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 Yu**[5], **Herve Marie-Nelly**[2,3], **David T. McSwiggen**[2,3], **Goran Kokic** [1], **Gina M. Dailey**[2,3], **Patrick Cramer** [1]*, **Xavier Darzacq**[2,3]* and **Markus Zweckstetter** [4,5]*
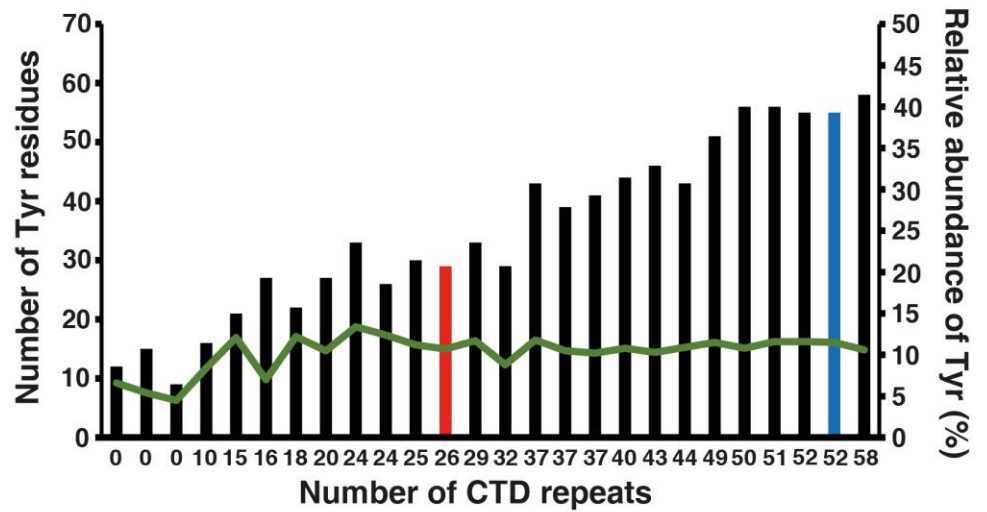
[1]Max Planck Institute for Biophysical Chemistry, Department of Molecular Biology, Göttingen, Germany. [2]Department of Molecular and Cell Biology, University of California, Berkeley, California, USA. [3]CIRM Center of Excellence, University of California, Berkeley, California, USA. [4]Max Planck Institute for Biophysical Chemistry, Department of NMR-based Structural Biology, Göttingen, Germany. [5]German Center for Neurodegenerative Diseases (DZNE), Göttingen, Germany. [6]These 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

a

	H. sapiens	S. cerevisiae
1	YSPTSPA	FSPTSPT
	YEPRSPGG	YSPTSPA
	YTPQSPS	YSPTSPS
	YSPTSPS	YSPTSPS
	YSPTSPS	YSPTSPS
	YSPTSPN	YSPTSPS
	YSPTSPS	YSPTSPS
	YSPTSPS	YSPTSPS
	YSPTSPS	YSPTSPS
10	YSPTSPS	YSPTSPS
	YSPTSPS	YSPTSPS
	YSPTSPS	YSPTSPS
	YSPTSPS	YSPTSPS
	YSPTSPS	YSPTSPS
	YSPTSPS	YSPTSPS
	YSPTSPS	YSPTSPS
	YSPTSPS	YSPTSPA
	YSPTSPS	YSPTSPS
	YSPTSPS	YSPTSPS
20	YSPTSPS	YSPTSPS
	YSPTSPS	YSPTSPS
	YSPTSPN	YSPTSPN
	YSPTSPN	YSPTSPS
	YTPTSPS	YSPTSPG
	YSPTSPS	YSPGSPA
	YSPTSPN	YSPKQDE
	YTPTSPN	QKHNENENSR
	YSPTSPS	
	YSPTSPS	
30	YSPTSPS	
	YSPSSPR	
	YTPQSPT	
	YTPSSPS	
	YSPSSPS	
	YSPTSPK	
	YTPTSPS	
	YSPSSPE	
	YTPTSPK	
	YSPTSPK	
40	YSPTSPK	
	YSPTSPT	
	YSPTTPK	
	YSPTSPT	
	YSPTSPV	
	YTPTSPK	
	YSPTSPT	
	YSPTSPK	
	YSPTSPT	
	YSPTSPKGST	
50	YSPTSPG	
	YSPTSPT	
	YSLTSPA	
	ISPDDSDEEN	

