptional proteins ϕ_p and RNA ϕ_r, which vary in space and time. For simplicity, all transcriptional proteins are combined into one pseudo-species. As noted above, our goal is not quantitative recapitulation of known experimental data but to obtain mechanistic insights into RNA-mediated non-equilibrium regulation of transcription that could be tested experimentally. Therefore, we first sought to develop a free-energy function that qualitatively recapitulates the observed reentrant phase behavior of RNA/protein mixtures. Following a long tradition in the physics of phase transitions, we employed a general Landau approach (Kardar, 2007; Landau, 1937) and expanded the free energy as a function of RNA and protein concentrations. We include terms to describe repulsive RNA-RNA interactions, favorable interactions among the transcriptional proteins that drive the condensate formation of transcriptional proteins in the absence of RNA (Figure 4C, equation 1, green) as well as a surface tension term important for describing condensate formation (Figure 4C, equation 1, blue; STAR Methods). The free-energy function also includes protein-RNA interactions that are described by a concentration-dependent interaction term, which is expanded in the standard Landau fashion (Figure 4C, equation 1, red; Kardar, 2007). Magnitudes of the coefficients of the various terms in the expansion account for the effective strength of RNA-protein interactions (STAR Methods), which implicitly include solvent effects. Although symmetry arguments do not preclude any specific terms in this expansion, analysis of the pertinent Jacobian matrix shows that the choice of $\chi>0$, $c > 0$, and $a,b<<1$ ensures a reentrant phase transition (schematic in Figures 4B and S6A; STAR Methods) with a minimal number of higher-order terms. Results using the Landau model (Figure 4C, equation 1) are recapitulated using a different method for obtaining the free energy (Flory-Huggins) to highlight the generality of our Landau approach (Figures S6A and S6B; STAR Methods). Given the universality of its application, easily characterizable phase behavior, and numerical ease of investigation (e.g., ~50 times faster than the Flory-Huggins to study coupled dynamics), we employed the Landau free energy in the rest of this work to study how the dynamics of transcriptional condensates are regulated by transcription.

We next developed a mathematical framework to study the temporal evolution of transcriptional condensates as transcription ensues. Most transcriptional proteins turn over with a half-life of several hours (Cambridge et al., 2011; Chen et al., 2016), which is longer than timescales of transcription-associated events, which range from seconds to minutes (Chen and Larson, 2016; Fukaya et al., 2016; Rodriguez and Larson, 2020). Hence, the overall amount of protein is conserved in the timescales of interest. Thus, the dynamics of the protein concentration ϕ_p are represented by standard model B dynamics (Figure 4C, equation 2; Hohenberg and Halperin, 1977). Under model B dynamics, gradients in the protein chemical potential, which depend on the spatial distribution of protein and RNA concentrations, drive diffusive protein fluxes, which, in turn, drive the spatiotemporal evolution of ϕ_p. Because RNA concentrations vary over transcription-associated timescales, the dynamics of f*r* are explicitly governed by a reaction-diffusion equation. The key features (schematic in Figure 4B) are that RNA diffuses with mobility *Mrna* and is synthesized and degraded with specific reaction rates: *kp* and *kd*, respectively. Because the RNA dynamics are far from equilibrium and the free-energy function noted above

depends on protein and RNA concentrations, the coupled temporal evolution of transcriptional proteins and RNA (Figure 4C, equations 1 and 2) cannot be obtained from near-equilibrium considerations of simply going downhill in free energy with time. We employ this mathematical framework to study the non-equilibrium regulation of transcriptional condensates.
We first sought to determine whether this model is consistent with previous studies (Cho et al., 2016, 2018). These studies have shown that transcriptional condensates at different genomic loci recruit a varying number of transcriptional proteins, which, in turn, correlates with condensate lifetimes. To explore this phenomenon, we numerically simulated equation 2 (Figure 4C) on 2D and 3D grids (STAR Methods). Locus-dependent recruitment of the transcriptional machinery can be mimicked in our model by varying the total transcriptional protein amount P_0 with all other parameters fixed because our simulation volume represents a local micro-environment (Figure 4A). Our simulations predict that loci that can recruit more transcriptional proteins (higher P_0) form relatively stable condensates, whereas condensates that recruit fewer proteins dissolve after a characteristic lifetime (Figure 4D). The model predictions for transcriptional condensate dynamics are qualitatively consistent with published data (Cho et al., 2016) and suggest that features encoded at genomic loci contribute to transcriptional condensate dynamics.

We next investigated how the sizes and lifetimes of transcriptional condensates change as a function of the effective rate of RNA synthesis k_p while keeping all other parameters fixed. In these simulations, the size of condensates initially increases and subsequently decreases with increasing effective rates of RNA synthesis (Figure 4E). Above a threshold rate of RNA synthesis, condensates dissolve (Figure 4E). The underlying reason for this result is the reentrant phase behavior of mixtures of transcriptional molecules and RNA (Figures 1 and 2). We also find that condensates with higher transcriptional activity dissolve faster, as measured by condensate lifetimes (Figure 4F). Condensate lifetimes do not vary over a range of RNA transcription rates that reflect RNA-transcriptional protein ratios that roughly correspond to the charge balance conditions (Figure 4F). The same qualitative results are recapitulated in 3D simulations (Figure S6C) as well as simulations employing the Flory-Huggins free energy (Figure S6D) and further reinforced by partition ratios computed from simulations (Figure S6E). Further, we carried out simulations that accounted for phase-dependent changes in diffusion of RNA; i.e., RNA diffusion was hindered in the dense phase because of crowding (Figure S6F). Predictions of the condensate size and lifetimes exhibited qualitative trends similar to simulations without this phase-dependent diffusion (Figure S6F; Figures 4E and 4F). Overall, our results suggest a model where low effective rates of RNA synthesis (or low transcription activity) stabilize transcriptional condensates, whereas higher rates promote condensate dissolution.

We then investigated the extent to which non-equilibrium effects underlying transcription regulate transcriptional condensate dynamics. RNA synthesis, degradation, and diffusion influence the spatial distribution of RNA, which, in turn, may feedback on transcriptional condensates. To explore this, we varied the diffusivity of RNA and the effective rates of RNA synthesis and degradation while holding the ratio of synthesis and degradation rates constant. The latter constraint ensures that the over- all RNA concentration is constant in the condensate as other parameters are varied; thus, any effect on condensate dynamics arises from purely non-equilibrium effects. Varying the parameters that control RNA synthesis/degradation rates and diffusion changes the relative timescales of these processes (tr and td, respectively) (STAR Methods), which in turn, influences the spatial distribution of RNA in the condensate. If diffusion is slower than synthesis/degradation ($tr < td$), then RNA will accumulate near transcription sites, leading to a higher local RNA concentration in the condensate. Conversely, if diffusion is faster than synthesis/degradation ($tr > td$), then RNA will diffuse away from transcription sites, leading

to a lower uniform RNA concentration in the condensate. The spatial distribution of RNA will affect condensates according to local charge balance. To study how varying spatial distributions of RNA affect transcriptional condensates, we simulated conditions where the overall RNA concentration was fixed close to the charge balance condition, promoting condensate formation at equilibrium. In these simulations, condensates that are stable when synthesis/degradation is slower than diffusion *tr* > *td* dissolve when RNA synthesis/degradation is faster than diffusion *tr* < *td* Þ (Figure 4G). When *tr* > *td*, RNA concentration is relatively uniform and low throughout the condensate, equilibrium effects dominate. Conversely, when *tr* < *td*; RNA is distributed non-uniformly with high local concentrations in the condensate, non-equilibrium effects dominate to result in condensate dissolution (Figure 4G). In the latter case, the localized high RNA concentrations exceed the charge balance condition because of non-equilibrium effects. Approximate estimates for the rates of RNA synthesis, degradation, and diffusion under physiological conditions (*td/tr*≈2-100, STAR Methods) suggest that transcriptional *tr* condensate dynamics are likely driven off equilibrium. We sought to synthesize our results so far to explore the effect of non-equilibrium dynamics on regulating transcriptional condensates across transcription initiation and productive elongation. Simulations were started at a relatively low effective rate of RNA synthesis, mimicking initiation, followed by an increase to a relatively high effective rate of RNA synthesis, mimicking productive elongation. The simulations predict that low effective rates of RNA synthesis enhance condensate formation, and these condensates subsequently dissolve upon ensuing higher effective rates of RNA synthesis (Figure 4H). Consistent with these simulations, Mediator condensates tend to be depleted in areas of high, RNA Pol II-driven nascent transcription (Figure S7A–S7C). These results suggest that non-equilibrium processes underlying RNA synthesis can potentially regulate formation and dissolution of transcriptional condensates.

Inhibition of RNA Elongation Leads to Enhanced Condensate Size and Lifetime in Cells

Transcriptional condensates in cells are highly dynamic, forming and dissolving at timescales ranging from seconds to minutes (Cho et al., 2018). We previously showed that condensate formation is associated with transcription activation and initiation (Cho et al., 2018). When transcriptional condensates are formed, the RNA-mediated condensate dissolution model predicts that inhibition of elongation should increase the size and lifetime of transcriptional condensates (Figure 5A). We used the physics-based model (Figure 4) to simulate the effects of elongation inhibition on transcriptional condensates and performed experiments to test the predictions from these simulations in cells (Figures 5B–5I). To ac- count for the locus-dependent ability to recruit the transcriptional machinery and RNA Pol II, we performed these simulations at a range of total protein concentrations (as in Figures 4D and 4E) but for conditions where the effective rate of RNA synthesis (k_p) was high (corresponding to elongation) and low (corresponding to inhibited elongation). The results of the simulations predict that a reduced effective rate of RNA synthesis should increase the size and lifetime of transcriptional condensates across a range of total protein concentrations (Figure 5B and 5F).

To experimentally test these predictions from the simulations, mESCs engineered with an endogenous, GFP-tagged subunit of Mediator (Med1-GFP) (Sabari et al., 2018) were treated for 30 min with Actinomycin-D (ActD) or 5,6-dichloro- benzimidazole riboside (DRB) (Figure 5C), which disrupt transcription elongation through DNA intercalation and inhibition of CDK9-mediated RNA Pol II pause release, respectively (Singh and Padgett, 2009; Sobell, 1985; Steurer et al., 2018). Consistent with the model predictions, after inhibition of elongation, Med1-GFP condensates increased in volume by ~2-fold as measured by 3D super-resolution microscopy (Figures 5D and 5E). Condensate lifetime could not be assessed in these cells

because of the long duration of image acquisition and consequent photobleaching, so we turned to time-correlated photo-activation localization microscopy (tcPALM) super-resolution microscopy in mESCs with an endogenous Med19-Halo tag (Cho et al., 2018; Cisse et al., 2013) to investigate the effects of elongation inhibition on condensate lifetime (Figure 5G). Cells were treated for 30 min with DRB to disrupt transcription elongation, and the lifetime of Med19 condensates was quantified. When transcription elongation was inhibited by DRB treatment, Med19 condensates exhibited significantly longer lifetimes than mock-treated cells (Figures 5H and 5I), and when DRB-treated cells were washed with fresh medium, the lifetimes of the Med19 condensates recovered to those of the mock-treated condition (Figures 5H and 5I). The *in silico* and experimental results show that suppression of elongation in cells leads to increased condensate size and lifetime, consistent with a model where a burst of RNA synthesis can promote dissolution of transcriptional condensates in cells.

Increasing the Levels of Local RNA Synthesis Reduces Condensate Formation and Transcription in Cells

The RNA-mediated feedback model suggests that modifying the concentration or size of RNA molecules should have a predict- able effect on transcriptional output. We developed complementary experimental and simulation approaches (Figure 6) where the levels of putative ‘‘feedback RNAs’’ could be increased artificially. We first used the physics-based model (Figure 4) to simulate the effect of increasing effective rates of RNA synthesis as well as varying lengths for the synthesized RNA on condensates (STAR Methods). The simulations predicted that increases in the production rate of shorter RNAs initially enhance and subsequently suppress transcriptional condensate size, whereas increases in the production rate of longer RNAs lead to reduced condensate size with increasing synthesis rates (Figure 6E).

To test this prediction, we investigated the effect of artificially increasing the levels of feedback RNAs on transcription of an adjacent luciferase reporter gene in cells (Figures 6A and 6B; Kirk et al., 2018). DNA molecules specifying RNAs of a range of sizes were cloned into this system to allow doxycycline (Dox)-inducible expression of these RNAs, and mESC lines were generated with clones of integrated constructs. Feedback RNAs were observed at loci of Mediator puncta under low-Dox stimulation, suggesting that these actively transcribed genes are associated with transcriptional condensates (Figures 6C and 6D). Elevated expression of feedback RNAs under higher-Dox stimulation reduced their co- localization with Mediator puncta, consistent with the model of RNA-mediated feedback on condensates (Figures 6C, 6D, and S7D). To study the effect of local RNA levels on transcription, additional cell lines harboring diverse feedback RNAs were then treated with increasing doses of Dox to induce feedback RNA expression (Figure 6F), and reporter expression was measured by luminescence (Figure 6G). The results were consistent with model predictions (Figure 6E); increases in the levels of short feedback RNAs initially enhanced reporter expression and then suppressed this, whereas progressive increases in the levels of the longer feedback RNAs reduced reporter expression more strongly (Figure 6G). We confirmed that changes in reporter expression arise from *cis* RNA-mediated effects by modifying the constructs, controlling for the global effects of Dox, and perturbing local RNA concentration (Figures S7E–S7K). These results support a role for RNA-mediated feedback in the control of transcriptional condensates.

DISCUSSION

The results described here indicate that transcription is a non- equilibrium process that provides dynamic feedback through its RNA product. The results support a model where RNA pro- vides positive and negative feedback on transcription via regulation of electrostatic interactions in transcriptional condensates. Transcriptional condensates, whose formation involves crowding of transcription factors by enhancer DNA (Shrinivas et al., 2019) and electrostatic and other interactions between the IDRs of transcription factors and coactivators (Boija et al., 2018; Sabari et al., 2018), engage RNA to both promote and dissolve the condensates. In this RNA feedback model, low levels of short RNAs produced during transcription initiation pro- mote formation of transcriptional condensates, whereas high levels of the longer RNAs produced during elongation can cause condensate dissolution (Figure 7).

An RNA-mediated feedback model for transcriptional regulation provides a potential explanation for the roles of enhancer and promoter-associated RNAs, which are evolutionarily conserved features of eukaryotes. These low-abundance short RNAs, transcribed bidirectionally from enhancers and promoters, have been reported to affect transcription from their associated genes through diverse postulated mechanisms. The diversity of sequences present in these short RNA species has made it difficult to postulate a common molecular mechanism for their effects on transcription. In this context, a model for RNA-mediated feedback regulation of condensates is attractive for several reasons. RNA molecules are known components of other biomolecular condensates, including the nucleolus, nu- clear speckles, paraspeckles, and stress granules, where they are known to play regulatory roles (Fay and Anderson, 2018; Roden and Gladfelter, 2020). RNA is a powerful regulator of condensates that are formed by electrostatic forces because it has a high negative charge density due to its phosphate backbone (Drobot et al., 2018; Frankel et al., 2016), explaining why the effects of diverse RNAs on transcriptional condensates are sequence independent. The functions of most noncoding RNAs remain a mystery, and this model suggests a mechanism by which some of these might participate in tuning local gene expression.

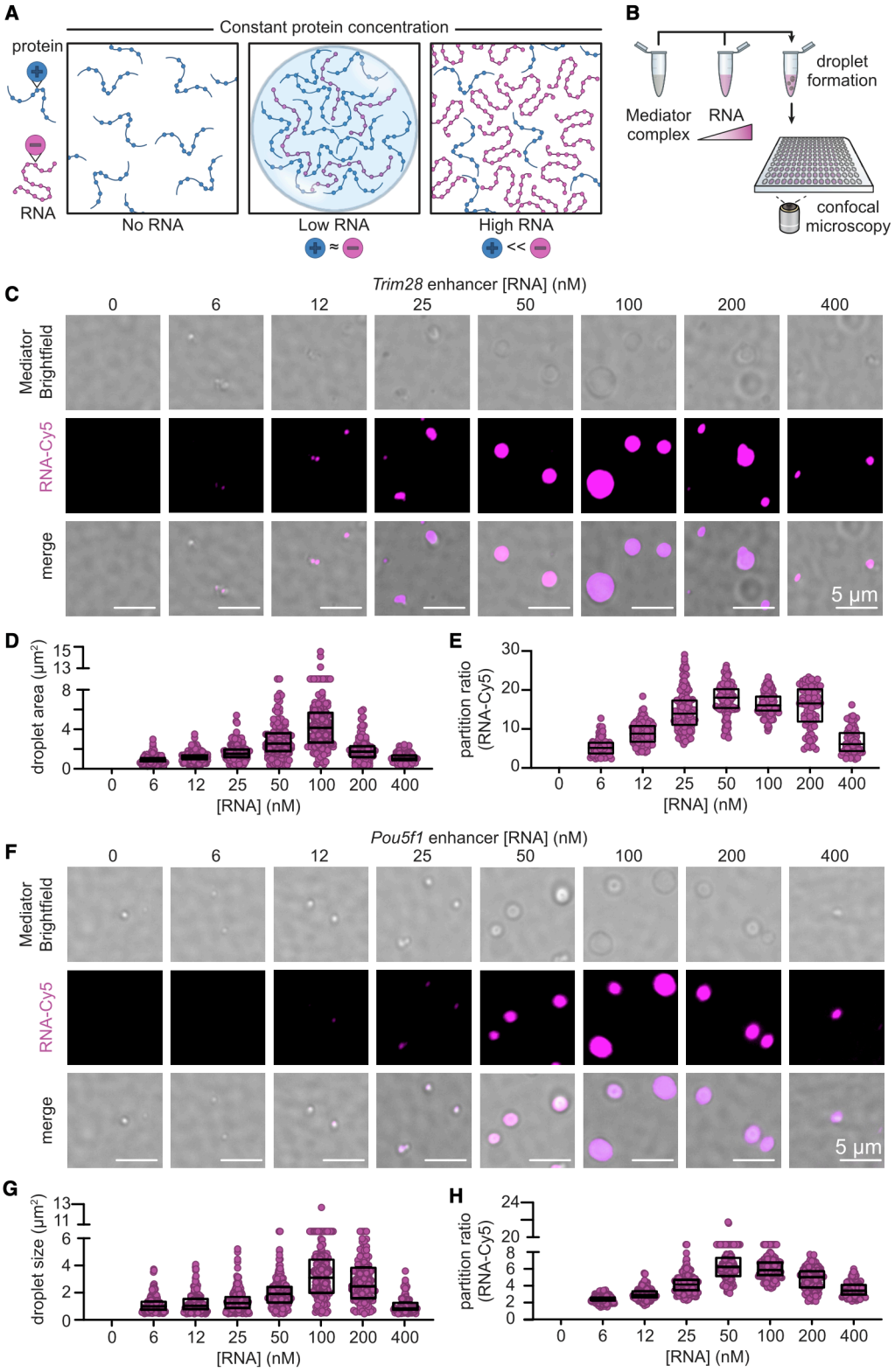
Recent studies indicate that transcription occurs in periodic bursts (~1–10 min in duration), where multiple molecules of RNA Pol II can be released from promoters within a short time frame and produce multiple molecules of mRNA (~1–100 molecules per burst). Multiple models explain such periodic bursts through stochastic gene activation events (Chen and Larson, 2016; Larsson et al., 2019; Raj et al., 2006; Rodriguez and Lar- son, 2020; Suter et al., 2011; Tunnacliffe and Chubb, 2020) but are often agnostic to the underlying mechanism or attribute these to rate-limiting transcription factor binding events. We suggest that a rapid and spatially localized change in charge balance, due to increased RNA synthesis at pause release of active RNA Pol II, may contribute to dissolution of transcriptional condensates and thus dynamic loss of the pool of transcriptional apparatus in those condensates. This would provide negative feedback to arrest transcription and a mechanism to account for the dynamic bursty behavior observed for transcription.

ACKNOWLEDGMENTS

We are grateful to Dora Tang and Ankur Jain for discussions on complex coacervates and their regulation by RNA. We also thank Mehdan Kardar and Young lab members for discussions; D. Reinberg, J.M. Calabrese, R. Jaenisch, and I. Cheeseman for reagents; W. Salmon of the W. M. Keck Microscopy Facility; E. Diel of the Harvard Center for Biomedical Imaging. This work was supported by NIH grants GM123511 (R.A.Y.), CA155258 (R.A.Y.), 1F32CA254216-01 (J.E.H.), GM134734 (I.I.C.), NSF grant PHY-1743900 (A.K.C., R.A.Y., and P.A.S.), and by the Gruss-Lipper Postdoctoral Fellowship and the Rothschild Postdoctoral Fellowship (I.S.).

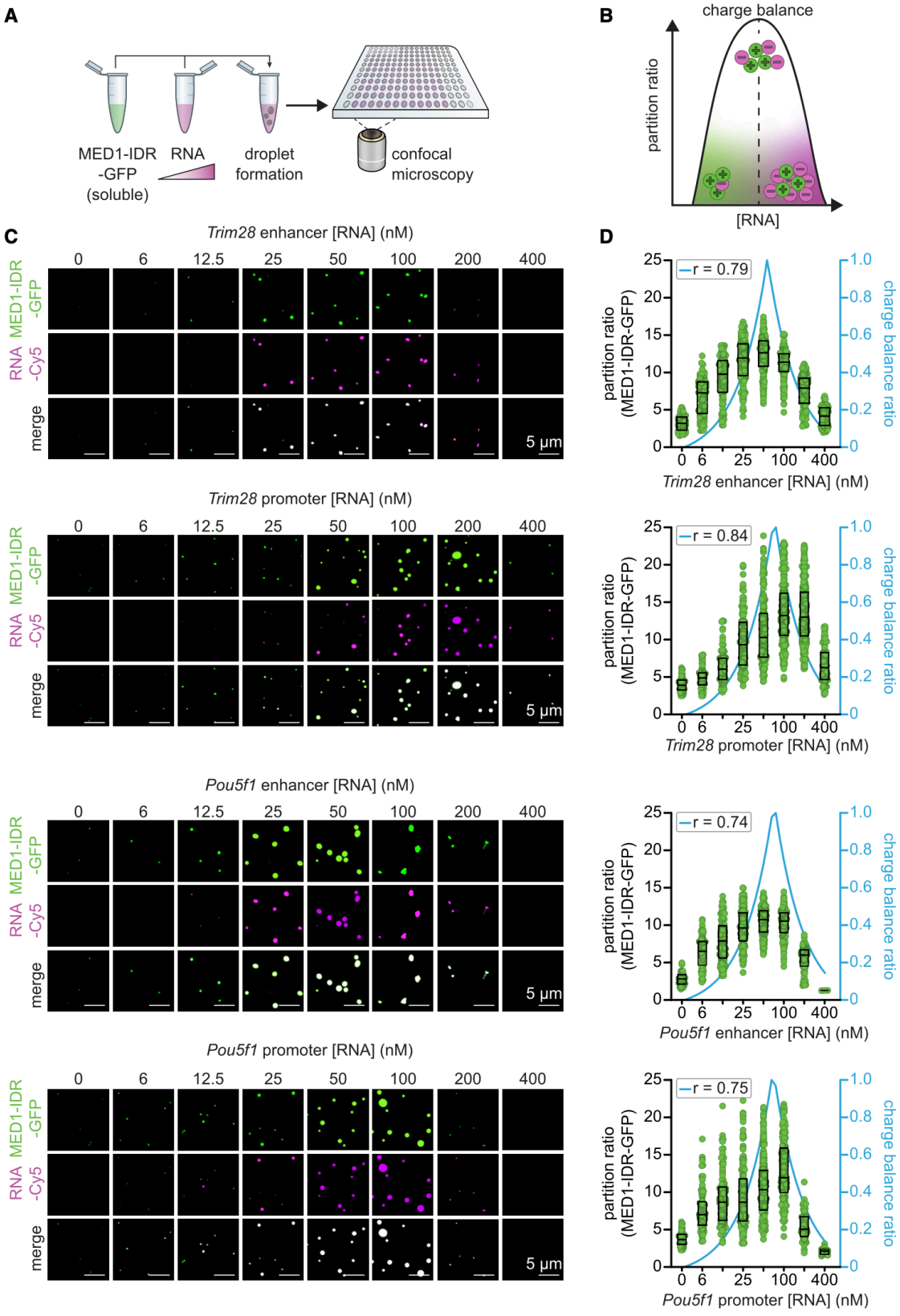
Figures and Tables

Figure 1. Low Levels of RNA Enhance and High Levels Dissolve Mediator Condensates



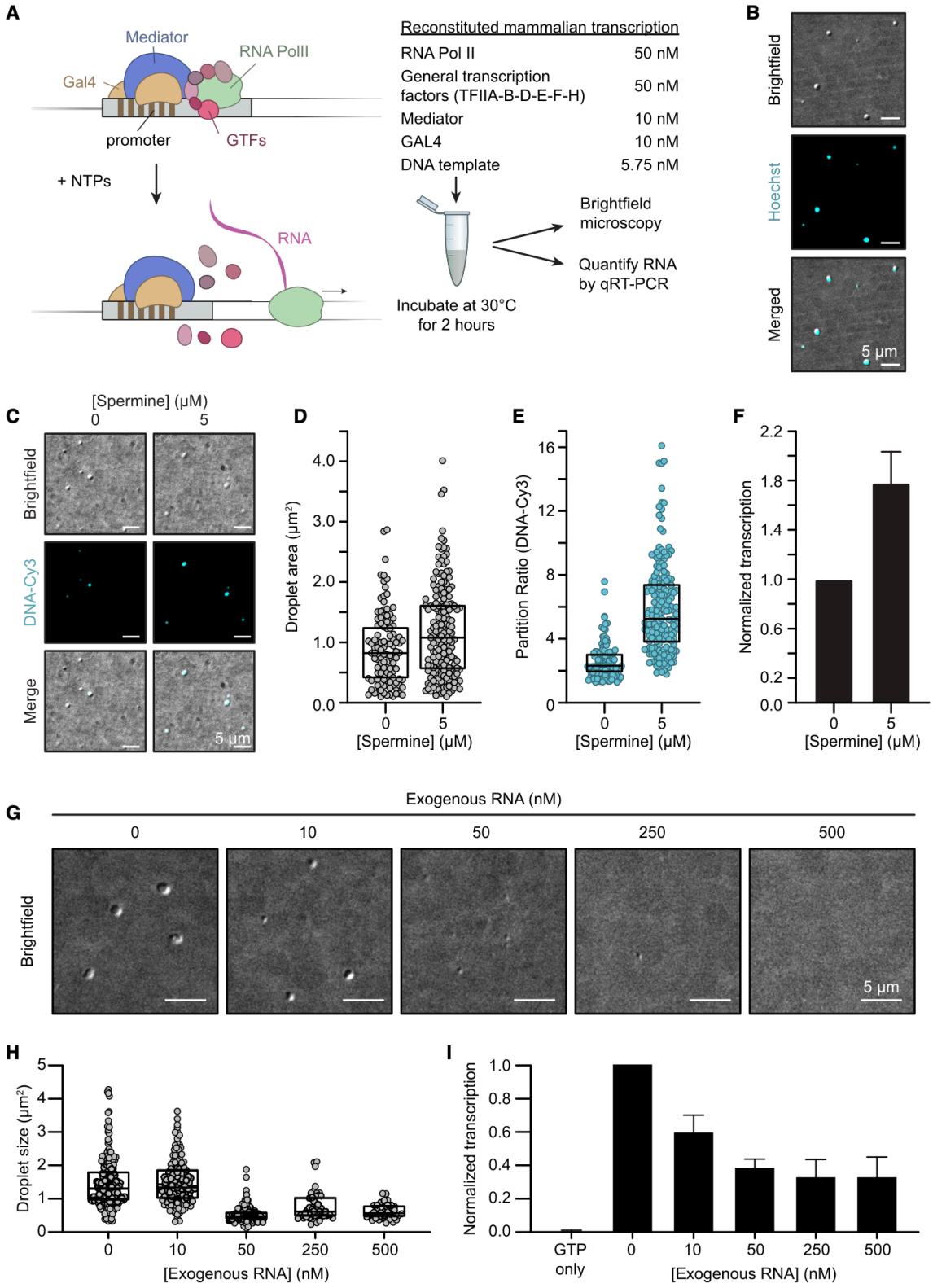
(A) Diagram of reentrant phase transition in response to increasing concentrations of RNA over constant protein concentration. The condensed fraction of protein peaks at the RNA concentration at which the charges between protein and RNA are balanced, whereas alteration of this charge balance in either direction decreases the condensed fraction.
(B) Experimental design for the *in vitro* droplet formation assay. Whole Mediator complex is mixed with increasing concentrations of RNA under physiologically relevant buffer conditions, and droplets are imaged using confocal microscopy.
(C) Representative images of droplets formed by the unlabeled whole Mediator complex (200 nM) and Cy5-labeled *Pou5f1* enhancer RNA at increasing concentrations (0–400 nM). Bright-field images of the Mediator complex were divided by a median-filtered image (pixels = 15) here and in the subsequent panels (scale bars = 5 mm).
(D) Droplet sizes in (C).
(E) Partition ratios of Cy5-labeled RNA within the droplets in (C).
(F) Representative images of droplets formed by the unlabeled whole Mediator complex (200 nM) and Cy5-labeled *Trim28* enhancer RNA at increasing concentrations (0–400 nM) (scale bars = 5 mm).
(G) Droplet sizes in (F).
(H) Partition ratios of Cy5-labeled RNA within the droplets in (F).
See also Figure S1.

Figure 2. Charge Balance Mediates Regulation of MED1-IDR Condensates by RNA



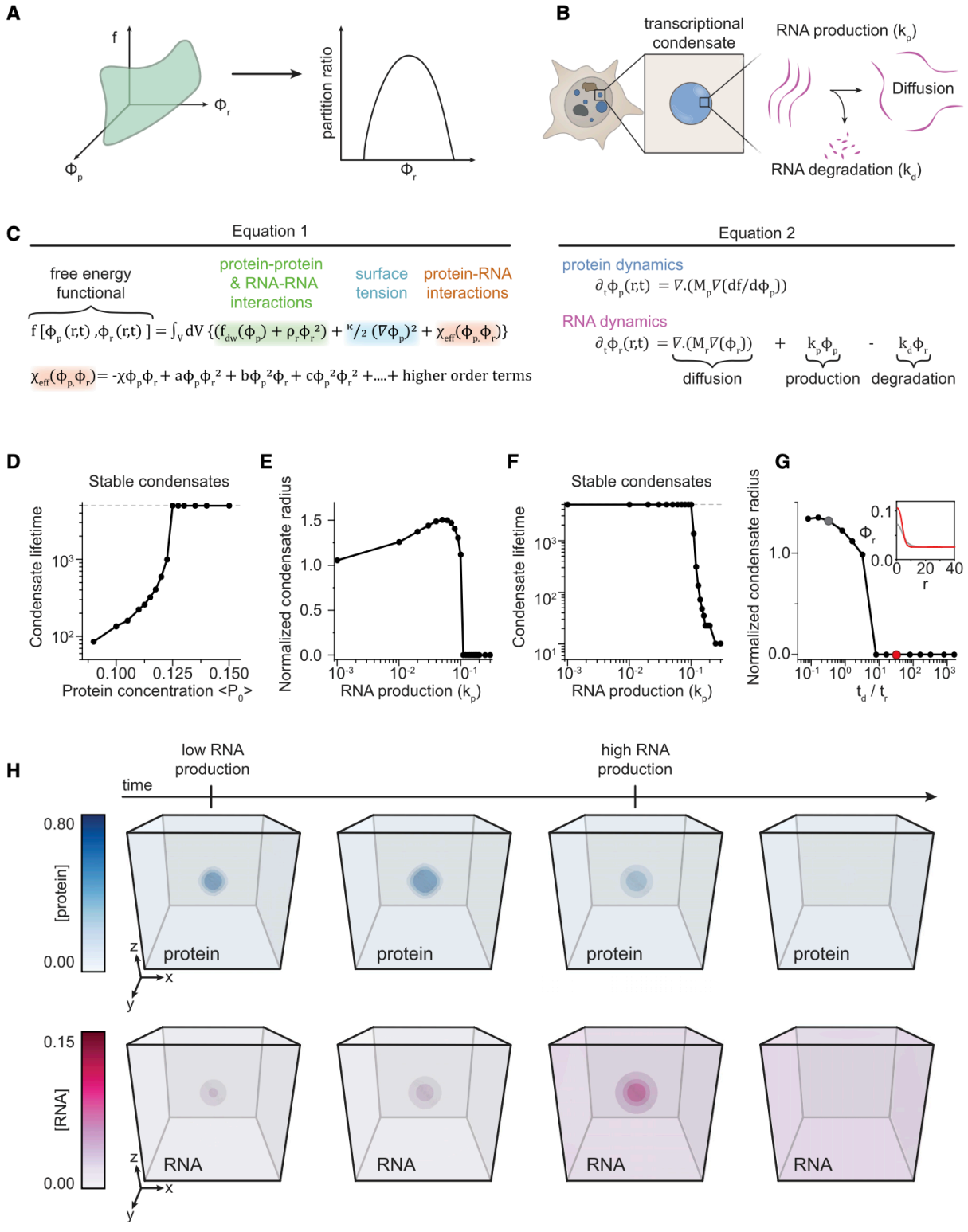
(A) Experimental design for the *in vitro* droplet formation assay. Soluble MED1-IDR-GFP is mixed with increasing concentrations of RNA under physiologically relevant buffer conditions, and droplets are imaged using confocal microscopy.
(B) Scheme of the charge balance ratio between constant protein concentration and increasing RNA concentrations.
(C) Representative images of droplets formed by increasing concentrations (0–400 nM) of the indicated RNAs mixed with 1 mM of MED1-IDR-GFP (scale bars = 5 mm). (D) Partition ratios of MED1-IDR-GFP within the droplets in (C) (left y axis). Charge balance ratios between MED1-IDR-GFP and increasing concentrations of the indicated RNAs are shown as blue lines (right y axis). Correlation between partition ratio and charge balance is determined by Pearson correlation (r).
See also Figures S2–S4.

Figure 3. RNA-Mediated Effects on Condensates in Reconstituted In Vitro Transcription Assays



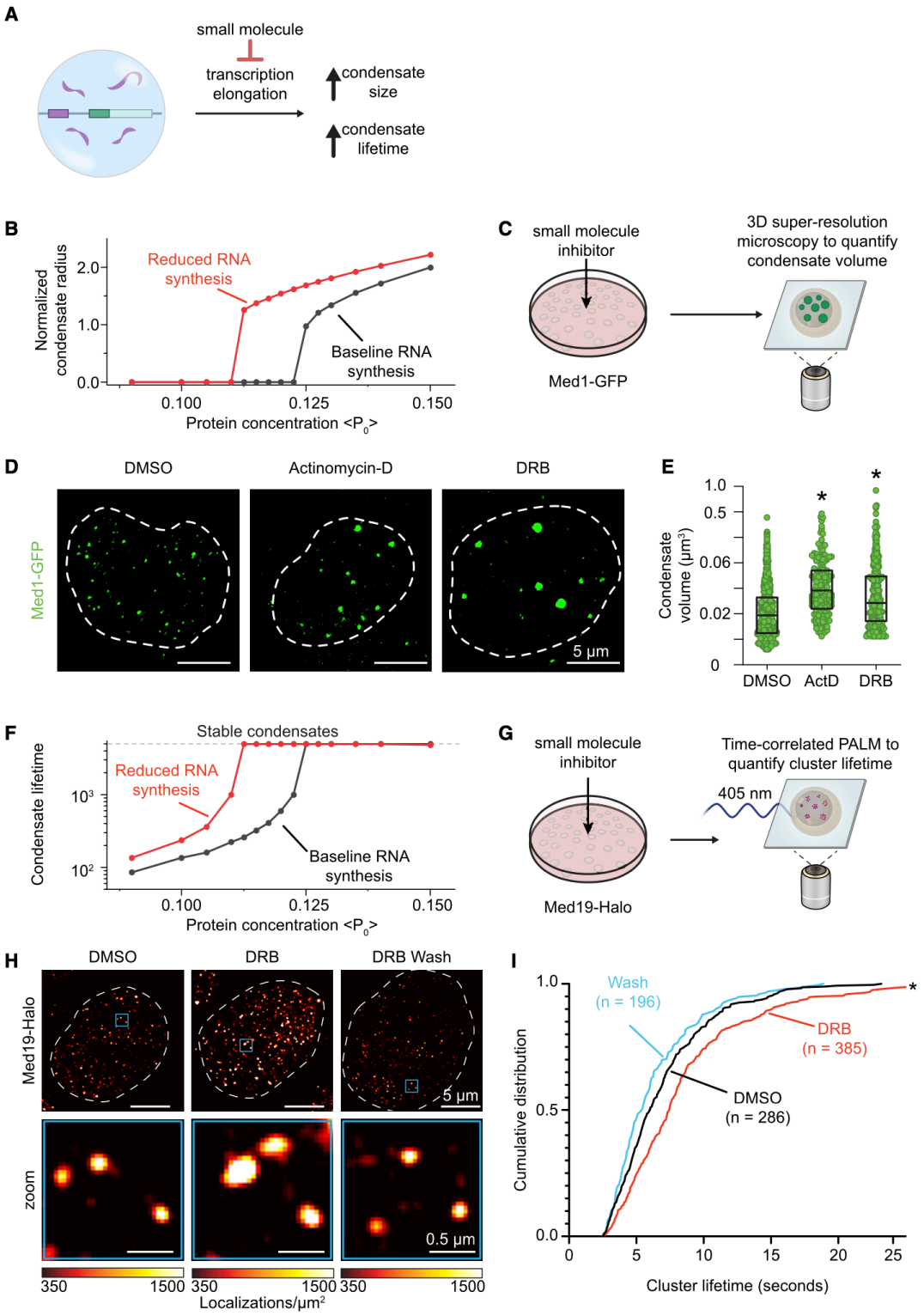
(A) Cartoon representation of the reconstituted *in vitro* mammalian transcription assay with purified components (left) and the design of the assay (right) (STAR Methods).
(B) Bright-field images of droplets formed within the *in vitro* transcription reaction. Droplets are stained with DNA dye (Hoechst). Bright-field images were white tophat filtered and smoothed here and in the subsequent panels (STAR Methods) (scale bars = 5 mm).
(C) Bright-field images of droplets formed within the *in vitro* transcription reaction performed in the presence of the indicated spermine concentrations. Template DNA is labeled with Cy3 (scale bars = 5 mm).
(D) Droplet sizes in (C) ($p = 0.0011$, Student's t test).
(E) Partition ratio of Cy3-labeled template DNA in the droplets in (C) ($p < 0.0001$, Student's t test).
(F) qRT-PCR of transcriptional output upon addition of spermine. The values are normalized to the no-spermine condition. The mean of 2 replicates is shown, and error bars depict SD ($p = 0.0477$, Student's t test).
(G) Representative images of droplets in the *in vitro* transcription reaction in the presence of the indicated amounts of exogenous RNA (scale bars = 5 mm). (H) Droplet sizes in (G) ($p = 0.9309$ 0 versus 10; $p < 0.001$ for 0 versus 50, 250, and 500; one-way ANOVA).
(I) qRT-PCR of transcriptional output upon addition of increasing concentrations of exogenous RNA. The values are normalized to the no-RNA condition. The mean of 2 replicates is shown, and error bars depict SD ($p = 0.0001$, GTP only versus 0; $p = 0.0111$, 0 versus 10; $p = 0.0013$, 0 versus 50; $p = 0.0008$, 0 versus 250; $p = 0.008$, 0 versus 500; one-way ANOVA).
See also Figure S5.

Figure 4. A Model of RNA-Mediated Non-equilibrium Feedback Control of Transcriptional Condensates



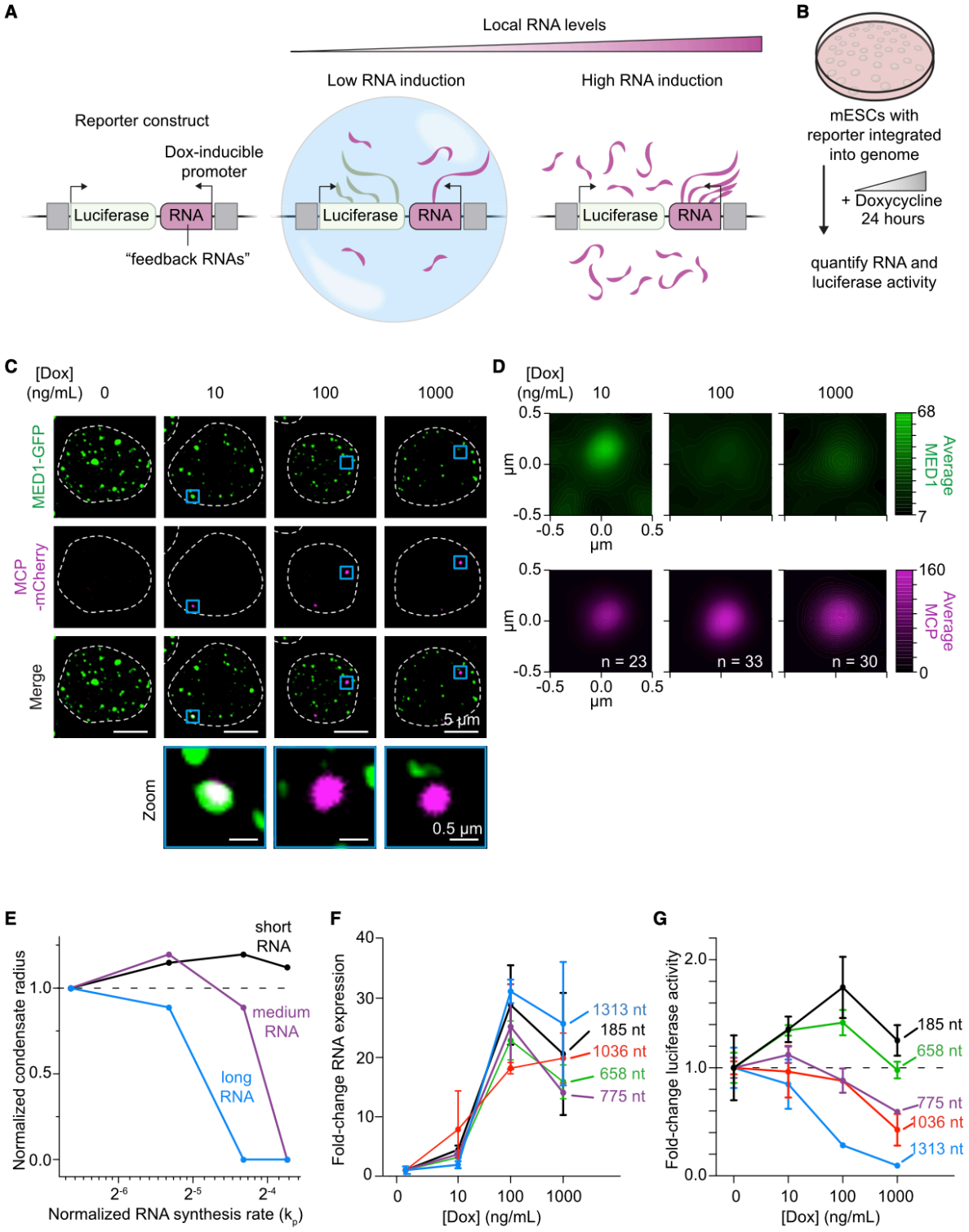
(A) Schematic of coarse-grained free energy (f, green surface), which depends on the transcriptional protein ϕ_p and RNA ϕ_r concentrations. This free energy recapitulates *in vitro* observations of an equilibrium reentrant transition.
(B) Schematic of the non-equilibrium model coupling transcriptional activity with transcriptional condensate dynamics. In the model framework, we focus on a local micro-environment near a single transcriptional condensate (blue). RNA (magenta) is synthesized and degraded and can diffuse.
(C) Equations underlying construction of the free-energy function (equation 1) and dynamics of protein and RNA (equation 2) (STAR Methods).
(D) Simulation predictions of transcriptional condensate lifetime with varying total protein concentrations (2D simulation grid). The dashed line represents the lifetime of condensates (in units of simulation time) that do not dissolve at steady state.
(E and F) Simulation predictions of transcriptional condensate radius (E) and lifetime (F) at varying effective rates of RNA synthesis (2D simulation grid). The radius values are normalized to r = 6.0 mesh units. The dashed line in (F) represents the lifetime of stable condensates in units of simulation time (STAR Methods).
(G) Variation of normalized condensate radius (ordinate, normalized to r = 6.0 mesh units) with changing relative timescales of reaction and diffusion (abscissa, $td = tr$) (2D simulation grid). In these simulations, the total effective concentration of RNA produced is held constant (see text). The inset graphs the distribution of RNA concentrations at early simulation times $t_{step} = 100$ for two different values of $td = tr$ (highlighted in the main panel with corresponding colors).
(H) Visualization of protein (blue) and RNA (magenta) concentration fields over simulation time for 3D simulations. The condensate is initialized (first panel) and then grows under low transcriptional activity (second panel). After a finite time $t_{sim} = 1000$, the effective rate of RNA synthesis k_p is increased by 2.5-fold, which, in turn, drives condensate shrinkage (third panel) and, ultimately, dissolution (fourth panel) (STAR Methods).
See also Figures S6 and S7.

Figure 5. Inhibition of RNA Elongation Leads to Enhanced Condensate Size and Lifetime in Cells



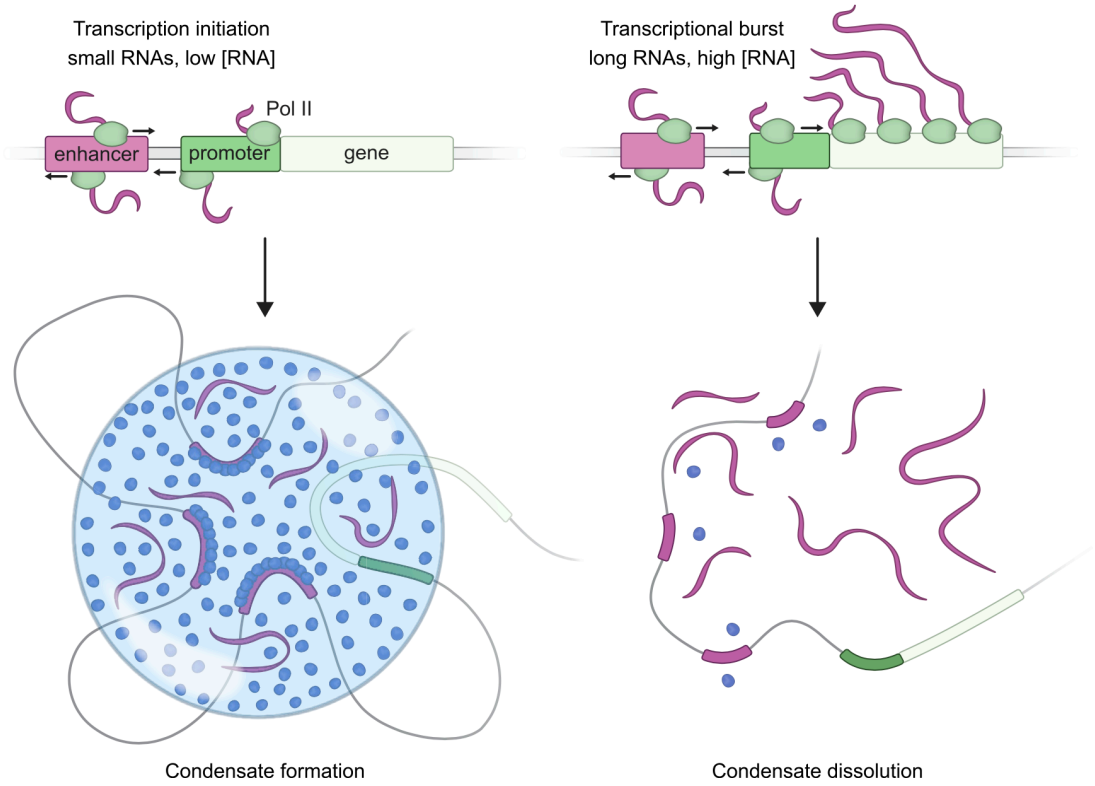
(A) Scheme for preventing condensate dissolution upon transcriptional burst by treatment with small molecules that inhibit transcriptional elongation.
(B) Simulation predictions show variation of normalized condensate radius with total protein amount (abscissa) in the absence (black, *kp* = 0.1) and presence (red,*kp* = 0:05) of RNA synthesis inhibition (2D simulation grid). The radius is normalized by the radius at *kp* = 0:05; < *P*0 > = 0:115.
(C) Experimental design to test the effect of transcriptional inhibition on the size of Mediator condensates. MED1-GFP mESCs are imaged by 3D super-resolution microscopy after treatment with small molecules.
(D) Maximum intensity projection images of single nuclei tagged with endogenous Med1-GFP in the presence of the indicated transcriptional inhibitors or DMSO control (scale bars = 5 mm).
(E) Volumes of Med1-GFP condensates in (D) (p for DMSO versus ActD < 0.0001 and p for DMSO versus DRB < 0.0001, one-way ANOVA).
(F) Simulation predictions show variation of condensate lifetime with total protein amount (abscissa) in the absence (black, *kp* = 0.1) and presence (red,*kp* = 0:05) of RNA synthesis inhibition (2D simulation grid). The lifetime is presented in units of simulation time.
(G) Experimental design to test the effect of DRB on the lifetime of Mediator clusters in Med19-tagged mESCs. Lifetimes are quantified by time-correlated PALM. (H) Representative heatmap of Med19-Halo localizations in a single nucleus upon addition of the transcriptional inhibitor DRB, DRB wash, or DMSO control (top scale bars = 5 mm; bottom scale bars = 0.5 mm).
(I) Cumulative distribution frequency plot of condensate lifetime in response to the indicated treatments (p < 0.0001, one-way ANOVA).
See also Figure S7.

