Investigating phase separation mechanisms for transcriptional control

Dissertation

for the award of the degree

“Doctor rerum naturalium”

of the Georg-August-Universität Göttingen

within the doctoral program IMPRS Molecular Biology

of the Georg-August University School of Science (GAUSS)

submitted by

Marc Böhning

from Fulda, Germany

Göttingen 2019

Thesis Advisory Committee

Prof. Dr. Patrick Cramer	Department of Molecular Biology
	Max Planck Institute for Biophysical Chemistry, Göttingen
Prof. Dr. Henning Urlaub	Research Group Bioanalytical Mass Spectrometry
	Max Planck Institute for Biophysical Chemistry, Göttingen
Prof. Dr. Steven A. Johnsen	Department of Gastroenterology and Hepatology
	Mayo Clinic, Rochester (USA)

Members of the Examination Board

Prof. Dr. Patrick Cramer (1st reviewer)	Department of Molecular Biology
	Max Planck Institute for Biophysical Chemistry, Göttingen
Prof. Dr. Henning Urlaub (2nd reviewer)	Research Group Bioanalytical Mass Spectrometry
	Max Planck Institute for Biophysical Chemistry, Göttingen

Further members of the Examination Board

Prof. Dr. Markus Zweckstetter	Translational Structural Biology in Dementia
	German Center for Neurodegenerative Diseases, Göttingen
Prof. Dr. Argyris Papantonis	Translational Epigenetics Laboratory
	University Medical Center Göttingen
Prof. Dr. Herbert Jäckle	Department of Molecular Developmental Biology
	Max Planck Institute for Biophysical Chemistry, Göttingen
Dr. Peter Lenart	Research Group Cytoskeletal Dynamics in Oocytes
	Max Planck Institute for Biophysical Chemistry, Göttingen

Date of oral examination: 20th of November 2019

Affidavit

I, Marc Böhning, hereby declare that my dissertation entitled ‘Investigating phase separation mechanisms for transcriptional control’ has been written independently and with no other sources and aids than quoted. This dissertation or parts thereof have not been submitted elsewhere for any academic award or qualification. The electronic version of this dissertation is congruent to the printed versions.

Göttingen, 30th of September 2019

(Marc Böhning)

“The living protoplasm […] is a liquid or rather a mixture of liquids in the form of a fine emulsion consisting of a continuous substance in which are suspended drops […] of different chemical nature.”

Edmund B. Wilson, Science, July 1899 [1]

Acknowledgements

First and foremost, I would like to thank Prof. Dr. Patrick Cramer for giving me the opportunity to conduct my PhD research in such an outstanding research environment. I'm extremely grateful for his continuous support and interest in all my projects. Patrick's instant trust and continuous encouragement have motivated me throughout the entire time of my PhD research. I deeply appreciated the granted freedom to develop and peruse my own ideas and, and, at the same time, was really thankful for many insightful discussions that have been instrumental to overcome the difficulties and challenges.

Parts of the work presented in this thesis were carried out in an extremely interactive and fruitful collaboration with the laboratory of Prof. Dr. Markus Zweckstetter, and would not have been possible in their entirety without the dedicated help of Dr. Marija Rankovic. I'm also very grateful to Markus Zweckstetter for his commitment and determination to push the project forward. A big thanks goes also to our oversea collaborators from the Tjian-Darzacq lab, especially to Dr. Claire Dugast-Darzacq and Prof. Dr. Xavier Darzacq, whose contributions have been crucial to the success of this work.

The second part of the work presented in this thesis was conducted in a great collaboration with Dr. Ritwick Sawarkar and Prashant Rawat. I'm grateful for their keen dedication and insatiable curiosity that facilitated very rapid progress, and for keeping up with my numerous enquiries and questions. I'm looking forward to seeing it published soon.

I'm grateful to all my lab internal collaborators, in particular Andrea Boltendahl for dedicated and excellent technical assistance, Dr. Seychelle Vos for numerous insightful discussions and for sharing her invaluable collection of protocols, plasmids and proteins, as well as Dr. Goran Kokic and Dr. Sandra Schilbach for sharing proteins. I thank all present and past members of Lab 116, Kerstin Maier, Petra Rus, Andrea Boltendahl, Dr. Sofia Battaglia, Dr. Carlo Bäjen, and Dr. Katharina Hofmann for daily discussions and for creating a truly enjoyable atmosphere. All work presented in this thesis was highly facilitated through an outstanding and reliable lab infrastructure provided and maintained by Kirsten Backs, Janine Blümel, Kerstin Maier, Petra Rus, Ute Neef, Thomas Schulz, Mario Klein, Angelika Kruse and Manuela Wenzel.

I’m thankful to all my great colleagues within the Cramer lab and on the entire MPI campus, many of which I can count as my friends. I thank them all for lots of fun, entertainment and (non-) scientific discussions inside and outside the lab.

Our weekly interdisciplinary phase separation journal club meetings at lunch with Dr. Johannes Söding, Dr. David Zwicker, Salma Sohrabi-Jahromi, Jan Kirschbaum and Matthew Grieshop have been a very fun and rewarding experience. I would like to especially thank Johannes Söding and Salma Sohrabi-Jahromi for numerous enlightening discussions.

I would like to thank Jan Böhning, Kerstin Maier, Salma Sohrabi-Jahromi, Dr. Katharina Hoffmann, Dr. Goran Kokic, and Prashant Rawat for critically reading of parts of this thesis.

I am grateful to the additional members of my Thesis Advisory Committee, Prof. Dr. Henning Urlaub and Prof. Dr. Steven Johnsen for providing valuable input and guidance during our meetings. I’m especially thankful to Henning Urlaub for examining this thesis as a second reviewer, for many interesting discussions and for granting me access to the mass spectrometers in his laboratory throughout the time of my PhD work.

Thanks to my additional examination committee members Prof. Dr. Markus Zweckstetter, Prof. Dr. Argyris Papantonis, Prof. Dr. Herbert Jäckle, and Dr. Peter Lenart for their interest and time evaluating this thesis.

Last but not least, I’m extremely grateful to my parents and my brother for their unconditional support, continuous encouragement, and sincere understanding throughout this time.

Publications

Part of this work has been published or is in the process of being published:

RNA polymerase II clustering through carboxy-terminal domain phase separation

M. Boehning[*], C. Dugast-Darzacq[*], M. Rankovic[*], A. S. Hansen, T. Yu, H. Marie-Nelly, D. T. McSwiggen, G. Kokic, G. M. Dailey, P. Cramer[#], X. Darzacq[#], M. Zweckstetter[#]

([*]) Equal contribution, ([#]) Corresponding author

Nature Structural and Molecular Biology **25**, 833–840 (2018)

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 TFIIH 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.

Stress-induced nuclear condensation of NELF drives transcriptional downregulation

P. Rawat[*#], **M. Boehning**[*], B. Hummel, F. Aprile-Garcia, A. S. Pandit, N. Eisenhardt, A. Khavaran, E. Niskanen, S. M. Vos, J. J. Palmivo, A. Pichler, P. Cramer[#], R. Sawarkar[#]

([*]) Equal contribution, ([#]) Corresponding author

Manuscript in revision.

Current author contributions: P.R. and R.S. initiated the project. P.R. performed imaging and molecular cell biology experiments. M.B designed and performed all in vitro experiments. S.M.V. purified recombinant NELF complex. B.H. performed all the computational analysis. F.A.G., A.S.P., A.K., N.E. and E.N. helped with the experiments. J.J.P., A.P., P.C. and R.S. supervised. P.R., M.B., P.C. and R.S. wrote the manuscript with inputs from all other authors.

A detailed summary of items excerpted from published manuscripts can be found in the Appendix (‘List of items from publications’, Page 149). Excerpts from published manuscripts are additionally indicated at the beginning of each section. Co-author contributions are stated in the figure captions and/or Methods Section. The presentation of unpublished data that was not generated by the author of this thesis has been authorized by the co-authors who generated the data.

Publications

Contributions to other publications:

Structure of activated transcription elongation complex Pol II-DSIF-PAF-SPT6

S. M. Vos, L. Farnung, **M. Boehning**, C. Wigge, A. Linden, H. Urlaub, P. Cramer[#]

([#]) Corresponding author

Nature **560**, 607-612 (2018)

Author contributions: S.M.V. designed and conducted all experiments unless stated otherwise. L.F. established and conducted SPT6 preparation and crystallized the SPT6 tSH2 domain. M.B. determined linker phosphorylation sites by mass spectrometry. C.W. assisted in cryo-EM data collection. A.L. performed crosslinking–mass spectrometry, supervised by H.U. P.C. supervised the research. S.M.V. and P.C. wrote the manuscript with input from L.F., M.B. and H.U.

Mechanisms for active regulation of biomolecular condensates

J. Söding[#], D. Zwicker, S. Sohrabi-Jahromi, **M. Boehning**, J. Kirschbaum

([#]) Corresponding author

Trends in Cell Biology **30**, 4-14 (2020)

Author contributions: J.S. and D.Z. initiated the study. J.S. designed and prepared main figures. J.S. wrote manuscript with input from all authors. S.S.J. drafted introduction and evidence supporting localization-induction. M.B. drafted evidence supporting enrichment-inhibition. D.Z. and J.K. contributed to theoretical modelling.

Summary

Transcription of protein-coding genes by RNA polymerase (Pol) II is a highly coordinated process. In metazoan cells, transcription is regulated both at the initiation step by recruitment of the Pol II machinery as well as during early elongation by promoter-proximal pausing.

Prior to transcription initiation, Pol II forms short-lived clusters near active gene promoters, but the underlying molecular basis has remained unknown. Pol II possesses a disordered C-terminal heptad repeat domain (CTD) that is essential for factor recruitment during the transcription cycle. CTD length is organism-specific with 52 repeats in human and 26 repeats in yeast. In this work, we report that the human and yeast CTD can undergo concentration-dependent liquid-liquid phase separation in vitro, based on weak multivalent repeat-repeat interactions. We show that this behavior strongly correlates with the repeat number, as the shorter yeast CTD forms less-stable droplets. Shortening of the CTD in human cells to the length of the yeast CTD reduces Pol II clustering and chromatin-association, while artificial extension has the contrary effect. Repeat-repeat interactions are sensitive to CTD phosphorylation by the transcription factor IIH kinase CDK7, which dissolves CTD droplets in vitro. Together these results imply a model for gene activation that involves CTD-mediated clustering of initiation-competent Pol II and release through CTD phosphorylation upon transcription initiation.

Heat shock causes the accumulation of the negative elongation factor (NELF) at chromatin, which stabilizes paused Pol II within the promoter-proximal region of downregulated target genes. In this work, we show that NELF clusters in nuclear puncta upon heat shock, which possess properties consistent with phase-separated condensates. In vitro, purified NELF complex self-interacts to form phase-separated droplets with liquid-like properties. We show that multivalent interactions between the disordered NELF tentacles are essential for NELF phase separation in vitro and stress-induced condensation in vivo. Phosphorylation by positive elongation factor b (P-TEFb) counteracts NELF phase separation in vitro and is prevented through the inactivation of P-TEFb upon heat shock in vivo. Sumoylation is further required for stress-induced NELF condensation, as NELF itself can be sumoylated in vitro and interacts with SUMO2/3 in a chain length-dependent manner. Together with published data, our results suggest a model that involves stress-induced sequestration of promoter-proximal paused Pol II by NELF near downregulated gene promoters.

Taken together, the findings presented in this work indicate that phase separation mechanisms regulate key steps of eukaryotic gene transcription and provide a basis to further analyze the role of phase separation within the Pol II transcription cycle, as well as investigate its modulation in the future.

Table of contents

1 Introduction

1.1 Key principles of eukaryotic gene transcription

The eukaryotic cell nucleus contains the vast majority of an organism's genetic information, which is stored in the form of the deoxyribonucleic acid (DNA) double helix[2]. In multicellular lifeforms, every cell contains the genetic blueprint to give rise to the entire organism[3]. Yet, cells can exhibit vast morphological and functional differences. Such variation arises through the activation of different genes in different cell types. The central dogma of molecular biology describes the directional flow of the genetic information[4]: All genetic information is encoded in the form of DNA, copied to the transient messenger ribonucleic acid (mRNA), and translated into a linear polypeptide chain that folds into a functional protein, the workhorse molecules of the cell that largely define its phenotype[4]. Thus, as the critical initial step, the regulated cell type-specific transcription of genes into RNA largely determines cellular identity[5].

Gene transcription is carried out by DNA-dependent RNA polymerases that utilize a DNA template to produce complementary single-stranded RNA molecules using nucleotide triphosphates as substrate[6]. While bacteria and archaea have only one type of RNA polymerase, transcription of the nuclear genome in eukaryotes requires up to five different RNA polymerases (Pol I-V), which synthesize functionally distinct transcripts[7, 8]. Pol I and Pol III catalyze the synthesis of ribosomal and transfer RNAs required for protein biosynthesis as well as some small non-coding RNAs[9-11]. The plant-specific enzymes Pol IV and Pol V produce small interfering RNAs (siRNAs) that are involved in gene silencing[12]. RNA polymerase II (Pol II) carries out the transcription of all protein-coding genes into precursor-messenger RNA (pre-mRNA)[13, 14]. Synthesized pre-mRNAs then undergo various co-transcriptional and post-transcriptional processing steps before they can serve in their mature form as templates for ribosomal protein synthesis in the cytoplasm[15, 16]. Since the repertoire of functional protein molecules results from the set of assembled mRNA transcripts, Pol II transcription is the central determinant of the cellular proteome and shapes its identity.

Pol II transcription can be divided into three key steps: Initiation, elongation and termination (Fig. 1.1). Each of these steps is tightly controlled through a myriad of accessory factors that

associate with Pol II in a transcription stage-dependent manner. The Pol II C-terminal domain (CTD), a unique tail-like C-terminal extension of the largest Pol II subunit RPB1, plays a pivotal role for the recruitment of many of these factors to the transcriptional machinery during the different phases of the transcription cycle.

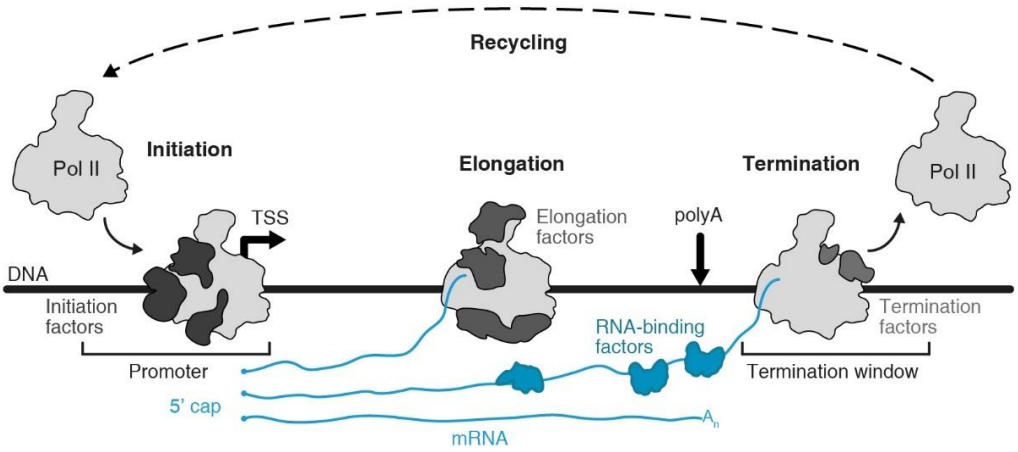

Figure 1.1 | Pol II transcription cycle.
During transcription initiation, Pol II binds together with the general transcription factors to the core promoter in close proximity to the transcription start site (TSS). After opening of the promoter DNA and promoter escape, elongation factors bind to Pol II to form a productive elongation complex that extends the mRNA (blue) in a processive manner. At the end of the transcription unit, mRNA becomes cleaved after reaching the cleavage and polyadenylation (polyA) site. Pol II continues elongation before it gets destabilized through binding of termination factors and ultimately dissociates from the template. During recycling, released Pol II is then prepared to engage in a new round of transcription. Co-transcriptional modification with a 5' cap structure and post-transcriptional modification with a polyA-tail (A_n) renders the produced RNA competent for nuclear export and translation. The Pol II C-terminal domain was omitted for clarity. Figure was adapted from Hantsche & Cramer (2016)[17].

1.1.1 RNA polymerase II carboxy-terminal domain (CTD)

Early studies in the 1980s uncovered the presence of an 'unusual' repetitive amino acid sequence at the carboxy-terminal end of the largest Pol II subunit RPB1 that was not present in any other bacterial, archaeal or eukaryotic RNA polymerase[18, 19]. Later studies revealed that the C-terminal domain of the largest RNA Pol II subunit RPB1 is instrumental to coordinate the association of accessory factors with Pol II during the transcription cycle (reviewed in[20-22]). The Pol II CTD forms a tail-like extension near the RNA exit size and consists of multiple heptad repeats with the conserved consensus sequence $Y_1S_2P_3T_4S_5P_6S_7$. The number of heptad

repeats within the CTD varies between species and roughly scales with the complexity of the organism[23]: While the CTD of the yeast *S. cerevisiae* contains 26 heptapeptide repeats, the human CTD is composed of 52 repeats (Fig. 1.2a). The human CTD can be divided in an N-terminal proximal half that consists largely of consensus repeats and closely resembles the yeast CTD, and a C-terminal distal half with many non-consensus heptads that diverge mostly at position 7 (Fig. 1.2a). Within the human CTD, Y_1 and P_6 show strongest conservation and are present in all 52 repeats. In addition, the tyrosine content is strongly conserved even between distantly related species (Fig. 1.2b).

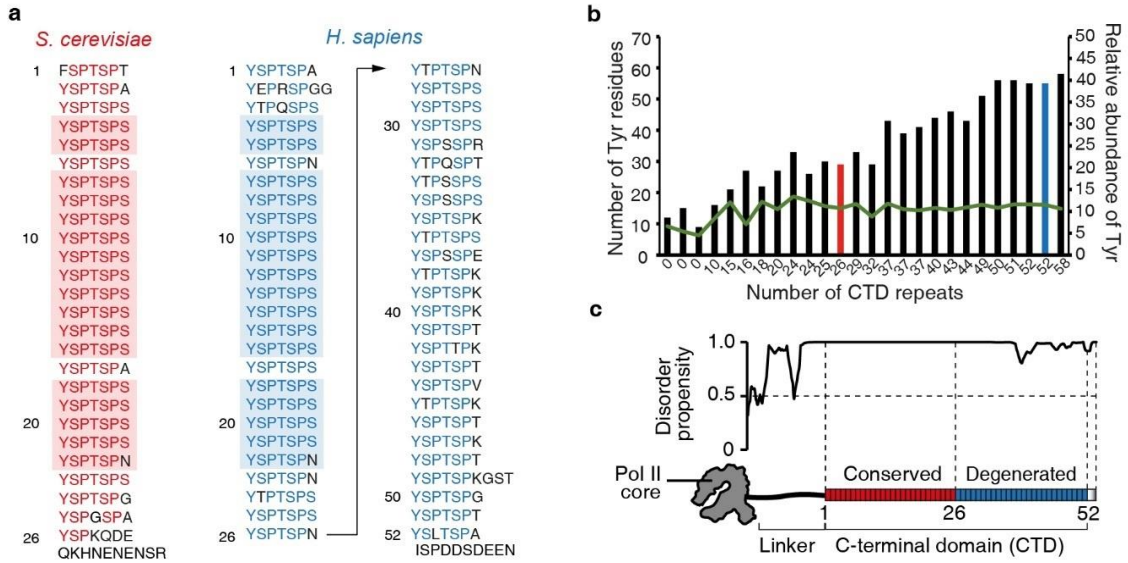

Figure 1.2 | The Pol II C-terminal domain.

a, Comparison of human and yeast Pol II CTD sequences. The human CTD comprises 52 heptad repeats and can be divided into an N-terminal part that consists mostly of repeats with the consensus sequence $Y_1S_2P_3T_4S_5P_6S_7$ and closely resembles the CTD of the yeast *S. cerevisiae* and a C-terminal part that contains mostly divergent repeats. **b,** Tyrosine content within the CTDs of distantly related species. Despite great variation in length and repeat number, the abundance of tyrosine within the CTD is strongly conserved in all eukaryotes. The species are from left to right: *Trichomonas vaginalis*, *Trypanosoma brucei*, *Leishmania donovani*, *Monoblepharis macrandra*, *Acanthamoeba castellanii*, *Plasmodium falciparum 3D7*, *Vairimorpha necatrix*, *Glaucosphaera vacuolata*, *Dictyostelium discoideum AX4*, *Nosema ceranae*, *Breviata anathema*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Monosiga brevicollis*, *Arabidopsis thaliana*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Culex quinquefasciatus*, *Branchiostoma floridae*, *Strongylocentrotus purpuratus*, *Ixodes scapularis*, *Tribolium castaneum*, *Danio rerio*, *Mus musculus*, *Homo sapiens* and *Hydra magnipapillata.* The red and blue bars correspond to *S. cerevisiae* and *H. sapiens*, respectively. **c**, Disorder prediction of the human CTD sequence using the PONDR tool[24] (for details see Section 2.3.1.7). The CTD is predicted to be disordered over its entire length. All panels of this figure were adapted from Boehning et al. (2018)[25].