**Consensus repeat:**

$Y_1S_2P_3T_4S_5P_6S_7$

b

c
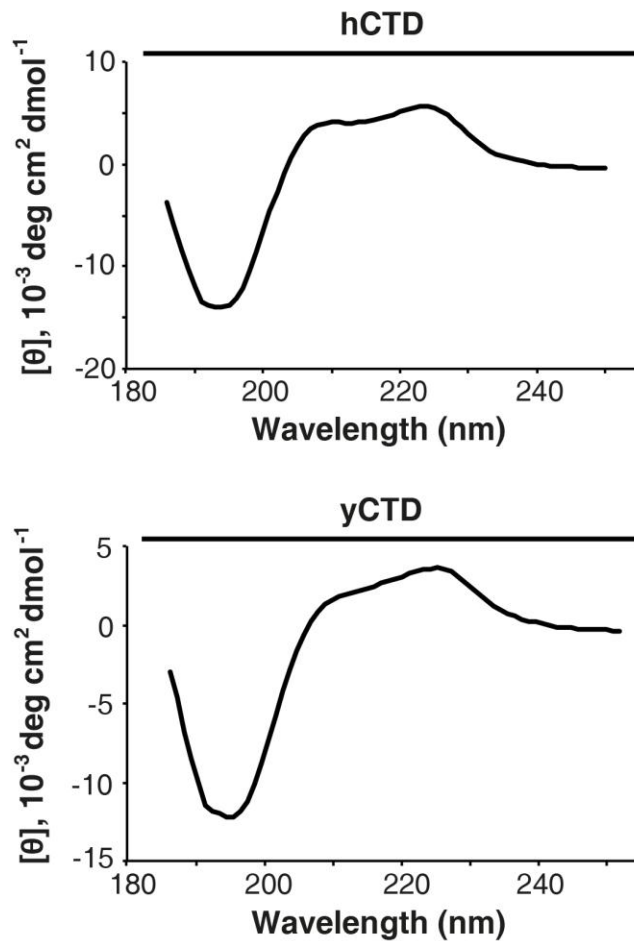
Supplementary Figure 1

Low-complexity nature of CTD sequences.

(**a**) Comparison of the CTD sequences from human (blue) and yeast (red). Consisting of 52 heptad repeats, the human CTD is twice the length of the yeast CTD. It can be divided in a N-terminal part, which is almost exclusively composed of consensus repeats ($Y_1S_2P_3T_4S_5P_6S_7$) and is highly similar to the yeast CTD (highlighted are identical regions between the two species), and a C-terminal part that is more degenerate (Hsin, J.P. & Manley, J.L., *Genes Dev.* **26**, 2119-37, 2012). Deviations from the consensus sequence are indicated in black. (**b**) Absolute (black bars) and relative (green line) abundance of tyrosine residues within the CTD sequences of various distantly related species (from left to right): *Trichomonas vaginalis* (Nucleotide ID: TVU20501, Protein ID: AAB50025), *Trypanosoma brucei* (Nucleotide ID: X13492, Protein ID: P17545), *Leishmania donovani* (Nucleotide ID: AF126254, Protein ID: AAF36438), *Monoblepharis macrandra* (Nucleotide ID: AF315822, Protein ID: AAK00312), *Acanthamoeba castellanii* (Nucleotide ID: ACU90211, Protein ID: AAC18417), *Plasmodium falciparum 3D7* (Nucleotide ID: XM_001351216, Protein ID: XP_001351252), *Vairimorpha necatrix* (Nucleotide ID: AF060234, Protein ID: AAD12604), *Glaucosphaera vacuolata* (Nucleotide ID: AF315820, Protein ID: AAK00310), *Dictyostelium discoideum AX4* (Nucleotide ID: XM_636643, Protein ID: XP_641735), *Nosema ceranae* (Nucleotide ID: ACOL01000247, Protein ID: EEQ81731), *Breviata anathema* (Nucleotide ID: AF083338, Protein ID: AAC62246), *Saccharomyces