Capturing Our Genome

by Ibrahim Cissé

In the Cissé lab, we develop high resolution imaging approaches, to go inside living cells and look at the scale of single biological molecules. We are particularly interested in the remarkable molecular machines that decode the genetic information from DNA into RNA. Here we discuss our discovery that these molecules interact transiently to exhibit liquid-like properties, and how these new insights are shaping our view of biological phase transition as a means for the cell to regulate the expression of our genetic material.

in Action

Imaging beyond the optical diffraction limit

Fluorescence microscopy is a time-honored method used for imaging living cells. The visible light being used (wavelength of ~500nm) is generally noninvasive to living matter. However, as light is a wave, diffraction limits the smallest features one can resolve to structures of the order of half a wavelength. This few hundred-nanometers resolution limit is orders of magnitude larger than the typical scale relevant for studies of single biomolecules, which is a few nanometers. Therefore, even with the most ideal optics, one would still be blind to the intricate world of molecular machines in living cells.

Approaches that overcome this limitation, and reveal biological organization and dynamics at sub-wavelength scales, employ so-called "super-resolution" fluorescence microscopes. The solution can be deceptively simple. For instance, by activating and localizing molecules one at a time, their centers can be localized with nanometer accuracy. Then, by localizing a large number of molecules one at a time, one can reconstruct the bigger picture as a pointillist image with very high localization precision for each individual point. Such "super-resolution" techniques based on single molecule detections were developed and popularized around 2006 and are known under acronyms like PALM (photo-activation localization microscopy) or

Biology Cheat Sheet for Physicists

In biophysics, we apply the theories and methods of physics to understand how biological systems work. But for many physicists, the last time they cracked open a biology textbook was decades ago as undergraduates. To help jog memories of those past days in molecular bio lab, here's a cheat sheet of fundamental biological terms.

The **CELL** is the fundamental unit of life. A cell has all the genetic material necessary to grow and divide into multiple cells.

An **ENZYME** is any substance that can help catalyze a biological or biochemical reaction.

A **GENE** is the unit of hereditary information; it's a portion of the long sequence of DNA, which encodes for specific genetic material (RNA and Proteins). Genetic is the adjective for anything related to genes or heredity.

DNA is the genetic molecule that encodes the set of all hereditary information.

RNA is the genetic molecule synthesized from the specific portion of DNA that is being decoded.

PROTEINS are genetic molecules synthesized from RNA. Proteins often fold (*i.e.,* collapse) into three-dimensional machines that do much of the heavy lifting in the cell. Proteins can do energetic work and act as enzymes to help catalyze all the biochemical reactions a cell needs.

STORM (stochastic optical reconstruction microscopy). Alternative approaches like STED (stimulated emissions depletion) and SIM (structured illumination) rely on non-linear properties of fluorescent molecules, without necessarily requiring single molecule sensitivity. These super-resolution approaches have transformed biophysical and biochemical studies, receiving the Nobel Prize in Chemistry in 2014 and the Breakthrough Prize in 2018.

A blind spot and adding the temporal dimension

Super-resolution approaches effectively trade time for achieving high spatial resolution. That is, it takes a long time to localize points individually and reconstruct the high spatial resolution images. For many years this was not considered to be a limitation as such super-resolution imaging was mostly performed on dead, chemically frozen (so-called fixed) biological samples. Or, if performed in live samples, the structures probed were relatively static over the course of imaging. However, this led to a blind spot in visualizing subdiffraction-sized macromolecular assemblies that are highly dynamic, as in living cells.

To overcome this limitation, we developed an approach known as *tcPALM* (time-correlated photo-activation localization microscopy) that analyses the temporal information from the super-resolution imaging performed directly on living biological samples *(Figure 1)*. The basic principle is to analyze the time series of the molecules detected in a region of interest. If the assembly were static the detections would exhibit a temporal correlation consistent with that of a stochastic process. Therefore, one can look for temporal correlations that are clearly distinct from that expected for the stochastic pointillist reconstruction process, to infer the dynamics of the biomolecular ensemble.