While the CTD in its entirety is dispensable for templated Pol II transcription in vitro, it is essential for viability in vivo[26, 27]. However, CTD truncation is tolerated to a limited extent in vivo, suggesting that repeats possess largely redundant functions: Mutational studies in yeast have revealed that at least 8 repeats are required for viability, but resulted in a slow growth phenotype[28]. Thirteen repeats resulted in normal yeast growth, but the cells exhibited defects in stimulus-activated transcription[29, 30]. Similarly, in mammalian cells about half of the repeats were sufficient to support growth[31], but reduced the responsiveness to enhancer-mediated transcriptional activation[32].

Despite its pivotal role during the transcription cycle, insights into the structure of the Pol II CTD have been limited. The CTD is absent from Pol II structures obtained by X-ray crystallography or cryo-electron microscopy (cryo-EM), indicating a large degree of inherent flexibility[21]. Consistent with this, computational algorithms predict the CTD to be disordered over its entire length (Fig. 1.2c). Disordered protein domains can assume various conformations that range from extended structures via a random coil to compact globules, depending on the degree of favorable intramolecular interactions that lead to chain compaction[33, 34]. Several lines of evidence indicate that the CTD assumes a rather compact conformation in dilute aqueous solutions[21]: The limited available space within the Pol II crystal lattice can only accommodate a compact CTD[13]. Moreover, negative stain electron microscopy analyses of wild-type and CTD-deficient Pol II measured a weak density difference of only ~100 Å in diameter[35] and provided indirect evidence for a compact conformation of a recombinant CTD fusion protein in complex with a submodule of the transcriptional regulator Mediator[36]. In agreement with previous data, recent cryo-EM analysis of the Pol II pre-initiation complex with Mediator suggested that only a compact CTD globule can be accommodated within an open space between Pol II and Mediator[37]. In support of this hypothesis, a rather compact structure was also inferred from biophysical analysis of recombinant CTD from various species using size-exclusion chromatography and small-angle X-ray scattering[38].

The Pol II CTD becomes extensively post-translationally modified during the transcription cycle[22]. Most prominently, all five CTD consensus amino acids that possess hydroxyl groups, Y_1, S_2, T_4, S_5, and S_7, have been shown to undergo reversible phosphorylation[39-42]. CTD phosphorylation marks are established through the dynamic interplay of transcriptional kinases and phosphatases that interact with Pol II in a transcription stage-dependent manner.

Specific phosphorylation patterns are recognized by CTD-binding factors that form binary interactions with short, 1-3 heptad comprising CTD segments and facilitate Pol II progression through the transcription cycle (see Section 1.1.2). The differential phosphorylation of up to 5 repeat positions, the isomerization of proline[43] as well as the methylation[44] and acetylation[44] of lysine, and the methylation[45] and citrullination[46] of arginine residues in distal non-consensus repeats were thus proposed to give rise to an elaborate 'CTD code'[47]. However, despite of the high conceivable complexity of such a code, recent mass spectrometric analyses of CTD phosphorylation in yeast and mammalian cells suggested that each heptad contains on average less than one phosphorylation and that S_2 and S_5 are the predominant phosphorylation sites in vivo[48-50].

Beyond its role in factor recruitment, CTD phosphorylation has also immediate effects on its structure. While the CTD assumes a rather compact conformation in the unphosphorylated state, its hydrodynamic radius increases upon phosphorylation resulting in a reduced electrophoretic mobility and elution volume in size-exclusion chromatography[51]. In agreement, this structural extension upon CTD phosphorylation leads also to an increased protease sensitivity[52].

1.1.2 The Pol II transcription cycle

1.1.2.1 Transcription initiation

In order to allow transcription initiation to take place, Pol II needs to obtain access to promoter DNA at the transcription start site of a gene. In vivo, DNA is compacted through binding to octameric histone complexes that act as spools[53]. Each histone octamer is composed of two copies of the histones H2A, H2B, H3, and H4, which together with 147 bp of encircling DNA form the nucleosome[53], the fundamental building block of chromatin. The accessibility of the DNA template is strongly impaired by chromatinization[54, 55], but can be regulated through chromatin remodelers that can slide or eject nucleosomes from the DNA[56, 57]. Promoters of active transcription units are generally located in nucleosome-depleted regions that provide access for the transcriptional machinery to the DNA. Transcription initiation involves the assembly of the pre-initiation complex (PIC) encompassing the general transcription factors (GTFs) TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH together with Pol II at the core promoter[58, 59] (Fig. 1.1). GTFs facilitate transcription initiation by positioning Pol II on the

promoter and enabling transcription start site (TSS) selection[17, 59]. TFIIH plays a key role in this process as it opens the double-stranded promoter DNA at the TSS through its ATP-dependent translocase activity ('open PIC'). This allows the translocation of the template strand into the Pol II active site, and promotes the polymerization of a complementary RNA strand via a conserved catalytic mechanism ('initially transcribing complex')[60-62]. Once the transcript exceeds a critical length (12-13 nt), the growing RNA chain clashes with TFIIB, strongly destabilizing the PIC[63-65]. Concomitantly, the trimeric TFIIH kinase module containing the CDK7 kinase (Kin28 in yeast) phosphorylates the Pol II CTD at the heptad positions S_5 and S_7, which further facilitates PIC disassembly and promoter escape[37, 66-68].

1.1.2.2 Transcription elongation

Upon promoter escape, initiation factors disassemble and a processive transcription elongation complex forms, which becomes increasingly stabilized by the growing DNA-RNA hybrid[69]. S_5-phosphorylated CTD recruits the capping enzyme that modifies the 5'-end of the nascent mRNA with a stabilizing methylated guanosine nucleotide[70-72]. In metazoans, the Pol II elongation complex temporarily pauses in the promoter-proximal region ~50 nt downstream of the TSS, representing a regulatory checkpoint for transcriptional control during elongation[73] (for details see Section 1.1.3.2). Pol II pausing is stabilized by the negative elongation factor (NELF) and DRB sensitivity-inducing factor (DSIF)[74]. Pause release by the CDK9 kinase subunit of the positive elongation factor b (P-TEFb) results in displacement of NELF by the elongation factor complex PAF, binding of the elongation factor SPT6, and S_2-phosphorylation of the Pol II CTD[22, 75-77]. While pause sites are generally located upstream of the first (+1) nucleosome[78, 79], further transcription of the gene body necessitates Pol II passage through nucleosomes[80, 81]. Recruitment of positive elongation factors like – among others[82-84] – the histone chaperone SPT6[83, 85], or the H3K36 histone methyltransferase SET2 to the S_2/S_5-hyperphosphorylated CTD[86] can enable efficient nucleosome passage[80, 87] to facilitate elongation velocities of up to ~60 bp/s[88, 89]. As transcription elongation proceeds through the gene body, the phosphorylated CTD coordinates co-transcriptional pre-mRNA maturation through the direct interaction with components of the splicing apparatus[90-93].

1.1.2.3 Transcription termination and recycling

Close to the termination site, the predominantly S_2-phophosphorylated CTD mediates interactions with the 3-end RNA processing machinery[42, 94, 95]. Concomitantly, other components of the 3'-end processing machinery bind to the emerging well-defined RNA cleavage and polyadenylation signal (5'-UUAUUU-3') in the nascent RNA and trigger transcript cleavage[96]. Release of the RNA reduces the stability of the Pol II elongation complex, ultimately resulting in termination of Pol II transcription (Fig. 1.1). Two non-exclusive models have been proposed for eukaryotic transcription termination[97, 98]: In the torpedo model, an exonuclease that degrades the unprotected 5'-end of the cleaved transcript displaces Pol II from the template strand[98-100]. In the allosteric model, transcription termination results from the indirect destabilization of the Pol II elongation complex through binding of 3'-end processing factors, transcript cleavage and sequences that induce Pol II pausing[101, 102]. After release from the template, CTD dephosphorylation renders Pol II competent for re-initiation[103, 104].

1.1.3 Regulation of the Pol II transcription cycle

The core transcriptional machinery is functionally well conserved across eukaryotes, although the mechanisms that underlie transcriptional regulation differ substantially between yeast and human[105]. While Pol II transcription in yeast is mostly regulated at the level of transcription initiation (Fig. 1.3a), elaborate mechanisms of elongation control have additionally evolved in metazoan organisms (Fig. 1.3b)[106].

1.1.3.1 Transcriptional activation through Pol II recruitment

Transcription initiation is largely regulated through the recruitment of the transcriptional machinery to gene promoters[107] (Fig. 1.3a). While most of the GTFs represent the minimal set that is sufficient to reconstitute transcription in vitro[108], in vivo gene transcription necessitates additional factors[109].

Transcriptional factors that can recruit the transcriptional machinery are essential for robust gene activation[110, 111]. Such transcriptional activators possess intrinsically disordered transactivation domains that can physically interact with Pol II and other transcriptional

coactivators to stimulate transcription initiation[112]. Transcription factors also contain DNA-binding domains, with which they bind in a sequence-specific manner to regulatory DNA elements that can be either located proximal or distal to the core promoter[113]. In yeast, short upstream activating sequences adjacent to the promoter harbor few closely spaced transcription factor-binding sites[114, 115]. Besides having similar proximal regulatory sequences, metazoans possess additional extended distal enhancer elements that can be localized spatially uncoupled, upstream or downstream from the target promoter (often >100 kb apart)[116-118], contain multiple transcription factor binding sites[118, 119], and can contact the core promoter through gene looping[116, 120, 121]. As a consequence, enhancer-promoter interactions are not mutually exclusive and multiple enhancers can activate transcription on a single promoter or single enhancers can co-activate transcription on multiple promoters simultaneously[122-125]. The frequency of enhancer-promoter contacts determines the transcriptional output[122, 126] and is highly controlled by the local genome organization within topologically-associated domains[116, 127]. For enhancer activity, transcriptional coactivators are required that interact with transcription factors and provide a functional link to translate activator binding into transcriptional activity[128]. Some coactivators possess chromatin-remodeling or histone-modifying activities that increase promoter accessibility. Others, such as the multi-subunit co-activator Mediator[129], additionally transiently interact with the gene promoter, serving as a functional and architectural bridge[130-132]. Association of Pol II with the PIC requires the disordered CTD in an unphosphorylated state[133]. CTD truncation in yeast[29] and CTD deletion in human cells[32] strongly diminished transcription at activated gene promoters, suggesting that the CTD is required for Pol II recruitment in vivo. Weak interactions of the CTD with transactivation domains of transcription factors[134, 135], mediator subunits[36, 136], other transcriptional co-activators[137], and general transcription factors[138] are reported. The transient nature and ill-defined stoichiometry of multi-protein assemblies that underlie gene activation have, however, hampered any thorough functional and structural characterization[139, 140].

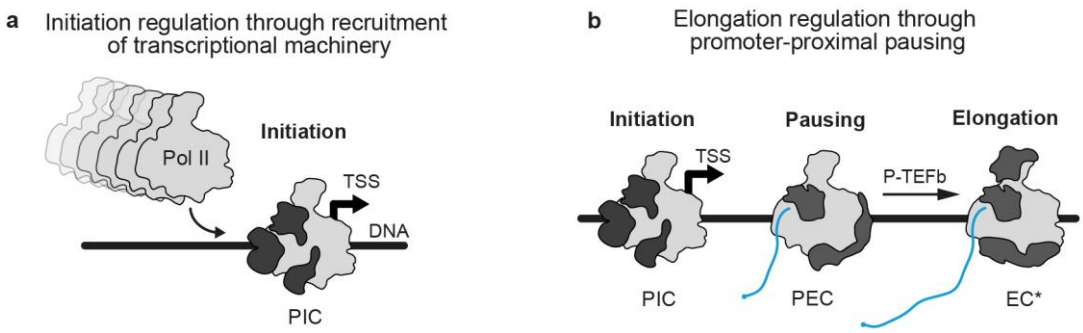

Figure 1.3 | Mechanisms of transcriptional regulation in higher eukaryotes.

a, Initiation regulation. Transcription initiation is largely regulated through the recruitment of the transcriptional machinery to gene promoters. Recruitment of the transcriptional machinery is facilitated by cis-regulatory elements such as enhancers. **b,** Elongation regulation. After transcription initiation, Pol II pauses 30-80 bp downstream of the transcription start site (TSS) through binding of negative elongation factor (NELF) and DRB sensitivity inducing factor (DSIF), which stabilize a tilted DNA-RNA hybrid conformation within the active site of Pol II. Pause release requires the positive elongation factor b (P-TEFb), which phosphorylates NELF, DISF and the Pol II CTD (omitted in the scheme for clarity). P-TEFb phosphorylation converts DSIF into a positive elongation factor, leads to the replacement of NELF by the PAF complex and binding of the histone chaperone SPT6, which facilitates Pol II transcription through chromatin. Figure is adapted from Hantsche & Cramer (2016)[17].

1.1.3.2 Promoter-proximal pausing and the role of negative elongation factor (NELF)

Early studies of transcriptional regulation in the model organism *S. cerevisiae* led to the notion that gene transcription is predominantly regulated through Pol II recruitment at the level of initiation[107]. However, the study of metazoan model systems revealed an additional previously unappreciated level of transcriptional regulation during early elongation, after Pol II commences transcription. Promoter-proximal pausing of Pol II was initially discovered at the HSP70 locus in the fruit fly *Drosophila melanogaster*[141, 142] (Fig. 1.3b). However, with the advent of approaches that map transcriptionally-engaged Pol II genome-wide, promoter-proximal pausing was readily appreciated as a widespread phenomenon and integral step of the transcription cycle in metazoan organisms[78, 88, 143, 144] (reviewed in[73]).

Many of the molecular determinants that lead to promoter-proximal pausing have been uncovered in the recent years, although their individual contribution is still part of on-going research. Pause sites appear to be localized in GC-rich regions flanked by an AT-rich

downstream sequence[89, 145, 146]. It was thus suggested that the high stability of the DNA-RNA hybrid reduces Pol II elongation rate and processivity[73, 145, 146]. Stable pausing, however, additionally requires binding of DSIF and NELF to the Pol II elongation complex[147, 148]. The emerging nascent transcript promotes the association of DSIF[149], a heterodimer composed of the conserved elongation factor SPT5 and the eukaryote-specific subunit SPT4[150], through contacts with RNA and the Pol II core[151, 152]. Similar to the Pol II CTD, SPT5 contains a C-terminal repeat (CTR) domain that contributes to factor recruitment during transcription elongation[153] (such as the PAF1 complex in yeast[154]). Binding of DSIF might then aid the recruitment of NELF[147, 152]. The metazoan-specific NELF complex is composed of four subunits[74], NELFA, NELFB, NELFC (or the nine amino acid shorter isoform NELFD) and NELFE, and is considered an essential facilitator of Pol II pausing[74, 76, 147]. While the DSIF subunit SPT5 underlies strong evolutionary conservation from bacteria to human, no NELF orthologs are present in yeast, nematodes, and plants, consistent with the absence of promoter-proximal pausing in these organisms[147, 155]. NELFA and –C as well as NELFB and –E form heterodimeric subcomplexes[156] that associate and form a three-lobed structure[76]. In addition, the NELFA and NELFE subunits possess large disordered C-terminal regions that were termed ‘tentacles’ due to their inherent flexibility[76]. Recent structural analysis of paused Pol II revealed that binding of NELF and DSIF allosterically stabilizes a tilted DNA-RNA hybrid conformation within the active site[76]. The tilted conformation of the DNA-RNA hybrid represents a non-productive state for nucleotide addition, as canonical base pairing of the next incoming nucleotide triphosphate with the template DNA is impaired[76]. Consequently, further extension of the pre-mRNA chain cannot occur, resulting in Pol II stalling. NELF additionally contacts the Pol II trigger loop[76], a mobile domain that facilitates nucleotide selection and catalysis[157], and restricts Pol II movement required to escape the paused state[76, 158].

The half-life of paused Pol II is often in the order of minutes, but can greatly vary between different transcription units[89, 159, 160] and under different developmental[78, 161] and environmental[162, 163] conditions.

Pause release requires the positive elongation factor b (P-TEFb) containing the CDK9 kinase together with a T-type cyclin, mainly T1[164-166]. P-TEFb triggers pause release through extensive phosphorylation of DSIF, NELF, the Pol II CTD and the CTD linker[75, 153, 167]. NELF becomes phosphorylated at numerous sites within the mobile tentacle regions, in particular at the NELFA

tentacle which contributes to pause stabilization[76, 87]. P-TEFb phosphorylation weakens the interaction of NELF with Pol II and promotes NELF displacement by the competitively binding elongation factor complex PAF[87]. Additionally, phosphorylation of the CTR helps to overcome the repressive effect of DSIF on transcription elongation and CTD linker phosphorylation results in recruitment of the histone chaperone SPT6. Thus, pause release involves the exchange of pausing factors by elongation factors (such as histone chaperones and remodelers) that enable Pol II transcription through chromatin (Fig. 1.3b).

While active in its free form, P-TEFb becomes inactivated by incorporation into the 7SK small nuclear ribonucleoprotein particle (snRNP)[168, 169]. P-TEFb sequestration involves the interaction with HEXIM1/2 (hexamethylene bisacetamide-inducible protein 1/2) and the non-coding 7SK RNA together with the stabilizing factors LARP7 (La-related protein 7) and MEPCE (methylphosphate capping enzyme)[170, 171]. Although the exact molecular mechanisms remain elusive, cells can tightly control P-TEFb activity by tuning the equilibrium between both forms in response to stimuli[168, 169, 172].

1.1.3.3 Transcriptional regulation in response to heat shock

In order to survive and thrive, cells need to rapidly sense and adapt to the ever-changing environmental conditions. Such adaptation involves an extensive gene-specific regulation of the transcription cycle. During heat stress, eukaryotic cells mount a rapid and conserved genome-wide response that involves the coordinated redistribution of the transcriptional machinery[173, 174]. It results in the upregulation of hundreds of genes, which encode pro-survival factors such as heat shock proteins (HSPs) and chaperones, and simultaneous downregulation of thousands of genes involved in anabolic processes[162, 175, 176].

Transcriptional upregulation upon heat shock in eukaryotic cells is mainly mediated through heat shock factor 1 (HSF1) (reviewed in[177, 178]). HSF1 binds as master regulator to conserved sequence elements within heat shock-activated gene promoters[179-181] and triggers gene activation by recruitment of the transcriptional machinery[182, 183] (Fig. 1.3a). The rapid transcriptional response at HSF1 target genes is facilitated through engaged, but transcriptionally paused Pol II, which maintains an open chromatin structure at the promoter[142, 184]. Binding of HSF1 triggers the successive accumulation of Pol II[185], positive

elongation factors such as P-TEFb[182] or SPT6[186], and chromatin remodelers[187] that can undergo local recycling to mediate efficient gene activation[188] (see also Section 1.2).