cerevisiae* (Nucleotide ID: X96876, Protein ID: CAA65619), *Schizosaccharomyces pombe* (Nucleotide ID: NM_001021568, Protein ID: NP_595673), *Monosiga brevicollis* (Nucleotide ID: AF315821, Protein ID: AAK00311), *Arabidopsis thaliana* (Nucleotide ID: X52954, Protein ID: CAA37130), *Caenorhabditis elegans* (Nucleotide ID: CELPOLII, Protein ID: AAA28126.1), *Drosophila melanogaster* (Nucleotide ID: NM_078569, Protein ID: NP_511124), *Culex quinquefasciatus* (Nucleotide ID: XM_001866696, Protein ID: XP_001866731), *Branchiostoma floridae* (Nucleotide ID: XM_002593172, Protein ID: XP_002593218), *Strongylocentrotus purpuratus* (Nucleotide ID: XM_001175967, Protein ID: XP_001175967), *Ixodes scapularis* (Nucleotide ID: XM_002407458, Protein ID: XP_002407502), *Tribolium castaneum* (Nucleotide ID: XM_968377, Protein ID: XP_973470), *Danio rerio* (Nucleotide ID: XM_677590, Protein ID: XP_682682), *Mus musculus* (Nucleotide ID: MMU37500, Protein ID: AAB58418), *Homo sapiens* (Nucleotide ID: NM_000937.3, Protein ID: NP_000928) and *Hydra magnipapillata* (Nucleotide ID: XM_002164318, Protein ID: XP_002164354). Although the number of repeats greatly differs among species, the tyrosine content is well conserved. Bars corresponding to *S. cerevisiae* und *H. sapiens* are highlighted in red and blue, respectively. The number/definition of repeats was taken from Simonti, C.N. *et al.*, *BMC Evol Biol.* **15**, 35, 2015. (**c**) Circular dichroism difference spectra of MBP-hCTD and MBP (upper panel), and MBP-yCTD and MBP (lower panel) show a minimum at ~194 nm that is characteristic for a random coil, confirming that the CTD is intrinsically disordered.
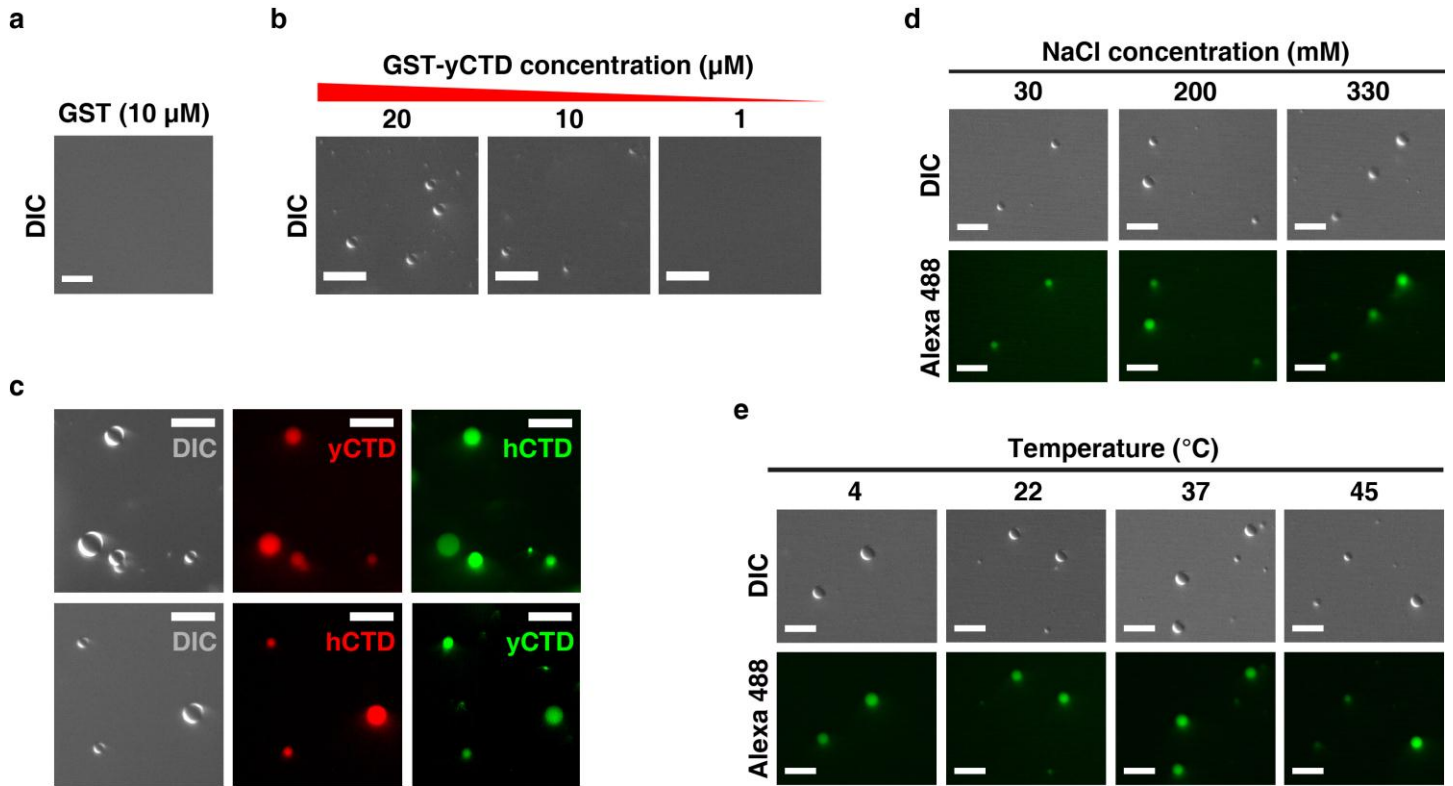
Supplementary Figure 2