FIGURE 1

***tcPALM and Pol II cluster lifetime distribution**. Left: Super-resolution image of Pol II distribution reveals sub-diffractive clustering in a human cell. A red-hot color code is used to represent the relative density. Center: Zoomed-in view of a few clusters of Pol II. Right: A temporal correlation analysis reveals the transient dynamics of the Pol II clusters in living cells. [Adaptation from Cissé et al Science 2013.]*
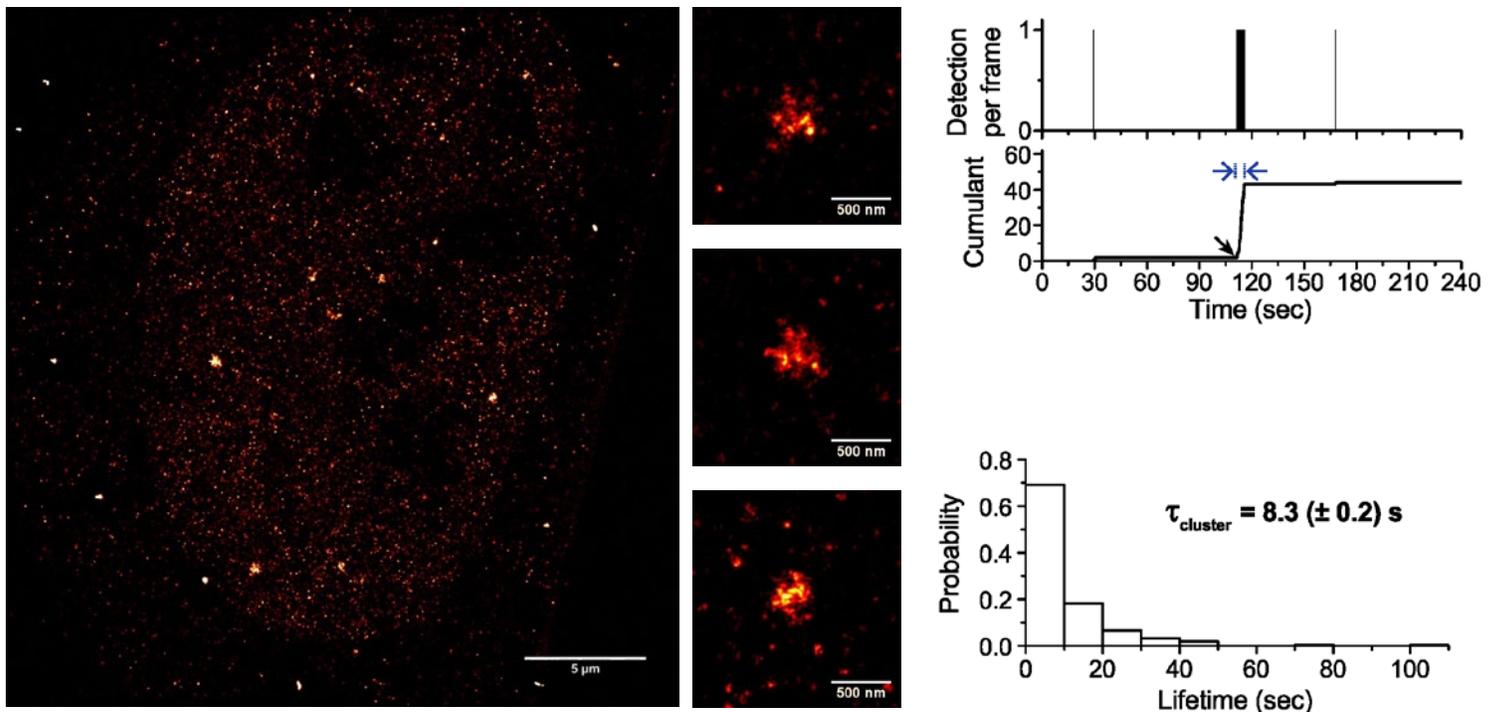
Pol II clusters too short-lived

In the first demonstration of tcPALM, we studied the dynamics of the enzyme Pol II, which is the molecular machine that is responsible for reading the information in DNA and converting it into a related polymer called RNA. This RNA molecule is a transient message that the cell uses to make proteins, which perform most of the work of the cell. Using tcPALM, we discovered that Pol II forms subdiffractive clusters that previously went unnoticed in living cells imaged by conventional fluorescence microscopes. Pol II had been hypothesized to form clusters in previous studies, but those observations were often dismissed as a possible artifact of the harsh chemical fixation used in those earlier studies, as there had been no evidence for such clusters in conventional live cell imaging. In addition to revealing the existence of such clusters in living cells, tcPALM could also measure their dynamics, and therein lay the surprise to us. Instead of being "stable factories," as previously hypothesized, we found that clusters of about 100 Pol II molecules assemble and then disassemble within a few seconds. And while a few seconds could sound like a long time, it is orders of magnitude faster than the time scale of Pol II activity.