Figure 6. Increasing the Levels of Local RNA Synthesis Reduces Condensate Formation and Transcription in Cells



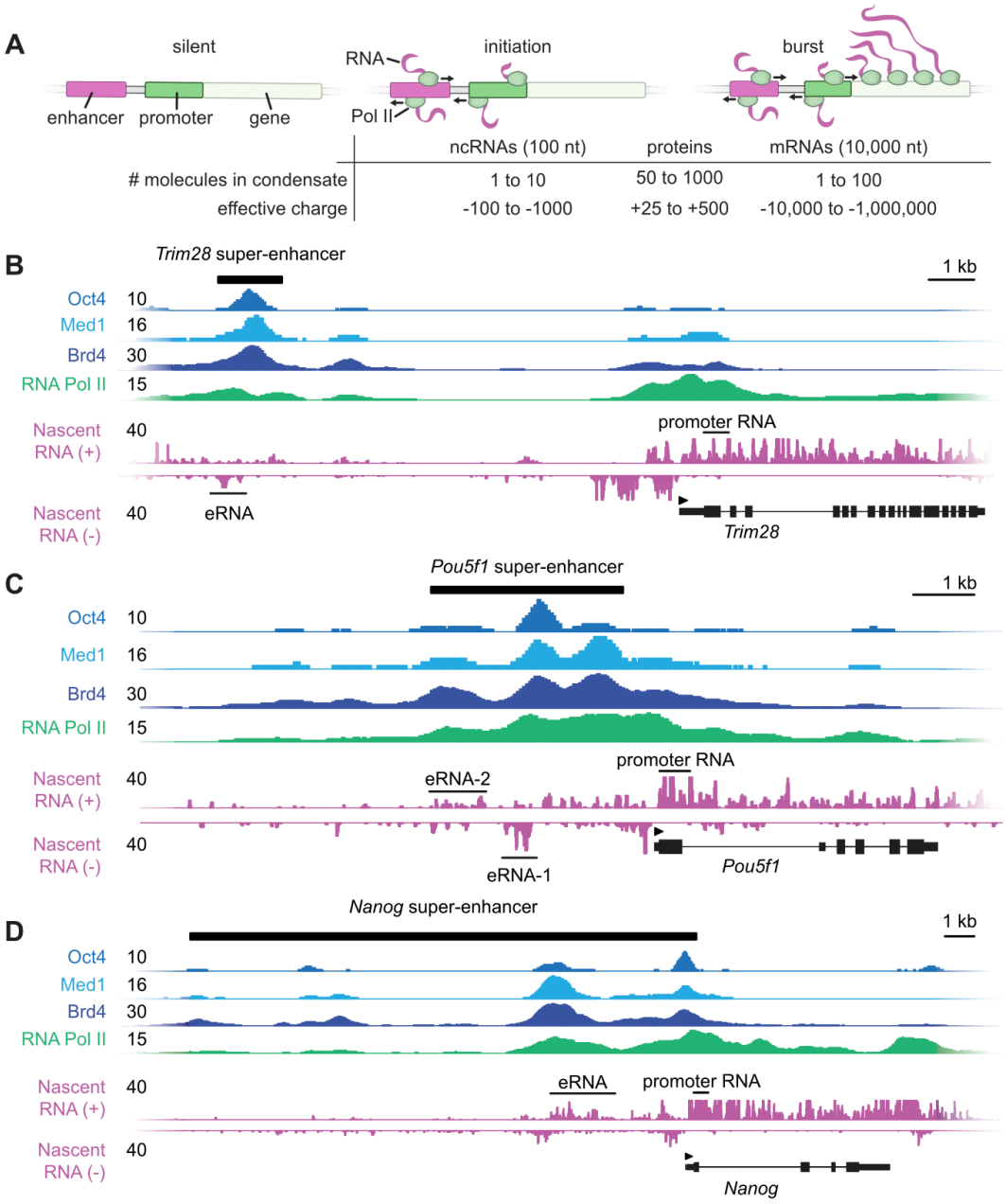
(A) Scheme depicting the reporter system (left) where local RNA expression near a luciferase reporter gene can be induced by Dox.
(B) Experimental design to test the effect of increasing local RNA levels on condensate formation and reporter gene expression.
(C) Live-cell imaging showing localization of Mediator condensates and MS2-tagged RNA expressed near the reporter gene with the indicated Dox stimulations. Med1-GFP mESCs have an integrated reporter system and 23-MS2 coat protein (MCP)-mCherry to visualize MS2-tagged RNA (2,456 nt). Representative images are maximum projections that have been subtracted by a median filter and smoothed (STAR Methods) (top scale bars = 5 mm; bottom scale bars = 0.5 mm).
(D) Average density of the MED1 signal centered at the RNA signal with the indicated Dox stimulations ($p = 0.066$, 10 ng/mL versus 100 ng/mL Dox; $p = 0.013$, 10 ng/mL versus 1000 ng/mL Dox; $p = 0.315$, 100 ng/mL versus 1000 ng/mL; 2-way Kolmogorov-Smirnoff test).
(E) Simulations predict the variation of condensate size with increasing effective rates of RNA synthesis (abscissa) (2D simulation grid). The condensate radius is normalized by value at rate = 1, and RNA synthesis rates are normalized to $kp = 0{:}02$ (STAR Methods).
(F) qRT-PCR of various ‘‘feedback RNAs’’ with increasing Dox concentrations. Markers show the mean of at least 3 replicates, and error bars depict the SD. (G) Luciferase luminescence with increasing Dox concentrations. Markers show the mean of at least 3 replicates, and error bars depict the SD.
See also Figure S7.

Figure 7. A Model for RNA-Mediated Feed-back Control of Transcriptional Condensates



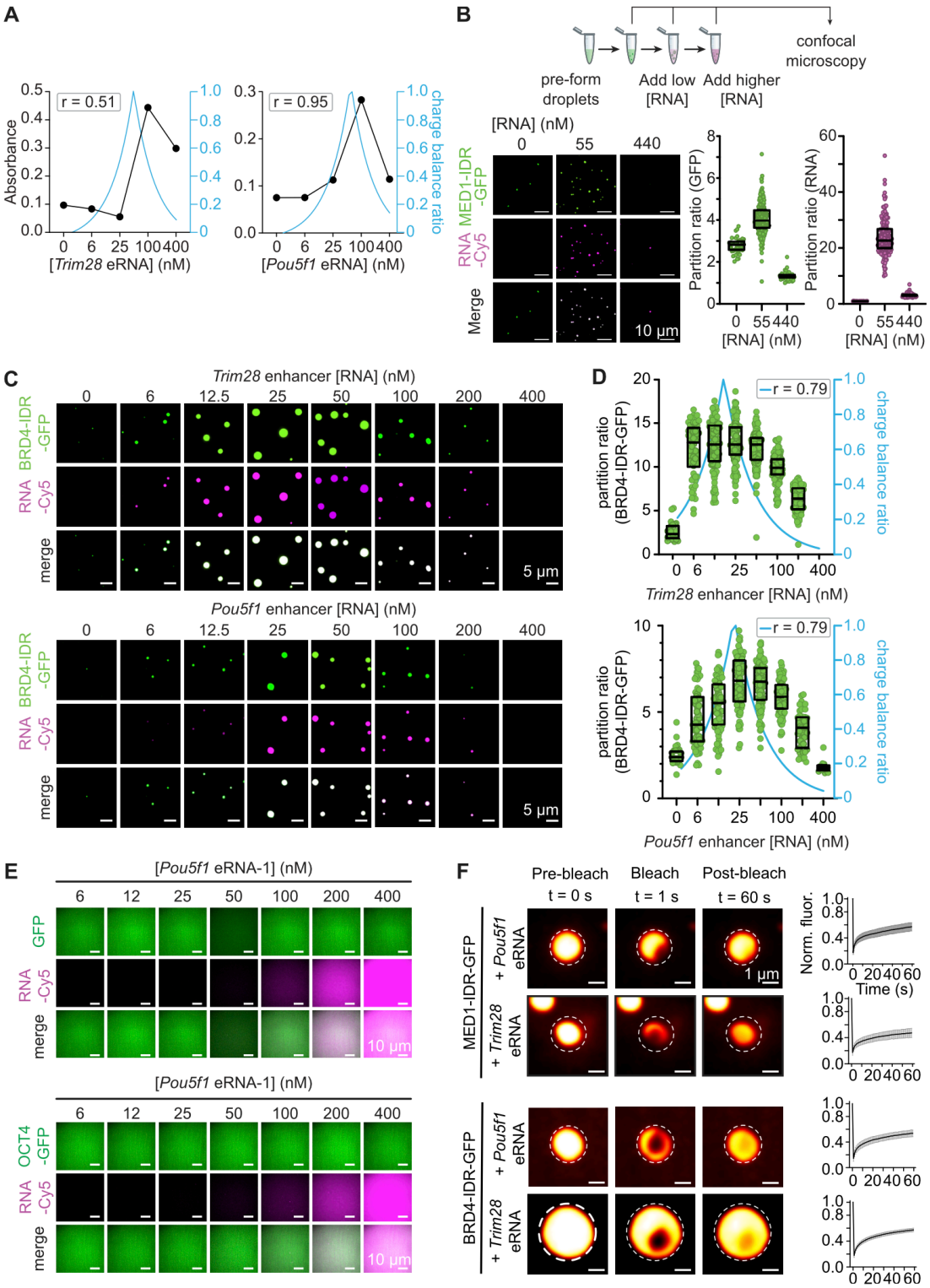
The cartoon depicts a model where low levels of RNA present at transcription initiation promote condensate formation, whereas high levels of RNA present during a transcriptional burst promote condensate dissolution.

Figure S1. Transcription Machinery and RNA at Active Genes in mESCs, Related to Figure 1



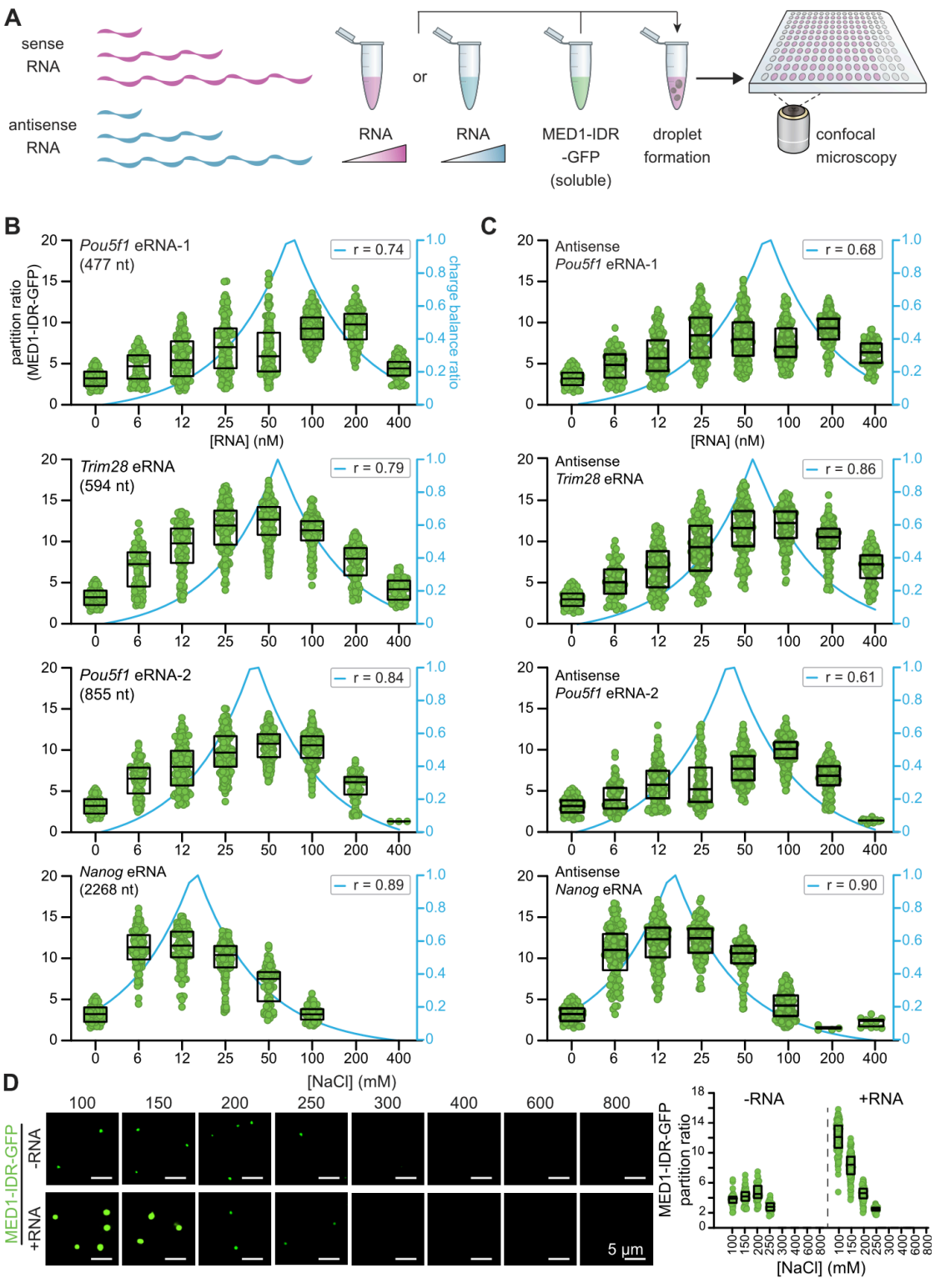
A. A scheme of transcription states and the number of molecules and their corresponding effective charge in a typical transcriptional condensate during initiation and bursts of transcription (STAR Methods)
B. C.D. Enrichment of transcription machinery and RNA at *Trim28* (A), *Pou5f1* (B) and *Nanog* (C) super-enhancers in mESCs. Gene tracks of ChIP-seq and nascent RNA-seq data at the indicated super-enhancers are shown. The enhancer- and promoter-derived (sense) RNAs that are used in this study are annotated in the gene tracks.
E. Nascent (left) or steady-state (right) levels of indicated RNAs at super and typical enhancers (eRNA = enhancer RNAs, uaRNA = upstream antisense promoter- associated RNAs, mRNA = messenger RNA).
F. Quantification of the number of enhancer RNA and pre-mRNA molecules in cells. Calculations are based on two biological replicates (STAR Methods).

Figure S2. Characterization of the Effect of RNA on Droplet Formation and Dissolution, Related to Figure 2



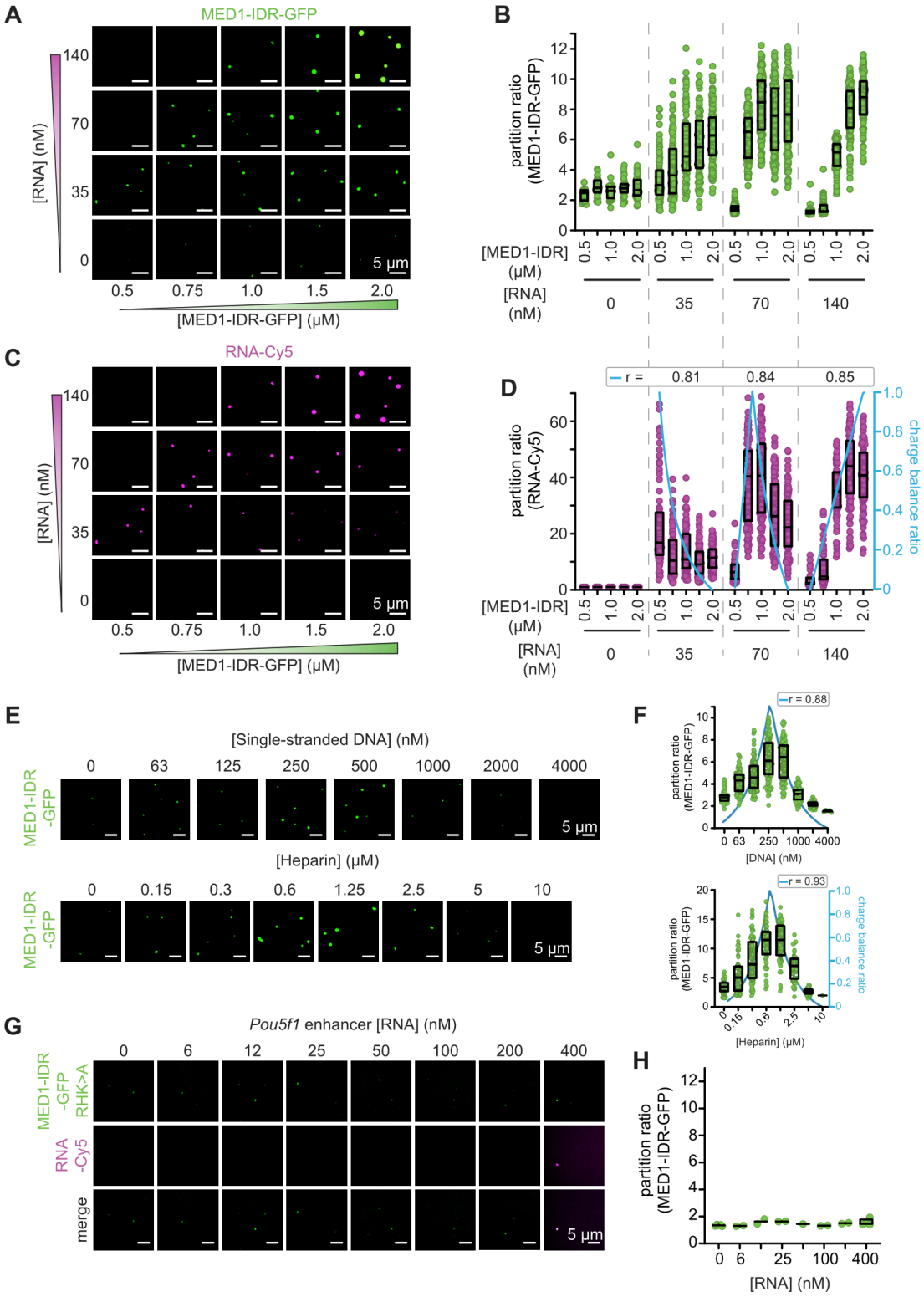
A. Turbidity measurements of droplets formed with MED1-IDR-GFP and indicated RNAs. Correlation between partition ratio and charge balance is determined by Pearson correlation (r).
B. Experimental design to test the effect of RNA on pre-formed MED1-IDR droplets (top). Representative images of MED1-IDR droplets and quantification of partition ratio of protein and RNA (bottom). Indicated concentrations of RNA were added after formation of droplets with 1 mM of MED1-IDR (scale bars = 10 mm).
C. Representative images of BRD4-IDR droplets at various RNA concentrations (scale bars = 5 mm).
D. Quantification of BRD4-IDR partition ratio from (C) and correlation with charge balance (blue lines). Correlation between partition ratio and charge balance is determined by Pearson correlation (r).
E. Purified GFP (top) or OCT4-GFP (bottom) was incubated with an enhancer RNA from the *Pou5f1* locus. Whereas this RNA could stimulate MED1-IDR-GFP condensate formation, it was unable to form droplets with GFP alone or OCT4-GFP. Images were adjusted to show signal and lack of droplet formation (scale bars = 10 mm).
F. FRAP analysis of droplets formed with MED1-IDR and RNA (top) or BRD4-IDR and RNA (bottom) (scale bars = 1 mm).

FigureS3. Modulation of Charge Balance between MED1-IDR and RNA Contributes to Stimulation and Dissolution of Condensates, Related to Figure 2



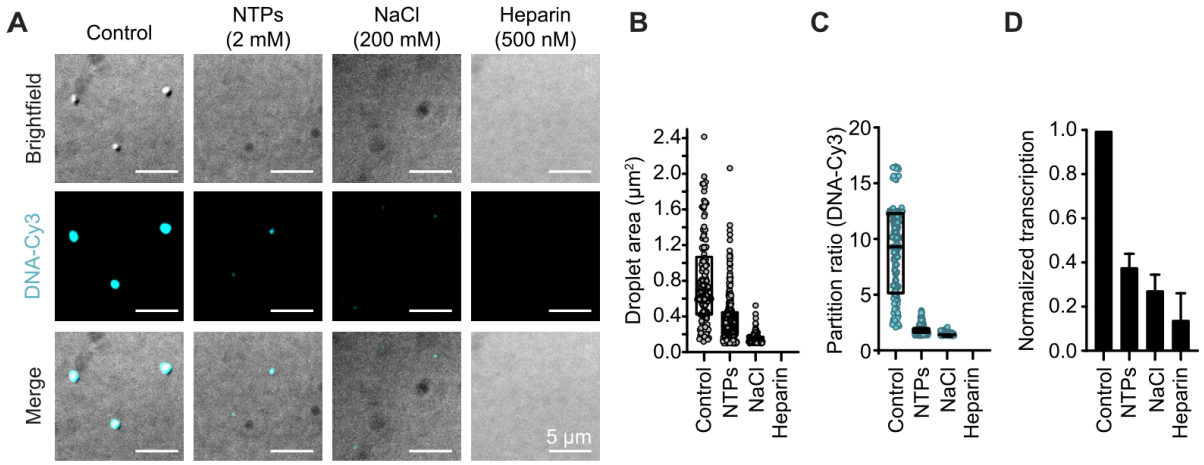
A. Experimental design for testing diverse sense and antisense RNAs of different lengths on formation of MED1-IDR-GFP droplets.
B. Quantification of the partition ratios of MED1-IDR-GFP within the droplets when incubated with RNAs of different lengths and sequences. Correlation between partition ratio and charge balance is determined by Pearson correlation (r).
C. Quantification of the partition ratios of MED1-IDR-GFP within the droplets when incubated with antisense versions of the RNAs in (B). Correlation between partition ratio and charge balance is determined by Pearson correlation (r).
D. Representative images of MED1-IDR droplets (left), which are formed with or without RNA and are subjected to increasing concentration of monovalent salt (NaCl). Quantification of partition ratios of MED1-IDR-GFP within the droplets are indicated (right) (scale bars = 5 mm).

Figure S4. Formation and Dissolution of MED1-IDR Droplets through Electrostatic Interactions, Related to Figure 2



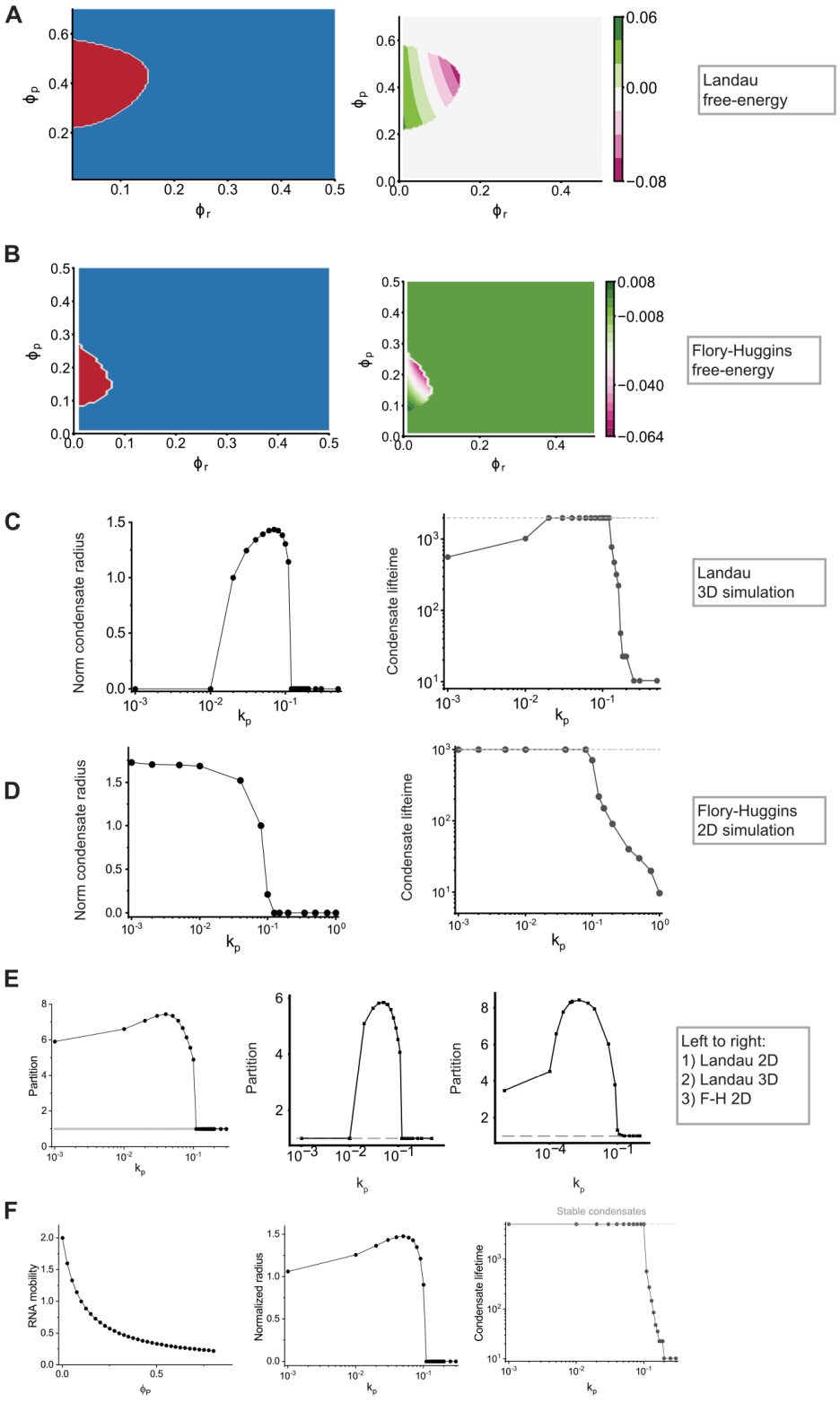
A. Representative images for MED1-IDR-GFP in a two-component phase diagram for MED1-IDR and *Pou5f1* eRNA (scale bars = 5 mm).
B. Quantification of MED1-IDR partition ratio for (A).
C. Representative images for RNA-Cy5 in a two-component phase diagram for MED1-IDR and *Pou5f1* eRNA in (A) (scale bars = 5 mm).
D. Quantification of RNA-Cy5 partition ratio and charge balance ratio for (C). Correlation between partition ratio and charge balance is determined by Pearson correlation (r).
E. Representative images of droplets formed with single-stranded DNA (top) or heparin (bottom) (scale bars = 5 mm).
F. Quantification of MED1-IDR partition ratio for images in (E). Correlation between partition ratio and charge balance is determined by Pearson correlation (r).
G. Representative images of droplets formed with MED1-IDR RHK > A and *Pou5f1* enhancer RNA (scale bars = 5 mm).
H. Quantification of MED1-IDR partition ratio for images in (G).

Figure S5. Charged Interactions in Reconstituted Mammalian Transcription and Droplets, Related to Figure 3



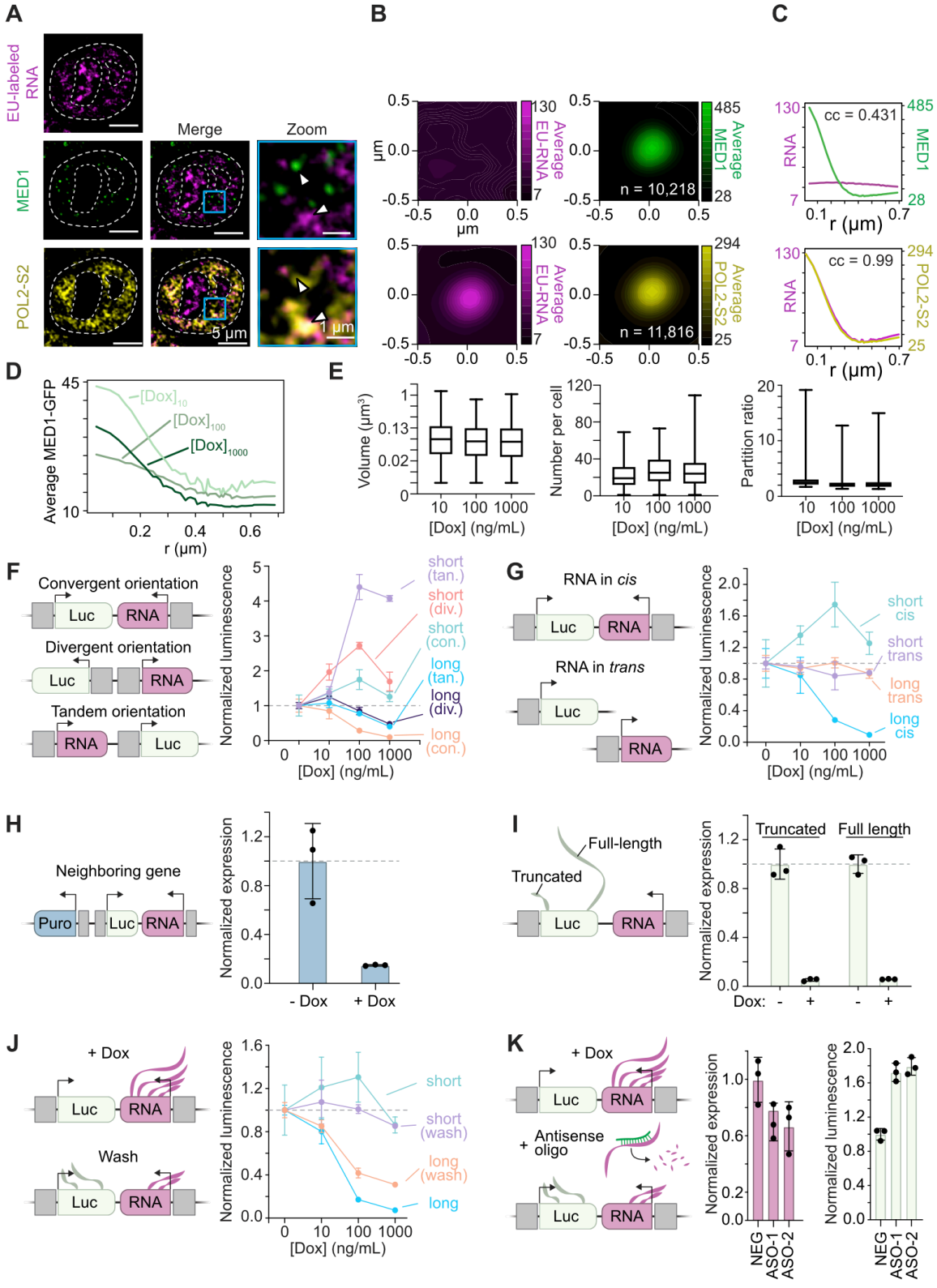
A. Representative images of transcription reactions with addition of 2mM NTPs, 200 mM NaCl, and 500 nM Heparin (scale bars = 5 mm).
B. Quantification of droplet area of droplets in (A).
C. Quantification of DNA-Cy3 partitioning from droplets in (A).
D. qRT-PCR measurement of template-derived RNA synthesis in the reconstituted transcription reactions from (A).

Figure S6. Computational Model of Non-equilibrium RNA Feedback on Transcriptional Condensates, Related to Figure 4



A. Regions where mixtures of protein and RNA phase separate spontaneously (red, left panel) are calculated from the Landau free-energy (Figure 4C) by analyzing the Jacobian (spinodal analyses; STAR Methods). As expected from the re-entry transition, increasing RNA concentration (abscissa) at fixed protein levels can start from a region promoting phase separation, and beyond a threshold, drive re-entry into dilute phase. The right panel shows the initial direction of the instability (STAR Methods), which indicates the RNA is enriched in protein condensates (value > 0, green shade), while at higher concentrations, RNA de-densifies the condensed phase (value < 0).
B. Similar analyses as in (A) are performed on a free-energy derived from Flory-Huggins model (STAR Methods).
C. Variation of condensate radius (left panel, normalized to value of R at *kp* = 0.02) and condensate lifetime (right panel) with effective rates of RNA synthesis (*kp* , abscissa) for simulations performed in 3D employing the Landau free-energy (Figure 4C; STAR Methods). Low values of *kp* promote condensate stability whereas higher rates drive dissolution. The dashed line in the right-panel represents the conditions under which condensates are stable in the simulations and condensate lifetime is presented in units of simulation time (STAR Methods).
D. Similar analyses as in (C) are performed on a free-energy derived from Flory-Huggins model on a 2D grid (STAR Methods). Values of the condensate radius are normalized to value of R at *kp* = 0:08..
E. Partition ratio, computed as maximum RNA concentration in condensate divided by dilute phase concentrations, are presented for simulations employing the Landau free-energy in 2 & 3-D as well as those employing the Flory-Huggins model in 2-D (left to right). When condensates are dissolved, the expected value of this ratio is 1 (as depicted by dashed gray lines). These calculations correspond to simulation data from Figures 4D, 4E, S6C, and S6D, respectively (left to right).
F. RNA diffusivity/mobility decreases with increasing protein concentration (abscissa) with an inverse correlation (left panel, MRNA f ϕP-1), consistent with ideal models of crowding. Variation of condensate radius (middle, normalized to r = 6.0 units) and lifetime (right) with effective rates of RNA synthesis under conditions of phase-dependent RNA mobility. The simulations recapitulate key trends predicted by models with constant diffusion/mobility that are shown in Figures 4E and 4F.

Figure S7. The Effect of Local RNA Synthesis on Transcriptional Condensates and Transcription in Cells, Related to Figures 4, 5, and 6



A. Immunofluorescence of MED1 and POL2-S2 with EU-labeled RNA (10-minute incubation) in WT mESCs. Representative images are a single z-plane that has been subtracted by a median-filtered image and smoothed (STAR Methods) (left scale bars = 5 mm; right scale bars = 1 mm).
B. Average signal analysis of EU-RNA, MED1 and POL2-S2. Average EU-RNA signal is centered at MED1 puncta (top) or at POL2-S2 puncta (bottom).
C. Radial distribution function with correlation between EU and MED1 (top) or POLII-S2 (bottom) channels for average IF signal in (B).
D. Radial distribution function of MED1-GFP signal for multiple dox concentrations from experiments in Figure 6C.
E. Global quantification of the volume (left), number per cell (middle), and partition ratio (right) of MED1-GFP condensates in reporter cells at multiple Dox concentrations.
F. Quantification of luciferase luminescence with increasing dox concentrations to stimulate expression of short (185 nt) and long (1313 nt) feedback RNAs with indicated orientations of feedback RNA and luciferase gene. The convergent data was collected as part of Figure 6G. Markers show the mean of at least 3 replicates and error bars depict the SD.
G. Quantification of luciferase luminescence with increasing dox concentration to stimulate expression of short (185 nt) and long (1313 nt) feedback RNAs with *cis* or trans integration of feedback RNA and luciferase reporter gene. The *cis* data was collected as part of Figure 6G. Markers show the mean of at least 3 replicates and error bars depict the SD.
H. RT-qPCR of neighboring puromycin resistance marker gene with and without stimulation of long (1313 nt) feedback RNA expression with dox.
I. RT-qPCR of luciferase mRNA gene with primers detecting truncated and full-length RNA with and without stimulation of long (1313 nt) feedback RNA expression with dox.
J. Quantification of luciferase luminescence with increasing dox concentration to stimulate expression of short (185 nt) and long (1313 nt) feedback RNAs, before and after dox washout.
K. The effect of antisense-oligo mediated degradation of feedback RNAs on luciferase luminescence. The expression of the feedback RNAs are measured by RT- qPCR (left bar graphs) and luciferase expression is measured by luminescence (right bar graphs). The luminescence values were first normalized to the no dox condition for that ASO, and then normalized to the dox condition of the negative control ASO.

Supplementary Table 1 - Primers used in this study, Related to Methods and Materials

Primers used to amplify templates for in vitro transcription assay	
The following sequences were added to the forward and reverse primers to add the bacterial polymerase promoters:	
T7 (add to 5' of sense or forward primer): 5'-TAATACGACTCACTATAGGG-3'	
SP6 (add to 5' of antisense or reverse primer): 5'-ATTTAGGTGACACTATAGAA-3'	
Primer ID	**Sequence (5' to 3')**
txn_eRNA_Oct4_002_F	GGCCTAGACAGCACTCTCCA
txn_eRNA_Oct4_002_R	TGGATCTCTGTGAGTTCAAG
txn_eRNA_Trim28_002_F	AAATCTTGGAGAGAGTAGGA
txn_eRNA_Trim28_002_R	GGGAAAAAGTTACAGTGACC
txn_eRNA_Oct4_003_F	CTTCCAGAACATCTGGATTT
txn_eRNA_Oct4_003_R	AAAACAAACAAAAAAGAGTC
txn_eRNA_Nanog_002_F	AGCCTGCCTTTTGGCTACCA
txn_eRNA_Nanog_002_R	AGAGTGCCAGGTCCCCTGGA
trx_pre-mRNA_Oct4_500_F	TAGGTGAGCCGTCTTTCCACC
trx_pre-mRNA_Oct4_500_R	CCCAATTCCCTTCACTGCTGC
trx_pre-mRNA_Trim28_500_F	CGGGCGGTGAGAAGCGT
trx_pre-mRNA_Trim28_500_R	AATGCATGCACACCCTCTGATT
Primers used to calculate number of RNA molecules in cells	
Primer ID	**Sequence (5' to 3')**
qPCR_15_eRNA_Trim28_002_F	AGAGGCTCTTCTGGGGTTGT
qPCR_16_eRNA_Trim28_002_R	GCGAACAAGTAGGGCCAGTT
qPCR_19_eRNA_Trim28_002_F	GCCCTGGATTGTACCTGTCC
qPCR_20_eRNA_Trim28_002_R	ACCTTCAAAGTGGGTAACGCT
qPCR_45_eRNA_Oct4_002_F	CAGGTTAGCCCTAAGCGTGC
qPCR_46_eRNA_Oct4_002_R	AGGCTAGGGCACATCTGTTT
qPCR_59_eRNA_Oct4_002_F	CCCTAAGCGTGCCTAGAGTAT
qPCR_60_eRNA_Oct4_002_R	ACCAGGCTAGGGCACATCT
qPCR_Trim28_intron_F3	GCTGCTGCCCTGTCTACATT
qPCR_Trim28_intron_R3	CTGGCCACCCAGTACTCACT
qPCR_Trim28_intron_F4	GAGTACTGGGTGGCCAGGT

qPCR_Trim28_intron_R4	CCCCCTCTTAAACCAGCAG
qPCR_Oct4_intron_F3	GTTGGAGAAGGTGGAACCAA
qPCR_Oct4_intron_R3	CCCAATTCCCTTCACTGCT
qPCR_Oct4_intron_F4	AGAGGGAACCTCCTCTGAGC
qPCR_Oct4_intron_R4	CAGCCAAGTCCCTTTCACTT
qPCR_mActb_F	GATCTGGCACCACACCTTCT
qPCR_mActb_R	TGGGGTGTTGAAGGTCTCA
qPCR_Tetris_5'F1	AGAATTCGAGCTCGGTAC
qPCR_Tetris_5'R1	GCgaattcCTAGTTAGCTAG
qPCR_Tetris_Luc_early_5'F	TTGCTCACGAATACGACGGT
qPCR_Tetris_Luc_early_5'R	CTGTACATCGGTGTGGCTGT
qPCR_Tetris_Luc_late_5'F	AAGAAGTGCTCGTCCTCGTC
qPCR_Tetris_Luc_late_5'R	TACGTTAACAACCCCGAGGC
qPCR_Tetris_Puro_5'F	GCTCGTAGAAGGGGAGGTTG
qPCR_Tetris_Puro_5'R	CACCAGGGCAAGGGTCTG
Primers used to quantify transcriptional output from reconstituted in vitro transcription assay	
Primer ID	**Sequence (5' to 3')**
qPCR_19_eRNA_Trim28_002_F (Spike-in control)	GCCCTGGATTGTACCTGTCC
qPCR_20_eRNA_Trim28_002_R (Spike-in control)	ACCTTCAAAGTGGGTAACGCT
GFP_qPCR_Fw (GFP primer)	cctgaagttcatctgcacca
GFP_qPCR_Rv (GFP primer)	gtcttgtaggtgccgtcgtc

Supplementary Table 2 - Lengths and charges of nucleic acids and proteins, Related to Figures 1, 2, 3

Nucleic acids	Strand sense	Length (bp)	Charge	Sequence
Nanog enhancer RNA	plus	2268	-2268	AGCCTGCCTTTTGGCTACCAGCCACCTCTTCGCTCGGATCTTTCACCAGAGACTCTCAAA GACACTAAAGAGGCAGGACAGGAATGGGGGTTGGGGAGGGATCCATCGCCGTCTCCTA AGCAGACTCCTTTGACCCGGAGCTGTGCGCCCTGTACCAAACCTTTGTAGAACTTGGGG TAAACTTAAGGCTATGGTGGCCTTGACTCCGTGGACCCAGAGGCAAGTTTCCTCCTTTA GAGGACTCGCATGCATTTTGTTTCTAATTTGAAATGAGAACCGGCTTAGAGCTTGAACCA GCCAGTTCTCTGGACTCCTCCCAGCTCTTACAATTCCTCTCCCGGACGGTTCCTAGAAG ACAAAGGCAAGCTTACCAAAATTACGTCGCCCTTGGGACACACCTAGGGTTCCCTGGTG GCATCTTTTTTTTTTCATTATAAACAGGAGTAAATTTTTGTAAGGGCAGAGCTGGTAGCTG AGGGAGAGGAACCCTTTGGCCTAGTGAAGGTAGTTTGCTGGGCTTTGTATCCCCGCCCC CACCTCCCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG AGAGAGAGAGAGAGAGAGAGAGAGACTACGTGGTTATTTCAAAAACTTGAGTGTGGC AAAAGTATGTAACTGGGATTAGTAAGCATTTCTTTCCTTAGTGAGATTGGAGTAGAGGGT GGGAAAGGACCTTAGAATCCTCGAATGTTGGGCTTAGGAATGGGGAGACAAGAGCCAT CACAGAATGCCTATTGTCCTTCAATATGTTAGCGATGGGCCCCGTGCTTTAGATTTTAGG CTTGTATTTTCTTTGTGTGTGTGTGTGTTTGTTTTGTTTCTGTTTCTTTAGGCAGTCTGGA GATCAGGCTGGCTTTCAACTCCCTGTGATGCCCCTACCTCTCCTGAGGTGTGAGTGGCA GAATGCTCAGCGGGTAAAACACACTTGTATAAGTGCAAGAGAACCAAGTTCCAAACATTG TCCTCTGACTTCCTCAGGCATGCCTTAGCTCTAACAAATAAAATTGAAGAAAAGGATTCC AAACCACGGTGAAGGTGCCACATCTTTGATCCCCTTAAGGCAGACAGAAGTAGAGGGAT CCCTGGGAGTTCAAAGCCAGGTTGGTCTGCATACTGAGTTGGCTAGCCAGGAATACACA CTGAGACTCTGTCTTTAAAAAAAAAAAAGTGTGAAGACTGCTTTCTCTGTCCCAGCACTT GGGAGAGAGGGAAAGAAACACAGCAGCCAGCCTTGTCTACACAGTGAGTTCCAGGCTA ACCAGATCTGACATAGTGAGGTCCTGACTCAAAATTAAAAATTGGCTACAGATAATACTG TAGCCCTTGGTTAGTCCGAGTACTTAACTCAACATACCATTCTTCGTTTAAGCAAACCAC GTGAAAGACTTTTCACTGAAGGCTGCAAGTCTTAAAATGACTTTGGTGATGCCTGCCTGG ACTGTCTACCCTCTGGAGCAGACTTACAAAGAATATTTTTTACTAAGCGCTGCATAAACC TTGATATTTTGAACGGCCTATTCATTCTTTGCCTAATGACAAGAATCACATCAGGGACATA TTTGTATTAGTCCAGCGAATAAGCAGAAGGTAGAACAGTTATTCTTTTGTTCTATGATTTC TACTTAGGGGCTTTAGTCCATCTGTTTATTGTTTAAAACCTTCATATCTCACCAAGTAGTG GTGGCCATCCCTTTAACCCCAGCATTCTGAGAGACAGAAGCAGGTGGGTCTCTGTGAGT TCGAGGCCAGCCTGGCCTACAGATCTAATTCCAGGATAGCCGGGGCTACATAGAGAAAC TCTGTCTCAAAATAAATAAGTAAATAAATAAATAAGTAAATGATATATTTACATTATTTATTA TATCACAAATTATATACGTGTTCTATAGAATATGTTACTATATACATATTTTTTTTGGTTTTT CGAGACAGGGTTTCATATAGCATTAGCTTCATATATATAATTTATATGCAAAAATATCTCT GAAAATGGAATCACTGGAACCCAATTCTAAAAGTATTGTTTTTTTTTGTTTTTAAAGTTTCC TCATACCTCAAAGTTGTCAGAGGAGGGCTTAAGAGATGGGCTAGAGGGGCTGGAGAGA TAGCTCAGCAGTTAAGAGCTCTGGTTGCTCTTCCAGAGGACCAGGGTTCAATTTCCAGT ACCTACATGACAGCTCATAACTATCTGTAATTTCAGTTCCAGGGGACCTGGCACTCT
Trim28 enhancer RNA	minus	594	-594	AAATCTTGGAGAGAGTAGGACCTGAGCTGTTTCCTATTTCTACTACTTCCATAGACATCC ATGGGTCTCCTTTGGTTCTTACCTTAAACTGCAGTTCTAGAGGCTCTTCTGGGGTTGTAA ACTGCCCTTCCCCAACTCAGGAGCTCCCATTCCCTCCCCTTCCTAAACTGGCCCTACTT GTTCGCTCTACTTTATCTTGCCAGGGGTTGTAACAGCCTCCAATCTGTCTTGGATCAGGA AGCAGTTTCCATGGGGTATTTTAAGTGTAAGTTCCCCGGTGAAATGGAAACAAAAAGAAT AAGGTATTCTGGGGACAGTGTAGGCCCACTAGGGTGTAGCTTGTTTTGTAAACACAGGA CCTGAGTTTTGGACTTTGCCCTGGATTGTACCTGTCCCAGTGGGTTACCTTAGATGACAC TAGCTTCATAATGCAAGAGGAGGTTAAGATGCTGGCTAGCGTTACCCACTTTGAAGGTA CTGGAAAAGGAAGACCTGATTTACCCAAAAGAGGGAGATCTCTGTGCTTCTACCGCTTTA ATTTTTTTTTTCTTTTTTTAAATATTTTTGCAGGACTGGTCACTGTAACTTTTTCCC
Pou5f1 enhancer RNA-1	minus	477	-477	GGCCTAGACAGCACTCTCCACCACAGTTCAAGTATGCCTGCAGCCCAGCAGTCCTGTCT GTATTCAATACCAACCTTGTCTTATGGATTGTGATTTCTCTTTTGGTGACTCACTGGCCAG GACAAGAGACATATTCTGAGTCCTTTAACTGCCTAGTGACTGGCTTTGCAACCAGGTTAG CCCTAAGCGTGCCTAGAGTATAATACAGTCCTTAACAGCAACTTTGTCTGAAGTCCCAAG TCTTCTTAGAACTAGCTTGAAACAGATGTGCCCTAGCCTGGTCCCGAGCTGTGAGCCTG GTGGCCCTGGAGATGGGACAGCAGACCTGTTGGCTCATCTGATCCAGTTTCTTGCCTCC TGGGTCTTAGAAAATCATCAGACTAACTTTTGGGTTTGATTCAAGGGTTCCCTACATAGC CTTAGATGGCTTGGAACTCACTCTGTAGACCAGGCTGGCCTTGAACTCACAGAGATCCA
Pou5f1 enhancer RNA-2	plus	855	-855	CTTCCAGAACATCTGGATTTGGGAAGAGACGTTGCTGGTCCCAGGGCGGCTGGGGGTT GGGGTTGGGGGAGGGGGATGCTAACCAGCAAGGAAGCTGTTCCTGGCTGGGGCAGGC CTGACTGAGCTCATGTCGCTGAAACTCCTCATTTCTCCCTATGGCTTCATAGGGAGACCC AGCCTGGATGCTAACACGAGTGATTTCCCTGCTCTAGTCTAGTGTCCTCCGTGAGTCCA TTTAACTGATCACCCAGTCTGTGAGGAGGTGGCTGAACTCACAGTAAGAAAGCTGTGGG GGTCAACGCCTATTGTTTGTTTGTTTTGTTTTAGACAAGGTCTCCTGCTGAGGCTGGCTC AAGCTGGCCTGGAGGACTCTTGTGTTTAAGGCTGGCCTTAAATTCTCTTTAAAAGAAAAT CATGTGTATATATGTGTGTCCATGTAACTGCAGATGACCACAGCAGCCAAGAGATTTCTG TTCTCCTAGCTGTAAACCACCCAATATGGGTGCTAGGAACAGAATTTTAAAAGGGTCCTC TGAAAGAGCAATGTACACTCAAATGCTGAGTTCTTTCCCAGCCCCTAGCCTTGGACCTTT GTTCTTATCACTTCCAACGCCCAAGGGCAGGCATTATAGGTGTGGCATTCCGCATCTGG CTTCCCAGGATACCTTTTCATGCTGGTGGACCATCTCTGGCTGGGGACGTGTGGGCTTC TCTGCTGTCTTTGGTTCTCCAGACAGAACTCCGAGACAGATCTTGACTTGGTTCTAAAAT

				ACAGGTGGTTTGTGGCAAGTTAACGAATTTTAGCTCAAATTTGGGGTATTTAAGATACCA TGGTGACTCTTTTTTGTTTGTTTT
Pou5f1 promoter RNA	plus	490	-490	TAGGTGAGCCGTCTTTCCACCAGGCCCCCGGCTCGGGGTGCCCACCTTCCCCATGGCT GGACACCTGGCTTCAGACTTCGCCTTCTCACCCCCACCAGGTGGGGGTGATGGGTCAG CAGGGCTGGAGCCGGGCTGGGTGGATCCTCGAACCTGGCTAAGCTTCCAAGGGCCTCC AGGTGGGCCTGGAATCGGACCAGGCTCAGAGGTATTGGGGATCTCCCCATGTCCGCCC GCATACGAGTTCTGCGGAGGGATGGCATACTGTGGACCTCAGGTTGGACTGGGCCTAG TCCCCCAAGTTGGCGTGGAGACTTTGCAGCCTGAGGGCCAGGCAGGAGCACGAGTGGA AAGCAACTCAGAGGGAACCTCCTCTGAGCCCTGTGCCGACCGCCCCAATGCCGTGAAG TTGGAGAAGGTGGAACCAACTCCCGAGGAGGTAAGTGAAAGGGACTTGGCTGGGCTGG CAGAGGCAGCAGTGAAGGGAATTGGG
Trim28 promoter RNA	plus	475	-475	CGGGCGGTGAGAAGCGTCCGGCTGCTTCCTcagccgcggcggcctctgcagccgcgtcgtcccctgcGG GGGGCGGTGGCGAGGCGCAGGAGCTTCTGGAGCACTGCGGCGTGTGTCGCGAGCGC CTGCGGCCCGAGCGGGATCCTCGGCTGCTGCCCTGTCTACATTCGGCCTGCAGTGCCT GCCTGGGCCCCGCTACACCCGCCGCAGCGAATAATTCGGGGGATGGCGGCTCGGCGG GCGACGGCGCTAGTGAGTACTGGGTGGCCAGGTGCCCCTCCCCCTCCTCGCAGCCCG TGCTCGGGACTGCGCCTGTGCGAGAGTATGGGGGCCCGGGTAGGGTTAAGTAGGCCT GCTGGTTTAAGAGGGGGCGGGGAACGGGTCCTGGCCTCTGCCAATGCCCGTTACCAGG TCTGGACACCGAGGTGCAGAATGTGATGGGAGATGTCAAGAAATCAGAGGGTGTGCAT GCATT
ssDNA	plus	117	-117	GGTTCTGCCGCAGGTGGATCCGGTATGTCCACCGCCACGACAGTCGCCCCCGCGGGG ATCCCGGCGACCCCGGGCCCTGTGAACCCACCCCCCCCGGAGGTCTCCAACCCCAGC AAG
Template DNA for IVT	NA	1000	-2000	ATGACCCTGCTGATTGGTTCGCTGACCATTTCCGGGTGCGGGACGGCGTTACCAGAAAC TCAGAAGGTTCGTCCAACCAAACCGACTCTGACGGCAGTTTACGAGAGAGATGATAGGG TCTGCTTCAGTAAGCCAGATGCTACACAATTAGGCTTGTACATATTGTCGTTAGAACGCG GCTACAATTAATACATAACCTTATGTATCATACACATACGATTTAGGTGACACTATAGAAT ACAAGCTTGCATGCCTGCAGGTCCTCGGAGGACAGTACTCCGCTCGGAGGACAGTACT CCGCTCGGAGGACAGTACTCCGCTCGGAGGACAGTACTCCGCTCGGAGGACAGTACTC CGACTCTAGAGGATCCCCGGTGTTCCTGAAGGGGGGCTATAAAAGGGGGTGGGGGCG CGTTCGTCCTCACTCTCTTCCCCTCCAAGCAAGGGCGAGGAGCTGTTCACCGGGGTGG TGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGCGCGG CGAGGGCGAGGGCGATGCCACCAACGGCAAGCTGACCCTGAAGTTCATCTGCACCACC GGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGT GCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCC GAAGGCTACGTCCAGGAGCGCACCATCTCCTTCAAGGACGACGGCACCTACAAGACCC GCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCA TCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTTCAACAG CCACAACGTCTATATCACGGCCGACAAGCAGAAGAACGGCATCAAGGCGAACTTCAAGA TCCGCCACAACGTCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACAC CCCCATC