Physicochemical properties of yCTD phase separation.

(**a**) GST alone (10 µM) did not undergo phase separation in 16% dextran. (**b**) Concentration-dependent phase separation of GST-yCTD in the presence of 150 mg/ml Ficoll. (**c- upper panels**) Recruitment of Alexa 488-labeled MBP-hCTD (green; right) to preformed droplets of GST-yCTD that were visualized by DIC microscopy (left) and by recruitment of a TMR-labeled YSPTSPS peptide (red; middle). (**c –lower panels**) Recruitment of Alexa 488-labeled GST-yCTD (green; right) to preformed droplets of MBP-hCTD. Preformed hCTD droplets were visualized by DIC microscopy (left) and by recruitment of the TMR-labeled YSPTSPS peptide (red; middle). (**d, e**) LLPS of MBP-yCTD in 16% dextran is not sensitive to ionic strength changes (d) or incubation of the sample for one hour at the indicated temperatures (e). All experiments were performed two times with similar outcome and representative images are shown. Scale bars correspond to 10 µm in all panels.
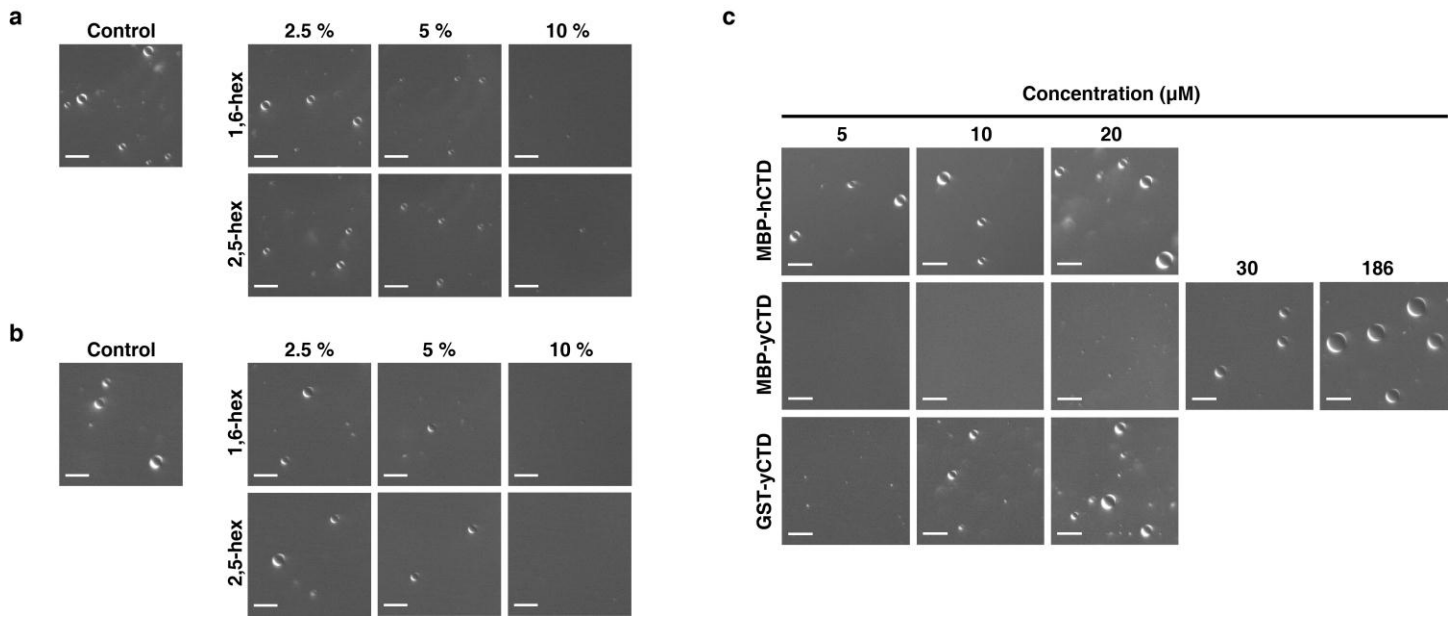
Supplementary Figure 3

Influence of aliphatic alcohols and solubility tags on CTD phase separation.