Pol II is a complex molecular motor that assembles and tracks on DNA, unwinding the double helix and copying it one basepair (bp) at a time into RNA transcripts. The speed of Pol II is ~2,500bp per minute. A portion of DNA that Pol II copies is called a gene, and in mammalian (human and mice) cells that we study in our laboratory a typical gene size is about 10,000 bp. So it takes several minutes for Pol II to finish even one transcript. It was therefore unclear whether the very transient, seconds-long clusters had any significant role in the biological function of Pol II. However, given that this stems from studying a fundamental process at unprecedented spatiotemporal resolutions in living cells, we suspected that perhaps these puzzling transient cluster dynamics could be pointing to previously hidden mechanisms of gene regulation.

Dynamics predict the number of transcripts made

We next sought to test the putative function of transient Pol II clusters in living cells. Toward this end we developed dual color super-resolution imaging in living cells. One color would be used to perform tcPALM on Pol II as we had done earlier, and the second color would be used to look at the RNA transcripts that are the product of Pol II activity at a specific gene of interest.

This idea worked wonderfully *(Figure 2)*. We were able to capture transient Pol II clustering, precisely at the locus of a gene that is actively being transcribed, and correlate those measurements directly with the gene output—that is, nascent RNA transcripts inside the living cells.

We discovered that the dynamics of clustering predictively correlate with how many RNA copies are made. The clusters form in the early stages, as the gene transcription process is initiated. The cluster likely helps to increase the local concentration of Pol II enzymes at the gene to be transcribed. In a biochemical reaction, a high concentration increases the collision rate between enzyme and substrate.

In our case, the data suggests that in the presence of the cluster, there is a burst of activity at a rate of one Pol II enzyme successfully engaging on the DNA every two seconds. Since one engaged Pol II makes at most one RNA copy, the duration of the cluster determines how many RNA transcripts are made. In support of this picture, we observe a linear correlation between the lifetime of Pol II clusters at a gene locus and the number of RNAs that are subsequently synthesized as a burst at the same gene locus. Thus the dynamic Pol II clusters represent a previously uncharacterized mechanism for regulation of genes in living cells.
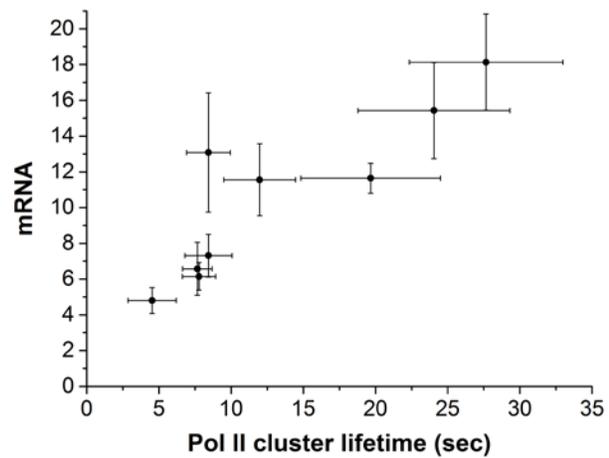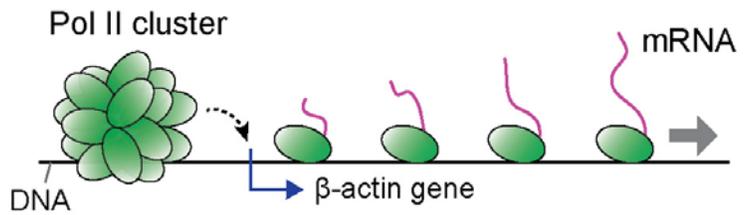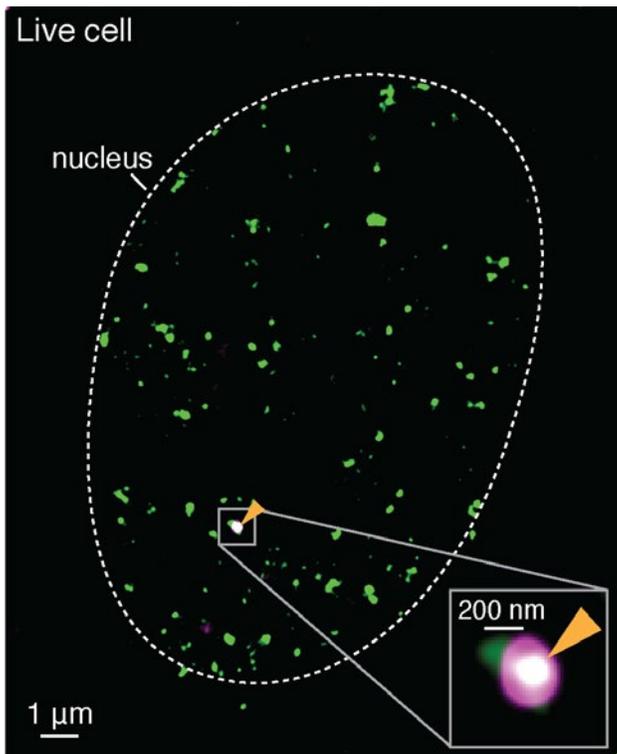
FIGURE 2