Recombinant proteins	**Length (aa)**	**Charge**	**Amino acids included**
mEGFP	237	-6	SKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFF KSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRH NVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYK
MED1-IDR	626	43	EHHSGSQGPLLTTGDLGKEKTQKRVKEGNGTSNSTLSGPGLDSKPGKRSRTPSNDGKSKDKPPKRKKADTEGKSPSHSSSN RPFTPPTSTGGSKSPGSAGRSQTPPGVATPPIPKITIQIPKGTVMVGKPSSHSQYTSSGSVSSSGSKSHHSHSSSSSSSASTSG KMKSSKSEGSSSSKLSSSMYSSQGSSGSSQSKNSSQSGGKPGSSPITKHGLSSGSSSTKMKPQGKPSSLMNPSLSKPNISPS HSRPPGGSDKLASPMKPVPGTPPSSKAKSPISSGSGGSHMSGTSSSSGMKSSSGLGSSGSLSQKTPPSSNSCTASSSSFSSS GSSMSSSQNQHGSSKGKSPSRNKKPSLTAVIDKLKHGVVTSGPGGEDPLDGQMGVSTNSSSHPMSSKHNMSGGEFQGKRE KSDKDKSKVSTSGSSVDSSKKTSESKNVGSTGVAKIIISKHDGGSPSIKAKVTLQKPGESSGEGLRPQMASSKNYGSPLISGST PKHERGSPSHSKSPAYTPQNLDSESESGSSIAEKSYQNSPSSDDGIRPLPEYSTEKHKKHKKEKKKVKDKDRDRDRDKDRDK KKSHSIKPESWSKSPISSDQSLSMTSNTILSADRPSRLSPDFMIGEEDDDL

MED1-IDR RHK-A	627	-50	EAASGSQGPLLTTGDLGAEATQAAVAEGNGTSNSTLSGPGLDSAPGAASATPSNDGASADAPPAAAAADTEGASPSASSSNA PFTPPTSTGGSASPGSAGASQTPPGVATPPIPAITIQIPAGTVMVGAPSSASQYTSSGSVSSSGSASAASASSSSSSSASTSGA MASSASEGSSSSALSSSMYSSQGSSGSSQSANSSQSGGAPGSSPITAAGLSSGSSSTAMAPQGAPSSLMNPSLSAPNISPSA SAPPGGSDALASPMAPVPGTPPSSAAASPISSGSGGSAMSGTSSSSGMASSSGLGSSGSLSQATPPSSNSCTASSSSFSSSG SSMSSSQNQAGSSAGASPSANAAPSLTAVIDALAAGVVTSGPGGEDPLDGQMGVSTNSSSAPMSSAANMSGGEFQGAAEAS DADASAVSTSGSSVDSSAATSESANVGSTGVAAIIISAADGGSPSIAAAVTLQAPGESSGEGLAPQMASSANYGSPLISGSTPA AEAGSPSASASPAYTPQNLDSESESGSSIAEASYQNSPSSDDGIAPLPEYSTEAAAAAAAEAAAVADADADADADADADADAAAS ASIAPESWSASPISSDQSLSMTSNTILSADAPSALSPDFMIGEEDDDLM
BRD4-IDR	678	18	CLRKKRKPQAEKVDVIAGSSKMKGFSSSESESSSESSSSDSEDSETEMAPKSKKKGHPGREQKKHHHHHHQQMQQAPAPVP QQPPPPPQQPPPPPPPQQQQQPPPPPPPPSMPQQAAPAMKSSPPPFIATQVPVLEPQLPGSVFDPIGHFTQPILHLPQPELPP HLPQPPEHSTPPHLNQHAVVSPPALHNALPQQPSRPSNRAAALPPKPARPPAVSPALTQTPLLPQPPMAQPPQVLLEDEEPPA PPLTSMQMQLYLQQLQKVQPPTPLLPSVKVQSQPPPPLPPPPHPSVQQQLQQQPPPPPPPQPQPPPQQQHQPPPRPVHLQP MQFSTHIQQPPPPQGQQPPHPPPGQQPPPPQPAKPQQVIQHHHSPRHHKSDPYSTGHLREAPSPLMIHSPQMSQFQSLTHQ SPPQQNVQPKKQELRAASVVQPQPLVVVKEEKIHSPIIRSEPFSPSLRPEPPKHPESIKAPVHLPQRPEMKPVDVGRPVIRPPE QNAPPPGAPDKDKQKQEPKTPVAPKKDLKIKNMGSWASLVQKHPTTPSSTAKSSSDSFEQFRRAAREKEEREKALKAQAEHA EKEKERLRQERMRSREDEDALEQARRAHEEARRRQEQQQQQRQEQQQQQQQQAAAVAAAATPQAQSSQPQSMLDQQRE LARKREQERRRREAMAATIDMNFQS
POU5F1	360	-5	MAGHLASDFAFSPPPGGGGDGPGGPEPGWVDPRTWLSFQGPPGGPGIGPGVGPGSEVWGIPPCPPPYEFCGGMAYCGPQ VGVGLVPQGGLETSQPEGEAGVGVESNSDGASPEPCTVTPGAVKLEKEKLEQNPEESQDIKALQKELEQFAKLLKQKRITLGY TQADVGLTLGVLFGKVFSQTTICRFEALQLSFKNMCKLRPLLQKWVEEADNNENLQEICKAETLVQARKRKRTSIENRVRGNLE NLFLQCPKPTLQQISHIAQQLGLEKDVVRVWFCNRRQKGKRSSSDYAQREDFEAAGSPFSGGPVSFPLAPGPHFGTPGYGSP HFTALYSSVPFPEGEAFPPVSVTTLGSPMHSN
GAL4-VP16	232	-16	MKLLSSIEQACDICRLKKLKCSKEKPKCAKCLKNNWECRYSPKTKRSPLTRAHLTEVESRLERLEQLFLLIFPREDLDMILKMDS LQDIKALLTGLFVQDNVNKDAVTDRLASVETDMPLTLRQHRISATSSSEESSNKGQRQLTVSPEFPGIWAPPTDVSLGDELHLD GEDVAMAHADALDDFDLDMLGDGDSPGPGFTPHDSAPYGALDMADFEFEQMFTDALGIDEYGG

Purified protein complexes	UniProt ID	Length (aa)	Charge	
TFIIA p55	P52655	376	-34	TFIIA
TFIIA p15	P52657	109	0	TFIIA
TFIIB	Q00403	316	5	TFIIB
TBP	P20226	339	13	TFIID
TFIIE p56	P29083	439	-35	TFIIE
TFIIE p34	P29084	291	16	TFIIE
TFIIF Rap74	P35269	517	0	TFIIF
TFIIF Rap30	P13984	249	7	TFIIF
ERCC3	P19447	782	-2	TFIIH
ERCC2	P18074	760	-3	TFIIH
p62	P32780	548	7	TFIIH
p52	Q92759	462	6	TFIIH
p44	Q13888	395	-7	TFIIH
CDK7	P50613	346	4	TFIIH
Cyclin H	P51946	323	-1	TFIIH
p34	Q13889	308	-1	TFIIH
MAT1	P51948	309	-6	TFIIH
Rpb1	P24928	1970	-3	RNA Pol II
Rpb2	P30876	1174	-8	RNA Pol II
Rpb3	P19387	275	-17	RNA Pol II

Rpb5	P19388	210	-4	RNA Pol II
Rpb7	P62487	172	-5	RNA Pol II
Rpb6	P61218	127	-20	RNA Pol II
Rpb4	O15514	142	-11	RNA Pol II
Rpb8	P52434	150	-12	RNA Pol II
Rpb9	P36954	125	-9	RNA Pol II
Rpb11	P52435	117	-3	RNA Pol II
Rpb10	P62875	67	1	RNA Pol II
Rpb12	P53803	58	5	RNA Pol II
MED1	Q15648	1581	19	Mediator
MED4	Q9NPJ6	270	-14	Mediator
MED6	O75586	246	3	Mediator
MED7	O43513	233	-10	Mediator
MED8	Q96G25	268	0	Mediator
MED9	Q9NWA0	146	0	Mediator
MED10	Q9BTT4	135	-3	Mediator
MED11	Q9P086	117	-2	Mediator
MED14	O60244	1454	22	Mediator
MED23	Q9ULK4	1368	-3	Mediator
MED15	Q96RN5	788	16	Mediator
MED24	O75448	989	-6	Mediator
MED16	Q9Y2X0	877	-1	Mediator
MED25	Q71SY5	747	5	Mediator
MED17	Q9NVC6	651	-1	Mediator
MED26	O95402	600	13	Mediator
MED18	Q9BUE0	208	-4	Mediator
MED19	A0JLT2	244	17	Mediator
MED20	Q9H944	212	-1	Mediator
MED21	Q13503	144	-12	Mediator
MED22	Q15528	200	-16	Mediator
MED27	Q6P2C8	311	9	Mediator
MED28	Q9H204	178	-4	Mediator
MED29	Q9NX70	200	-2	Mediator
MED30	Q96HR3	178	2	Mediator
MED31	Q9Y3C7	131	3	Mediator

Materials and Methods

RESOURCE AVAILABILITY

Materials Availability
All unique/stable reagents generated in this study are available from the Lead Contact upon reasonable request with a completed Materials Transfer Agreement.

Data and Code Availability
The code generated during this study is available at:
https://github.com/krishna-shrinivas/2020_Henninger_Oksuz_Shrinivas_RNA_feedback.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines
The Jaenisch laboratory of the Whitehead Institute gifted the V6.5 murine embryonic stem cells. These cells are male cells derived from a cross of C57Bl/6(F) x 129/sv(M).

Cell culture conditions
ES cells were maintained at 37°C with 5% CO2 in a humidified incubator on 0.2% gelatinized (Sigma, G1890) tissue-culture plates in 2i medium with LIF, which was made according to the following recipe: 960 mL DMEM/F12 (Life Technologies, 11320082), 5 mL N2 supplement (Life Technologies, 17502048; stock 100X), 10 mL B27 supplement (Life Technologies, 17504044; stock 50X), 5 mL additional L-glutamine (Gibco 25030-081; stock 200 mM), 10 mL MEM nonessential amino acids (Gibco 11140076; stock 100X), 10 mL penicillin-streptomycin (Life Technologies, 15140163; stock 10^4 U/mL), 333 µL BSA fraction V (Gibco 15260037; stock 7.50%), 7 µL β-mercaptoethanol (Sigma M6250; stock 14.3 M), 100 µL LIF (Chemico, ESG1107; stock 10^7 U/mL), 100 µL PD0325901 (Stemgent, 04-0006-10; stock 10 mM), and 300 µL CHIR99021 (Stemgent, 04-0004-10; stock 10 mM). For confocal and PALM imaging, cells were grown on glass coverslips (Carolina Biological Supply, 633029) that had been coated with the following: 5 µg/mL of poly-L-ornithine (Sigma P4957) at 37°C for at least 30 minutes followed by 5 µg/mL of laminin (Corning, 354232) at 37°C for at least 2 hours. Cells were passaged by washing once with 1X PBS (Life Technologies, AM9625) and incubating with TrypLE (Life Technologies, 12604021) for 3-5 minutes, then quenched with serum-containing media made by the following recipe: 500 mL DMEM KO (Gibco 10829-018), MEM nonessential amino acids (Gibco 11140076; stock 100X), penicillin-streptomycin (Life Technologies, 15140163; stock 10^4 U/mL), 5 mL L-glutamine (Gibco 25030-081; stock 100X), 4 µL β-mercaptoethanol (Sigma M6250; stock 14.3 M), 50 µL LIF (Chemico, ESG1107; stock 10^7 U/mL), and 75 mL of fetal bovine serum (Sigma, F4135). Cells were passaged every 2 days.

METHOD DETAILS

ChIP-seq analysis
As described in Sabari et al., 2018, ChIP-seq browser tracks for MED1, Pol II, BRD4, and OCT4 were generated by aligning reads to NCBI37/mm9 using Bowtie with the following settings: “-p 4 --best -k 1 -m 1 --sam -l 40”. WIG files represent counts (in reads per million, floored at 0.1) of aligned reads within 50 bp bins. Each read was extended by 200nt in the direction of the alignment.
(Source: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE112808)

GRO-seq analysis
For generation of the GRO-seq browser tracks, GRO-seq reads were processed as described in (Sigova et al., 2015). The GRO-seq .sra file corresponding to GEO accession number GSM1665566 (Sigova et al., 2015) was converted to .fastq using the SRA toolkit (Leinonen et al., 2011). Reads were aligned to the mouse genome (NCBI37/mm9) using Bowtie v1.2.2 (Langmead et al., 2009) with the following settings "-e 70 -k 1 -m 10 -n 2 --best". The reads corresponding to each one of the features (super-enhancers, typical enhancers, proximal promoter regions, genes) were counted using featureCounts v1.6.2 (Liao et al., 2014) with default settings. The coordinates for typical enhancers and super-enhancers in mouse embryonic stem cells (mESCs) were acquired from (Whyte et al., 2013). The coordinates for genes (transcription start and end sites) were acquired using the UCSC Table Browser (Karolchik et al., 2004). The upstream antisense promoter regions were defined as genomic areas containing 1 kb upstream of each TSS. Their coordinates were retrieved by using BEDTools v.2.26.0 (Quinlan and Hall, 2010) and the TSS coordinates as input (to the slop function). Reads were normalized with the size of the corresponding feature they aligned to.

RNA-seq analysis
The RNA-seq .sra file corresponding to GEO accession number GSM2686137b (Chiu et al., 2018) was converted to .fastq using the SRA Toolkit RNA-seq analysis was performed using the nf-core RNA-seq pipeline (v1.4.2) (Ewels et al., 2020) with default settings and NCBI37/mm9 as reference genome. Nextflow v20.01.0 was used as a workflow tool on an LSF High-Performance Computing environment (Di Tommaso et al., 2017). STAR v2.6.1d (Dobin et al., 2013) was used for the alignment of reads. Aligned reads were assigned to the aforementioned intervals (typical enhancers, super-enhancers, proximal promoter regions and genes) by using featureCounts v1.6.4, with the default settings.

Calculation number of RNA molecules in cells
Known concentrations of in vitro transcribed enhancer RNAs and pre-mRNAs from *Trim28* and *Pou5f1* loci are used as standards to approximate the number of molecules in cells. These RNAs are converted to cDNAs by reverse-transcription and mixed at equal concentrations. For each RNA species, a standard curve of qRT-PCR Ct value to RNA amount was generated using serial dilutions, with two different primer sets in technical duplicates. Next, qRT-PCR reactions using the same primer sets were performed for biological duplicates of mESCs. Actb-normalized Ct values were then used to determine the amount of RNA species in the reaction based on the standard curves above. To calculate the number of RNA molecules per cell, the amount of RNA (g) was divided by the molar weight of each species (~350 (g mol^{-1} nt^{-1}) × length of in vitro transcribed RNA (nt)), multiplied by Avogadro's number (6.022×10^{23} mol^{-1}), and divided by the approximate number of cells used in each reaction (10,000 cells). Melting curves were analyzed to confirm primers specificity. Non-reverse-transcribed (–RT) controls were included to rule out the amplification of genomic DNA. Primer sequences are indicated in Table S1.

In vitro droplet assay
Recombinant GFP fusion proteins were concentrated to a desired protein concentration using Amicon Ultra centrifugal filters (30K MWCO, Millipore). Droplet reactions with the recombinant proteins were performed in 10 ul volumes in PCR tubes under the following buffer condition: 30 mM Tris HCl pH 7.4, 100 mM NaCl, 2% Glycerol and 1 mM DTT. The same buffer containing 55 mM NaCl was used for BRD4-IDR-GFP. Droplet reactions with the Mediator complex were performed under the following buffer condition: 30 mM HEPES pH 7.4, 65 mM NaCl, 2% Glycerol and 1 mM DTT. For all droplet reactions, protein and buffer were mixed first and RNA or ssDNA or heparin (Sigma, H3393) was added later. The reactions were incubated at room temperature for 1 hr without any shaking or rotating. The reactions were then individually

transferred into 384 well-plate (Cellvis P384-1.5H-N) by using a micropipette (2-20 µL) 5 minutes prior to imaging by confocal microscopy at 150X magnification or prior to turbidity measurements on a plate reader (Tecan) at 350 nm absorbance at room temperature (Banerjee et al., 2017). The concentration of proteins and RNAs in the droplet reactions are indicated in the figure legends. For brightfield Mediator experiments (Figure 1), representative images were subtracted by a median filtered image (px=15) using ImageJ to remove camera artifacts discovered by taking images of blank wells.

Fluorescence recovery after photobleaching (FRAP)
FRAP was performed on an Andor Revolution Spinning Disk Confocal microscope with 488-nm laser. Droplets were bleached using 30% laser power with 20 µs dwell time for 5 pulses, and images were collected every second for 60 seconds. Fluorescence intensity at the bleached spot and a control unbleached spot was measured using ImageJ. Values are normalized to the unbleached spot to control for photobleaching during image acquisition and then normalized to the first time point intensity.

MATLAB™ scripts were written to process the intensity data, and post bleach FRAP recovery data was normalized to pre-bleach intensity (FRAP(t)) and fit to:
$FRAP(t) = M(1-\exp(-t/\tau))$

Where, M (mobile fraction) and τ (half-life of recovery) are inferred in-built MATLAB functions. These values are inferred for each replicate and averaged to provide a range for the apparent diffusion coefficients, which is computed as:

$D_{app}=(\text{Bleach radius})^2/\tau$

In vitro droplet analysis
To analyse in vitro droplet experiments, we used a previously reported pipeline (Guo et al., 2019). The code for this analysis is available at the Github link in the Data/Code Availability section. All droplets were segmented from average images of captured channels on various criteria: (1) an intensity threshold that was three s.d. above the mean of the image; (2) size thresholds (20 pixel minimum droplet size); and (3) a minimum circularity ($\text{circularity}=4\pi\cdot(\text{area})/(\text{perimeter}^2)$) of 0.8 (1 being a perfect circle). After segmentation, mean intensity for each droplet was calculated while excluding pixels near the phase interface. Hundreds of droplets identified in (typically) ten independent fields of view were quantified. The mean intensity within the droplets (C-in) and in the bulk (C-out) were calculated for each channel. The partition ratio was computed as (C-in)/(C-out).

Droplet size, partition ratio, and condensed fraction measure distinct properties of droplet formation, and these three metrics show similar trends upon RNA-mediated reentrant phase transitions. When a protein or RNA is fluorescently-labeled in our experiments, we favor measuring the partition ratio. This is because the partition ratio can be measured on a per-droplet basis, and unlike condensed fraction, which varies depending on the number of droplets per field, the partition ratio is more independent of the field that is imaged.

For the size analysis of droplets formed in the reconstituted transcription assays (Figure 3), brightfield images were subtracted by a median-filtered image (px=21), and droplets were manually segmented and their areas measured using ImageJ.

Synthesis of RNA by in vitro transcription

Enhancer and promoter sequences for RNAs were obtained from super-enhancer-regulated genes *Pou5f1*, *Nanog*, and *Trim28*. For promoter sequences, the first 475-490bp from the first exon were selected from mm10. For enhancer sequences, GROseq reads (Sigova et al., 2015) from both + and - strands aligned to mm9 were overlapped with called super-enhancers (Whyte et al., 2013). Contiguous regions of read density above background were manually selected (Figure S1). Primers were designed to amplify the selected promoter and enhancer sequences from genomic DNA isolated from V6.5 mESCs (Table S1). The following sequences were added to the forward and reverse primers to add the bacterial polymerase promoters:

T7 (add to 5' of sense or forward primer): 5'-TAATACGACTCACTATAGGG-3'
SP6 (add to 5' of antisense or reverse primer): 5'-ATTTAGGTGACACTATAGAA-3'

Phusion polymerase (NEB) is used to amplify the products with the bacterial promoters, and products are run on a 1% agarose gel, gel-purified using the Qiaquick Gel Extraction Kit (Qiagen), and eluted in 40 µL H2O. Templates were sequenced to verify their identity. A volume of 8 µL of each template (10-40 ng/µL) was transcribed using the MEGAscript T7 (Invitrogen; sense) or MEGAscript SP6 (Invitrogen; antisense) kits according to the manufacturer's instructions. For visualization of the RNA by microscopy, reactions included a Cy5-labeled UTP (Enzo LifeSciences ENZ-42506) at a ratio of 1:10 labeled UTP:unlabeled UTP. The in vitro transcription was incubated overnight at 37°C, then 1 µL TURBO DNAse (supplied in kit) was added, and the reaction was incubated for 15 minutes at 37°C. The MEGAclear Transcription Clean-Up Kit (Invitrogen) was used to purify the RNA following the manufacturer's instructions and eluting in 40 µL H2O. RNA was diluted to 2 µM and aliquoted to limit freeze/thaw cycles, and RNA was run on 1% agarose gels in TBE buffer to verify a single band of correct size.

Recombinant protein purification
Recombinant protein purifications were performed as previously reported (Boija et al., 2018; Guo et al., 2019; Sabari et al., 2018; Shrinivas et al., 2019; Zamudio et al., 2019). pET expression plasmids containing 6xHIS tag and genes of interest or their IDRs tagged with either mEGFP or mCherry were transformed into LOBSTR cells (gift of I. Cheeseman Lab). Expression of proteins was induced by addition of 1mM IPTG either at 16°C for 18 hours or at 37°C for 5 hours. Extracts were prepared as previously described (Boija et al., 2018). Cell pellets were resuspended in 15 ml of denaturing buffer (50 mM Tris 7.5, 300 mM NaCl, 10 mM imidazole, 8M Urea) with protease inhibitors (Roche,11873580001). After complete resuspension, the lysates were sonicated for ten cycles (15 s on and 60 s off) and subjected to centrifugation at 12,000 g for 30 minutes. The supernatant was transferred into fresh tube and the lysates were incubated with 1 ml of pre-equilibrated Ni-NTA agarose beads (Invitrogen, R901-15) with denaturing buffer at 4°C for 1.5 hours. After washing the beads with 15 volumes of the denaturing buffer, proteins were eluted with 50mM Tris pH 7.4, 500mM NaCl, 250mM imidazole buffer containing complete protease inhibitors (Roche,11873580001). Proteins were dialyzed against 50mM Tris pH 7.4, 125 mM NaCl, 10% glycerol and 1mM DTT at 4°C for BRD4-IDR-GFP, OCT4-GFP and GFP alone and the same buffer containing 500 mM NaCl for MED1-IDR-GFP.

Purification of human Mediator complex from HeLa nuclear extract.
HeLa nuclear protein extract (4g) was prepared as described in (Dignam et al., 1983). Nuclear extract was dialyzed against BC100: BC buffer, pH 7.5 + 100mM KCl (20 mM Tris-HCl, 20 mM β-Mercaptoethanol, 0.2 mM PMSF, 0.2 mM EDTA, 10% glycerol (v/v) and 100 mM KCl). The extract was fractionated on a phosphocellulose column (P11) with BC buffer containing 0.1, 0.3, 0.5 and 1M KCl. The Mediator complex eluted in the 0.5M KCL (BC500) fraction. This fraction was dialyzed against BC100 and loaded on a DEAE Cellulose column and sequentially

fractionated with BC buffer containing 0.1, 0.3 and .5M KCl. The Mediator did not bind the DEAE Cellulose resin and was collected in the flow through fraction 0.1M KCl (BC100). This fraction was then directly loaded onto a DEAE-5PW column (TSK) and eluted with a linear KCl gradient from 0.1 to 1M KCl in BC buffer. The Mediator complex eluted between 0.4 and 0.6M KCl. The fractions containing Mediator were pooled and dialyzed against BD700: BD buffer, (20 mM Hepes pH 7.5, 20 mM β-Mercaptoethanol, 0.2 mM PMSF, 0.2 mM EDTA, 10% glycerol, and 700 mM (NH4)2SO4). This fraction was then loaded onto a Phenyl-Sepharose Hydrophobic Interaction Chromatography (HIC) column and eluted with a linear reverse gradient from 0.7 to 0.025M (NH4)2SO4 in BD buffer. The Mediator complex eluted between 0.3 and 0.1M (NH4)2SO4. The Mediator-containing fractions were again pooled and dialyzed against BA100: BA buffer, pH 7.5 + 100 mM NaCl (20 mM Hepes, 20 mM β-Mercaptoethanol, 0.2 mM PMSF, 0.2 mM EDTA, 10% glycerol and 100 mM NaCl) and loaded onto a Heparin Agarose column. The column was washed with BA100 and step-eluted with BA buffer containing 0.25, 0.5, 1M and 1M NaCl. The Mediator complex eluted in the 0.5M NaCl (BA500) fraction. A portion of this fraction was then loaded on a Superose-6 (gel filtration column) that was equilibrated and run in BC100. The Mediator complex eluted from the gel filtration column with a mass range between 1-2MDa.

Reconstituted in vitro transcription assay

The reconstituted in vitro transcription by RNA polymerase II was performed as previously described (Flores et al., 1992; LeRoy et al., 2008, 2019; Orphanides et al., 1998) with some modifications. A 1000 bp template DNA (unlabeled or Cy-3 labeled at 3' end) containing adenovirus major late promoter, five Gal4 binding sites, TATA-box sequence and 561 bp from eGFP sequence was used. First, pre-initiation complex was assembled at RT for 15 min by mixing the following components: 50 nM RNA polymerase II enriched for hypophosphorylated CTD, 50 nM general transcription factors (TFIIA-B-D-E-F-H), and 5.75 nM template DNA, in a buffer containing 10 mM HEPES pH 7.5, 65 mM NaCl, 6.25 mM MgCl2, and 6.25 mM Sodium butyrate. Next, 10 nM Mediator complex and 10 nM GAL4 (Gal4 DNA binding domain fused to activation domain of VP16) were added to the reaction. Last, nucleotide mix containing 0.375 mM ATP, CTP, UTP, GTP (Invitrogen), 0.01 U RNase Inhibitor (Invitrogen), 1.25 % PEG-8000 were added together with one of the following: a) various amounts of purified exogenous *Pou5f1* RNA (0-500 nM) b) spermine (Sigma, S4264) c) extra NTPs (Invitrogen) d) extra NaCl e) heparin (Sigma, H3393). The reaction was incubated at 30°C for 2 hr. RNA isolation was performed using RNeasy kit (Qiagen) by including a spike-in RNA control and an RNA carrier. Purified RNAs were treated with ezDNase (Invitrogen) for 30 min at 37°C to eliminate the template DNA. Reverse transcription was performed using Superscript IV (Invitrogen) and qPCR was performed with SYBR Green Real Time PCR master mix (Invitrogen) to quantify the template derived transcriptional output. The Ct values of the reactions were normalized to the spike-in RNA control. The concentration of template derived transcriptional output was calculated by using a standard curve of qRT-PCR Ct values generated by known amounts of serially diluted GFP RNA. The sequence of primers used for qRT-PCR are indicated in Table S1.

To visualize the droplets formed in the reconstituted transcription assay, using a micropipette (2-20 µL), 5 µL of the reactions were loaded onto a homemade chamber, which was prepared by attaching coverslips to a glass slide by parallel strips of double-sided tape (Sabari et al., 2018). After the droplets were settled on the glass coverslip, the images were collected by using RPI Spinning Disk confocal microscope with a 100x objective. To account for camera artifacts in the images, brightfield Images of droplets from reconstituted assays were subjected to a white tophat filter with a disk element radius of 21 using the MorphoLib plugin in ImageJ, then a Gaussian filter (sigma=1) was applied.

Constructing a free-energy for RNA-protein phase behavior

Our goal in this section is to develop a simplified and coarse-grained model that captures the qualitative physics of RNA-protein mixtures. Based on phenomenological observations of transcriptional proteins and RNA (Figure 2), such a model must recapitulate the following key features:

- Transcriptional proteins phase separate in the absence of RNA through other types of interactions, albeit at higher concentrations.
- At fixed protein concentrations, addition of RNA initially promotes de-mixing and at higher levels drive a re-entry into the mixed phase.

Motivated by the evidence that transcriptional condensates recruit diverse coactivators, transcription factors, and other proteins of the transcriptional apparatus (Boija et al., 2018; Guo et al., 2019; Sabari et al., 2018; Shrinivas et al., 2019), we define an effective protein component P that lumps together different transcriptional molecules. Similarly, while different species of RNA are likely present within these condensates, we define an effective RNA species (R).

Landau model

First, we approach this problem by constructing a phenomenological free-energy with 2 order-parameters that represent scaled concentrations of protein ($\phi_p(\vec{r},t)$) and RNA ($\phi_r(\vec{r},t)$). We define the free-energy (normalized to $k_BT = 1$) as:

$$f[\phi_p,\phi_r] = \int_V d^dV \left(f_{dw}\left(\phi_p(\vec{r},t)\right) + \rho_r\phi_r^2 + \chi_{eff}\left(\phi_p(\vec{r},t),\phi_r(\vec{r},t)\right) + \frac{\kappa}{2}\left(\nabla\phi_p\right)^2\right)$$

Here, $f_{dw}\left(\phi_p(\vec{r},t)\right) = \rho_s\left(\phi_p-\alpha\right)^2\left(\phi_p-\beta\right)^2$ is a standard double-well potential that ensures protein components phase separate without RNA with co-existence concentrations specified by α,β. Choice of $\kappa > 0$ ensures that there is finite surface tension for the protein condensate. The second-order term for RNA ($\rho_r > 0$) states that within this model-framework, RNA cannot phase-separate in the absence of protein. Given that electrostatic interactions at physiological salt conditions are fairly short-ranged (Debye length ~ 1nm), we capture the non-linear nature of RNA-protein interactions in an effective interaction term χ_{eff}. We define this interaction term in the spirit of the Landau-Ginzburg approach as an expansion in powers of the order parameters:

$$\chi_{eff}\left(\phi_p,\phi_r\right) = -\chi\phi_p\phi_r + a\phi_p\phi_r^2 + b\phi_p^2\phi_r + c\phi_p^2\phi_r^2 + \cdots + H.O.T$$

While symmetry arguments often dictate or exclude certain types of terms (odd powers in Ising models for example) in such an expansion, there are no obvious symmetry constraints for this system. Hence, our modeling approach is to minimize the number of higher-order terms that need to be included to recapitulate the experimentally observed reentrant phase transition. Our experimental results suggest that low concentrations of RNA promote phase separation, and thus the lowest order term ($-\chi\phi_p\phi_r, \chi > 0$) lowers the free-energy. However, higher-order terms must counter this and below we outline how we determine which terms to include. In general, the stability of a mixture described by such a free-energy can be ascertained from the Jacobian matrix J. For our model, the elements of this 2×2 matrix are:

$$J_{pp} = \frac{\partial^2 f}{\partial\phi_p^2} = 2\rho_p\left(6\phi_p^2 - 6\phi_p(\beta+\alpha) + (\alpha-\beta)^2\right) + 2b\phi_r + 2c\phi_r^2$$

$$J_{pr} = \frac{\partial^2 f}{\partial\phi_p\partial\phi_r} = -\chi + 2a\phi_r + 2b\phi_p + 4c\phi_p\phi_r$$

$$J_{rr} = \frac{\partial^2 f}{\partial\phi_r^2} = 2\rho_r + 2a\phi_p + 2c\phi_p^2$$

The mixed phase is no longer stable to perturbations when at least one eigen value of J becomes negative (spinodal instability). In the absence of RNA, the spinodal satisfies $J_{pp} = 0$. If only the pair-wise interaction terms were considered ($-\chi\phi_p\phi_r$), the spinodal region broadens i.e. phase separation is promoted at lower protein concentrations when RNA is present. We next characterized the effect of an additional higher-order term (only one of a, b or c is non-zero) on the Jacobian matrix. We ascertained that:

- $a > 0$: While the free-energy is dominated by repulsive interactions at higher RNA concentrations, the Jacobian matrix predicts a continuous underlying instability. Instead of suppressing phase separation at higher RNA concentrations and promoting re-entry to dilute phase, this term would instead change the composition of the demixed phases.
- $b > 0$: While this term promotes a reentrant behavior, the resulting regions of instability demix RNA away from protein for most values of b.
- $c > 0$: For values of c that are not too large (i.e. $c <\approx \rho_r$), the resulting phase diagram mirrors a reentrant shape with RNA enrichment in the protein condensate. If c is moderately large, then a second de-mixing transition (similar to case 2 i.e. b>0) is observed at high values of ϕ_p, ϕ_r. Since we are interested in the limit of relatively low protein/RNA concentrations, and the values of ϕ_p, ϕ_r represent qualitative proxies of protein/RNA concentrations, we choose to explore our model in this parameter regime.

While cubic and higher-order terms are required to recapitulate complete phase-behavior, we explored our model with $c > 0$, assuming the coefficients a, b are small. In the simulations reported in Figures 4-6, the free-energy parameters are $\alpha = 0.1, \beta = 0.7, \chi = 1.0, c = 10.0, \kappa = 0.5, \rho_s = 1.0, \rho_r = 10.0, a = b = 0$. All free-energy calculations were performed with *Python* and code is available at:
https://github.com/krishna-shrinivas/2020_Henninger_Oksuz_Shrinivas_RNA_feedback.

Flory-Huggins model

In this approach, rather than employ a phenomenological model, we parametrize a microscopic model motivated by Flory-Huggins polymer-solution theory (Flory, 1942). The simplified F-H model contains 3 components - protein, RNA, and the solvent (s), whose volume fractions are defined as $\phi_p(\vec{r}, t), \phi_r(\vec{r}, t), 1 - \phi_p(\vec{r}, t) - \phi_r(\vec{r}, t)$ respectively. The free-energy (normalized as before) is defined as:

$$f = \sum_i \frac{\phi_i}{r_i} log(\phi_i) + \sum_{i,j>i} \chi_{ij}\phi_i\phi_j$$

Here, r_i are the solvent-equivalent polymerization lengths of the RNA & protein (assumed to be equal for simplicity) and χ_{ij} are the various pairwise interaction terms. As before, we assume these interactions to be short-ranged at physiological salt levels. Choice of $\chi_{pr} > \chi_{ps} > 0$ and $\chi_{rs} < 0$ recapitulate the attractive contributions of protein-protein/protein-RNA interactions and repulsive RNA-RNA interactions. With these choices of constraints, the resulting free-energy looks similar to the phase diagram from the Landau approach with $c > 0$ (Figure S6A-B) where the key F-H parameters are χ_{pr}=1.1,χ_{ps}=0.75,$\chi_{rs} = -0.6, and\ r_p = r_r = 30$.

Numerical phase-field simulations

Numerical investigations of the coupled-equations outlined in Figure 4C were performed with the FiPy package (Guyer et al., 2009). Simulations were performed on a 2-D/3-D square lattice ($L_x = L_y = 200, dx = 0.3; L_x = L_y = L_z = 40, dx = 1.0$) and with adaptive time-stepping ($dt_{min} = 1e-8, dt_{max} = 5e-1$) until steady state is reached (which typically requires ~ 10000 simulation steps).

The chemical potential for the protein components is calculated as:

$$\mu_p = \frac{df}{d\phi_p} = 2\rho_s(\phi_p - \alpha)(\phi_p - \beta)(2\phi_p - \alpha - \beta) + \kappa\nabla^2\phi_p - \chi\phi_r + 2c\phi_r^2\phi_p$$

The radius of condensates was inferred from the volume of mesh regions where $\phi_p \geq \frac{\alpha+\beta}{2}$. The mobility of RNA and protein were chosen to be 1.0 unless mentioned elsewhere. The raw data for all figures from simulation data are provided along with the manuscript.

Design of Simulations to vary RNA features and rates of RNA synthesis

We designed simulations (Figure 6E) to study the effect of RNA features and rates of effective synthesis on condensate size. The rates of synthesis were changed by increasing k_p by multiplicative factors (see x-axis in Figure 6E). Since RNA length is not explicitly incorporated in the model framework, we defined the effective local synthesis rates of longer RNA as a product of k_p and an additional multiplicative factor (1,2, and 4x for short, medium, and long RNA respectively) to mimic increased local concentrations of RNA.

Calculation of number of charged molecules in condensates

In estimating the number and charge of transcriptional proteins (Figure S1), we use previous estimates (Cho et al., 2018) that suggest key transcriptional proteins such as Mediator are present at 10-100 molecules in transcriptional condensates. Further, molecules such as MED1 or BRD4 contain large disordered domains with net positive charge of +5 to to+40. This provides a highly approximate estimate of 25-500 as the effective positive charge. Since there are many more transcriptional proteins and most proteins tend to contain net positive charges, it is likely that this estimate represents lower bounds on the range. Steady-state levels of nascent eRNA (Figure S1) suggest a range of 0.2-10 molecules, and since super-enhancers typically contain clusters of such active enhancers, we approximate the typical range of eRNA molecules at a transcriptional condensate between 1-10. Since RNA carries a charge of around -1 per nt (Banerjee et al., 2017) and eRNAs are short (<1 kb), we estimate the effective negative charge during initiation to be in the range 10-1000. During productive elongation, mRNAs are produced in bursts ranging from few to tens (1-50) and are typically longer (>1kb), suggesting a conservative estimate of the effective charge to range from (1000-100,000). It is important to stress that our approximations are performed with the aim of obtaining order-of-magnitude estimates and do not account for factors such as local composition of different proteins or extent to which nascent mRNAs may be coated by RNA-binding proteins. With the above numbers, we estimate concentrations based on a typical transcriptional condensate of size r=0.25 μm (Cho et al., 2018) that suggests that eRNA concentrations range about 10-200 nM and transcriptional proteins range 1-20 μM within the condensate.

Reactive/diffusive time-scales and estimates in cells

As defined in the model (Figure 4B), the key rates of synthesis/degradation reactions are kp/kd, which have units of s^{-1}, and thus the relevant time-scales are $t_r = {k_p}^{-1} (or\ {k_d}^{-1})$. Timescales of RNA transport depend on both the diffusivity as well as the size of the condensate (L) and is defined as $t_d = L^2/M_{rna}$. We approximated the range of diffusivity of the nascent transcript at the lower end by the diffusivity of of chromatin, which ranges from $10^{-3.5} - 10^{-2}\ \mu m^2/s$ (Gu et al., 2018) and on the higher end by those of freely diffusing mRNPs, which can be upto $5 \times 10^{-2}\ \mu m^2/s$ (Niewidok et al., 2018). By assuming a typical eRNA of size 100nt and Pol II transcription rates as $\sim 20 - 70\,\frac{nt}{s}$ (Maiuri et al., 2011) we inferred typical synthesis rates of $\sim 0.5\ eRNA$ s^{-1}Pol II^{-1}. In our previous work (Cho et al., 2018), we have seen that clusters that contain multiple polymerases (>5), are typically around $r \approx 200 - 400nm$. Since super-enhancers typically contain clusters of enhancers with multiple sites of eRNA synthesis (~5),

this gives an effective synthesis rate of $k_p \approx 2.5$ s^{-1}Pol II^{-1}. This allows us to approximately obtain the ratio of diffusive and reactive time-scales as $\frac{t_d}{t_r} = \frac{kr^2}{M} \approx 2 - 1000$ over the range of parameters including diffusivity and radii of cluster.

Calculation of charge balance

Charge-balance calculations were performed (Figures 2 S2, S3 and S4) employing the following method. Net protein charge per molecule was calculated as $C_p = \#(R,K) - \#(D,E)$ for the relevant sequence including the GFP tag. RNA charge per molecule was calculated as $C_r = -(\#\ of\ bp)$, assuming an approximate charge of -1 per nucleotide (Lin et al., 2019). Next, the charge balance ratio was computed at a particular RNA and protein concentration as:

$$Charge - balance\ ratio = \frac{min\left(C_p[P], C_r[R]\right)}{max(C_p[P], C_r[R])}$$

The effective concentration of MED1-IDR in our assays was $1000\ nM$. Our results were not quantitatively affected by inclusion/exclusion of the partial charge on Histidine residues, partly due to their low frequency on the protein sequences. For Heparin, a charge of roughly -3 per monomer was employed (Lin et al., 2020) and for single-stranded DNA, a charge of -1 per nt was employed. A comprehensive listing of charges of various species employed in this study are provided in Table S2. The Pearson correlation coefficient (r) was calculated between the median droplet partition value at different concentrations and the relevant charge-balance ratios and reported in Figures 2, S2, S3, and S4. A higher correlation implies that experimental data follow a similar qualitative trend as the estimated charge-balance curves. The code for performing these calculations are available at:
https://github.com/krishna-shrinivas/2020_Henninger_Oksuz_Shrinivas_RNA_feedback.

Transcription inhibition by small molecules

For small molecule inhibition experiments, cells were treated with 100 μM DRB (Sigma), or 1 μM Actinomycin-D (Sigma) in 2i media (detailed above) for 30 minutes, then imaged. For wash-out experiments, media was replaced with fresh 2i media and cells were allowed to recover for 1 hour, then the cells were imaged.

Condensate size

Cells with endogenously-tagged Med1-GFP (Sabari et al., 2018) were plated on glass-bottom dishes (Mattek) coated with poly-L-ornithine (Sigma) and laminin (ThermoFisher). Mock (DMSO) and treated cells were imaged on a LSM 880 Confocal Microscope with Airyscan to obtain super-resolution z-stacks for at least 8 different fields containing multiple cells. For quantification, a manual threshold was applied equally across all conditions to remove background, and the size of Med1-GFP puncta was quantified in 3D using the 3D object counter plugin (Fiji/ImageJ).

Condensate lifetime

HaloTag was endogenously knocked into 5'-end of Med19 via homology-directed repair (HDR) in mouse embryonic stem cells (R1 mESCs). Three single-guide RNAs (sgRNAs) targeting +/- 100 bps from the start codons of Med19 gene were designed using the web-based CRISPR Design tool (http://crispr.mit.edu) and integrated into a *Streptococcus pyogenes* Cas9 vector (Addgene #62988) for standard CRISPR/Cas9 editing. Single positive colonies were sorted by fluorescence-activated cell sorting (FACS) and validated under the microscope.

Cells were cultured in serum-free 2i medium on poly-L-ornithine (PLO) and Laminin-coated flasks for more than two days and then were transferred onto coated imaging dishes for another day. Before imaging, cells were stained with (PA)-JF549-HaloTAG dye (a gift from Luke Lavis Lab, Janelia Research Campus) of 100nM concentration for 2 hours followed by a 60-minute wash in fresh 2i medium. Lastly, dishes were filled in with 2ml Leibovitz's L-15 Medium (no phenol red, Thermo Fisher) and brought to the microscope for imaging.

Photo-activation localization microscopy (PALM) imaging was performed using a Nikon Eclipse Ti microscope with a 100x oil immersion objective (NA 1.40) (Nikon, Tokyo, Japan). A 405nm beam of 100mW power (attenuated with 25% AOTF) and a 561nm beam of 500mW power were columnated and superposed to perform simultaneous activation and excitation. The combined beam was expanded and re-collimated with an achromatic beam expander (AC254-040-A and AC508-300-A, THORLABS) to improve the uniformity of illumination across the whole region of interest (ROI 256^2 pixels). Images were acquired with an Andor iXon Ultra 897 EMCCD camera (gain 1000, exposure time 50ms) interfaced through Micro Manager 1.4. 2400 frames were acquired for each imaging cycle. The cells were maintained at 37°C in a temperature-controlled platform (InVivo Scientific, St. Louis, MO) on the microscope stage during image acquisition. Med19-Halo cluster lifetimes were calculated as previously described using the qSR software (dark time tolerance = 20 frames, min cluster size=50) (Andrews et al., 2018), and a cumulative distribution was generated using Prism software (GraphPad).

Nascent RNA imaging

For the nascent RNA experiments in Figure S7A-7C, $1.25x10^5$ wildtype mESCs were plated on coverslips coated with poly-L-ornithine (Sigma) and Laminin (ThermoFisher). After overnight plating, nascent RNA labeling with 2.5 mM EU was done with the Click-iT™ RNA Alexa Fluor™ 594 Imaging Kit (Thermofisher) according to manufacturer instructions for 10 minutes. After incubation, cells were immediately fixed with 4% paraformaldehyde for 10 minutes, washed 3X with PBS, then permeabilized with 0.5% TritonX-100 in PBS for 15 minutes. After the Click-iT reaction, coverslips were blocked with 4% RNase-free BSA in PBS for 10 minutes at room temperature. Coverslips were incubated with primary antibodies (1:500; rabbit Abcam ab64965 for MED1 and rat Millipore Sigma 04-1571 for Pol II-S2) in 4%BSA/PBS at room temperature overnight. The next day, coverslips were washed 3X with PBS, then incubated in secondary antibody (1:500; goat anti-rabbit AlexaFluor-488 Thermofisher A11008, goat anti-rat AlexaFluor-647 Invitrogen A21247) for 1 hour at room temperature. After washing 3X with PBS, coverslips were stained with 1:1000 Hoechst 33342 in PBS, incubated for 15 minutes at room temperature, washed 3X with PBS, and mounted on imaging slides with Vectashield Mounting Media. Images were collected on the RPI Spinning Disk confocal. Representative images in Figure S7A are single z-planes of median-subtracted (px=10) and Gaussian smoothed (sigma=1) channels to correct for uneven illumination and background.

For analysis of these images, nuclei were segmented using the Cellpose algorithm (Stringer et al., 2020) on the 405 Hoechst channel images. For average image analysis in Figure S7B, all channel images were maximally projected, subtracted by median filter (px=10), and Gaussian smoothed (sigma=1). The center of MED1 and Pol II-S2 puncta were segmented as follows. The Laplace of Gaussian transformation (sigma=3) was applied to the images using the scikit-image package in python, and puncta were identified above a threshold intensity 3 standard deviations above the mean of the image. All spots were confirmed to be in nuclei. A 1μm by 1μm box was centered on the spots, and the box subimage was collected for that region in both the processed MED1 and Pol II-S2 channel images. These subimages from >10 imaged fields were stacked and averaged, which was the input for the contour plots in Figure S7B. Radial intensity plots in Figure S7C show the distribution of these averaged signals as a

function of the distance from the center of the spot, along with their correlation to EU RNA signal..

Reporter assay to determine the effect of local RNA synthesis on transcription
Vectors used in the reporter assay are modified from pTETRIS-cargo vector, gift from J. M. Calabrese (Kirk et al., 2018). 6X STOP codon sequence was cloned into NotI digested pTETRIS-cargo vector using Gibson cloning strategy by following the manufacturer's instructions (NEB). This vector is called pTETRIS-cargo-STOP. The feedback gene and the reporter gene have their own polyA termination signal (200-300 bp) to terminate transcription. There is 51 bp between these two polyA signals that are facing each other. The reporter gene is regulated by a phosphoglycerate kinase (PGK) promoter. Various versions of the pTETRIS-cargo-STOP using Gibson cloning strategy (NEB): i) the relative orientations of the feedback RNA and luciferase reporter were altered (tandem or divergent orientations) ii) feedback RNAs and luciferase reporter were cloned into separate vectors. Using Gibson cloning strategy (NEB), various RNA sequences were cloned downstream of the 6X STOP sequence to prevent translation of these RNAs. Stable cell lines for individual RNAs were generated by transfecting Med1-GFP mESCs with the following vectors: 1.0 µg pTETRIS-cargo-STOP containing individual RNAs, 1.0 µg rTTA-cargo, gift from J. M. Calabrese (Kirk et al., 2018), and 1 µg piggyBAC transposase (Systems Biosciences). Cells were selected on puromycin (2 µg/ml) and G418 (200 µg/ml) for 1 week for successful integrations. For luciferase assays, $1x10^5$ cells of each genotype were plated in triplicate on 0.2%-gelatin-coated 24-well plates and allowed to settle overnight. Cells were treated with doxycycline (Sigma) and harvested after 24 h to measure either luciferase activity or to purify RNA. Luciferase activity was measured using the Luciferase Assay System (Promega) according to manufacturer instructions. Luciferase signal was normalized to total protein content, measured by BCA protein assay kit (Invitrogen, #23227), and then normalized to a control not treated with doxycycline. To measure RNA expression, RNA was purified using the Qiagen RNeasy Mini kit (Qiagen) according to manufacturer instructions, cDNA was generated by Superscript III (Invitrogen) according to manufacturer instructions, and 10 ng of cDNA was used in a qRT-PCR SYBR-green reaction (Life Technologies) with primers specific to a common sequence shared across the vectors (qPCR_Tetris, Table S1). Ct values were normalized to a housekeeping gene (qPCR_mActb, Table S1) and a control condition with no doxycycline treatment.

For the washout experiments in Figure S7J, reporter cells were plated as described above. After 24 hours of dox treatment, media was replaced with fresh media, whereas control cell media was replaced with dox-containing media. After an additional 24 hours, luciferase levels were measured as described above. For the antisense oligo experiments of Figure S7J, antisense oligos (LNA gapmers, Qiagen) were designed using the Qiagen GeneGlobe tool against the feedback RNA. A negative scrambled control was also included. Reporter cells were plated as described above in triplicate, and cells were transfected with 25 nM ASO with Lipofectamine-3000 (and no P3000 enhancer agent). After overnight transfection, cell media was replaced with dox-containing or fresh 2i media as a control. After 24 hour dox treatment, RNA and luciferase levels were quantified by qRT-PCR and luminescence, respectively, as described above. For the analysis of luciferase rescue, luminescence values of the dox conditions were first normalized to the no dox condition for that ASO, and then normalized to the dox condition of the negative scrambled control.

For imaging experiments in Figures 6C-6D, the reporter construct was modified using Gibson cloning to include a 24X-MS2 hairpin (Cho et al., 2018) at the 5’ end of the RNA sequence (2,456 nt total). Cell lines with this construct and double MS2 capsid protein fused to an mCherry tag (2xMCP-mCherry) were generated as detailed above in a mESC background with

endogenously-tagged Med1-GFP (Sabari et al., 2018). $1x10^6$ reporter cells were plated on glass-bottom dishes (Mattek) coated with poly-L-ornithine (Sigma) and Laminin (ThermoFisher) After overnight plating, cells were treated with 10, 100, or 1000 ng/mL doxycycline for 24 hours. Cells were imaged on an RPI Spinning Disk Confocal with the following laser powers and exposure times: 488 70% 500 ms, 561 40% 300 ms. Images were maximum projected, median subtracted (px=10), and Gaussian filtered (sigma=1) to correct for uneven illumination and background subtraction. For analysis of these images in Figure 6D, nuclei were segmented using the Cellpose algorithm (Stringer et al., 2020) on images from the 561 channel that had been subjected to a maximum and median filter (px=10). For average image analysis, both the RNA and MED1-GFP channel images were maximally projected, subtracted by median filter (px=10), and Gaussian smoothed (sigma=1). The centers of RNA spots in a maximum projection of the 561 channel were manually marked using ImageJ. All spots were confirmed to be in nuclei. A 1μm by 1μm box was centered on the RNA spot, and the box subimage was collected for that region in both the processed RNA and MED1-GFP channel images. These subimages from >10 imaged fields were stacked and averaged, which was the input for the contour plots in Figure 6D. Radial intensity plots in Figure S7D show the distribution of these averaged signals as a function of the distance from the center of the spot. To control for global Dox effects, we quantified the size, number, and partition ratio of MED1-GFP condensates in all conditions by using a threshold of 3 standard deviations above the mean intensity of the image to segment condensates. Partition ratio for each condensate was calculated as the average intensity inside the condensate divided by the average intensity of the nucleoplasm.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis of in vitro condensate assays:
The Pearson correlation coefficient (r) was calculated between the median droplet partition value at different concentrations and the relevant charge-balance ratios (see Calculation of charge balance under Method Details. The computed values are reported in Figures 2, S2, S3, and S4. In Figure S2A, the correlation coefficient was computed between the median droplet turbidity at different conditions and the relevant charge-balance ratios. A higher correlation implies that experimental data follow a similar qualitative trend as the estimated charge-balance curves. The code for performing these calculations are available at: https://github.com/krishna-shrinivas/2020_Henninger_Oksuz_Shrinivas_RNA_feedback.

