

RNA polymerase II clustering through carboxy-terminal domain phase separation

Marc Boehning^{1,6}, Claire Dugast-Darzacq^{2,3,6}, Marija Rankovic^{4,6}, Anders S. Hansen^{2,3}, Taekyung Yu⁵, Herve Marie-Nelly^{2,3}, David T. McSwiggen^{2,3}, Goran Kokic¹, Gina M. Dailey^{2,3}, Patrick Cramer^{1*}, Xavier Darzacq^{2,3*} and Markus Zweckstetter^{1,4,5*}

The carboxy-terminal domain (CTD) of RNA polymerase (Pol) II is an intrinsically disordered low-complexity region that is critical for pre-mRNA transcription and processing. The CTD consists of hepta-amino acid repeats varying in number from 52 in humans to 26 in yeast. Here we report that human and yeast CTDs undergo cooperative liquid phase separation, with the shorter yeast CTD forming less-stable droplets. In human cells, truncation of the CTD to the length of the yeast CTD decreases Pol II clustering and chromatin association, whereas CTD extension has the opposite effect. CTD droplets can incorporate intact Pol II and are dissolved by CTD phosphorylation with the transcription initiation factor IIH kinase CDK7. Together with published data, our results suggest that Pol II forms clusters or hubs at active genes through interactions between CTDs and with activators and that CTD phosphorylation liberates Pol II enzymes from hubs for promoter escape and transcription elongation.

Cellular processes often require clustering of molecules to facilitate their interactions and reactions^{1,2}. During transcription of protein-coding genes, Pol II clusters in localized nuclear hubs³. Whereas Pol II concentration in the nucleus is estimated to be ~1 μM , it increases locally by several orders of magnitude⁴. Such high Pol II concentrations are reminiscent of the clustering of proteins in membraneless compartments such as P granules, Cajal bodies, and nuclear speckles^{1,2,5,6}. These cellular compartments are stabilized by interactions between intrinsically disordered low-complexity domains (LCD) and depend on liquid–liquid phase separation (LLPS)^{1,2,6–11}. However, the molecular basis of Pol II clustering remains unknown.

The largest subunit of Pol II, RPB1, contains a C-terminal low-complexity domain, CTD, that is critical for pre-mRNA synthesis and co-transcriptional processing¹². The CTD is conserved from humans to fungi, but differs in the number of its heptapeptide repeats, with the consensus sequence $\text{Y}_1\text{S}_2\text{P}_3\text{T}_4\text{S}_5\text{P}_6\text{S}_7$ ^{13,14}. Human CTD (hCTD) contains an N-terminal half, which comprises 26 repeats and resembles the CTD from the yeast *Saccharomyces cerevisiae* (yCTD), and a C-terminal half containing 26 repeats of more divergent sequence (Supplementary Fig. 1a). CTD sequences from different species all contain high numbers of tyrosine, proline, and serine residues (Supplementary Fig. 1b)^{13,15}. The most conserved CTD residues are Y_1 and P_6 , which are present in all 52 repeats of hCTD. Truncating the CTD of RPB1 in *S. cerevisiae* to fewer than 13 repeats leads to growth defects, and a minimum of eight repeats is required for yeast viability¹⁶. The CTD forms a mobile, tail-like extension from the core of Pol II¹⁴ that is thought to facilitate the binding of factors for co-transcriptional RNA processing and histone modification^{13,14}.

Despite its extremely high conservation, its essential functions, and the large number of related published studies, the unique

CTD structure and properties have remained enigmatic. Here we show that CTD can undergo cooperative LLPS in vitro, driven by weak multivalent interactions. We further show that CTD is critical for the formation of hubs of Pol II in human cells. Together with published results, we arrive at a model for gene activation that involves CTD-mediated Pol II clustering at active gene promoters and release of initiated polymerases from these clusters after CTD phosphorylation.

Results

CTD of Pol II phase separates into liquid-like droplets. To investigate whether LLPS of CTD may underlie Pol II clustering, we expressed and purified hCTD and yCTD from *Escherichia coli*. We used a prokaryotic expression system to prevent eukaryotic post-translational modifications. The biophysical properties of the purified CTD proteins were characterized using circular dichroism (Supplementary Fig. 1c). Circular dichroism spectroscopy showed that hCTD and yCTD are intrinsically disordered in solution (Supplementary Fig. 1c)^{17–20}, consistent with the low complexity of CTD sequences (Fig. 1a)¹⁴.

Next we investigated the ability of CTD to undergo LLPS using a combination of differential interference contrast microscopy and fluorescence microscopy. Differential interference contrast microscopy revealed the formation of micrometer-sized droplets at a concentration of 20 μM hCTD in the presence of 5–10% of the molecular-crowding agent dextran (Fig. 2a). Fluorescence microscopy demonstrated that hCTD molecules were strongly concentrated within the droplet interior compared to the surrounding milieu (Fig. 2a). At higher dextran concentration (16%), droplets could be detected at a concentration of 5 μM hCTD (Fig. 1b,c). The number of droplets increased with increasing protein concentration (Fig. 1c), consistent with the general concentration dependence of

¹Max Planck Institute for Biophysical Chemistry, Department of Molecular Biology, Göttingen, Germany. ²Department of Molecular and Cell Biology, University of California, Berkeley, California, USA. ³CIRM Center of Excellence, University of California, Berkeley, California, USA. ⁴Max Planck Institute for Biophysical Chemistry, Department of NMR-based Structural Biology, Göttingen, Germany. ⁵German Center for Neurodegenerative Diseases (DZNE), Göttingen, Germany. ⁶These authors contributed equally and are listed alphabetically: Marc Boehning, Claire Dugast-Darzacq, Marija Rankovic. *e-mail: Patrick.Cramer@mpibpc.mpg.de; darzacq@berkeley.edu; Markus.Zweckstetter@dzne.de

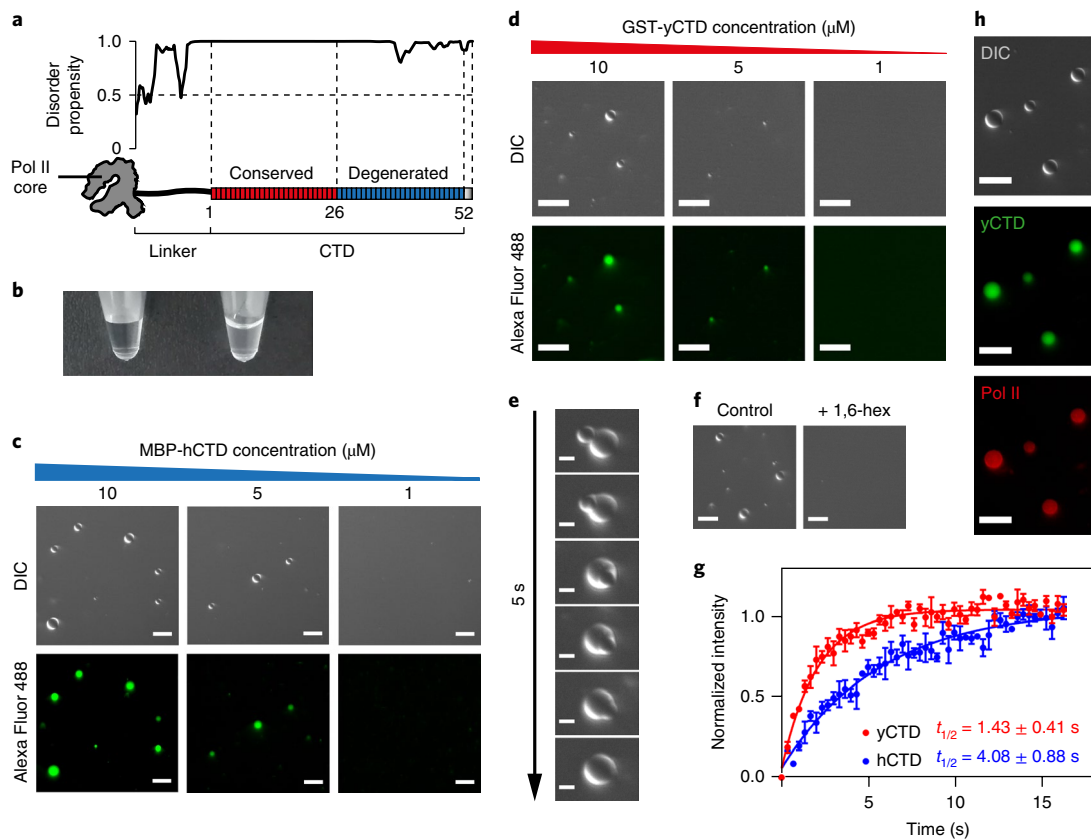


Fig. 1 | Phase separation of Pol II CTD into liquid-like droplets. **a**, Disorder analysis (top) and schematic view of the low-complexity hCTD sequence of Pol II, comprising 52 conserved heptapeptides (bottom). The N-terminal half (red) is composed almost exclusively of consensus repeats (Y₂S₂P₃T₄S₃P₆S₇) and is highly similar to yCTD. **b**, Addition of 16% dextran to a 20 μ M solution of MBP-hCTD turns the solution turbid, a characteristic property of liquid phase separation. **c**, Differential interference contrast (DIC) and fluorescence microscopy demonstrate the concentration-dependent formation of liquid droplets of MBP-hCTD in the presence of 16% dextran. Images are representative of three independent experiments. **d**, Concentration-dependent liquid phase separation of glutathione S-transferase (GST)-tagged yCTD (GST-yCTD) in the presence of 16% dextran. Images are representative of three independent experiments. **e**, GST-yCTD droplets rapidly fuse upon contact into one spherical droplet (25 μ M GST-yCTD in 16% dextran). **f**, Liquid phase separation of yCTD is sensitive to 1,6-hexanediol (1,6-hex; 10%). Images from at least five representative images taken for both conditions are shown. **g**, FRAP kinetics of photobleaching a spot within hCTD (blue) and yCTD (red) droplets, which were formed in the presence of 16% dextran. Data points represent mean values across three independent replicates and error bars show the standard error. **h**, Pol II (red, Alexa Fluor 594) is concentrated in preformed yCTD droplets (green, Alexa Fluor 488). Representative images from one of three independent experiments are shown. Scale bars, 10 μ m in **c,d,f,h** and 2.5 μ m in **e**.

liquid phase separation²¹. In addition, hCTD formed droplets in the presence of another molecular-crowding agent, the polysaccharide Ficoll (Fig. 2b). hCTD also underwent LLPS after cleavage of the maltose-binding protein (MBP) tag, while MBP alone did not form droplets in the presence of molecular-crowding agents (Fig. 2c). hCTD droplet formation was robust against changes in ionic strength (Fig. 2d) and against incubation of the sample for 1 h at different temperatures (Fig. 2e). Like hCTD, yCTD formed droplets in a concentration-dependent manner (Fig. 1d and Supplementary Fig. 2a,b). Contacts of both hCTD and yCTD droplets led to fusion and formation of a single spherical droplet (Fig. 1e and Supplementary Videos 1 and 2). At concentrations subcritical for LLPS, yCTD was incorporated into preformed hCTD droplets and hCTD was included into preformed yCTD droplets (Supplementary Fig. 2c), in agreement with the ability of CTD to be trapped into droplets and hydrogels of LCD proteins^{22,23}. Formation of yCTD droplets was also resistant against changes in ionic strength (Supplementary Fig. 2d) and temperature (Supplementary Fig. 2e), similarly to hCTD. The combined data show that the CTD of Pol II formed LCD–LCD interactions and readily underwent LLPS to form liquid-like droplets in solution.

Liquid droplets and cellular puncta are held together by weak, distributed interactions between LCDs that are sensitive to aliphatic alcohols^{6,24,25}. As expected for such interactions, liquid phase separation of yCTD and hCTD was counteracted by addition of 5–10% 1,6-hexanediol (Fig. 1f and Supplementary Fig. 3a,b). Addition of 5–10% of the hexanediol isomer 2,5-hexanediol also inhibited CTD droplet formation (Supplementary Fig. 3a,b). Because it has been shown that 2,5-hexanediol is less efficient in dissolving droplets and hydrogels²⁶, the data indicate that CTD droplets are more sensitive to aliphatic alcohols than other LCD–LCD interactions. On the contrary, CTD phase separation is robust to changes in ionic strength (Fig. 2d and Supplementary Fig. 2d).