Cluster dynamics correlate with RNA output

Left: Dual color super-resolution imaging shows Pol II (false-colored green) clusters, and nascent RNA molecules (false-colored magenta) being synthesized at a particular gene inside a living cell. The nucleus of cell (white dash) is where genetic material (DNA, not pictured) is compartmentalized and where Pol II activity of transcribing DNA into RNA molecules takes place inside the living cell. Most genes are not labeled, only one gene locus (the so-called Beta-actin gene) is visualized here in magenta. Zoomed-in view: Pol II cluster is captured on the actively transcribing gene (green + magenta = white indicates colocalization of the two signals).

Right, top: Cartoon depicting how a Pol II (green) cluster can increase local concentration and help the loading of a few Pol II molecules, which ultimately track the DNA to copy the gene into RNA transcripts (magenta).

Right, bottom: Data shows that despite their transient lifetime, the longer Pol II cluster stays at the gene locus, proportionately more mRNA transcripts are made.

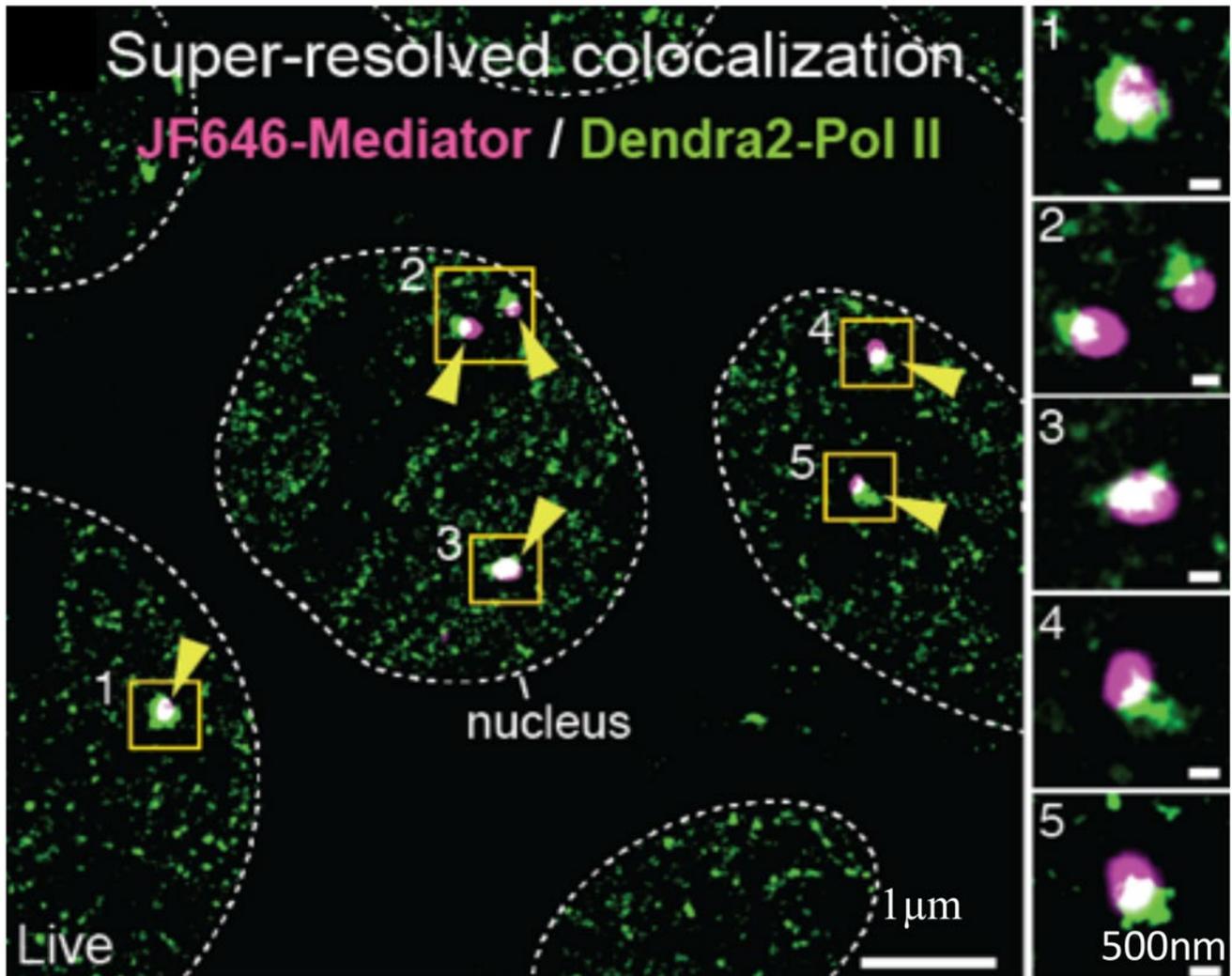[Credit: Adaption from Cissé Lab ms.: Cho et al Elife 2016.]