Statistical analysis of in vitro transcription assays:
The Student's t-test was used to determined whether the addition of spermine to the in vitro transcription assays led to statistically different outcomes in the mean of droplet size, partition ratio, and normalized transcription and the corresponding p-values are reported in the figure legends (Figure 3D-F). In the normalized transcription assay (Figure 3F), The values are normalized to the mean of the no spermine condition. The one-way ANOVA test is used to determine whether addition of exogenous RNA or absence of all NTPs leads to statistically different outcomes in droplet area or normalized transcription and the corresponding pairwise p-values (compared to control conditions i.e. no addition of exogeneous RNA) are reported in the figure legends (Figure 3H-I). These tests were performed using PRISM.

Statistical analysis of transcription inhibition experiments:
PRISM was used to compute a one-way ANOVA comparison to test whether the mean of the condensate volumes (see Method Details) was statistically different upon inhibition by DRB or ACT-D versus control (DMSO) and the corresponding p-value is reported in the Figure legend (Figure 5E). Similarly, a one-way ANOVA comparison was used to test whether the mean of the

condensate lifetimes (see Method Details) were statistically different from control conditions (DMSO) and the corresponding p-values are reported in the Figure Legend (Figure 5I).

Statistical analysis of luciferase reporter experiments:
A 2-way Kolmongorov-Smirnoff test was used to determine whether the cumulative distribution functions of the average MED1 intensity centered at RNA (see Method Details) arising from replicates (values of n are represented in Figure 6D) at different Dox concentrations. The p-values are reported in the Figure legend (Figure 6D). Markers show the mean of at least 3 replicates and error bars depict the S.D. in Figures 6F-G and supplementary S7. The Pearson correlation coefficient between the radial intensity distributions, computed from the averaged signal analyses (see S7B, Method Details) was computed and this value is reported in S7C.

No methods were used to determine whether the data met assumptions of the statistical approach.

References

1. Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat Rev Genet *13*, 720–731.
2. Adhikari, S., Leaf, M.A., and Muthukumar, M. (2018). Polyelectrolyte complex coacervation by electrostatic dipolar interactions. J Chem Phys *149*, 163308.
3. Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al. (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455–461.
4. Andrews, J.O., Conway, W., Cho, W.-K., Narayanan, A., Spille, J.-H., Jayanth, N., Inoue, T., Mullen, S., Thaler, J., and Cissé, I.I. (2018). qSR: a quantitative super-resolution analysis tool reveals the cell-cycle dependent organization of RNA Polymerase I in live human cells. Scientific Reports *8*, 1–10.
5. Aumiller, W.M., Pir Cakmak, F., Davis, B.W., and Keating, C.D. (2016). RNA-Based Coacervates as a Model for Membraneless Organelles: Formation, Properties, and Interfacial Liposome Assembly. Langmuir *32*, 10042–10053.
6. Azofeifa, J.G., Allen, M.A., Hendrix, J.R., Read, T., Rubin, J.D., and Dowell, R.D. (2018). Enhancer RNA profiling predicts transcription factor activity. Genome Res *28*, 334–344.
7. Banani, S.F., Lee, H.O., Hyman, A.A., and Rosen, M.K. (2017). Biomolecular condensates: Organizers of cellular biochemistry. Nature Reviews Molecular Cell Biology *18*, 285–298.
8. Banerjee, P.R., Milin, A.N., Moosa, M.M., Onuchic, P.L., and Deniz, A.A. (2017). Reentrant Phase Transition Drives Dynamic Substructure Formation in Ribonucleoprotein Droplets. Angewandte Chemie International Edition *56*, 11354–11359.
9. Bergot, M.O., Diaz-Guerra, M.J., Puzenat, N., Raymondjean, M., and Kahn, A. (1992). Cis-regulation of the L-type pyruvate kinase gene promoter by glucose, insulin and cyclic AMP. Nucleic Acids Res *20*, 1871–1877.
10. Blair, D.G.R. (1985). Activation of mammalian RNA polymerases by polyamines. International Journal of Biochemistry *17*, 23–30.
11. Boeynaems, S., Holehouse, A.S., Weinhardt, V., Kovacs, D., Lindt, J.V., Larabell, C., Bosch, L.V.D., Das, R., Tompa, P.S., Pappu, R.V., et al. (2019). Spontaneous driving forces give rise to protein−RNA condensates with coexisting phases and complex material properties. PNAS *116*, 7889–7898.
12. Boija, A., Klein, I.A., Sabari, B.R., Dall'Agnese, A., Coffey, E.L., Zamudio, A.V., Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M., et al. (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell *175*, 1842-1855 e16.
13. Brandman, O., and Meyer, T. (2008). Feedback Loops Shape Cellular Signals in Space and Time. Science *322*, 390–395.
14. Bruhat, A., Jousse, C., Carraro, V., Reimold, A.M., Ferrara, M., and Fafournoux, P. (2000). Amino Acids Control Mammalian Gene Transcription: Activating Transcription Factor 2 Is Essential for the Amino Acid Responsiveness of the CHOP Promoter. Mol Cell Biol *20*, 7192–7204.
15. Cambridge, S.B., Gnad, F., Nguyen, C., Bermejo, J.L., Krüger, M., and Mann, M. (2011). Systems-wide Proteomic Analysis in Mammalian Cells Reveals Conserved, Functional Protein Turnover. J. Proteome Res. *10*, 5275–5284.
16. Carey, M.F., Peterson, C.L., and Smale, S.T. (2009). Transcriptional Regulation in Eukaryotes: Concepts, Strategies, and Techniques, Second Edition.
17. Catarino, R.R., and Stark, A. (2018). Assessing sufficiency and necessity of enhancer activities for gene expression and the mechanisms of transcription activation. Genes Dev. *32*, 202–223.
18. Chen, H., and Larson, D.R. (2016). What have single-molecule studies taught us about gene expression? Genes Dev. *30*, 1796–1810.
19. Chen, W., Smeekens, J.M., and Wu, R. (2016). Systematic study of the dynamics and half-lives of newly synthesized proteins in human cells. Chem. Sci. *7*, 1393–1400.
20. Chiu, A.C., Suzuki, H.I., Wu, X., Mahat, D.B., Kriz, A.J., and Sharp, P.A. (2018). Transcriptional Pause Sites Delineate Stable Nucleosome-Associated Premature Polyadenylation Suppressed by U1 snRNP. Molecular Cell *69*, 648-663.e7.

21. Cho, W.-K., Jayanth, N., English, B.P., Inoue, T., Andrews, J.O., Conway, W., Grimm, J.B., Spille, J.-H., Lavis, L.D., Lionnet, T., et al. (2016). RNA Polymerase II cluster dynamics predict mRNA output in living cells. ELife *5*, e13617.
22. Cho, W.-K.K., Spille, J.-H.H., Hecht, M., Lee, C., Li, C., Grube, V., Cisse, I.I., Lee, C., Hecht, M., Cho, W.-K.K., et al. (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science *361*, 412–415.
23. Chubb, J.R., Trcek, T., Shenoy, S.M., and Singer, R.H. (2006). Transcriptional Pulsing of a Developmental Gene. Current Biology *16*, 1018–1025.
24. Cisse, I.I., Izeddin, I., Causse, S.Z., Boudarene, L., Senecal, A., Muresan, L., Dugast-Darzacq, C., Hajj, B., Dahan, M., and Darzacq, X. (2013). Real-Time Dynamics of RNA Polymerase II Clustering in Live Human Cells. Science *341*, 664–667.
25. Core, L., and Adelman, K. (2019). Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. Genes Dev. *33*, 960–982.
26. Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A., and Lis, J.T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nat Genet *46*, 1311–1320.
27. Cramer, P. (2019). Organization and regulation of gene transcription. Nature *573*, 45–54.
28. Delaney, K.T., and Fredrickson, G.H. (2017). Theory of polyelectrolyte complexation-Complex coacervates are self-coacervates. J Chem Phys *146*, 224902.
29. Di Tommaso, P., Chatzou, M., Floden, E.W., Barja, P.P., Palumbo, E., and Notredame, C. (2017). Nextflow enables reproducible computational workflows. Nature Biotechnology *35*, 316–319.
30. Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res *11*, 1475–1489.
31. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.
32. Drobot, B., Iglesias-Artola, J.M., Le Vay, K., Mayr, V., Kar, M., Kreysing, M., Mutschler, H., and Tang, T.-Y.D. (2018). Compartmentalised RNA catalysis in membrane-free coacervate protocells. Nat Commun *9*, 3643.
33. Dunlap, J.C. (1999). Molecular Bases for Circadian Clocks. Cell *96*, 271–290.
34. Ebert, B.L., and Bunn, H.F. (1999). Regulation of the erythropoietin gene. Blood *94*, 1864–1877.
35. Elowitz, M.B., and Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. Nature *403*, 335–338.
36. Ewels, P.A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M.U., Di Tommaso, P., and Nahnsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines. Nature Biotechnology *38*, 276–278.
37. Fay, M.M., and Anderson, P.J. (2018). The role of RNA in biological phase separations. J Mol Biol *430*, 4685–4701.
38. Flores, O., Lu, H., and Reinberg, D. (1992). Factors involved in specific transcription by mammalian RNA polymerase II. Identification and characterization of factor IIH. J. Biol. Chem. *267*, 2786–2793.
39. Flory, P.J. (1942). Thermodynamics of High Polymer Solutions. J. Chem. Phys. *10*, 51–61.
40. Frankel, E.A., Bevilacqua, P.C., and Keating, C.D. (2016). Polyamine/Nucleotide Coacervates Provide Strong Compartmentalization of Mg2+, Nucleotides, and RNA. Langmuir *32*, 2041–2049.
41. Fukaya, T., Lim, B., and Levine, M. (2016). Enhancer Control of Transcriptional Bursting. Cell *166*, 358–368.
42. Gardini, A., and Shiekhattar, R. (2015). The many faces of long noncoding RNAs. The FEBS Journal *282*, 1647–1657.
43. Gardner, T.S., Cantor, C.R., and Collins, J.J. (2000). Construction of a genetic toggle switch in Escherichia coli. Nature *403*, 339–342.
44. Gu, B., Swigut, T., Spencley, A., Bauer, M.R., Chung, M., Meyer, T., and Wysocka, J. (2018). Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. Science *359*, 1050–1055.

45. Guo, Y.E., Manteiga, J.C., Henninger, J.E., Sabari, B.R., Dall'Agnese, A., Hannett, N.M., Spille, J.-H., Afeyan, L.K., Zamudio, A.V., Shrinivas, K., et al. (2019). Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. Nature *572*, 543–548.
46. Guyer, J.E., Wheeler, D., and Warren, J.A. (2009). FiPy: Partial Differential Equations with Python. Comput. Sci. Eng. *11*, 6–15.
47. Henriques, T., Scruggs, B.S., Inouye, M.O., Muse, G.W., Williams, L.H., Burkholder, A.B., Lavender, C.A., Fargo, D.C., and Adelman, K. (2018). Widespread transcriptional pausing and elongation control at enhancers. Genes Dev. *32*, 26–41.
48. Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A Phase Separation Model for Transcriptional Control. Cell *169*, 13–23.
49. Hohenberg, P.C., and Halperin, B.I. (1977). Theory of dynamic critical phenomena. Rev. Mod. Phys. *49*, 435–479.
50. http://crispr.mit.edu Guide Design Resources.
51. Jangi, M., and Sharp, P.A. (2014). Building Robust Transcriptomes with Master Splicing Factors. Cell *159*, 487–498.
52. Jin, Y., Eser, U., Struhl, K., and Churchman, L.S. (2017). The Ground State and Evolution of Promoter Region Directionality. Cell *170*, 889-898.e10.
53. Kardar, M. (2007). Statistical physics of fields (Cambridge: Cambridge Univ. Press).
54. Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., and Kent, W.J. (2004). The UCSC Table Browser data retrieval tool. Nucleic Acids Res. *32*, D493-496.
55. Kim, T.-K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182–187.
56. Kirk, J.M., Kim, S.O., Inoue, K., Smola, M.J., Lee, D.M., Schertzer, M.D., Wooten, J.S., Baker, A.R., Sprague, D., Collins, D.W., et al. (2018). Functional classification of long non-coding RNAs by k-mer content. Nature Genetics *50*, 1474–1482.
57. Klein, I.A., Boija, A., Afeyan, L.K., Hawken, S.W., Fan, M., Dall'Agnese, A., Oksuz, O., Henninger, J.E., Shrinivas, K., Sabari, B.R., et al. (2020). Partitioning of cancer therapeutics in nuclear condensates. Science *368*, 1386–1392.
58. Lahav, G., Rosenfeld, N., Sigal, A., Geva-Zatorsky, N., Levine, A.J., Elowitz, M.B., and Alon, U. (2004). Dynamics of the p53-Mdm2 feedback loop in individual cells. Nat Genet *36*, 147–150.
59. Lai, F., Orom, U.A., Cesaroni, M., Beringer, M., Taatjes, D.J., Blobel, G.A., and Shiekhattar, R. (2013). Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature *494*, 497–501.
60. Landau, L.D. (1937). ON THE THEORY OF PHASE TRANSITIONS. Zh. Eksp. Teor. Fiz *11*, 19.
61. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods 9, 357–359.

62. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology *10*, R25.
63. Larsson, A.J.M., Johnsson, P., Hagemann-Jensen, M., Hartmanis, L., Faridani, O.R., Reinius, B., Segerstolpe, Å., Rivera, C.M., Ren, B., and Sandberg, R. (2019). Genomic encoding of transcriptional burst kinetics. Nature *565*, 251–254.
64. Leinonen, R., Sugawara, H., and Shumway, M. (2011). The Sequence Read Archive. Nucleic Acids Res *39*, D19–D21.
65. LeRoy, G., Rickards, B., and Flint, S.J. (2008). The Double Bromodomain Proteins Brd2 and Brd3 Couple Histone Acetylation to Transcription. Molecular Cell.
66. LeRoy, G., Oksuz, O., Descostes, N., Aoi, Y., Ganai, R.A., Kara, H.O., Yu, J.-R., Lee, C.-H., Stafford, J., Shilatifard, A., et al. (2019). LEDGF and HDGF2 relieve the nucleosome-induced barrier to transcription in differentiated cells. Sci Adv *5*, eaay3068.
67. Li, C.H., Coffey, E.L., Dall'Agnese, A., Hannett, N.M., Tang, X., Henninger, J.E., Platt, J.M., Oksuz, O., Zamudio, A.V., Afeyan, L.K., et al. (2020). MeCP2 links heterochromatin condensates and neurodevelopmental disease. Nature.
68. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.

69. Li, W., Notani, D., and Rosenfeld, M.G. (2016). Enhancers as non-coding RNA transcription units: recent insights and future perspectives. Nat Rev Genet *17*, 207–223.
70. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.
71. Lin, Y., McCarty, J., Rauch, J.N., Delaney, K.T., Kosik, K.S., Fredrickson, G.H., Shea, J.-E., and Han, S. (2019). Narrow equilibrium window for complex coacervation of tau and RNA under cellular conditions. ELife *8*, e42571.
72. Lin, Y., Fichou, Y., Zeng, Z., Hu, N.Y., and Han, S. (2020). Electrostatically Driven Complex Coacervation and Amyloid Aggregation of Tau Are Independent Processes with Overlapping Conditions. ACS Chem Neurosci *11*, 615–627.
73. Maharana, S., Wang, J., Papadopoulos, D.K., Richter, D., Pozniakovsky, A., Poser, I., Bickle, M., Rizk, S., Guillén-Boixet, J., Franzmann, T.M., et al. (2018). RNA buffers the phase separation behavior of prion-like RNA binding proteins. Science *360*, 918–921.
74. Maiuri, P., Knezevich, A., De Marco, A., Mazza, D., Kula, A., McNally, J.G., and Marcello, A. (2011). Fast transcription rates of RNA polymerase II in human cells. EMBO Rep *12*, 1280–1285.
75. Mikhaylichenko, O., Bondarenko, V., Harnett, D., Schor, I.E., Males, M., Viales, R.R., and Furlong, E.E.M. (2018). The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. Genes Dev *32*, 42–57.
76. Milin, A.N., and Deniz, A.A. (2018). Reentrant Phase Transitions and Non-Equilibrium Dynamics in Membraneless Organelles. Biochemistry *57*, 2470–2477.
77. Monod, J., and Jacob, F. (1961). General Conclusions: Teleonomic Mechanisms in Cellular Metabolism, Growth, and Differentiation. Cold Spring Harb Symp Quant Biol *26*, 389–401.
78. Moruzzi, G., Barbiroli, B., Moruzzi, M.S., and Tadolini, B. (1975). The effect of spermine on transcription of mammalian chromatin by mammalian deoxyribonucleic acid-dependent ribonucleic acid polymerase. Biochem J *146*, 697–703.
79. Mountain, G.A., and Keating, C.D. (2020). Formation of Multiphase Complex Coacervates and Partitioning of Biomolecules within them. Biomacromolecules *21*, 630–640.
80. Muthukumar, M. (2016). Electrostatic correlations in polyelectrolyte solutions. Polym. Sci. Ser. A *58*, 852–863.
81. Nair, S.J., Yang, L., Meluzzi, D., Oh, S., Yang, F., Friedman, M.J., Wang, S., Suter, T., Alshareedah, I., Gamliel, A., et al. (2019). Phase separation of ligand-activated enhancers licenses cooperative chromosomal enhancer assembly. Nature Structural & Molecular Biology *26*, 193–203.
82. Niewidok, B., Igaev, M., Pereira da Graca, A., Strassner, A., Lenzen, C., Richter, C.P., Piehler, J., Kurre, R., and Brandt, R. (2018). Single-molecule imaging reveals dynamic biphasic partition of RNA-binding proteins in stress granules. J. Cell Biol. *217*, 1303–1318.
83. Nott, T.J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T.D., Bazett-Jones, D.P., Pawson, T., Forman-Kay, J.D., et al. (2015). Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. Molecular Cell *57*, 936–947.
84. Orphanides, G., LeRoy, G., Chang, C.-H., Luse, D.S., and Reinberg, D. (1998). FACT, a Factor that Facilitates Transcript Elongation through Nucleosomes. Cell *92*, 105–116.
85. Overbeek, J.T.G., and Voorn, M.J. (1957). Phase separation in polyelectrolyte solutions. Theory of complex coacervation. Journal of Cellular and Comparative Physiology *49*, 7–26.
86. Pai, D.A., Kaplan, C.D., Kweon, H.K., Murakami, K., Andrews, P.C., and Engelke, D.R. (2014). RNAs nonspecifically inhibit RNA polymerase II by preventing binding to the DNA template. RNA *20*, 644–655.
87. Pefanis, E., Wang, J., Rothschild, G., Lim, J., Kazadi, D., Sun, J., Federation, A., Chao, J., Elliott, O., Liu, Z.-P., et al. (2015). RNA Exosome-Regulated Long Non-Coding RNA Transcription Controls Super-Enhancer Activity. Cell *161*, 774–789.
88. Peran, I., and Mittag, T. (2020). Molecular structure in biomolecular condensates. Current Opinion in Structural Biology *60*, 17–26.
89. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

90. Rahnamoun, H., Lee, J., Sun, Z., Lu, H., Ramsey, K.M., Komives, E.A., and Lauberth, S.M. (2018). RNAs interact with BRD4 to promote enhanced chromatin engagement and transcription activation. Nature Structural & Molecular Biology *25*, 687–697.
91. Raj, A., and van Oudenaarden, A. (2008). Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. Cell *135*, 216–226.
92. Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA Synthesis in Mammalian Cells. PLoS Biol *4*.
93. Reinberg, D., and Roeder, R.G. (1987). Factors involved in specific transcription by mammalian RNA polymerase II. Transcription factor IIS stimulates elongation of RNA chains. J. Biol. Chem. *262*, 3331–3337.
94. Roden, C., and Gladfelter, A.S. (2020). RNA contributions to the form and function of biomolecular condensates. Nature Reviews Molecular Cell Biology 1–13.
95. Rodriguez, J., and Larson, D.R. (2020). Transcription in Living Cells: Molecular Mechanisms of Bursting. Annual Review of Biochemistry *89*, null.
96. Roeder, R.G. (2019). 50+ years of eukaryotic transcription: an expanding universe of factors and mechanisms. Nature Structural & Molecular Biology *26*, 783–791.
97. Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A.V., Manteiga, J.C., et al. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. Science *361*, eaar3958.
98. Sabari, B.R., Dall'Agnese, A., and Young, R.A. (2020). Biomolecular Condensates in the Nucleus. Trends in Biochemical Sciences *0*.
99. Schaukowitch, K., Joo, J.-Y., Liu, X., Watts, J.K., Martinez, C., and Kim, T.-K. (2014). Enhancer RNA Facilitates NELF Release from Immediate Early Genes. Molecular Cell *56*, 29–42.
100. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nature Methods 9, 676–682.
101. Scruggs, B.S., Gilchrist, D.A., Nechaev, S., Muse, G.W., Burkholder, A., Fargo, D.C., and Adelman, K. (2015). Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin. Molecular Cell *58*, 1101–1112.
102. Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent Transcription from Active Promoters. Science *322*, 1849–1851.
103. Sellick, C.A., and Reece, R.J. (2003). Modulation of transcription factor function by an amino acid: activation of Put3p by proline. EMBO J *22*, 5147–5153.
104. Shin, Y., and Brangwynne, C.P. (2017). Liquid phase condensation in cell physiology and disease. Science *357*, eaaf4382–eaaf4382.
105. Shrinivas, K., Sabari, B.R., Coffey, E.L., Klein, I.A., Boija, A., Zamudio, A.V., Schuijers, J., Hannett, N.M., Sharp, P.A., Young, R.A., et al. (2019). Enhancer Features that Drive Formation of Transcriptional Condensates. Molecular Cell *75*, 549-561.e7.
106. Sigova, A.A., Abraham, B.J., Ji, X., Molinie, B., Hannett, N.M., Guo, Y.E., Jangi, M., Giallourakis, C.C., Sharp, P.A., and Young, R.A. (2015). Transcription factor trapping by RNA in gene regulatory elements. Science *350*, 978–981.
107. Sing, C.E. (2017). Development of the modern theory of polymeric complex coacervation. Advances in Colloid and Interface Science *239*, 2–16.
108. Singh, J., and Padgett, R.A. (2009). Rates of in situ transcription and splicing in large human genes. Nat Struct Mol Biol *16*, 1128–1133.
109. Smith, K.N., Miller, S.C., Varani, G., Calabrese, J.M., and Magnuson, T. (2019). Multimodal Long Noncoding RNA Interaction Networks: Control Panels for Cell Fate Specification. Genetics *213*, 1093–1110.
110. Sobell, H.M. (1985). Actinomycin and DNA transcription. Proc Natl Acad Sci U S A *82*, 5328–5331.
111. Srivastava, S., and Tirrell, M.V. (2016). Polyelectrolyte complexation. Advances in Chemical Physics 499–544.
112. Steurer, B., Janssens, R.C., Geverts, B., Geijer, M.E., Wienholz, F., Theil, A.F., Chang, J., Dealy, S., Pothof, J., van Cappellen, W.A., et al. (2018). Live-cell analysis of endogenous

GFP-RPB1 uncovers rapid turnover of initiating and promoter-paused RNA Polymerase II. Proc Natl Acad Sci USA *115*, E4368–E4376.
113. Stringer, C., Wang, T., Michaelos, M., and Pachitariu, M. (2020). Cellpose: a generalist algorithm for cellular segmentation. BioRxiv 2020.02.02.931238.
114. Strom, A.R., and Brangwynne, C.P. (2019). The liquid nucleome – phase transitions in the nucleus at a glance. J Cell Sci *132*.
115. Struhl, K. (2007). Transcriptional noise and the fidelity of initiation by RNA polymerase II. Nat. Struct. Mol. Biol. *14*, 103–105.
116. Suter, D.M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., and Naef, F. (2011). Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics. Science *332*, 472–474.
117. Taylor, N.O., Wei, M.-T., Stone, H.A., and Brangwynne, C.P. (2019). Quantifying Dynamics in Phase-Separated Condensates Using Fluorescence Recovery after Photobleaching. Biophysical Journal 117, 1285–1300.
118. Tunnacliffe, E., and Chubb, J.R. (2020). What Is a Transcriptional Burst? Trends in Genetics *36*, 288–297.
119. Umbarger, H.E. (1956). Evidence for a Negative-Feedback Mechanism in the Biosynthesis of Isoleucine. Science *123*, 848–848.
120. Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. Cell *153*, 307–319.
121. Zamudio, A.V., Dall'Agnese, A., Henninger, J.E., Manteiga, J.C., Afeyan, L.K., Hannett, N.M., Coffey, E.L., Li, C.H., Oksuz, O., Sabari, B.R., et al. (2019). Mediator Condensates Localize Signaling Factors to Key Cell Identity Genes. Molecular Cell.
122. Zhang, P., Shen, K., Alsaifi, N.M., and Wang, Z.-G. (2018). Salt Partitioning in Complex Coacervation of Symmetric Polyelectrolytes. Macromolecules *51*, 5586–5593.
123. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based Analysis of ChIP-Seq (MACS). Genome Biology 9, R137.

Chapter 5:
Dynamic clustering of insulin receptor underlies its signaling and is disrupted in insulin resistance

Originally published in *Nature Communications* volume 13, Article number: 7522 (2022)

Alessandra Dall'Agnese[1,#], Jesse M. Platt[1,2,#], Ming M. Zheng[1,3], Max Friesen[1], Giuseppe Dall'Agnese[1,4], Alyssa M. Blaise[1], Jessica B. Spinelli[1], Jonathan E. Henninger[1], Erin N. Tevonian[5], Nancy M. Hannett[1], Charalampos Lazaris[1], Hannah K. Drescher[2], Lea M. Bartsch[2], Henry R. Kilgore[1], Rudolf Jaenisch[1,6], Linda G. Griffith[5,7,8], Ibrahim I. Cisse[3,6], Jacob F. Jeppesen[1,9], Tong I. Lee[1,*], Richard A. Young[1,6,*]

[1]Whitehead Institute for Biomedical Research; Cambridge, MA 02142, USA.
[2]Division of Gastroenterology, Department of Medicine, Massachusetts General Hospital; Boston, MA 02114, USA.
[3]Department of Physics, Massachusetts Institute of Technology; Cambridge, MA 02139, USA.
[4]Department of Medicine, University of Udine; Udine, 33100, Italy
[5]Department of Biological Engineering, Massachusetts Institute of Technology; Cambridge, MA 02139, USA.
[6]Department of Biology, Massachusetts Institute of Technology; Cambridge, MA 02139, USA.
[7]Department of Mechanical Engineering, Massachusetts Institute of Technology; Cambridge, MA, USA.
[8]Center for Gynepathology Research, Massachusetts Institute of Technology; Cambridge, MA, USA.
[9]Global Drug Discovery, Novo Nordisk, Copenhagen, Denmark
*Corresponding authors. Email: young@wi.mit.edu, tlee@wi.mit.edu
[#]These authors contributed equally to this work

https://doi.org/10.1038/s41467-022-35176-7

Abstract

Insulin receptor (IR) signaling is central to normal metabolic control and is dysregulated in metabolic diseases such as type 2 diabetes. We report here that IR is incorporated into dynamic clusters at the plasma membrane, in the cytoplasm and in the nucleus of human hepatocytes and adipocytes. Insulin stimulation promotes further incorporation of IR into these dynamic clusters in insulin-sensitive cells but not in insulin-resistant cells, where both IR accumulation and dynamic behavior are reduced. Treatment of insulin-resistant cells with metformin, a first-line drug used to treat type 2 diabetes, can rescue IR accumulation and the dynamic behavior of these clusters. This rescue is associated with metformin's role in reducing reactive oxygen species that interfere with normal dynamics. These results indicate that changes in the physico-mechanical features of IR clusters contribute to insulin resistance and have implications for improved therapeutic approaches.

Main Text

Insulin signaling controls cell growth and metabolism, and dysregulation of this pathway is a common feature of type 2 diabetes (T2D), obesity and metabolic syndrome[1, 2]. Insulin binds at the cell surface to the insulin receptor (IR), a receptor tyrosine kinase (RTK)[3, 4]. Insulin binding induces IR autophosphorylation and IR phosphorylation of IR substrate (IRS) and src homology 2 (SHC) proteins, which activate PI3K-AKT and ERK signaling, respectively[1, 2, 3, 4, 5, 6, 7, 8]. These pathways regulate glucose uptake, lipogenesis, gluconeogenesis, glycogen synthesis and cellular proliferation[1, 2, 9]. The active IR is internalized by endocytosis and is either degraded in lysosomes, recycled back to the plasma membrane, or transported into the nucleus where it becomes associated with insulin-responsive genes[10, 11, 12, 13, 14, 15, 16].

Insulin resistance is a heterogeneous disorder common to type 2 diabetes (T2D), obesity, and metabolic syndrome[17, 18]. Multiple cell-extrinsic and cell-intrinsic factors can blunt the cellular response to insulin and thus contribute to insulin resistance[1, 2, 19]. These include alterations in insulin signaling components as a consequence of chronic hyperinsulinemia, nutritional excess, inflammation, oxidative stress, ER stress, fatty acid accumulation and mitochondrial dysfunction[1, 2, 17, 18, 20, 21, 22, 23].

Recent reports indicate that signaling factors can form dynamic clusters with properties and characteristics expected of biomolecular condensates[24, 25, 26, 27, 28, 29, 30, 31, 32, 33]. Biomolecular condensates are cellular compartments wherein proteins and nucleic acids concentrate without being physically delimitated by a membrane[34]. Condensate formation and condensate properties have been shown to contribute to diverse types of cellular signaling[24, 25, 26, 27, 28, 29, 30, 31, 32, 33]. For example, evidence suggests that T cell receptor activation causes formation of condensate compartments at the plasma membrane that incorporate signaling components and promote signaling[24, 25] and similar observations were recently reported for various RTKs[30]. The terminal components of the Wnt, Lif and TGFb developmental signaling pathways are directed to key developmental genes through integration into transcriptional condensates at those genes[26]. In addition, the dynamic properties of condensates have been shown to correlate with the activity and function of the molecules within the condensates[28, 31, 35, 36, 37, 38]. This previous evidence for dynamic clusters of signaling factors, coupled with the observation that insulin receptor can be seen as puncta when visualized in live cells, led us to investigate whether insulin signaling involves dynamic clustering and whether dysregulation of such clustering contributes to insulin resistance.

Here we report that IR is incorporated into dynamic clusters at the plasma membrane, in the cytoplasm and in the nucleus of human hepatocytes and adipocytes. Insulin stimulation promotes further incorporation of IR into these clusters in insulin-sensitive cells but not in insulin-resistant cells, where IR molecules within clusters exhibit less dynamic behavior. Metformin treatment of insulin-resistant cells rescues IR cluster dynamics and insulin responsiveness. Insulin-resistant cells are subjected to high levels of oxidative stress, which we find to cause reduced cluster dynamics, and treatment of these cells with metformin reduces levels of ROS and returns IR clusters to their normal dynamic behavior.

Insulin receptor bodies in human liver cells

Clusters of proteins can be visualized as punctate bodies in cells, and IR has previously been observed in punctate bodies in diverse cultured cells[16, 39, 40]. We investigated whether IR puncta occur in healthy human liver tissue and whether such puncta differ in T2D patients treated with

and without metformin, the front-line drug for T2D. We examined 23 human liver tissue samples, comprising seven from healthy donors, seven from donors with T2D and nine from donors with T2D under treatment with metformin (Fig. 1, Supplementary Table 1). These liver tissues exhibited histologic and metabolic features, as well as redox states, expected for healthy donors, donors with T2D and donors with T2D under metformin treatment[18, 41, 42, 43, 44, 45] (Supplementary Fig. 1a-c, Supplementary Table 1). Immunofluorescence for CK18 was used to assess tissue quality, cell morphology and as a marker for hepatocytes (Fig. 1a). Imaging of these tissues with a validated antibody for IR (Supplementary Fig. 1d,e) revealed that IR occurs in punctate bodies in hepatocytes from healthy donors, but these signals were significantly reduced in tissues from T2D donors that were not treated with metformin (Fig. 1a,b, Supplementary Fig. 1f,g). It was notable that hepatocytes from metformin-treated donors with T2D had IR punctate signals similar to those observed in healthy tissues. These differences were evident in puncta formed in the plasma membrane, the cytoplasm and the nucleus (Fig. 1, Supplementary Fig. 1g). The total levels of IR protein spanned a similar range in donor tissues from healthy and T2D donors (Supplementary Fig. 1h), suggesting that the reduced punctate signal in the tissue of T2D donors lacking metformin treatment is not simply due to a difference in the overall level of IR protein. These results suggest that the incorporation of IR into puncta in human hepatocytes is attenuated in T2D and is rescued to some extent by metformin treatment.

Insulin receptor bodies in HepG2 cells

To further investigate the features of IR puncta in hepatocytes, we turned to HepG2 cells because of their demonstrated utility in the study of insulin signaling and resistance, and because they are amenable to genetic modification[11, 46, 47]. Cells were cultured in media containing physiological concentrations of glucose (5mM) and insulin (0.1nM)[48, 49, 50] (Fig. 2a). Cell culture conditions were selected to mimic those experienced by hepatocytes *in situ* and cell viability and the ability of cells to clear insulin remained high under these conditions (Supplementary Fig. 2a,b). To confirm that these cells were insulin-sensitive, conventional assays of insulin sensitivity were performed. Acute insulin stimulation induced IR phosphorylation (Supplementary Fig. 2c), AKT and ERK pathway activation (Supplementary Fig. 2c), upregulation of lipogenic genes and downregulation of gluconeogenesis genes (Supplementary Fig. 2d,e), increased lipogenesis (Supplementary Fig. 2f), decreased glucose production (Supplementary Fig. 2g,h), and increased GSK3 phosphorylation (Supplementary Fig. 2i). Thus, the HepG2 cells cultured in this fashion exhibit the conventional features associated with insulin sensitivity.

IR localization was monitored in HepG2 cells by immunofluorescence super-resolution microscopy and was found to be incorporated into puncta at the plasma membrane, cytoplasm and nucleus of HepG2 cells in the absence of insulin stimulation, and this signal was elevated with insulin stimulation (Fig. 2a,b, Supplementary Fig. 3). Western blot analysis indicated HepG2 cells contain ~300,000 IR molecules/cell (Supplementary Fig. 4a), consistent with estimates for human hepatocytes[51], and showed that the unstimulated and insulin-stimulated cells contain similar levels of IR (Supplementary Fig. 4b). Given the similar levels of IR in unstimulated and insulin-stimulated cells, we infer from the imaging results that the increased IR signal in puncta reflects increased incorporation of IR molecules into these bodies from the surrounding intracellular environment, and not changes in the overall level of the protein. Active IR has been reported to localize at plasma membrane microdomains or signalosomes with other signaling proteins, enter the cytoplasm via endocytosis and become associated with lysosomes or be recycled to the plasma membrane, and enter the nucleus and bind to insulin responsive genes[10, 11, 12, 13, 14, 15, 16, 39, 40]. Our observations with IR bodies in the plasma

membrane, cytoplasm and nucleus of HepG2 cells are consistent with these prior reports (Fig. 2a; Supplementary Fig. 5). Electron microscopy with IR-specific antibodies confirmed that IR can be found near the plasma membrane, in the cytoplasm (some associated with membranes and some not) and in the nucleus (Supplementary Fig. 5a). Super-resolution microscopy confirmed that IR puncta can colocalize with a portion of the insulin signaling proteins AKT and PI3K (Supplementary Fig. 5b), that IR puncta can colocalize with clathrin vesicles and lysosomes (Supplementary Fig. 5c), and that IR puncta can be found at the periphery of endosome vesicles (Supplementary Fig. 5d). These results suggest that IR puncta are not simply concentrations of IR constrained as a consequence of being fully enveloped by membranes, but instead can sometimes be partially associated with membranes, consistent with previously published results for IR[52] and other protein assemblies associated with plasma membranes and endosomes such as those formed by other signaling factors and neuronal post-synaptic densities[27, 30, 53, 54, 55, 56]. In the nucleus, IR puncta were found colocalized with markers of the transcriptional machinery (MED1 and RNA Polymerase II) at the insulin responsive genes *FASN*, *SREBF1* and *TIMM22* (Supplementary Fig. 5e), and this was confirmed by ChIP-seq analysis of these proteins at these genes (Supplementary Fig. 5f).

To investigate whether IR puncta are altered in insulin resistance, we compared the insulin-sensitive HepG2 cells to cells in which an insulin-resistant state was induced by hyperinsulinemia. HepG2 cells were exposed to either physiologic levels (0.1nM) or pathologic levels (3nM) of insulin[48, 49, 50, 57] for two days (Supplementary Fig. 6a). Cells exposed to pathologic levels of insulin showed hallmarks of insulin resistance that are observed after insulin stimulation: reduced phosphorylation of IR, IRS1, AKT and ERK (Supplementary Fig. 6b-e), unchanged expression of the lipogenic gene *FASN* (Supplementary Fig. 6f), impaired suppression of glucose production and impaired promotion of lipogenesis (Supplementary Fig. 6g-i), and decreased phosphorylation of GSK3 (Supplementary Fig. 6j). These results are consistent with recent evidence that, in the liver of insulin-resistant patients, hepatocellular insulin signaling is blocked at the level of phosphorylation of insulin receptor and there is impaired insulin-mediated suppression of glucose production and impaired insulin-mediated promotion of lipogenesis[20]. Insulin-sensitive and resistant cells contained similar amounts of IR in whole cell extracts and at the cell surface (Supplementary Fig. 6k-m). Insulin binding was also similar between insulin-sensitive and resistant cells (Supplementary Fig. 6n). These results suggest that the attenuated IR signaling was not due to a substantial change in IR levels in cells and at the plasma membrane, nor due to changes in the ability of insulin receptor to bind insulin.

Immunofluorescence imaging of IR in insulin-resistant cells revealed that it is incorporated into puncta at the plasma membrane, cytoplasm and nucleus in a manner similar to that observed for insulin-sensitive cells (compare Fig. 2c and 2a, 0nM insulin). However, in these insulin-resistant cells, acute treatment with insulin (3nM) did not promote incorporation of additional IR into puncta (Fig. 2b,c), in contrast to the effects observed in insulin-sensitive cells (Fig. 2a,b). If the observed IR puncta defects are common features of insulin-resistant cells, then cells treated with other conditions expected to induce insulin resistance, such as chronic inflammation and high nutrient levels[17, 18], should exhibit IR puncta defects that phenocopy those caused by hyperinsulinemia. Treatment of cells with pathological concentrations of TNFA□or with high nutrients also caused a decrease in insulin-stimulated IR incorporation into puncta similar to that observed for hyperinsulinemia (Supplementary Fig. 7). These results suggest that IR puncta dysfunction, defined here with respect to accumulation of molecules in puncta, may be a common feature of insulin resistance induced by diverse factors.

To confirm these observations and enable imaging of IR in live cells, HepG2 cells were engineered to express endogenous IR as a fusion protein with monomeric enhanced green

fluorescent protein (IR-GFP) (Supplementary Fig. 8a). IR-GFP was expressed in these homozygous cells at the same levels as WT IR and was functional, as cells expressing this fusion protein maintained insulin-induced phosphorylation of IR and insulin signaling proteins (Supplementary Fig. 8b,c). A time course of insulin stimulation using live-cell imaging provided further evidence that insulin stimulation promotes an increase in IR-GFP signal in IR puncta (Supplementary Fig. 9a, b), as well as an increase in the number of IR puncta in the nucleus and cytoplasm (Supplementary Fig. 9c).

Given our observation that hepatocytes in liver tissue from metformin-treated T2D patients have IR puncta that resemble those in healthy donors, we investigated whether metformin could rescue the reduction in IR punctate signal seen in insulin-resistant HepG2 cells. We again observed that IR-GFP HepG2 cells rendered insulin resistant showed reduced insulin-promoted incorporation of IR into puncta (Fig. 2d,e) and found that treatment of these insulin-resistant cells with metformin partially restored IR signal in these puncta (Fig. 2d,e and Supplementary Fig. 10a). Treatment of insulin-sensitive HepG2 cells with metformin had little or no effect on IR puncta (Supplementary Fig. 10b). The rescue of IR puncta phenotype in insulin-resistant cells was not due to changes in IR levels (Supplementary Fig. 10c). IR puncta rescue was evident at 12.5μM metformin (Supplementary Fig. 10a), which approximates the concentration of metformin in the plasma of T2D patients[58, 59, 60]. These results indicate that the insulin-resistant state in these cells is associated with reduced IR incorporation in puncta, and that this dysfunction can be reversed to some extent by metformin, as observed in human liver tissue (compare Fig. 1a with Fig. 2d).

Insulin receptor bodies in primary hepatocytes and adipocytes

We next investigated whether similar IR puncta occur in human primary hepatocytes, whether these are altered in insulin resistance, and studied the effects of metformin on such puncta. Primary human hepatocytes can form three-dimensional spheroids and can be cultured for days with physiological or pathological concentrations of insulin while maintaining their cell identity and function (Supplementary Fig. 11a,b). These hepatocyte spheroids are insulin-sensitive if cultured with physiological concentrations of insulin and insulin-resistant if subjected to insulin levels characteristic of chronic hyperinsulinemia (Supplementary Fig. 11c,d). In insulin-sensitive human liver spheroids, IR was found in puncta at the plasma membrane, cytoplasm and nucleus (Supplementary Fig. 11e). As observed with HepG2 cells, insulin stimulation of hepatocyte spheroids produced an increase in IR signal intensity in cytoplasmic and nuclear puncta (Supplementary Fig. 11e). In insulin-resistant spheroids, by contrast, IR incorporation into cytoplasmic and nuclear puncta was diminished (Supplementary Fig. 11e), and metformin treatment partially rescued this attenuation of IR puncta signal (Supplementary Fig. 11e). These results show that the phenotypes observed for IR puncta in insulin-sensitive and insulin-resistant HepG2 cells also occur in primary human hepatocyte spheroids.

Adipocytes are among the cell types that exhibit insulin-resistant behavior, so we also investigated whether primary human adipocytes exhibit IR puncta phenotypes similar to those observed in hepatocytes (Supplementary Fig. 12). Primary human pre-adipocytes were first differentiated into adipocytes (Supplementary Fig. 12a) and then cultured for five days with either physiological concentrations of insulin or pathological concentration of insulin known to induce insulin-resistance in adipocytes[61] (Supplementary Fig. 12b). As observed with insulin-sensitive hepatocytes, IR-associated puncta were found at the plasma membrane, in the cytoplasm, and in nuclei of insulin-sensitive adipocytes, and insulin stimulation promoted further IR incorporation into these puncta (Supplementary Fig. 12c). In the insulin-resistant adipocytes,

insulin stimulation was less able to promote further IR incorporation into puncta and this reduction in signal was reversed by metformin (Supplementary Fig. 12c). These results show that primary human adipocytes exhibit IR puncta phenotypes similar to those observed in hepatocytes.

Characterization of insulin receptor bodies

The appearance of iR in punctate bodies suggests IR may be forming dynamic clusters similar to those seen with other signaling pathways, where such clusters can exhibit physical changes such as deformation, fission, and fusion[62, 63, 64, 65, 66, 67]. Super-resolution microscopy of HepG2 IR-GFP cells revealed IR-GFP puncta in the plasma membrane, cytoplasm and nucleus do indeed undergo deformation, fission and fusion (Fig. 3a). To investigate whether these clusters undergo dynamic formation and dissolution, we used time-correlated photoactivation localization microscopy (tc-PALM)[35, 62, 68] with a HepG2 cell line engineered to express endogenous IR as a fusion protein with Dendra2 (IR-Dendra2) (Supplementary Fig. 8a). IR-Dendra2 was expressed at the same levels as WT IR and was functional, as this fusion protein maintained its kinase activity (Supplementary Fig. 8b,c). IR-Dendra2 cells were subjected to tc-PALM and clusters of IR molecules were studied (Fig. 3b); several control analyses of the single molecule photochemistry were performed to validate the statistics of the molecular clusters examined here (Supplementary Fig. 13). The results revealed that IR forms dynamic clusters at the plasma membrane, cytoplasm and nucleus that exhibit various lifetimes, consistent with formation and dissolution times seen for other dynamic biomolecular assemblies (Fig. 2b,c). In cells with and without insulin stimulation, the majority of IR clusters (~85%) were short-lived (lifetime <100s) and had an average lifetime of 6-12s (Fig. 3c), comparable to those measured for other dynamic biomolecular assemblies in various cell types[35, 62, 68]. A smaller fraction of clusters (~15%) were present for considerably longer lifetimes (>100s) (Fig. 3c). Insulin stimulation resulted in an increase in the number of IR clusters in the cytoplasm and nucleus of these cells (Fig. 3d) and an increase in the number of IR detections in clusters in insulin-sensitive cells (Supplementary Fig. 14). The average number of IR-Dendra2 detections per cluster was estimated to be 22 (range 4–609) in unstimulated cells, and 27 (range 4–539) in insulin-stimulated cells (Supplementary Fig. 14). A similar trend was observed in IR clusters at the plasma membrane, in the cytoplasm and in the nucleus (Fig. 3e). These results suggest that multiple molecules of IR are incorporated into dynamic clusters, and that insulin stimulation leads to an increase in both the number of IR-containing clusters and the number of IR molecules per cluster in the cytoplasm and nucleus.

We next sought evidence that IR molecules present in clusters are functionally active. If IR kinase activity occurs in these clusters, then we would expect that the level of phosphorylated IR substrate IRS1 would increase in these IR-associated clusters upon insulin stimulation. Indeed, immunofluorescence microscopy with an antibody specific for phosphorylated IRS1 (pIRS1) showed that insulin stimulation increased the intensity of pIRS1 at IR clusters (Fig. 3f). IR incorporation into clusters positively correlated with signal intensity of pIRS1 in these clusters (Supplementary Fig. 15a) and pIRS1 was more concentrated inside IR clusters than outside (Supplementary Fig. 15b). In addition, acute stimulation of insulin-sensitive cells with a range of insulin concentrations produced a non-linear transition in IR-incorporation into clusters (Fig. 3g-h). The sharp increase in IR signal in clusters occurred coincident with insulin receptor activity and function measured by IRS1 phosphorylation (Fig. 3g-i), which is expected if the IR molecules incorporated into assemblies are functional.

Altered insulin receptor dynamics in insulin-resistant cells and rescue by metformin

Chronic signaling was recently shown to reduce the dynamic properties of other dynamic clusters formed by signaling factors[36] so we investigated whether the dynamics of IR clusters are altered in insulin-resistant cells (Fig. 4). HepG2 cells expressing IR-Dendra2 were exposed to physiologic levels of insulin (0.1nM) to maintain insulin sensitivity, or pathologic levels of insulin (3nM)[48, 49, 50, 57] to promote insulin resistance (Fig. 4a). tc-PALM was used to measure IR cluster dynamics in the insulin-sensitive and resistant cells. The results showed that IR molecules remained in clusters for longer lifetimes in the cytoplasm and nucleus in insulin-resistant cells relative to insulin-sensitive cells (Fig. 4b). The average lifetime of short-lived IR clusters in sensitive versus insulin-resistant cells increased from 6.8s to 11.8s at the plasma membrane, from 10.0s to 15.8s in the cytoplasm and from 7.0s to 12.9s in the nucleus. The percentage of long-lived IR clusters also increased in the plasma membrane, cytoplasm and nucleus (Fig. 4b). These results suggest that the insulin-resistant state is associated with reduced IR cluster dynamics, reflected in the longer lifetime of these clusters, which may account for the attenuated responses observed during insulin stimulation.

We wondered whether IR cluster dynamics are also decreased in other models of insulin resistance. Treatment of cells with pathological concentrations of TNFα or with high nutrients decreased IR dynamics (Supplementary Fig. 16a,b). These results indicate that IR cluster dysfunction, defined here with respect to accumulation and dynamics of molecules, may be a common feature of insulin resistance induced by diverse factors.

We next examined the effect of metformin treatment on IR cluster dynamics. Metformin treatment of insulin-resistant cells rescued IR cluster lifetimes in the plasma membrane, cytoplasm and nucleus to times that were similar to those in insulin-sensitive cells (Fig. 4b). For example, while ~40% of cytoplasmic IR clusters in insulin-resistant cells had a lifetime of 0 – 13s, ~60% of IR clusters in the cytoplasm of insulin-sensitive and metformin-treated resistant cells had a lifetime of 0 – 13s (Fig. 4b). Similarly, the frequency of plasma membrane and nuclear IR clusters with 0 – 13s lifetimes, which was reduced in the resistant cells relative to sensitive cells, was increased by the metformin treatment (Fig. 4b). In contrast, metformin did not decrease IR cluster lifetime in insulin-sensitive cells (Supplementary Fig. 17). Thus, metformin treatment rescues the dynamic properties of IR-containing clusters that occur in insulin-resistant cells.

We next investigated whether IR kinase activity differs in clusters in insulin-resistant cells and in these cells treated with metformin. Imaging experiments revealed reduced levels of phosphorylated IRS1 in IR-containing clusters in insulin-resistant cells as compared to insulin-sensitive cells (Figure 4c). Metformin treatment partially rescued the levels of phosphorylated IRS1 in IR-containing clusters in insulin-resistant cells (Figure 4c). These results were further supported by western blotting experiments that revealed a partial rescue of IRS1 phosphorylation in insulin-resistant cells by metformin treatment (Supplementary Fig. 18). Taken together, these results suggest that IR kinase activity is reduced in IR clusters in insulin-resistant cells and that metformin treatment can partially reverse this effect.

We next explored whether changes in the dynamics of IR-containing clusters might have a direct effect on IR kinase activity. To decrease IR molecule dynamics within clusters, we fused IR-GFP to 4 tandem repeats of FK506 binding protein (FKBP; IR-FKBP), which interact with each other only in the presence of the small molecule AP1903[69]. IR-FKBP was expressed in HepG2 cells and these cells were treated with AP1903 or control DMSO (Figure 4d-h). Treating HepG2 cells expressing IR-FKBP with AP1903 significantly increased the lifetime of IR clusters,

consistent with a reduction in IR molecule dynamics in these clusters (Fig. 4f). Western blotting and imaging experiments revealed that IR was less functionally active in cells expressing IR-FKBP treated with AP1903 (Figure 4g,h, Supplementary Fig. 19). Taken together, these results indicate that a decrease in insulin receptor cluster dynamics can produce a decrease in IR activity.

High ROS levels promote IR cluster dysregulation

Several observations led us to test the hypothesis that high levels of reactive oxygen species (ROS) contribute to dysregulated IR clusters in insulin-resistant cells. Insulin stimulation causes a transient increase in H_2O_2 levels[70, 71, 72, 73, 74, 75]. Many cell-extrinsic factors that promote insulin resistance, including hyperinsulinemia, TNFα and high nutrients lead to excessive production of ROS[19, 76, 77, 78] (Supplementary Fig. 7c). Insulin-resistant cells and patients with T2D have been shown to have elevated levels of ROS[43, 44] and high ROS is a known cell-intrinsic factor that promotes insulin resistance[19, 76, 79, 80]. Metformin has been proposed to decrease ROS levels by multiple mechanisms, including inhibition of the mitochondrial complex I respiratory chain[45], inhibition of the redox shuttle enzyme mitochondrial glycerophosphate dehydrogenase[42], and upregulation and activation of anti-oxidants[81]. Importantly, oxidative stress has previously been shown to affect the dynamic behaviors of other cluster-forming proteins[82, 83, 84, 85, 86, 87].
To test this idea, we first determined if insulin-resistant cells are subjected to higher levels of oxidative stress than insulin-sensitive cells. Imaging of NRF2, a marker of oxidative stress[88], revealed that insulin-resistant cells experienced higher levels of oxidative stress than insulin-sensitive cells (Fig. 5a). Quantification of ROS using a ROS-sensitive dye revealed that ROS levels were higher in insulin-resistant cells and, furthermore, that metformin treatment of these cells reduced ROS levels to those found in insulin-sensitive cells (Fig. 5b).