CTD length influences CTD phase separation in vitro. A characteristic property of liquid-like droplets is fast diffusion of molecules in their interior¹. We used fluorescence recovery after photobleaching (FRAP) to compare diffusion kinetics of hCTD and yCTD molecules within droplets. MBP-tagged hCTD and yCTD proteins were fluorescently labeled on a single cysteine residue that is present C-terminally to the tobacco etch virus protease cleavage site. After cleavage of the MBP tag and droplet formation, circular regions

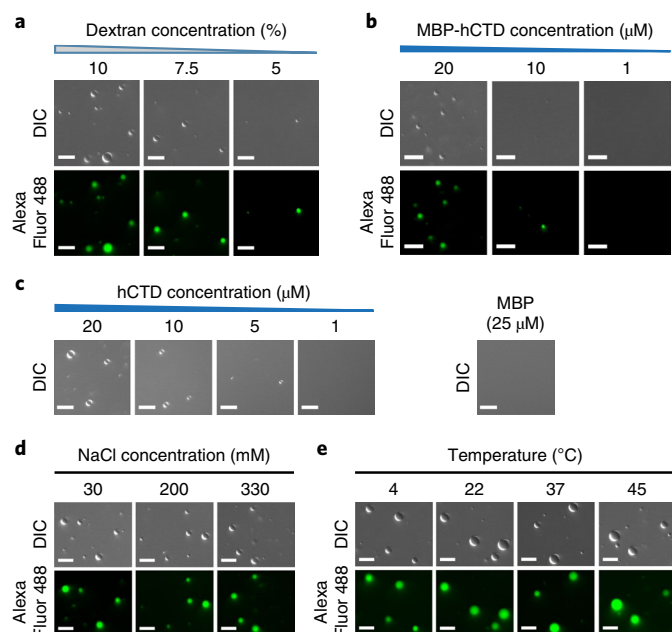


Fig. 2 | Physicochemical properties of hCTD phase separation.

a, Phase separation of MBP-hCTD (20 μ M) in the presence of different concentrations of dextran. **b**, Concentration-dependent LLPS of MBP-hCTD in the presence of Ficoll (150 mg/mL). **c**, Concentration-dependent LLPS of hCTD after cleavage of the MBP tag with tobacco etch virus (TEV) protease (left) in 16% dextran. MBP alone (25 μ M) did not undergo phase separation (right) under these conditions. For panels **a–c**, images shown are representative of at least five images recorded for each condition. **d,e**, LLPS of MBP-hCTD in 16% dextran shows little sensitivity to ionic strength changes (**d**) and is stable during incubation of the sample at different temperatures for 1 h (**e**). Images shown are representative of two independent experiments. Scale bars, 10 μ m.

in the interior of CTD droplets were bleached. Within hCTD droplets, the bleached fluorescence recovered with a half time of 4.08 ± 0.88 s (Fig. 1g). For yCTD, recovery was faster, with a half time of 1.43 ± 0.41 s (Fig. 1g).

These results demonstrate that CTD molecules within droplets were generally highly dynamic, confirming the liquid-like nature of CTD droplets. The difference in fluorescence recovery between hCTD and yCTD further suggests that the higher number of repeats in hCTD strengthened CTD–CTD interactions. This observation is consistent with the concentration-dependent ability of hCTD and yCTD to undergo LLPS when fused to the MBP tag. MBP-hCTD phase separated at a concentration of 5 μ M (Fig. 1c and Supplementary Fig. 3c). In contrast, LLPS of MBP-yCTD started only at a four- to sixfold higher protein concentration (Supplementary Fig. 3c). When the smaller, dimerizing glutathione S-transferase tag was used to replace the more soluble MBP tag²⁷, the critical concentration for yCTD phase separation decreased to approximately 5 μ M (Fig. 1d and Supplementary Fig. 3c). These results suggest that the solubilizing effect of MBP counteracts droplet formation. This effect is more easily overcome by hCTD because the higher repeat number and valency results in stronger CTD–CTD interactions compared to yCTD. We conclude that the length of CTD influences the stability and dynamics of LLPS droplets, with a longer CTD leading to stronger CTD–CTD interactions and less-dynamic droplets.

CTD droplets recruit intact Pol II. The above results indicate that CTD–CTD interactions within liquid droplets may underlie Pol II clustering. However, we could not test directly whether intact Pol

II forms LLPS droplets *in vitro* because it was impossible to prepare Pol II at a sufficient concentration in the presence of dextran or Ficoll. We could, however, test whether Pol II could be trapped within CTD droplets. We purified Pol II from yeast cells, labeled it with the fluorescent dye Alexa Fluor 594, and added it to preformed CTD droplets at a concentration of 0.02 μ M. Fluorescence microscopy showed that Pol II located to CTD droplets (Fig. 1h).

CTD length controls Pol II clustering in human cells. To explore whether CTD-based LLPS may underlie Pol II clustering in cells, we engineered two human cell lines that express a fluorescent Dendra2-tagged version of RPB1. To create these cell lines, we transfected cells with a plasmid containing an α -amanitin-resistant RPB1 variant (N792D) and selected cells in the presence of α -amanitin, which leads to the degradation of endogenous RPB1³. Such cell lines are known to recapitulate the behavior of endogenous wild-type Pol II^{3,28–30}. One cell line contained the full-length CTD with 52 repeats (RPB1-52R), whereas the other cell line contained a truncated CTD with 25 repeats (RPB1-25R) that closely resembles the yCTD sequence (Fig. 3a,b). The two cell lines remained viable upon degradation of endogenous RPB1 after treatment with α -amanitin and expressed similar levels of the Dendra2-tagged exogenous Pol II, as assessed by western blotting (Fig. 3c), confocal imaging, and fluorescence-activated cell sorting (Fig. 3d,e). The two cell lines also grew at similar rates (Fig. 3f).

We then studied clustering of Pol II in these human cell lines with the help of three-dimensional photoactivated localization microscopy (3D-PALM) using induced astigmatism by a cylindrical lens (Fig. 4)^{3,31}. Compared to cells with full-length CTD (52R), cells with the truncated, yeast-like CTD (25R) showed less Pol II clustering (Fig. 4a,b). These results suggested that CTD interactions underlie Pol II clustering in cells and that the CTD length influences clustering. To test this directly, we further created a cell line containing an artificially extended CTD (RPB1-70R; see Methods). This cell line was also viable and grew at a similar rate as the other two lines upon degradation of endogenous RPB1 (Fig. 3f), though it expressed RPB1 at a lower level (Fig. 3c). Despite this difference in expression level, the 70R cell line showed even more Pol II clustering than cells with wild-type, full-length CTD (Fig. 4a,b), strongly supporting our findings.

For all three cell lines, differences in CTD-dependent cluster density were supported by quantitative analysis on the basis of a modified Ripley function, $L(r)$, which compares the spatial distribution of localizations to complete spatial randomness ($L(r)=0$ for all r)³². In all cells, $L(r)$ curves showed strong clustering signatures (Fig. 4c,d). Whereas the sharp increase observed at scales less than 100 nm can be influenced by photophysical effects, such as blinking of Dendra2³³, the continuous increase at larger spatial distances is representative of Pol II clustering at multiple length scales. Taken together, these results demonstrate that Pol II clustering in cells depends on the CTD and increases with increasing CTD length.

CTD length influences Pol II dynamics in cells. We next investigated the impact of CTD length on Pol II dynamics *in vivo* using two orthogonal approaches, live-cell single-particle tracking (SPT)³⁴ and FRAP experiments. Because these methods require a high signal-to-noise ratio and a photostable fluorescent label, we established cell lines with a Halo tag on RPB1-25R, RPB1-52R, and RPB1-70R (Supplementary Fig. 4). We then tracked single molecules of Pol II in live cells as demonstrated by single-step photoactivation and photobleaching (Fig. 5a,b and Supplementary Videos 3–5). Subsequent two-state kinetic modeling analysis assuming a free and bound state (Fig. 5c, Supplementary Fig. 5a, and Supplementary Note) revealed that 29.1% of wild-type Pol II (RPB1-52R) in live cells was immobile and therefore presumably chromatin-associated. The bound Pol II fraction was decreased to 21% in RPB1-25R cells and was increased

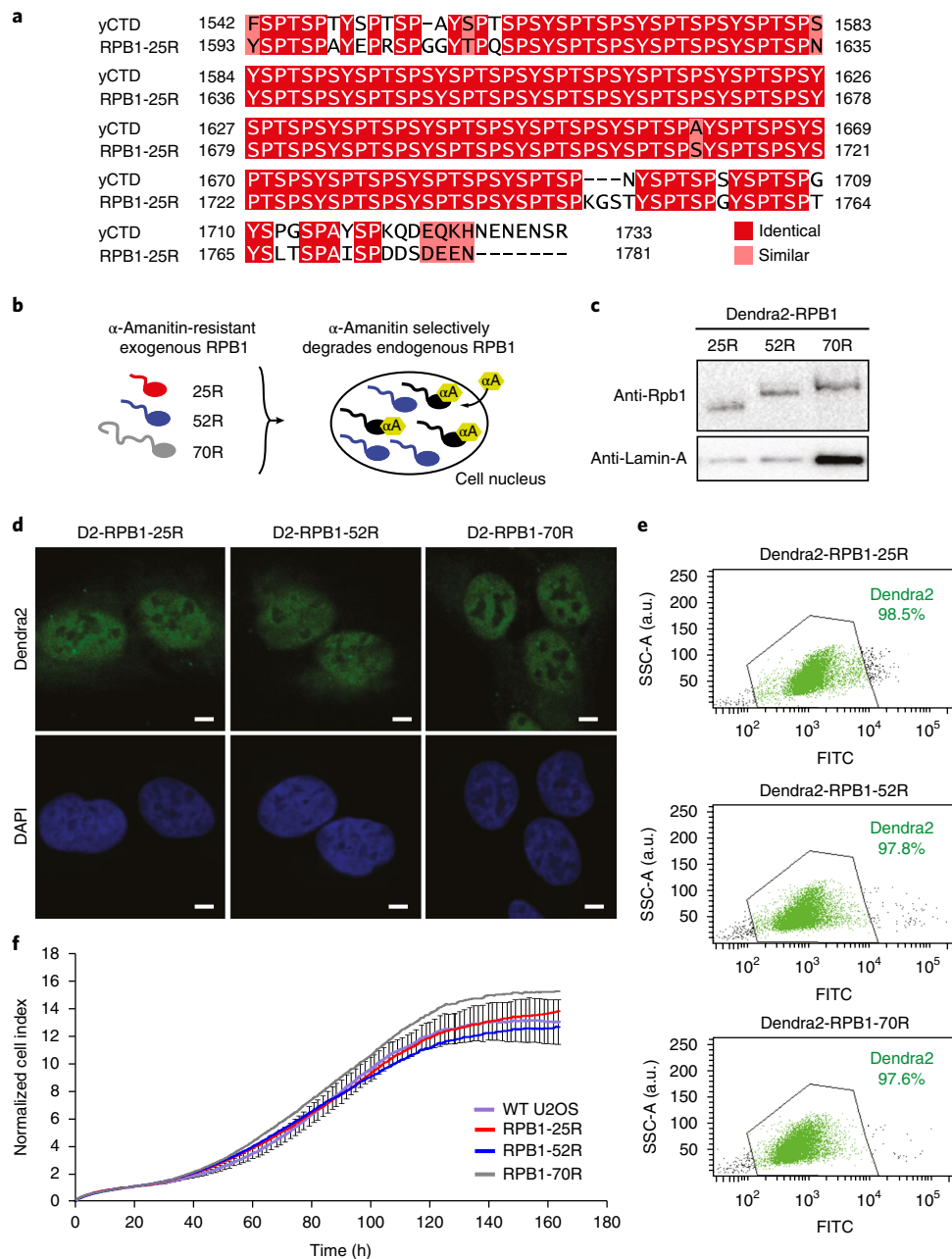


Fig. 3 | Human Dendra2-RPB1 cell lines for imaging CTD-dependent Pol II clustering. a, Pairwise alignment of CTD sequences from *S. cerevisiae* and human Dendra2-RPB1-25R cell line. The RPB1-25R cell line encodes a truncated version of human RPB1 comprising only 25 CTD repeats. This hCTD truncation closely resembles the yeast CTD in length (189 amino acids (aa) in RPB1-25R vs. 192 aa in yCTD) and amino acid composition (similarity, 87.2%; identity, 83.2%). The alignment was generated using the EMBOSS needle tool⁴⁴ with default settings, and aligned residues are colored red or pink according to their degree of similarity. **b**, Overview of the strategy used to establish the RPB1-25R, RPB1-52R, and RPB1-70R cell lines. Cells are transfected with a plasmid encoding an α -amanitin-resistant RPB1 variant (N792D). Upon growth in α -amanitin-containing medium, endogenous RPB1 is degraded and functionally replaced by the exogenous RPB1 variant. α A, α -amanitin. **c**, Western blot analysis of the expression level of Dendra2-RPB1-25R, -52R and -70R. RPB1-70R is expressed at lower levels than the other two proteins. Image representative of an experiment performed five times. Uncropped blot is shown in Supplementary Dataset 1. **d**, Confocal images of RPB1-25R (left), RPB1-52R (middle), and RPB1-70R (right) cell lines showing the nuclear localization of Dendra2-RPB1 in all three cell lines. Images are representative of five images taken for each cell line. D2, Dendra2. Scale bar, 5 μ m. **e**, Fluorescence-activated cell sorting (FACS) analysis to evaluate expression levels of the different Dendra2-RPB1 cell lines. One of three representative experiments is shown. FITC, fluorescence intensity in green fluorescein isothiocyanate channel; SSC-A, side scatter area. **f**, Growth-curve analysis of the three different Dendra2-tagged RNA Pol II cell lines in comparison to the osteosarcoma U2OS wild-type cell line (WT U2OS). The growth rates of all cell lines are similar. Growth curves are representative of an experiment performed independently three times and show the mean across six replicates. Error bars represent the s.d. and are shown only for the WT U2OS cell line to aid readability.