Transcriptional downregulation of genes involved in metabolism, protein synthesis and cell cycle is the prevalent consequence of heat stress, and by far outnumbers the upregulated genes[162, 175, 189]. It is accompanied by an enhanced recruitment of negative elongation factors such as NELF to chromatin that accumulate near repressed gene promoters[175]. Consistently, paused Pol II becomes stabilized within the promoter-proximal region of these genes upon stress, resulting in increased pause duration[162, 163, 189] (Fig. 1.3b). Since the presence of paused Pol II prevents new transcription initiation, enhanced Pol II pausing can facilitate swift transcriptional repression[89, 190]. But at the same time, it might keep the transcriptional machinery in a competent state that allows rapid reactivation after the heat stress ceases[176]. In contrast to stress-induced activation, the molecular mechanisms that cause genome-wide transcriptional downregulation upon heat shock are far less well understood[175].

1.2 Spatiotemporal organization of Pol II transcription

Each human diploid cell contains 23 pairs of chromosomes that encompass together about six billion base pairs DNA with a total length of ~2 m. To accommodate the genetic information in the cell nucleus that is about five orders of magnitude smaller, the DNA is highly packaged at multiple levels. This degree of compaction is equivalent to accommodating a DNA strand encircling the earth for >6000 times inside a chicken egg[191]. Given this highly crowded nuclear environment the question arises how the manifold factors involved in Pol II transcription can efficiently encounter each other in a spatiotemporally controlled manner. In the middle of the 1990s this puzzling question was first addressed in pioneering studies by Peter Cook and colleagues who observed that Bromo-UTP labelled nascent transcripts in fixed human cells were not evenly distributed throughout the entire nucleus, but localized to distinct focal sites that they termed ‘transcription factories’[192]. Several follow-up studies reported similar observations using different nucleotide analogs and electron microscopy, and detected the co-localization of Pol II with labelled nascent transcripts in foci[193-195]. About 2400 of such foci were detected, each estimated to contain on average about 30 engaged polymerases[195]. Using similar techniques, even distant genes spaced several megabases apart were observed to

colocalize with Pol II foci, which in turn colocalized with fluorescence in situ hybridization signals from the produced transcripts[196]. Correspondingly, these and other[197-199] early studies led to the concept that stable pre-assembled transcription factories, dedicated nuclear sites for RNA synthesis with high concentrations of transcriptional components such as Pol II, exist, to which genes must translocate in order to become transcribed[200-203] (Fig. 1.4a). However, it was argued at the same time that results obtained in these studies could have been affected by intrinsic methodological limitations. For example, it was criticized that chemical cell fixation might have introduced artificial aggregation artefacts[204]. Similarly, the number of Pol II molecules might be overestimated through indirect Pol II immunolabeling with antibodies targeting the repetitive CTD as multiple antibody molecules bind a single Pol II enzyme[205]. Importantly, the dynamics of (dis-)assembly of the detected transcription factories could not be explored due to cell fixation, impeding conclusions regarding their stability. Subsequent attempts to detect stable clusters of transcriptionally active Pol II in living mammalian cells using GFP-tagged Pol II and confocal microscopy were not successful[206, 207]. Rather, initiating and elongating forms of Pol II were observed to possess a distinct but adjacent nuclear localization[204]. More recent single-molecule super-resolution microscopy approaches suggest that the majority of Pol II molecules are solitary and spaced on average >200 nm away from each other[205], arguing against the predominant occurrence of Pol II in large stable transcription factories.

Using an elegant super-resolution microscopy approach that focuses on transiently (~50 ms) immobile Pol II molecules, Cisse et al. (2013) showed that a small fraction of Pol II molecules indeed forms transient clusters in live human cells[208]. For these experiments, the authors used a stable human cell line encoding RPB1 that was N-terminally tagged with the photo-switchable fluorescent protein Dendra2. Successive cycles of photoactivation and localization allowed time-resolved counting of detections used then for pair-correlation analysis[208-210]. Interestingly, the detected Pol II clusters possessed highly transient lifetimes of only few seconds (5.1 ± 0.4 s in[208], 8.3 ± 0.2 s in[211], and 12.9 ± 1.4 s in[212]) and average sizes below the diffraction limit, representing potential reasons why they could not be detected in previous studies. As estimated in fixed cells, an average cluster contains ~80 Pol II molecules[211]. Live-cell two-color imaging of Pol II and mRNA produced from the β-actin locus revealed that transient Pol II clustering precedes mRNA synthesis[211], consistent with a notable stabilization of cluster

lifetimes after inhibition of transcription elongation[208, 211]. Together, these findings provided compelling evidence that small populations of Pol II transiently form high local concentrations in close proximity to gene promoters prior to transcription initiation (Fig. 1.4b).

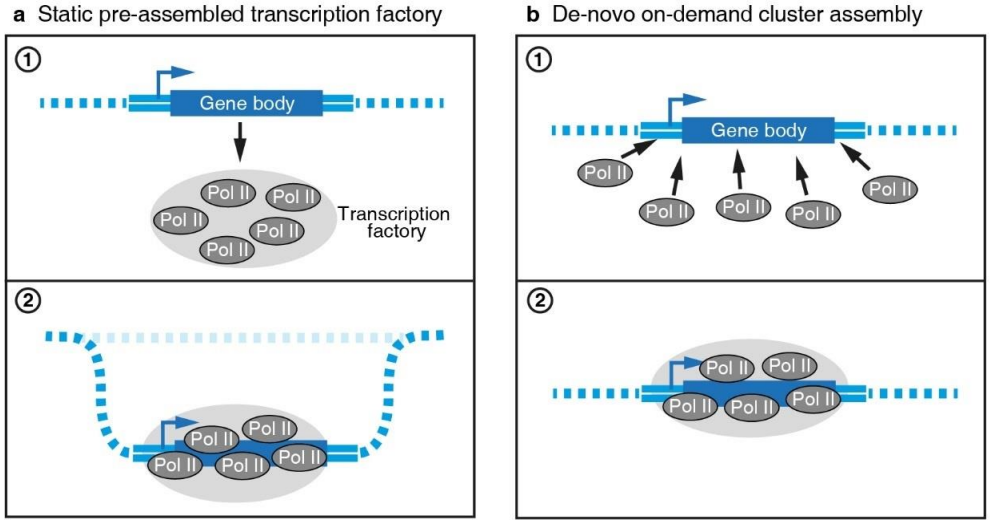

Figure 1.4 | Models for the spatiotemporal organization of gene transcription.

a, Gene transcription requires the translocation into static pre-assembled transcription factories containing high concentrations of relevant factors (i.e. RNA Pol II). **b**, Nucleoplasmic pool of Pol II surrounds the gene and dynamically forms high concentration clusters upon transcriptional activation. Figure concept was adapted from Buckley & Lis (2016)[200].

The rapid Pol II clustering kinetics also match residence times observed for several transcription factors (TFs) on their target sites remarkably well: FRAP and recent single-molecule tracking experiments showed that the large majority of TF molecules occupy fast-diffusing states[213-216] and that just a small percentage of molecules is bound at specific target sites. At the same time each binding event persists for only few seconds[216]. Consistent with the kinetics of Pol II clustering, the coactivator complex Mediator also forms transient clusters at enhancer elements with average lifetimes of 11.1 ± 0.9 s[212]. In agreement with the transient assembly/disassembly of Pol II and co-activator clusters, recent analysis of transcription in single-cells revealed that transcription initiation is not a constant continuous process, but occurs in short 'bursts' followed by long periods of transcriptional inactivity[126, 217-220]. Transcriptional

bursts generate convoys of closely spaced Pol IIs, which transcribe the gene body[220]. Bursts are triggered when enhancer elements come in close proximity to gene promoters through DNA looping[116, 117, 120, 121].

Cellular stress such as heat shock causes the dynamic nuclear redistribution of the Pol II machinery. The heat shock response has been extensively studied on *Drosophila* polytene chromosomes, where heat shock stress causes local chromatin decondensation at transcriptionally active loci called puffs[221]. Because of the naturally amplified HSP70 gene cluster at polytene chromosomes it is possible to image transcriptional activation at high signal-to-noise using diffraction-limited fluorescence microscopy[203, 222]. Heat stress-induced transcriptional activation caused the sequential accumulation of heat shock factor HSF1, Pol II and other positive transcription elongation factors (i.e. P-TEFb, SPT6, and chromatin remodelers) at the HSP70 locus[185, 188, 223]. Prolonged gene activation resulted in sustained recruitment of Pol II and elongation factors beyond the amount that can bind to the transcription unit and the ADP-ribosylation-dependent compartmentalization of the locus[185] that facilitated the local recycling of these factors over the time of activation[185, 188].

Taken together, these insights into the spatiotemporal organization of transcription in living cells suggests a very dynamic regulation involving transient high local concentrations of Pol II and relevant cofactors during gene activation in steady state and upon stress. While these studies suggest the functional importance of transient macromolecular assemblies encompassing Pol II, a mechanistic understanding of the molecular principles that govern such factor concentration only begins to emerge.

1.3 Intrinsic disorder in Pol II transcription

Which mechanisms might underlie the formation of such transient macromolecular assemblies? Proteome-wide computational analyses revealed that factors involved in eukaryotic gene transcription contain a high proportion of intrinsically disordered protein regions (IDRs) (Fig. 1.5). Such protein domains exist as a heterogeneous ensemble of rapidly interconverting conformations[34]. Because IDRs do not fold into stable three-dimensional structures, they are generally devoid of hydrophobic amino acids that drive the higher-order folding of proteins. Rather, they are often enriched for polar amino acids (in particular glycine (G), serine (S),

threonine (T), and glutamine (N)) and charged amino acids (arginine (R), lysine (K), aspartic acid (D), and glutamic acid (E))[224, 225]. In addition, disordered regions often appear to contain interspersed aromatic amino acids (in particular tyrosine (Y) and phenylalanine (F)) and have sometimes a high proline (P) content[224, 225]. Since the biased amino acid composition and the strong overrepresentation of certain amino acids is often indicative of disorder, many IDRs are referred to as low-complexity domains[226].

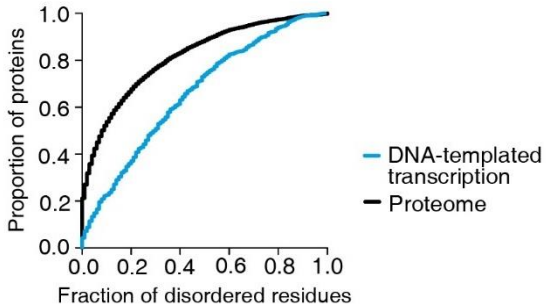

Figure 1.5 | Intrinsic disorder in gene transcription.

The cumulative proportion of proteins is shown in dependence on the fraction of amino acids in disordered protein regions for the entire proteome (black line) or proteins involved in gene transcription (blue line; GO:0006351, DNA-templated transcription) for *S. cerevisiae*. While only 32% of proteins in the entire proteome (n =8610) possess more than 20% of the residues in disordered regions, over 63% of the proteins involved in gene transcription (n=528) possess more than 20% disorder. Figure was plotted based on the data presented in Herzel et al. (2017)[90].

From early on, the importance of disordered regions was recognized for transcriptional activation[140]. Transcription factors often possess extended disordered transactivation domains that promote transcription initiation through ill-defined intermolecular interactions and were thus referred to as 'negative noodles'[140]. The human proteome comprises over 1,600 transcription factors[113], with many of them containing disordered transactivation domains[227]. Pol II itself contains an extended disordered C-terminal domain, which comprises over 350 residues with a molecular mass of ~40 kDa in mammals. Being composed almost exclusively of the four amino acids Y, S, P, and T, the CTD is certainly one of the most prominent low-complexity sequences in the proteome. Apart from Pol II itself, several other factors that

regulate Pol II progression through the transcription cycle possess substantial disorder[46, 228, 229, 230]. One such example is the pausing factor NELF that contains two large unstructured 'tentacle regions', which are in part required for stabilization of Pol II pausing in vitro[76].

1.4 Liquid-liquid phase separation

The emerging concept of liquid-liquid phase separation can provide a mechanistic basis how intrinsically disordered regions can mediate the local concentration of proteins in so-called membraneless organelles or biomolecular condensates[225, 231]. The underlying concepts are based on fundamental physical properties of polymers[232, 233]: Molecules are soluble in solution until their concentration reaches a solubility threshold. At concentrations above the solubility limit, some of the molecules cannot remain dissolved in solution and distribute into a distinct separate phase[232, 233]. Hyman and colleagues were the first to recognize that identical principles also apply to biological polymers such as proteins in aqueous solutions, which then can give rise to two liquid phases with different properties[234]. Such liquid-liquid phase separation of proteins has subsequently emerged as a fundamental principle of intracellular organization in the absence of bounding membranes[225, 231].

1.4.1 Physical basis of liquid-liquid phase separation

Liquid-liquid phase separation (LLPS) is a concentration-dependent process and results above the polymer saturation concentration c_{sat} in the demixing of a homogeneous solution into a condensed, polymer-rich phase with liquid-like properties that co-exists with a dilute (polymer-poor) phase (Fig. 1.6a)[225, 231, 235-237]. Phase separation results from governing thermodynamic principles by which a system strives to reach the lowest energy state possible[237]. On a molecular level, different polymers possess varying tendencies to interact with solvent molecules as well as with other polymer molecules. Such interactions between biological polymer molecules are typically low-affine but multivalent[231, 235, 238]. The solubility of a polymer results from the balance between polymer-solvent and polymer-polymer interactions. If polymer-solvent interactions are stronger than the tendency of polymer molecules to interact with each other, then the polymer molecules remain soluble in solution, independent of their concentration[231]. Under these conditions, the polymer molecules will distribute uniformly within the solution to

maximize the entropy of the system. If, on the other hand, polymer-polymer interactions are stronger than polymer-solvent interactions, polymer solubility decreases and the propensity of phase separation increases. Such a system can undergo phase separation when favorable polymer-polymer interactions become more and more likely with increasing concentration until their energetic contribution overcomes the entropic tendency of the system to stay uniformly mixed. Under these conditions phase separation into two phases is thermodynamically favorable as the sum of the free energy of both phases is smaller than the free energy of the single (mixed) phase (Fig 1.6b). Since in the two-phase regime polymer-polymer and solvent-solvent interactions are energetically more favorable, the condensed phase assumes a droplet-like structure to minimize unfavorable polymer-solvent interactions. The formation of such droplets can be analyzed using differential contrast interference or fluorescence microscopy methods[239]. Importantly, liquid-like droplets possess similar characteristics as known from ideal liquids and can coalesce and fuse (Fig 1.6c). Since the nature of the interactions within the condensed phase is weak, molecules can diffuse dynamically and are in constant exchange with the dilute phase (Fig 1.6d).

On the basis of the previous considerations, it follows that the breadth of interaction between polymer molecules crucially determines whether a polymer undergoes phase separation at a given concentration. In this regard, Rosen and colleagues were the first to provide experimental evidence that multivalency – the ability of a single molecule to engage in interactions with multiple other molecules – is a key parameter that promotes phase separation[240]. IDRs that do not fold into well-defined three-dimensional structures but possess conformational heterogeneity can provide the underlying basis for such multivalent intermolecular interactions. While it is not well understood on a molecular level how IDRs can promote LLPS of the protein they are attached to[241], a ‘stickers and spacers model’ that was developed from polymer theory[242, 243] by Pappu and colleagues[244, 245] has proven helpful in this regard. The model predicts the existence of distributed associative motifs called ‘stickers’ that can promote LLPS through intermolecular sticker-sticker interactions and are interspersed by inert ‘spacers’ (Fig. 1.7a). Sticker motifs appear to interact through three key types of molecular interactions which are combined/arranged in various patterns[224, 225, 246]: Electrostatic interactions between blocks of oppositely charged amino acids[247-251], cation-π interactions between basic amino acids and predominantly aromatic π electron systems[245, 252], and π-π interactions especially between

aromatic groups[134, 135, 253] (Fig. 1.7b). Interactions between polar amino acids are generally considered to have only minimal direct effect on phase separation. Instead, polar amino acids often appear to occupy spacer regions that enhance solubility and determine flexibility of sticker motifs[224, 244, 245] (Fig. 1.7a). Aliphatic amino acids are generally depleted in IDRs, yet proline constitutes a notable exception: Although proline-rich domains were reported to decrease the saturation concentration for LLPS in a few cases[254-256], an important reason for the overrepresentation of proline might be its ability to suppress the formation of regularly structured elements together with the accompanying increase in the conformational flexibility of the peptide backbone[257, 258].

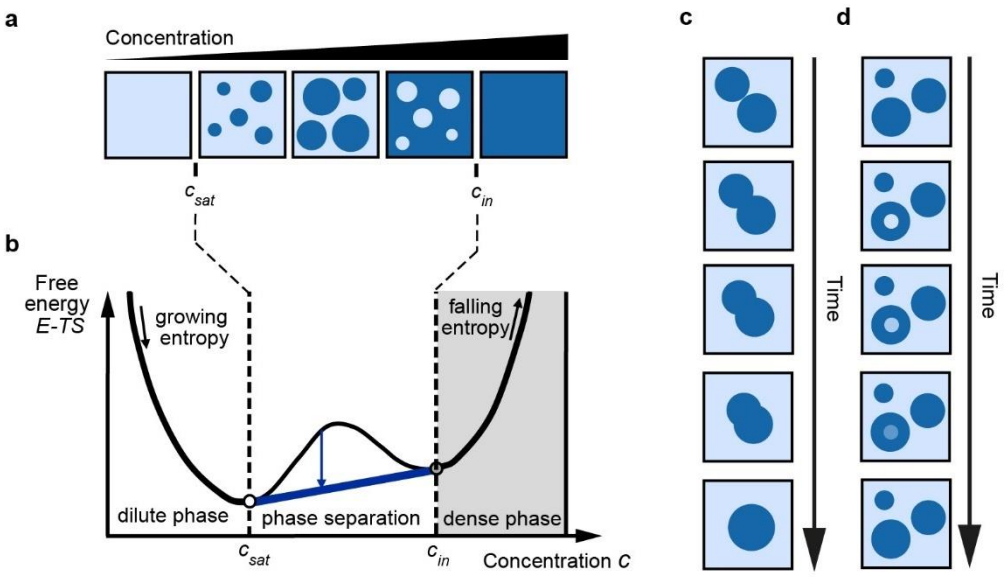

Figure 1.6 | Thermodynamic basis of liquid-liquid phase separation and emerging properties.
a, Biological polymers such as proteins can undergo liquid-liquid phase separation in aqueous solutions if their concentration exceeds the solubility limit (c_{sat}), also sometimes referred to as the critical concentration. Above c_{sat}, the solution demixes into two liquid phases: A condensed polymer-rich phase co-exists with a dilute phase with the concentration c_{sat}. At a polymer concentration that is equal or higher to the concentration inside the droplets (c_{in}), the system returns to the one-phase regime. **b,** Free energy as a function of the protein concentration. Above the saturation concentration (c_{sat}), polymer-polymer and solvent-solvent interactions are more favorable than polymer-solvent interactions and drive phase separation. Demixing happens when the sum of the free energy of both phases (head of blue arrow) is smaller than the free energy of the mixed phase (base of blue arrow). The figure was kindly provided by Dr. Johannes Soeding (MPI for Biophysical Chemistry, Göttingen) and adapted based on Soeding et al. (2019)[236]. **c,** Droplet fusion. Droplets formed by liquid-liquid phase separation have properties known from ideal liquids and coalesce upon contact. **d,** Condensed polymers in liquid-like droplets are in constant exchange with the surrounding solution. Photobleaching of material in the condensed phase using fluorescence recovery after photobleaching (FRAP) results in rapid recovery of fluorescence.