(**a, b**) Influence of aliphatic alcohols on CTD LLPS. 1,6-hexanediol (1,6-hex) and 2,5-hexanediol (2,5-hex) counteract LLPS of hCTD (a) or yCTD (b) in a concentration-dependent manner. Images from ≥5 representative images taken for each condition are shown. (**c**) The solubilizing effect of a N-terminal MBP-tag inhibits droplet formation of yCTD at low protein concentrations. Concentration-dependent LLPS of MBP-hCTD, MBP-yCTD and GST-yCTD in 16% dextran. MBP-hCTD undergoes liquid phase separation already at a concentration of 5 µM (top), while a ≥ 4-fold higher concentration of MBP-yCTD is required for droplet formation (middle). Replacement of MBP by a GST-tag decreases the critical concentration for yCTD phase separation to approximately 5 µM (bottom). Shown images are representative of two independent experiments. Scale bars correspond to 10 µm in all panels.
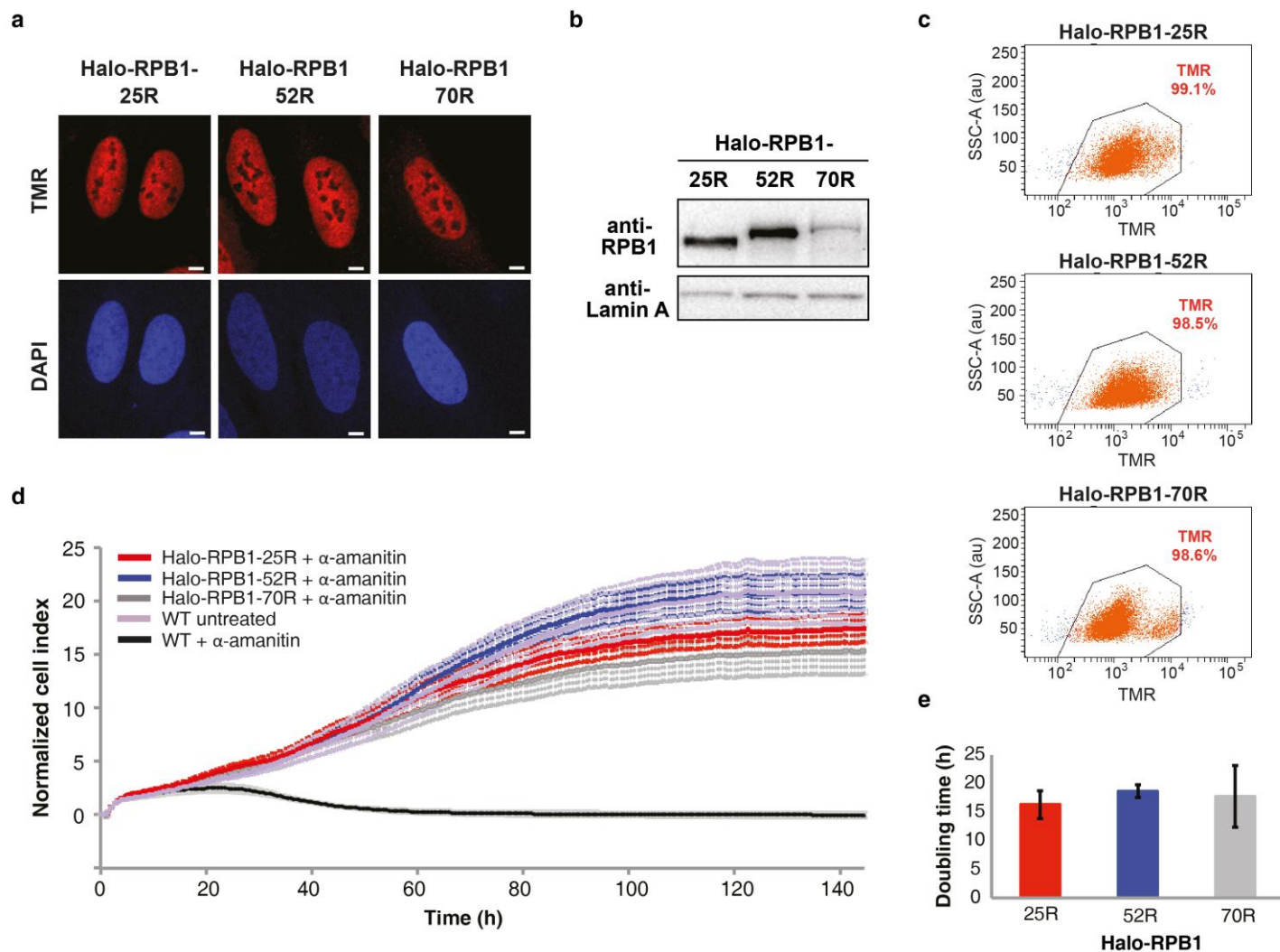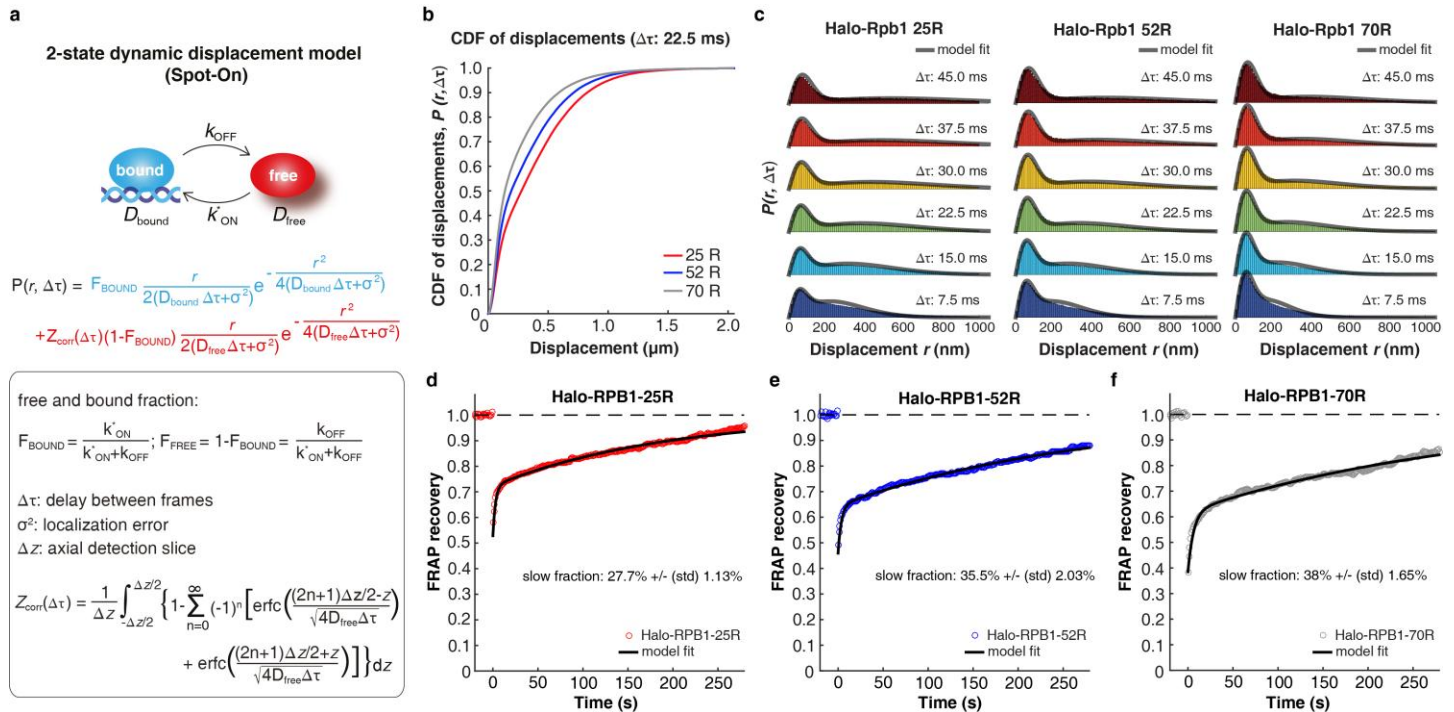
Supplementary Figure 4

Characterization of Halo-RPB1 cell lines.