to 38.4% in RPB1-70R cells (Fig. 5d and Supplementary Fig. 5b,c). In addition, the diffusion coefficients for free Pol II were higher and lower, respectively, for RPB1-25R and RPB1-70R cells. Free diffusion

coefficients of 3.74, 2.97, and 2.34 μ m²/s were measured in RPB1-25R, RPB1-52R, and RPB1-70R cells, respectively (Fig. 5e). These large differences in diffusion coefficients cannot be explained solely

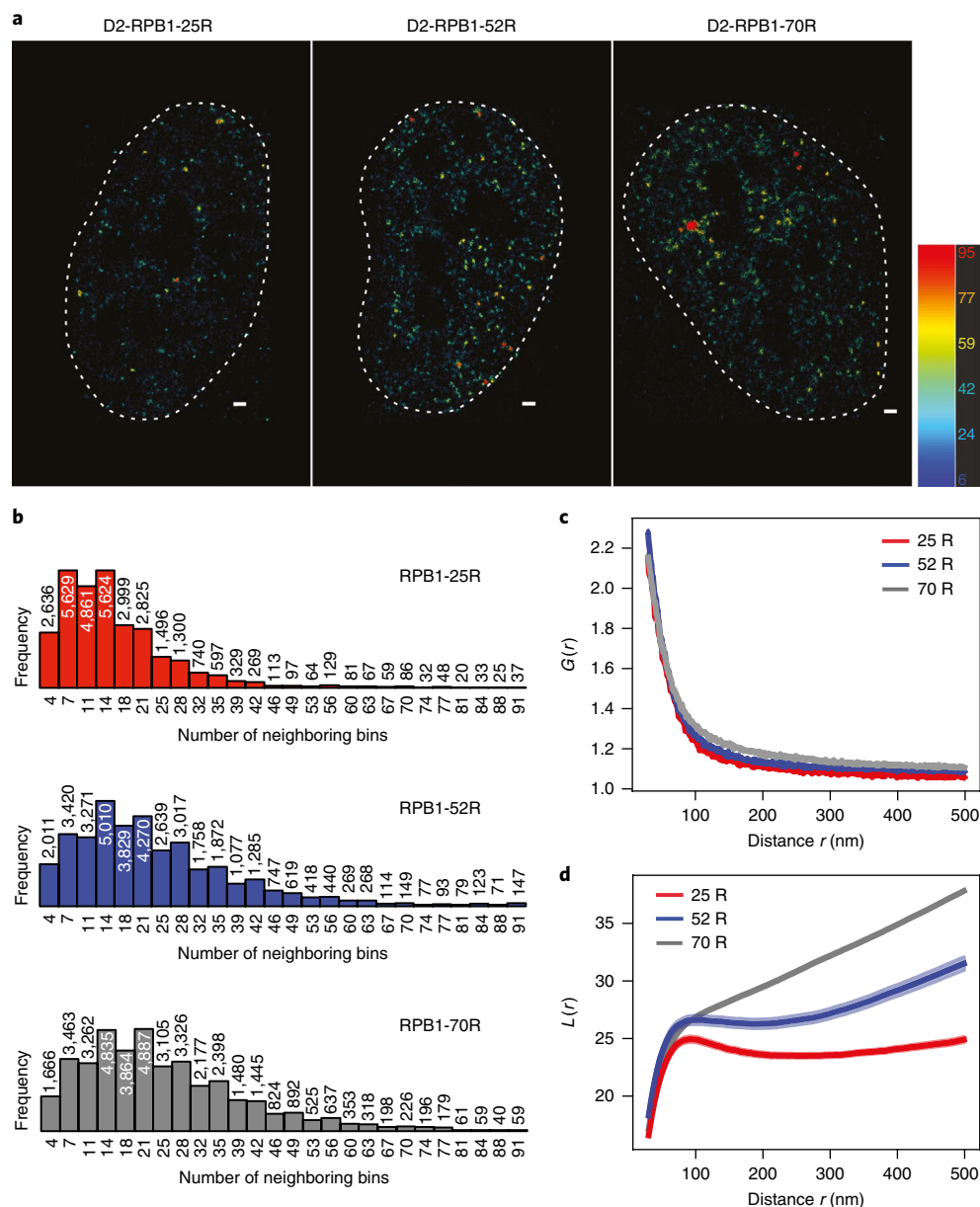


Fig. 4 | CTD-dependent Pol II clustering in human cells. **a**, 3D-PALM reconstruction of RPB1-25R (left), RPB1-52R (middle), and RPB1-70R (right). Each detection is color-coded by the number of detections within a surrounding radius of 120 nm (number of detections per 120-nm disk). Images are representative of six images taken for each cell line. Scale bars, 500 nm. **b**, Local density distribution (radius = 120 nm). Histograms of the average number of detections in a 120-nm-radius disk of Dendra2-RPB1-25R (top), Dendra2-RPB1-52R (middle), and Dendra2-RPB1-70R (bottom). Histograms are representative of six images taken for each cell line. **c**, G -pair correlation function (see Methods). The null hypothesis of complete spatial randomness is rejected because the curves strongly deviate from 1. A strong clustering signal is displayed for $r \geq 100$ nm. All things being equal (blinking, localization accuracy), Dendra2-RPB1-70R exhibits stronger clustering power than Dendra2-RPB1-52R, which is stronger than Dendra2-RPB1-25R ($P = 1.08 \times 10^{-21}$, Kolmogorov–Smirnov test). Analysis based on $n = 6$ independent measurements for each cell line. **d**, L -modified Ripley function. Linearized representation of the classic Ripley function. The null model of complete spatial randomness is rejected because the curves positively deviate from zero. All three curves exhibit strong clustering at all scales. Analysis based on $n = 6$ independent measurements for each cell line.

by differences in mass or size (Supplementary Note). Therefore, our results indicate that CTD length strongly influences Pol II mobility in vivo, with shorter and longer CTDs leading to higher and lower mobility, respectively.

These findings in cells match our observed length dependence of CTD–CTD interactions in vitro (Fig. 1g and Supplementary Fig. 3c). Indeed, FRAP recovery curves in human cells depended on CTD length (Fig. 5f), consistent with differences in FRAP recovery kinetics observed between hCTD and yCTD droplets in vitro (Fig. 1g).

Analysis of these FRAP recovery curves by a reaction-dominant two-state model^{35,36} further showed that the fraction that did not recover within a few seconds increased from 27.7% in RPB1-25R cells to 35.5% in RPB1-52R cells and to 38% in RPB1-70R cells (Supplementary Fig. 5d–f). This trend is consistent with the SPT results (Fig. 5d), which also showed a higher chromatin-associated fraction for Pol II with a longer CTD. Notably, both SPT and FRAP analysis showed that this putative chromatin-associated fraction of Pol II decreased to similar levels in all three cell lines after

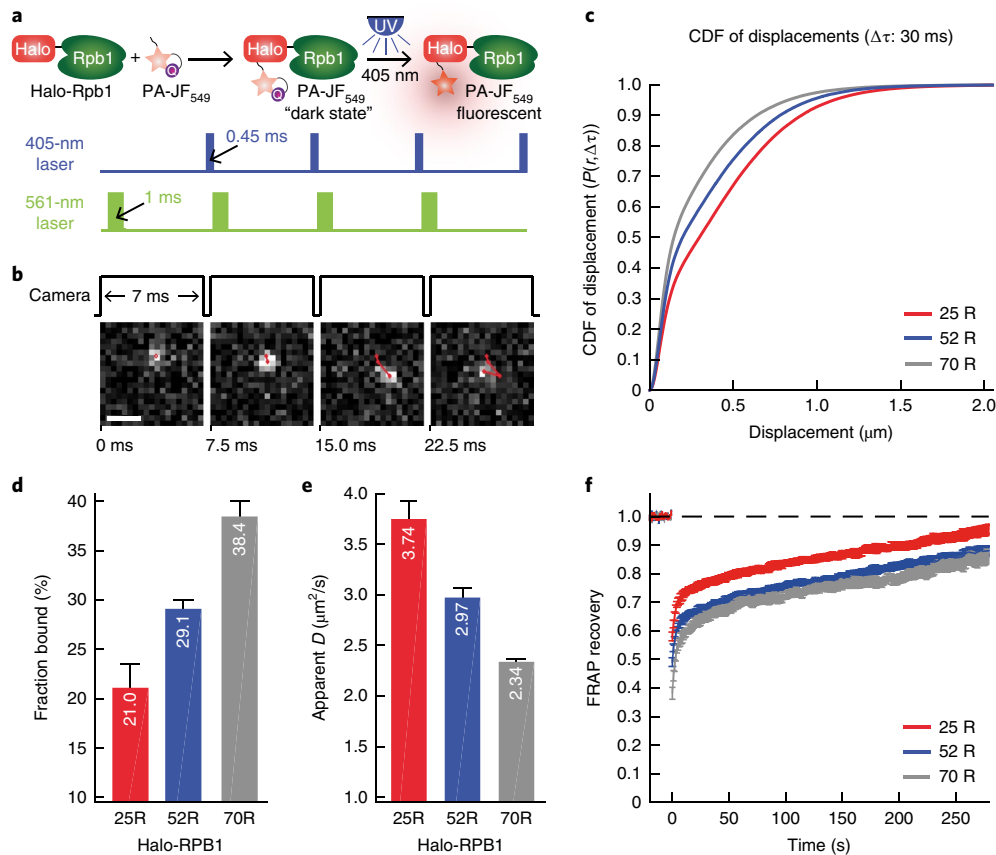


Fig. 5 | CTD-dependent Pol II dynamics in human cells. a, Overview of stroboscopic photoactivation SPT (spaSPT) at -133 Hz. Halo-RPB1 labeled with photoactivatable Janelia Fluor 549 (PA-JF₅₄₉) is photoactivated with a 405-nm laser and excited with 1-ms stroboscopic pulses of a 561-nm laser. This simultaneously minimizes motion-blurring by strobing the excitation laser and minimizes tracking errors by maintaining a low average density of ~1 localization per frame. **b**, Representative spaSPT images with overlaid trajectories. Scale bar, 1 μm . **c**, Cumulative distribution functions (CDF) for displacements. CDF of displacements for the representative time-lag $\Delta\tau = 30$ ms is shown for Halo-RPB1-25R, Halo-RPB1-52R, and Halo-RPB1-70R. Data shown are merged from three independent replicates ($n = 29, 30,$ and 26 cells in total for Halo-RPB1-25R, -52R, and -70R, respectively). **d**, Bound fractions of Halo-RPB1-25R, -52R, and -70R. The bound fraction was inferred from two-state model fitting to the spaSPT displacement data using Spot-On³⁴. Each of three independent replicates was fitted separately, and bar graphs show the mean and standard error. **e**, Diffusion coefficients (D) of the free populations of Halo-RPB1-25R, -52R, and -70R. Free diffusion coefficients were inferred from a two-state model fitting to the spaSPT displacement data using Spot-On³⁴. Each of three independent replicates was fitted separately, and bar graphs show the mean and standard error. **f**, FRAP dynamics. Mean drift and photobleaching-corrected FRAP recoveries are shown for Halo-RPB1-25R, Halo-RPB1-52R, and Halo-RPB1-70R. FRAP data were collected at 1 frame per s for 300 s, and bleaching was performed before frame 21. FRAP curves show means across three independent replicates ($n = 15, 15,$ and 8 cells in total for Halo-RPB1-25R, -52R, and -70R, respectively), and error bars show the standard error.

flavopiridol treatment, which blocks the transition into productive elongation by targeting positive transcription elongation factor B (P-TEFb; Supplementary Fig. 6). This favors an interpretation in which the CTD-length-dependent bound fraction is linked to polymerase activity. Together, our data show that longer CTDs result in more clustered Pol II and more chromatin association in vivo, reflecting the influence of CTD length on LLPS in vitro.