FIGURE 3

Pol II partner, Mediator, also clusters

To understand how the clusters form and what could control their dynamics, we looked at the molecular partners of Pol II. One such partner is the so-called *Mediator complex*. Mediator binds at regions on the DNA that are called an *enhancer*, which are distant from where Pol II binds and starts transcribing (which is called a *promoter*). The DNA is then believed to loop around so that the Mediator can directly contact Pol II. But when these models of Mediator–Pol II contacts were developed, it was believed that only a single Pol II complex would be present, not a cluster of 100 Pol II. We therefore wondered whether, in fact, Mediator also forms clusters similar to the Pol II clusters.

We found that indeed Mediator forms clusters similarly to Pol II. In fact, Mediator and Pol II clusters co-localize in the same loci in living cells *(Figure 3)*. Therefore, the dynamic clustering we discovered with Pol II may be general to many other regulatory molecules beyond Pol II.

Super-enhancers and mediator clusters, and liquid-like droplets

The recent studies of Mediator binding at enhancer sites used techniques that analyze the DNA sequence footprint in a population of chemically fixed cells. These techniques are so-called "ChIP-Seq." Our colleagues in the MIT Biology Department, Profs. Rick Young and Phil Sharp, utilized sequencing analysis and discovered that for many key genes, it is not a single Mediator binding site but tandem repeats of Mediator binding sites that are known as "super-enhancers." Together with Physics Prof. Arup Chakraborty, Young and Sharp and their trainees recently proposed that super-enhancers could be highly dynamic protein clusters, such as the ones we routinely observe in my lab. They further proposed that if these clusters exist they might exhibit properties of a liquid-like droplet, or a phase-separated biomolecular condensate.

We collaborated with Young, Sharp and Chakraborty, along with their trainees, to test these ideas directly *(Figure 4)*. While the ChIP-Seq technique used previously had been in chemically fixed cells (thereby not providing any dynamic information), we hypothesized that some of the clusters we can observe in live cells with tcPALM may correspond to super-enhancers. And indeed, we were success-