(**a**) Confocal image of RPB1-25R (left), RPB1-52R (middle) and RPB1-70R (right) cell lines showing the nuclear localization of Halo-RPB1. Before fixation the cells were labeled for 30 min with HALO-TMR ligand (500 nM final concentration). Scale bars correspond to 5 μm. (**b**) Western blot analysis of the Halo-RPB1-25R, -52R and -70R expression level for the different cell lines. (**c**) FACS analysis to evaluate Halo-RPB1 expression levels of the different cell lines. (**d**) Growth curve analysis of the Halo-tagged Pol II cell lines. The growth rate of the three cell lines is overall similar albeit the RPB1-25R (red) and RPB1-70R (grey) cell lines grew slightly slower than the U2OS WT (purple) and RPB1-52R (blue) cell lines. Once treated with alpha-amanitin, the WT cells (black) die while the growth rate of the other lines is unaffected. Growth curves show mean across n = 6 independent samples and error bars show the standard deviation. Growth curves of a representative experiment performed independently five times are shown. (**e**) Doubling time analysis of the Halo-tagged Pol II cell lines. The three cell lines have relatively similar doubling times. The mean across n = 3 independent replicates is shown and error bars show the standard deviation.
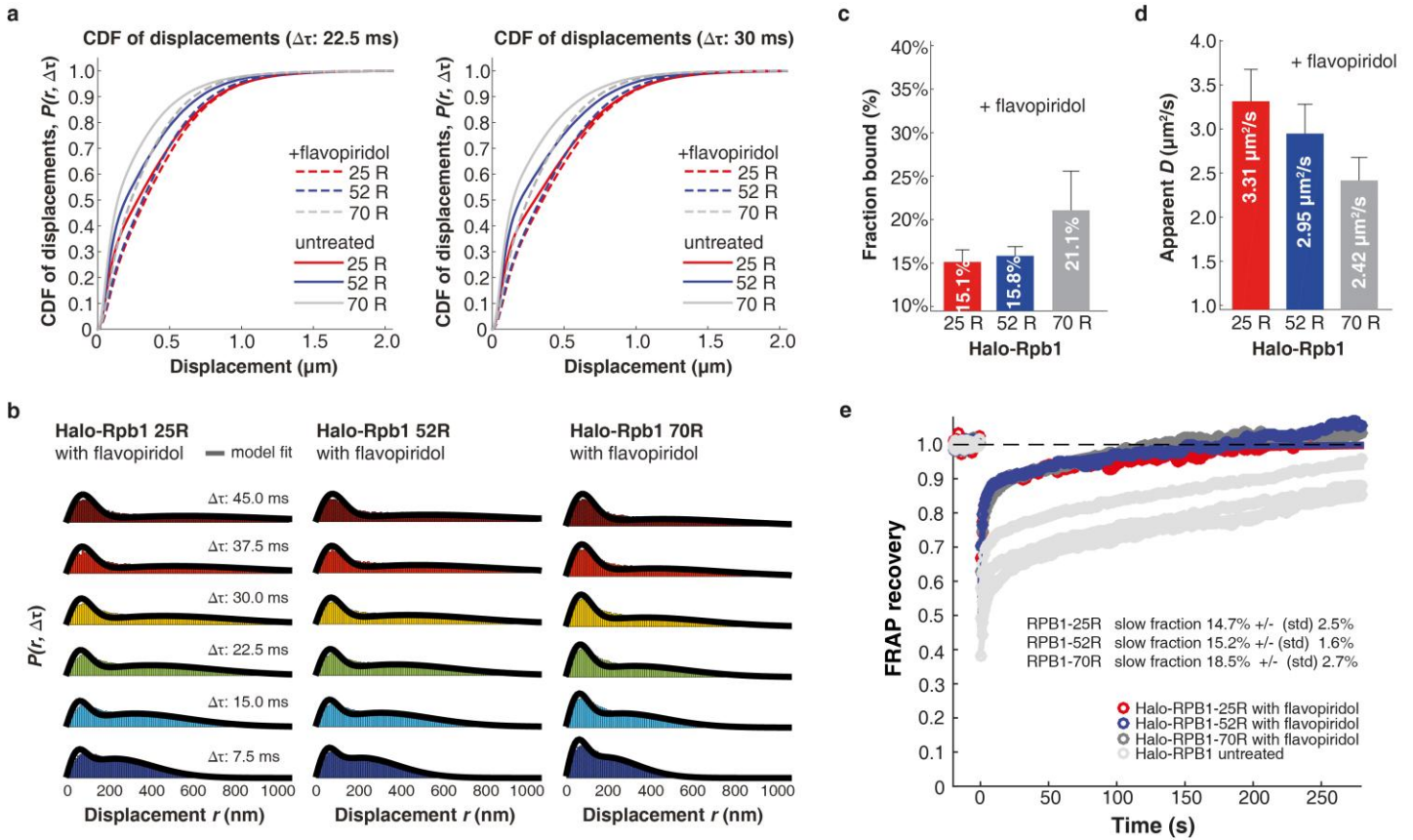
Supplementary Figure 5

spaSPT and FRAP.