CTD phosphorylation dissolves droplets. Finally, we investigated whether CTD phosphorylation impacts phase separation. It has long been known that assembly of the pre-initiation complex at Pol II promoters requires an unphosphorylated CTD and that subsequent CTD phosphorylation at S₅ CTD residues by the cyclin-dependent kinase 7 (CDK7) in transcription factor IIH (TFIIH) stimulates the transition of Pol II into active elongation^{37,38}. We treated hCTD with recombinant human TFIIH subcomplex containing CDK7 kinase³⁹ and ATP, leading to S₅ phosphorylation of hCTD (Fig. 6a). The resulting CDK7-phosphorylated hCTD was no longer able to form droplets, whereas prior incubation with ATP alone did not

inhibit LLPS (Fig. 6b). Phosphorylation of γ CTD by the yeast TFIIH kinase subcomplex also inhibited phase separation (Supplementary Fig. 7). In addition, phosphorylation of preformed hCTD droplets by human CDK7 caused gradual shrinking and ultimately disappearance of hCTD droplets (Fig. 6c and Supplementary Video 6). Therefore, phosphorylation at S₅ is incompatible with CTD phase separation and transfers the CTD from the highly concentrated state within droplets to the dispersed pool.

Discussion

Here we show that the Pol II CTD can undergo length-dependent LLPS in vitro and that it controls Pol II clustering and mobility in vivo. Whereas CTD function is generally thought to depend on defined binary interactions of short CTD regions (1–3 repeats) with CTD-binding proteins, our results suggest that CTD function can additionally depend on weak homo- and heterotypic LCD–LCD interactions and that these interactions may dominate Pol II localization and dynamics in vivo. Although it was previously shown that CTD can interact with preformed LLPS droplets and hydrogels of

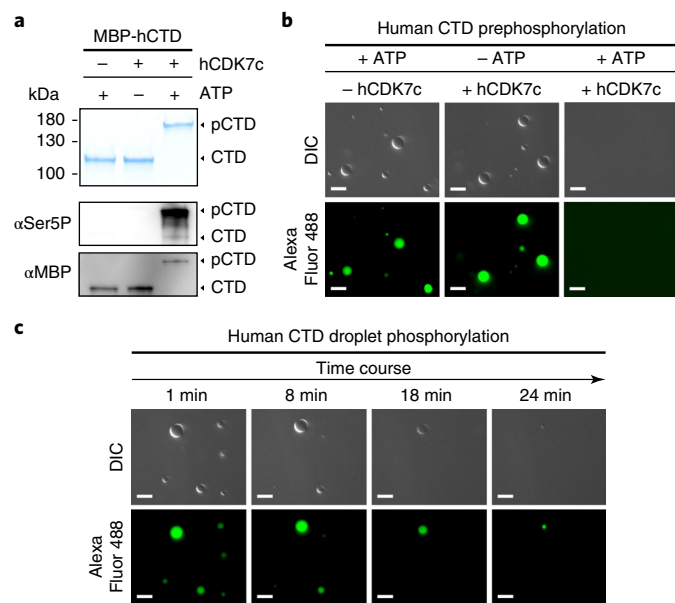


Fig. 6 | Liberation through initiation-coupled CTD phosphorylation.

a, Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and western blot analysis of phosphorylated MBP-hCTD fusion protein. MBP-hCTD was treated with recombinant human CDK7 complex. The hCTD substrate became highly phosphorylated, resulting in a pronounced mobility change during polyacrylamide gel electrophoresis in comparison to the nonphosphorylated substrate (control reactions without ATP and without kinase). Western blotting confirms phosphorylation of heptad position Ser5. Corresponding loading controls are shown to correct for potential differences in blotting efficiency. Western blot analysis of Ser5 phosphorylation was performed in duplicate. Uncropped gel and western blot images are shown in Supplementary Dataset 1. **b**, hCTD phase separation is inhibited upon CTD phosphorylation by the human TFIIF subcomplex containing the CDK7 kinase. This effect is caused neither by hydrophobic properties of ATP⁴⁵ nor by the simple presence of the kinase, since MBP-hCTD readily forms droplets in control reactions containing ATP or kinase alone. Images from at least ten representative images taken for each condition are shown. **c**, CDK7 phosphorylation dissolves preformed hCTD droplets over time. Images are representative of two independent experiments. Scale bars, 10 μ m.

FET (FUS-EWS-TAF15) proteins^{22,23}, our results demonstrate for the first time that the CTD alone, in absence of other proteins, can undergo phase separation. Whereas the correlation between our in vitro and in vivo data is striking, our data cannot demonstrate directly that Pol II undergoes phase separation in vivo. However, we provide strong evidence that the weak and multivalent interactions driving phase separation in vitro are the same as those that drive polymerase hubs within the nucleoplasm of living cells.

Our findings have implications for understanding Pol II transcription in eukaryotic cells and suggest a simple model for gene activation and the initiation–elongation transition during early transcription (Fig. 7). Unphosphorylated Pol II clusters, forming nucleoplasmic hubs in cells, mediated by CTD–CTD interactions. Pol II hubs may be recruited by transcriptional activators that bind to regulatory sites such as enhancers^{40,41}. Transcriptional activators can also undergo LCD interactions⁴² and might assist in Pol II hub formation when Pol II concentration is subcritical. Formation of Pol II hubs near gene promoters may provide a reservoir of Pol II, enabling high initiation rates during activated transcription. When a Pol II enzyme is incorporated into a pre-initiation complex, its CTD is phosphorylated by the TFIIF kinase CDK7. Phosphorylation

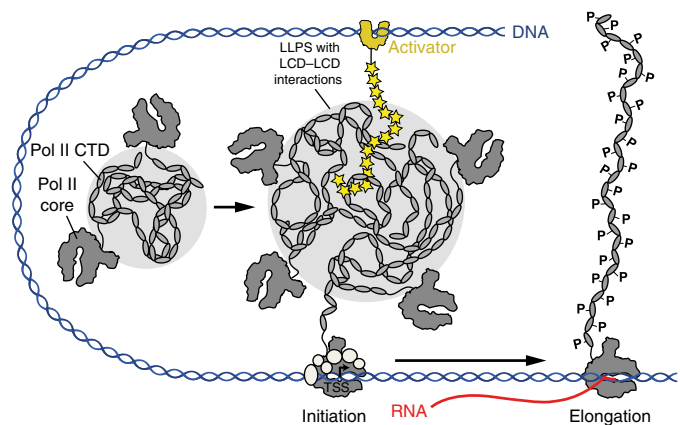


Fig. 7 | Model for the role of CTD-driven phase separation in activated transcription. Activators may recruit or nucleate Pol II hubs near gene promoters. Initiation-coupled CTD phosphorylation removes individual Pol II enzymes for transcription elongation.

removes this Pol II enzyme from the hub, liberates it to escape the promoter, and enables the transition into active transcription elongation (Fig. 7).

While our manuscript was under review, the Zhou laboratory published a paper showing that phase separation of the intrinsically disordered region of cyclin T1, a subunit of P-TEFb, promotes the inclusion of Pol II CTD into hubs⁴³. This effect was observed exclusively for the CTD in its CDK7-phosphorylated form, but not for its unphosphorylated form⁴³. Together with the current work, this suggests that Pol II CTD can undergo phase separation via at least two different mechanisms: in its unphosphorylated form, CTD phase separation is based mainly on weak hydrophobic interactions (Fig. 1f and Supplementary Fig. 3). Phosphorylation disrupts CTD–CTD interactions (Fig. 6) and successively allows the CTD to engage in novel multivalent interactions with selected factors. The latter interactions are likely primarily electrostatic in nature and can also lead to phase separation, as demonstrated for the phosphorylated CTD, together with the positively charged, intrinsically disordered region of cyclin T1⁴³. Our results and those of Lu et al.⁴³ are thus complementary and provide a starting point for analyzing the chemical basis of CTD phase separation, its possible modulation by nucleic acids and protein factors, and its specific roles in transcription regulation and the coordination of the transcription cycle. More generally, this provides a powerful and highly specific and regulated local protein-sorting mechanism modulating local proteomes within cells.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at <https://doi.org/10.1038/s41594-018-0112-y>.

Received: 5 May 2018; Accepted: 17 July 2018;

Published online: 20 August 2018

References

1. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
2. Boeynaems, S. et al. Protein phase separation: a new phase in cell biology. *Trends Cell Biol.* **28**, 420–435 (2018).
3. Cisse, I. I. et al. Real-time dynamics of RNA polymerase II clustering in live human cells. *Science* **341**, 664–667 (2013).
4. Cook, P. R. The organization of replication and transcription. *Science* **284**, 1790–1795 (1999).

5. Brangwynne, C. P. et al. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324**, 1729–1732 (2009).
6. Molliex, A. et al. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**, 123–133 (2015).
7. Han, T. W. et al. Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. *Cell* **149**, 768–779 (2012).
8. Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A phase separation model for transcriptional control. *Cell* **169**, 13–23 (2017).
9. Li, P. et al. Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012).
10. Martin, E. W. & Mittag, T. Relationship of sequence and phase separation in protein low-complexity regions. *Biochemistry* **57**, 2478–2487 (2018).
11. Csizmek, V., Follis, A. V., Kriwacki, R. W. & Forman-Kay, J. D. Dynamic protein interaction networks and new structural paradigms in signaling. *Chem. Rev.* **116**, 6424–6462 (2016).
12. Zaborowska, J., Egloff, S. & Murphy, S. The pol II CTD: new twists in the tail. *Nat. Struct. Mol. Biol.* **23**, 771–777 (2016).
13. Hsin, J. P. & Manley, J. L. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev.* **26**, 2119–2137 (2012).
14. Meinhardt, A., Kamenski, T., Hoepfner, S., Baumli, S. & Cramer, P. A structural perspective of CTD function. *Genes Dev.* **19**, 1401–1415 (2005).
15. Simonti, C. N. et al. Evolution of lysine acetylation in the RNA polymerase II C-terminal domain. *BMC Evol. Biol.* **15**, 35 (2015).
16. West, M. L. & Corden, J. L. Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics* **140**, 1223–1233 (1995).
17. Gibbs, E. B. et al. Phosphorylation induces sequence-specific conformational switches in the RNA polymerase II C-terminal domain. *Nat. Commun.* **8**, 15233 (2017).
18. Portz, B. et al. Structural heterogeneity in the intrinsically disordered RNA polymerase II C-terminal domain. *Nat. Commun.* **8**, 15231 (2017).
19. Janke, A. M. et al. Lysines in the RNA polymerase II C-terminal domain contribute to TAF15 fibril recruitment. *Biochemistry* **57**, 2549–2563 (2018).
20. Cagas, P. M. & Corden, J. L. Structural studies of a synthetic peptide derived from the carboxyl-terminal domain of RNA polymerase II. *Proteins* **21**, 149–160 (1995).
21. Hyman, A. A., Weber, C. A. & Jülicher, F. Liquid-liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014).
22. Burke, K. A., Janke, A. M., Rhine, C. L. & Fawzi, N. L. Residue-by-residue view of in vitro FUS granules that bind the C-terminal domain of RNA polymerase II. *Mol. Cell* **60**, 231–241 (2015).
23. Kwon, I. et al. Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell* **155**, 1049–1060 (2013).
24. Brangwynne, C. P., Tompa, P., Pappu, R. V. Polymer physics of intracellular phase transitions. *Nat. Phys.* **11**, 899–904 (2015).
25. Pak, C. W. et al. Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein. *Mol. Cell* **63**, 72–85 (2016).
26. Kato, M. & McKnight, S. L. A solid-state conceptualization of information transfer from gene to message to protein. *Annu. Rev. Biochem.* **87**, 351–390 (2018).
27. Kapust, R. B. & Waugh, D. S. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* **8**, 1668–1674 (1999).
28. Darzacq, X. et al. In vivo dynamics of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **14**, 796–806 (2007).
29. Becker, M. et al. Dynamic behavior of transcription factors on a natural promoter in living cells. *EMBO Rep.* **3**, 1188–1194 (2002).
30. Cho, W. K. et al. Super-resolution imaging of fluorescently labeled, endogenous RNA Polymerase II in living cells with CRISPR/Cas9-mediated gene editing. *Sci. Rep.* **6**, 35949 (2016).
31. Huang, B., Wang, W., Bates, M. & Zhuang, X. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science* **319**, 810–813 (2008).
32. Nicovich, P. R., Owen, D. M. & Gaus, K. Turning single-molecule localization microscopy into a quantitative bioanalytical tool. *Nat. Protoc.* **12**, 453–460 (2017).
33. Annibale, P., Vanni, S., Scarselli, M., Rothlisberger, U. & Radenovic, A. Identification of clustering artifacts in photoactivated localization microscopy. *Nat. Methods* **8**, 527–528 (2011).
34. Hansen, A. S. et al. Robust model-based analysis of single-particle tracking experiments with Spot-On. *eLife* **7**, e33125 (2018).
35. Sprague, B. L., Pego, R. L., Stavreva, D. A. & McNally, J. G. Analysis of binding reactions by fluorescence recovery after photobleaching. *Biophys. J.* **86**, 3473–3495 (2004).
36. Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R. & Darzacq, X. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *eLife* **6**, e25776 (2017).
37. Feaver, W. J., Svejstrup, J. Q., Henry, N. L. & Kornberg, R. D. Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIF/TFIIK. *Cell* **79**, 1103–1109 (1994).
38. Morgan, D. O. Principles of CDK regulation. *Nature* **374**, 131–134 (1995).
39. Conaway, J. W., Shilatifard, A., Dvir, A. & Conaway, R. C. Control of elongation by RNA polymerase II. *Trends Biochem. Sci.* **25**, 375–380 (2000).
40. Scafe, C. et al. RNA polymerase II C-terminal repeat influences response to transcriptional enhancer signals. *Nature* **347**, 491–494 (1990).
41. Gerber, H. P. et al. RNA polymerase II C-terminal domain required for enhancer-driven transcription. *Nature* **374**, 660–662 (1995).
42. Chong, S. et al. Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science*. <https://doi.org/10.1126/science.aar2555> (2018).
43. Lu, H. et al. Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. *Nature* **558**, 318–323 (2018).
44. Rice, P., Longden, I. & Bleasby, A. EMBL: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277 (2000).
45. Patel, A. et al. ATP as a biological hydrotrope. *Science* **356**, 753–756 (2017).