If oxidative stress causes IR cluster dysregulation, then treatment of insulin-sensitive cells with concentrations of an oxidizing agent known to cause oxidative stress might be expected to phenocopy the effects seen with insulin resistance. Similarly, if metformin acts by relieving the effects of oxidative stress, treatment of insulin-resistant cells with a reagent that reduces oxidative stress might phenocopy the effects of metformin. Indeed, we found that treating insulin-sensitive cells for 30 minutes with a concentration of H_2O_2 known to cause oxidative stress[89] caused a reduction in the incorporation of IR into clusters with insulin stimulation and altered IR cluster dynamics, phenocopying the IR cluster dysregulation seen in insulin-resistant cells (Fig. 5c,d, Supplementary Fig. 20a). Furthermore, treatment of insulin-resistant cells with clinically relevant concentrations of N-acetyl cysteine (NAC)[90, 91] partially rescued the dynamic behavior of IR clusters (Fig. 5e,f, Supplementary Fig. 20b). Together, these results suggest that chronic hyperinsulinemia leads to excess levels of ROS in insulin-resistant HepG2 cells, that high levels of ROS alter IR incorporation into clusters, and that anti-oxidants can partially rescue the behavior of IR-clusters as a consequence of reducing ROS levels.

DISCUSSION

Recent studies have shown that the components of diverse signaling pathways, including those involving receptor tyrosine kinases, T cell receptor, WNT, TGF-β, and JAK/STAT, involve the assembly of protein molecules into condensates at the plasma membrane, in the cytoplasm and nucleus[24, 25, 26, 27, 28, 29, 30, 31, 32, 33]. Our evidence indicates that this is also the case for the insulin receptor, as the IR clusters observed here have characteristics expected of condensates. IR clusters form punctate bodies in cells[34, 92], undergo fusion, fission and deformation[62, 63, 64, 65, 66, 67]

and typically exhibit the short lifetimes described for other signaling condensates[62], so we propose that these IR clusters are biomolecular condensates (Fig. 6).

Our results reveal that IR is incorporated into clusters at the plasma membrane, in the cytoplasm and in the nucleus. Acute insulin stimulation promotes further incorporation of IR into clusters in insulin-sensitive cells. In insulin-resistant cells, however, the ability of insulin stimulation to promote further IR incorporation into clusters is attenuated. Furthermore, IR cluster dynamics is altered in the insulin resistant cells, but can be rescued with metformin treatment. In insulin-resistant cells, prolonged elevation of ROS levels appears to account for altered IR cluster dynamics because it can be phenocopied by H_2O_2 treatment of insulin-sensitive cells and rescued by NAC treatment of insulin-resistant cells. Metformin likely rescues IR cluster dynamics in insulin-resistant cells by reducing ROS levels.

We find that prolonged elevation of ROS levels in chronic hyperinsulinemia reduces the ability of insulin to promote further incorporation of IR molecules into clusters and extends the lifetime of IR molecules within the existing clusters. The known effects of ROS on proteins provides a mechanism to explain these findings. Transient insulin-induced H_2O_2 formation is essential for mediating insulin signaling[70, 71, 72, 73, 74, 75], but ROS can cause protein oxidation, which can alter protein conformation and change the ability of proteins to be incorporated into clusters[82, 93]. It is possible that ROS-induced alteration of proteins may be a common mechanism in the pathogenesis of insulin resistance-associated diseases, including T2D, metabolic syndrome, non-alcoholic fatty liver disease (NAFLD), polycystic ovarian syndrome (PCOS), and Alzheimer's Disease. Indeed, metformin has been shown to decrease ROS production and improve patient outcomes in T2D and other diseases characterized by high ROS levels[42, 81, 94, 95, 96]. The mitochondrial respiratory chain complex 1 has been reported to be the primary target of metformin[45], but it may be the reduction in ROS levels and the consequent benefit to protein cluster dynamics that is key to normal IR signaling.

The proposal that insulin resistance is associated with IR cluster dysfunction is consistent with prior evidence that implicates defects of insulin signaling pathways in hepatocytes in vivo[20] and specific cellular stresses in both insulin resistance and cluster dysregulation. Some systemic and intracellular stresses that have been reported to induce insulin resistance, including oxidative stress and mitochondrial dysfunction, have independently been shown to influence the formation or behavior of other cellular condensates[93, 97, 98]. Further study of the molecular components of IR clusters and their oxidative modification should provide more detailed insights into the physicochemical properties that are altered in clusters by oxidative stress and mitochondrial dysfunction.

The work presented here proposes a mechanism that may explain, in part, how attenuation of insulin signaling occurs at a physicochemical level under conditions of insulin resistance in cells. Insulin resistance is a complex phenotype that can be considered at different scales - from cell, to tissue, to whole organism - and insulin resistance-related phenotypes will likely be best understood as the net result of cell and organism-level effects working together over time[1, 2, 19, 99]. For example, while attenuated insulin signaling in insulin resistant liver creates the expected effects on metabolic processes like gluconeogenesis and fat accumulation at the cellular level, cross-talk in the form of increased flux of metabolic substrates from insulin-resistant adipose tissue appears to have additive or compensatory effects, thus resulting in the final net phenotypes of hepatocyte gluconeogenesis and fat accumulation[20, 100, 101, 102, 103, 104, 105, 106, 107]. For these reasons, cell-based models may provide valuable insights into the mechanisms of dysregulated insulin signaling, but it will be important to integrate insights from these models

with those obtained with organism-level studies for a more complete understanding of the insulin-resistant state.

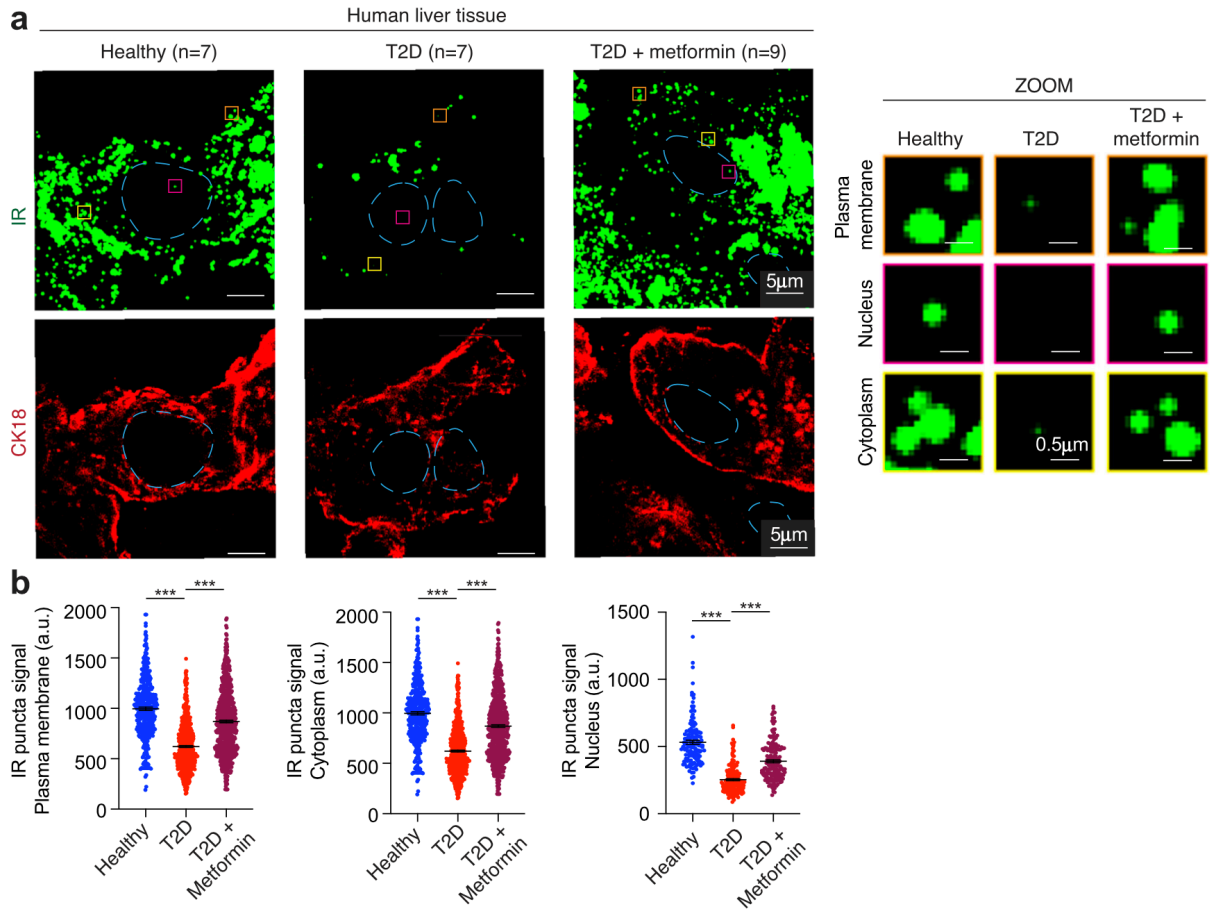
The model described here for IR dysfunction has implications for development of novel therapeutics for T2D. For example, the assays described here might be leveraged to develop new therapeutics that improve clinical outcomes for patients who cannot tolerate metformin or become resistant to the drug with prolonged use. Such therapeutics might also provide benefits to patients with other diseases where condensate dysregulation is also thought to play a role[97, 108, 109].

ACKNOWLEDGMENTS

We thank Drs. Niels Blume, Rita Slaaby, Ozgur Oksuz, and members of the Young Lab for helpful discussions, Wendy Salmon, Brandyn Braswell and Cassandra Rogers at the W.M. Keck Microscopy facility for imaging support, Dr. Paloma Cejas, Sudeepa Syamala and Dr. Henry Long at the Dana Farber Cancer Institute for ChIP-seq, Thomas Volkert at the Whitehead Institute Genome Technology Core for RNA-seq support, the Whitehead Metabolite Profiling Core Facility for metabolomics support, Nicki Watson from Harvard CNS for Electron Microscopy support, and Dr. Fabian Schulte at the Whitehead Quantitative Proteomics Core for proteomics experiment. Funding sources include: National Institute of Health grant R01 GM123511 (R.A.Y.); National Science Foundation grant PHY1743900 (R.A.Y.); Novo Nordisk (R.A.Y, L.G.G. and R.J); Hope Funds for Cancer Research fellowship (A.D.); National Institute of Health T32 5T32DK007191-45 (J.M.P.); National Institute of Health KL2 KL2TR002542 (J.M.P.); National Cancer Institute F32 5F32CA254216 (J.E.H.); National Science Foundation fellowship 1745302 (E.N.T.); National Institute of Health grant R01 1DP2CA195769-01 (I.I.C.); German Research Foundation DR 1161/1.1 (H.K.D); German Research Foundation DR BA7175/1.1 (L.M B.).

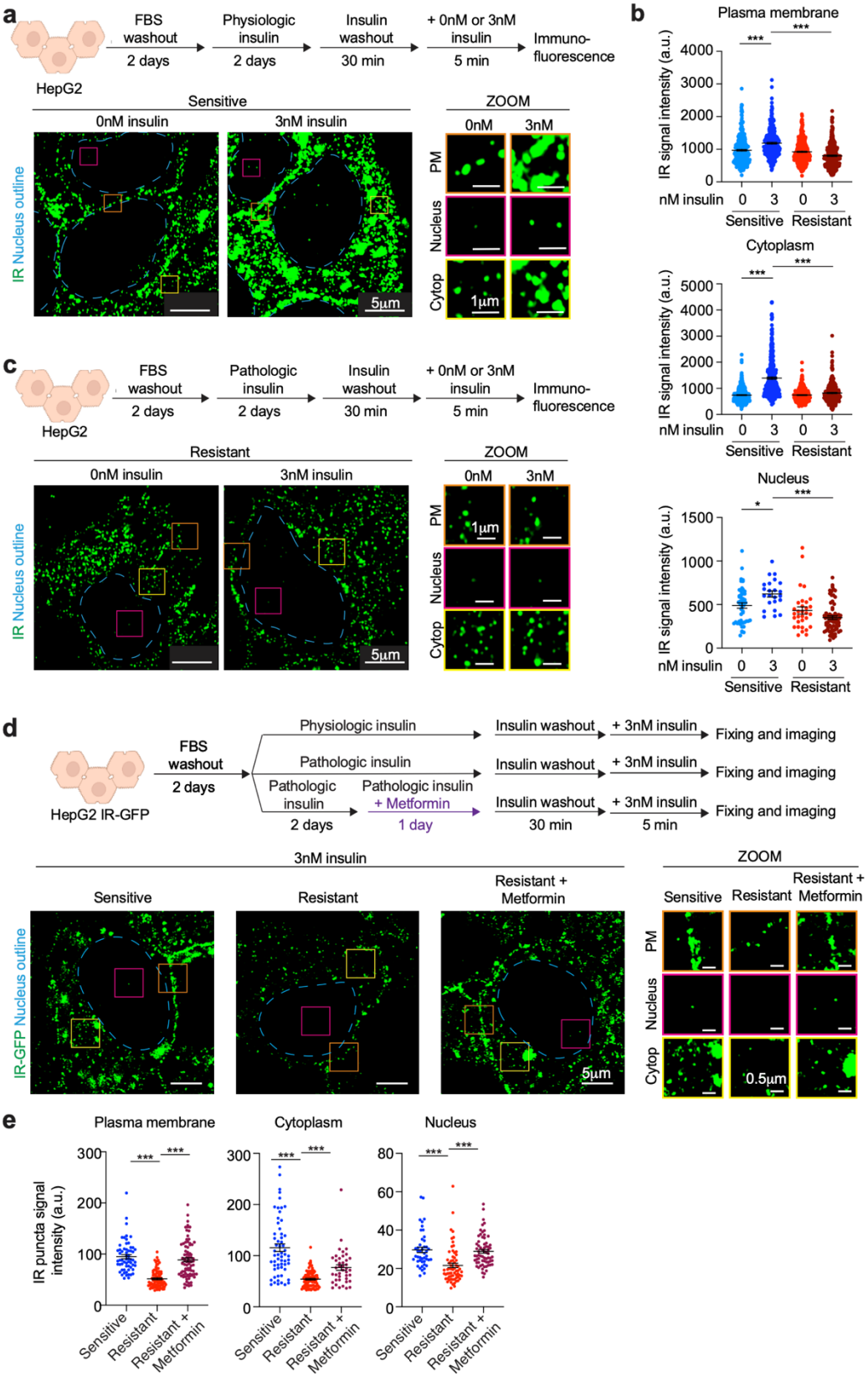
Figures and Tables

Fig. 1: Insulin receptor bodies in human liver cells



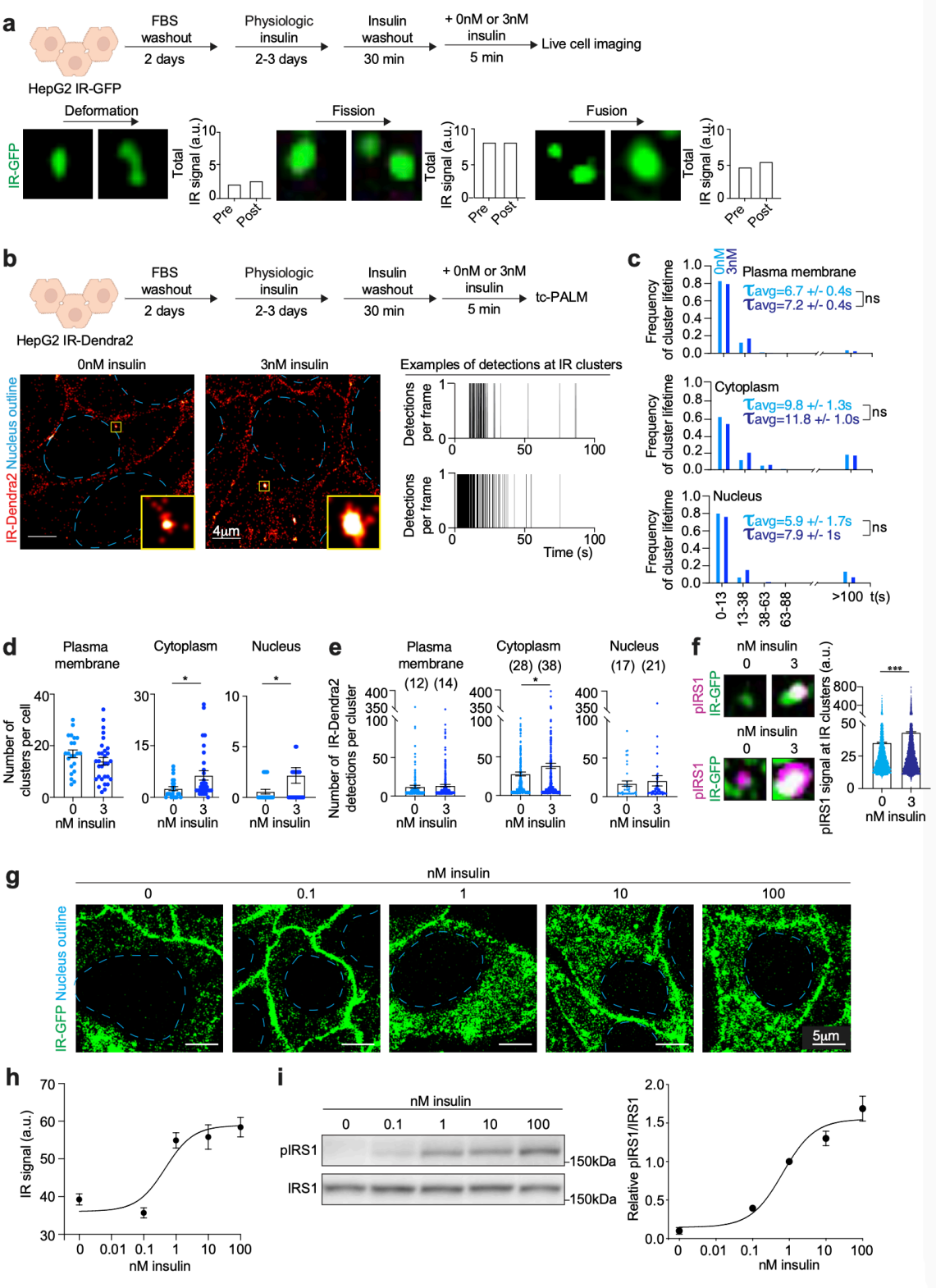
a, Representative immunofluorescence images for IR and CK18 in liver tissue from a healthy donor (Healthy), a donor with T2D (T2D), and a donor with T2D who had been treated with metformin (T2D + metformin). Dashed light blue lines represent the nuclear outline. Orange, magenta and yellow boxes represent regions at the plasma membrane, nucleus and cytoplasm, respectively, that are magnified on the right (ZOOM). Scale bars are indicated in the images. **b**, Quantification of IR signal in puncta at the plasma membrane, cytoplasm and nucleus for 7 healthy donors (blue), 7 donors with T2D (red) and 9 donors with T2D who had been treated with metformin (purple). Data is represented as mean +/- standard error of the mean (SEM). Number of puncta analyzed: Healthy Plasma membrane 399 puncta, Cytoplasm 304 puncta, Nucleus 137 puncta; T2D Plasma membrane 618 puncta, Cytoplasm 283 puncta, Nucleus 187 puncta; T2D Metformin Plasma membrane 716 puncta, Cytoplasm 350 puncta, Nucleus 173 puncta. Unpaired two-sided t-test was used for statistical analysis.

Fig. 2: Insulin receptor bodies in HepG2 cells



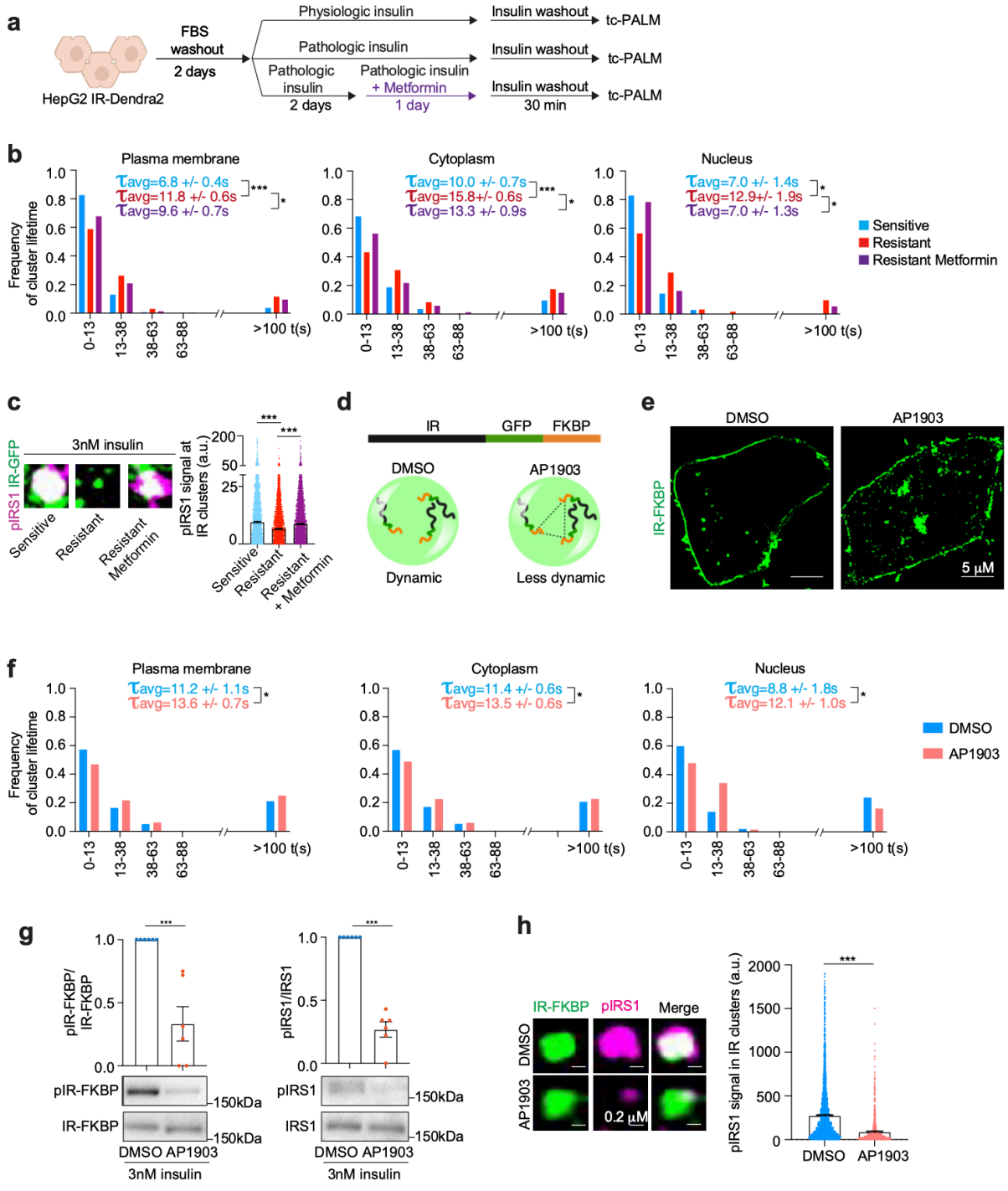
a, Schematic of cell treatments (top). Representative immunofluorescence images of IR (green) in cells stimulated with (3nM) or without (0nM) insulin for 5 min (bottom). Dashed light blue lines represent the outline of the nucleus. Orange, magenta and yellow boxes represent regions at the plasma membrane (PM), nucleus and cytoplasm (Cytop), respectively, that are magnified (ZOOM, middle). Scale bars are indicated. **b,** Quantification of IR signal intensity in puncta at the plasma membrane (top), cytoplasm (middle) and nucleus (bottom) in insulin-sensitive cells stimulated (3nM, blue) or not (0nM, light blue) with insulin and in insulin-resistant cells stimulated (3nM, dark red) or not (0nM, red) with insulin. Data is represented as mean +/- SEM. Number of puncta analyzed: Sensitive 0nM insulin Plasma membrane 342 puncta, Cytoplasm 366 puncta, Nucleus 48 puncta; Sensitive 3nM insulin Plasma membrane 272 puncta, Cytoplasm 378 puncta, Nucleus 23 puncta; Resistant 0nM insulin Plasma membrane 406 puncta, Cytoplasm 295 puncta, Nucleus 31 puncta; Resistant 3nM insulin Plasma membrane 328 puncta, Cytoplasm 283 puncta, Nucleus 69 puncta. Unpaired two-sided t-test was used for statistical analysis. **c,** Schematic of cell treatments to model insulin resistance (top). Representative immunofluorescence images for IR (green) in insulin-resistant cells acutely stimulated with (3nM) or without (0nM) insulin (bottom). Orange, magenta and yellow boxes represent regions at the plasma membrane (PM), nucleus and cytoplasm (Cytop), respectively, that are magnified (ZOOM, middle). Scale bars are indicated. **d**, Schematic of cell treatments (top). Cells expressing endogenous IR tagged with GFP (IR-GFP) were used. Metformin concentration used is 12.5μM. Representative images for IR-GFP in insulin-sensitive, insulin-resistant and metformin-treated insulin-resistant cells stimulated with insulin (3nM) for 5 minutes (bottom). Orange, magenta and yellow boxes represent regions at the plasma membrane (PM), cytoplasm (Cytop) and nucleus, respectively, that are magnified (ZOOM). Scale bars are indicated. **e**, Quantification of IR-GFP signal intensity in IR puncta at the plasma membrane, cytoplasm and nucleus of insulin-sensitive (blue), insulin-resistant (red) and metformin-treated insulin-resistant (purple) cells acutely stimulated with (3nM) insulin. Data is represented as mean +/- SEM. Number of puncta analyzed: Sensitive 3nM insulin Plasma membrane 60 puncta, Cytoplasm 59 puncta, Nucleus 45 puncta; Resistant 3nM insulin Plasma membrane 90 puncta, Cytoplasm 90 puncta, Nucleus 65 puncta; Resistant Metformin 3nM insulin Plasma membrane 82 puncta, Cytoplasm 41 puncta, Nucleus 67 puncta. Unpaired two-sided t-test was used for statistical analysis. Source data are provided as a Source Data file.

Fig. 3: Characterization of insulin receptor clusters



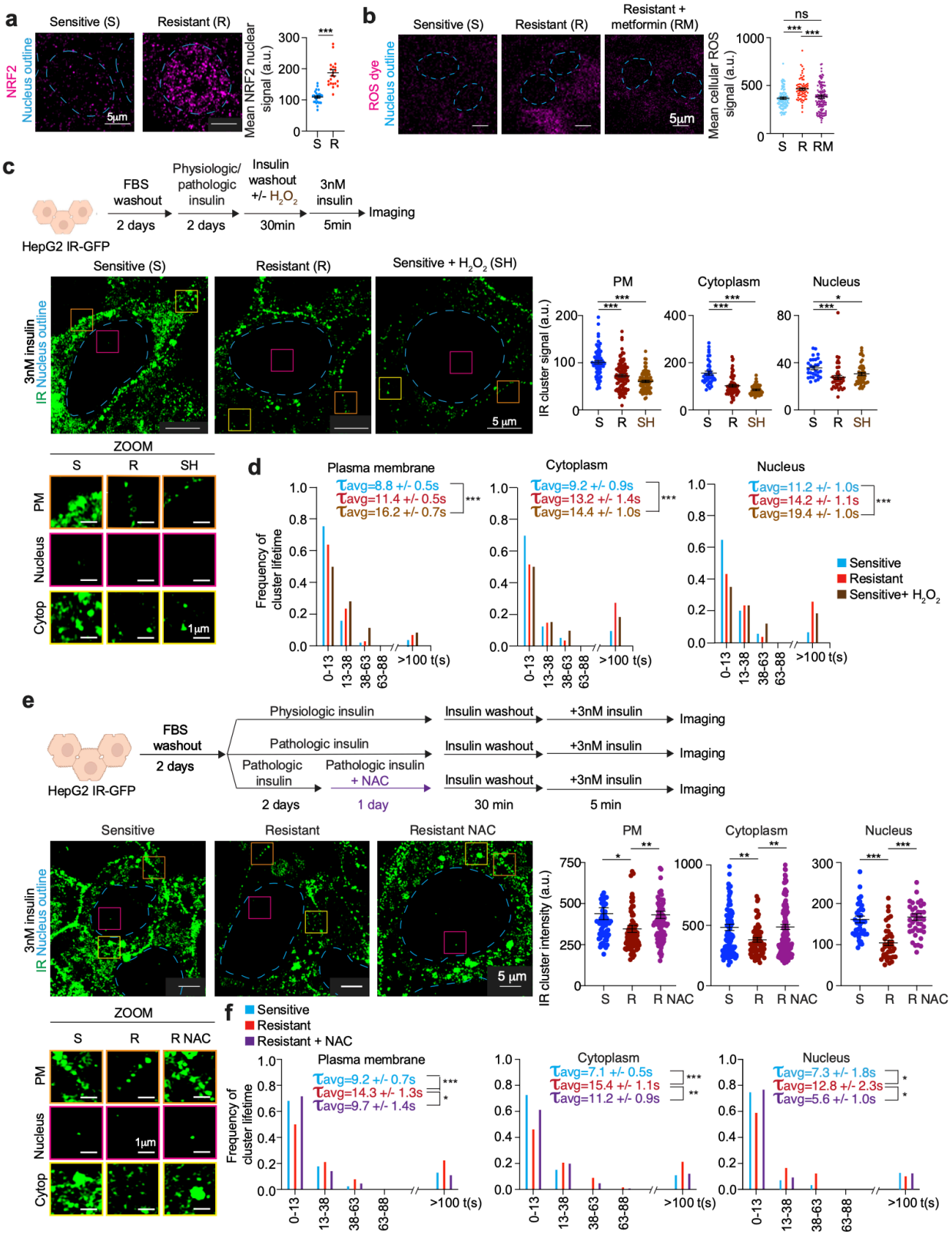
a, Schematic of cell treatments (top). Representative images of IR puncta undergoing deformation (bottom left), fission (bottom center) and fusion (bottom right). Quantification of total IR signal intensity over puncta pre- and post-deformation, fission or fusion. Images were taken 0.2s or 0.5s apart. **b,** Schematic of cell treatments (top). Representative tcPALM images of cells expressing IR-Dendra2 stimulated acutely with (3nM) or without (0nM) insulin for 5 min (bottom left). Scale bars are indicated. Representative tc-PALM traces (bottom right). **c,** Frequency of IR cluster lifetime in cells not acutely stimulated with insulin (0nM insulin, light blue) and acutely stimulated with insulin for 5min (3nM insulin, dark blue). Average lifetime (t_{avg}) of short-lived IR clusters +/- SEM is reported. Number of IR clusters analyzed: Sensitive 0nM insulin Plasma membrane 385 puncta, Cytoplasm 136 puncta, Nucleus 15 puncta; Sensitive 3nM insulin Plasma membrane 430 puncta, Cytoplasm 231 puncta, Nucleus 72 puncta. Unpaired two-sided t-test was used for statistical analysis. **d,** Number of IR clusters per cell in insulin-sensitive cells not acutely stimulated with insulin (light blue) and acutely stimulated with insulin (dark blue). Data is represented as mean +/- SEM. Number of cells analyzed: Sensitive 0nM insulin Plasma membrane 22 cells, Cytoplasm 22 cells, Nucleus 22 cells; Sensitive 3nM insulin Plasma membrane 30 cells, Cytoplasm 30 cells, Nucleus 30 cells. Unpaired two-sided t-test was used for statistical analysis for the cytoplasm and unpaired one-sided t-test was used for the nucleus. **e,** Number of IR-Dendra2 detections per IR cluster in insulin-sensitive cells not acutely stimulated with insulin (light blue) and acutely stimulated with insulin (dark blue). Average number of IR detections per IR cluster is reported on top of each histogram. Data is represented as mean +/- SEM. Number of clusters analyzed: Sensitive 0nM insulin Plasma membrane 430 clusters, Cytoplasm 499 clusters, Nucleus 37 clusters; Sensitive 3nM insulin Plasma membrane 573 clusters, Cytoplasm 551 clusters, Nucleus 57 clusters. Unpaired two-sided t-test was used for statistical analysis. **f**, Representative images of IR-GFP and phosphorylated IRS1 (pIRS1) in insulin-sensitive HepG2 cells stimulated acutely with (3nM) or without (0nM) insulin for 5 min (left). Quantification of pIRS1 signal in IR clusters in insulin-sensitive HepG2 cells stimulated acutely with (blue) or without (dark blue) insulin for 5 min (right). Data is represented as mean +/- SEM. Number of IR clusters analyzed: Sensitive 0nM insulin 7,640 clusters; Sensitive 3nM insulin 10,979 clusters. Unpaired two-sided t-test was used for statistical analysis. **g**, Representative images of insulin-sensitive cells stimulated with the reported concentrations of insulin for 5 min. **h**, Quantification of IR signal in clusters in the cytoplasm. Data is represented as mean +/- SEM. Number of IR clusters analyzed: 0nM insulin 141 clusters, 0.1nM insulin 164 clusters, 1nM insulin 163 clusters, 10nM insulin 101 clusters, 100nM insulin 100 clusters. **i**, Immunoblot and quantification of pIRS1 over total IRS1 in insulin-sensitive cells stimulated with the reported concentrations of insulin for 5 min. Data is represented as mean +/- SEM. Number of biologically independent samples analyzed: 9 per condition. Source data are provided as a Source Data file.

Fig. 4: Altered insulin receptor dynamics in insulin-resistant cells and rescue by metformin



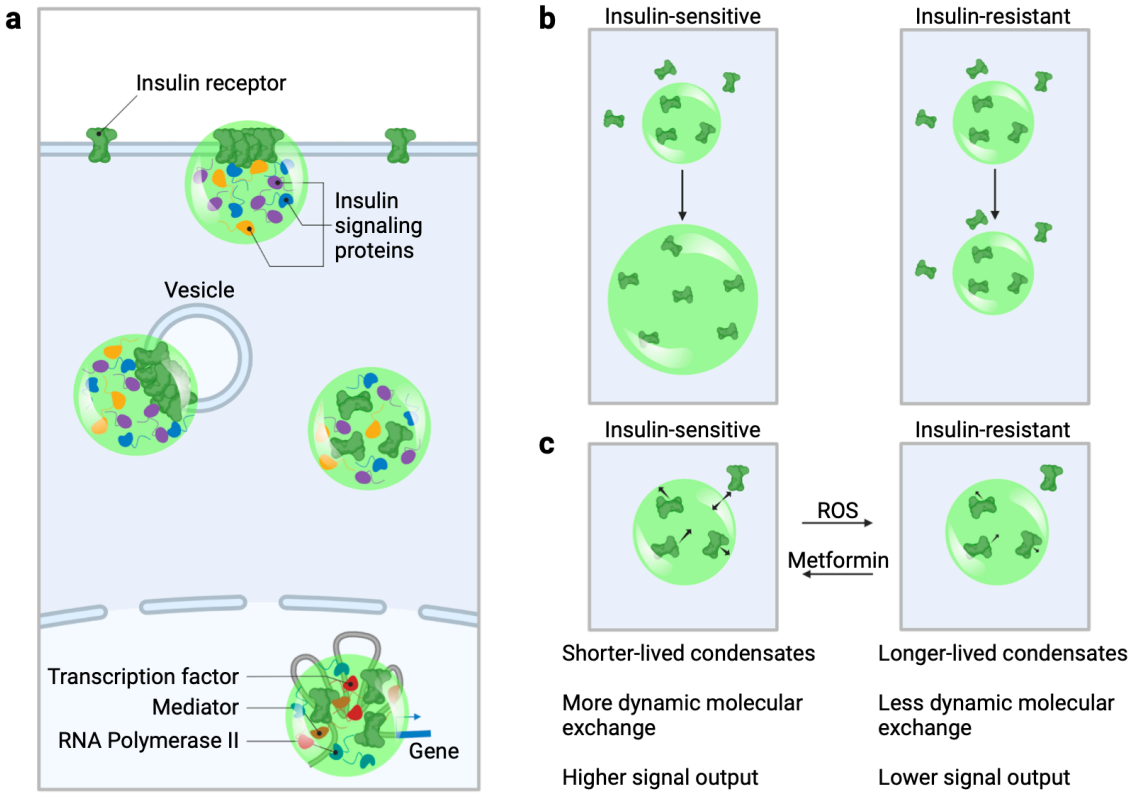
a, Schematic of cell treatments. **b,** Frequency of IR cluster lifetime in insulin-sensitive (light blue), insulin-resistant (red) and metformin-treated insulin-resistant (purple) cells. The concentration of metformin was 12.5μM. Cells were imaged after insulin washout. Average lifetime (t_{avg}) of short-lived IR clusters +/- SEM is reported. Number of IR clusters analyzed: Sensitive Plasma membrane 294 clusters, Cytoplasm 230 clusters, Nucleus 35 clusters; Resistant Plasma membrane 491 clusters, Cytoplasm 734 clusters, Nucleus 62 clusters; Resistant Metformin Plasma membrane 230 clusters, Cytoplasm 309 clusters, Nucleus 37 clusters. Unpaired two-sided t-test was used for statistical analysis of the of short-lived clusters. **c**, Example images of IR-GFP (green) and pIRS1 (magenta) in insulin-sensitive (Sensitive, light blue), insulin-resistant (Resistant, red) and metformin-treated insulin-resistant (Resistant Metformin, purple) cells stimulated acutely with 3nM insulin for 5 min (left). Quantification of pIRS1 signal in IR clusters (right). Data is represented as mean +/- SEM. IR clusters analyzed: Sensitive 4,859 clusters, Resistant 3,557 clusters, Resistant Metformin 8,964 clusters. Unpaired two-sided t-test was used for statistical analyses. **d**, Schematic representation of IR-GFP-FKBP (IR-FKBP) construct and the effect of DMSO and AP1903 on IR-FKBP clusters. **e**, Representative images of IR-FKBP in cells treated with DMSO or AP1903 for 16 hours. **f**, Frequency of IR-Dendra2-FKBP cluster lifetime in cells treated with DMSO (light blue) or AP1903 (red) for 16 hours. Average lifetime (t_{avg}) of short-lived IR clusters +/- SEM is reported. Number of clusters analyzed: DMSO Plasma membrane 194 clusters, Cytoplasm 499 clusters, Nucleus 50 clusters; AP1903 Plasma membrane 544 clusters, Cytoplasm 737 clusters, Nucleus 129 clusters. Unpaired two-sided t-test was used for statistical analysis for the cytoplasm, unpaired one-sided t-test was used for statistical analysis for the plasma membrane and nucleus. **g**, Immunoblot and quantification for phosphorylated IR-FKBP (pIR-FKBP) and pIRS1 over total protein. Cells expressing IR-FKBP were treated with DMSO (light blue) or AP1903 (red) for 16 hours. 6 biological replicates were analyzed for DMSO and AP1903-treated cells. Data is represented as mean +/- SEM. Unpaired two-sided t-test was used for statistical analysis. **h**, Representative images of IR-FKBP and pIRS1 in cells expressing IR-FKBP that were treated with DMSO (light blue) or AP1903 (red) for 16 hours. Data is represented as mean +/- SEM. 3,181 IR-FKBP clusters were analyzed in the DMSO condition and 985 IR-FKBP clusters in the AP1903 condition. Unpaired two-sided t-test was used for statistical analysis. Source data are provided as a Source Data file.

Fig. 5: High ROS levels promote IR cluster dysregulation



a, Representative immunofluorescence images for NRF2 (magenta) in insulin-sensitive or resistant cells (left). Dashed light blue lines represent nuclear outline. Quantification of mean NRF2 signal intensity in nuclei of insulin-sensitive (Sensitive, S) or resistant (Resistant, R) cells (right). Data is represented as mean +/- SEM. Number of cells analyzed: insulin-sensitive 27 cells and insulin-resistant 22 cells. Unpaired two-sided t-test was used for statistical analysis. **b,** Representative images of cells treated with ROS-sensitive dye (left). Dashed light blue lines represent nuclear outline. Quantification of mean ROS signal in insulin-sensitive (Sensitive, S), insulin-resistant (Resistant, R) and metformin-treated insulin-resistant cells (Resistant + metformin, RM) (right). Metformin concentration used was 12.5 μM. Data is represented as mean +/- SEM. Number of cells analyzed: Sensitive 107 cells, Resistant 70 cells, Resistant + metformin 134 cells. Unpaired two-sided t-test was used for statistical analysis. **c,** Schematic of cell treatments (top). Representative images of IR-GFP in insulin-sensitive cells (Sensitive, S), insulin-resistant cells (Resistant, R) or insulin-sensitive cells treated with H_2O_2 (Sensitive + H_2O_2, SH) (middle left). Dashed light blue lines represent nuclear outline. Orange, magenta and yellow boxes represent regions at the plasma membrane (PM), nucleus and cytoplasm (Cytop), respectively, that are magnified at the bottom (ZOOM). Scale bars are indicated. Quantification of IR-GFP signal intensity in IR clusters at the plasma membrane (PM), cytoplasm and nucleus in insulin-sensitive (S), insulin-resistant (R) and H_2O_2-treated insulin-sensitive (SH) cells (right). Data is represented as mean +/- SEM. Number of clusters analyzed: Sensitive Plasma membrane 68 clusters, Cytoplasm 40 clusters, Nucleus 30 clusters; Resistant Plasma membrane 96 clusters, Cytoplasm 60 clusters, Nucleus 37 clusters; Sensitive + H_2O_2 Plasma membrane 79 clusters, Cytoplasm 45 clusters, Nucleus 44 clusters. Unpaired t-test was used for statistical analysis. **d,** Frequency of IR cluster lifetime in insulin-sensitive (Sensitive, light blue), insulin-resistant (Resistant, red) and H_2O_2-treated insulin-sensitive (Sensitive + H_2O_2, brown) cells. Average lifetime (t_{avg}) of short-lived IR clusters +/- SEM is reported. Number of short-lived clusters analyzed: Sensitive Plasma membrane 432 clusters, Cytoplasm 247 clusters, Nucleus 181 clusters; Resistant Plasma membrane 547 clusters, Cytoplasm 154 clusters, Nucleus 222 clusters; Sensitive + H_2O_2 Plasma membrane 564 clusters, Cytoplasm 379 clusters, Nucleus 314 clusters. Unpaired two-sided t-test was used for statistical analysis. **e,** Schematic of cell treatments (top). Representative images of IR-GFP in insulin-sensitive cells (S, Sensitive), insulin-resistant cells (R, Resistant) or insulin-resistant cells treated with NAC (R NAC, Resistant NAC) (middle left). Dashed light blue lines represent nuclear outline. Orange, magenta and yellow boxes represent regions at the plasma membrane (PM), nucleus and cytoplasm (Cytop), respectively, that are magnified at the bottom (ZOOM). Scale bars are indicated. Quantification of IR-GFP signal intensity in IR clusters in insulin-sensitive (S), resistant (R) and NAC-treated insulin-resistant (R NAC) cells (middle right). Data is represented as mean +/- SEM. Number of IR clusters analyzed: Sensitive Plasma membrane 66 clusters, Cytoplasm 109 clusters, Nucleus 40 clusters; Resistant Plasma membrane 74 clusters, Cytoplasm 73 clusters, Nucleus 40 clusters; Resistant NAC Plasma membrane 91 clusters, Cytoplasm 183 clusters, Nucleus 41 clusters. Unpaired two-sided t-test was used for statistical analysis. **f**, Frequency of IR cluster lifetime in insulin-sensitive (Sensitive, light blue), insulin-resistant (Resistant, red) and NAC-treated insulin-resistant (Resistant + NAC, purple) cells. Average lifetime (t_{avg}) of short-lived IR clusters +/- SEM is reported in the graphs. Number of short-lived clusters analyzed: Sensitive Plasma membrane 143 clusters, Cytoplasm 168 clusters, Nucleus 53 clusters; Resistant Plasma membrane 159 clusters, Cytoplasm 309 clusters, Nucleus 47 clusters; Resistant + NAC Plasma membrane 94 clusters, Cytoplasm 232 clusters, Nucleus 31 clusters. Unpaired two-sided t-test was used for statistical analysis for all comparisons, except for nucleus sensitive vs resistant for which unpaired one-sided t-test was used for statistical analysis. Source data are provided as a Source Data file.

Fig. 6: A proposed condensate model for insulin signaling and resistance



a, Insulin receptor (green) is incorporated into condensates at the plasma membrane, at vesicle membranes, in the cytosol and in the nucleus, together with other insulin signaling proteins and, in the nucleus, with proteins involved in transcription (transcription factors, Mediator, RNA Polymerase II). **b**, Insulin stimulation promotes IR incorporation into condensates in insulin-sensitive cells and this effect is attenuated in insulin resistance. **c**, In insulin-resistant cells, IR condensates are longer lived and have less dynamic molecular exchange than those in insulin-sensitive cells, and this difference in IR condensate dynamics correlates with signal output. Source data are provided as a Source Data file.

.

Materials and Methods

Human liver donor samples

Samples of human livers were purchased from BioIVT or shared by collaborators at MGH (Hannah K Drescher and Lea M Bartsch). Informed consent was obtained by BioIVT or MGH from all human research participants. Sample ID numbers and donor information are obtained from either BioIVT or MGH and reported in Supplementary Table 1. Frozen samples were embedded in OCT compound (Tissue-Tek, 4583), re-frozen on dry ice and stored at -80°C. Embedded samples were sectioned using the cryostat at the W.M. Keck Microscopy Facility, MIT. Sectioning was performed at -21°C to generate 10µm-thick slices that were then placed on a Superfrost Plus VWR Micro Slides (VWR, 48311-703) and stored at -20°C. Images of hematoxylin and eosin (H&E) stained liver tissue were obtained from BioIVT.
IRB 1999P004983 and IRB 2019P001245, COUHES E3272 and COUHES E3665.

Cell culture

HepG2 cells (ATCC HB-8065™) were used because of their demonstrated utility in the study of insulin signaling and resistance, and because they are amenable to genetic modification[11, 46, 47]. HepG2 cells were cultured in EMEM (ATCC 30-2003) supplemented with 10% FBS (Sigma Aldrich, F4135) at 37°C with 5% CO_2 in a humidified incubator. For passaging, cells were washed in PBS (Gibco, 10010-023) and TrypLE Express Enzyme (Life Technologies, 12604021) was used to detach cells from plates and dissociate cell clumps. To ensure proper cell dissociation, cells were incubated with TrypLE at 37°C with 5% CO_2 in a humidified incubator for 5 minutes; they were then mechanically dissociated by pipetting them up and down 8 times using a 5mL serological pipette attached to an unfiltered 200µl pipette tip. The 5 minutes incubation and mechanical dissociation were repeated one more time. TrypLE was quenched with EMEM supplemented with 10% FBS and cells were plated in new tissue culture-grade plates.
HEK293T cells (ATCC, CRL-3216) were used for the production of purified IRb protein. HEK293T cells were cultured in DMEM (GIBCO, 11995-073) supplemented with 10% FBS (Sigma Aldrich, F4135), 2mM L-glutamine (Gibco, 25030) and 100 U/mL penicillin-streptomycin (Gibco, 15140), at 37°C with 5% CO_2 in a humidified incubator.
Primary pre-adipocytes from (ATCC, PCS-210-010) were cultured in in Fibroblast Growth Kit-Low Serum (ATCC PCS-201-041), as per manufacturer's instructions. Cells for experiments were dissociated and plated at 18,000 cells/cm2. Two days later pre-adipocytes were differentiated in adipocyte differentiation media (ATCC PCS-500-050), as per manufacturer's instructions. Briefly, cells were washed and medium was replaced with adipocyte differentiation initiation medium. After 48 hours, half the medium was replaced. At day 4, medium was changed to adipocyte differentiation maintenance medium, and replaced every three days. At day 12, cells were rinsed and incubated in DMEM (Thermo Fisher, 11885084) with 0.1 or 3nM of insulin for 5 days, replacing medium every other day. In the last 24 hours, 12.5mM metformin was added. On day 17, cells were prepared for the assays by rinsing, and 30 minutes of washing. Afterwards, for imaging, cells were exposed to 3nM insulin for 5 minutes, rinsed with PBS and fixed in 4%PFA for 15 minutes at room temperature. For pAKT ELISAs cells were

exposed to 3nM insulin for 15 minutes and harvested in cell lysis buffer (Cell Signaling Technology, #9803) with phosphatase inhibitor (Thermo Fisher Scientific, 78442).
For human liver spheroids, primary human hepatocytes from a 50-year-old male donor (BioIVT; lot #SMC) were used. Cells were thawed in Cryopreserved Hepatocyte Recovery Media (CHRM, ThermoFisher), spun down at 100xg for 8 minutes, and resuspended in seeding medium (William's E with 5.5mM glucose, 2mM GlutaMax, 15mM HEPES, 5% FBS, 1% Pen/Strep, 100nM hydrocortisone, and insulin 200pM or 800pM corresponding to the proper experimental group). Spheroids were formed using custom alginate microwells. In brief, 120,000 cells were seeded per well and spun at 50xg for 2 minutes to seed microwells, and cultured in a volume of 300µL seeding medium. After 24 hours, cells were switched to maintenance media for the remainder of the experiment. This maintenance media was composed of William's E plus 6.25 µg/ml transferrin, 6.25ng/ml selenium, 0.125% fatty acid-free BSA, 20µM linoleic acid, 5.5mM glucose, 2mM GlutaMax, 15mM HEPES, 0.5% Pen/Strep, and 100nM hydrocortisone. Insulin was supplemented with concentrations adjusted to mimic healthy and disease-inducing states, either 200pM for physiological or 800pM for pathologic insulin levels. Media was exchanged every 48 hours throughout the experiment.

Endogenously-tagged cell line generation
A CRISPR/Cas9 system was used to generate genetically modified HepG2 cell lines. Target sequences were cloned into a plasmid containing sgRNA backbone, a codon-optimized version of Cas9 and mCherry. For IR targeting, two Cas9 gRNAs were used. For the generation of the IR-mEGFP, IR-Dendra2, and IR-Dendra2-FKBP endogenously tagged lines, homology directed repair templates were cloned into pUC19 using NEBuilder HiFi DNA Assembly Master Mix (NEB, E2621S). For IR-mEGFP and IR-Dendra2 cell lines, the homology repair template consisted of mEGFP or Dendra2 cDNA sequence flanked on either side by 800 bp homology arms amplified from HepG2 genomic DNA using PCR (Supplementary Fig. 7a). For the IR-Dendra2-FKBP cell line, the homology repair template consisted of Dendra2 cDNA sequence followed by four FK506 binding protein (FKBP) binding domains[69] flanked on either side by 800 bp homology arms amplified from HepG2 genomic DNA using PCR. The following sgRNA sequences with PAM sequence in parentheses were used for CRISPR/Cas9 targeting:

sgRNA_IR_C-term_1: cacggtaggcactgttagga(agg)
sgRNA_IR_C-term_2: taggcactgttaggaaggat(tgg)

To generate genetically modified cell lines, $2x10^6$ cells were transfected with 500 ng of Cas9 plasmid 1, 500ng of Cas9 plasmid 2, and 1,000 ng of non-linearized homology repair template using Lipofectamine 3000 (Invitrogen, L3000). Cells were sorted 48 hours after transfection for the presence of the mCherry fluorescent protein encoded on the Cas9 plasmid to enrich for transfected cells. This population of cells was allowed to expand for 1.5 to 2 weeks before sorting a second time for the presence of mEGFP or Dendra2 and single cells were plated into individual wells of a 96-well plate. The single cells were cultured in conditioned EMEM media (described below) for 1-1.5 months. 20-30 colonies were screened for successful targeting using PCR genotyping to confirm insertion. PCR genotyping was performed using Phusion polymerase (Thermo Scientific, F531S). Using the following primers, PCR products were amplified according to manufacturer specifications:

IR_fwd: GGAGAATGTGCCCCTGGAC

IR_rev: ttggtaaccaaacgagtccacct

To make conditioned media, we cultured HepG2 cells in fresh EMEM media (ATCC, 30-2003) supplemented with 10% FBS (Sigma Aldrich, F4135) for 3 days and saved the media (old EMEM media). The composition of conditioned EMEM media is as follows: 50% fresh EMEM media and 50% old EMEM media. The conditioned media was filter sterilized prior to use.
HepG2 cells expressing IR-mEGFP were used for super-resolution microscopy with LSM880 or LSM980 with Airyscan detector.
HepG2 cells expressing IR-Dendra2 were used for single-molecule super-resolution microscopy, Dendra2 is a green-to-red photo-switchable protein that allows for single-molecule imaging.