1.4.2 Regulation of phase separation

The ability of cells to inducibly and rapidly alter the chemical properties of amino acids through post-translational modifications (PTMs) and thereby modulate interactions between IDRs has emerged as a key mechanism to regulate phase separation[236, 259, 260]. PTMs can affect phase separation in various ways through the modulation of sticker-sticker interactions. For example, direct post-translational modification of sticker motifs can prevent IDR-IDR interactions with other sticker motifs[251, 261, 262]. In other cases, PTMs can also create new sticker motifs resulting in novel interactions that promote phase separation[247, 263, 264]. Similarly, the ability of different sticker motifs to interact can be sterically influenced through post-translational modifications of spacer sequences[265]. Phosphorylation of serine, threonine, and tyrosine residues is the most common PTM in eukaryotes targeting about half of all human proteins[266]. Interestingly, most of the phosphorylation sites detected so far are localized in disordered regions[267, 268]. It comes thus with no surprise that phosphorylation has been implicated in the regulation of various biomolecular condensates, both positively and negatively[240, 247, 265, 269-271] (for overview see[236, 260] and references therein). Nevertheless, other PTMs such as methylation[251, 261, 262, 272], sumoylation[263, 273] or ADP-ribosylation[264, 274] have also been shown to regulate phase separation.

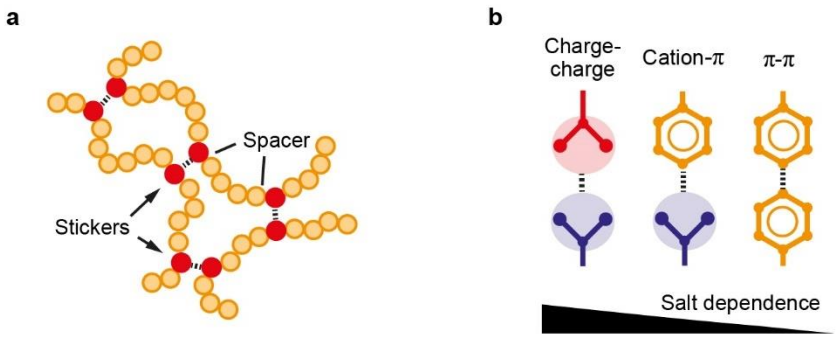

Figure 1.7 | Molecular basis for liquid-phase separation of IDRs.
a, Concept of the 'stickers and spacers' model[244, 245]. Stickers engage in intermolecular interactions with other stickers, while spacers are inert and form no such interactions. Formation of noncovalent intermolecular sticker-sticker interactions gives rise to crosslinked network and promotes LLPS. In the context of IDRs, stickers correspond to single amino acids or short motifs, but note that the concept can be further extended to entire folded or disordered domains[224, 244, 245]. **b,** Molecular nature of sticker-sticker interactions between IDRs. Charge-charge, cation-π, or π-π interactions appear to be the prevalent driving force for LLPS. Note that LLPS of IDRs might not strictly depend on a single type of interaction but combinations thereof. Figure is adapted based on Brangwynne et al. (2015)[246].

1.5 Aims and scope of this work

1.5.1 RNA polymerase II clustering through CTD phase separation

Eukaryotic gene transcription is highly regulated at the step of initiation, which involves the recruitment of the transcriptional machinery to the gene promoter[107]. Prior to transcription initiation, Pol II enzymes cluster in short-lived nuclear 'hubs' near active gene promoters[208, 211]. However, the molecular mechanisms that may underlie the dynamic formation and disassembly of Pol II clusters have remained enigmatic. Liquid-liquid phase separation, which is based on weak multivalent interactions between disordered protein domains, constitutes a fundamental mechanism to concentrate proteins inside cells[225, 231, 235, 275]. Pol II possesses a disordered C-terminal repeat domain that is essential for factor recruitment during the transcription cycle[21, 22, 276]. In the unphosphorylated form, the Pol II CTD assumes a compact, globular conformation[13, 36-38], indicating intramolecular interactions between different CTD repeats. Given the existence of weak repeat-repeat interactions and the intrinsic flexibility of the CTD structure, the question arises whether CTD molecules might be able to engage in multivalent intermolecular interactions that could lead to liquid-liquid phase separation and whether such interactions might underlie Pol II clustering in vivo. Because the CTD becomes strongly phosphorylated throughout the transcription cycle, it is important to investigate how phosphorylation might affect these interactions. This work aimed to provide answers to these questions. First, homotypic CTD phase separation was successfully reconstituted in a simplified system in vitro. Results from these experiments demonstrated that the human CTD can undergo phase separation through weak multivalent intermolecular interactions. They allowed then to further probe the dependence of CTD length on LLPS by comparing the human with the half as long yeast CTD. The strong dependence of LLPS on CTD repeat number that became apparent from in vitro phase separation assays then served as the basis to investigate the effect of CTD length on Pol II clustering as well as its dynamics in live human cells. The subsequent microscopic analyses of different cell lines with varying CTD repeat number revealed that, similar to our in vitro results, CTD length also determines Pol II clustering and dynamics in human cells. Together the combined findings indicated that interactions of unphosphorylated CTDs underlie Pol II clustering near active gene promoters in living cells through a phase separation mechanism. Finally, the direct effect of initiation-coupled phosphorylation on phase separation of human and yeast CTD was studied with purified components in vitro.

The obtained results imply a model for gene activation that involves CTD-mediated clustering of initiation-competent Pol II and release through CTD phosphorylation upon transcription initiation.

1.5.2 NELF condensation accompanies stress-induced transcriptional downregulation

Promoter-proximal pausing of Pol II during early elongation constitutes a metazoan-specific regulatory checkpoint[73], which plays a key role during genome-wide heat stress-induced transcriptional downregulation[162, 163, 189]. In human cells, the enhanced recruitment of the pausing factor NELF to downregulated gene promoters[175] leads to the genome-wide stabilization of paused Pol II[162, 175, 176]. However, the underlying molecular basis of the increased residence time of NELF at chromatin upon stress is unclear. In Drosophila, heat stress-induced transcriptional activation of pro-survival genes results in the formation of a transcriptional compartment that facilitates local factor recycling and retardation[185, 188]. This work aimed to investigate whether an analogous phase separation mechanism could underlie the increased residence time of NELF at downregulated gene promoters, and to understand its heat stress-specific regulation. To attain this aim, it was first necessary to construct human cell lines with fluorescently labelled NELF and examine the nuclear redistribution upon heat shock. The results revealed that NELF clustered in numerous puncta upon heat shock that possessed properties consistent with phase-separated condensates. To enable further investigation of the underlying molecular mechanism, the four-subunit NELF complex was recombinantly produced to reconstitute phase separation using a simplified in vitro system. This allowed to identify the disordered tentacles as drivers of NELF self-association. The combination of in vivo and in vitro approaches then enabled the investigation of mechanisms that underlie the heat stress-specific regulation of NELF condensation. The experiments unveiled that heat stress-induced inactivation of the pause release factor P-TEFb counteracts phase separation through phosphorylation of disordered NELF regions. In addition, they revealed that stress-induced sumoylation contributes to NELF condensation. Taken together, these results describe the molecular basis for stress-induced NELF condensation and suggest a mechanistic framework that controls stress-induced genome-wide transcriptional downregulation.

2 Material and methods

2.1 Materials

2.1.1 *E. coli* strains

Table 1 | *E. coli* strains used in the study.

Strain	Genotype	Supplier
XL1-blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F' *proAB lacI*^{q} ZΔ*M15* Tn*10* (Tet^r)]	Agilent
5-alpha	*fhuA2*a(*argF-lacZ*)*U169 phoA glnV44 a80a* (*lacZ*) M15 *gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	NEB
BL21-CodonPlus (DE3) RIL	F$^-$ *ompT hsdS* ($r_B^-m_B^-$) dcm$^+$ Tet^r *gal* λ(DE3) *endA* Hte [*argU ileY leuW* Camr]	Agilent
BL21 Rosetta 2 (DE3) pLysS	F$^-$ *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* (DE3) pLysSRARE2 (Cam^R)	Novagen
DH10EMBacY	F- *mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80dlacZΔM15 ΔlacX74 endA1 recA1 deoR Δ(ara, leu)7697 araD139 galU galK λ- rpsL nupG / bMON14272‡ yfp+/ pMON7124*	Geneva Biotech

2.1.2 *S. cerevisiae* strains

Table 2 | *S. cerevisiae* strains used in the study.

Strain	Genotype	Source
BJ5464 RPB3-His$_6$-Bio	*BJ5464 MATa ura3-52 trp1 leu2-Δ1 his3-Δ200 pep4::HIS3 bprb1Δ1.6R can1 GAL;* 6×His-Bio-RPB3 *URA3*	M. Kashlev [277]

2.1.3 Insect cell lines

Table 3 | Insect cell lines used in the study.

Cell line	Species	Source	Supplier
Sf9	*Spodoptera frugiperda*	Immortalized ovarian cells (clonal derivative of Sf21 cell line)	Thermo Fisher Scientific
Sf21	*Spodoptera frugiperda*	Immortalized ovarian cells	Thermo Fisher Scientific
High Five (Hi5)	*Trichoplusia ni*	Immortalized ovarian cells	Expression Systems

2.1.4 Culture media

Table 4 | Growth media for *E. coli* and yeast cell cultures.

Medium	Organism	Composition	Supplier
LB	*E. coli*	1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl (1.5% (w/v) agar for plates)	homemade
2xYT	*E. coli*	1.6% (w/v) tryptone, 1% (w/v) yeast extract, 0.5% (w/v) NaCl	homemade
SOC	*E. coli*	2% (w/v) peptone, 0.5% (w/v) yeast Extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose	NEB
YPD	*S. cerevisiae*	2% (w/v) peptone, 2% (w/v) glucose, 1.5% (w/v) yeast extract (1.8% (w/v) agar for plates)	homemade

Table 5 | Media for insect cell culture.

Medium	Cell line	Supplier
Gibco® Sf-900 III SFM	*Spodoptera frugiperda* Sf9/Sf21	Thermo Fisher Scientific
ESF921	*Trichoplusia ni* Hi5	Expression Systems

2.1.5 Antibiotics and additives

Table 6 | Antibiotics and other additives.

Substance	Working concentration	Stock concentration	Solvent
Ampicillin	100 µg/mL	100 mg/mL	100% EtOH
Chloramphenicol	34 µg/mL	34 mg/mL	70% EtOH
Gentamycin	10 µg/mL	10 mg/mL	Ultrapure water
IPTG	0.5 mM	1 M	Ultrapure water
Kanamycin	50 µg/mL	50 mg/mL	Ultrapure water
Tetracycline	10 µg/mL	10 mg/mL	Ultrapure water
X-gal	150 µg/mL	150 mg/mL	DMSO

2.1.6 Expression plasmids

Table 7 | Bacterial expression plasmids used in the study. Detailed description of the constructs with the protein regions they contain for expression, the respective vector backbone and the cloning method used to generate them. The pET24d(+) plasmid was initially obtained from EMD Biosciences, the 1C vector was obtained from Addgene (#29654). 1-C* denotes a 1-C-derivative plasmid, in which the Asn_{10}-linker was replaced by a SGGS-linker. All plasmids possess a kanamycin-resistance cassette. NA, not applicable.

Name	Construct	Residues	Vector backbone	Cloning method	Source
MBP-hCTD	His_6-MBP-Asn_{10}-TEV-hCTD	*H. sapiens* RPB1 residues 1593-1970	1-C	NA	Seychelle Vos [87]
MBP-yCTD	His_6-MBP-Asn_{10}-TEV-yCTD	*S. cerevisiae* RPB1 residues 1542–1733	1-C	Gibson assembly	This study
MBP	His_6-MBP-Asn_{10}-TEV	None	1-C	Deletion mutagenesis	This study
MBP-hCTD YF	His_6-MBP-Asn_{10}-TEV-hCTD YF	*H. sapiens* RPB1 residues 1593-1970, tyrosine replaced by phenylalanine in every repeat	1-C	Restriction enzyme-based	This study
MBP-hCTD YL	His_6-MBP-Asn_{10}-TEV-hCTD YL	*H. sapiens* RPB1 residues 1593-1970, tyrosine replaced by leucine in every repeat	1-C	Restriction enzyme-based	This study
GST-yCTD	His_6-GST-TEV-yCTD	*S. cerevisiae* RPB1 residues 1542–1733	pET24d	Insertion mutagenesis	This study
GST	His_6-GST-TEV	None	pET24d	Deletion mutagenesis	This study
GFP-NELFA tentacle	His_6-mEGFP-TEV-NELFA IDR	*H. sapiens* NELFA residues 189-528	1-C*	Gibson assembly	This study
GFP-NELFE tentacle	His_6-mEGFP-TEV-NELFE IDR	*H. sapiens* NELFE residues 139-380	1-C*	Gibson assembly	This study
NELFE tentacle-GFP-NELFA tentacle	His_6-TEV-NELFE IDR-mEGFP-NELFA IDR	*H. sapiens* NELFE residues 139-380	1-C*	Gibson assembly	This study
SUMO2	His_6-TEV-SUMO2	*H. sapiens* SUMO2 full-length	1-C*	Restriction-enzyme based	This study

Table 8 | Insect cell expression plasmids used in this study. Detailed description of the constructs with the protein regions and tags they contain for expression as well as the respective vector backbone.

Name	Protein variant	Tag	Vector backbone	Source
TFIIH kinase module	*H. sapiens* CDK7 full-length *H. sapiens* CCNH full-length *H. sapiens* MAT1 full-length	His_6-TEV- His_6-TEV- His_6-TEV-	438-B	Goran Kokic [278]
NELF complex	*H. sapiens* NELFA full-length *H. sapiens* NELFB full-length *H. sapiens* NELFD full-length *H. sapiens* NELFE full-length	None None His_6-TEV- None	438-B	Seychelle Vos [76]
NELF complex ΔA tentacle	*H. sapiens* NELFA 1-188 *H. sapiens* NELFB full-length *H. sapiens* NELFD full-length *H. sapiens* NELFE full-length	None None His_6-TEV- None	438-B	Seychelle Vos [76]
NELF complex ΔE tentacle	*H. sapiens* NELFA full-length *H. sapiens* NELFB full-length *H. sapiens* NELFD full-length *H. sapiens* NELFE 1-138	None None His_6-TEV- None	438-B	Seychelle Vos [76]
P-TEFb WT	*H. sapiens* CDK9 1-372 *H. sapiens* CCNT1 1-272	His_8-TEV- GST-TEV-	pACEBac1	Matthias Geyer [76]
P-TEFb kinase-dead	*H. sapiens* CDK9 1-372 D149N *H. sapiens* CCNT1 1-272	His_8-TEV- GST-TEV-	pACEBac1	Matthias Geyer [76]

2.1.7 Common buffers and solutions

Table 9 | Common buffers and solutions used in this study.

Buffer	Composition
1x PBST	10 mM Na_2HPO_4 pH 7.4, 1.76 mM KH_2PO_4 pH 7.4, 137 mM NaCl, 2.7 mM KCl, 0.1% Tween-20
1x TAE running buffer	250 mM Tris-acetate, 5 mM EDTA pH 8.0 at 20°C
1x MES SDS running buffer	50 mM MES pH 7.3, 50 mM Tris base, 0.1% SDS, 1 mM EDTA
1x MOPS SDS running buffer	50 mM MOPS pH 7.7, 50 mM Tris base, 0.1% SDS, 1 mM EDTA
100x Protease inhibitor cocktail	0.028 mg/mL leupeptin, 0.137 mg/mL pepstatin A, 17 mg/mL PMSF, 33 mg/mL benzamidine in 100% EtOH

2.1.8 Kits and consumables

Table 10 | Kits and consumables used in this study.

Kit	Supplier
Qiaprep Spin Miniprep kit, Plasmid Plus Midi kit, Plasmid Plus Maxi kit	Qiagen
MinElute DNA extraction kit	Qiagen
MinElute PCR purification kit	Qiagen

2.1.9 Antibodies

Table 11 | Antibodies used in this study.

Antibody	Reactivity and description	Host	Clonality	Dilution	Supplier
Anti-Ser5P	Ser5-phosphorylated Pol II CTD, clone 3E8	Rat	Monoclonal	1:60	Dirk Eick [39]
Anti-MBP	Maltose-binding protein, HRP-coupled	Mouse	Monoclonal	1:1000	Abcam (ab49923)
Anti-GST	Glutathione S-transferase, HRP-coupled	Goat	Polyclonal	1:5000	GE healthcare (RPN1236)
Anti-NELFA	*H. sapiens* NELFA C-terminus (499-549)	Rabbit	Polyclonal	1:500	Bethyl Laboratories (A301-910A)
Anti-NELFB	*H. sapiens* NELFB C-terminus (550-580)	Rabbit	Polyclonal	1:500	Bethyl Laboratories (A301-912A)
Anti-NELFCD	*H. sapiens* NELFCD N-terminus (1-300)	Mouse	Monoclonal	1:500	Santa Cruz (sc-393972)
Anti-NELFE	*H. sapiens* NELFE N-terminus (75-125)	Rabbit	Polyclonal	1:1000	Bethyl Laboratories (A301-913A)
Anti-ZNF451	*H. sapiens* ZNF451 N-terminus	Rabbit	Polyclonal	1:200	Ray Biotech (102-18338)
Anti-rabbit IgG-HRP	Rabbit IgG, HRP-coupled secondary antibody	Donkey	Polyclonal	1:5000	GE healthcare (NA934)
Anti-mouse IgG-HRP	Mouse IgG, HRP-coupled secondary antibody	Goat	Polyclonal	1:3000	Abcam (ab5870)
Anti-rat IgG-HRP	Rat IgG, HRP-coupled secondary antibody	Goat	Polyclonal	1:5000	Sigma Aldrich (A9037)

2.2 Methods

2.2.1 General techniques in molecular biology

2.2.1.1 Quantitation of nucleic acids and proteins

DNA and protein concentrations were determined spectrophotometrically using a NanoDrop device (Peqlab). Double-stranded DNA concentrations were calculated based on the assumption that an absorbance at 260 nm of 1 corresponds to a concentration of 50 µg/µL. Protein concentrations were determined according to the Lambert-Beer law based on the absorbance at 280 nm and the theoretical molar extinction coefficient calculated with the software tool Protean (DNAStar Lasergene Suite, version 12.0).

2.2.1.2 Transformation of chemically competent *E. coli* cells

Chemically competent *E. coli* cells were thawed on ice for 10 min. For each transformation, 50 µL *E. coli* cell suspension were mixed with 1 µL plasmid DNA (50-100 ng) or up to 5 µL ligation/assembly mix (as indicated below) and incubated for 10-30 min on ice. The cells were then heated to 42 °C for 45 s, followed by a 2 min incubation on ice. Immediately afterwards 450 µL pre-warmed SOC recovery medium were added and the cells were incubated for 45-90 min at 37 °C. The recovered cells were then spread on LB agar plates containing appropriate antibiotics (see Table 6) and incubated overnight at 37 °C.

2.2.1.3 Polymerase chain reaction (PCR)

PCR reactions were performed using Phusion High-Fidelity (HF) 2x PCR master mix (NEB) using a T3000 thermocycler (Biometra) or a Lab Cycler Basic (SensoQuest). A typical PCR reaction contained 50-100 ng template DNA, both primers at a final concentration of 0.5 µM, 3% DMSO, and 1x Phusion HF master mix in a total volume of 50 µL. Double-stranded template DNA was initially denatured at 98 °C for 30 s and then amplified in 25-30 consecutive cycles with three steps each: (1) Denaturation (98 °C, for 10-30 s) of the DNA double strand, (2) primer annealing to the single-stranded DNA (depending on the specific melting temperature of the primer-template hybrid or in a gradient setting with annealing temperatures varying between 50-65 °C, for 30 s), and (3) primer extension (72 °C, for 30 s per kb,

depending on the product length). Afterwards, a final extension step was carried out (72 °C, 10 min). For each PCR reaction, a negative control reaction without template DNA was conducted in parallel to exclude contamination of PCR components with active amplicons. Reactions were analyzed using agarose gel electrophoresis and PCR products with the correct size were recovered from the gel as described in Section 2.2.1.4.