Acknowledgements

We thank S.M. Vos for the MBP-hCTD expression plasmid and advice on purification, S. Schilbach for the yeast TFIIF kinase module, and A. Boltendahl and C. Burzinski for help with cloning and purification. We are grateful to S. Ambadipudi for help with initial microscopy measurements and discussions about LLPS. Microscopic analysis of phase separation was conducted at the Facility for Innovative Light Microscopy at the MPI for Biophysical Chemistry. We thank L. Lavis for generously providing JF dyes. We thank A. Robles, C. Inouye, S. Zheng, M. Haggart, and J. Lim for technical and administrative assistance. In vivo FRAP experiments were conducted at the CRL Molecular Imaging Center, supported by the Gordon and Betty Moore Foundation. We thank all current members of the Cramer, Zweckstetter, and Tjian–Darzacq labs for discussions and R. Tjian for critical reading of the manuscript. M.Z. was supported by the Deutsche Forschungsgemeinschaft (SFB860; project B02), by the Cluster of Excellence and DFG Research Center Nanoscale Microscopy and Molecular Physiology of the Brain, and the advanced grant ‘787679-LLPS-NMR’ from the European Research Council. P.C. was supported by the Deutsche Forschungsgemeinschaft (SFB860; project A13), the advanced grant ‘TRANSREGULON’ from the European Research Council, and the Volkswagen Foundation. X.D. was supported by NIH grant U01-EB021236 and the California Institute of Regenerative Medicine grant LA1-08013.

Author contributions

M.B. designed experiments, generated constructs, and prepared proteins unless otherwise noted. C.D.-D. designed experiments, established and characterized the RPB1 cell lines, and performed and analyzed the in vivo FRAP and SPT experiments. M.R. designed experiments, performed all phase separation assays, DIC and fluorescence microscopy, in vitro FRAP measurements and data analysis. A.S.H. designed, performed, and analyzed SPT experiments and helped with the in vivo FRAP analysis. H.M.-N. designed, performed, and analyzed 3D-PALM experiments. D.T.McS. performed cell-viability experiments and helped in performing 3D-PALM experiments. G.M.D. designed and cloned the different RPB1 expression vectors. G.K. prepared human TFIIF kinase complex. T.Y. performed CD and NMR experiments. C.D.-D., X.D., P.C., and M.Z. designed and supervised research. M.B., M.R., C.D.-D., P.C., X.D., and M.Z. prepared the manuscript with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41594-018-0112-y>.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to P.C. or X.D. or M.Z.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Cloning and protein expression. A plasmid encoding the human Pol II CTD sequence (hCTD; RPB1 residues 1,593–1,970) fused C-terminally to a sequence encoding 6xHis-tagged maltose-binding protein (MBP) directly followed by a flexible linker (10xAsn) and a tobacco etch virus (TEV) protease cleavage site was a kind gift from S.M. Vos (Max Planck Institute for Biophysical Chemistry, Göttingen). Derivative constructs, in which the hCTD sequence was replaced by the sequence coding for the *S. cerevisiae* Pol II CTD (yCTD; RPB1 residues 1,542–1,733) or entirely removed, were generated using a Gibson Assembly (New England Biolabs) and through deletion mutagenesis, respectively. All proteins contain a single cysteine residue C-terminal to the TEV protease cleavage site, allowing for site-specific labeling. MBP-tagged proteins were overexpressed in *E. coli* BL21 (DE3) RIL cells (Stratagene) cultured in LB medium containing 50 mg/L kanamycin and 34 mg/L chloramphenicol. After reaching an OD₆₀₀ of ~0.8, 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added and proteins were expressed for 3–4 h at 37°C. Cells were harvested by centrifugation and resuspended in lysis buffer LB300 (20 mM HEPES, pH 7.4, 300 mM NaCl, 30 mM imidazole, 10% glycerol, 1 mM DTT, 0.284 μg/mL leupeptin, 1.37 μg/mL pepstatin A, 0.17 mg/mL PMSE, 0.33 mg/mL benzamidine). The cell suspension was snap-frozen and stored at –80°C.

The sequence coding for the yCTD was additionally cloned into a pET24-derived plasmid, directly C-terminal to a GST-tag followed by the TEV protease cleavage site. An N-terminal 6xHis-tag was introduced by site-directed mutagenesis. From the obtained plasmid, a second expression vector encoding only 6xHis-GST-TEV was constructed through deletion mutagenesis. GST-tagged proteins were overexpressed in *E. coli* BL21 Rosetta 2(DE3)pLysS cells (Stratagene) grown in 2xYT medium containing 50 mg/L kanamycin and 34 mg/L chloramphenicol. After the culture reached an OD₆₀₀ of 0.6–0.8, IPTG was added to a final concentration of 0.5 mM. GST-yCTD was overexpressed for 16 h at 18°C and GST for 3 h at 18°C. Cells were harvested by centrifugation, resuspended in lysis buffer LB150 (20 mM HEPES, pH 7.4, 150 mM NaCl, 30 mM imidazole, 10% glycerol, 1 mM DTT, 0.284 μg/mL leupeptin, 1.37 μg/mL pepstatin A, 0.17 mg/mL PMSE, 0.33 mg/mL benzamidine), flash-frozen in liquid nitrogen and stored at –80°C.

Sequences encoding the full-length subunits of the human TFIIF kinase module (CDK7, MAT1, and cyclin H) were separately transferred into MacroBac 438B vectors⁴⁶ and combined into a single construct by ligation-independent cloning. All subunits contained an N-terminal 6xHis-tag followed by a TEV protease cleavage site. Insect cell expression was performed as described⁴⁷.

Protein purification. All purification steps were performed at 4°C. Frozen *E. coli* cell suspension was thawed, lysed by sonication, cleared from insoluble material by centrifugation (27,000 g, 45 min, 4°C), and filtered through 0.8-μm syringe filters.

For the purification of MBP-tagged proteins, cleared *E. coli* lysate was loaded onto a 5-mL HisTrap HP column (GE Healthcare) equilibrated in LB300. The HisTrap column was washed extensively using high-salt buffer HSB1000 (20 mM HEPES, pH 7.4, 1 M NaCl, 30 mM imidazole, 10% glycerol, 1 mM DTT, 0.284 μg/mL leupeptin, 1.37 μg/mL pepstatin A, 0.17 mg/mL PMSE, 0.33 mg/mL benzamidine) and equilibrated again in LB300. The column was then attached inline to a LB300-equilibrated XK-16 column (GE Healthcare), which was packed with amylose resin (New England Biolabs). Bound proteins were eluted directly onto the amylose column using nickel elution buffer 300 (20 mM HEPES, pH 7.4, 300 mM NaCl, 500 mM imidazole, 10% glycerol, 1 mM DTT, 0.284 μg/mL leupeptin, 1.37 μg/mL pepstatin A, 0.17 mg/mL PMSE, 0.33 mg/mL benzamidine). The HisTrap column was subsequently removed and the amylose column was washed again extensively with HSB1000 buffer. MBP-tagged proteins were eluted using amylose elution buffer (20 mM HEPES, pH 7.4, 300 mM NaCl, 10% glycerol, 1 mM DTT, 117 mM maltose, 0.284 μg/mL leupeptin, 1.37 μg/mL pepstatin A, 0.17 mg/mL PMSE, 0.33 mg/mL benzamidine) and concentrated with a 30-kDa MWCO Amicon Ultra filter unit (Merck). The concentrate was then subjected to size-exclusion chromatography using a Superdex 200 10/300 Increase column (GE Healthcare) pre-equilibrated in SE300 buffer (20 mM HEPES, pH 7.4, 300 mM NaCl, 10% glycerol, 1 mM TCEP). Pure fractions, as assessed by SDS-PAGE and Coomassie staining, were pooled and concentrated using a 30-kDa MWCO Amicon Ultra centrifugal filter. The protein concentration was calculated based on the absorbance at 280 nm and the predicted molar extinction coefficient (DNASTar Lasergene Suite). Aliquots were frozen in liquid nitrogen and stored at –80°C.

6xHis-GST-TEV-yCTD was purified following a similar scheme as described previously⁴⁸, with the following modifications. The clarified extract was applied to a 5-mL HisTrap HP column equilibrated in lysis buffer LB150. The column was extensively washed using high-salt buffer HSB800 (20 mM HEPES, pH 7.4, 800 mM NaCl, 30 mM imidazole, 10% glycerol, 1 mM DTT, 0.284 μg/mL leupeptin, 1.37 μg/mL pepstatin A, 0.17 mg/mL PMSE, 0.33 mg/mL benzamidine) and equilibrated again in LB150. A pre-equilibrated 5-mL HiTrap Q HP column (GE Healthcare) was attached inline to the HisTrap column, which was subsequently eluted using a linear gradient from 0–100% nickel elution buffer 150 (20 mM HEPES, pH 7.4, 150 mM NaCl, 500 mM imidazole, 10% glycerol, 1 mM DTT, 0.284 μg/mL leupeptin, 1.37 μg/mL pepstatin A, 0.17 mg/mL PMSE, 0.33 mg/mL benzamidine). The flow-through fractions were analyzed by SDS-PAGE and Coomassie staining,

pooled, and concentrated using a 30-kDa MWCO Amicon Ultra centrifugal filter unit. The sodium chloride concentration was adjusted to 50 mM and the protein was applied to a 1-mL HiTrap S column (GE Healthcare). The flow-through was concentrated using a 30-kDa MWCO Amicon Ultra concentrator and then separated on an equilibrated Superdex 200 10/300 Increase column with buffer SE300. Individual fractions were analyzed by SDS-PAGE and Coomassie staining, and pure fractions were pooled and concentrated with a 30-kDa MWCO Amicon Ultra filter unit. *E. coli* extract from the 6xHis-GST-TEV expression was applied to a 5-mL HisTrap HP column, washed with HSB800, and eluted with nickel elution buffer 150. The protein was concentrated using a 10-kDa MWCO Amicon filter unit and directly subjected to size-exclusion chromatography as described above. Concentrated protein solutions were aliquoted, flash-frozen in liquid nitrogen, and stored at –80°C.