Constructs

For experiments that forced reduction in IR cluster dynamics, the vector used in this assay was modified from pJH135_pb_MCPx2_mCherry_rTTA vector[110]. IR-mEGFP-FKBP, which consists of the insulin receptor cDNA, flexible linker 1, mEGFP, flexible linker 2, and four copies of FKBP, was cloned into PmeI and NheI digested pJH135_pb_MCPx2_mCherry_rTTA using Gibson cloning by following the manufacturer's instructions (NEB, E2621S). This vector is called pJP204_pb_TetON_INSR_2A_GFP_Dmbr4.

Cell treatments

For insulin sensitivity and resistance experiments in HepG2 cells, cells were washed once with EMEM alone, without any supplements (ATCC, 30-2003) and cultured in EMEM for two days. Cells were then treated for two days with either physiological (0.1nM) or pathological (3nM) levels of insulin (Sigma Aldrich, I9278-5ML) in EMEM supplemented with 1.25% fatty acid-free bovine serum albumin (BSA; Sigma Aldrich, A8806-5G). Media was replenished every 12 hours. To wash out insulin, cells were washed with EMEM seven times, including: three quick washes, three 5-minute washes and a long 20 minute-wash in EMEM at 37°C. In order to investigate insulin response, cells were acutely treated for 5 minutes with insulin diluted in EMEM supplemented with 1.25% fatty acid-free BSA at 37°C with 5% CO_2 in a humidified incubator. Concentration of insulin used varied and is reported in the figures.
For TNFA treatment, cells were cultured in EMEM BSA containing 10pg/ml[111] of Human TNF-alpha Recombinant Protein (Thermo Fisher Scientific, PHC3016) for 2 days. Media was replenished every 12 hours. Insulin washout and insulin stimulation was performed as above.
For high nutrient condition, cells were cultured for 2 days in EMEM containing either: 1) 10mM glucose, 45μM oleic acid (CAYMAN CHEMICAL, 29557), 30μM palmitic acid (CAYMAN CHEMICAL, 29558) and 3nM insulin (called in the text "pathologic glucose, pathologic fat, and pathologic insulin (GFI)") or 2) 10mM glucose, 45μM oleic acid, 30μM palmitic acid and 0.1nM insulin (called in the text "pathologic glucose, pathologic fat, and physiologic insulin (GF)"). Control cells were cultured for 2 days with EMEM containing BSA control (CAYMAN CHEMICAL, 29556). Media was replaced every 12 hours. Insulin washout and insulin stimulation was performed as above.
For metformin treatment, metformin (Sigma Aldrich, D150959-5G) was resuspended in sterile water to a concentration of 1M and diluted in cell media to the reported concentrations. Insulin-resistant cells were treated with pathological concentrations of insulin and metformin at various

concentration reported in the figures. Media was replenished very 12 hours. Insulin washout and insulin stimulation was performed as above.
For N-acetyl cysteine treatment, insulin-resistant cells were treated with pathological concentrations of insulin and 1mM N-acetyl cysteine (Sigma Aldrich, A9165-25G) for 24 hours. Media was replenished very 12 hours. Insulin washout and insulin stimulation was performed as above.
For oxidative stress, insulin sensitive cells were treated with 20mM H_2O_2 (Sigma Aldrich, H1009) for 30 minutes. Insulin stimulation was performed as above.
For adipocytes, following differentiation, cells were cultured with EMEM for 2 days and with EMEM containing either physiological (0.1nM) or pathological (3nM) concentrations of insulin for 2 days. Cells were then cultured with EMEM containing either physiological (0.1nM) or pathological (3nM) of insulin or with pathological (3nM) concentrations of insulin and 12.5mM of metformin for 1 day. Cells were then washed with EMEM and acutely stimulated with or without 3nM insulin for 5 minutes prior to cell collection for immunofluorescence or ELISA.
For experiments that forced reduction in IR cluster dynamics, $1x10^5$ cells/cm^2 HepG2 cells were transfected with 0.07 µg/cm^2 pJP204_pb_TetON_INSR_2A_GFP_Dmbr4 using Lipofectamine 3000 (Invitrogen, L3000). On day 2, the cells were treated with 100ng/ml doxycycline (Sigma, D9891-5G). On day 3, the cells were treated with EMEM, 100ng/ml doxycycline containing either 5 µM AP1903 (MedChemExpress, NC1416062) or 5 µM DMSO (Sigma, D2650-100ML) for 16 hours and then harvested for imaging and western blot.

Cell viability
Cells were detached from plates and dissociated from clumps using TrypLE as described above. TrypLE was quenched with EMEM supplemented with 10% FBS. Dead cells were stained with trypan blue (Life Technologies, T10282) and the percentage of cell viability was then measured using the Countess II FL (Applied Biosystems, A27977) according to the manufacturer's specifications.

Insulin clearance
Insulin sensitive HepG2 cells were cultured in EMEM for 30 minutes and then in 3nM insulin for 0, 5 or 24 hours. Culture media was collected and insulin concentration was measured at all timepoints using Human/Canine/Porcine Insulin DuoSet ELISA kit (R&D Systems, DY8056-05) according to the manufacturer's specifications. Clearance fraction was calculated by dividing the measured insulin concentration in cultured media by the measured insulin concentration in the cell-free control wells.
Insulin clearance in human liver spheroids was evaluated by collecting media after 48 hours in culture media was removed and insulin concentration was measured using Human/Canine/Porcine Insulin DuoSet ELISA kit (R&D Systems, DY8056-05) according to the manufacturer's specifications. Clearance fraction was calculated by dividing the measured insulin concentration in cultured media by the measured insulin concentration in the cell-free control wells.

Glucose production

Insulin sensitive and resistant cells were cultured in EMEM for 30 minutes and then cells were treated with 0, 0.1, 1 and 10nM insulin in glucose production media, containing DMEM (Thermo Fisher Scientific, A1443001), 15mM HEPES (Gibco, 15630-080), 1mM pyruvate (Sigma Aldrich, P5280), 20mM lactate (Sigma Aldrich, L7022-5G) for 4-5 hours. Media was removed and glucose production was measured using Amplex™ Red Glucose/Glucose oxidase assay kit (Thermo Fisher Scientific, A22189) according to the manufacturer's specifications.
Measurement of glucose production in human liver spheroids was performed as follows. At Day 10 in culture, spheroids were washed 5 times with glucose free William's E media (Thermo Fisher, ME18082L1), followed by culture for 24 hours in glucose free William's E maintenance media, supplemented 1mM pyruvate, 20mM lactate, and between 0 to 10nM insulin stimulation. After 24 hours, media was collected and glucose quantified with the Amplex Red Glucose Assay Kit (Thermo Fisher, A22189) according to manufacturer instructions.

Albumin quantification

To assess hepatocyte spheroid function, media was collected during every media exchange and albumin secretion was assayed via ELISA kit (Bethyl Laboratories, E80-129) following the manufacturer's instructions.

siRNA experiments

HepG2 cells were reverse transfected using Lipofectamine™ RNAiMAX Transfection reagent (Thermo Fisher Scientific, 13778100) following the manufacturer's instructions. Cells were dissociated using TrypLE as previously described then seeded in 6-multiwells in 1ml EMEM supplemented with 10% FBS and the transfection reagent. Cells were cultured with the transfection reagent for 2-3 days prior to collection for Western blot and immunofluorescence. The INSR siRNA pool (Dharmacon Inc, L-003014-00-0005) and the ON-TARGETplus Non-targeting Control Pool (Horizon Discovery, D-001810-01-05) were used.

Western blot

Cells were washed with ice-cold PBS (Life Technologies, AM9625) and lysed in Cell Lytic M (Sigma Aldrich C2978) supplemented with protease and phosphatase inhibitors (Sigma Aldrich, 11873580001 and 4906837001) directly on the wells. Lysates were placed into a 1.5ml tube and mixed at 4°C for 20 minutes, sonicated and then centrifuged at 12,000xg for 15 minutes. Super natant was collected and protein concentration was determined using a BCA Protein Assay Kit (Life Technologies, 23250) according to the manufacturer's instructions. Equal amounts of protein (5-50µg per sample) were separated on 10% or 12% Bis-Tris gels in 5% XT MOPS running buffer (Bio-Rad Laboratories, 1610788) at 100V until dye front reached the end of the gel. Protein was then transferred to a 0.45µM PVDF membrane (Millipore, IPVH00010) in ice cold transfer buffer (25mM Tris, 192mM glycine, 20% methanol) at 300mA for 1 hour or 250mA for 2 hours at 4°C. After the transfer, membranes were blocked in either 5% non-fat milk (LabScientific, M0842) dissolved in TBST (2% Tris HCl pH 8.0, 1.3% 5M NaCl, 0.05% Tween 20) or 5% BSA (VWR, 102643-516) in 1X TBST for 15 minutes to 1 hour at room temperature with shaking. Membranes were then incubated overnight at 4°C in 1:1000 primary antibody (specific antibodies listed below) in 5% non-fat milk in TBST or 5% BSA in TBST. BSA was used for immunoblotting phosphorylated proteins, otherwise milk was used. Membranes were then

washed three times for 5 minutes in TBST shaking at room temperature prior to incubation in 1:10,000 secondary antibody (specific antibodies listed below) in 5% non-fat milk in TBST for 1 hour at room temperature. This was followed by three 10-minute washes in TBST. Membranes were developed with ECL substrate (Thermo Scientific, 34080) and imaged using a CCD camera (BIO RAD, 1708265). Immunoblot quantification was performed using the “analyze gel” tool on Fiji/ImageJ v2.1.0/153c.
The following primary antibodies were used for WB: anti-phosphorylated insulin receptor (Abcam, ab60946; Cell Signaling, 3026, dilution 1:1000), anti-insulin receptor beta (Cell Signaling, 23413 dilution 1:1000; Bethyl, A303-712A; Cell Signaling, 3025 dilution 1:1000), anti-insulin receptor alpha (Cell Signaling, 74118 dilution 1:1000), anti-phosphorylated IRS1 (Cell Signaling, 3070 dilution 1:1000), anti-IRS1 (Cell Signaling, 2382 dilution 1:1000), anti-phosphorylated AKT (Cell Signaling, 4056 dilution 1:1000), anti-AKT (Cell Signaling, 9272 dilution 1:1000), anti-phosphorylated ERK (Cell Signaling, 4377 dilution 1:1000), anti-ERK (Cell Signaling, 9102 dilution 1:1000), anti-pGSKa Cell Signaling, 8566 dilution 1:1000), anti-GSKa (Cell Signaling, 4337 and 12456 dilution 1:1000), anti-beta Actin (Sigma Aldrich, A5441 dilution 1:10000), and anti-GAPDH (Abcam, ab8245 dilution 1:1000). The following secondary antibodies were used: donkey anti-rabbit IgG (Cytiva Life Sciences, NA934-1ML, dilution 1:10000) and sheep anti-mouse IgG (Sigma Aldrich, NXA931V, dilution 1:10000).
For quantitative western blot analysis, equal numbers of cells were cultured in each well of on a 6-well plate. To estimate the number of cells per well, cells in two wells were dissociated with TrypLE (Life Technologies, 12604021) and counted using the Countess II (Applied Biosystems, A27977). Cells from another well were lysed on the plate as described above. A dilution series of purified IRb-mCherry and HepG2 cellular lysate was separated on 10% Bis-Tris gels in 5% XT MOPS running buffer. Immunoblotting was performed as above. Bands were quantitated using Fiji/ImageJ v2.1.0/153c, from which we calculated the estimated number of molecules of IRb per HepG2 cell.

Proteolytic surface shaving experiment
To compare IR amounts in whole cells and at the plasma membrane, proteolytic surface shaving experiment was performed. Equal numbers of cells were cultured in each well of on a 6-well plate and cultured with either 0.1nM insulin or 3nM insulin for 2 days. To estimate the relative IR amounts in the whole cell, cells were washed in PBS and lysed in ice-cold Cell Lytic M supplemented with protease and phosphatase inhibitors as previously described. To estimate the relative IR amounts at the plasma membrane (labeled in the figure as “Digested”), cells were digested with TrypLE for 10 minutes at 37˚C and quenched with EMEM 10% FBS. Cells were spun down at 300g for 5min, washed in PBS and ice-cold Cell Lytic M supplemented with protease and phosphatase inhibitors. Samples were then processed for Western blot as previously described and immunoblotted for insulin receptor alpha and beta actin.

Metabolomics
Metabolite isolation from liver tissue: Flash frozen tissues were pulverized with a mortar and pestle in a liquid nitrogen bath. Tissue powder was transferred into Eppendorf tubes and re-suspended in 800 uL ice-cold LC-MS grade 60:40 methanol:water (ThermoFisher). Samples were vortexed for 10 minutes at 4ºC. Then, 500 uL of ice-cold LC-MS grade chloroform

(provided by the Metabolomics core) was added to the lysate and samples were vortexed for an additional 10 minutes at 4°C. Samples were centrifuged at 16,000g for 10 minutes at 4°C, creating three layers: the top layer containing polar metabolites, the bottom layer containing non-polar metabolites, and the middle layer containing protein. The top layer was transferred to a new tube, dried down in a speedvac, and subsequently stored at -80°C until they were analyzed by LC-MS.

Stable isotope tracing for lipogenesis in HepG2 cells: Cells were cultured with 0.1nM or 3nM insulin (Sigma Aldrich, I9278-5ML) for 2 days as detailed above. To wash out insulin, cells were washed with EMEM (ATCC 30-2003) seven times, including: three quick washes, three 5-minute washes and a long 20 minute-wash in EMEM at 37°C. Cells were then cultured in EMEM supplemented with 1mM Sodium acetate-$^{13}C_2$ (Sigma Aldrich, 282014) and with either 0nM insulin or 1nM insulin for 36 hours. Media was replenished after 24 hours. Cells were then processed for metabolite isolation (see below).

Stable isotope tracing for gluconeogenesis in HepG2 cells: Cells were cultured with 0.1nM or 3nM insulin (Sigma Aldrich, I9278-5ML) for 2 days as detailed above. To wash out insulin, cells were washed with EMEM (ATCC 30-2003) seven times, including: three quick washes, three 5-minute washes and a long 20 minute-wash in EMEM at 37°C. Cells were washed with glucose-free RPMI (Gibco, 11879-020) then cultured in glucose-free RPMI for 3 hours. Cells were then cultured in glucose-free RPMI supplemented with 5mM sodium pyruvate-$^{13}C_3$ (Cambridge Isotope Laboratories, NC1345852) and 5mM Sodium L-lactate (Sigma Aldrich, L7022) and with either 0nM, 1nM, 10nM or 100nM insulin (Sigma Aldrich, I9278-5ML) for 16 hours. Cells were then processed for metabolite isolation (see below).

Metabolite isolation from HepG2 cells: cells were washed in ice-cold PBS (Life Technologies, AM9625), 500 μl of cold 80% MeOH (shared by the Metabolite Profiling Core Facility) was added per well of a 6-well plate and the plate was placed at -80°C for at least 15 minutes. Following the -80 incubation, the plate was scraped on dry ice and the solution was transferred to a 1.5ml tube and then vortexed for 5 minutes. To remove cellular debris, the samples were centrifuged at maximum speed for 10 minutes at 4°C and the supernatant was transferred to a new 1.5mL tube on dry ice. To remove solvents, the samples were lyophilized using Refrigerated CentriVap Benchtop Vacuum Concentrator connected to a CentriVap-105 Cold Trap (Labconco). Metabolite pellets were re-suspended in LC-MS grade water (ThermoFisher) and vortexed for 10 minutes at 4°C. Samples were centrifuged at 16,000g for 10 minutes at 4°C and supernatant was moved into LC-MS vials. Liquid Chromatography and Mass Spectrometry was performed by the Whitehead metabolomics core.

Immunofluorescence

HepG2 cells, human liver spheroids, human primary adipocytes and human tissue liver sections were fixed in 4% PFA (VWR, BT140770-10X10) in PBS (Life Technologies, AM9625) for 10 minutes at room temperature. Cells were washed three times for 5 minutes in PBS, permeabilized with 0.5% TritonX100 (Sigma Aldrich, X100) in PBS, washed three times for 5 minutes in PBS, and then blocked with 4% IgG-free BSA (VWR, 102643-516) for 15-60 minutes at room temperature. Afterwards, the cells were incubated with 1:500 or 1:1000 primary antibody (specific antibodies listed below) in 4% IgG-free BSA in PBS at 4˚C overnight. The next day, cells were washed three times with PBS and incubated with 1:500 or 1:1000 secondary

antibodies (specific antibodies listed below) in 4% IgG-free BSA at room temperature for 1hr covered in foil. Cells were washed three times with PBS for 5 minutes. DNA was stained using 1:5000 Hoechst (Thermo Fischer Scientific, 3258) in PBS for 5 minutes at RT. Cells were washed three times with PBS for 5 minutes, stored at 4˚C until imaging. For tissue sections, samples were mounted using Vectashield mounting media (Vector Laboratories, Inc, H-1000). LSM880 or LSM980 microscope with Airyscan detector (ZEISS) was used for image acquisition. Images were then processed using Fiji/ImageJ v2.1.0/153c.
Primary antibodies used were anti-insulin receptor beta (Cell Signaling, 23413), anti-NRF2 antibody (Abcam, ab62352, 1:500 dilution), anti-cytokeratin 18 (CK18) (Abcam, ab668, 1:500 dilution), anti-PI3K (Abcam, ab135253, ab62352, 1:500 dilution), anti-AKT (Cell Signaling, 2920, ab62352, 1:500 dilution), anti-clathrin (Abcam, ab24578, ab62352, 1:500 dilution), anti-LAMP1 (abcam, ab25630, ab62352, 1:500 dilution), and anti-EEA1 (Abcam, ab70521, ab62352, 1:500 dilution), anti-pIRS1 (Abcam, ab4873, ab62352, 1:1000 dilution), anti-perilipin (Sigma, P1873, 1:500 dilution). Secondary antibodies used were Alexa Fluor 488 goat anti-rabbit IgG (Thermo Fischer Scientific, A11008), Alexa Fluor 647 goat anti-rabbit IgG (Thermo Fischer Scientific, A21244), Alexa Fluor 568 goat anti-mouse IgG (Thermo Fischer Scientific, A11031).
Images were acquired at LSM880 or LSM980 Microscope with Airyscan detector with 63x objective using Zen Black software (ZEISS) at the W.M. Keck Microscopy Facility, MIT. Images were then processed using Fiji/ImageJ v2.1.0/153c. Scale bars were determined using Fiji/ImageJ v2.1.0/153c and, when scale bars were obscured by fluorescence intensity, a black background was added to improve visibility.

Live-cell imaging
Cells expressing endogenous IR tagged with GFP were grown on 35 mm glass bottom dishes (MatTek Corporation, P35G-1.5-20-C). Cells were imaged at 37°C using the LSM880 or LSM980 Microscope with Airyscan detector with 63x objective and Zen Black software (ZEISS) at the W.M. Keck Microscopy Facility, MIT. Images were then processed using Fiji/ImageJ v2.1.0/153c.

ROS staining and live-cell imaging
After culturing the HepG2 cells with physiological insulin concentrations or pathological insulin concentrations or with TNFa or high nutrients for 2 days, media was removed and cells were cultured with ROS Deep Red Stock Solution (Abcam, ab186029) diluted to 1X in Dulbecco's PBS (Gibco, 14040-133). Cells were incubated at 37°C with 5% CO_2 in a humidified incubator for 30 minutes. Cells were imaged at 37°C using the LSM880 Microscope with Airyscan detector with 63x objective and Zen Black software (ZEISS) at the W.M. Keck Microscopy Facility, MIT. Images were then processed using Fiji/ImageJ v2.1.0/153c.

RNA FISH
Pipettes and laboratory bench were treated with RNaseZap (Life Technologies, AM9780). Cells were fixed with 4% PFA (VWR, BT140770-10X10) in PBS (Life Technologies, AM9625) for 10 minutes at RT. Cells were washed three times with PBS for 5 minutes. Cells were permeabilized with 0.5% TritonX100 (Sigma Aldrich, X100) in 1X RNasefree PBS (Invitrogen, AM9625) for 10 minutes at room temperature. Cells were washed three times with RNase-free PBS for 5

minutes. Cells were washed once with 20% Stellaris RNA FISH Wash Buffer A (Biosearch Technologies, Inc., SMF-WA1-60), 10% Deionized Formamide (EMD Millipore, S4117) in RNase-free water (Life Technologies, AM9932) for 5 minutes at room temperature. Cells were then hybridized with 90% Stellaris RNA FISH Hybridization Buffer (Biosearch Technologies, SMF-HB1-10), 10% Deionized Formamide, 12.5 µM Stellaris RNA FISH probes designed to hybridize intronic regions of each transcript (*FASN, SREBF1,* and *TIMM22*; probes listed below*)*. Hybridization was performed overnight at 37°C. Cells were then washed twice with Wash Buffer A for 30 minutes at 37°C and once with Stellaris RNA FISH Wash Buffer B (Biosearch Technologies, SMF-WB1-20) for 5 minutes at room temperature. Images were acquired at LSM880 or LSM980 Microscope with Airyscan detector with 63x objective using Zen Black software (ZEISS) at the W.M. Keck Microscopy Facility, MIT.
Stellaris® FISH Probes, Custom Assay with TAMRA Dye (LGC Bioserch, SMF-1001-5).

Imaging analyses

Fiji/ImageJ v2.1.0/153c was used to quantify IR fluorescence intensity per cell for the IR antibody validation experiment. With the polygon selection tool, a polygon was drawn around a cell outline. The average fluorescence intensity in the polygon (= in the cell) was determined using the measure tool on Fiji/ImageJ v2.1.0/153c. Background was then subtracted by a threshold determined by averaging the background intensity in a rectangular region outside of the cells.
To manually quantify IR fluorescent signal in puncta and clusters, Fiji/ImageJ v2.1.0/153c was used. A circle or an oval was drawn around IR puncta using the oval selection tool and the average fluorescence intensity in the circle or oval (= in the puncta) was determined using the measure tool on Fiji. Background was then subtracted as previously described. To quantify IR fluorescent signal in puncta and clusters in various cellular compartments, we identified the location of the plasma membrane, cytoplasm, and nucleus as follows. The plasma membrane location was identified based on IR immunofluorescence signal, IR-GFP fluorescent signal or CK18 immunofluorescence signal or cell edge. The nucleus was determined by the Hoechst stain for immunofluorescence and tc-PALM experiments. For IR-GFP experiments, Hoechst dye could not be used, because of bleed-through of the Hoechst fluorescence into the GFP channel confounded the identification of IR-GFP puncta. In these cases, the nuclear outline was inferred based on the very clear IR signal difference between the nucleus and the cytoplasm.
To computationally measure IR fluorescent signal in puncta and clusters, Airyscan images from all conditions were maximally-projected in the z-plane and background subtracted by a threshold determined by averaging the background intensity in a rectangular region outside of the cells. For segmenting IR puncta, the images were first subtracted by a median-filtered image (10 px) and then subjected to a Laplace of Gaussian filter (sigma=1). Filtered images were then thresholded on signal intensity (intensity > mean image intensity + 2*standard deviation of image intensity). Thresholded binary images were then subjected to a morphological opening operation with a 3x3 filled structuring element to remove small objects. The mean intensity of the background-subtracted raw image was then measured for each segmented puncta (c-in), and background intensity (c-out) was calculated from the mean intensity of an inverted mask of the called puncta.

To quantify pIRS1 fluorescent signal in IR puncta/clusters manually, Airyscan images were opened on Fiji as composite images. A circle or an oval was drawn around IR puncta using the oval selection tool and the average fluorescence intensity of pIRS1 in the circle or oval (= in the puncta/clusters) was determined using the measure tool on Fiji.
To quantify pIRS1 fluorescent signal in IR puncta/clusters computationally using Fiji, 3D object counter tool was used. Briefly, images were thresholded on signal intensity and IR puncta/clusters were identified. 3D object counter tool then determined the intensity pIRS1 channel in the identified IR puncta. pIRS1 signal intensity was then background subtracted by a threshold determined by averaging the background intensity in a rectangular region outside of the cells.
To estimate the number of IR puncta at the plasma membrane, cytoplasm, nucleus in an entire cell, IR puncta were initially counted at various cellular locations in a cell slice using Fiji as described above. The number obtained from the cell slice was then multiplied based on the estimated surface area of the plasma membrane, volume of the cytoplasm, volume of the nucleus or volume of the entire cell, which were obtained considering the length and width of the cell under investigation and the estimated height (~5μm).
Quantification of the ROS dye fluorescence intensity per cell was performed using Fiji. Using the polygon selection tool on Fiji/ImageJ v2.1.0/153c, a polygon was drawn around a cell outline, which was identified by looking at the IR-GFP channel. The average ROS dye fluorescence intensity in the polygon (= in the cell) was determined using the measure tool on Fiji/ImageJ v2.1.0/153c.
Fusion, fission, or deformation events were identified in time-lapse images of the endogenously-tagged IR-GFP HepG2 line. To confirm *bona fide* deformation, fusion or fission events, we quantified the total IR intensity before and after the event as a product of IR fluorescence intensity and area of the IR puncta. The total intensity was conserved in *bona fide* deformation, fusion and fission events.

ELISA

PathScan® Total Insulin Receptor β Sandwich ELISA kit (Cell Signaling, 7069) was used to quantify insulin receptor levels, PathScan Phospho-Akt2 (Ser474) and Total Akt2 Sandwich ELISA kits were used to quantify AKT2 levels (Cell Signaling Technology, #7048 and #7046, respectively) as per the manufacturer's instructions, by colorimetric reading at 450 nm on a Thermo Fisher Multiskan Go plate reader.

Chromatin immunoprecipitation-sequencing (ChIP-seq)

ChIP-seq experiments were performed by the Center for Functional Cancer Epigenetics (CFCE) at the Dana-Farber Cancer Institute. For ChIP-seq analysis, cells were cross-linked with 2mM DSG (VWR, PI20593) for 45 minutes at room temperature followed by fixation for 10 minutes with 1% formaldehyde (Tousimis Research Corporation, 1008A) at room temperature on a shaker at 850rpm. Crosslinked nuclei were quenched with 0.125M glycine (Sigma Aldrich, G7126) for 5 minutes at room temperature and washed with PBS (Life Technologies, AM9625) that contained protease inhibitor (Roche, 11836170001) and HDAC inhibitor Sodium Butyrate. After fixation, pellets were resuspended in 200μl of 1% SDS, 50mM Tris-HCl pH 8, 10mM EDTA

and sonicated in 1ml AFA fiber millitubes (Covaris, 520135) for 25 minutes using a Covaris E220 instrument (setting: 140 peak incident power, 5% duty factor and 200 cycles per burst) 600seconds per sample. Chromatin was diluted 5 times with ChIP Dilution buffer (1%Triton X-100, 2mM EDTA pH 8, 150mM NaCl, 20mM Tris-HCl pH 8) and was immunoprecipitated with 10µg of primary antibody against IR (Bethyl, A303-712A) and Dynabeads® Protein A/G (Thermo Fisher, 10015D). ChIP-seq libraries were constructed using NEBNext UltraTM II kit (NEB, E7645S) according to the manufacturer's specifications. 75-bp paired-end reads were sequenced on a NextSeq instrument. 75-bp single-end reads were sequenced on an Illumina NextSeq instrument.

MED1 ChIP-seq was used from GEO: GSM2040029
RPB1 ChIP-seq was used from GEO: GSM2864931 *(14)*

ChIP-seq analysis
ChIP-seq bioinformatics analysis for insulin receptor was performed on the Whitehead High-Performance Computing Facility using the nf-core ChIP-seq pipeline v1.2.1[112] with Nextflow v20.04.1. Quality control of .fastq files was performed with FastQC v0.11.9. Trim Galore! v0.6.4_dev was used to trim low quality reads. Alignment was performed against the hg19 genome assembly using BWA v0.7.17-r1188[113]. Peak calling was performed using MACS2 v2.2.7.1[114]. Preseq v2.0.3[115] and MultiQC v1.9 were used for quality control. Browser tracks were prepared to represent reads per million per basepair (rpm/bp).

RT-qPCR and RNA-sequencing
RNA was extracted using TRIzol™ reagent (Thermo Fisher Scientific, 15596026) following the manufacturer's instructions. cDNA synthesis was performed using qScript cDNA Supermix (QuantaBio, 95048-500) according to the manufacturer's instructions, using 1000ng RNA as starting material.
qPCR was performed on a Thermo Fisher Scientific QuantStudio 6 machine using Fast SYBR™ Green Master Mix (Thermo Fisher, 4385618) and primers (listed below) according to the manufacturer's instructions. Expression data is presented after calculating the relative expression compared with the housekeeping gene RPLP0, using the equation Relative Quantification (RQ) = 100/(2ˆ(Target Gene Ct – RPLP0 Ct). When data is reported relative to a sample condition, the condition of reference was set as 1 and the data of the other conditions were reported as a ratio (condition/condition of reference).
RNA sequencing was performed by the Whitehead Institute Genome Technology Core. Libraries were prepared using the KAPA HyperPrep stranded RNA kit (Roche, KK8540) following manufacturer's instructions. Samples were sequenced on a HiSeq2500 in High-Output mode generating 50 bases, single-end reads.

RT-qPCR primers

RPLP0_qF	gcagcatctacaaccctgaag
RPLP0_qR	gcagacagacactggcaaca
FASN_qF	CCGAGACACTCGTGGGCTA
FASN_qR	CTTCAGCAGGACATTGATGCC

PCK1_qF GCTGGTGTCCCTCTAGTCTATG
PCK1_qR GGTATTTGCCGAAGTTGTAG

RNA-sequencing analysis

RNA-sequencing (RNA-seq) bioinformatics analysis was performed on the Whitehead High-Performance Computing Facility using the nf-core RNA-seq pipeline v1.4.2 [112] with Nextflow v20.04.1. Quality control of .fastq files was performed with FastQC v0.11.8. The reads were single-end and the strandedness was set to reverse. Low quality sequences were trimmed using Trim Galore! v0.6.4. Alignment was performed against the hg19 genome assembly using STAR v2.6.1d [116] and duplicates were marked using Picard MarkDuplicates v2.21.1. Quantification of transcripts was performed using featureCounts v1.6.4[117]. Differential expression analysis was performed using edgeR v3.26.5[118]. deepTools v3.3.1[119], dupRadar v1.14.0[120], Qualimap v.2.2.2-dev[121], and MultiQC v1.7 were used for quality control.

Functional profiling of RNA-sequencing

For functional profiling of RNA-sequencing, differentially expressed genes (based on adjusted p-value < 0.05 and no log2FC cut-off) from the RNA-seq experiment were uploaded to the online version of g:Profiler[122]. Over-representation analysis was performed using g:GOSt selecting Homo sapiens (Human) as organism and treating the query set as unordered. The selected statistical domain scope was “Only annotated genes” and the significance threshold (g:SCS) was set to 0.05. Significant KEGG pathways were selected for visual representation.

Time-correlated photoactivation localization microscopy (tc-PALM)

Widefield, live-cell, super-resolution imaging was performed in a photo-activation localization microscopy (PALM) approach using a Nikon Eclipse Ti microscope with a 100× oil immersion objective. The 405nm and 561nm laser beams were combined in an external platform with customized power densities to image Dendra2-tagged molecules as previously reported[62]. Cells were cultured on imaging dishes (MatTek, P35G-1.5-20-C) and then imaged while maintaining both the temperature at 37°C with a temperature-controlled platform and the level of CO_2 at 5% with Leibovitz's L-15 Medium with no phenol red (Thermo Fisher, 21083027). During each imaging cycle, a 2400-frame video stream including a $(256\ \text{pixel})^2$ region of interest (ROI) was recorded in 20Hz acquisition rate with the EM-gain setting as 1000 on an Andor iXon Ultra 897 EMCCD. Each pixel conjugates with a $(160\text{nm})^2$ area on the sample side. After PALM imaging, the Hoechst-stained nuclei of the same ROI were imaged using a stronger 405nm excitation through DAPI filter. For insulin stimulation, cells were first imaged in 1.5ml insulin-free L-15 medium for 15-20 minutes. Afterwards, the cells were stimulated with insulin by adding 1.5ml of prewarmed and freshly-made L-15 medium, 6nM insulin to the same dish containing the original 1.5ml of insulin-free L-15 medium. Following a 5-minute wait, the cells were imaged for 15-20 minutes. For the insulin-unstimulated condition, cells were imaged in 1.5ml insulin-free L-15 Medium for 15-20 minutes. For the insulin-treated condition, 1.5ml fresh-made, prewarmed L-15 medium containing 2× insulin (6nM) was directly added to the same dish while it was still on the platform, followed by a 5-minute wait, then cells were imaged for 15-20 minutes.

tc-PALM Analysis

Detection localization: For each frame of a raw image, Gaussian particles were identified by pixelwise test of hypotheses, whose peak positions were individually fitted at subpixel resolution by maximum-likelihood regression with Gauss-Newton method[123]. An additional deflation loop was performed to avoid missing dimmer particles when they were overshadowed by neighboring brighter ones. This multi-particle detection localization procedure has been integrated in a published, open-source MATLAB software called MTT[123].

Spatial clustering. DBSCAN and "manual selection" hybridized approach was applied to group spatially clustered detections via the qSR software[124]. Firstly, DBSCAN was performed to generalize a proposal map of spatially clustered detections. Given that IR clusters can be tiny and transient, a "loose" parameter setting was used when performing the DBSCAN (length-scale = 120nm, N_min = 4). This parameter combination was determined by comparing the rendered, super-resolved reconstructions with the color-coded cluster maps until the clustering results visually make sense for most ROIs. Second, individual clusters were manually selected based on the clustering proposal map from previous step. Custom MATLAB code was used to reconstruct the IR distribution of each ROI superposed with the corresponding nuclei image, which was further cross-compared with the corresponding cluster map to determine which region each cluster belongs to (i.e., plasma-membrane, cytoplasm, or nucleus).

Temporal clustering: For each spatial cluster, time-correlated PALM (tc-PALM) analysis was performed along the time axis to extract the truly colocalized, time-correlated multi-molecule bursting events. The lifetime of a burst is simply defined as the timespan from the first to the last detections. More details about the quantitative validations and statistics of tc-PALM analysis can be found in Supplemental Text and Supplementary Figure 13.

Cryo-Immuno Electron Microscopy (EM)[125]

The cells were fix mildly Using PLP (paraformaldehyde / lysine / sodium periodate) fixative for 4 hours. Cells were pelleted. Infused with a cryo-protectant for at least one hour (PVP/ sucrose). Blocks were mounted onto cryo-pins, and snap frozen in liquid nitrogen cooled ethane. Ultrathin sections were cut at -140 degrees C with a Leica UC7 equipped with a FC7 cryo-stage using a glass knife, and immunolabeled, stained and embedded using the Tokuyasu technique. The material was examined using a Hitach 7800.

Antibody used were anti-insulin receptor beta antibody (Cell Signaling, 23413) for WT HepG2 cells or anti-GFP (Abcam, ab6556) for HepG2 cells expressing endogenous IR tagged with GFP).

Protein purification

Human cDNA encoding the beta subunit of the insulin receptor (IRb; residues 763-1382) was cloned into a mammalian expression vector. The base vector was engineered to include sequences encoding an N-terminal FLAG tag followed by mCherry and a 14 amino acid linker sequence "GAPGSAGSAAGGSG." cDNA sequences were inserted in-frame following the linker sequence using NEBuilder HiFi DNA Assembly Master Mix (NEB, E2621S). The expression construct was subjected to Sanger sequencing to confirm the sequence.

For protein expression, IRb-mcherry plasmid was transfected into HEK293T cells (ATCC, CRL-3216) using Polyethylenimine (Fisher Scientific, NC1014320). Cells were cultured for 72 hours,

scraped off the plate, and washed with ice-cold PBS (Life Technologies, AM9625). Cells were centrifuged at 500xg for 5 minutes and the cell pellet was stored at -80ºC.
The cell pellet was resuspended in 35ml Lysis Buffer (20mM HEPES pH7.4, 150mM NaCl, 1mM EDTA, 0.5% NP40, with fresh inhibitors and 1mM DTT). Cell lysate was rocked for 30 minutes at 4°C and spun down at 12,000 x g 15 minutes. 35ml of supernatant was removed to a fresh tube and centrifuged again if cloudy. 300µl of washed Anti-Flag M2 magnetic beads (Sigma Aldrich, M8823) was added to the lysate, which was then rotated overnight at 4°C. The next day beads were pelleted at 500rpm for 5 minutes, washed with 35ml BD Buffer (10mM HEPES, 450mM NaCl, 5% glycerol with fresh inhibitors), transferred to Eppendorf tube. Tubes were then placed in a magnetic rack to pellet beads and washed 3 – 5 times with BD Buffer, with resuspension of the pellet for each wash. Elution was performed overnight with 500µl Dialysis Buffer (50mM HEPES, 150mM NaCl, 5mM MgCl2, 5% glycerol) plus 50µl Flag peptide (5mg/ml stock solution). The next day the sample was eluted with the magnetic rack and washed with 250µl Dialysis buffer with no peptide. The sample was dialyzed with 500ml buffer, which was changed 1 to 2 times at 4°C.

Insulin binding assay
Insulin-sensitive and insulin-resistant cells were washed in EMEM for 30 minutes as previously described, incubated on ice at 4ºC for 30 minutes and treated with 3nM insulin for 60 minutes on ice at 4ºC with gentle agitation. Cells were washed five times with ice-cold PBS. TrypLE was added to the wells and cells were gently detached from the wells and added to a 1.5ml tube. Cells were incubated at 37ºC for 10 minutes with gentle shaking. Cells were then pelleted and supernatant was collected into new 1.5ml tube and incubated O/N at 37ºC with gentle shaking. Samples were then used for proteomics.

Proteomics
SDB-RPS extraction of tryptic peptides from cell cultures. For the preparation of tryptic peptides for mass spectrometry, 50 µl protein extract was reduced by adding 2 µl 250 mM TCEP and the solution was incubated for 15 min at 55 ºC. Then 5 µl 500 mM IAA was added and the proteins were alkylated for 30 min at RT. The extracts were acidified (0.5 % w/v TFA) and extracted by using custom-made SDB-RPS tips (CDS Analytical, Oxford, PA, USA) following the descriptions by Rappsilber et al. (2007)[126]. Peptides were eluted from SBD-RPS filters with 50 % (v/v) ACN and 5 % (v/v) ammonium hydroxide, dried in a lyophilizer, and taken up in 20 µL 0.2 % (v/v) formic acid. Insoluble material was removed from reconstituted peptide solutions by centrifugation for 10 min at 20,000g at 4°C prior to analysis with mass spectrometry.
nanoLC-MS and data analysis. The LC-MS/MS analysis was performed using an Easy-nLC 1200 system connected to an Orbitrap Exploris 480 mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) equipped with an Easy Spray ESI source together with FAIMS for the ionization of eluting fractions. Peptide separation, collection of MS1 and MS2 profiles, identification and quantitation of protein abundances, and statistical data analyses were carried out as described by Schulte et al. (2019)[127].

Statistics and Reproducibility

Statistical analysis was performed using Prism Version 9.4.0 (GraphPad, La Jolla, CA) and Excel Version 16.66.1. The statistical test used is reported in the figure legend. Data is represented as individual values and mean ± SEM or as mean ± SEM. The exact p-value is reported in the Source Data file (given the limitation of Excel, very low p-values are reported as 0). Information on the number of independent experiments performed is reported in Supplementary Table 2. All imaging experiments have been repeated independently at least two times as reported in Supplementary Table 2. All tc-PALM experiments were performed three times using biologically independent replicates and results were merged. Results reported in Figure 1 derive from three independent experiments. All imaging of IR (immunofluorescence and live-cell imaging) in HepG2 were performed three times using biologically independent samples, unless otherwise stated in Supplementary Table 2. All imaging of pIRS1 in IR clusters have been performed twice using biologically independent samples. Imaging of NRF2 and ROS have been performed twice using biologically independent samples, unless otherwise stated in Supplementary Table 2. Imaging of IR in human primary hepatocytes was performed once with biologically independent samples. Imaging of IR in human primary adipocytes was performed twice with biologically independent samples. Western blotting, ELISA and RT-qPCR experiments have been performed ≥2 times using biologically independent samples. RNA-seq, metabolomics experiments and proteomics experiments were performed once with biologically independent samples. Quantification of the number of IR molecules was performed twice. Survival assay was performed three times using biologically independent samples.

Schematics

BioRender was used to make the graphics reported in the figures (BioRender.com).

Data availability: ChIP-sequencing and RNA-sequencing data generated in this study have been deposited in GEO: GSE181096
https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE181096

Code availability: All codes are available to researchers by the corresponding authors. Email: young@wi.mit.edu and tlee@wi.mit.edu.

References

1. Batista TM, Haider N, Kahn CR. Defining the underlying defect in insulin action in type 2 diabetes. *Diabetologia* **64**, 994-1006 (2021).

2. James DE, Stockli J, Birnbaum MJ. The aetiology and molecular landscape of insulin resistance. *Nat Rev Mol Cell Biol*, (2021).

3. Freychet P, Roth J, Neville DM, Jr. Insulin receptors in the liver: specific binding of (125 I)insulin to the plasma membrane and its relation to insulin bioactivity. *Proc Natl Acad Sci U S A* **68**, 1833-1837 (1971).

4. Kasuga M, Zick Y, Blithe DL, Crettaz M, Kahn CR. Insulin stimulates tyrosine phosphorylation of the insulin receptor in a cell-free system. *Nature* **298**, 667-669 (1982).

5. Yu KT, Czech MP. Tyrosine phosphorylation of the insulin receptor beta subunit activates the receptor-associated tyrosine kinase activity. *J Biol Chem* **259**, 5277-5286 (1984).

6. Yu KT, Pessin JE, Czech MP. Regulation of insulin receptor kinase by multisite phosphorylation. *Biochimie* **67**, 1081-1093 (1985).

7. White MF, Maron R, Kahn CR. Insulin rapidly stimulates tyrosine phosphorylation of a Mr-185,000 protein in intact cells. *Nature* **318**, 183-186 (1985).

8. Pronk GJ, McGlade J, Pelicci G, Pawson T, Bos JL. Insulin-induced phosphorylation of the 46- and 52-kDa Shc proteins. *J Biol Chem* **268**, 5748-5753 (1993).

9. Furtado LM, Somwar R, Sweeney G, Niu W, Klip A. Activation of the glucose transporter GLUT4 by insulin. *Biochem Cell Biol* **80**, 569-578 (2002).

10. Goldfine ID, Smith GJ. Binding of insulin to isolated nuclei. *Proc Natl Acad Sci U S A* **73**, 1427-1431 (1976).

11. Hancock ML*, et al.* Insulin Receptor Associates with Promoters Genome-wide and Regulates Gene Expression. *Cell* **177**, 722-736 e722 (2019).

12. Nelson JD, LeBoeuf RC, Bomsztyk K. Direct recruitment of insulin receptor and ERK signaling cascade to insulin-inducible gene loci. *Diabetes* **60**, 127-137 (2011).

13. Fagerholm S, Ortegren U, Karlsson M, Ruishalme I, Stralfors P. Rapid insulin-dependent endocytosis of the insulin receptor by caveolae in primary adipocytes. *PLoS One* **4**, e5985 (2009).

14. Hall C, Yu H, Choi E. Insulin receptor endocytosis in the pathophysiology of insulin resistance. *Exp Mol Med* **52**, 911-920 (2020).

15. Iraburu MJ, Garner T, Montiel-Duarte C. Revising Endosomal Trafficking under Insulin Receptor Activation. *Int J Mol Sci* **22**, (2021).

16. Amaya MJ*, et al.* The insulin receptor translocates to the nucleus to regulate cell proliferation in liver. *Hepatology* **59**, 274-283 (2014).

17. Czech MP. Insulin action and resistance in obesity and type 2 diabetes. *Nat Med* **23**, 804-814 (2017).

18. DeFronzo RA*, et al.* Type 2 diabetes mellitus. *Nat Rev Dis Primers* **1**, 15019 (2015).

19. Petersen MC, Shulman GI. Mechanisms of Insulin Action and Insulin Resistance. *Physiol Rev* **98**, 2133-2223 (2018).

20. Ter Horst KW*, et al.* Hepatic Insulin Resistance Is Not Pathway Selective in Humans With Nonalcoholic Fatty Liver Disease. *Diabetes Care* **44**, 489-498 (2021).

21. Furnica RM, Istasse L, Maiter D. A severe but reversible reduction in insulin sensitivity is observed in patients with insulinoma. *Ann Endocrinol (Paris)* **79**, 30-36 (2018).

22. Lemmer IL, Willemsen N, Hilal N, Bartelt A. A guide to understanding endoplasmic reticulum stress in metabolic disorders. *Mol Metab* **47**, 101169 (2021).

23. Hurrle S, Hsu WH. The etiology of oxidative stress in insulin resistance. *Biomed J* **40**, 257-262 (2017).

24. Su X*, et al.* Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science* **352**, 595-599 (2016).

25. Ditlev JA*, et al.* A composition-dependent molecular clutch between T cell signaling condensates and actin. *Elife* **8**, (2019).

26. Zamudio AV*, et al.* Mediator Condensates Localize Signaling Factors to Key Cell Identity Genes. *Mol Cell* **76**, 753-766 e756 (2019).

27. Jaqaman K, Ditlev JA. Biomolecular condensates in membrane receptor signaling. *Curr Opin Cell Biol* **69**, 48-54 (2021).

28. Case LB, Ditlev JA, Rosen MK. Regulation of Transmembrane Signaling by Phase Separation. *Annu Rev Biophys* **48**, 465-494 (2019).

29. Zeng M*, et al.* Reconstituted Postsynaptic Density as a Molecular Platform for Understanding Synapse Formation and Plasticity. *Cell* **174**, 1172-1187 e1116 (2018).

30. Lin C*, et al.* Receptor tyrosine kinases regulate signal transduction through a liquid–liquid phase separated state. *bioRxiv*, (2019).

31. Huang WYC*, et al.* A molecular assembly phase transition and kinetic proofreading modulate Ras activation by SOS. *Science* **363**, 1098-1103 (2019).

32. Zeng M, Shang Y, Araki Y, Guo T, Huganir RL, Zhang M. Phase Transition in Postsynaptic Densities Underlies Formation of Synaptic Complexes and Synaptic Plasticity. *Cell* **166**, 1163-1175 e1112 (2016).

33. Gao XK*, et al.* Phase separation of insulin receptor substrate 1 drives the formation of insulin/IGF-1 signalosomes. *Cell Discov* **8**, 60 (2022).

34. Banani SF, Lee HO, Hyman AA, Rosen MK. Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol* **18**, 285-298 (2017).

35. Cho WK*, et al.* RNA Polymerase II cluster dynamics predict mRNA output in living cells. *Elife* **5**, (2016).

36. Nair SJ*, et al.* Phase separation of ligand-activated enhancers licenses cooperative chromosomal enhancer assembly. *Nat Struct Mol Biol* **26**, 193-203 (2019).

37. Li W*, et al.* Biophysical properties of AKAP95 protein condensates regulate splicing and tumorigenesis. *Nat Cell Biol* **22**, 960-972 (2020).

38. Niaki AG*, et al.* Loss of Dynamic RNA Interaction and Aberrant Phase Separation Induced by Two Distinct Types of ALS/FTD-Linked FUS Mutations. *Mol Cell* **77**, 82-94 e84 (2020).

39. Boothe T*, et al.* Inter-domain tagging implicates caveolin-1 in insulin receptor trafficking and Erk signaling bias in pancreatic beta-cells. *Mol Metab* **5**, 366-378 (2016).

40. Gustavsson J*, et al.* Localization of the insulin receptor in caveolae of adipocyte plasma membrane. *FASEB J* **13**, 1961-1971 (1999).

41. S K, P T, AD A. Histopathology of Liver in Diabetes Mellitus - An Autopsy Study. *International Journal of Scientific Study* **4**, (2016).

42. Madiraju AK*, et al.* Metformin suppresses gluconeogenesis by inhibiting mitochondrial glycerophosphate dehydrogenase. *Nature* **510**, 542-546 (2014).

43. Butkowski EG, Jelinek HF. Hyperglycaemia, oxidative stress and inflammatory markers. *Redox Rep* **22**, 257-264 (2017).

44. KAYAL R., BHATTACHARJEE D., BANERJEE S.k., B. K. RELATIONSHIP BETWEEN PROTEIN CARBONYLATION AND TYPE-2 DIABETES MELLITUS IN KOLKATA, INDIA. *International Journal of Advances in Science Engineering and Technology* **6**, 2321 –9009 (2018).

45. Batandier C*, et al.* The ROS production induced by a reverse-electron flux at respiratory-chain complex 1 is hampered by metformin. *J Bioenerg Biomembr* **38**, 33-42 (2006).

46. Yadollah S, Kazemipour N, Bakhtiyari S, Nazifi S. Palmitate-induced insulin resistance is attenuated by Pioglitazone and EGCG through reducing the gluconeogenic key enzymes expression in HepG2 cells. *J Med Life* **10**, 244-249 (2017).

47. Sefried S, Haring HU, Weigert C, Eckstein SS. Suitability of hepatocyte cell lines HepG2, AML12 and THLE-2 for investigation of insulin signalling and hepatokine gene expression. *Open Biol* **8**, (2018).

48. Song SH, McIntyre SS, Shah H, Veldhuis JD, Hayes PC, Butler PC. Direct measurement of pulsatile insulin secretion from the portal vein in human subjects. *J Clin Endocrinol Metab* **85**, 4491-4499 (2000).

49. Storch M, Rössle M, Kerp L. Pulsatile Insulinsekretion in die Pfortader bei Leberzirrhose. *Med Wochenschr* **118**, 134-138 (1993).

50. Porksen N*, et al.* Human insulin release processes measured by intraportal sampling. *Am J Physiol Endocrinol Metab* **282**, E695-702 (2002).

51. Watanabe M, Hayasaki H, Tamayama T, Shimada M. Histologic distribution of insulin and glucagon receptors. *Braz J Med Biol Res* **31**, 243-256 (1998).

52. Ramakrishnan G, Arjuman A, Suneja S, Das C, Chandra NC. The association between insulin and low-density lipoprotein receptors. *Diab Vasc Dis Res* **9**, 196-204 (2012).

53. Wu X, Ganzella M, Zhou J, Zhu S, Jahn R, Zhang M. Vesicle Tethering on the Surface of Phase-Separated Active Zone Condensates. *Mol Cell* **81**, 13-24 e17 (2021).

54. Emperador-Melero J*, et al.* PKC-phosphorylation of Liprin-alpha3 triggers phase separation and controls presynaptic active zone structure. *Nat Commun* **12**, 3057 (2021).

55. Bergeron-Sandoval LP*, et al.* Endocytic proteins with prion-like domains form viscoelastic condensates that enable membrane remodeling. *Proc Natl Acad Sci U S A* **118**, (2021).

56. Kozak M, Kaksonen M. Condensation of Ede1 promotes the initiation of endocytosis. *Elife* **11**, (2022).

57. Saad MF, Knowler WC, Pettitt DJ, Nelson RG, Mott DM, Bennett PH. Sequential changes in serum insulin concentration during development of non-insulin-dependent diabetes. *Lancet* **1**, 1356-1359 (1989).