FIGURE 4

Pol II-Mediator fuse like droplets. *Left: Snapshot of video showing that the two clusters diffusing, upon contact, coalesce like droplets into a single cluster. Trajectory of the two clusters drawn in orange and blue respectively; the trajectory of the co-diffused merged cluster is drawn in red. Video available at https://science.sciencemag.org/highwire/filestream/711819/ field_highwire_adjunct_files/3/aar4199_s1.mov. Right: Quantitative analysis shows the intensity of the two clusters, and upon coalescence the merged cluster instantaneously has sum of the two intensities. [Adaption from Cissé Lab ms.: Cho et al Science 2018.]*
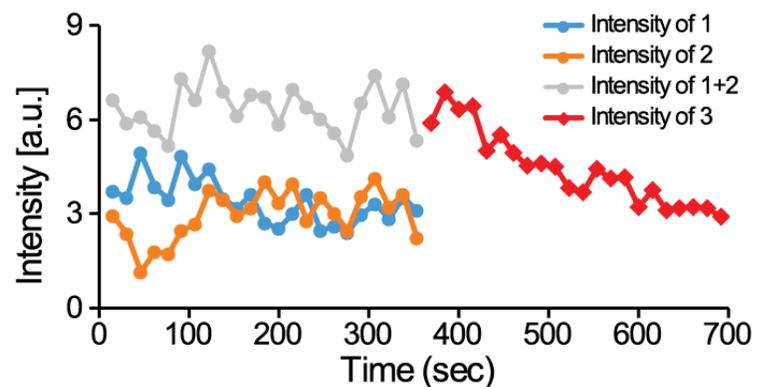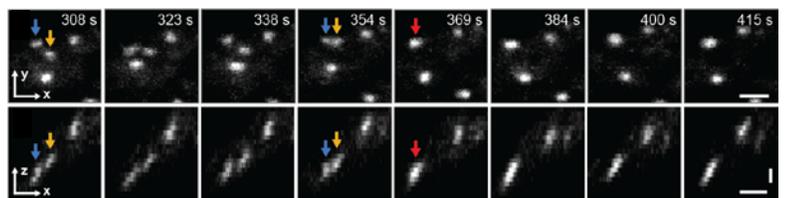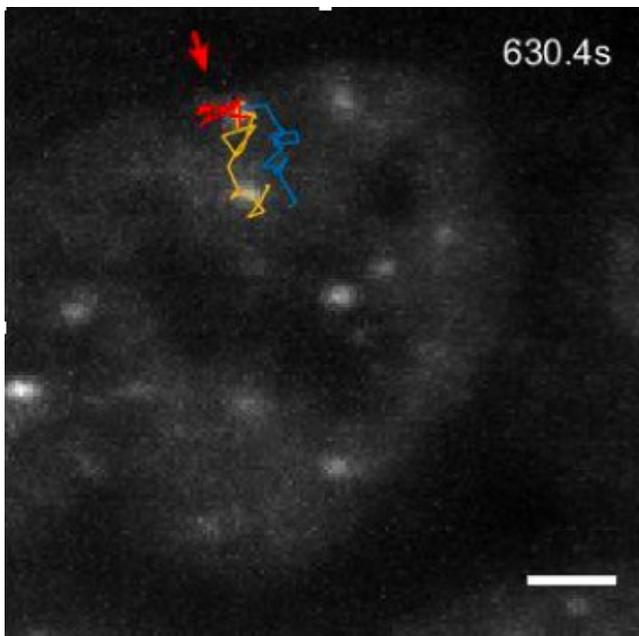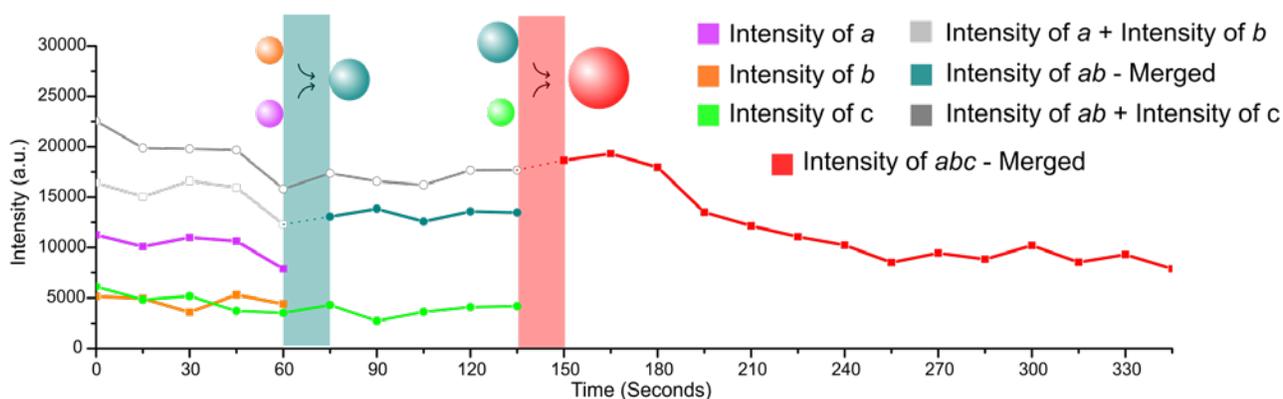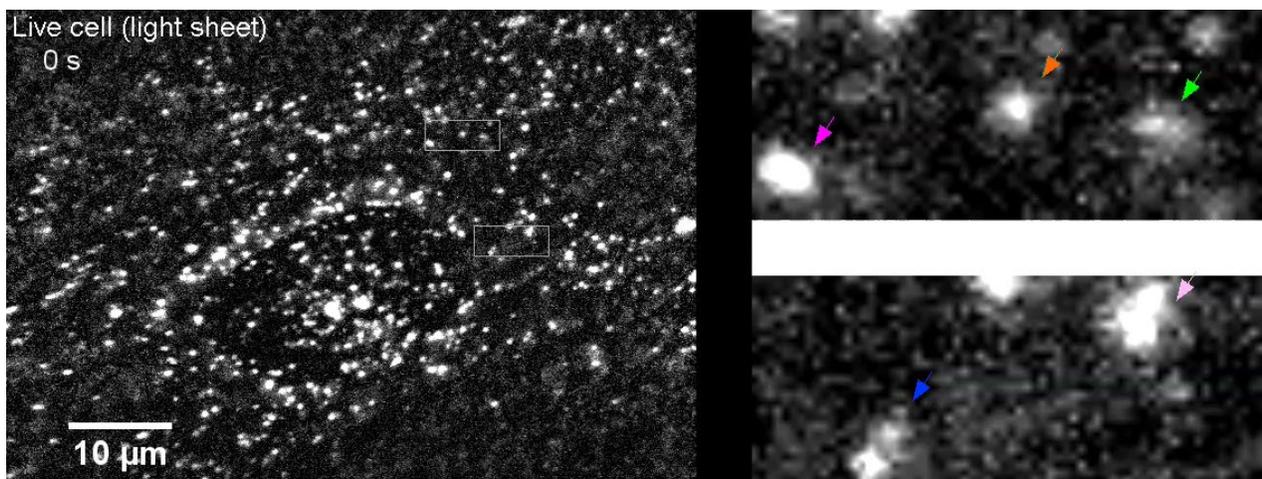
FIGURE 5