The recombinant *S. cerevisiae* TFIIF kinase module consisting of the subunits Kin28, Ccl1, and Tfb3, was prepared as described⁴⁹. For purification of the three-subunit human TFIIF kinase module (CDK7, cyclin H, and Mat1), insect cells were lysed by sonication in lysis buffer (20 mM K-HEPES, pH 7.0, 400 mM KCl, 10% glycerol, 1 mM MgCl₂, 10 μM ZnCl₂, 5 mM β-mercaptoethanol, 30 mM imidazole, pH 8, 0.284 μg/mL leupeptin, 1.37 μg/mL pepstatin A, 0.17 mg/mL PMSE, 0.33 mg/mL benzamidine). Clarified cell lysate was applied onto a HisTrap HP 5-mL column (GE Healthcare), washed with 20 CV of lysis buffer, and eluted with a linear gradient of 0–100% of elution buffer (20 mM K-HEPES, pH 7, 400 mM KCl, 10% glycerol, 1 mM MgCl₂, 10 μM ZnCl₂, 5 mM β-mercaptoethanol, 500 mM imidazole, pH 8, 0.284 μg/mL leupeptin, 1.37 μg/mL pepstatin A, 0.17 mg/mL PMSE, 0.33 mg/mL benzamidine) in 10 CV. Peak fractions were combined, supplemented with 2 mg of 6xHis-tagged TEV protease, and dialyzed overnight against 2 L dialysis buffer (20 mM K-HEPES, pH 7, 400 mM KCl, 10% glycerol, 1 mM MgCl₂, 10 μM ZnCl₂, 5 mM β-mercaptoethanol). The dialyzed solution was applied to a HisTrap HP 5-mL column pre-equilibrated in dialysis buffer. The trimeric complex was eluted with 10% elution buffer and concentrated using an Amicon Ultra 15-mL, 30-kDa MWCO centrifugal concentrator. The sample was applied to a Superdex 200 10/300 GL size exclusion column (GE Healthcare) pre-equilibrated in storage buffer (20 mM K-HEPES, pH 7, 350 mM KCl, 10% glycerol, 1 mM MgCl₂, 10 μM ZnCl₂, 5 mM β-mercaptoethanol). Peak fractions containing stoichiometric kinase trimer were pooled, concentrated using an Amicon Ultra 15-mL, 30-kDa MWCO centrifugal concentrator to 130 μM, aliquoted, flash-frozen in liquid nitrogen, and stored at –80°C. The identity of all purified proteins was confirmed by LC-MS/MS analysis.

Pol II preparation and fluorescent labeling. Pol II was prepared from the *S. cerevisiae* strain BJ5464 as described⁵⁰ and treated with lambda phosphatase during purification. The Pol II subunit RPB3 contains an N-terminal biotin acceptor peptide, which can be biotinylated in vitro by the bacterial biotin-protein ligase BirA and used for site-specific labeling with fluorescent streptavidin conjugates. For this, 200 μg Pol II were incubated with 6 μg BirA, 100 μM D(+)-biotin and 2 mM ATP for 2 h at 20°C in Pol II buffer (10 mM HEPES, pH 7.2, 200 mM KCl, 5% glycerol, 2 mM DTT). Excess biotin was removed using a Micro Bio-Spin 6 column (Biorad) according to the manufacturer's suggestions. A small fraction of biotinylated Pol II was bound to streptavidin-coupled Dynabeads M-280 (Thermo Fisher Scientific) to confirm quantitative biotinylation. The remaining biotinylated Pol II was reacted with Alexa Fluor 594-coupled streptavidin (Thermo Fisher Scientific, ~2x molar excess) for 20 min at 20°C. Pol II was then separated from unbound streptavidin by size-exclusion chromatography using a Superose 6 10/300 column (GE Healthcare) equilibrated in Pol II buffer. Pol II-containing fractions were pooled and concentrated (100-kDa MWCO Amicon Ultra spin filter unit), and flash-frozen aliquots were stored in the dark at –80°C.

CTD phosphorylation. GST-yCTD was phosphorylated using the recombinant *S. cerevisiae* TFIIF kinase module. For this, 50 μM GST-yCTD were incubated with 0.4 μM kinase module and 3 mM ATP for 1 h at 30°C in kinase reaction buffer (20 mM HEPES, pH 7.4, 200 mM NaCl, 5 mM MgCl₂, 10% glycerol, 1 mM TCEP). Upon completion, the phosphorylation reaction was quenched by addition of EDTA to a final concentration of 10 mM. Phosphorylation of MBP-hCTD was performed using the recombinant human TFIIF kinase module. For this, MBP-hCTD (100 μM) was incubated with 2 μM kinase module in reaction buffer (20 mM HEPES, pH 7.4, 260 mM NaCl, 20 mM MgCl₂, 20 μM ZnCl₂, 10% glycerol, 2 mM TCEP). The reaction was started by addition of 8 mM ATP, incubated for 1 h at 30°C, and quenched by addition of 40 mM EDTA. Control reactions lacking either the kinase or ATP were conducted in both cases under identical conditions. After completion of GST-yCTD and MBP-hCTD phosphorylation experiments, all reactions were mixed with 20% dextran (in buffer containing 20 mM HEPES, pH 7.4, 200 mM NaCl) at a ratio of 1:4 (vol/vol) and then analyzed microscopically (as described below). To study phosphorylation-induced dissolution of preformed CTD droplets, MBP-hCTD was mixed at a final concentration of 20 μM into 16% dextran containing 20 mM HEPES, pH 7.4, 220 mM NaCl, 1.6 mM ATP, 4 mM MgCl₂, 20 μM ZnCl₂, and 1 mM TCEP to induce phase separation. Immediately before imaging, the reaction was started by addition of human TFIIF kinase module to a final concentration of 0.4 μM and immediately analyzed by microscopy.

Kinase activity assay. Kinase activity was analyzed by mobility shift assays. One microgram of CTD fusion protein from kinase and control reactions was separated on 4–15% Tris-glycine Protean TGX polyacrylamide gels (Biorad) and stained with Coomassie solution (InstantBlue, Expedeon). Phosphorylation of the CTD substrates by human and yeast TFIIF kinase modules results in a pronounced decrease of electrophoretic mobility. Phosphorylation of the CTD residue Ser5 was confirmed by immunoblotting. For this, samples (100 ng/lane) were separated on 4–15% Tris-glycine Protean TGX gels and blotted onto a PVDF membrane with a Trans-Blot Turbo Transfer System (Bio-Rad). The membrane was blocked for 1–2 h at room temperature (20–24 °C) with 5% (w/v) milk powder in phosphate-buffered saline containing 0.1% Tween-20 (PBST). The blocked membrane was then incubated with either anti-MBP HRP conjugate (ab49923; Abcam) or anti-GST HRP conjugate (RPN1236; GE Healthcare) for 2 h at room temperature. SuperSignal West Pico Chemiluminescent Substrate (Thermo Fisher) was used to develop the membrane before scanning with a ChemoCam Advanced Fluorescence imaging system (Intas Science Imaging). For immunoblot analysis of CTD phosphorylation, the membrane was subsequently stripped by incubation in stripping buffer (200 mM glycine-HCl, pH 2.2, 0.1% SDS, 1% Tween-20), blocked with 5% (w/v) milk powder in PBST, and probed overnight at 4 °C with primary CTD antibody against phosphorylated Ser5 (3E8; diluted 1:60 in 2.5% (w/v) milk powder in PBST). The anti-Ser5 CTD antibody was a kind gift of D. Eick (Molecular Epigenetics Research Unit, Helmholtz Center, Munich). The membrane was then incubated with an anti-rat HRP-conjugate (A9037, Sigma-Aldrich) in 2.5% milk-PBST for 1 h at room temperature and developed as describe above.

Disorder prediction. Recent cryo-EM analysis of mammalian RNA polymerase II could derive an atomic model only to RPB1 position P1487³¹, indicating a high conformational flexibility of the following RPB1-linker and the C-terminal repeat domain. We thus used the VLXT predictor implemented in PONDR³² to calculate the disorder propensity for the human RPB1 residues 1,488–1,970.

CD spectroscopy. Far-UV CD measurements were performed on a Chirascan spectrometer (Applied Photophysics, Ltd) at 25 °C using a 0.2-mm path length cuvette. The concentration of MBP-hCTD and MBP-yCTD was 5 μM in 20 mM NaPi, pH 7.4. CD spectra were recorded from 180 to 280 nm with an integration time of 0.5 s, and experiments were repeated three times. The spectra of human and yeast CTD were obtained through subtraction of the spectrum of MBP and correction of the baseline using buffer. Data are expressed in terms of the mean residual ellipticity (θ) in deg/(cm² dmol).

NMR spectroscopy. Peptides comprising one (1R-CTD; seven residues), two (2R-CTD; 14 residues), and three (3R-CTD; 21 residues) YSPTSPS-repeats were synthesized by GenScript with acetyl- and amide-protection groups at the N and C termini, respectively. NMR spectra were recorded at 5 °C on Bruker 600- and 700-MHz spectrometers with triple-resonance cryogenic probes for 1.0 mM 3R-CTD peptide (20 mM HEPES, pH 7.4, 200 mM NaCl, and 90% H₂O/10% D₂O) and 0.5 mM phosphorylated 3R-CTD peptide. For phosphorylation, 3R-CTD was incubated with 2 μM TFIIF kinase at 37 °C for 18 h (20 mM HEPES, pH 7.4, 200 mM NaCl, 3 mM MgCl₂, 3 mM ATP, and 90% H₂O/10% D₂O). Spectra were processed with TopSpin software (Bruker) and analyzed using CCPN Analysis³³. Sequence-specific backbone and sidechain resonance assignments of nonphosphorylated and phosphorylated 3R-CTD peptide were achieved through ¹H-¹⁵N heteronuclear single quantum coherence (HSQC) and ¹H-¹³C HSQC experiments at natural abundance, together with 2D ¹H-¹H TOCSY (100-ms mixing time) and 2D ¹H-¹H NOESY (200- and 300-ms mixing times) experiments. Resonance assignments of 3R-CTD were further validated through comparison with NMR spectra recorded for 1R-CTD and 2R-CTD peptides.

Differential interference contrast (DIC) and fluorescence microscopy. Droplet formation of protein samples was monitored by DIC and fluorescence microscopy. Samples were fluorescently labeled using Alexa Fluor 488 Microscale Protein Labeling Kit (Thermo Fisher Scientific, #A30006) according to the manufacturer's instructions. Small amounts (<0.5 μM) of labeled protein, which are not sufficient to induce droplet formation by itself, were mixed with unlabeled protein to the final concentration indicated in the text. In experiments with Ficoll PM 400 (Sigma, #F4375) at a final concentration of 150 mg/mL (buffer containing 20 mM HEPES, pH 7.4, 200 mM NaCl) was used. In experiments using dextran T500 (Pharmacosmos) as a crowding agent, dextran was added to reach the indicated final concentrations in 20 mM HEPES, pH 7.4, 220 mM NaCl. In all experiments, reducing conditions were maintained during droplet formation through the presence of TCEP generally at a final concentration of 0.2 mM. Five to 10 μL of samples were loaded onto glass slides, covered with ø18 mm coverslips, and sealed. DIC and fluorescent images were acquired on a Leica DM6000B microscope with a 63× objective (water immersion) and processed using Fiji software (NIH). In experiments requiring MBP-tag removal, fusion proteins were incubated with TEV protease in molar ratio TEV:protein = 1:25 for 3 h at 25 °C. Complete tag removal was confirmed by SDS-PAGE analysis and Coomassie staining of the digested samples.

In experiments with aliphatic alcohols, the MBP-tag was cleaved off from MBP-yCTD and MBP-hCTD as indicated above, followed by addition of the

protein to a premix containing dextran (final concentration 16%) and either 1,6-hexanediol (Sigma, #240117) or 2,5-hexanediol (Sigma, #H11904). The final protein concentration in the sample was 50 μM for yCTD and 20 μM for hCTD, and hexanediol concentrations varied from 2.5 to 10%. Samples were imaged by DIC microscopy as indicated above.

All experiments with droplet formation were performed at room temperature except when the influence of temperature was tested. In the latter case, MBP-hCTD or MBP-yCTD was mixed with small amounts (<0.2 μM) of the corresponding Alexa Fluor 488-labeled protein, from which the MBP-tag was cleaved off using TEV protease as described above. Final protein concentrations in the samples were 20 μM for MBP-hCTD and 40 μM for MBP-yCTD in 20 mM HEPES, pH 7.4, 220 mM NaCl, 0.2 mM TCEP with 16% dextran. Samples were then incubated for 1 h on ice (4 °C), at room temperature (22 °C), or in an incubator at 37 °C or 45 °C before microscopy analysis. Labeled (without MBP tag) and unlabeled (MBP-tagged) proteins were also mixed in experiments testing the influence of ionic strength. Final protein concentrations were 10 μM for MBP-hCTD and 40 μM for MBP-yCTD, and samples contained indicated NaCl concentrations in 20 mM HEPES, pH 7.4, 0.2 mM TCEP and 16% dextran.

Co-recruitment experiments. For CTD co-recruitment experiments, droplets were made with 20 μM MBP-hCTD or GST-yCTD in 20 mM HEPES, pH 7.4, 220 mM NaCl, 0.2 mM TCEP containing 16% dextran. Droplets were visualized through addition of 0.6 μM of tetramethylrhodamine (TMR)-labeled peptide with the sequence YSPTSPS, i.e., corresponding to one consensus heptad repeat. Subsequently, small amounts (<0.5 μM) of Alexa Fluor 488-labeled GST-yCTD or MBP-hCTD were added to preformed MBP-hCTD or GST-yCTD droplets, respectively. Co-recruitment was assessed by imaging on a Leica DM6000B microscope as described above, using DIC in combination with red and green channels for fluorescence (GFP and N3 filter cubes).