58. Kajbaf F, De Broe ME, Lalau JD. Therapeutic Concentrations of Metformin: A Systematic Review. *Clin Pharmacokinet* **55**, 439-459 (2016).

59. Timmins P, Donahue S, Meeker J, Marathe P. Steady-state pharmacokinetics of a novel extended-release metformin formulation. *Clin Pharmacokinet* **44**, 721-729 (2005).

60. Frid A*, et al.* Novel assay of metformin levels in patients with type 2 diabetes and varying levels of renal function: clinical recommendations. *Diabetes Care* **33**, 1291-1293 (2010).

61. Friesen M., Khalil A.S., Barrasa M.I., Jeppesen J.F., Mooney D.J., R. J. Development of a physiological insulin resistance model in human stem cell–derived adipocytes. *Science Advances* **8**, 1-9 (2022).

62. Cho WK*, et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412-415 (2018).

63. Brangwynne CP*, et al.* Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324**, 1729-1732 (2009).

64. Hyman AA, Weber CA, Julicher F. Liquid-liquid phase separation in biology. *Annu Rev Cell Dev Biol* **30**, 39-58 (2014).

65. Feric M*, et al.* Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell* **165**, 1686-1697 (2016).

66. Patel A*, et al.* A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **162**, 1066-1077 (2015).

67. Brangwynne CP, Mitchison TJ, Hyman AA. Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. *Proc Natl Acad Sci U S A* **108**, 4334-4339 (2011).

68. Cisse, II*, et al.* Real-time dynamics of RNA polymerase II clustering in live human cells. *Science* **341**, 664-667 (2013).

69. Huoh YS*, et al.* Dual functions of Aire CARD multimerization in the transcriptional regulation of T cell tolerance. *Nat Commun* **11**, 1625 (2020).

70. Mukherjee SP, Lane RH, Lynn WS. Endogenous hydrogen peroxide and peroxidative metabolism in adipocytes in response to insulin and sulfhydryl reagents. *Biochem Pharmacol* **27**, 2589-2594 (1978).

71. May JM, de Haen C. Insulin-stimulated intracellular hydrogen peroxide production in rat epididymal fat cells. *J Biol Chem* **254**, 2214-2220 (1979).

72. Krieger-Brauer HI, Kather H. Human fat cells possess a plasma membrane-bound H2O2-generating system that is activated by insulin via a mechanism bypassing the receptor kinase. *J Clin Invest* **89**, 1006-1013 (1992).

73. Mahadev K*, et al.* The NAD(P)H oxidase homolog Nox4 modulates insulin-stimulated generation of H2O2 and plays an integral role in insulin signal transduction. *Mol Cell Biol* **24**, 1844-1854 (2004).

74. Krieger-Brauer HI, Medda PK, Kather H. Insulin-induced activation of NADPH-dependent H2O2 generation in human adipocyte plasma membranes is mediated by Galphai2. *J Biol Chem* **272**, 10135-10143 (1997).

75. Iwakami S*, et al.* Concentration-dependent dual effects of hydrogen peroxide on insulin signal transduction in H4IIEC hepatocytes. *PLoS One* **6**, e27401 (2011).

76. Ge X, Yu Q, Qi W, Shi X, Zhai Q. Chronic insulin treatment causes insulin resistance in 3T3-L1 adipocytes through oxidative stress. *Free Radic Res* **42**, 582-591 (2008).

77. Fischer R, Maier O. Interrelation of oxidative stress and inflammation in neurodegenerative disease: role of TNF. *Oxid Med Cell Longev* **2015**, 610813 (2015).

78. Yu D*, et al.* High fat diet-induced oxidative stress blocks hepatocyte nuclear factor 4alpha and leads to hepatic steatosis in mice. *J Cell Physiol* **233**, 4770-4782 (2018).

79. Rudich A, Kozlovsky N, Potashnik R, Bashan N. Oxidant stress reduces insulin responsiveness in 3T3-L1 adipocytes. *Am J Physiol* **272**, E935-940 (1997).

80. Boden G*, et al.* Excessive caloric intake acutely causes oxidative stress, GLUT4 carbonylation, and insulin resistance in healthy men. *Sci Transl Med* **7**, 304re307 (2015).

81. Hou X*, et al.* Metformin reduces intracellular reactive oxygen species levels by upregulating expression of the antioxidant thioredoxin via the AMPK-FOXO3 pathway. *Biochem Biophys Res Commun* **396**, 199-205 (2010).

82. Huang X*, et al.* ROS regulated reversible protein phase separation synchronizes plant flowering. *Nat Chem Biol* **17**, 549-557 (2021).

83. Guillen-Boixet J*, et al.* RNA-Induced Conformational Switching and Clustering of G3BP Drive Stress Granule Assembly by Condensation. *Cell* **181**, 346-361 e317 (2020).

84. Rayman JB, Karl KA, Kandel ER. TIA-1 Self-Multimerization, Phase Separation, and Recruitment into Stress Granules Are Dynamically Regulated by Zn(2). *Cell Rep* **22**, 59-71 (2018).

85. Chong PA, Forman-Kay JD. Oxidative Inhibition of Pbp1 Phase Separation. *Biochemistry* **58**, 3057-3059 (2019).

86. Kato M, Yang YS, Sutter BM, Wang Y, McKnight SL, Tu BP. Redox State Controls Phase Separation of the Yeast Ataxin-2 Protein via Reversible Oxidation of Its Methionine-Rich Low-Complexity Domain. *Cell* **177**, 711-721 e718 (2019).

87. Yang YS*, et al.* Yeast Ataxin-2 Forms an Intracellular Condensate Required for the Inhibition of TORC1 Signaling during Respiratory Growth. *Cell* **177**, 697-710 e617 (2019).

88. Ahmed SM, Luo L, Namani A, Wang XJ, Tang X. Nrf2 signaling pathway: Pivotal roles in inflammation. *Biochim Biophys Acta Mol Basis Dis* **1863**, 585-597 (2017).

89. Peng TI, Lin MS, Jou MJ. Dual phases of respiration chain defect-augmented mROS-mediated mCa 2+ stress during oxidative insult in normal and rho 0 RBA1 astrocytes. *Oxid Med Cell Longev* **2013**, 159567 (2013).

90. Rehman T, Fought J, Solomon R. N-acetylcysteine effect on serum creatinine and cystatin C levels in CKD patients. *Clin J Am Soc Nephrol* **3**, 1610-1614 (2008).

91. Radtke KK, Coles LD, Mishra U, Orchard PJ, Holmay M, Cloyd JC. Interaction of N-acetylcysteine and cysteine in human plasma. *J Pharm Sci* **101**, 4653-4659 (2012).

92. Forman-Kay JD, Ditlev JA, Nosella ML, Lee HO. What are the distinguishing features and size requirements of biomolecular condensates and their implications for RNA-containing condensates? *RNA* **28**, 36-47 (2022).

93. La Cunza N*, et al.* Mitochondria-dependent phase separation of disease-relevant proteins drives pathological features of age-related macular degeneration. *JCI Insight* **6**, (2021).

94. Soberanes S*, et al.* Metformin Targets Mitochondrial Electron Transport to Reduce Air-Pollution-Induced Thrombosis. *Cell Metab* **29**, 503 (2019).

95. Ashabi G, Khalaj L, Khodagholi F, Goudarzvand M, Sarkaki A. Pre-treatment with metformin activates Nrf2 antioxidant pathways and inhibits inflammatory responses through induction of AMPK after transient global cerebral ischemia. *Metab Brain Dis* **30**, 747-754 (2015).

96. Prasad S*, et al.* Role of Nrf2 and protective effects of Metformin against tobacco smoke-induced cerebrovascular toxicity. *Redox Biol* **12**, 58-69 (2017).

97. Alberti S, Hyman AA. Biomolecular condensates at the nexus of cellular stress, protein aggregation disease and ageing. *Nat Rev Mol Cell Biol* **22**, 196-213 (2021).

98. Shin Y, Brangwynne CP. Liquid phase condensation in cell physiology and disease. *Science* **357**, (2017).

99. Defronzo RA. Banting Lecture. From the triumvirate to the ominous octet: a new paradigm for the treatment of type 2 diabetes mellitus. *Diabetes* **58**, 773-795 (2009).

100. Samuel VT, Shulman GI. The pathogenesis of insulin resistance: integrating signaling pathways and substrate flux. *J Clin Invest* **126**, 12-22 (2016).

101. Perry RJ*, et al.* Hepatic acetyl CoA links adipose tissue inflammation to hepatic insulin resistance and type 2 diabetes. *Cell* **160**, 745-758 (2015).

102. Kahn CR*, et al.* The syndromes of insulin resistance and acanthosis nigricans. Insulin-receptor disorders in man. *N Engl J Med* **294**, 739-745 (1976).

103. Wertheimer E, Lu SP, Backeljauw PF, Davenport ML, Taylor SI. Homozygous deletion of the human insulin receptor gene results in leprechaunism. *Nat Genet* **5**, 71-73 (1993).

104. Michael MD*, et al.* Loss of insulin signaling in hepatocytes leads to severe insulin resistance and progressive hepatic dysfunction. *Mol Cell* **6**, 87-97 (2000).

105. Lu M*, et al.* Insulin regulates liver metabolism in vivo in the absence of hepatic Akt and Foxo1. *Nat Med* **18**, 388-395 (2012).

106. Vatner DF*, et al.* Insulin-independent regulation of hepatic triglyceride synthesis by fatty acids. *Proc Natl Acad Sci U S A* **112**, 1143-1148 (2015).

107. Melvin A, O'Rahilly S, Savage DB. Genetic syndromes of severe insulin resistance. *Curr Opin Genet Dev* **50**, 60-67 (2018).

108. Tsang B, Pritisanac I, Scherer SW, Moses AM, Forman-Kay JD. Phase Separation as a Missing Mechanism for Interpretation of Disease Mutations. *Cell* **183**, 1742-1756 (2020).

109. Hayashi Y, Ford LK, Fioriti L, McGurk L, Zhang M. Liquid-Liquid Phase Separation in Physiology and Pathophysiology of the Nervous System. *J Neurosci* **41**, 834-844 (2021).

110. Henninger JE*, et al.* RNA-Mediated Feedback Control of Transcriptional Condensates. *Cell* **184**, 207-225 e224 (2021).

111. Alzamil H. Elevated Serum TNF-alpha Is Related to Obesity in Type 2 Diabetes Mellitus and Is Associated with Glycemic Control and Insulin Resistance. *J Obes* **2020**, 5076858 (2020).

112. Ewels PA*, et al.* The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol* **38**, 276-278 (2020).

113. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).

114. Zhang Y*, et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).

115. Daley T, Smith AD. Predicting the molecular complexity of sequencing libraries. *Nat Methods* **10**, 325-327 (2013).

116. Dobin A*, et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).

117. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).

118. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).

119. Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187-191 (2014).

120. Sayols S, Scherzinger D, Klein H. dupRadar: a Bioconductor package for the assessment of PCR artifacts in RNA-Seq data. *BMC Bioinformatics* **17**, 428 (2016).

121. Okonechnikov K, Conesa A, Garcia-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32**, 292-294 (2016).

122. Reimand J, Kull M, Peterson H, Hansen J, Vilo J. g:Profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res* **35**, W193-200 (2007).

123. Serge A, Bertaux N, Rigneault H, Marguet D. Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes. *Nat Methods* **5**, 687-694 (2008).

124. Andrews JO*, et al.* qSR: a quantitative super-resolution analysis tool reveals the cell-cycle dependent organization of RNA Polymerase I in live human cells. *Sci Rep* **8**, 7424 (2018).

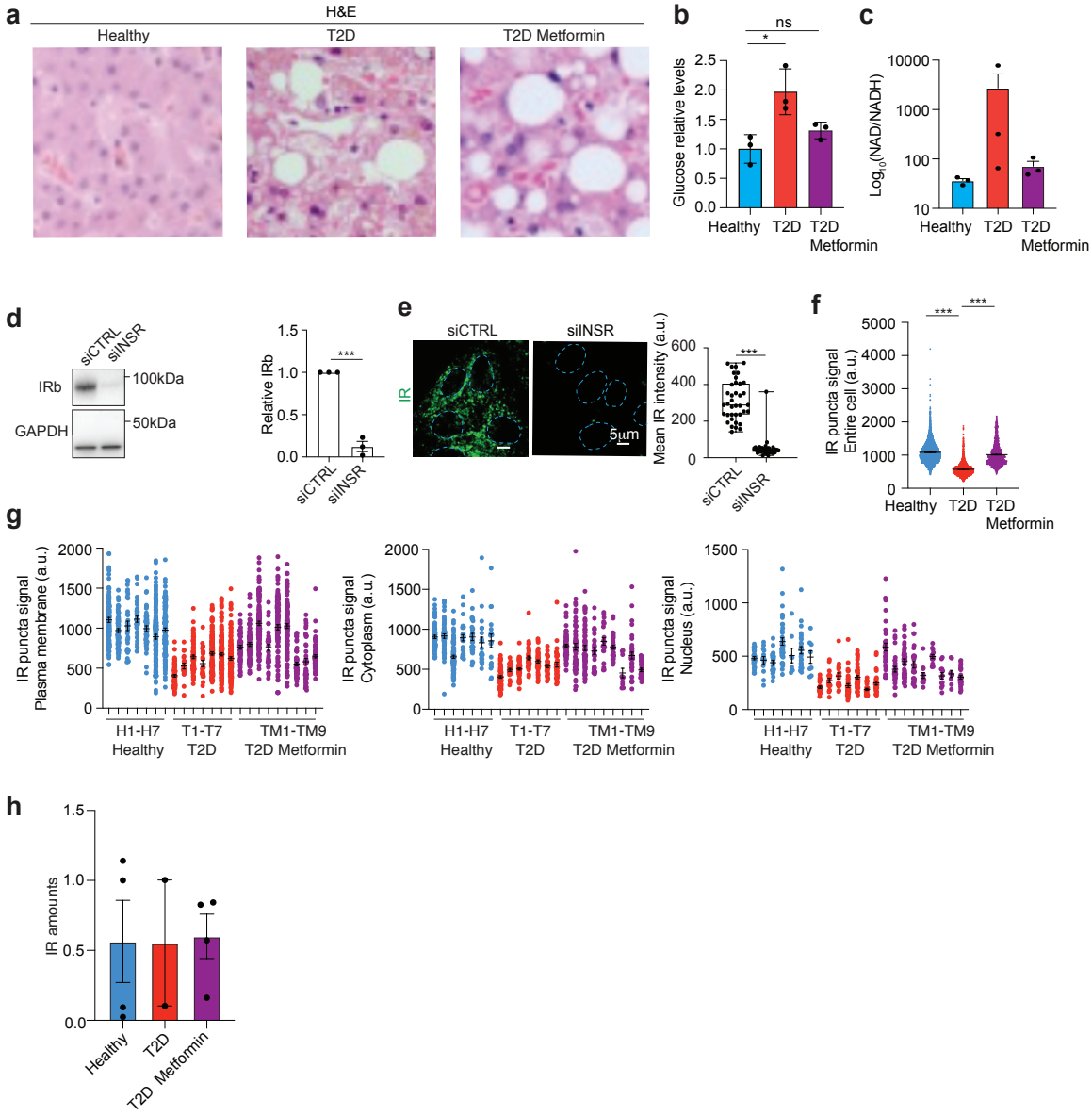
125. Tokuyasu KT. Immunochemistry on ultrathin frozen sections. *Histochem J* **12**, 381-403 (1980).

126. Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* **2**, 1896-1906 (2007).

127. Schulte F, Hasturk H, Hardt M. Mapping Relative Differences in Human Salivary Gland Secretions by Dried Saliva Spot Sampling and nanoLC-MS/MS. *Proteomics* **19**, e1900023 (2019).

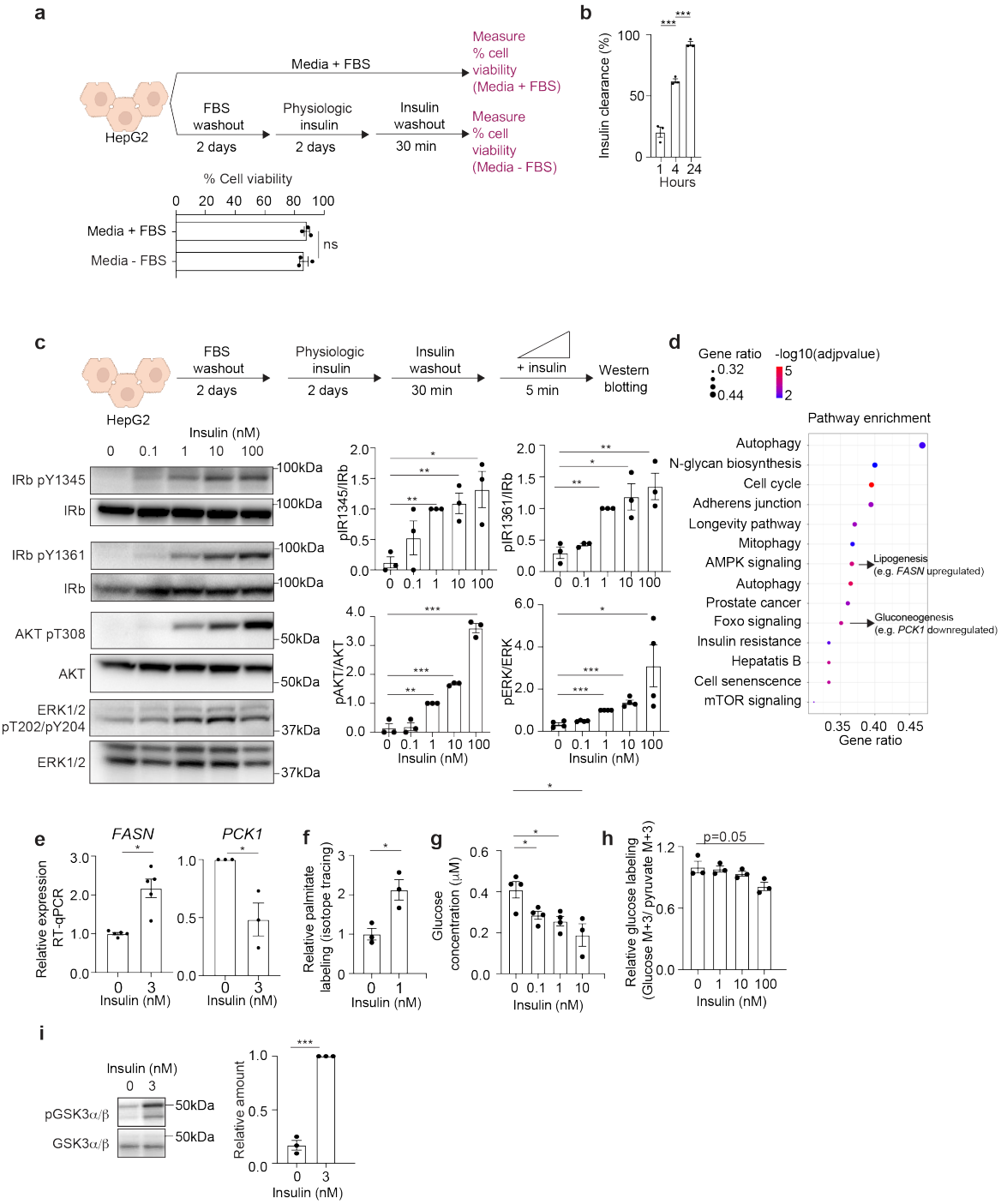
Supplementary Material

SUPPLEMENTARY FIGURES AND TABLES



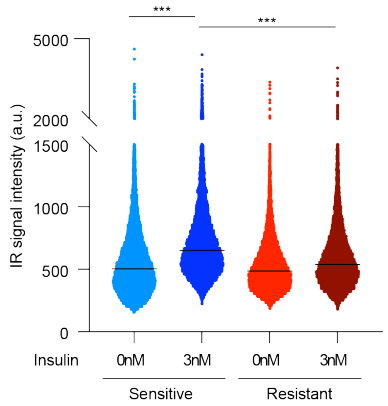
Supplementary Fig. 1. Human liver characterization and antibody validation. a, Representative hematoxylin and eosin (H&E) images of human livers from a healthy donor (Healthy), a donor with T2D (T2D) and a donor with T2D who had been treated with metformin (T2D Metformin). **b**, Quantification of relative glucose levels in livers from healthy donors (Healthy, light blue), donors with T2D (T2D, red) and donors with T2D who had been treated with metformin (T2D Metformin, purple) as determined by metabolomics. Data is represented as individual values and as mean +/- SEM. Liver samples from 3 donors were analyzed per condition. Unpaired two-sided t-test was used for statistical analysis. **c**, Quantification of NAD/NADH ratio in livers from healthy donors (Healthy, light blue), donors with T2D (T2D, red) and donors with T2D who had been treated with metformin (T2D Metformin, purple) as determined by metabolomics. Data is represented as individual values and as mean +/- SEM. Liver samples from 3 donors were analyzed per condition. **d,e,** Validation of the antibody against IR by immunoblot (d) and immunofluorescence (e) and quantification. Data is represented as individual values and as mean +/- SEM. For immunoblot validation, 3 biological replicates were analyzed per condition. For

immunofluorescence validation, 38 siCTRL cells and 36 siINSR cells were analyzed. Unpaired two-sided t-test was used for statistical analysis. siCTRL minima 142, maxima 518, centre 288, 25th percentile 235, 75th percentile 404; siINSR minima 13, maxima 360, centre 41, 25th percentile 31; 75th percentile 51. **f**, Automated quantification of IR signal in puncta in entire cells (without specifying cellular subcompartments) of healthy donors (Healthy, light blue), donors with T2D (T2D, red) and donors with T2D who had been treated with metformin (T2D Metformin, purple). Data is represented as mean +/- SEM. Number of IR puncta analyzed: Healthy 5891 puncta, T2D 3118 puncta, T2D Metformin 1271. Unpaired two-sided t-test was used for statistical analysis. **g**, Quantification of IR signal in puncta at the plasma membrane, cytoplasm and nucleus in healthy donors (H1-H7, Healthy, light blue), donors with T2D (T1-T7, T2D, red) and donors with T2D who had been treated with metformin (TM1-TM9, T2D Metformin, purple). Quantification for each individual donor is shown. Data is represented as mean +/- SEM. Number of puncta analyzed: Plasma membrane, going from left (sample H1) to right (sample TM9) along the x-axis = 97, 69, 31, 36, 45, 147, 135, 70, 19, 74, 24, 135, 161,135, 54, 77, 138, 37, 120, 118, 57, 41, 74; Cytoplasm, going from left (sample H1) to right (sample TM9) along the x-axis = 68, 55, 91, 26, 25, 21, 18, 69, 20, 52, 37, 38, 30, 37, 67, 57, 68, 28, 22, 36, 10, 38, 24; Nucleus, going from left (sample H1) to right (sample TM9) along the x-axis = 26, 13, 19, 36, 8, 27, 8, 21, 19, 17, 36, 44, 21, 29, 38, 31, 44, 27, 12, 13, 12, 9, 25. **h**, Quantification of relative IR levels by immunoblot. IR level was normalized to CK18 and represented relative to sample Healthy 3 (H3). Data is represented as mean +/- SEM. Number of liver samples analyzed: Healthy 4 liver samples, T2D 2 liver samples, T2D Metformin 4 liver samples. Source data are provided as a Source Data file.

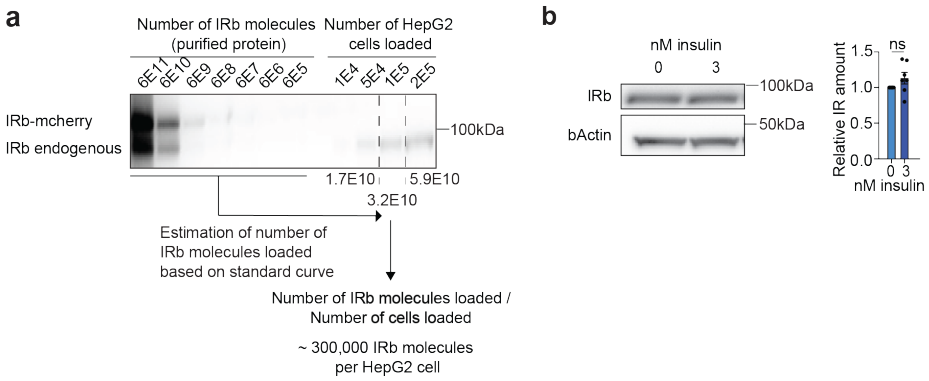


Supplementary Fig. 2. Validation of insulin-sensitive HepG2 cell model. a, Schematic of cell treatment (top). Percent viability of cells cultured in cell expansion media (Media + FBS, n=3 biological replicates) or in media containing physiological concentrations of insulin (Media – FBS, n=3 biological replicates) is reported in the graph (bottom). Data is reported as mean +/- SEM. Unpaired two-sided t-test was used for statistical analysis. **b**, Quantification of insulin clearance at 1, 5 or 24 hours in insulin-sensitive cells treated with 3nM insulin (1 hours, n=3 biological replicates; 5 hours, n=3 biological replicates, 24 hours n=3 biological replicates). Data is reported as mean +/- SEM. Unpaired two-sided t-test was used for statistical analysis. **c**, Experimental protocol (top) and immunoblot with quantitation (bottom) to measure phosphorylated insulin signaling proteins (pIRb, pAKT, pERK) over total insulin

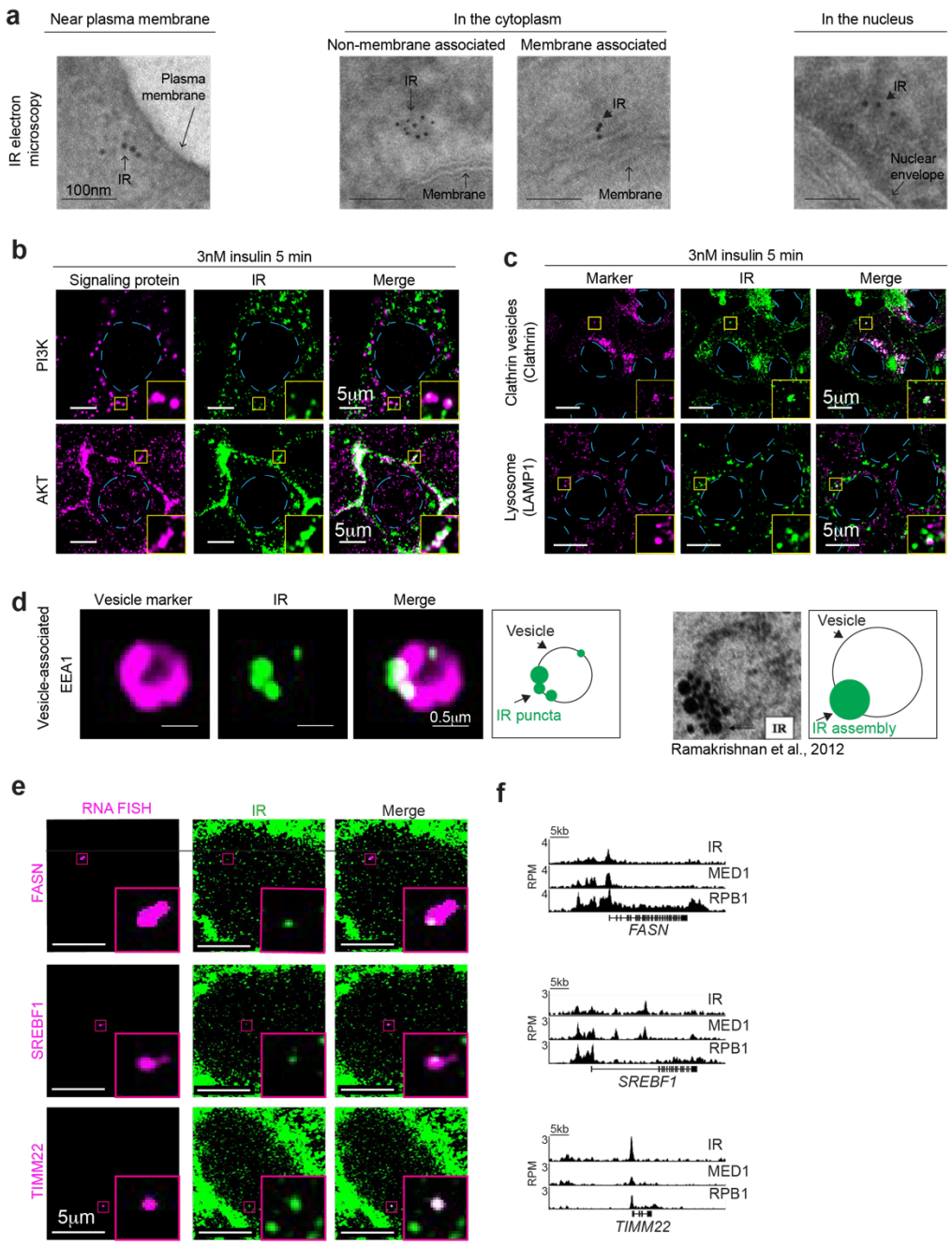
signaling proteins (IRb, AKT, ERK). 3 biological replicates per condition were analyzed for pIR/IR and for pAKT/AKT, 4 biological replicates were analyzed for pERK/ERK. Data is represented as mean +/- SEM. Unpaired two-sided t-test was used for statistical analysis. **d**, Gene ontology of the differentially expressed genes after 4 hours of 3nM insulin stimulation (3 biological replicates). The y-axis corresponds to the KEGG pathways. The x-axis and the point size represent the "Gene Ratio" defined as the fraction of differentially expressed genes in each given ontology term (in this case KEGG pathway). The color corresponds to -$\log_{10}$(adjusted p-value). **e**, Relative expression of *FASN* (n=5 biological replicates) and *PCK1* (n=3 biological replicates) in HepG2 cells acutely stimulated with (3) or without (0) insulin for 4 hours. Data is represented as mean +/- SEM. Two-sided unpaired t-test was used for statistical analysis. **f**, Isotope tracing experiment showing relative palmitate labeling in HepG2 cells acutely stimulated with 0nM or 1nM insulin for 36 hours. Data is represented as mean +/- SEM. 3 biological replicates were analyzed per condition and two-sided unpaired t-test was used for statistical analysis. **g**, Quantification of glucose production in cells stimulated with 0, 0.1, 1 or 10nM insulin for 5 hours. Data is represented as mean +/- SEM. 4 biologically independent samples were analyzed for conditions 0nM insulin, 0.1nM insulin and 1nM insulin, while 3 biologically independent samples were analyzed for condition 10nM insulin. Two-sided unpaired t-test was used for statistical analysis. **h**, Isotope tracing experiment showing relative glucose labeling in HepG2 cells acutely stimulated with 0, 1, 10 or 100nM insulin for 24 hours. Data is represented as mean +/- SEM. 3 biological replicates were analyzed per condition and unpaired two-sided t-test was used for statistical analysis. **i**, Immunoblot to quantify phosphorylated GSKa/b (pGSKa/b) over total GSKa/b protein in HepG2 cells acutely stimulated with 0nM or 3nM insulin for 5 minutes. 3 biological replicates were analyzed per condition, data is represented as mean +/- SEM and unpaired two-sided t-test was used for statistical analysis. This is the same experiment as in Supplementary Fig. 6j. Source data are provided as a Source Data file.



Supplementary Fig. 3. Automated quantification of IR signal intensity in puncta. a, Quantification of IR signal intensity in puncta in entire cells (without specifying cellular subcompartments), relative to Fig. 2b. Data is represented as individual values and mean +/- SEM. Number of IR puncta analyzed: Sensitive 0nM insulin (light blue) 16,984 puncta, Sensitive 3nM insulin (blue) 16,397 puncta, Resistant 0nM insulin (red) 15,296 puncta, Resistant 3nM insulin (dark red) 14,708 puncta. Unpaired two-sided t-test was used for statistical analysis. Source data are provided as a Source Data file.

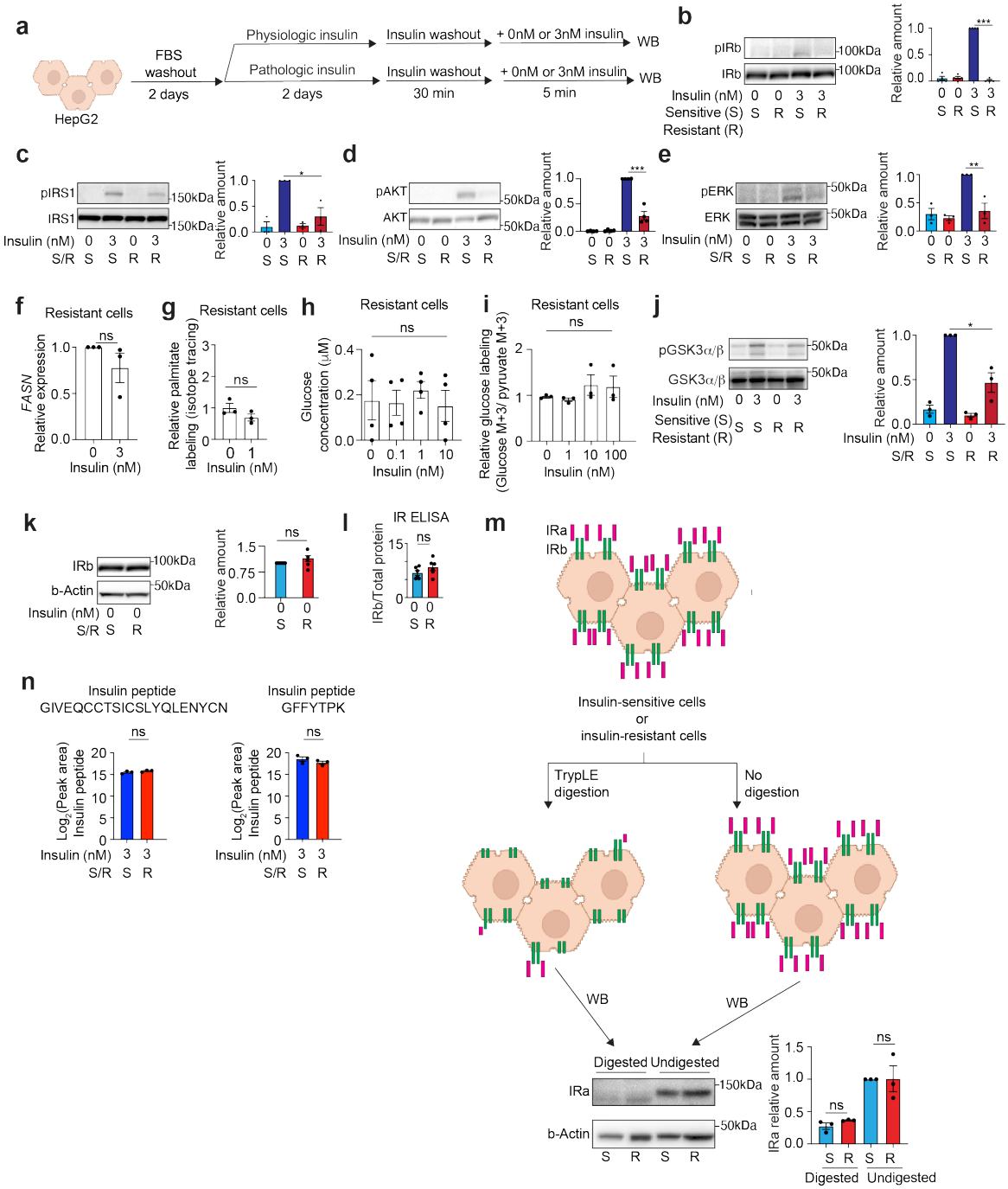


Supplementary Fig. 4. Quantification of the number of IR molecules in HepG2 cells.
a, Quantitative western blot with standard curve of purified IRbeta mCherry fusion protein (IRb-mCherry; first 7 lanes) and cell lysate containing a specific number of cells (last four lanes). **b,** Immunoblot for IRbeta (IRb) and beta-actin (bActin) in cells treated acutely with 0nM or 3nM insulin (left). Quantification of relative IRb levels in HepG2 cells without (0nM, light blue) and with (3nM, dark blue) acute insulin stimulation (right). IRb level was normalized to beta-actin. 7 biological replicates were analyzed per condition. Data is represented as individual values and as mean +/- SEM. Unpaired two-sided t-test was used for statistical analysis. Source data are provided as a Source Data file.



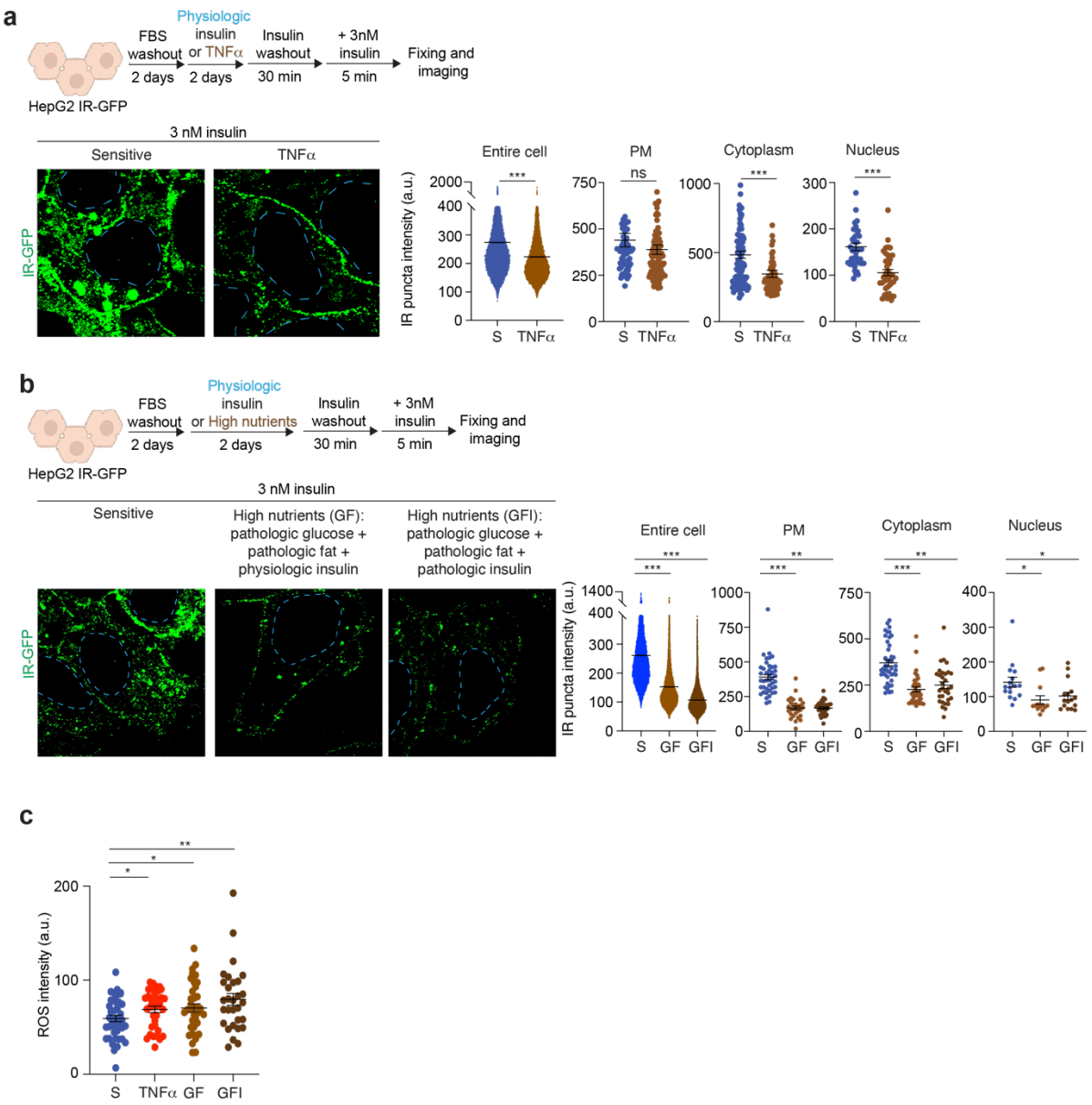
Supplementary Fig. 5. IR puncta in various cellular compartments in insulin-sensitive cells. a, Representative electron microscopy images for IR showing its presence near the plasma membrane, in the cytoplasm and in the nucleus. **b**, Representative immunofluorescence images for PI3K or AKT (magenta) together with IR (green) in insulin-sensitive HepG2 cells acutely stimulated with insulin for 5 minutes. Dashed light blue lines represent nuclear outline. Representative colocalization area (yellow box) is magnified at the bottom right corner of each image. **c**, Representative immunofluorescence images for clathrin or LAMP1 together with IR (green) in insulin-sensitive HepG2 cells acutely stimulated with insulin for 5 minutes. IR was detected either by immunofluorescence or by imaging endogenous IR-GFP. Dashed light blue lines represent nuclear outline. Representative colocalization area (yellow box) is magnified at the bottom right corner of each image. **d**, Representative immunofluorescence images for EEA1 (endosome marker) and IR, with a schematic representation of IR puncta associated with a portion

of the vesicle membrane (left). Published electron microscopy image of IR and another receptor associated with a portion of the membrane of a vesicle[1], with a schematic representation of IR puncta associated with the vesicle (right). Reuse of the published image[1] is granted under STM guidelines. **e,** Colocalization of IR and nascent RNA of *FASN, SREBF1* and *TIMM22* determined by imaging IR-GFP and *FASN, SREBF1* and *TIMM22* intronic RNA FISH in cells stimulated with 3nM insulin. Colocalization area (magenta box) is magnified at the bottom right corner of each image. Scale bars are indicated in the images. *FASN, SREBF1* and *TIMM22* are known insulin-responsive genes[2, 3, 4, 5, 6, 7]. If the fluorescence made the scale bar hard to see, a black box was added behind the scale bar. **f**, ChIP-seq tracks of IR, MED1 and RPB1 at *FASN, SREBF1* and *TIMM22* loci.



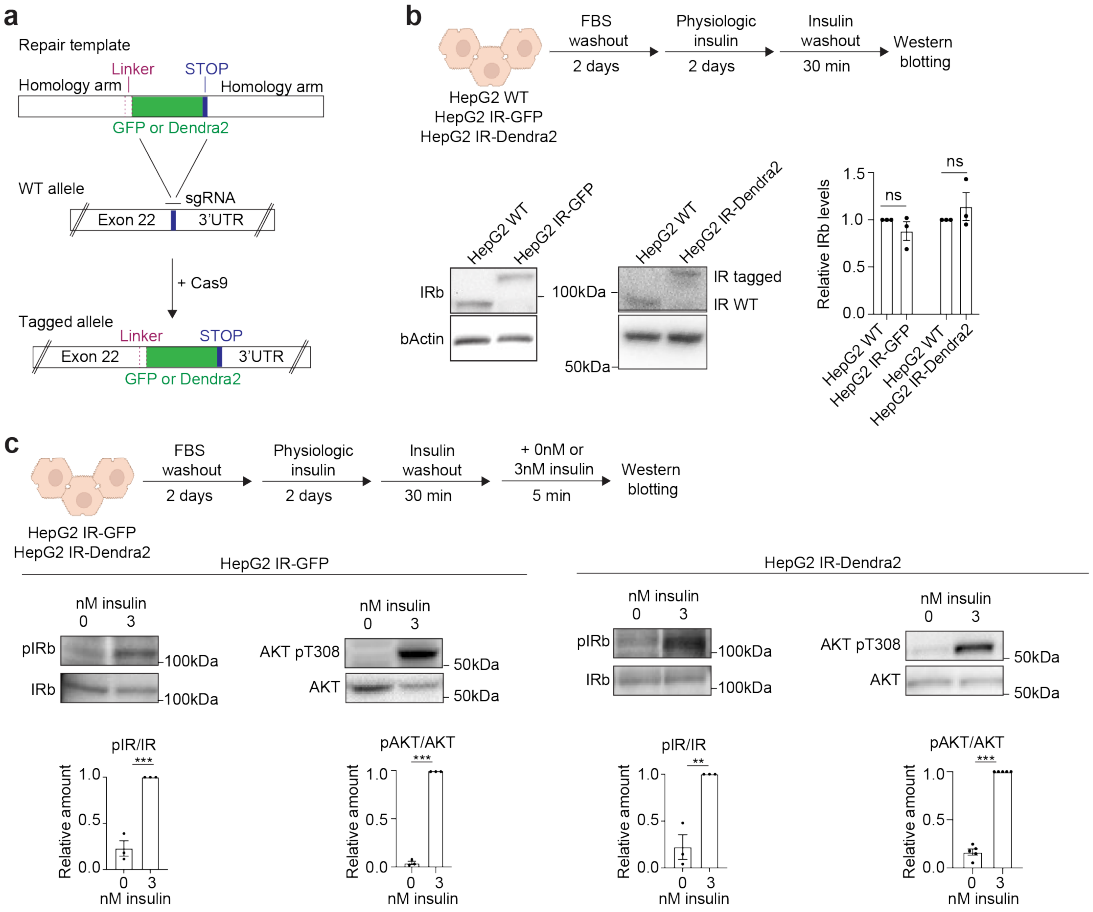
Supplementary Fig. 6. Validation of insulin-resistant HepG2 cell model. a, Schematic of cell treatments. **b-e**, Immunoblot with quantitation to measure phosphorylated insulin signaling proteins (pIRb, pIRS1, pAKT, pERK) over total insulin signaling proteins (IRb, IRS1, AKT, ERK) in insulin-sensitive (Sensitive, S, light blue) and insulin-resistant (Resistant, R, red) cells stimulated with 0nM (light color) or 3nM (dark color) insulin for 5 minutes. For figures b and d four biological replicates were analyzed, for figures c and e three biological replicates were analyzed. Individual replicates are shown in the graphs and bar graphs represent mean +/- SEM. Unpaired two-sided t-test was used for statistical analysis. **f**, Relative expression of *FASN* in insulin-resistant HepG2 cells acutely stimulated with 0nM or 3nM insulin for 4 hours. 3 biological replicates were analyzed. Individual data points are represented as well as the mean +/- SEM. Unpaired two-sided t-test was used for statistical analysis. **g**, Isotope tracing experiment showing relative palmitate labeling in insulin-resistant HepG2 cells acutely stimulated with 0nM or 1nM

insulin for 36 hours. 3 biological replicates were analyzed. Individual values are reported, bar graph represents mean +/- SEM. Unpaired two-sided t-test was used for statistical analysis. **h**, Quantification of glucose production in insulin-resistant HepG2 cells stimulated with 0, 0.1, 1 or 10nM insulin for 5 hours. 4 biological replicates were analyzed. Individual values are reported, bar graph represents mean +/- SEM. Unpaired two-sided t-test was used for statistical analysis. **i**, Isotope tracing experiment showing relative glucose labeling in insulin-resistant HepG2 cells acutely stimulated with 0, 1, 10 or 100nM insulin for 24 hours. 3 biological replicates were analyzed. Individual values are reported, bar graph represents mean +/- SEM. Unpaired two-sided t-test was used for statistical analysis. **j**, Immunoblot to quantify phosphorylated GSKA/B (pGSKA/B) over total GSKA/B protein in insulin-sensitive (Sensitive, S, blue) and insulin-resistant (Resistant, R, red) HepG2 cells acutely stimulated with 0nM (light color) or 3nM (dark color) insulin for 5 minutes. 3 biological replicates were analyzed. Individual values are reported, bar graph represents mean +/- SEM. Unpaired two-sided t-test was used for statistical analysis. **k**, Immunoblot for IRbeta (IRb) and beta-actin (b-Actin) in insulin-sensitive (light blue) and insulin-resistant (red) cells unstimulated with insulin (left). Quantification of relative IRb levels (right). 5 biological replicates were analyzed. Individual values are reported, bar graph represents mean +/- SEM. Unpaired two-sided t-test was used for statistical analysis. **l**, Enzyme-linked immunoassay (ELISA) for IRbeta (IRb) relative to total protein in insulin-sensitive (S, light blue) and insulin-resistant (R, red) cells unstimulated with insulin. 6 biological replicates were analyzed. Individual values are reported, bar graph represents mean +/- SEM. Unpaired two-sided t-test was used for statistical analysis. **m**, Schematic of proteolytic shaving experiment. Insulin-sensitive or resistant cells were either treated with TrypLE to digest the portions of proteins at the cell surface (Digested) or not (Undigested). Immunoblot with quantitation to measure IRalpha (IRa) and beta-actin (b-Actin) in digested and undigested insulin-sensitive (S, light blue) and insulin-resistant (R, red) cells. 3 biological replicates were analyzed. Individual values are reported, bar graph represents mean +/- SEM. Unpaired two-sided t-test was used for statistical analysis. **n**, Proteomic quantification of insulin binding in insulin-sensitive (S, blue) and insulin-resistant (R, red) cells treated with 3nM insulin at 4^{o}C. Peak area quantification is reported for two insulin peptides: GIVEQCCTSICSLYQLENYCN (insulin A-chain) and GFFYTPK (insulin B-chain). 3 biological replicates were analyzed. Individual values are reported, bar graph represents mean +/- SEM. Unpaired two-sided t-test was used for statistical analysis. Source data are provided as a Source Data file.