Other clusters fuse with similar biophysical properties as Pol II. Top: Screen shot from video showing molecules of interests to Parkinson's disease also form small clusters that fuse, which can now for the first time be studied with the same technologies that we had developed to study Pol II. For video, scroll down to Video 1 at https://elifesciences.org/articles/39695/figures. Bottom: Data showing the intensities of three clusters coalescing into one. Similarly to what we observed with other small biological droplets, here too upon merging, the resulting cluster has an intensity equal to the sum of the individual coalescing clusters. [Adaptation from Cissé Lab ms.: Narayanan et al eLife 2016]

ful in linking Mediator and Pol II clusters to super-enhancer regions, directly in living cells. Moreover, we were able to image the clusters of Mediator and Pol II and confirmed that they exhibited many properties of liquid-like biomolecular condensate—indeed, when they persisted long enough they ultimately merged like droplets.

On the quest for physical mechanisms describing the formation of biomolecular condensates in living cells

Moving forward, we are excited about investigating the physical mechanisms by which highly dynamic clusters, with liquid-like properties, form and dissolve in

living cells. Biomolecular condensates are in fact becoming ubiquitous *(Figure 5)* in many different aspects of cell function besides Pol II transcription. And there are likely general physical mechanisms that govern these phase transitions inside living cells.

We recently discovered that classical nucleation theory describing the condensation of a system in a steady state, super-saturation system describes how some diffraction-sized liquid-like condensates form and grow in living cells.

Classical nucleation theory: a first order phase transition mechanism for biomolecular condensates formation in living cells

Briefly, *classical nucleation theory* describes how a super-saturated system, say water vapor at high relative humidity, condenses into liquid droplets. It predicts that small clusters or drops will spontaneously form but rapidly dissolve due to a high surface tension. If the vapor concentration is higher than a critical (saturation) concentration, there emerges a maximum in the free energy, that is a critical size above which a nucleating cluster is stable enough to grow spontaneously. However, such a free energy is thought nearly impossible to measure experimentally: that is because the super-saturated state is a very transient metastable state. As soon as a few clusters reach the critical size, the clusters rapidly absorb monomers to grow, the ambient vapor concentration decreases rapidly, and the free energy maximum vanishes. Nonetheless, we suspected that with our high spatiotemporal resolution in living cells we might have a shot at capturing the small nucleating clusters, and perhaps measure for the first time directly the free energy of a super-saturated system in the living cells.
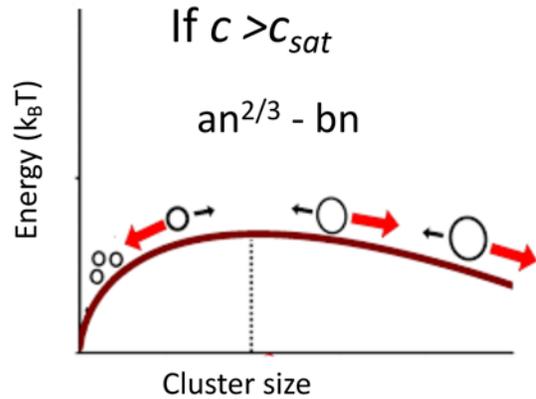
Szilard's demon and steady state super saturation

A toy model proposed by Hungarian-American physicist Leo Szilard suggested that super-saturation could be maintained in steady state, in theory, if one could imagine a "Maxwell's demon" that surveys nucleating clusters and removes those that reach the critical cluster size. However, Szilard's model was thought never to exist in natural systems where phase transitions were more commonly observed in physics. Nonetheless, we posited that in biological phase transitions one could conceive that, through evolution, living cells evolve such a Szilard demon as a regulation mechanism, especially in cases where large clusters can be detrimental to the cell function. We found this to be exactly the case in clusters associated with Parkinson's disease, a neurological disease.