For Pol II co-recruitment experiments, Alexa Fluor 594-labeled Pol II (final concentration 0.02 μM) was mixed with preformed GST-yCTD droplets (final concentration 25 μM) that were visualized by addition of Fluor Alexa 488-labeled GST-yCTD (final concentration 2.3 μM) in 20 mM HEPES, pH 7.4, 220 mM NaCl, 0.2 mM TCEP containing 16% dextran. Co-recruitment was documented by DIC and fluorescent microscopy using red and green channels (GFP and N3 filter cubes) on a Leica DM6000B microscope as described.

In vitro FRAP experiments. The dynamics of human and yeast CTD molecules in the phase-separated state were investigated by fluorescence recovery after photobleaching (FRAP). MBP-tagged human and yeast CTD proteins were labeled on a single Cys residue that is present C-terminal to the TEV protease cleavage site (see above) using Alexa Fluor 488 C5 maleimide dye (Thermo Fisher Scientific, #A10254) according to the manufacturer's recommendations. Briefly, proteins were incubated in a light-protected Eppendorf tube with the dye freshly dissolved in DMSO in a molar ratio of 1:15 = protein:dye in 20 mM HEPES, pH 7.4, 300 mM NaCl, 1 mM TCEP, 10% glycerol for 3 h at room temperature. Excess label and salt were removed by desalting samples twice with 0.5-mL 7000 MWKO Zeba spin desalting columns (Thermo Fisher Scientific, #89882). The MBP-tag was then cleaved from labeled and unlabeled human and yeast CTD using TEV protease as indicated above. Droplets for FRAP measurements were made in 16% dextran T500 in 20 mM HEPES, pH 7.4, 220 mM NaCl, 0.2 mM TCEP by adding mixtures of labeled and unlabeled yCTD (or hCTD) in a molar ratio of 1:100 to the final CTD concentration of 20 μM. To minimize droplet movement, FRAP recordings were done after approximately 30 min, which is the time required for freshly formed droplets to settle down on the glass slide and become less mobile.

FRAP experiments were recorded on a Leica TCS SP8 confocal microscope using a 63× objective (water immersion) at a zoom corresponding to a pixel size of 96 nm × 96 nm and using the 488-nm argon laser line. A circular region of ~1.4 μm in diameter was chosen in a region of homogenous fluorescence away from the droplet boundary and bleached with 10 iterations of full laser power. Recovery was imaged at low laser intensity (0.057%). Fifty frames were recorded, with 1 frame per 330 ms. Pictures were analyzed in Fiji (NIH), and FRAP recovery curves were calculated using standard methods. For calculating half time recoveries, normalized values from each recording were separately fit to a single exponential model, and half time recoveries were presented as mean ± standard error.

Cell line establishment and characterization. Human U2OS osteosarcoma cells (Research Resource Identifier (RRID): CVCL_0042) were used in this study. The parental U2OS cell line was authenticated by the UC Berkeley cell culture facility on 05/05/2017 by STR analysis. The result was a 100% match with the U2OS cell line reference. The parental U2OS cell line was tested for mycoplasma contamination before establishing the RPB1 cell lines and thereafter regularly (approximately every 6 months) checked to confirm mycoplasma-negativity. Cells were grown in a Sanyo copper-alloy IncuSafe humidified incubator (MCO-18AIC(UV)) at 37 °C/5.5% CO₂ in low-glucose DMEM with 10% FBS (full recipe: 500 mL DMEM (ThermoFisher #10567014), 50 mL FBS (HyClone FBS SH30910.03 lot #AXJ47554), and 5 mL penicillin-streptomycin (ThermoFisher #15140122)) and were passaged every 2–4 d before reaching confluency. Plasmids expressing N-terminally tagged (either Dendra2 or Halo) α-amanitin-resistant mutated

(N792D) human RPB1 were stably transfected into U2OS cells using Fugene 6 following the manufacturer's instruction (Promega #E2692). The RPB1-52R vectors encode the 52 CTD repeats originally present in the endogenous RPB1 cDNA. The RPB1-25R expressing vectors contain only 25 repeats of the original 52, corresponding to repeats 1 to 21 and repeats 49 to 52. The RPB1-70R cell lines express either a Dendra2-RPB1 protein containing 66 repeats in its CTD (repeats 1 to 51, then repeats 38 to 52) or a Halo-RPB1 protein containing 70 repeats in its CTD (repeats 1 to 47, then repeats 42 to 47, then repeats 38 to 52), as assessed by sequencing of the RPB1 mRNA expressed in these cells. Details of cloning strategies are available upon request. α -Amanitin (SIGMA #A2263) was used during the stable selection process at a concentration of 2 μ g/mL and was used thereafter in permanence in the culture of the cells at a concentration of 1 μ g/mL to avoid endogenous RPB1 re-expression as described in ref.¹. Even though these lines cannot genotypically be considered as endogenously tagged (the endogenous wild-type RPB1 gene is still present; a cDNA expressing the tagged version of RPB1 is incorporated in the genome), phenotypically they can, as the expression of endogenous RPB1 protein is replaced by the tagged version of the protein at all time. RT-PCR analysis (Superscript III with oligo (dT)₂₀, Invitrogen (#18080093), and NEB Phusion High-Fidelity DNA Polymerase (#M0530S) followed by sequencing was performed to confirm the sequence of the RPB1-CTD expressed in the various cell lines (more details of all the molecular biology characterizations are available upon request).

Western blot. Cells were collected after ice-cold PBS wash by scraping into 0.5 mL/10 cm plate of high-salt lysis buffer (0.5 M NaCl, 50 mM HEPES, 5 mM EDTA, 0.5% NP-40, and protease inhibitors), with 125 U/mL of benzonase (Novagen, EMD Millipore), passed through a 25-gauge needle, rocked at 4°C for 30 min, and centrifuged at maximum speed at 4°C for 20 min. Supernatants were quantified by the Bradford method. The same amount of proteins was loaded onto 7% Bis-Tris SDS-PAGE gel, transferred onto nitrocellulose membrane (Amersham Protran 0.45- μ m NC, GE Healthcare) for 2 h at 80 V, blocked in TBS-Tween with 5% milk for at least 1 h at room temperature, and blotted overnight at 4°C with primary antibodies (anti-Pol II (N20) from SantaCruz, #sc-899; anti-Lamin A from Abcam, #ab26300) in TBS-T with 5% milk. HRP-conjugated secondary antibodies were diluted 1:5,000 in TBS-T with 5% milk and incubated at room temperature for 1 h.

FACS analysis. Expression of the exogenous RPB1 protein was assessed by flow cytometry analysis on live cells on a BD LSR Fortessa, performed according to the manufacturer's protocols. For the Halo-tagged line, Halo-TMR labeling (500 nM) was performed for 30 min at 37°C before harvesting the cells.

xCELLigence analysis. The Cell Index (a representation of cell growth and viability) was measured in real time using the RTCA-SP (Acea Biosciences) according to manufacturer's instructions. Cells were seeded at a density of 2,000 cells/well (Dendra2-tagged cell lines) or 4,000 cells/well (Halo-tagged cell lines), respectively. The Cell Index was normalized at 3 h after seeding to account for slight variations in the number of counted cells between various lines.

Doubling time analysis. Doubling time analysis was performed (using FarRed CFSE from a CellTrace CFSE Cell Proliferation Kit, ThermoFisher Scientific #C34554) to compare the growth capacities of the different lines. More precisely, for doubling time analysis, data was collected on a BD Bioscience LSR Fortessa; the geometric fluorescent mean intensity of each sample for each timepoint (day 1 to day 5) was extracted from FlowJo, and the average change over the 5-d period was calculated. The average change was then converted to log scale to calculate the doubling time.

Cell imaging conditions. For live-cell imaging, the medium was identical, except DMEM without phenol red was used (ThermoFisher #31053028). U2OS cells expressing α -amanitin-resistant Halo-RPB1-25R, Halo-RPB1-52R, or Halo-RPB1-70R were grown overnight with α -amanitin on plasma-cleaned 25-mm circular no. 1.5 H cover glasses (Marienfeld High-Precision 0117650). For the flavopiridol experiments, cells were treated for 30–45 min with flavopiridol (2 μ M final concentration), and then imaged for a maximum of 30–45 min. Prior to all experiments, the cover glasses were plasma-cleaned and then stored in isopropanol until use. For live-cell FRAP experiments, cell preparation was identical except cells were grown on glass-bottomed (thickness #1.5) 35-mm dishes (MatTek P35G-1.5-14-C).

PALM imaging. Six videos of ~50,000 frames were acquired for each condition at 30 ms/frame. The axial drift was corrected in real time with a perfect-focus system. A cylindrical lens was added to the system to induce astigmatism in the point-spread function (PSF) of the optical setup. 300,000 detections were collected on average per video. Single-molecule detection and localization was performed with a modified version of the multiple-target tracking algorithm. The 3D position of single detections was inferred from the lateral elongation of the PSF. The lateral drift of the sample was corrected by using fluorescent beads (TetraSpeck microspheres). To correct for blinking of the Dendra2 fluorophore, detections in a disk of 30 nm diameter and adjacent in time were grouped and averaged.

Nuclei and nucleoli were automatically detected and segmented for further processing. $N(r)$ is the estimate of the expected number of neighbors within a distance r of a given point of the sample

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

where P is the set of all detections, and N_p is the total number of detections. The f function^{54,55} corrects for biases generated by points located at short distances to the borders (nucleus or nucleoli)

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

where $d(i, j)$ is the distance between i and j , and C_{in} is the length of that part of a circle of radius $d(i, j)$ centered on i which is inside the area of study, the nucleus.

The null hypothesis, complete spatial randomness (CSR), is a homogeneous Poisson process with intensity λ , equal to the density of detection in the area of study A : $\lambda = \frac{N_p}{A}$.

We estimated four spatial statistics based on $N(r)$: $n(r)$, $K(r)$, $L(r)$ and $G(r)$ ^{55,56}. The local neighbor density function, is defined as $n(r) = N(r)/\pi r^2$. The K-Ripley function is defined as $K(r) = N(r)/\lambda$. The linearized K-Ripley function is given by $L(r) = \sqrt{K(r)/\pi} - r$. The pair density function $G(r)$ is simply the derivative of $K(r)$.

Under CSR, the expected value taken by $n(r)$ (resp. $K(r)$, $L(r)$, and $G(r)$) is λ (resp. πr^2 , 0, and 1). Triangulation of the areas was performed with a custom python script and we used the ADS R package⁵⁷ to estimate the four spatial statistics. To estimate the s.d. and standard error associated with these measurements, we performed a bootstrapping analysis of the dataset. We randomly selected 10,000 detections from each original dataset 100 times and fed these subsampled data to the R script computing the spatial statistics.

Single-molecule imaging (spaSPT). After overnight growth, cells were labeled with 50 nM PA-JF₅₄₉⁵⁸ for ~15–30 min and washed twice (first wash: medium removed; second wash: PBS). At the end of the final wash, the medium was replenished and changed to phenol red-free medium, keeping all other aspects of the medium the same (and adding back α -amanitin). Single-molecule imaging was performed on a custom-built Nikon TI microscope equipped with a 100 \times /NA 1.49 oil-immersion TIRF objective (Nikon Apochromat CFI Apo TIRF 100 \times Oil), EM-CCD camera (Andor iXon Ultra 897; frame-transfer mode; vertical shift speed: 0.9 μ s; -70°C), a perfect focusing system to correct for axial drift and motorized laser illumination (Ti-TIRF; Nikon) and which allows incident angle adjustment to achieve highly inclined and laminated optical sheet illumination⁵⁹. An incubation chamber maintained a humidified 37°C atmosphere with 5% CO₂, and the objective was also heated to 37°C. Excitation was achieved using a 561-nm (1 W, Genesis Coherent) laser for PA-JF₅₄₉. The excitation laser was modulated by an acousto-optic tunable filter (AA Opto-Electronic, AOTFnc-VIS-TN) and triggered by the camera TTL exposure output signal. The laser light was coupled into the microscope by an optical fiber, reflected using a multiband dichroic (405 nm/488 nm/561 nm/633 nm quad-band, Semrock), and focused in the back focal plane of the objective. Fluorescence emission light was filtered using a single-bandpass filter placed in front of the camera using a Semrock 593/40-nm bandpass filter. The microscope, cameras, and hardware were controlled through NIS-Elements software (Nikon).

We recorded single-molecule tracking movies using our previously developed technique, stroboscopic photoactivation single-particle tracking (spaSPT)^{34,36}. Briefly, 1-ms, 561-nm excitation (100% AOTF) of PA-JF₅₄₉ was delivered at the beginning of the frame to minimize motion blurring; 405-nm photoactivation pulses were delivered during the camera integration time (~447 μ s) to minimize background, and their intensity optimized to achieve a mean density of ~1 molecule per frame per nucleus. We recorded 30,000 frames per cell per experiment. The camera exposure time was 7 ms, resulting in a frame rate of approximately 134 Hz (7 ms + ~447 μ s per frame).

spaSPT data was analyzed (localization and tracking) and converted into trajectories using a custom-written Matlab implementation of the MTT algorithm⁶⁰ and the following settings: localization error: $10^{-6.25}$; deflation loops: 0; blinking (frames): 1; max competitors: 3; max D ($\mu\text{m}^2/\text{s}$): 20.