2.2.1.4 Agarose gel electrophoresis and preparative DNA isolation

DNA fragments were analyzed using agarose gel electrophoresis. Agarose was dissolved at a concentration of 1% (w/v) by heating in 1xTAE buffer. The molten agarose solution was mixed with SYBR safe DNA stain (Invitrogen; 6 µL per 100 mL agarose solution) in an appropriately-sized casting tray and polymerized at room temperature. DNA samples, pre-mixed with DNA loading dye (NEB), were separated along with a DNA ladder (GeneRuler 1 kb Plus, Thermo Fisher Scientific) on the agarose gel at a suitable voltage (~10 V/cm gel) in 1x TAE buffer. After a sufficient separation was achieved (usually after 30-45 min), DNA fragments were visualized using a Gel IX20 Imager System (Intas). In cases where isolation of PCR products was required for downstream applications, DNA fragments with the correct size were visualized under blue light with a Bio Transilluminator (BioStep) and excised from the gel using a scalpel. DNA was subsequently isolated from the gel using the MinElute gel extraction kit (Qiagen) according to the manufacturer's instructions, and eluted in 10-12 µL ultrapure water.

2.2.1.5 Restriction digestion

Plasmid DNA was digested for restriction endonuclease-based cloning (see Section 2.2.1.6) or to produce a linearized template for polymerase chain reaction. For this, 1-5 µg plasmid DNA were digested with 50 units of the appropriate restriction endonuclease(s) (NEB) in an enzyme-specific 1x digestion buffer in smallest possible volume for 2 h at 37 °C. If the plasmid DNA was used for restriction enzyme-based cloning, 5'-ends of respective vector fragments were dephosphorylated for 30 min at 37 °C by subsequent addition of 2 µL FastAP alkaline phosphatase (Thermo Fisher Scientific) directly to the digestion reaction. Plasmid digestion reactions, in which only a single restriction enzyme was used, were cleaned up using the

MinElute PCR purification kit according to the manufacturer's instructions. Double digest reactions were analyzed on an agarose gel and correctly-sized fragments were excised and purified as described in Section 2.2.1.4.

2.2.1.6 DNA ligation

Purified digested DNA fragments with complementary sticky ends were ligated using T4 DNA ligase. For a standard ligation reaction, 30 fmol vector and 60 fmol insert fragments were mixed in the smallest possible volume. Then, an equivalent volume of 2x ligase mix consisting of 1/5 vol T4 DNA ligase stock solution (NEB), 1/5 vol 10x T4 DNA ligase buffer and 3/5 ultrapure water was added to the DNA fragments. A background control without insert fragment was carried out in parallel to control for self-ligation of vector fragments. The ligation reaction was incubated for 30 min at 37 °C, before heat inactivation of T4 DNA ligase for 10 min at 70 °C. Half of the ligation reaction was then used to transform chemically competent *E. coli* cells (see Section 2.2.1.2).

2.2.1.7 Gibson assembly reaction

Gibson assembly[279] was used for seamless, restriction endonuclease-independent cloning of insert fragments into a vector backbone. Primers with an appropriate 5'-overhang were used to generate insert and vector fragments containing overlapping ends of 15-20 bp. The primers were designed using the NEBuilder Assembly Tool (NEB). For a typical Gibson assembly reaction, 0.1-0.5 pmol of each fragment (insert(s) and vector) were mixed with 2x Gibson assembly master mix (NEB) in a total volume of 10 µL and incubated for 15 min at 50 °C. As a background control, a reaction lacking Gibson assembly mix was conducted in parallel. Generally, 5 µL of the reaction mix were used for transformation of chemically competent *E. coli* 5-alpha cells.

2.2.1.8 Site-directed mutagenesis

PCR-based site-directed mutagenesis was used to introduce small insertions into a given vector sequence (insertion mutagenesis) or delete defined DNA sequences from a parental vector

(deletion mutagenesis). If an appropriate restriction endonuclease site was available, the parental vector was first linearized and then amplified by PCR. In case of deletion mutagenesis, the primers were designed to hybridize on both sites of the sequence to be deleted, facing it with their 5'-ends. In case of insertion mutagenesis, the primers are designed to bind at the precise location where the insertion shall be introduced. Either one or both of the primers possesed an overhang at the 5'-end encoding the corresponding insertion. PCR amplification (see Section 2.2.1.3) then generated a linear product fragment carrying the corresponding modification. In order to allow for the re-circularization of the blunt-ended PCR product by T4 ligase, the 5'-ends of the fragments were phosphorylated using T4 polynucleotide kinase. Due to similar buffer requirements of both enzymes, phosphorylation and ligation was carried within the same reaction. For this, 200-500 ng of the purified PCR product were incubated with 0.4 µL T4 polynucleotide kinase (NEB), 1 µL T4 ligase (NEB) in a total volume of 10 µL 1x T4 ligase buffer (NEB) for 30 min at 37 °C. T4 ligase was subsequently inactivated through a 10 min incubation at 70 °C. As background control, a parallel reaction without T4 ligase was conducted. Due to the relatively small size difference between the unmodified template and the PCR product that carries the insertion or deletion, low amounts of template DNA are often co-excised upon isolation of the PCR product from the agarose gel and might cause a substantial background. For this reason, the ligation reaction was treated with the restriction endonuclease Dpn1, which cleaves specifically the *E. coli* methylated template DNA. For this, 8 µL of the ligation reaction were mixed with 1 µL 10x CutSMART buffer (NEB) and 1 µL Dpn1 (NEB) and incubated for 2 h at 37 °C. After completion, 5 µL of the reaction mixture were directly transformed in chemically competent *E. coli* strains XL1-blue or 5-alpha.

2.2.1.9 Plasmid DNA preparation and sequencing

To identify positive clones from cloning reactions, 6-8 different non-overlapping colonies were selected and grown overnight in test tubes containing 8 mL LB medium and appropriate antibiotics. Preparation of plasmid DNA was generally performed using the Qiaprep Spin Miniprep kit (see Table 10) according to the manufacturer's instructions. Plasmid sequences were verified by Sanger sequencing (Microsynth Seqlab, Göttingen). From plasmids that were frequently used in cloning reactions, larger quantities were prepared using the Qiagen Plasmid Plus Midi or Maxi kit respectively (see Table 10), according to the manufacturer's instructions.

2.2.1.10 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Protein samples were separated according to their molecular weight using SDS-PAGE[280]. For this, protein samples were mixed with 4x NuPAGE LDS sample buffer (Thermo Fisher Scientific) and boiled at 95 °C for 1-10 min. Samples were spun down and loaded on a precast NuPAGE 4-12% Bis-Tris protein gel (Invitrogen) together with 2 µL PageRuler Prestained protein ladder (Thermo Fisher Scientific). The protein samples were separated at a constant voltage of 160-200 V for 45-60 min using either 1x MOPS or 1x MES running buffer (Thermo Fisher Scientific). Protein bands were then stained using InstantBlue staining solution (Expedeon) or silver staining. For silver staining, the gel was incubated for 20 min in 50% EtOH, 20 min in 5% EtOH and soaked with 35 µM DTT for 5 min. The gel was then incubated for 10 min with silver nitrate solution (containing 1 mg/mL silver nitrate and 0.1 µL/mL 37% formaldehyde in H_2O), stained with developing solution (30 mg/mL sodium carbonate and 0.5 µL/mL 37% formaldehyde in H_2O) to a desired intensity and quenched using solid citric acid monohydrate.

2.2.2 Insect cell culture techniques

2.2.2.1 Cloning of vectors for insect cell expression

Single subunits of multisubunit protein complexes were cloned first individually into the pFastBac-derived-MacroBac 438 vector series[281] using ligation-independent cloning. Open reading frames encoding single subunits were either cloned into the vector 438A (no tag), or in-frame into the vectors 438B (6xHis-TEV-tag) or 438C (6xHis-MBP-TEV-tag). For this, in either case, SspI-linearized vector and insert fragments (50-500 ng each) were treated with 2 units T4 DNA polymerase in the presence of 2.5 mM dGTP or dCTP, respectively, and 1x T4 DNA polymerase buffer containing 5 mM DTT for 40 min at 25 °C. Reactions were combined and incubated at room temperature for 30 min to allow for fragment annealing, followed by transformation in *E. coli* XL1-blue cells (see Table 1). Plasmid DNA was prepared using the Qiaprep Spin Miniprep kit and verified as described in 2.2.1.9. In order to combine single subunits from two different vectors, restriction digestion with Pme1 was used to excise the prospective subunit-encoding insert fragment. The second vector encoding a different complex subunit served as recipient and was concomitantly digested with Swa1. Vector and

insert fragments were treated again for 40 min at 25 °C with 2 units T4 DNA polymerase in the presence of 2.5 mM dGTP or dCTP, respectively. Fragments were annealed for 30 min at room temperature and transformed in *E. coli* XL1-blue cells. The obtained two-subunit encoding plasmid might then serve as the recipient vector for the addition of another subunit-encoding insert following the same strategy. Successive rounds of this ligation-independent cloning strategy allowed the combination of all subunits of multi-subunit complexes in a single plasmid. Expression of each subunit is controlled by a separate PolH promoter and followed by a SV40 termination site. In addition, the outermost subunit-encoding open reading frames (ORFs) are flanked together with a gentamycin resistance cassette by Tn7 transposase target sites.

2.2.2.2 Recombination in *E. coli* DH10αBacY and bacmid isolation

About 0.5 µg plasmid DNA was transformed in *E. coli* DH10αBacY cells using a MicroPulser electroporator (BioRad), which was set to 25 µF and 1.8 kV. This *E. coli* strain carries a bacmid that encodes lacZ with an internal Tn7 transposase insertion site (attTn7) and YFP under control of a PolH promoter as well as a helper plasmid encoding the Tn7 transposase. This allows the transfer of the target gene cassettes and gentamycin resistance marker, which are both flanked by Tn7 target sites from the MacroBac 438 series into the bacmid vector, thereby disrupting the lacZ ORF. Immediately after transformation, LB medium was added and cells were recovered for 5-16 h at 37 °C and plated on LB agar plates containing X-gal, IPTG, and gentamycin. Cells from white colonies, which are indicative of successful transposition, were used to inoculate a 5 mL LB overnight culture containing gentamycin. Cells were then harvested by centrifugation, resuspended in 250 µL buffer P1 (50 mM Tris pH 8.0, 10 mM EDTA, 50 mM glucose, 0.1 mg/ml RNase A), and lysed by addition of 250 µL lysis buffer P2 (0.2 M NaOH, 1% SDS). After the solution turned clear, 350 µL neutralization buffer P3 (4 M KOAc pH 5.5) were added and the lysate was cleared from cell debris twice by centrifugation (20,000g, 10 min, room temperature). The supernatant was then removed and mixed with 700 µL isopropanol to precipitate DNA by incubation at -20 °C for several hours. After centrifugation (20,000 g, 30 min, 4 °C), the DNA pellet was washed with 500 µL 70% ethanol, covered with 70% ethanol and stored at -20 °C until further use.

2.2.2.3 Expression of recombinant proteins in insect cells

Sf9 and Hi5 insect cell stocks were constantly maintained in culture, protected from light and at constant temperature (27 °C) and agitation (60 rpm). Cell density, diameter and viability were analyzed regularly using a CASY TT150 cell counter (Omni Life Science). Cell morphology was regularly monitored using light microscopy. Table 8 gives an overview of the combined vectors that were used in this study for insect cell expression.

Sf9 insect cells grown in SF900-III SFM medium to a density of ~10^6 cells/mL were transfected with reconstituted bacmid DNA using X-tremeGENE9 transfection reagent. After 48 h incubation in the dark without agitation, cells were inspected under a fluorescence microscope for YFP expression, which is indicative of successful transfection and resulting virus production. If YFP expression was detected for at least 5 cells, the virus-containing supernatant V_0 was collected 72 h after transfection. Subsequently, 0.15-3 mL V_0 virus were used to infect 25 mL insect cell culture grown to a density of ~10^6 cells/mL. Insect cells were grown until proliferation arrest and then incubated for additional 48-72 h. The V_1 virus-containing supernatant was then separated from cells by mild centrifugation (320 g, 10 min, 4 °C) and either stored in the dark at 4 °C or used directly for infection of an expression culture.

For protein expression, 0.2-2 mL V1 virus solution were used to infect 600 mL Hi5 cell culture that was grown in ESF921 medium to a density of 10^6 cells/mL. If low expression yield was expected, multiple expression cultures were prepared in parallel as required. Cells were incubated under standard conditions (27 °C, 60 rpm), and culture density, cell diameter, cell viability and YFP expression were monitored regularly in 24 h intervals. After the viability decreased to 80-90% (usually after 48-72 h), the cells were harvested by centrifugation (238 g, 30 min, 4 °C), resuspended in an appropriate lysis buffer, flash-frozen in liquid nitrogen and stored at -80 °C until further use.

2.2.2.4 Lysis and preparation of cleared insect cell extract

Insect cell suspension from the expression culture was thawed and lysed by sonication. For this, the cell suspension was transferred to a metal beaker and sonicated on ice with a Branson 250 Digital Sonifier (Marshall Scientific) at 30% power for 2 min, with ON and OFF times of 0.6 s and 0.4 s respectively. The obtained lysate was then first cleared from cell debris

by centrifugation at 26,195 g for 30 min at 4 °C, followed by ultracentrifugation at 158,420 g for 1 h at 4 °C. The clarified lysate was passed through a syringe filter with 0.8 µm pore size, before target proteins were further purified using an ÄKTA Pure chromatography system (GE healthcare).

2.2.3 *E. coli* cell culture techniques

2.2.3.1 Expression of recombinant proteins in *E. coli*

All bacterial expression constructs were expressed in *E. coli* BL21 CodonPlus (DE3)-RIL cells. The cells carry additional copies of the argU, ileY and LeuW tRNA genes on a chloramphenicol-resistant plasmid to facilitate a more efficient translation of heterologous proteins, and are thus always cultured in the presence of chloramphenicol. Table 7 gives an overview of the vectors that were used in this study for recombinant protein expression in *E. coli*.

Chemically competent *E. coli* cells were transformed with the respective expression plasmid (see Section 2.2.1.2, Table 7). The recovered culture was unevenly spread on LB agar containing appropriate antibiotics and incubated overnight at 37 °C. A single *E. coli* colony was then used to inoculate a 400 mL LB-preculture containing antibiotics, which was incubated overnight at 37 °C and shaking at 110 rpm. On the next day, 2 L LB medium containing appropriate antibiotics were inoculated with 60 mL preculture and shaken in a baffled 5 L flask at 37 °C and 110 rpm until an OD_{600} of ~0.8 was reached. The volume of the expression culture was scaled up as required, based on the expected expression level (see Table 12). Expression temperature and duration varied also for different constructs and are summarized in Table 12. Protein expression was induced at the respective expression temperature with 0.5 mM IPTG and further incubated with agitation (110 rpm). Cells were then harvested by centrifugation (15 min, 7000 rpm, 4 °C) and cell pellets were resuspended in the respective lysis containing protease inhibitors buffer (see Section 2.3, Table 9). The *E. coli* cell suspension was flash-frozen in liquid nitrogen and stored at -80 °C.

Table 12 | Conditions for overexpression of recombinant proteins in *E. coli*. Overexpression temperature and duration as well as culture volume for all purified proteins. For more information on the expression construct refer to Table 7.

Construct	Temperature	Duration	Volume	*E. coli* strain
MBP-hCTD WT/YF/YL	37°C	3-4 h	6x 2 L	BL21-CodonPlus (DE3) RIL
MBP-yCTD	37°C	4 h	6x 2 L	BL21-CodonPlus (DE3) RIL
MBP	37°C	4 h	2 L	BL21-CodonPlus (DE3) RIL
GST-yCTD	18°C	16 h	6x 2 L	Rosetta 2 (DE3) pLysS
GST	18°C	3 h	2 L	Rosetta 2 (DE3) pLysS
GFP-NELFA tentacle	22°C	3.5 h	6x2 L	BL21-CodonPlus (DE3) RIL
GFP-NELFE tentacle	22°C	3.5 h	6x2 L	BL21-CodonPlus (DE3) RIL
NELFE tentacle-GFP-NELFA tentacle	22°C	4 h	6x2 L	BL21-CodonPlus (DE3) RIL
SUMO2	18 °C	6 h	3x2 L	BL21-CodonPlus (DE3) RIL

2.2.3.2 Lysis and preparation of cleared *E. coli* extract

Frozen *E. coli* cell suspension was thawed and lysed by sonication using a Branson 250 Digital Sonifier. Sonication was conducted in metal beakers placed in an ice-water bath. Cells were lysed with alternating pulses of 10 s at an amplitude of 60% followed by an OFF period of 50 s, for a total ON time for 2.30 min. The cell lysate was then centrifuged for 1 h at 27.000 g and 4 °C. The cleared cell lysate was recovered and filtered through 0.8 µm syringe filters prior to application to an ÄKTA pure chromatography system (GE healthcare).

2.3 Project-specific techniques

2.3.1 RNA polymerase II clustering through carboxy-terminal domain phase separation

---

This section describes project-related methods and techniques that were specifically developed and applied in this study. Several protocols presented in this section have been published:

RNA polymerase II clustering through carboxy-terminal domain phase separation

M. Boehning*, C. Dugast-Darzacq*, M. Rankovic*, A. S. Hansen, T. Yu, H. Marie-Nelly, D. T. McSwiggen, G. Kokic, G. M. Dailey, P. Cramer, X. Darzacq, M. Zweckstetter

Nature Structural and Molecular Biology **25**, 833–840 (2018)

A detailed list of published items can be found in the Appendix (‘List of items from publications’, Page 149). Published methods presented within this section are marked with an asterisk (*). Contributions from co-authors of the publication are stated below the caption and can be found on Page VI. In some cases the caption deviates from the publication and different sections were subdivided or combined to improve clarity.

2.3.1.1 Purification of MBP-tagged human and yeast CTD and MBP

6xHis-MBP-hCTD, 6xHis-MBP-yCTD or 6xHis-MBP were overexpressed in *E. coli* BL21-CodonPlus (DE3) RIL cells, harvested and resuspended in lysis buffer (LB) 300 (20 mM Na-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 PMSF, 0.33 mg/mL benzamidine) as described in Section 2.2.3. The cleared and filtered *E. coli* lysate was applied to a 5 mL HisTrap HP column (GE healthcare) that was equilibrated in LB300. The HisTrap column was washed extensively with 20 CV of high-salt buffer HSB1000 (20 mM Na-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 PMSF, 0.33 mg/mL benzamidine). Afterwards the column was equilibrated again in LB300 and then attached inline to a XK-16 column (GE healthcare) that was filled with ~20 mL amylose beads (NEB) and pre-equilibrated in LB300. The HisTrap column was then developed with nickel elution buffer 300 (20 mM Na-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 PMSF, 0.33 mg/mL benzamidine) to elute bound proteins directly onto the amylose column. The HisTrap column was subsequently removed and the amylose column was washed again with 4-5 CV of HSB1000 buffer. The column was then equilibrated in LB300 and MBP-tagged proteins were eluted with amylose elution buffer (20 mM Na-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 PMSF, 0.33 mg/mL benzamidine). All elution factions were pooled and concentrated using a 30 kDa MWCO Amicon Ultra centrifugal filter (Merck). The protein solution was then applied to a Superdex 200 10/300 Increase column (GE healthcare) that was equilibrated in SE300 buffer (20 mM Na-HEPES, pH 7.4, 300 mM NaCl, 10% glycerol, 1 mM TCEP). The purity of the elution fractions was assessed using SDS-PAGE and Coomassie staining (as described in 2.2.1.10). Fractions containing pure protein were pooled and concentrated using a 30 kDa MWCO Amicon Ultra concentrator. Protein concentrations were determined based on the predicted molar extinction coefficient (section 2.2.1.1). Concentrated protein solutions (>100 µM) were divided in small aliquots (5-10 µL), frozen in liquid N_2 and stored at -80°C until further use.