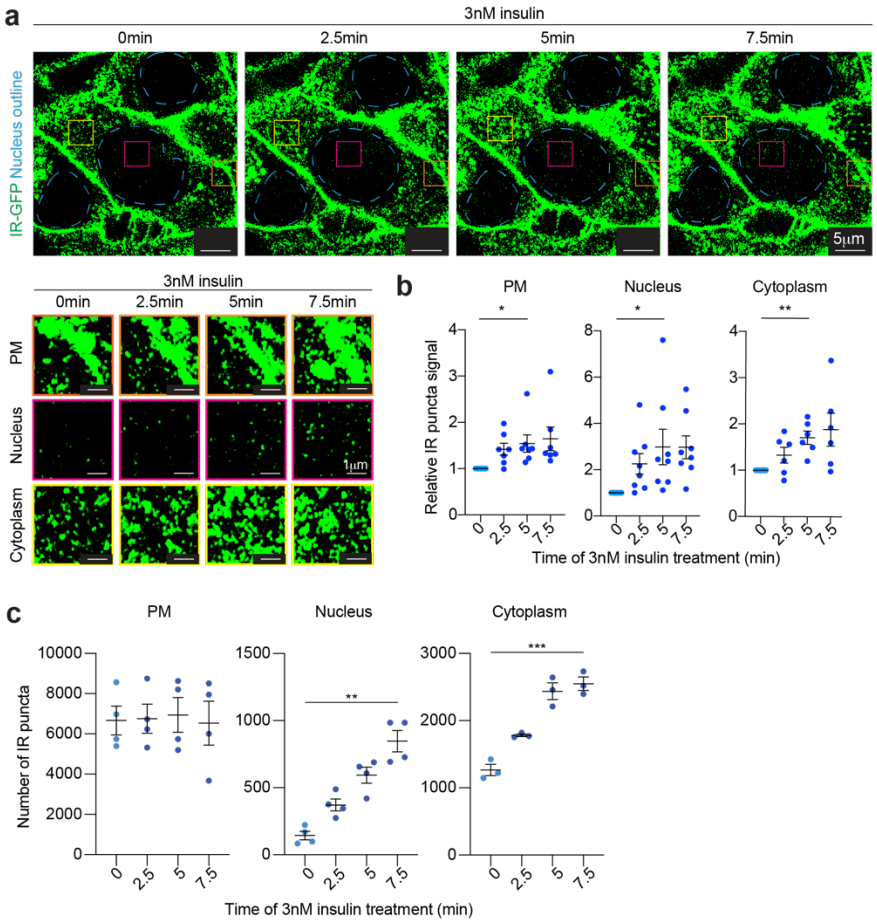
Supplementary Fig. 7. Other models of insulin resistance. a, Schematic of cell treatments (top). Imaging of IR-GFP in HepG2 cells treated with physiological concentrations of insulin (Sensitive, S, blue) or pathological concentration of TNFA (TNFA, brown) and acutely stimulated with 3nM insulin for 5 minutes (bottom left). Quantification of IR signal intensity in IR puncta in the entire cell (automated quantification, without specifying cellular subcompartments), at the plasma membrane (PM), cytoplasm or nucleus of cells (bottom right). In the graph individual values and the mean +/- SEM are reported. Number of IR puncta analyzed: Entire cell Sensitive 29,398 puncta, TNFA 31,083 puncta; Plasma membrane Sensitive 66 puncta, TNFA 112 puncta; Cytoplasm Sensitive 109 puncta, TNFA 76 puncta; Nucleus Sensitive 40 puncta, TNFA 40 puncta. Unpaired two-sided t-test was used for statistical analysis. **b**, Schematic of cell treatments (top). Imaging of IR-GFP in HepG2 cells treated with physiological concentrations of insulin (Sensitive, S, blue) or with high nutrients (either 1) pathological concentrations of glucose and fat and physiological concentration of insulin (high nutrients, GF, brown) or 2) pathological concentrations of glucose, fat and insulin (high nutrients, GFI, dark brown) and acutely stimulated with 3nM insulin for 5 minutes (bottom left). Quantification of IR signal intensity in IR puncta in the entire cell (automated quantification, without specifying cellular subcompartments), at the plasma membrane (PM), cytoplasm or nucleus of cells (bottom right). In the graph, individual values and the mean +/- SEM are

reported. Number of IR puncta analyzed: Entire cell Sensitive 21,382 puncta, GF 17,715 puncta, GFI 20,514 puncta; Plasma membrane Sensitive 42 puncta, GF 29 puncta, GFI 32 puncta; Cytoplasm Sensitive 44 puncta, GF 41 puncta, GFI 31 puncta; Nucleus Sensitive 16 puncta, GF 13 puncta, GFI 16 puncta. Unpaired two-sided t-test was used for statistical analysis. **c**, ROS intensity in insulin-sensitive HepG2 cells (blue), in cells treated with TNFA (red), or in cells treated with high nutrients (either 1) pathological concentrations of glucose and fat and physiological concentration of insulin (GF, brown) or 2) pathological concentrations of glucose, fat and insulin (GFI, dark brown). Physiological concentration of insulin corresponds to 0.1nM, pathological concentration of insulin corresponds to 3nM, pathological concentration of TNFA corresponds to 100pg/ml, pathological concentration of fat corresponds to 30µM palmitic acid and 45µM oleic acid, pathological concentration of glucose corresponds to 10mM. In the graph, individual values and the mean +/- SEM are reported. Number of cells analyzed: Sensitive 42 cells, TNFA 35 cells, GF 36 cells, GFI 30 cells. Unpaired two-sided t-test was used for statistical analysis. Source data are provided as a Source Data file.

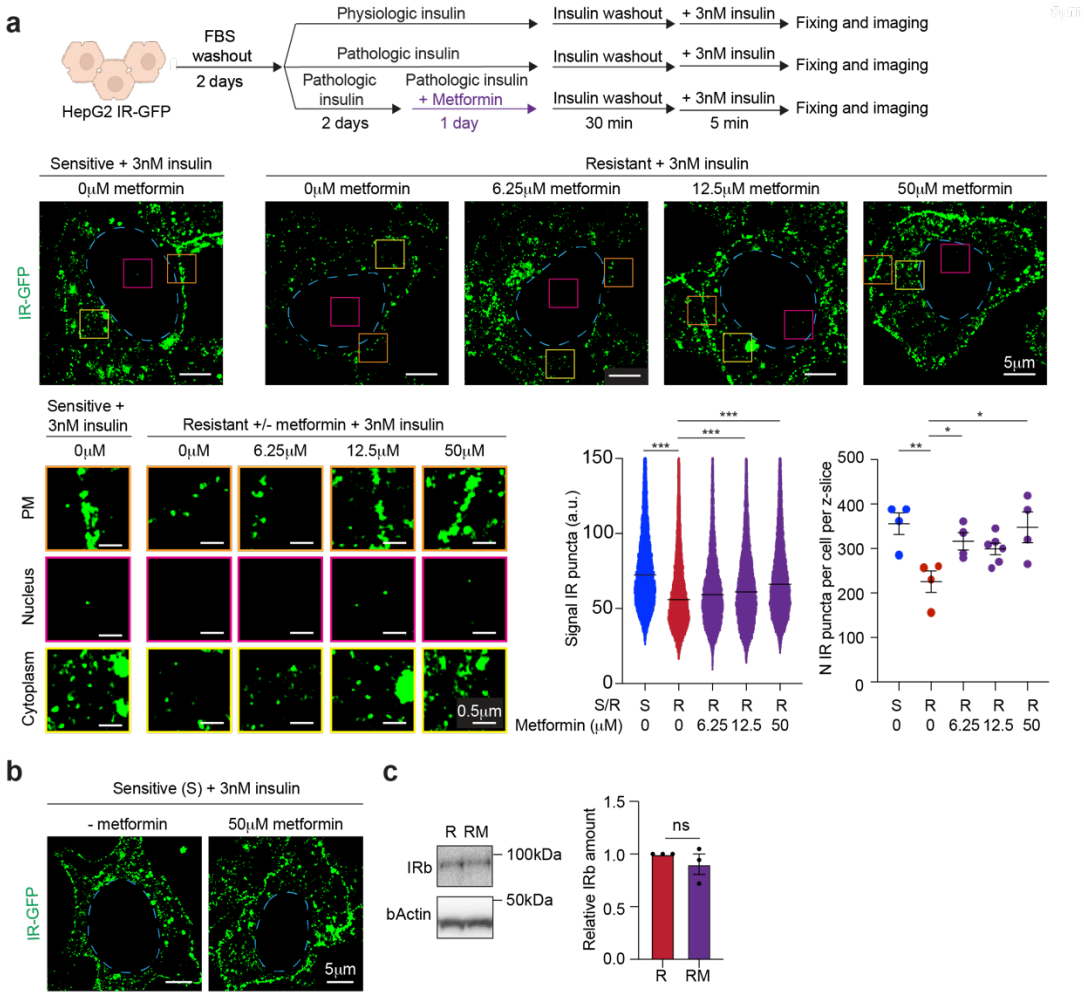


Supplementary Fig. 8. Homozygous HepG2 cell lines expressing functional endogenous IR tagged with GFP or Dendra2. a, Schematic of knock-in strategy. **b**, Schematic of cell treatments (top). Immunoblot for IRbeta (IRb) and beta-actin (bActin) control in WT, IR-GFP and IR-Dendra2 cell lines (bottom left). The shift in molecular weight is the expected size for the GFP or Dendra2 fusion with IR. Quantitation of IRb levels (bottom right). Individual values are reported and the bar graphs represent mean +/- SEM. 3 biological replicates were analyzed and unpaired two-sided t-test was used for statistical analysis. **c,** Schematic of cell treatments (top). Immunoblot with quantitation to measure phosphorylated insulin signaling proteins (pIRb and pAKT) over total insulin signaling proteins (IRb and AKT) in IR-GFP and IR-Dendra2 cells stimulated with 0nM or 3nM insulin for 5 minutes (bottom). Individual values are reported and the bar graphs represent mean +/- SEM. 3 biologically independent replicates were analyzed for pIRb/IRb in HepG2 IR-GFP cells and HepG2 IR-Dendra2 cells. 3 biologically independent replicates were analyzed for pAKT/AKT in IR-GFP cells and 4 biologically independent replicates were

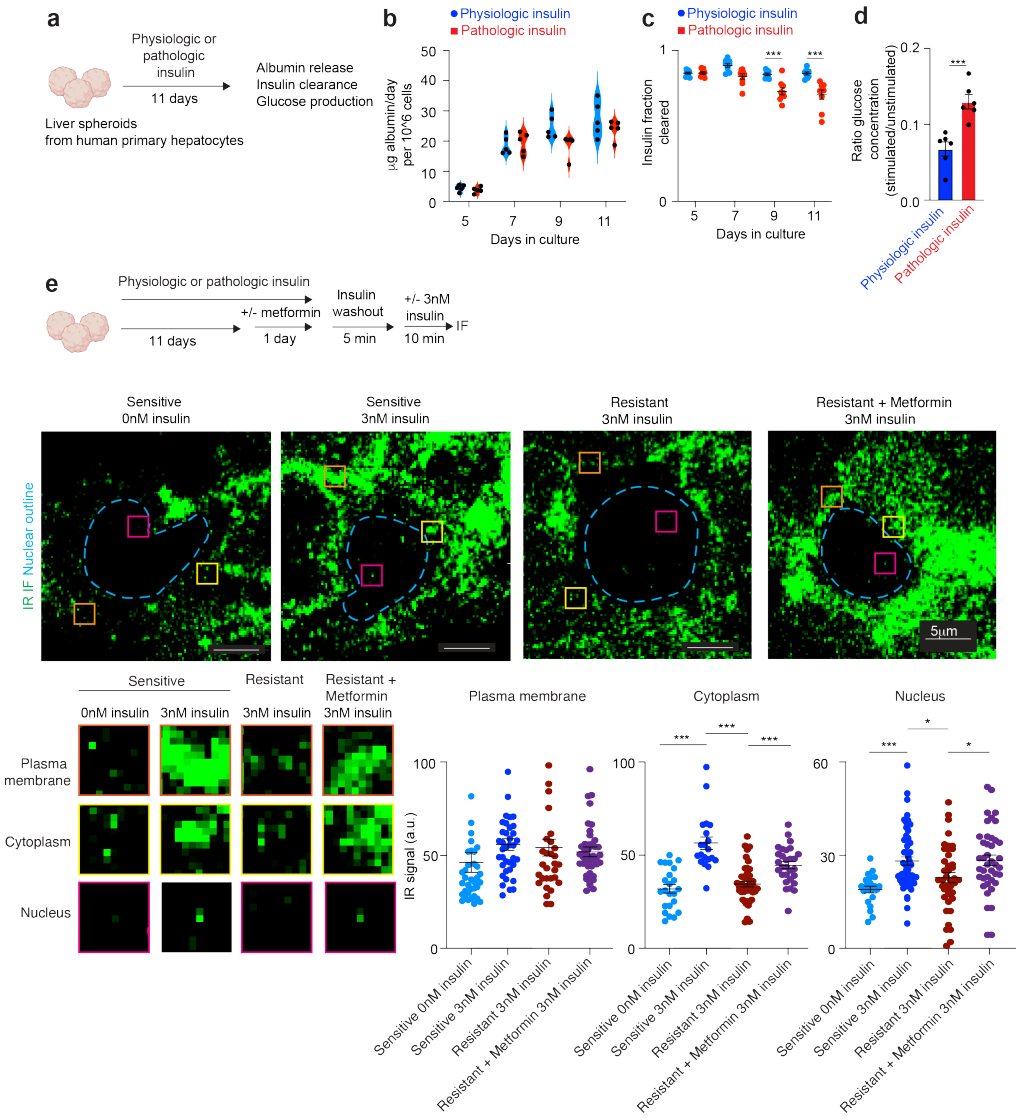
analyzed for pAKT/AKT in IR-Dendra2 cells. Unpaired two-sided t-test was used for statistical analysis. Source data are provided as a Source Data file.



Supplementary Fig. 9. Live-cell imaging of IR puncta in HepG2 cells. a, Live imaging time course of HepG2 cells expressing endogenous IR tagged with GFP during insulin stimulation. Time of acquisition is reported above images. Dashed light blue lines represent nuclear outline and scale bar are indicated in the images. Representative images of three cells (top). Orange, magenta and yellow boxes represent regions at the plasma membrane (PM), nucleus and cytoplasm, respectively, that are magnified at the bottom. If the fluorescence made the scale bar hard to see, a black box was added behind the scale bar. **b,** Quantification of IR puncta signal at the plasma membrane (PM), nucleus and cytoplasm of IR-GFP cells stimulated with 3nM insulin for 0, 2.5, 5 and 7.5 minutes. Data is represented as "relative to 0 minutes". In the graphs, individual values and the mean +/- SEM are reported. Number of regions analyzed: Plasma membrane 7 regions, Cytoplasm 6 regions, Nucleus 8 regions. Unpaired two-sided t-test was used for statistical analysis. **c**, Quantification of number of IR puncta at the plasma membrane (PM), nucleus and cytoplasm of IR-GFP cells stimulated with 3nM insulin for 0, 2.5, 5 and 7.5 minutes. In the graphs, individual values and the mean +/- SEM are reported. Number of cells analyzed: Plasma membrane 4 cells, Cytoplasm 3 cells, Nucleus 4 cells. Unpaired two-sided t-test was used for statistical analysis. Source data are provided as a Source Data file.

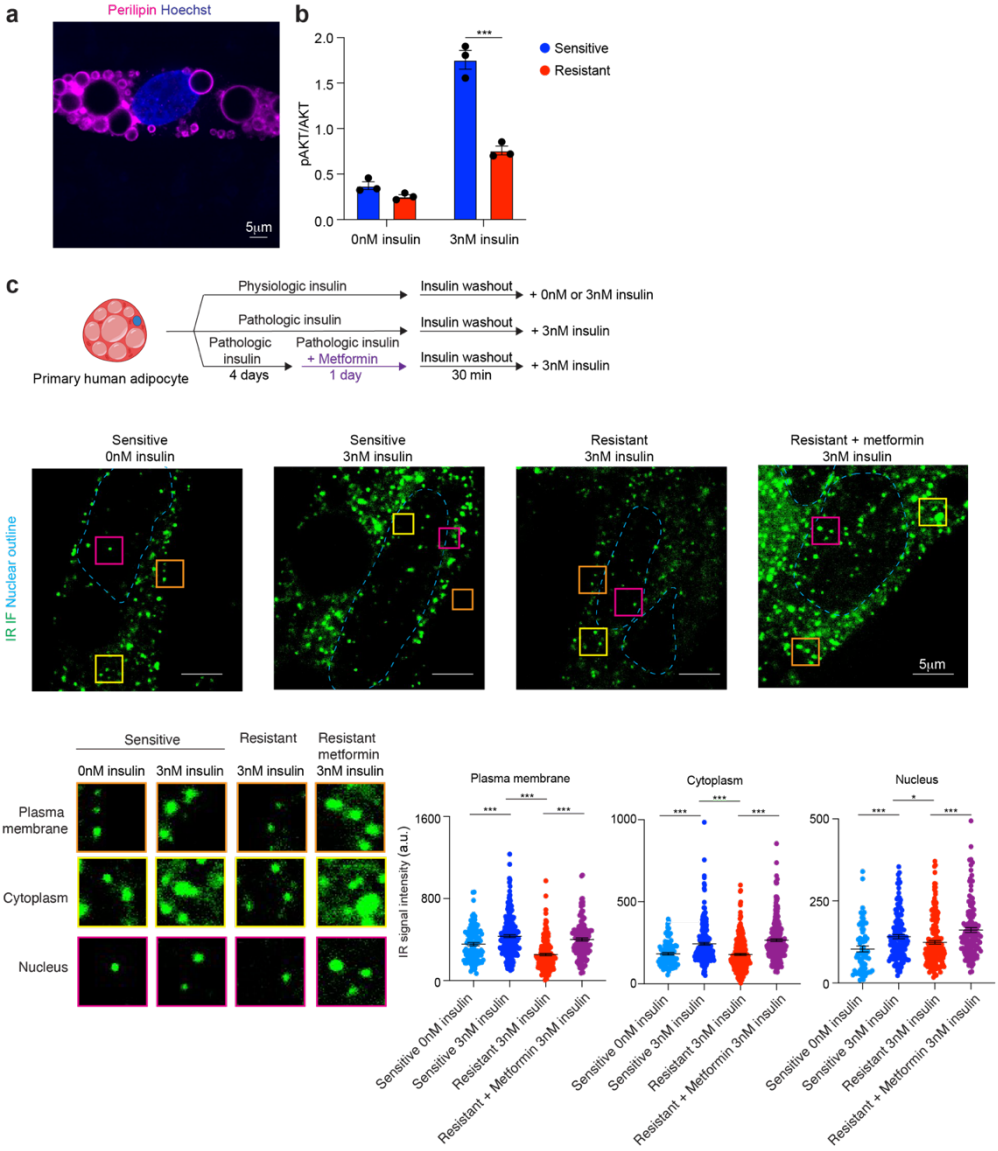


Supplementary Fig. 10. Metformin effect on IR puncta. a, Schematic of cell treatments (top). Imaging of IR-GFP in insulin-sensitive and insulin-resistant cells treated with or without metformin (middle). Metformin concentration is reported above the images. IR-GFP fluorescence signal is shown in green. Dashed light blue lines represent nuclear outline. Orange, magenta and yellow boxes represent regions at the plasma membrane (PM), nucleus and cytoplasm, respectively, that are magnified (bottom left). Scale bars are indicated in the images. This is the same experiment as in Fig. 2d and thus the same images for insulin-sensitive cells, insulin-resistant cells and insulin-resistant cells treated with 12.5μM metformin are reported in Fig. 2d. Quantification of IR signal in puncta (automated quantification) and the number of IR puncta in insulin-sensitive (blue) or insulin-resistant cells treated with (purple) or without metformin (red) (bottom right). In the graphs, individual values and the mean +/- SEM are reported. Number of IR puncta analyzed: Sensitive 13,128 puncta, Resistant 14,327 puncta, Resistant 6.25μM Metformin 12,948 puncta, Resistant 12.5μM Metformin 13,867 puncta, Resistant 50μM Metformin 20,817 puncta. Number of cells analyzed to quantitate the number of IR puncta per cell: Sensitive 4 cells, Resistant 4 cells, Resistant 6.25μM Metformin 4 cells, Resistant 12.5μM Metformin 6 cells, Resistant 50μM Metformin 4 cells. Unpaired two-sided t-test was used for statistical analysis. **b**, Imaging of IR-GFP in insulin-sensitive cells treated with or without 50μM metformin and acutely stimulated with 3nM insulin for 5 minutes. Dashed light blue lines represent nuclear outline. **c,** Immunoblot for IRbeta (IRb) and beta-actin (bActin) in cells cultured in pathologic levels of insulin treated with (RM) or without (R) 12.5μM metformin (left). Quantification of relative levels of IRb in insulin-resistant cells (red) and insulin-resistant cells treated with metformin (purple) (right). 3 biological replicates were analyzed. In the graph, individual values and the mean +/- SEM are reported. Unpaired two-sided t-test was used for statistical analysis. Source data are provided as a Source Data file.



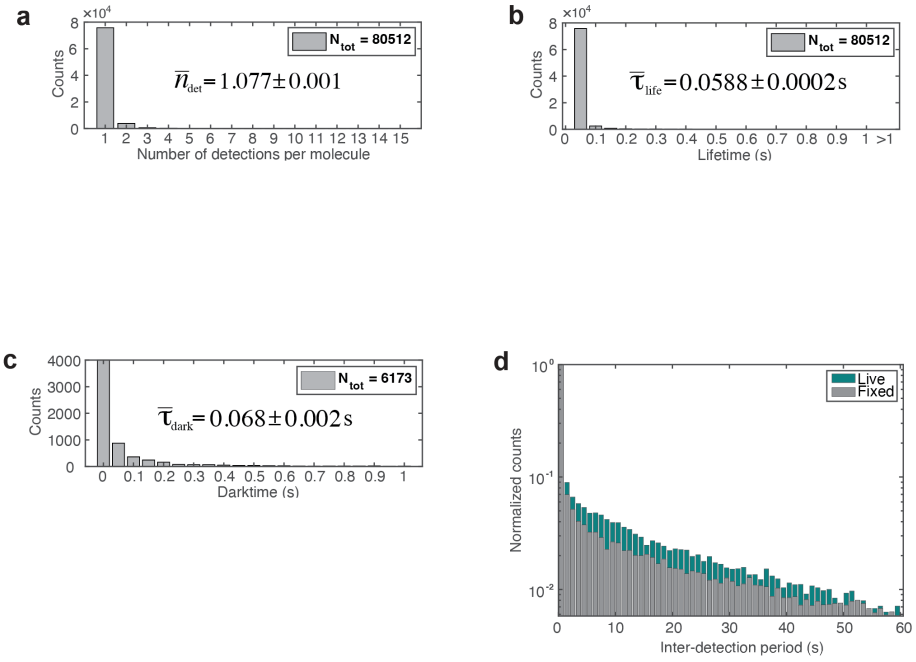
Supplementary Fig. 11. IR puncta in human primary hepatocytes, a, Schematic of cell treatments. **b,** Enzyme-linked immunoassay (ELISA) quantification of albumin production by human liver spheroids cultured with physiologic (blue) or pathologic (red) concentrations of insulin. Individual values are reported in the graph. 5 biological replicates were analyzed. **c,** Enzyme-linked immunoassay (ELISA) quantification of insulin clearance by human liver spheroids cultured with physiologic (blue) or pathologic (red) concentrations of insulin. Individual values and the mean +/- SEM are reported in the graph. 9 biological replicates were analyzed and unpaired two-sided t-test was used for statistical analysis. **d,** Quantification of glucose production in human liver spheroids cultured with physiologic (blue) or pathologic (red) concentrations of insulin. Individual values and the mean +/- SEM are reported in the graph. 6 biological replicates were analyzed and unpaired two-sided t-test was used for statistical analysis. **e,** Schematic of cell treatments (top). Immunofluorescence for IR in insulin-sensitive, insulin-resistant and metformin-treated insulin-resistant human liver spheroids acutely treated with 0nM or 3nM insulin for 10 minutes (middle). Dashed light blue lines represent nuclear outline. Orange, yellow and magenta boxes represent regions at the plasma membrane, cytoplasm and nucleus, respectively, that are magnified (bottom left). Quantification of IR signal at IR puncta at the plasma membrane, cytoplasm and nucleus of insulin-sensitive hepatocytes (light blue), insulin-sensitive hepatocytes acutely stimulated with insulin (blue), insulin-resistant hepatocytes acutely stimulated with insulin (red) and insulin-resistant hepatocytes treated with metformin and acutely stimulated with insulin (purple) (bottom right). Individual values and the mean

+/- SEM are reported in the graph. Number of IR puncta analyzed: Sensitive 0nM insulin Plasma membrane 37 puncta, Cytoplasm 23 puncta, Nucleus 21 puncta; Sensitive 3nM insulin Plasma membrane 38 puncta, Cytoplasm 20 puncta, Nucleus 55 puncta; Resistant 3nM insulin Plasma membrane 32 puncta, Cytoplasm 46 puncta, Nucleus 41 puncta; Resistant + Metformin 3nM insulin Plasma membrane 45 puncta, Cytoplasm 29 puncta, Nucleus 41 puncta. Unpaired two-sided t-test was used for statistical analysis. If the fluorescence made the scale bar hard to see, a black box was added behind the scale bar. Source data are provided as a Source Data file.



Supplementary Fig. 12. IR puncta in human primary adipocytes. **a,** Representative immunofluorescence image of perilipin (magenta) in human primary adipocyte. Nucleus is counterstained using Hoechst. **b**, Enzyme-linked immunoassay (ELISA) quantification of pAKT over AKT in human primary adipocytes treated with physiological (Sensitive, blue) or pathological (Resistant, red) concentrations of insulin for 5 days and acutely stimulated (3nM insulin) or not (0nM insulin) with insulin for 15 minutes. 3 biological replicates were analyzed. Individual values are reported in the graph and the bar graph represent mean +/- SEM. Unpaired two-sided t-test was used for statistical analysis. **c,** Schematic of cell treatments (top). Immunofluorescence for IR in insulin-sensitive (Sensitive), insulin-resistant (Resistant) and metformin-treated insulin-resistant (Resistant + Metformin) human primary adipocytes acutely treated with 0nM or 3nM insulin for 5 minutes (middle). Orange, yellow and magenta

boxes represent regions at the plasma membrane, cytoplasm and nucleus, respectively, that are magnified at the bottom left. Quantification of IR signal at IR puncta at the plasma membrane, cytoplasm and nucleus of insulin-sensitive adipocytes (light blue), insulin-sensitive adipocytes acutely stimulated with insulin (blue), insulin-resistant adipocytes acutely stimulated with insulin (red) and insulin-resistant adipocytes treated with metformin and acutely stimulated with insulin (purple) (bottom right). (bottom right). Individual values and the mean +/- SEM are reported in the graph. Number of IR puncta analyzed: Sensitive 0nM insulin Plasma membrane 107 puncta, Cytoplasm 91 puncta, Nucleus 67 puncta; Sensitive 3nM insulin Plasma membrane 209 puncta, Cytoplasm 238 puncta, Nucleus 135 puncta; Resistant 3nM insulin Plasma membrane 178 puncta, Cytoplasm 274 puncta, Nucleus 187 puncta; Resistant + Metformin 3nM insulin Plasma membrane 148 puncta, Cytoplasm 279 puncta, Nucleus 129 puncta. Unpaired two-sided t-test was used for statistical analysis. Source data are provided as a Source Data file.



e

Statistics of overall single molecule (SM)	Live	Fixed	Units
Avg. number of detections per SM	1.077 ± 0.001	1.204 ± 0.004	counts
Avg. duration time of overall SM	0.059 ± 0.000	0.075 ± 0.001	second
Avg. inter-detection period of multi-detection SM	0.068 ± 0.002	0.079 ± 0.002	second

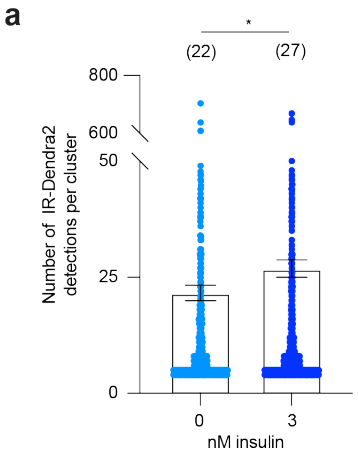
f

Statistics of identified multimolecule bursts	AVG.		STD.	Units
Number of identified bursts per 10,000 detections	67.02	±	2.26	counts
Avg. duration time ("lifetime") of bursts	14.52	±	0.23	second
Avg. inter-detection period of bursts	1.43	±	0.02	second
Avg. number of detections per burst	22.46	±	0.78	counts
Lifetime of bursts — 0.05 quantile	0.85	±	0.03	second
Number of detections per burst — 0.05 quantile	4.00	±	0.00	counts

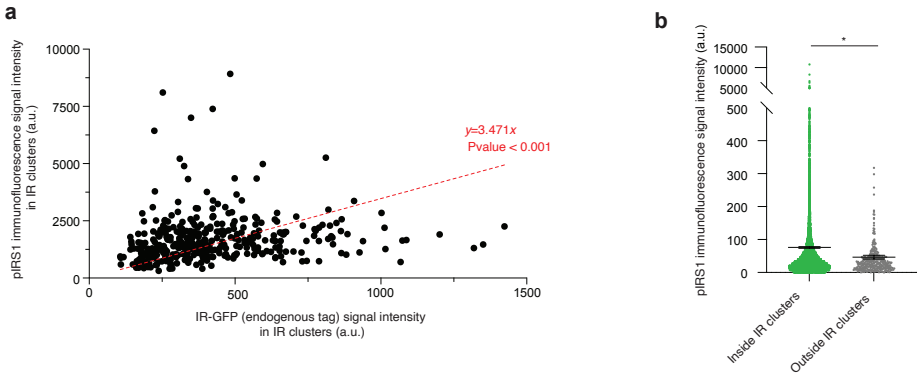
Statistics of outlier SM (t_min=0.85s, n_min=4)	AVG.		STD.	Units
Number of false positive SM per 10,000 detections	4.67	±	0.66	counts
Avg. duration time of false positive SM	1.21	±	0.06	second
Avg. inter-detection period of false positive SM	0.24	±	0.02	second
Avg. number of detections per false positive SM	5.65	±	0.37	counts

TPR>90%

Supplementary Fig. 13. Single-molecule statistics and validation of tc-PALM analysis. a, Distribution of the number of detections of single molecules in live cells. Total number of 80512 single molecules are collected for plotting the histogram. **b**, Distribution of the lifetime of single molecules. **c,** Distribution of the inter-detection period (dark-time) of single molecules with more than one detection. Total number of 6173 multi-detection single molecules are collected for plotting the histogram. **d,** Histogram of inter-detection period of identified transient clusters in live cells and pseudo-transient clusters in fixed cells selected with the same procedure as in live cells. The counts of each bin are normalized to the first bin, which mostly consists of counts of blinking events from single molecules (given that most single molecules have a lifetime span shorter than 1s). **e,** Statistics of single molecules in live and fixed samples. **f,** Statistics of identified multimolecule bursts and outlier single molecules. Ideally, the true positive rate (TPR) can go beyond 90% based on the estimation of cut-offs (0.05 quantile) from real bursts. Source data are provided as a Source Data file.

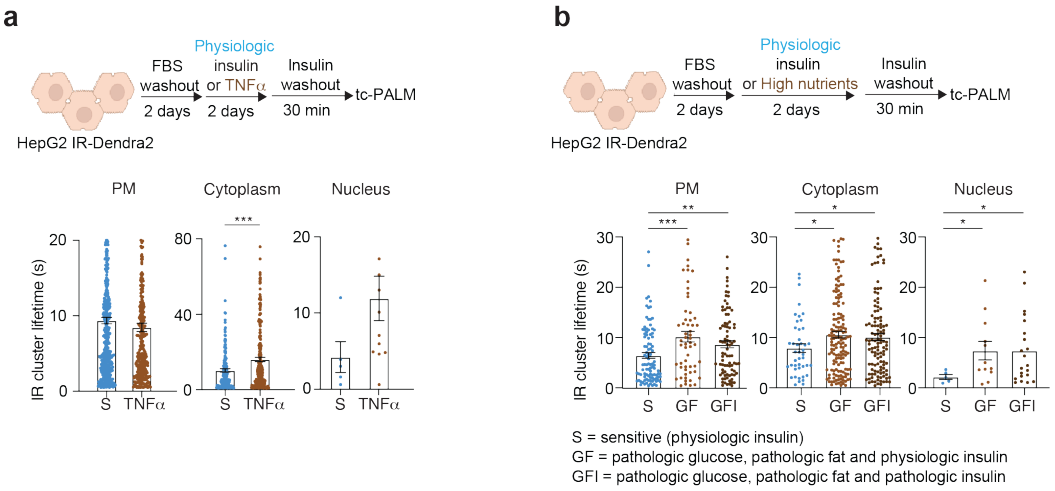


Supplementary Fig. 14. IR-Dendra2 detections in clusters throughout the cell. a, Quantification of the number of IR-Dendra2 detections per IR cluster in insulin-sensitive cells stimulated with (3nM, dark blue) and without (0nM, light blue) insulin for 5 minutes. Average number of IR-Dendra2 detections per IR cluster is reported in parenthesis on top of each histogram. Histograms represent mean +/- SEM. Number of clusters analyzed: Sensitive 0nM insulin 908 clusters, Sensitive 3nM insulin 1,116 clusters. Unpaired two-sided t-test was used for statistical analysis. Source data are provided as a Source Data file.

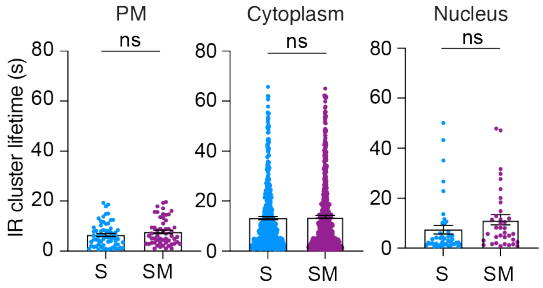


Supplementary Fig. 15. Correlation between IR and pIRS1 signal intensity in clusters. a, Quantification of pIRS1 and IR signal in clusters. To obtain IR clusters with different levels of IR molecules, HepG2 cells expressing endogenous IR tagged with GFP (IR-GFP) were treated with siControl or siRNA for INSR for 18 hours or 24 hours. 391 IR clusters were analyzed. Linear regression

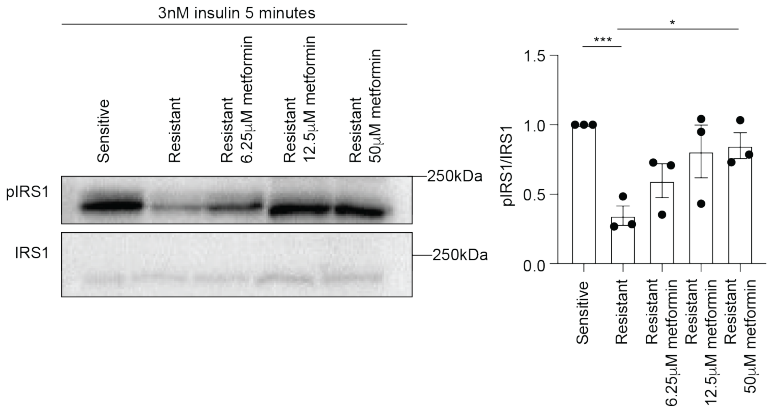
was used to generate trendline. Equation: y=3.47x. **b,** Quantification of pIRS1 signal inside (green) and outside (grey) IR clusters. Individual values and the mean +/- SEM are reported in the graph. 12,447 IR clusters and 397 regions outside of IR clusters were analyzed. Unpaired two-sided t-test was used for statistical analysis. Source data are provided as a Source Data file.



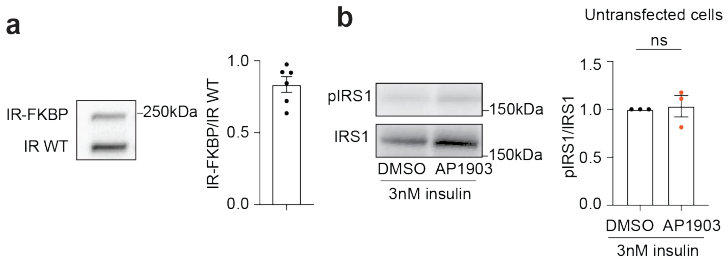
Supplementary Fig. 16. Increased IR cluster lifetime by inflammation and high nutrients. a, Schematic of cell treatments (top). Tc-PALM quantification of IR cluster lifetime at the plasma membrane (PM), cytoplasm and nucleus in HepG2 cells expressing IR-Dendra2 treated with physiological concentrations of insulin (Sensitive, S, light blue) or pathological concentration of TNFA (TNFA, brown). Individual values and the mean +/- SEM are reported in the graph. Number of short-lived clusters analyzed: Plasma membrane Sensitive 440 clusters, TNFa 307 clusters; Cytoplasm Sensitive 170 clusters, TNFa 212 clusters; Nucleus Sensitive 5 clusters, TNFa 22 clusters. Unpaired two-sided t-test was used for statistical analysis. **b**, Schematic of cell treatments (top). Tc-PALM quantification of IR cluster lifetime at the plasma membrane (PM), cytoplasm and nucleus in HepG2 cells expressing IR-Dendra2 treated with physiological concentrations of insulin (Sensitive, S, light blue) or with high nutrients either 1) pathological concentrations of glucose and fat and physiological concentration of insulin (high nutrients, GF, brown) or 2) pathological concentrations of glucose, fat and insulin (high nutrients, GFI, dark brown). Physiological concentration of insulin corresponds to 0.1nM, pathological concentration of insulin corresponds to 3nM, pathological concentration of TNFA corresponds to 100pg/ml, pathological concentration of fat corresponds to 30µM palmitic acid and 45µM oleic acid, pathological concentration of glucose corresponds to 10mM. Number of short-lived clusters analyzed: Plasma membrane Sensitive 97 clusters, GF 56 clusters, GFI 85 clusters; Cytoplasm Sensitive 44 clusters, GF 148 clusters, GFI 117 clusters; Nucleus Sensitive 5 clusters, GF 12 clusters, GFI 21 clusters. Unpaired two-sided t-test was used for statistical analysis (Plasma membrane Sensitive vs GF, Sensitive vs GFI; Cytoplasm Sensitive vs GF) or unpaired one-sided t-test was used for statistical analysis for cytoplasm Sensitive vs GFI and nucleus Sensitive vs GF, Sensitive vs GFI. Source data are provided as a Source Data file.



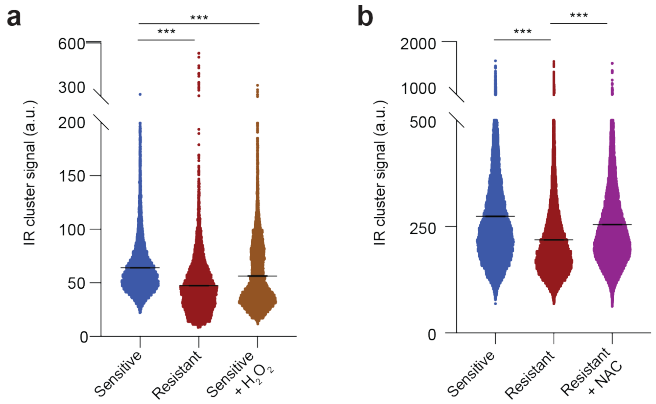
Supplementary Fig. 17. Metformin does not decrease IR cluster lifetime in insulin-sensitive cells. Tc-PALM quantification of IR cluster lifetime at the plasma membrane (PM), cytoplasm and nucleus in insulin-sensitive HepG2 cells expressing IR-Dendra2 treated with (purple) and without 12.5μM metformin (light blue) for 1 day. Data is represented as mean +/- SEM. Number of IR short-lived clusters analyzed: Sensitive (S) Plasma membrane 310 clusters, Cytoplasm 456 clusters, Nucleus 43 clusters; Sensitive + Metformin (SM) Plasma membrane 447 clusters, Cytoplasm 609 clusters, Nucleus 36 clusters. Unpaired two-sided t-test was used for statistical analysis.



Supplementary Fig. 18. Metformin partially rescues phosphorylation of IRS1. Immunoblot and quantification of pIRS1 over total IRS1 in insulin-sensitive, insulin-resistant and metformin-treated insulin-resistant cells. Metformin concentrations used in the experiment are reported in the image. Data is represented as single values and bar graphs (mean +/- SEM). Three biological replicates were analyzed. Unpaired two-sided t-test was used for statistical analysis. Source data are provided as a Source Data file.



Supplementary Fig. 19. Effect of AP1903 on HepG2 cells. a, Immunoblot and quantification of the relative levels of expression of WT IR (IR WT) and IR-GFP-FKBP (IR-FKBP). Data is represented as single values and a bar graph (mean +/- SEM). 6 biological replicates were analyzed. **b**, Immunoblot and quantification of the relative levels of phosphorylated IRS1 and total IRS1 in untransfected, wildtype HepG2 cells treated with DMSO or AP1903. 3 biological replicates per condition were analyzed. Data is represented as single values and bar graphs (mean +/- SEM). Unpaired two-sided t-test was used for statistical analysis. Source data are provided as a Source Data file.



Supplementary Fig. 20. **Oxidative stress effect on IR incorporation into clusters. a**, Quantification of IR signal intensity in IR clusters in the entire cell (automated quantification, without specifying cellular subcompartments) relative to Fig. 5c. Data points from sensitive cells are represented in blue, data points from resistant cells are represented in dark red, data points from resistant cells treated with NAC are represented in dark red. Single values and mean +/- SEM are shown. Number of clusters analyzed: Sensitive 11,110 clusters, Resistant 8,861 clusters, Sensitive + H_2O_2 8,068 clusters. Unpaired two-sided t-test was used for statistical analysis. **b**, Quantification of IR signal intensity in IR clusters in the entire cell (automated quantification, without specifying cellular subcompartments) relative to Fig. 5e. Single values and mean +/- SEM are shown. Number of clusters analyzed: Sensitive 29,398 clusters, Resistant 30,600 clusters, Resistant + NAC 46,992 clusters). Unpaired two-sided t-test was used for statistical analysis. Source data are provided as a Source Data file.

Supplementary Table 1. Donor characteristics.

Healthy/T2D	Specimen ID	Case ID	Age At Excision	Sex	Etnicity	Biosample Diagnosis	BMI	Metformin Treatment	Source
Healthy	1208572F	87808	77	Female	Caucasian	Normal	23.14	No	BioIVT
Healthy	1208568F	87808	77	Female	Caucasian	Normal	23.14	No	BioIVT
Healthy	FHU-L-102319	N/A	37	Female	Caucasian	Normal	23.82	No	BioIVT
Healthy	AM-092	N/A	29	Female	Caucasian	Normal	24	No	MGH
Healthy	20-018 / OL-001	N/A	67	Male	Caucasian	Normal	23	No	MGH
Healthy	21-015/OL-035	N/A	50	Male	N/A	Normal	25	No	MGH
Healthy	AM-019/OL-028	N/A	50	Female	Caucasian	Normal	28	No	MGH
T2D	1214825F	99117	53	Female	Caucasian	Steatosis	27.18	No	BioIVT
T2D	1137920F	47111	70	Male	Caucasian	Steatosis	34.42	No	BioIVT
T2D	1143147F	49893	76	Female	N/A	Congestion	37.6	No	BioIVT
T2D	AM-026/OL-038	N/A	70	Male	Caucasian	Steatosis	38	No	MGH
T2D	20-024 / OL-004	N/A	73	Male	Caucasian	Unkown	23	No	MGH
T2D	OL-019	N/A	73	Male	Caucasian	Steatosis	22	No	MGH
T2D	20210519	N/A	62	Male	Hispanic	Congestion	36.6	No	MGH
T2D	1096575F	48612	66	Male	Caucasian	Steatosis	30.86	Yes	BioIVT
T2D	1153543F	52473	77	Female	Native American or Alaskan Native	Steatosis	25.7	Yes	BioIVT
T2D	27534H1	14426	59	Female	Caucasian	Steatosis	20.9	Yes	BioIVT
T2D	AM-011 / OL-025	N/A	54	Male	Caucasian	Steatosis	42	Yes	MGH
T2D	OL-013	N/A	72	Male	Caucasian	Steatosis	33.5	Yes	MGH
T2D	21-006/OL-029	N/A	69	Male	Caucasian	Steatosis	31.5	Yes	MGH
T2D	21-134/OL-063	N/A	67	Male	Caucasian	Steatosis	31	Yes	MGH
T2D	20-054 / OL-024	N/A	41	Male	Caucasian	Steatosis	36	Yes	MGH
T2D	21-088/OL-052	N/A	69	Female	Caucasian	Steatosis	28	Yes	MGH

SUPPLEMENTARY TEXT

Photochemistry of single Dendra2 molecules

Given that there could be an ambiguous mapping from the number of detections to the number of Dendra2 molecules, several control analyses of the single molecule photochemistry have to be done to validate the statistics of the real clusters (which ideally consist of colocalized, time-correlated, multimolecule bursts). Imaging of IR in either fixed or live IR-Dendra2 cells was performed in L-15 medium using the same laser setups as described above. After the same ROI was imaged for a long time, most Dendra2 molecules were photo-converted and bleached, whereupon the rest of intact single molecules were sparsely photo-converted and recorded, and the consequent colocalized detections from the same molecule can be well spatiotemporally isolated and grouped. The statistics of live-cell Dendra2 single molecules are shown in Supplementary Figure 13a-c. The comparisons of Dendra2 single molecules in live and fixed samples are shown in Supplementary Figure 13e. 94% of the single molecules only generate one detection (Supplementary Fig. 13a), which results in the average number of detections per molecule being close to one ($\bar{n}_{\mathrm{det}} \approx 1.077$). The average lifetime of single molecules is 0.059s, and only 1% of them has a lifetime longer than 0.25s (Supplementary Fig. 13b). Among those multiple-detection molecules, 65% of them result in the same emitting event occupying two adjacent frames (Supplementary Fig. 10c), and the real average dark-time between blinking events is around 0.2s.

Validation of the existence of dynamic clustering in live cells

We identified pseudo-transient clusters in fixed cells with the exact procedures and criteria as for searching transient clusters in live cells. For spatially clustered structures, significantly larger dark times in live cells, compared to fixed cells under identical condition, is a sign of the bursting dynamics in live cells [8]. This is exactly what we observed (Supplementary Fig. 13d), and such larger dark times of clusters in live cells cannot be explained by longer intrinsic inter-detection period of Dendra2 single molecules in live-cell samples (Supplementary Fig. 13e). Furthermore, we normalized the number of tc-PALM identified bursts by the total number of detections of the same ROI, thus are able to estimate the number of identified bursts per 10,000 detections as 67.02 (Supplementary Fig. 13f). Meanwhile, among the tc-PALM identified bursts, we obtained the number of detections and lifetime of the 0.05 quantile at the lower-bound side as 4 and 0.85s, respectively (Supplementary Fig. 13f). If we use these two numbers as the cut-off for the set of Dendra2 single molecules we measured in live samples, only 4.67 molecules among 10,000 detections can pass the threshold. This indicates that the true positive rate (TPR) can easily go beyond 90%: 67.02÷(67.02+4.67)=93.5%; even in the worst case (all the bursts below the 0.05 quantile were single molecules), the corresponding TPR is 93.5%×95%≈89%. Even for the outlier single molecules that pass the cut-off, their statistics (including duration time, inter-detection period, and number of detections) are still quite different from the of tc-PALM identified bursts (Supplementary Fig. 13f). In another extreme test, we applied several additional high cut-offs to the tc-PALM identified bursts (in some cases, the TPR was pushed to 98%), whereupon we are still able to recapitulate all the significant trends of lifetime-shifting in cytoplasm and nuclei after different perturbations. This observation is reasonable: given that IR molecules are much less abundant in the cytoplasm and nuclei, un-clustered background of randomly bound IR molecules can be safely ignored. Therefore, any time-corelated, multi-detection events inside in the cytoplasm or nuclei are very likely to result from real clusters, which are insensitive to the FPR cut-off. Gathering all these evidences together, we are able to validate the existence of multi-molecule dynamical clustering of IR molecules in live cells, which yields transient bursting dynamics with distinct properties than single molecules and are robustly, physiologically responsive.

SUPPLEMENTARY REFERENCES

1. Ramakrishnan G, Arjuman A, Suneja S, Das C, Chandra NC. The association between insulin and low-density lipoprotein receptors. *Diab Vasc Dis Res* **9**, 196-204 (2012).

2. Hancock ML*, et al.* Insulin Receptor Associates with Promoters Genome-wide and Regulates Gene Expression. *Cell* **177**, 722-736 e722 (2019).

3. Foretz M, Guichard C, Ferre P, Foufelle F. Sterol regulatory element binding protein-1c is a major mediator of insulin action on the hepatic expression of glucokinase and lipogenesis-related genes. *Proc Natl Acad Sci U S A* **96**, 12737-12742 (1999).

4. Hillgartner FB, Salati LM, Goodridge AG. Physiological and molecular mechanisms involved in nutritional regulation of fatty acid synthesis. *Physiol Rev* **75**, 47-76 (1995).

5. Fortez M*, et al.* ADD1/SREBP-1c is required in the activation of hepatic lipogenic gene expression by glucose. *Mol Cell Biol* **19**, 3760-3768 (1999).

6. Claycombe KJ*, et al.* Insulin increases fatty acid synthase gene transcription in human adipocytes. *Am J Physiol* **274**, R1253-1259 (1998).

7. Dif N, Euthine V, Gonnet E, Laville M, Vidal H, Lefai E. Insulin activates human sterol-regulatory-element-binding protein-1c (SREBP-1c) promoter throug h SRE motifs. *Biochem J* **400**, 179-188 (2006).

8. Cisse, II*, et al.* Real-time dynamics of RNA polymerase II clustering in live human cells. *Science* **341**, 664-667 (2013).

Chapter 6:
Concluding Remarks

In this thesis, I have investigated protein spatiotemporal dynamics at two levels: the apparent diffusion of proteins as single molecules, and the formation and disassembly of condensates based on proteins' collective behaviors. Both levels of protein spatiotemporal dynamics are studied in the context of gene regulation and disease pathology.

Gene regulation at the transcription level requires meticulous coordination of transcriptional proteins in space and time. In Chapter 3, using transcription factors (TFs) as an example, I demonstrate that TFs interact with RNA to promote the transition from the fast diffusive state to subdiffusive and immobile states. Such transition may reflect enhanced chromatin occupancy needed for activating genes. In Chapter 4, in addition to single-molecule diffusion dynamics, we demonstrate that RNA synthesis is coupled to the condensate dynamics of formation and disassembly. Low levels of RNA synthesis at regulatory elements promote condensate formation, and high levels of RNA synthesis from gene transcription can dissolve condensates. Both studies reveal highly dynamic interactions between proteins and RNA, occurring at time and length scales previously uninvestigated, which play key roles in gene regulation. To conceptualize these interactions at a fundamental level, we applied the basic physical principle where like charges repel and unlike charges attract. However, it's crucial to acknowledge that the actual interactions are often more complex and influenced by other molecular factors. The next question will be: is there any specificity in such highly dynamic interactions? In other words, are proteins' spatiotemporal dynamics differentially affected by such interactions depending on peptide and RNA sequences? If so, is such specificity based on structured interactions, or is an integrated part of the emergent local chemistry established by the weak, multivalent interactions?

Abnormal protein spatiotemporal dynamics may be linked with protein dysfunction and thus disease pathogenesis. In Chapter 5, using the insulin receptor (IR) as an example, I demonstrate that IR exhibits diminished dynamics in condensate formation and disassembly under pathogenic conditions. This is inferred from the observed decrease in condensate size and increase in condensate lifetime. Meanwhile, IR's signaling function is dysregulated. Treatment of pathogenic cells with metformin, a first-line drug used to treat type 2 diabetes, can rescue IR's condensate dynamics and signaling function. In Chapter 2, in addition to IR, more proteins show compromised spatiotemporal dynamics under pathogenic conditions, indicated by decreased protein mobility. Such mobility responsiveness appears to be associated with proteins' oxidizable amino acids and the intracellular redox shift. We propose intermolecular disulfide bonding under oxidative-stress-associated pathogenic conditions as an underlying molecular mechanism. This study proposes a novel pathological hypothesis whereby subtle changes in common environmental factors may significantly impair protein mobility within cells. Such widespread impairment in protein mobility could potentially manifest as disease syndromes at the anatomical level, indicating a direct link between molecular dysfunction and broader physiological disruptions. In this thesis, this hypothesized pathology is only studied at the level of cell models and has been constrained to oxidative stress-associated diseases. For future research, I would like to propose further protein-mobility studies in animal models and patient tissues. I would also like to explore additional disease-relevant environmental perturbations besides oxidative stress that decrease protein mobility. Finally, I would like to use screening-based approaches to identify that control global protein mobility in normal conditions, and further identify the subset of such pathways that are disrupted in disease-relevant conditions.

Generally speaking, protein spatiotemporal dynamics—including localization, mobility, and collective behaviors—are determined by the interactions of a protein with its surrounding environment. There must be a variety of natural pathways—encompassing several factors and proteins—that directly or indirectly account for establishing intracellular physiology and controlling global protein spatiotemporal dynamics. Therefore, perturbation of such natural pathways will broadly influence the spatiotemporal dynamics of diverse proteins. For future research, screen-based approaches will be extremely useful in identifying natural pathways of protein spatiotemporal dynamics. Those pathways may be pervasive targets in diseases whose pathogenic mechanisms are not yet fully investigated from a dynamic point of view. In addition, for proteins with different biochemistries, their spatiotemporal dynamics may show differential sensitivity to certain environmental perturbations. Thus, future research in identifying protein biochemistries subject to various environmental changes is also intriguing and may pave the way for new generations of therapeutic solutions.

In conclusion, as a key parameter in defining a protein’s overall functionality, protein spatiotemporal dynamics provide a critical angle in viewing cellular processes and disease pathology. I illustrate such an angle through the lens of apparent diffusion and condensate dynamics of formation and disassembly, where the gene activation process and some cell models of chronic diseases are examined. Future studies incorporating other variations of this theme that cover different length/time scales, cellular processes, and complex diseases remain desirable.