We took a high-resolution snapshot of the clusters associated with Parkinson's, thereby measuring the probability distribution of cluster sizes. From that distribution, which the theory predicts should be Boltzmann's distribution for cluster sizes below criticality, we can measure the associated free energy function. To our surprise, the free energy was exactly as predicted in classical nucleation theory with a term driven by surface tension (which dominates at very small cluster sizes), balanced by a second term attributable to relative super-saturation of the system.

Theory

If $c > c_{sat}$

$an^{2/3} - bn$

Energy (k_BT)

Cluster size

We could therefore measure the energy barrier for nucleation (7 k_BT); the critical cluster radius (160nm); bounds for the surface tension, and more. Importantly for disease research, we discovered that the pre-nucleating cluster is in fact a liquid-like biomolecular condensate, and we uncovered Szilard's mechanism, involving a protein that seems to preferentially remove clusters above the critical size *(Figure 6)*.

Moving ahead

Armed with the excitement that such biomolecular phase transition mechanisms can be quantitatively addressed by our experimental approaches, we are hopeful that we can ultimately uncover the hidden steps by which Pol II condenses into clusters, and how cells might have evolved to regulate these clusters. We suspect this is a place where cells have evolved to utilize a hidden and complex mixture of physics, biology and chemistry to control one of the most fundamental processes of life: decoding the hereditary information we inherit in the form of DNA.

Data

$R_c = n_c^{1/3} = 162 \text{ nm} \pm 4 \text{ nm}$

$\Delta G_{barrier} = 7.2 \pm 0.5 \ k_B T$
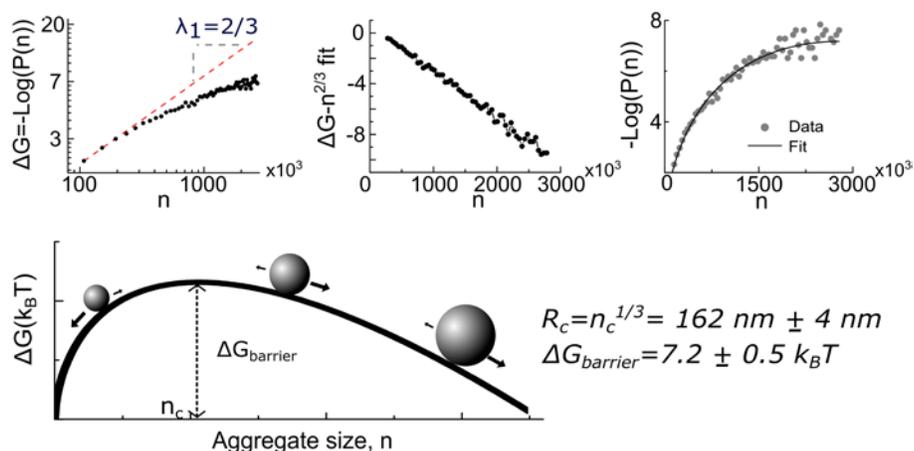
IBRAHIM CISSÉ *is an Associate Professor of Physics and Biology at the Massachusetts Institute of Technology. He is a native of Niger, Africa, where he lived until moving to the U.S. for college. He received his BS in Physics in 2004 from North Carolina Central University, and during that time his research was to investigate the packing of ellipsoids using M&M candies, under the guidance of Prof. Paul M. Chaikin (then at Princeton University). Cissé received his PhD in physics from the University of Illinois at Urbana-Champaign in 2009. His graduate training was at the interface between physics and biology, working under Prof. Taekjip Ha. Cissé then moved to Paris, where he was a postdoctoral fellow from 2010-2012 at Ecole Normale Supérieure, jointly in the departments of physics and biology, as a Pierre Gilles de Gennes fellow and a European Molecular Biology Organization long-term fellow, respectively. He returned to the US in 2013 to work as a research specialist at the Howard Hughes Medical Institute's Janelia Research Campus, before joining the MIT Department of Physics as an assistant professor in January 2014. In July 2020 he was promoted to associate professor with tenure. Cissé has received multiple national and international awards, including the Young Fluorescence Investigator award from the Biophysical Society, The Pew Biomedical Scholars, and the National Institutes of Health Director's New Innovator Award. In 2019, he received the Institute's Everett Moore Baker Award for Excellence in Undergraduate Teaching, the only teaching award at MIT where faculty are nominated and chosen by students.*