2.3.1.2 Purification of GST-tagged yeast CTD and GST

Purification of 6xHis-GST-yCTD was conducted using a similar strategy as described previously[282] with minor modifications. The protein was overexpressed in *E. coli* BL21 Rosetta 2(DE3)pLysS cells as described in Section 2.2.3.2 and Table 12. Afterwards the cells were collected by centrifugation, resuspended in lysis buffer (LB) 150 (20 mM Na-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 PMSF, 0.33 mg/mL benzamidine) and lysed by sonication (Section 2.2.3.2). The cleared and filtered *E. coli* extract was loaded onto a 5 mL HisTrap HP column that was pre-equilibrated with LB150. The HisTrap column was then

washed extensively with high salt buffer (HSB) 800 (20 mM Na-HEPES, pH 7.4, 800 mM NaCl, 10% glycerol, 1 mM DTT, 0.284 µg/mL leupeptin, 1.37 µg/mL pepstatin A, 0.17 mg/mL PMSF, 0.33 mg/mL benzamidine). The column was equilibrated again with LB150 and connected inline to a pre-equilibrated 5 mL HiTrap Q HP column (GE healthcare). The HisTrap column was then developed over a 18 CV-linear gradient ranging from 0-100% nickel elution buffer (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 PMSF, 0.33 mg/mL benzamidine) and flow-through fractions were collected. Flow-through fractions were analyzed using SDS-PAGE and Coomassie staining and pooled accordingly. The sodium chloride concentration of the pooled protein solution was adjusted to 50 mM through dilution in no-salt buffer (20 mM Na-HEPES, pH 7.4, 10% glycerol, 1 mM DTT, 0.284 µg/mL leupeptin, 1.37 µg/mL pepstatin A, 0.17 mg/mL PMSF, 0.33 mg/mL benzamidine) and concentrated using a 30 kDa Amicon Ultra spin filter. The concentrated protein solution was then applied to a 1 mL HiTrap S column (GE healthcare) pre-equilibrated in LB50 (20 mM Na-HEPES, pH 7.4, 50 mM NaCl, 10% glycerol, 1 mM DTT, 0.284 µg/mL leupeptin, 1.37 µg/mL pepstatin A, 0.17 mg/mL PMSF, 0.33 mg/mL benzamidine). The flow-through fractions were analyzed by SDS-PAGE and Coomassie staining, pooled accordingly and concentrated using a 30-kDa MWCO Amicon Ultra spin concentrator. The protein was then subjected to size-exclusion chromatography using a Superdex 200 Increase 10/300 column (GE healthcare) equilibrated in SE300 buffer. Peak fractions were analyzed using SDS-PAGE and Coomassie staining and appropriate fractions were pooled and concentrated using a 30 kDa MWCO Amicon centrifugal filter. Clarified and 0.8 µm-filtered *E. coli* extract from overexpression of 6xHis-GST-TEV was loaded on a 5 mL HisTrap HP column. The column was washed with HSB800, and then developed with a linear gradient of 0-100% nickel elution buffer 150. The eluate was concentrated using a 10 kDa MWCO Amicon concentrator and then applied to a Superdex 200 Increase 10/300 column as described above. Purified proteins were concentrated, aliquoted, flash-frozen at stored at -80 °C.

2.3.1.3 Purification of TFIIH kinase modules (*)

Purified yeast TFIIH kinase module was a kind gift of Dr. Sandra Schilbach (Department of Molecular Biology, MPI for Biophysical Chemistry). Purification of the human TFIIH kinase module was conducted by Dr. Goran Kokic (Department of Molecular Biology, MPI for Biophysical Chemistry).

The recombinant *S. cerevisiae* TFIIH kinase module consisting, of the subunits Kin28, Ccl1, and Tfb3, was prepared as described[37]. For purification of the three-subunit human TFIIH 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 $MgCl_2$, 10 µM $ZnCl_2$, 5 mM β-mercaptoethanol, 30 mM imidazole, pH 8, 0.284 µg/mL leupeptin, 1.37 µg/mL pepstatin A, 0.17 mg/mL PMSF, 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 $MgCl_2$, 10 µM $ZnCl_2$, 5 mM β-mercaptoethanol, 500 mM imidazole, pH 8, 0.284 µg/mL leupeptin, 1.37 µg/mL pepstatin A, 0.17 mg/mL PMSF, 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 $MgCl_2$, 10 µM $ZnCl_2$, 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 $MgCl_2$, 10 µM $ZnCl_2$, 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.

2.3.1.4 Pol II preparation and fluorescent labeling (*)

Pol II was prepared from the *S. cerevisiae* strain BJ5464 as described[283] 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, ~2× 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.

2.3.1.5 CTD phosphorylation (*)

GST-yCTD was phosphorylated using the recombinant *S. cerevisiae* TFIIH 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 $MgCl_2$, 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 TFIIH 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 $MgCl_2$, 20 µM $ZnCl_2$, 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 (v/v) 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 $MgCl_2$, 20 µM $ZnCl_2$, and 1 mM TCEP to induce phase separation. Immediately before imaging, the reaction was started by addition of human TFIIH kinase module to a final concentration of 0.4 µM and immediately analyzed by microscopy.

2.3.1.6 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 TFIIH 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.

2.3.1.7 Disorder prediction (*)

Recent cryo-EM analysis of mammalian RNA polymerase II could derive an atomic model only to RPB1 position P1487[284], indicating a high conformational flexibility of the following RPB1-linker and the C-terminal repeat domain. We thus used the VLXT predictor implemented in PONDR[285] to calculate the disorder propensity for the human RPB1 residues 1,488–1,970.

2.3.1.8 Differential interference contrast (DIC) and fluorescence microscopy (*)

These experiments were conducted by Dr. Marija Rankovic (Dept. of Translational Structural Biology in Dementia, DZNE).

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.

2.3.1.9 Pol II co-recruitment experiments (*)

These experiments were conducted by Dr. Marija Rankovic (Dept. of Translational Structural Biology in Dementia, DZNE).

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.

2.3.1.10 In vitro FRAP experiments (*)

These experiments were performed by Dr. Marija Rankovic (Dept. of translational structural biology in dementia, DZNE).

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.

2.3.1.11 Analysis of human CTD Y_1 mutants

To investigate the influence of tyrosine substitution on CTD phase separation, MBP-hCTD WT, the $Y_1 \rightarrow F_1$- and $Y_1 \rightarrow L_1$-variants were labelled on a single cysteine residue with Alexa Fluor 647 C2 maleimide (Thermo Fisher Scientific). For this, MBP-hCTD was incubated in a light protected tube on ice at a final concentration of 25 µM for 4-5 h with 10x excess dye in buffer containing 20 mM Na-HEPES, 300 mM NaCl, 10% glycerol and 5 mM TCEP. Excess dye was

removed in a first step using either PD10 desalting columns (GE healthcare) or Micro Bio-Spin P6 gel columns (Biorad) that were equilibrated in SE300 buffer (20 mM Na-HEPES, pH 7.4, 300 mM NaCl, 10% glycerol, 1 mM DTT) and then through at least five diafiltration steps using an 0.5 mL 30 kDa MWCO Amicon concentrator.

Differential interference contrast and fluorescence microscopy was used to examine droplet formation. For this, Alexa Fluor 647-labelled and unlabelled proteins were combined at a molar ratio of 1:10 and then mixed with dextran T500 to reach a final concentration of 20 µM in 20 mM Na-HEPES, pH 7.4, 220 mM NaCl, 2% glycerol, 1 mM TCEP, 16% dextran. 5 µL microliters of the resulting solution were trapped between a passivated glass slide and a coverslip, which was subsequently sealed. Samples were imaged using a Leica DM6000B microscope equipped with a 63x water-immersion objective and a 60R filter cube (wavelength/bandwidth: 600/37 nm (excitation), 675/67 nm (emission)). FRAP experiments were performed using a Leica SP8 confocal microscope. Imaging was performed in passivated 50-well CultureWell chambered coverslides (Grace Bio-Labs) (see Section 2.3.2.3). Phase separation was induced as described above and 3 µL of the solution were transferred to the coverslide. The plate was then sealed using transparent tape (Scotch) and incubated at RT for 2 h prior to imaging to allow the droplets to settle on the coverslip surface. FRAP movies were recorded at a resolution of 256x256 pixels with a pixel size of 96 nm x 96 nm and a scan speed of 1400 Hz. Bidirectional scanning was used with a phase X offset of -30.01. The 633 nm helium neon laser line was used at full intensity to photobleach a circular spot of 1 µm in the center of 6.5-8 µm-sized droplets. Recovery of fluorescence was recorded every 300 ms for 200 frames at low laser intensity (0.3 %). Fluorescence recovery of the bleached spot was analyzed using FIJI (version 1.52h), background subtracted, corrected for acquisition bleaching using an unbleached reference droplet and normalized to the mean pre-bleach intensity. A double exponential model was used to fit the obtained curves using GraphPad Prism (version 5.03).

2.3.1.12 Cell line establishment and characterization (*)

These experiments were conducted by Dr. Claire Dugast-Darzacq together with Dr. David T. McSwiggen (UC Berkeley, Dept. of Molecular and Cell Biology).

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% CO_2 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 µg/mL and was used thereafter in permanence in the culture of the cells at a concentration of 1 µg/mL to avoid endogenous RPB1 re-expression as described in ref.[208]. 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.

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 (Amershan Protran 0.45-µm 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, Thermo Fisher 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.

2.3.1.13 Cell imaging conditions (*)

These experiments were conducted by Dr. Claire Dugast-Darzacq, Dr. Anders Hansen, and Dr. Hervé Marie-Nelly (UC Berkeley, Dept. of Molecular and Cell Biology).

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 μM 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).

2.3.1.14 PALM imaging (*)

These experiments were conducted by Dr. Hervé Marie-Nelly (UC Berkeley, Dept. of Molecular and Cell Biology).

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

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

where P is the set of all detections, and N_p is the total number of detections. The f function[286,287] corrects for biases generated by points located at short distances to the borders (nucleus or nucleoli)

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

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 $G(r)$[287, 288]. 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 package[289] 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.

2.3.1.15 Single-molecule imaging (spaSPT) (*)

These experiments were performed by Dr. Claire Dugast-Darzacq together with Dr. Anders S. Hansen (UC Berkeley, Dept. of Molecular and Cell Biology).

After overnight growth, cells were labeled with 50 nM PA-JF_{549}[290] 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×/NA 1.49 oil-immersion TIRF objective (Nikon Apochromat CFI Apo TIRF 100× Oil), EM-CCD camera (Andor iXon Ultra 897; frame-transfer mode; vertical shift speed: 0.9 µs; –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 illumination[291]. An incubation chamber maintained a humidified 37 °C atmosphere with 5% CO_2, and the objective was also heated to 37 °C. Excitation was achieved using a 561-nm (1 W, Genesis Coherent) laser for PA-JF_{549}. 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)[292, 293]. Briefly, 1-ms, 561-nm excitation (100% AOTF) of PA-JF_{549} was delivered at the beginning of the frame to minimize motion blurring; 405-nm photoactivation pulses were delivered during the camera integration time (~447 µs) 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 µs per frame).

spaSPT data was analyzed (localization and tracking) and converted into trajectories using a custom-written Matlab implementation of the MTT algorithm[294] and the following settings: localization error: $10^{-6.25}$; deflation loops: 0; blinking (frames): 1; max competitors: 3; max D ($\mu m^2/s$): 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 598,941 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.

2.3.1.16 FRAP in cells (*)

These experiments were conducted by Dr. Claire Dugast-Darzacq together with Dr. Anders S. Hansen (UC Berkeley, Dept. of Molecular and Cell Biology).

FRAP experiments were performed and analyzed as previously described[292]. 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% CO_2, 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 Matlab[292]. 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 ($r = 0.6$ μm), 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.

2.3.2 NELF condensation accompanies stress-induced transcriptional downregulation

2.3.2.1 Purification of full-length NELF and tentacle deletion variants

Purification of NELF was conducted together with Dr. Seychelle Vos (Department of Molecular Biology, MPI for Biophysical Chemistry).

NELF complex was overexpressed in 3x600 mL Hi5 insect cells as described in see Section 2.2.2. The cells were harvested by centrifugation (238 g, 30 min, 4 °C) and resuspended in lysis buffer (20 mM Na-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 PMSF, 0.33 mg/mL benzamidine). The cell suspension was lysed by sonication and a cleared extract was prepared as described in Section 2.2.2.4. NELF complex was purified essentially as described[87, 156]. In brief, the cleared insect cell lysate was loaded on a 5 mL HisTrap HP column (GE healthcare) pre-equilibrated in lysis buffer (20 mM Na-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 PMSF, 0.33 mg/mL benzamidine) and washed with high salt buffer (20 mM Na-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 PMSF, 0.33 mg/mL benzamidine). The column was then attached inline to a 5 mL HiTrap Q column (GE healthcare), equilibrated in low salt wash buffer (20 mM Na-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 PMSF, 0.33 mg/mL benzamidine) and eluted using elution buffer (20 mM Na-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 PMSF, 0.33 mg/mL benzamidine). Flow-through fractions were analyzed using SDS-PAGE and Coomassie staining, appropriate fractions were pooled and mixed with 2 mg 6xHis-tagged TEV protease and 0.8 µM Lambda protein phosphatase. The protein solution was then transferred to a Slide-A-Lyzer cassette (Thermo Fisher Scientific) and dialyzed overnight against lysis buffer (20 mM Na-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 PMSF, 0.33 mg/mL benzamidine) containing 1 mM $MnCl_2$. The dialysate was applied to a 5 mL HisTrap column pre-equilibrated in lysis buffer (20 mM Na-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 PMSF, 0.33 mg/mL benzamidine) to remove TEV protease and the 6xHis tag. The flow-through fractions were then concentrated using a 100 kDa MWCO Amicon spin filter (Merck) and subjected to gel filtration using a HiLoad 16/600 Superdex 200 pg column in NELF size-exclusion buffer (20 mM Na-HEPES, pH 7.4, 150 mM NaCl, 10% glycerol, 1 mM DTT). Fractions were analyzed by SDS-PAGE and Coomassie staining. Pure fractions were pooled and concentrated with a 100 kDa MWCO Amicon filter (Merck). To produce P-TEFb-treated NELF, an aliquot was removed and incubated with 0.4 µM gluthathion S-transferase-tagged P-TEFb, 6 mM $MgCl_2$, and 3 mM ATP for 2 h at 30 °C. GST-tagged P-TEFb was bound to pre-equilibrated GSTrap 4B resin (GE healthcare) and the NELF-containing supernatant was subjected again to size-exclusion chromatography on a HiLoad 16/600 Superdex 200 pg column.

2.3.2.2 Purification of GFP-NELF tentacle fusion proteins

GFP-NELF tentacle fusion proteins were overexpressed in *E. coli* BL21-CodonPlus (DE3) RIL cells as described in Section 2.2.3. The cleared and filtered *E. coli* lysate was loaded on a 5 mL HisTrap HP column pre-equilibrated in lysis buffer (20 mM Na-HEPES, pH 7.4, 300 mM NaCl, 10% glycerol, 1 mM DTT, 0.284 µg/mL leupeptin, 1.37 µg/mL pepstatin A, 0.17 mg/mL PMSF, 0.33 mg/mL benzamidine). The column was washed extensively with high salt buffer (20 mM HEPES, pH 7.4, 1000 mM NaCl, 10% glycerol, 1 mM DTT, 0.284 µg/mL leupeptin, 1.37 µg/mL pepstatin A, 0.17 mg/mL PMSF, 0.33 mg/mL benzamidine) and eluted in a linear gradient over 20 CV with elution buffer (20 mM Na-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 PMSF, 0.33 mg/mL benzamidine). Purity of the elution fractions was assessed by SDS-PAGE and Coomassie staining and appropriate fractions were pooled. Single tentacle GFP fusion proteins were concentrated using a 30 kDa MWCO Amicon Ultra centrifugal filter and directly applied to a Superdex 200 10/300 Increase column equilibrated in NELF size-exclusion buffer (20 mM Na-HEPES, pH 7.4, 150 mM NaCl, 10% glycerol, 1 mM DTT). The eluate containing the double tentacle GFP fusion protein was mixed with 6xHis-tagged TEV protease to cleave the N-terminal hexahistidine tag and dialyzed overnight in a 10 kDa Slide-A-Lyzer cassette against lysis buffer (20 mM Na-HEPES, pH 7.4, 300 mM NaCl, 10% glycerol, 1 mM DTT, 0.284 µg/mL leupeptin, 1.37 µg/mL pepstatin A, 0.17 mg/mL PMSF, 0.33 mg/mL

benzamidine). The protein was then applied to a 5 ml HisTrap column pre-equilibrated in lysis buffer (20 mM Na-HEPES, pH 7.4, 300 mM NaCl, 10% glycerol, 1 mM DTT, 0.284 µg/mL leupeptin, 1.37 µg/mL pepstatin A, 0.17 mg/mL PMSF, 0.33 mg/mL benzamidine). Flow-through fractions were collected and concentrated using a 30 kDa MWCO Amicon Ultra spin filter (Merck) and then applied onto a Superdex 200 10/300 Increase column (GE healthcare) that was pre-equilibrated in NELF size exclusion buffer (20 mM HEPES, pH 7.4, 150 mM NaCl, 10% glycerol, 1 mM DTT). Peak fractions were assessed by SDS-PAGE and Coomassie staining. Pooled fractions were concentrated and protein concentration was calculated based on the absorbance at 280 nm and the predicted molar extinction coefficient. Small aliquots were flash-frozen in liquid nitrogen and stored at -80°C.

2.3.2.3 Purification of P-TEFb

Purification of P-TEFb was conducted by Dr. Seychelle Vos (Department of Molecular Biology, MPI for Biophysical Chemistry).

Wild-type P-TEFb and catalytically inactive P-TEFb (D149N) were expressed in Hi5 insect cells as described in Section 2.2.2. P-TEFb was purified from 4 L expression culture, essentially as described[87]. In brief, the clarified and filtered cell lysate was loaded on a 5 mL HisTrap HP that was pre-equilibrated in lysis buffer (20 mM Na-HEPES, pH 7.4, 300 mM NaCl, 10% glycerol, 30 mM imidazole, 1 mM DTT, 0.284 µg/mL leupeptin, 1.37 µg/mL pepstatin A, 0.17 mg/mL PMSF, 0.33 mg/mL benzamidine). The column was washed with HSB1000 (20 mM Na-HEPES, pH 7.4, 1 M NaCl, 10% glycerol, 1 mM DTT, 0.284 µg/mL leupeptin, 1.37 µg/mL pepstatin A, 0.17 mg/mL PMSF, 0.33 mg/mL benzamidine), equilibrated in lysis buffer and washed with low-salt buffer LB150 (20 mM Na-HEPES, pH 7.4, 150 mM NaCl, 10% glycerol, 1 mM DTT, 0.284 µg/mL leupeptin, 1.37 µg/mL pepstatin A, 0.17 mg/mL PMSF, 0.33 mg/mL benzamidine). The column was then connected inline to a HiTrap S column (GE healthcare) that was pre-equilibrated in LSB150. The HisTrap column was eluted using a linear gradient from 0-100% nickel elution buffer 150 (20 mM Na-HEPES, pH 7.4, 150 mM NaCl, 10% glycerol, 500 mM imidazole, 1 mM DTT, 0.284 µg/mL leupeptin, 1.37 µg/mL pepstatin A, 0.17 mg/mL PMSF, 0.33 mg/mL benzamidine). Collected flow-through fractions were analyzed by SDS-PAGE and Coomassie staining and P-TEFb peak fractions were pooled, mixed with 6xHis-tagged TEV protease and dialyzed overnight in a 10 kDa MWCO Slide-A-

Lyzer cassette (Thermo Fisher Scientific). The protein solution was then applied to a 5 mL HisTrap column that was equilibrated in lysis buffer. Flow-through fractions were pooled, concentrated using a 10 kDa MWCO Amicon Ultra spin filter and applied to a HiLoad S200 16/600 pg column pre-equilibrated in SE300 buffer (20 mM Na-HEPES, pH 7.4, 300 mM NaCl, 10% glycerol, 1 mM TCEP). Pure peak fractions, as assessed by SDS-PAGE and Coomassie staining, were pooled, concentrated in a 10 kDa MWCO Amicon Ultra concentrator and aliquots were flash-frozen and stored at -80 °C until further use.