(**a**) Overview of the 2-state model used in fitting the displacement data from spaSPT (Spot-On) (Hansen, A.S. *et al.*, *Elife* **7**, 2018). Diffusion is modeled as Brownian and arising from a bound/immobile population and a freely diffusing population assuming no state transitions at the short time-scale of observation. A correction is applied to the free population to correct for "defocalization": since the 2D imaging only captures a ~700 nm axial slice of the nucleus, the free population rapidly moves out of focus at later time points. (**b**) Cumulative distribution functions (CDF) for displacements. The CDF of displacements for the representative time-lag $\Delta\tau$ = 22.5 ms is shown for Halo-RPB1-25R, Halo-RPB1-52R and Halo-RPB1-70R. The data shown is merged from three independent replicates (n = 29, 30, 26 cells in total for Halo-RPB1-25R, -52R, -70R, respectively). (**c**) Model fit to displacement histograms. Raw displacements from spaSPT data for six different time-lags are shown for Halo-RPB1-25R, Halo-RPB1-52R and Halo-RPB1-70R. Model-fitting from a two-state (bound vs. free) model is overlaid, from which the diffusion constants and subpopulation sizes were calculated. The data shown is merged from three independent replicates (n = 29, 30, 26 cells in total for Halo-RPB1-25R, -52R, -70R, respectively). (**d-f**) FRAP data of Halo-RPB1-25R (d), Halo-RPB1-52R (e) and Halo-RPB1-70R (f) were fitted to a reaction dominant two-state model (Sprague, B.L. *et al.*, *Biophys J.* **86**, 3473-95, 2004; Hansen, A.S. *et al.*, *Elife* **6**, 2017). We performed 50 iterations using 50% of the data in each to estimate the error (standard deviation of the subsampling) on the bound fraction.
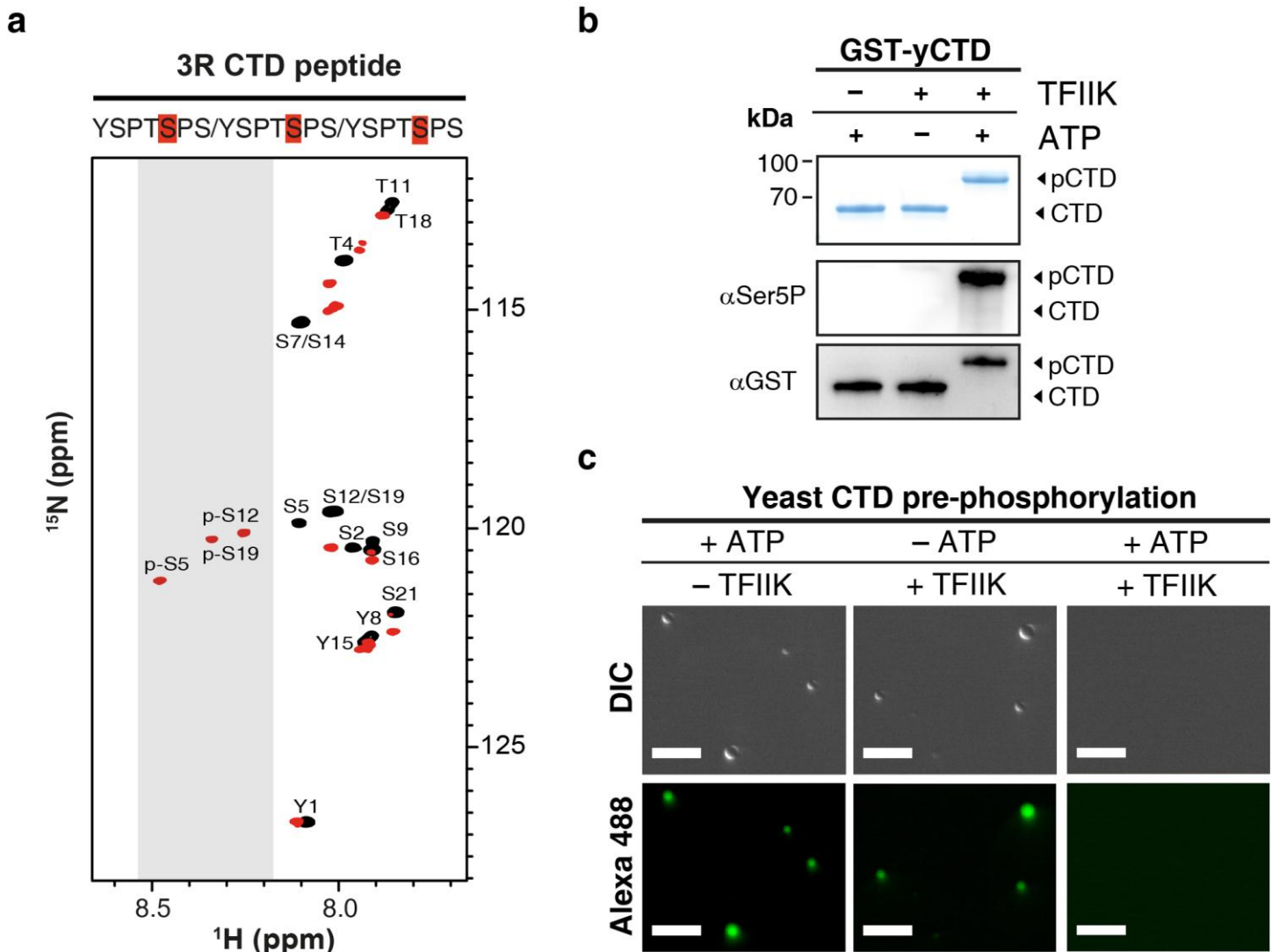
Supplementary Figure 6