We recorded ~5–10 cells per replicate and performed three independent replicates on three different days. Specifically, across three replicates we imaged 29 cells for 25 R Halo-RPB1 and obtained 448,362 trajectories with 690,682 unique displacements at a mean density of 1.2 localizations per frame. Similarly, we imaged 30 cells for 52 R Halo-RPB1 and obtained 324,928 trajectories with 619,247 unique displacements at a mean density of 1.1 localizations per frame. Finally, we imaged 26 cells for 70 R Halo-RPB1 and obtained 333,720 trajectories with 571,345 unique displacements at a mean density of 1.0 localization per frame. In the flavopiridol treated experiment, we imaged 13 cells for 25 R Halo-RPB1 and obtained 598,947 trajectories with 926,057 unique displacements at a mean density of 2.4 localizations per frame. We imaged 15 cells for 52 R Halo-RPB1 and

obtained 395,206 trajectories with 671,492 unique displacements at a mean density of 1.5 localizations per frame. Finally, we imaged 28 cells for 70 R Halo-RPB1 and obtained 616,088 trajectories with 1,030,523 unique displacements at a mean density of 1.9 localizations per frame.

FRAP in cells. FRAP experiments were performed and analyzed as previously described³⁶. Briefly, FRAP was performed on an inverted Zeiss LSM 710 AxioObserver confocal microscope equipped with a motorized stage, a full incubation chamber maintaining 37 °C/5% CO₂, a heated stage, an X-Cite 120 illumination source as well as several laser lines. Halo-TMR was excited using a 561-nm laser. Images were acquired on a 40× Plan NeoFluar NA1.3 oil-immersion objective at a zoom corresponding to a 100 × 100-nm pixel size, and the microscope was controlled using the Zeiss Zen software. In FRAP experiments, 300 frames were acquired at 1 frame per s, allowing 20 frames to be acquired before the bleach pulse to accurately estimate baseline fluorescence. A circular bleach spot ($r = 10$ pixels) was chosen in a region of homogenous fluorescence at a position at least 1 μm from nuclear and nucleolar boundaries. The spot was bleached using maximal 561-nm laser intensity and pixel dwell time corresponding to a total bleach time of ~1 s. We generally collected data from 3–5 cells per cell line per condition per day and all presented data is from at least three independent replicates on different days.

To quantify and drift-correct the FRAP movies, we used a previously described custom-written pipeline in Matlab³⁶. Briefly, we manually identify the bleach spot. The nucleus was automatically identified by thresholding images after Gaussian smoothing and hole-filling (to avoid the bleach spot being misidentified as not belonging to the nucleus). We use an exponentially decaying threshold (from 100% to ~85% (measured) of initial over one video) to account for whole-nucleus photobleaching during the time-lapse acquisition. Next, we quantified the bleach spot signal as the mean intensity of a slightly smaller circle ($r = 0.6 \mu\text{m}$), which is more robust to lateral drift. The FRAP signal was corrected for photobleaching using the measured reduction in total nuclear fluorescence (~15% over 300 frames at the low laser intensity used after bleaching) and internally normalized to its mean value during the 20 frames before bleaching. We corrected for drift by manually updating a drift vector quantifying cell movement during the experiment. Finally, drift- and photobleaching corrected FRAP curves from each single cell were averaged to generate a mean FRAP recovery. We used the mean FRAP recovery in all figures; error bars show s.e.m.

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability. The in vitro datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request. PALM and SPT data are available at <https://doi.org/10.5281/zenodo.1188488>.

Raw spaSPT data is available in Spot-On readable CSV format in the form of single-molecule trajectories. The Spot-On Matlab code is available together with a step-by-step guide at GitHub: <https://gitlab.com/tjian-darzacq-lab/spot-on-matlab>. For additional documentation, please see also the Spot-On website at <https://SpotOn.berkeley.edu> and previous publications^{34,36}. All code and data generated in this study are available from the corresponding author on reasonable request.

References

46. Gradia, S. D. et al. MacroBac: new technologies for robust and efficient large-scale production of recombinant multiprotein complexes. *Methods Enzymol.* **592**, 1–26 (2017).
47. Farnung, L., Vos, S. M., Wigge, C. & Cramer, P. Nucleosome-Chd1 structure and implications for chromatin remodelling. *Nature* **550**, 539–542 (2017).
48. Battaglia, S. et al. RNA-dependent chromatin association of transcription elongation factors and Pol II CTD kinases. *eLife* **6**, e25637 (2017).
49. Schilbach, S. et al. Structures of transcription pre-initiation complex with TFIID and Mediator. *Nature* **551**, 204–209 (2017).
50. Sydow, J. F. et al. Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Mol. Cell* **34**, 710–721 (2009).
51. Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J. M. & Cramer, P. Structure of transcribing mammalian RNA polymerase II. *Nature* **529**, 551–554 (2016).
52. Obradovic, Z. et al. Predicting intrinsic disorder from amino acid sequence. *Proteins* **53**(Suppl 6), 566–572 (2003).
53. Vranken, W. F. et al. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696 (2005).
54. Goreaud, F. & Pelissier, R. On explicit formulas of edge effect correction for Ripley's K-function. *J. Veg. Sci.* **10**, 433–438 (1999).
55. Ripley, B. D. Modeling spatial patterns. *J. R. Stat. Soc. Series B Stat. Methodol.* **39**, 172–212 (1977).
56. Dixon, P. M. Ripley's K function. *Encyclopedia of Environmetrics* **3**, 1796–1803 (2002).
57. Pelissier, R. & Goreaud, F. Ads package for R: a fast unbiased implementation of the K-function family for studying spatial point patterns in irregular-shaped sampling windows. *J. Stat. Softw.* **63**, 1–18 (1999).
58. Grimm, J. B. et al. Bright photoactivatable fluorophores for single-molecule imaging. *Nat. Methods* **13**, 985–988 (2016).
59. Tokunaga, M., Imamoto, N. & Sakata-Sogawa, K. Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nat. Methods* **5**, 159–161 (2008).
60. Sergé, A., Bertaux, N., Rigneault, H. & Marguet, D. Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes. *Nat. Methods* **5**, 687–694 (2008).

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☒ ☐ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☒ ☐ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☐ ☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☐ ☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☐ ☒ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated
- ☐ ☒ Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

The in vitro datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request. The PALM and SPT data are available at <http://doi.org/10.5281/zenodo.1188488>. The raw spaSPT data is available in Spot-On readable CSV format in the form of single-molecule trajectories.

Data analysis

The Spot-On Matlab code is available together with a step-by-step guide at Gitlab: <https://gitlab.com/tjian-darzacq-lab/spot-on-matlab>. For additional documentation, please see also the Spot-On website <https://SpotOn.berkeley.edu> and previous publications 28,30. All other codes are available from corresponding author on reasonable request.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The in vitro datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request. The PALM and SPT data are available at <http://doi.org/10.5281/zenodo.1188488>. The raw spaSPT data is available in Spot-On readable CSV format in the form of single-molecule trajectories. The Spot-On Matlab code is available together with a step-by-step guide at Gitlab: <https://gitlab.com/tjian-darzacq-lab/spot-on-matlab>. For additional documentation, please see also the Spot-On website <https://SpotOn.berkeley.edu> and previous publications (Ref. 34, Ref. 36). All other codes used in this study are available from corresponding author on reasonable request.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For single molecule analysis, sample size was analysed and described in Hansen et al, 2018. For FRAP and Clustering analysis statistical reproducibility was determined by datas resampling.
Data exclusions	No data were excluded.
Replication	All attempts at replication were successful.
Randomization	Randomization was not required in this study.
Blinding	Blinding was not required in this study and investigators were not blinded to group allocation.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Unique biological materials
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

anti-Ser5P CTD antibody (kind gift of Dirk Eick (Helmholtz Center Munich), monoclonal antibody, clone 3E8, supplied as hybridoma supernatant)
 anti-MBP conjugate (Abcam, ab49923, HRP-coupled, monoclonal antibody, MBP-17, GR251034-1)
 anti-GST antibody (GE healthcare, RPN1236, HRP-coupled, monoclonal antibody)
 anti-rat antibody (Sigma Aldrich, A9037, HRP-coupled, polyclonal antibody, SLBL0586V)
 anti-Pol II antibody (N20, Santa Cruz, sc-899)
 anti-Lamin A antibody (abcam, ab26300)

Validation

Anti-Ser5P CTD antibody, anti-GST-HRP antibody, and anti-MBP-HRP antibody were used to probe TFIIH kinase activity. In the

study antibodies were used to detect highly purified proteins. Validity of detection was cross-confirmed by simultaneous polyacrylamide gel electrophoresis and subsequent coomassie staining, based on the known migration behavior of the unphosphorylated protein species. Identity of size-shifted (i.e. phosphorylated) CTD-proteins was confirmed through the use of two antibodies, which bind the target in an orthogonal manner (e.g. anti-MBP and anti-Ser5P). Anti-Ser5P CTD antibody is of monoclonal origin (clone 3E8) and was thoroughly validated with synthetic peptides to require phosphorylated Ser5 for binding (Chapman et al. (2007), Science 318, 1780-1782). The antibody was further characterized in previous, peer-reviewed publications (e.g. Clemente-Blanco et al. (2011), Nat Cell Biol. 13, 1450-1456; Xu et al. (2012), Dev. Cell 23, 1059-71; Manafra et al. (2014), Plos One 9, e99603).

N20 anti-Pol II antibody was used to assess the RPB1 level expressed in the different human cell lines. Another anti-RPB1 antibody was used to compare the size of the detected bands with the ones detected by the N20 antibody. The result was consistent. Due to changes in the CTD length between the different cell lines, antibodies against the RPB1 C-terminal domain cannot be used. The N20 antibody was additionally successfully used in numerous peer-reviewed studies (e.g. Emmett et al., Nature 546, 544-548, 2017; Dieuleveult et al., Nature 530, 113-6, 2016; Forget et al., NAR 41, 6881-6891, 2015).

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	Human U2OS osteosarcoma cells (Research Resource Identifier: RRID:CVCL_0042)
Authentication	The parental U2OS cell line was authenticated by the UC Berkeley Cell culture facility on 5/5/2017 by STR analysis . The result was a 100% match with U2OS cell line.
Mycoplasma contamination	The parental U2OS cell lines were tested for mycoplasma contamination before establishing the RPB1 cell lines. Then the cells are tested every 6 months or so. We never found mycoplasma contamination for these lines.
Commonly misidentified lines (See ICLAC register)	NA

Flow Cytometry

Plots

Confirm that:

- ☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☒ All plots are contour plots with outliers or pseudocolor plots.
- ☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	For the Halo-tagged line, Halo-TMR labelling (500 nM) was performed for 30mn at 37°C before harvesting the cells. The cells were harvested by Trypsinization after wash with PBS. Then Cells were resuspended in media, span down and resuspended in PBS for analysis.
Instrument	BD LSRFortessa
Software	Diva software
Cell population abundance	10,000 cells were analysed per sample
Gating strategy	Gate between positive and negatively labelled cells were fixed running WT non labelled cells treated the same way the day of each experiment. The gating strategy was used only to compare the level of expression in between cells. Same gate used for all the conditions. The gate is represented on each graph. This is more a qualitative than quantitative analysis.

- ☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

► Data presentation

For all flow cytometry data, confirm that:

- ☒ 1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ 2. The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☒ 3. All plots are contour plots with outliers or pseudocolor plots.
- ☒ 4. A numerical value for number of cells or percentage (with statistics) is provided.

► Methodological details

- | | |
|--|---|
| 5. Describe the sample preparation. | For the Halo-tagged line, Halo-TMR labelling (500 nM) was performed for 30mn at 37°C before harvesting the cells. The cells were harvested by Tripsynisation after wash with PBS. Then Cells were resuspended in media, span down and resuspended in PBS for analysis |
| 6. Identify the instrument used for data collection. | BD LSRFortessa |
| 7. Describe the software used to collect and analyze the flow cytometry data. | Diva software |
| 8. Describe the abundance of the relevant cell populations within post-sort fractions. | 10,000 cells were analysed per sample |
| 9. Describe the gating strategy used. | gate between positive and negatively labelled cells were fixed running WT non labelled cells treated the same way the day of each experiment.
The gating strategy was used only to compare the level of expression in between cells...same gate used for all the conditions. The gate is represented on each graph. This is more a qualitative than quantitative analysis. |

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information. ☒