2.3.2.4 Purification of SUMO2/3 and E1/E2/E3 enzymes

6xHis-TEV-SUMO2 was overexpressed in *E. coli* BL21-CodonPlus (DE3) RIL cells as described in Section 2.2. Cells were harvested and resuspended in low-salt buffer 150 (20 mM Na-HEPES, pH 7.4, 150 mM NaCl, 10% glycerol, 30 mM imidazole, 1 mM DTT, 0.284 µg/mL leupeptin, 1.37 µg/mL pepstatin A, 0.17 mg/mL PMSF, 0.33 mg/mL benzamidine). The cleared and 0.8 µm-filtered lysate was applied to a 5 mL HisTrap HP column equilibrated in low-salt buffer. The column was then washed with high-salt buffer 1000 (20 mM Na-HEPES, pH 7.4, 1 M NaCl, 10% glycerol, 30 mM imidazole, 1 mM DTT, 0.284 µg/mL leupeptin, 1.37 µg/mL pepstatin A, 0.17 mg/mL PMSF, 0.33 mg/mL benzamidine) and equilibrated in LSB150. The column was then developed over a linear 20 CV-gradient with nickel 150 elution buffer (20 mM Na-HEPES, pH 7.4, 150 mM NaCl, 10% glycerol, 500 mM imidazole, 1 mM DTT, 0.284 µg/mL leupeptin, 1.37 µg/mL pepstatin A, 0.17 mg/mL PMSF, 0.33 mg/mL benzamidine). The protein solution was incubated overnight with 6xHis-TEV protease at a mass ratio of 40:1 and then applied to a 5 mL HisTrap column equilibrated in low-salt buffer. Flow-through fractions were analyzed by SDS-PAGE and Coomassie staining and concentrated using a 3 kDa MWCO Amicon Ultra centrifugal filter. The concentrated protein solution was then applied to a Superdex 75 10/300 GL column that was equilibrated in size-exclusion buffer (20 mM Na-HEPES, pH 7.4, 150 KOAc, 5% glycerol, 1 mM TCEP). Fractions were analyzed by SDS-PAGE and Coomassie staining and pure peak fractions were pooled and concentrated using a 3 kDa MWCO Amicon Ultra spin filter. A part of the purified protein solution was fluorescently labelled using Alexa Fluor 647 maleimide as described in Section 2.3.2.5. The remaining protein solution was aliquoted, snap-frozen and stored at -80°C. HA-tagged 4xSUMO3, His-Aos1-His-Uba2 (E1), Ubc9 (E2), and ZNF451-1-N 1-246 (E3) were a kind gift

of Dr. Andrea Pichler (MPI for Immunobiology and Epigenetics, Freiburg) and purified as previously described[295].

2.3.2.5 Fluorescent labeling of proteins

NELF complex was chemically labeled using fluorescent TFP-Alexa Fluor 488 dye (Molecular Probes). For this, NELF complex was incubated with a 10-fold molar excess of the dye for 1 h on ice in the dark. The reaction was quenched by addition of a 10-fold molar excess of L-lysine (in 50 mM Tris-HCl pH 7.0) to the dye. To remove unreacted dye, the reaction was desalted using a Micro Bio-Spin P6 gel column (Biorad) pre-equilibrated in NELF size-exclusion buffer according to the manufacturer's instructions followed by multiple diafiltration steps using a 30 kDa MWCO Amicon Ultra spin filter unit. The protein was then concentrated and small aliquots were flash-frozen and stored at -80 °C. The protocol results in a labeling density of ~1.0-1.2 fluorophores per NELF molecule.

The native cysteine residues of SUMO2 (C48) and SUMO3 (C47) were used for labeling with Alexa Fluor 647 C2 maleimide (Molecular Probes). For Alexa Fluor 647 C2 maleimide labeling of MBP the engineered single cysteine residue C-terminal of the TEV cleavage site was used for labeling[87]. Proteins were incubated with a 10-fold molar excess of the dye for 4 h on ice in the dark. Unbound dye was removed using PD-10 desalting columns (GE healthcare) according to the manufacturer's instructions and extensive diafiltration using an Amicon Ultra centrifugal concentrator (Merck) with an appropriate molecular weight cut off.

2.3.2.6 Phase separation assays and microscopy

Phase separation assays were conducted in 50-well CultureWell chambered coverslides (Grace Bio-Labs) to which an additional silicon gasket was attached. To minimize nonspecific adsorption of molecules to the glass surface, the coverslide was passivated through a covalently attached layer of polyethylene glycol, following a similar procedure as described[272]. For this, the coverslides were washed with 2% Hellmanex III solution (Hellma Analytics) for 2 h, then extensively rinsed with ultrapure water, and dried. The coverglass surface was etched with 1 M NaOH for 1 h, extensively washed with ultrapure water, dried and incubated overnight with

mPEG5K-silane (25 mg/mL in 95% EtOH; Nanosoft Polymers). The glass slide was subsequently rinsed extensively with ultrapure water, dried and sealed with crystal clear tape.

For phase separation assays with NELF complex, unlabeled and Alexa Fluor 488-labeled protein were mixed at a molar ratio of 1:9 and diluted to the appropriate protein concentration using NELF size-exclusion buffer (20 mM Na-HEPES, pH 7.4, 150 mM NaCl, 10% glycerol, 1 mM DTT). Generally, 1.7 µL protein solution were then diluted with 3.4 µL LLPS buffer (20 mM Na-HEPES, pH 7.4, 10 % glycerol, 1 mM DTT) to induce phase separation. The plate was sealed with crystal clear tape to minimize sample evaporation. The plate was generally incubated for 1-2 h in the dark prior to imaging to assure that formed droplets have settled down quantitatively onto the coverslide surface. All images were acquired directly above the coverslide surface using an inverted Leica TCS SP8 laser scanning confocal microscope equipped with a HC PL APO 63x/1.40 CS2 objective (oil immersion) at room temperature (22±1 °C). For excitation of Alexa Fluor 488 and GFP, the 488 nm argon laser line was used at low intensity (usually 1 %). For NELF complex samples, emitted fluorescence between 492-550 nm was detected using a PMT detector gain of 750-850 V. Emitted light from GFP-tagged fusion proteins was detected between 500-600 nm with a typical PMT detector gain of 420-510 V. At least 5 images per condition were taken in non-overlapping regions, and are considered representative for the droplet distribution on the entire slide. Images were further analyzed and processed using FIJI (version 1.52h).

2.3.2.7 Hexanediol sensitivity assays

To test the sensitivity of phase separation towards the aliphatic alcohol 1,6-hexanediol, 1.3 µL of a 77 µM solution of the NELFE-GFP-NELFA double tentacle fusion protein were mixed with 3.7 µL LLPS assay buffer (20 mM Na-HEPES, pH 7.4, 15 mM NaCl, 10% glycerol, 1 mM DTT) either containing 13.6% or no 1,6-hexanediol. The coverslide was subsequently sealed with tape, incubated in the dark for 90 min to allow the droplets to settle on the coverslide surface and imaged as described above.

2.3.2.8 In vitro FRAP experiments

Fluorescence recovery after photobleaching (FRAP) was used to determine the influence of P-TEFb phosphorylation on the mobility of molecules in the condensed phase. Two types of FRAP experiments can be distinguished: Fluorescence recovery after bleaching only small parts of the condensed phase is largely dependent on the mobility of the molecules in the condensed phase, while the recovery after full droplet bleaching is a measure of the exchange between droplet material and the bulk solution over the phase boundary. Lambda phosphatase (λPPase)-treated NELF as well as NELF that was first treated with λPPase and afterwards with P-TEFb during purification were diluted in LLPS buffer to induce LLPS. A final NELF concentration of 2 µM and 2.5 µM was used for partial and full droplet FRAP experiments, respectively. After induction of phase separation, the plate was incubated for 1 h in the dark prior to imaging. FRAP movies were recorded with 1400 Hz scan speed at a resolution of 256 x 256 pixels with a pixel size of 96 x 96 nm. Bidirectional scanning was enabled using a phase X offset of -30.01. The 488 nm line was used at full laser intensity (80% output) to photobleach a defined region of interest to ~10-20% of its initial fluorescence. For partial droplet bleaching, a 1 µm circular region was bleached in the center of droplets with a diameter of 6.0-7.5 µm. Fluorescence recovery was imaged every 500 ms over a period of 250 s. For full droplet FRAP, the entire area of droplets with a diameter of 5.2-5.8 µm was photobleached and 600 frames were recorded with a frame rate of 0.5 s^{-1}. Fluorescence recovery of the bleached region was analyzed using FIJI (version 1.52h). For this, the fluorescence intensity of the bleached spots was background subtracted, normalized to the fluorescence intensity of the first post-bleach image, then corrected for acquisition bleaching using an unbleached reference droplet of similar size within the frame, and normalized to the mean prebleach intensity. The resulting recovery curves were fit to a double-exponential recovery model using GraphPad Prism (version 5.03).

2.3.2.9 Real-time P-TEFb droplet phosphorylation

For real-time P-TEFb droplet phosphorylation experiments, Alexa Fluor 488-labeled and unlabeled NELF complex were mixed at a molar ratio of 1:9 at a final concentration of 3 µM. Phase separation was then induced by mixing 1.7 µL of the NELF solution with 3.4 µL of LLPS buffer (20 mM HEPES, pH 7.4, 10 % glycerol, 1 mM DTT) containing 1.5 mM ATP and 4.5 mM $MgCl_2$. The required P-TEFb co-substrate ATP and cofactor $MgCl_2$ both affect NELF

phase separation and were thus included already during induction of LLPS. The coverslide was then sealed and incubated in the dark on the microscope stage for 2 h to assure that the NELF droplets have settled quantitatively on the coverslide surface. Subsequently, either active P-TEFb or a catalytically inactive P-TEFb variant (containing the D149N substitution in CDK9)[87] were added gently in a small volume (0.4 µL) to a final concentration of 0.2 µM. Immediately after P-TEFb addition, a series of images (usually 10-12) with variable z-position was acquired in a representative area spanning 92 x 92 µm as described above. After an incubation of 30 min, 60 min and 120 min, the same region was imaged again similarly over a comparable range of different focal planes. For the different time points, the images in which the focal plane intersects approximately the center of the largest droplet within a specified region were selected manually, cropped and contrast-adjusted using identical settings. After the experiment was completed, 3 µL 4x LDS sample buffer were added and the sample was retrieved from the well to confirm protein integrity using SDS-PAGE with subsequent silver staining.

2.3.2.10 P-TEFb phosphorylation of NELF double tentacle-GFP fusion protein

NELF double tentacle-GFP fusion protein was mixed at a final concentration of 50 µM with either 1 µM wild-type or catalytically inactive P-TEFb in kinase buffer (20 mM Na-HEPES, pH 7.4, 140 mM NaCl, 6 mM $MgCl_2$, 4% glycerol, 1 mM DTT). The reaction was started by addition of ATP to a final concentration of 3 mM and incubated at 30 °C for 4 h. To induce LLPS, 1.7 µL of the reaction were mixed with 3.4 µL LLPS buffer (20 mM Na-HEPES, pH 7.4, 10% glycerol, 1 mM DTT) containing 4.5 mM EDTA. The plate was sealed, incubated for 2 h in the dark and then imaged as described above.

2.3.2.11 Mass spectrometric analysis of P-TEFb phosphorylation

In-gel digestion, phosphopeptide enrichment and mass spectrometric analysis were kindly performed by Monika Raabe and Annika Reinelt (Laboratory of Prof. Dr. Henning Urlaub, MPI for Biophysical Chemistry).

Dephosphorylated or P-TEFb-treated NELF samples (see Section 2.3.2.1) were separated on a SDS-PAGE gel and stained with Coomassie solution (see Section 2.2.1.10). Appropriate bands

were selected for in-gel digestion and phosphopeptides were enriched using TiO_2 chromatography (GL Sciences) as described[296]. Desalted peptides were analyzed using a Dionex UltiMate 3000 nano liquid-chromatography system (Thermo Fisher Scientific) coupled to a Q-Exactive HF mass spectrometer (Thermo Fisher Scientific) as described[296]. Raw files were searched against the human proteome database using the Andromeda search engine of MaxQuant (version 1.5.2.8)[297]. Default settings were used, except serine, threonine, and tyrosine phosphorylation, methionine oxidation, and carbamidomethlyation of cysteine were set as variable modifications. Identified phosphorylation sites were filtered for high confidence (posttranslational modification (PTM) score > 0.75) and only sites that were exlusively detected in the P-TEFb-treated sample were considered relevant.

2.3.2.12 Droplet interaction assays

Alexa Fluor 647-labeled HA-$(SUMO3)_4$, SUMO2 or MBP were diluted in NELF size exclusion buffer and mixed at a final concentration of 3 µM with 9 µM of Alexa 488-labeled NELF complex. The solution was incubated for 5 min at room temperature before phase separation was induced through a 1:2 (v/v) dilution in LLPS buffer. The plate was sealed with transparent tape and incubated for 1 h prior to imaging as described in Section 2.3.2.6. In addition to the 488 nm line, the 633 nm helium neon laser line was used to excite Alexa Fluor 647 in sequential scanning mode. Emitted fluorescence was detected between 652-750 nm using a hybrid detector.

2.3.2.13 In vitro sumoylation assay

For in vitro sumoylation reactions, NELF was mixed at a final concentration of 1 µM with 0.1 µM E1, 0.1 µM E2, 0.5 µM E3 and 10 µM SUMO2 in assay buffer (20 mM K-HEPES, pH 7.4, 110 mM KOAc, 2 mM MgOAc, 0.2 mg/mL ovalbumin, 0.05 % Tween-20, 1 mM DTT). The reaction was started by addition of ATP to a final concentration of 5 mM and incubated at 30 °C for 60 min. Control reactions without ATP or E3-ligase were conducted in parallel. The reactions were carried out in 30 µL volumes in low protein binding tubes. The reaction was quenched by addition of 10 µL 4x LDS sample buffer. Sumoylation causes an apparent mass shift of ~20 kDa in SDS-PAGE, which was evaluated using western blotting.

For this, 4 µL of each reaction were separated on 4-15% Tris-glycine Protean TGX gels (Biorad) run in 1x TGS buffer (Biorad). Proteins were transferred onto a PVDF membrane using a Trans-Blot Turbo Transfer System (Biorad). The membrane was then blocked at RT for 1 h with 5% (w/v) milk powder in 1x PBS containing 0.1% Tween-20 (PBST, Table 9). The membrane was then incubated at RT for 3 h with NELF subunit-specific antibodies (Table 11) that were diluted in 2.5% milk powder/PBST. The membrane was washed 3x for 5 min with 1x PBST and then incubated at RT for 1 h with the appropriate HRP-conjugates secondary antibody (Table 11) in 2.5% milk powder/PBST. The membrane was washed again 3x for 5 min with 1x PBST and was then developed using SuperSignal West Pico Chemiluminescence Substrate (Thermo Fisher Scientific). Chemiluminescence was detected using a ChemoCam Advanced Fluorescence imaging system (Intas Science Imaging).

2.3.2.14 Cell line establishment and cell culture

These experiments were conducted by Prashant Rawat (Laboratory of Dr. Ritwick Sawarkar, MPI for Immunobiology and Epigenetics).

Human HeLa Flp-In T-Rex cells (kind gift from Prof. Dr. Marc Timmers, University of Freiburg) and HEK293 Flp-In T-Rex cells were cultured in DMEM high glucose medium supplemented with 10% fetal bovine serum (FBS) (Sigma Aldrich, F0804), 2 mM L-glutamine (Sigma Aldrich, P4333) and 1% penicillin-streptomycin (Sigma Aldrich, G7513) in a humidified incubator at 37 °C and 5% CO_2. Cells were routinely checked for mycoplasma contamination by PCR.

Human open reading frames cloned into pDONOR vectors were obtained from BIOSS (University of Freiburg). Coding sequences were then transferred into pDestination vectors encoding either C-terminal GFP tag (obtained from Prof. Dr. Marc Timmers, University of Freiburg) or mCherry tag (obtained from Dr. Robin Shaw, addgene #31907) using the Gateway LR clonase II enzyme kit (Life Technologies, 11791020) according to the manufacturer's instructions. HeLa cells were transfected with Lipofectamine 3000 (Life Technologies, L3000) as suggested by the manufacturer. For transient expression, HeLa cells were transfected with NELFA-mCherry WT and a NELFA-mCherry dIDR variant (lacking amino acids 321-460). To generate the stable cell lines, HeLa Flp-In T-Rex cells that were cultured in the presence of

100 µg/mL Zeocin (Invitrogen, R25001) for seven days were transfected with NELFA-GFP plasmid DNA according to the manufacturer's instructions. Selection of stable positive clones was carried out over a period of three weeks with 15 µg/mL blasticidin S (Carl Roth CP14.1) and 100 µg/mL hygromycin B. Additionally, Flp-In HEK293 T-Rex cell lines that stably express NELFA-GFP and CDK9-GFP were constructed in the same way.

2.3.2.15 Heat shock treatment and cell imaging

These experiments were conducted by Prashant Rawat (Laboratory of Dr. Ritwick Sawarkar, MPI for Immunobiology and Epigenetics).

Adherent HeLa cells were used for imaging. Prior to each experiment, cells were grown on Nunc Lab-Tek 8-well chambers (Thermo Fisher Scientific, 155411) for 24 h at 37 °C. Cells were then either transfected transiently or expression of the stable cell line was induced for 24 h with 1 µg/mL tetracycline (Sigma Aldrich, T7660). For heat shock experiments, cells were either shifted to a 43 °C incubator for 30 min or left at 37 °C. Prior to imaging, cells were treated with fixative solution (Invitrogen, FB002) for 10 min at 37°C and washed three times with 1x PBS (Sigma-Aldrich, D8537). DAPI (Serva, 18860) was added for 30 min at RT immediately before imaging using a Zeiss LSM780 confocal microscope. Images of stable cell lines were acquired at constant laser power. For transiently transfected cell lines, laser power was adjusted within a narrow window to avoid oversaturation of the images. Images were contrast adjusted using the Min/Max intensity function of the Zeiss Zen software (version 2012).

Treatments. Prior to the hexanediol treatment, expression was induced for 24 h with 1 µg/mL tetracycline. The cells were either incubated at 43°C without or with 10% 1,6-hexanediol for 15 min and imaged as described above. The number of puncta per cell were determined automatically with Imaris software (version 9.3) using a custom-written script. Two replicates were done on different days and one of the representative replicates is shown. Treatment with the SUMO E1 inhibitor ML-792 (kind gift from UbiQ) was performed at a final concentration of 1 µM for 1 h immediately before induction of heat shock. To induce oxidative stress, the cells were treated with 100 µM arsenic trioxide for 1 h prior to fixation.

FRAP in live cells. FRAP was done in live cells using a Zeiss LSM780 confocal microscope with 405 nm and 488 nm laser. A strip of the nucleus was bleached at 100% laser power for 2x30 µs and recovery was imaged at low laser intensity every 600 ms over 260 frames. The fluorescence intensity of the bleached region, an unbleached reference region and the background were determined using Zeiss Zen software. Normalized recovery curves were fit to a double exponential model using easyFRAP[298].