Effect of flavopiridol treatment on RNA Pol II dynamics.

(**a**) Cumulative distribution functions (CDF) for displacements. The CDF of displacements for the representative time-lags $\Delta\tau$ = 22.5 ms and $\Delta\tau$ = 30.0 ms are shown for Halo-RPB1-25R, Halo-RPB1-52R and Halo-RPB1-70R after treatment with flavopiridol. The data shown is merged from three independent replicates. (**b**) Model fit to displacement histograms. Raw displacements from spaSPT data for six different time-lags are shown for Halo-RPB1-25R, Halo-RPB1-52R and Halo-RPB1-70R. Model-fitting from a two-state (bound vs. free) model is overlaid, from which the diffusion constants and subpopulation sizes were calculated. The data shown is merged from three independent replicates (n = 13, 15, 28 cells in total for Halo-RPB1-25R, -52R, -70R, respectively). (**c**) Bound fractions of Halo-RPB1-25R, -52R and -70R after flavopiridol treatment. The bound fraction was inferred from two-state model-fitting to the spaSPT displacement data using Spot-On (Hansen, A.S. *et al.*, *Elife* **7**, 2018). Each of three independent replicates was fitted separately and bar graphs show the mean and standard error. (**d**) Diffusion coefficients of the free population of Halo-RPB1-25R, -52R and -70R. Free diffusion coefficients were inferred from two-state model-fitting to the spaSPT displacement data using Spot-On (Hansen, A.S. *et al.*, *Elife* **7**, 2018). Each of three independent replicates was fitted separately and bar graphs show the mean and standard error. (**e**) FRAP data of Halo-RPB1-25R (red), Halo-RPB1-52R (blue) and Halo-RPB1-70R (grey) after flavopiridol treatment. The data were fitted to a reaction dominant two-state model (Sprague, B.L. *et al.*, *Biophys J.* **86**, 3473-95, 2004; Hansen, A.S. *et al.*, *Elife* **6**, 2017). FRAP curves represent the mean across n = 5 independent measurements for each cell line and error bars show the standard error. We performed 50 iterations using 50% of the data in each to estimate the error (standard deviation of the subsampling) on the bound fraction. For comparison purposes, the FRAP curves of untreated cells (presented in **Fig. 5f**) are overlaid (light grey).

Supplementary Figure 7

CTD phosphorylation by recombinant yeast TFIIK complex (Kin28–Ccl1–Tfb3).

(**a**) Superposition of ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectra of a 21-residue peptide comprising three consecutive YSPTSPS repeats in its non-phosphorylated form (black) and after phosphorylation by recombinant yeast TFIIK complex (red). Sequence-specific resonance assignments were achieved by a combination of two-dimensional NOESY and TOCSY spectra. Unambiguous assignments were obtained for all residues of the non-phosphorylated peptide and most of the residues of the phosphorylated peptide. The NMR signals of phosphorylated S12 (p-S12) and phosphorylated S19 (p-S19) can be assigned to either the cross-peaks at 8.28 or 8.34 ppm ^{1}H frequency, respectively. The shift of the three NMR signals of S5, S12 and S19 from its non-phosphorylated positions to the region from approximately 8.2 to 8.6 ppm (shown in grey) demonstrates that the three Ser5-positions in the YSPTSPS hepta-repeats were fully phosphorylated. In addition, the NMR data show that no other residue in this 21-residue peptide was phosphorylated by TFIIK. (**b**) Sodium dodecyl sulfate polyacrylamide gel electrophoresis and western blot analysis of phosphorylated GST-yCTD fusion protein. GST-yCTD was treated with recombinant yeast TFIIK. The CTD substrate became highly phosphorylated, resulting in a pronounced mobility change during polyacrylamide gel electrophoresis in comparison to the non-phosphorylated substrate (–ATP and –kinase control reactions). 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. (**c**) Phosphorylation of GST-yCTD by recombinant TFIIK inhibits phase separation. GST-yCTD was pre-incubated with TFIIK and ATP leading to its phosphorylation. A 10 µM solution of TFIIK-phosphorylated GST-yCTD does not undergo phase separation in 16 % dextran. This effect is neither caused by hydrotropic properties of ATP nor the pure presence of the kinase, since GST-yCTD readily forms droplets in control reactions containing ATP or the kinase alone. Images from ≥5

representative images taken for each condition are shown. Scale bars correspond to 10 µm.