2.3.2.16 Analysis of CDK9 interactome and NELFA phosphorylation

These experiments were conducted by Prashant Rawat (Laboratory of Dr. Ritwick Sawarkar, MPI for Immunobiology and Epigenetics).

Quantitative mass spectrometry was carried out using the SILAC approach[299]. For the analysis of the CDK9 interactome, Flp-In T-Rex HEK293 cells that stably express C-terminally GFP-tagged CDK9 or GFP (kind gift of Dr. Asifa Akhtar[300]) were cultured in normal medium ('light'), medium containing $^{2}H_4$-lysine and $^{13}C_6$-arginine ('medium'), or $^{15}N_2{}^{13}C_6$-lysine and $^{15}N_4{}^{13}C_6$-arginine ('heavy') supplemented with dialyzed FBS (Thermo Fisher Scientific, 88440) and non-essential amino acids (Gibco, 11140050) for at least five cell divisions (starting density <1%). Light-, medium- and heavy-labelled cells were then either exposed to heat shock conditions (43 °C, 1.5 h) or left at 37 °C. Three replicate experiments were conducted following the label-swap strategy[301] (Park et al., 2012). After the treatment, cells were chemically crosslinked using 1% formaldehyde for 10 min and quenched for 5 min using excess L-glycine. Cells were then lysed in 1xRIPA buffer (50 mM Na-HEPES, pH 7.4, 150 mM NaCl, 0.5% sodium deoxycholate, 0.1% SDS, 0.1% Triton X-100, 5 mM $MgCl_2$, 10% glycerol) and combined in equal amounts. The cell lysate was sonicated using a BioRuptor system and cleared by centrifugation. The cleared cell lysate was then applied to GFP-trap magnetic agarose beads (ChromoTek, gtma 20) for 1 h at 4°C. Beads were then washed with high salt buffer (50 mM Tris-HCl pH 7.4, 1 M NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 1 mM EDTA) and non-denaturing buffer (50 mM Tris-HCl pH 7.4, 100 mM NaCl, 0.1% Tween-20). Proteins were digested sequentially on the beads with LysC and then with trypsin in solution. Peptides were desalted and analyzed using a Q-Exactive mass spectrometer (Thermo Fisher Scientific) coupled to an nLC 1000 Nano UHPLC (Thermo Fisher Scientific). Raw files were processed using MaxQuant[297] (version 1.5.7.4) and analyzed using Perseus (version 1.5.2.4). First, decoy

and contaminant entries were removed. Next, specific CDK9 interactors were discriminated from background interactors under steady-state and heat shock conditions. For this, Student's t-tests were performed based on the iBAQ intensities of proteins quantified in GFP and GFP-CDK9 pulldowns. Proteins exhibiting ≥2-fold enrichment and a t-test p-value ≥ 0.1 were defined as GFP-CDK9 specific interactors in both conditions. Subsequently, SILAC ratios of common CDK9 interactors under steady state and heat shock conditions were extracted and normalized so that the SILAC ratio of CDK9 equals one to account for slight variations in IP efficiency. Simultaneous Western blots of the input material were performed to confirm that the intracellular CDK9 level did not change upon heat shock.

For analysis of NELFA phosphorylation, Flp-In T-Rex HEK293 T-Rex cells that stably express C-terminally GFP-tagged NELFA were used. All procedures were conducted in triplicates as described above. Three different phosphorylation sites on NELFA were detected in each of the replicates with high confidence (PTM score > 0.99).

3 RNA polymerase II clustering through carboxy-terminal domain phase separation

3.1 Results

We concluded from published findings that CTD repeats must have a weak attractive interaction to each other[13, 21, 35-38]. This prompted us to ask whether CTD molecules could engage also in multivalent intermolecular interactions that result in liquid-liquid phase separation. In order to investigate whether such CTD-mediated LLPS could be the molecular mechanism underlying Pol II clustering, we expressed human CTD (hCTD) and yeast CTD (yCTD) in *Escherichia coli* fused to the solubility enhancing maltose-binding protein (MBP) tag or to the MBP and glutathione S-transferase (GST) tags, respectively. We chose a prokaryotic expression system to prevent the covalent attachment of any eukaryotic post-translational modifications during protein expression. CTD fusion proteins were purified to homogeneity (Fig. 3.1a) and then fluorescently labeled with Alexa Fluor 488.

Except for Section 3.1.3, the results presented in this Chapter have been published.

RNA polymerase II clustering through carboxy-terminal domain phase separation

M. Boehning*, C. Dugast-Darzacq*, M. Rankovic*, A. S. Hansen, T. Yu, H. Marie-Nelly, D. T. McSwiggen, G. Kokic, G. M. Dailey, P. Cramer, X. Darzacq, M. Zweckstetter

Nature Structural and Molecular Biology **25**, 833–840 (2018)

The published text was adapted to match the style of this thesis. Numbering and references to figures as well as references to the literature thus deviate from the published version. A detailed list of published items can be found in the Appendix (‘List of items from publications’, Page 149). Co-author contributions are stated on Page VI.

3.1.1 CTD of Pol II phase separates into liquid-like droplets

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. 3.2a). Fluorescence microscopy demonstrated that hCTD molecules were strongly concentrated within the droplet interior compared to the surrounding milieu (Fig. 3.2a, lower panels). At higher dextran concentration (16%), droplets could be detected already at a concentration of 5 μM hCTD (Fig. 3.1b-c). The number of droplets increased with increasing protein concentration (Fig. 3.1c), consistent with the general concentration dependence of liquid phase separation[235]. In addition, hCTD formed droplets in the presence of another molecular-crowding agent, the polysaccharide Ficoll (Fig. 3.2b). hCTD also underwent LLPS after cleavage of the maltose-binding protein (MBP) tag, while MBP alone did not form droplets in presence of molecular-crowding agents (Fig. 3.2c). hCTD droplet formation was robust against changes in ionic strength (Fig. 3.2d), and against incubation of the sample for 1 h at different temperatures (Fig. 3.2e). Like hCTD, yCTD formed droplets in a concentration-dependent manner (Fig. 3.1d; Fig. 3.4d). Contacts of both hCTD and yCTD droplets led to fusion and formation of a single spherical droplet (Fig. 3.1e-f). At concentrations subcritical for LLPS, yCTD was incorporated into preformed hCTD droplets and hCTD was included into preformed yCTD droplets (Fig. 3.3c), in agreement with the ability of CTD to be trapped into droplets and hydrogels of LCD proteins [134, 135]. Formation of yCTD droplets was also resistant against changes in ionic strength (Fig. 3.2d) and temperature (Fig. 3.2e), 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.

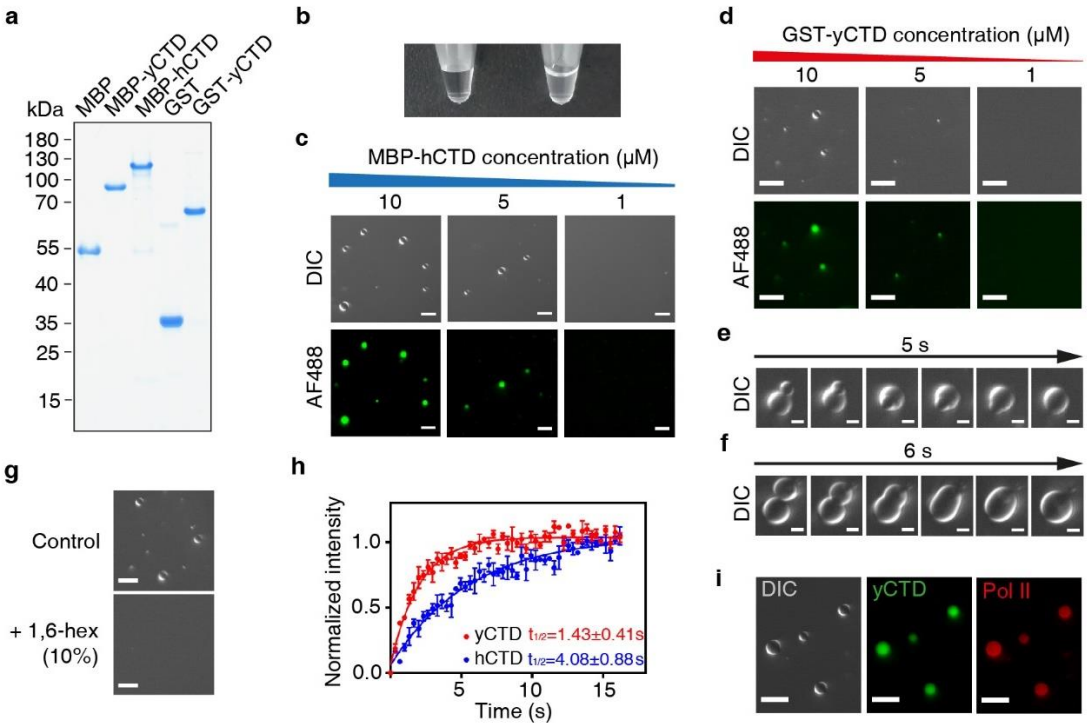

Figure 3.1 | Phase separation of Pol II CTD into liquid-like droplets.

a, Representative SDS-PAGE analysis of purified proteins used in this study reveals high purity and homogeneity. Due to low expression yields, several batches of the CTD fusion proteins were purified. **b**, Addition of 16% dextran to a 20 μM 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**,**f**, GST-yCTD (e) and MBP-hCTD (f) droplets rapidly fuse upon contact into one spherical droplet. **g**, 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. **h**, 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. **i**, 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 μm in **c**, **d**, **f**, **h** and 2.5 μm in **e, f**.

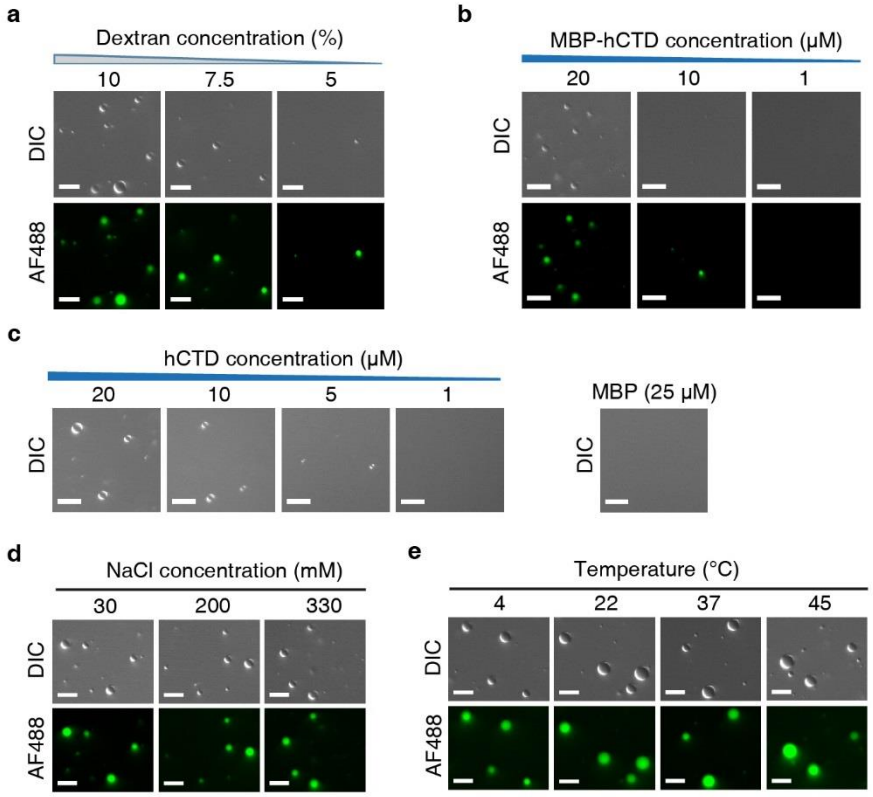

Figure 3.2 | Physicochemical properties of hCTD phase separation.

a, Phase separation of MBP-hCTD (20 μM) 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 μM) 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 μm.

Liquid droplets and cellular puncta are held together by weak, distributed interactions between LCDs that are sensitive to aliphatic alcohols[252, 302, 303]. As expected for such interactions, liquid phase separation of yCTD and hCTD was counteracted by addition of 5-10% 1,6-hexanediol (Fig. 3.1g; Fig. 3.4a-b, upper panels). Addition of 5-10% of the hexanediol isomer 2,5-hexanediol also inhibited CTD droplet formation (Fig. 3.4a-b, lower panels). Because it was shown that 2,5-hexanediol is less efficient in dissolving droplets and hydrogels[139], 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. 3.2d; Fig. 3.3c).

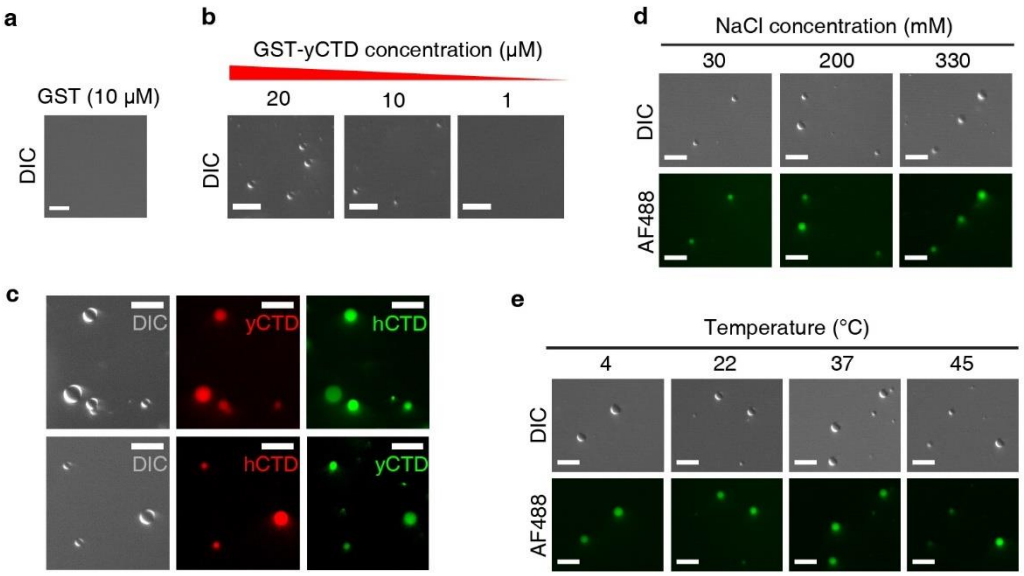

Figure 3.3 | Physicochemical properties of yCTD phase separation.

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,** LLPS of MBP-yCTD in 16% dextran is not sensitive to ionic strength changes. **e**, LLPS of MBP-yCTD in 16% dextran is robust against incubation of the sample for one hour at the indicated temperatures. All experiments were performed two times with similar outcome and representative images are shown. Scale bars correspond to 10 µm in all panels.

3.1.2 CTD length influences CTD phase separation in vitro

A characteristic property of liquid-like droplets is fast diffusion of molecules in their interior[231]. 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 in the interior of CTD droplets were bleached. Within hCTD droplets, the bleached fluorescence recovered with a half time of 4.08 s ± 0.88 s (Fig. 3.1h). For yCTD, recovery was faster with a half time of 1.43 s ± 0.41 s. (Fig. 3.1h).

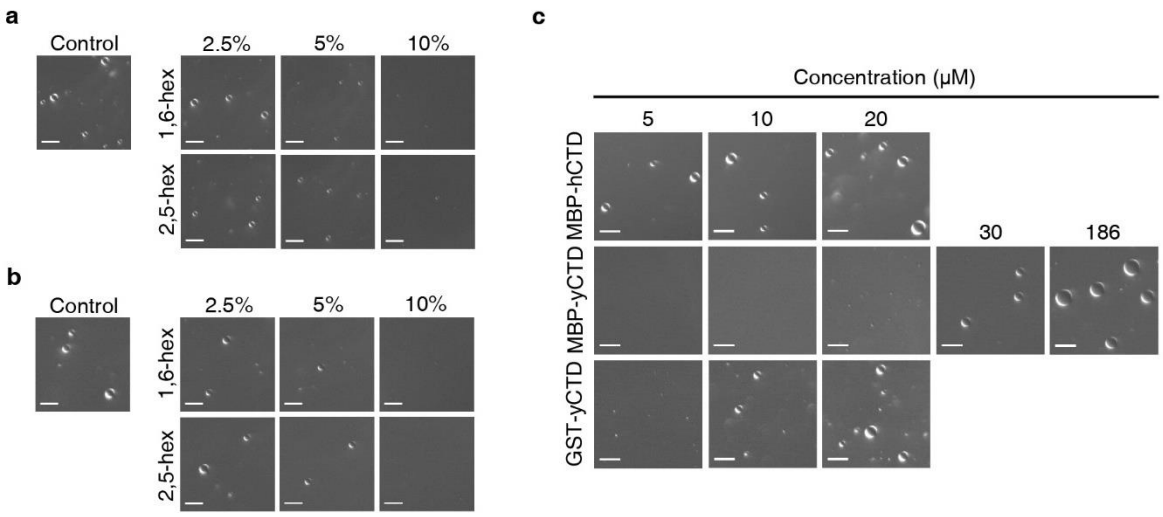

Figure 3.4 | Influence of aliphatic alcohols and solubility tags on CTD phase separation.
a-b, Influence of aliphatic alcohols on CTD LLPS. 1,6-hexanediol (1,6-hex) and 2,5-hexanediol (2,5-hex) counteract LLPS of hCTD (**a**) or yCTD (**b**) in a concentration-dependent manner. Images from ≥ 5 representative images taken for each condition are shown. (**c**) The solubilizing effect of an 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.

These results demonstrate that CTD molecules within droplets are 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 µM (Fig. 3.1c; Fig. 3.4c). In contrast, LLPS of MBP-yCTD started only at a four- to sixfold higher protein concentration (Fig. 3.4c). When the smaller, dimerizing glutathione S-transferase tag was used to replace the more soluble MBP tag[304], the critical concentration for yCTD phase separation decreased to approximately 5 µM (Fig. 3.4c; Fig. 1d). 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.

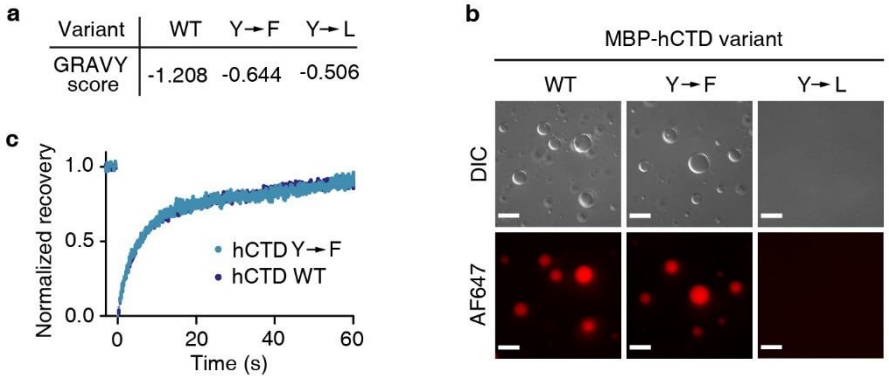

Figure 3.5 | Aromatic interactions drive CTD phase separation.

a, Hydrophobicity prediction of the wild-type (WT) CTD sequence in comparison to the Y→F and Y→L variants with the ProtParam tool (Expasy suite) based on Kyte & Doolittle[305]. The higher the GRAVY score, the higher the hydrophobicity. The variants are predicted to have approximately similar hydrophobicity and are more hydrophobic than the wild-type CTD sequence. **b**, WT hCTD and the Y→F variant form droplets at a final concentration of 20 µM, while the Y→L variant does not undergo phase separation. The depicted result is representative for two independent experiments. Scale bar, 10 µm. **c**, FRAP analysis of hCTD WT and Y→F droplets. Curves show the average normalized recovery (mean ± standard error) of five different droplets each.

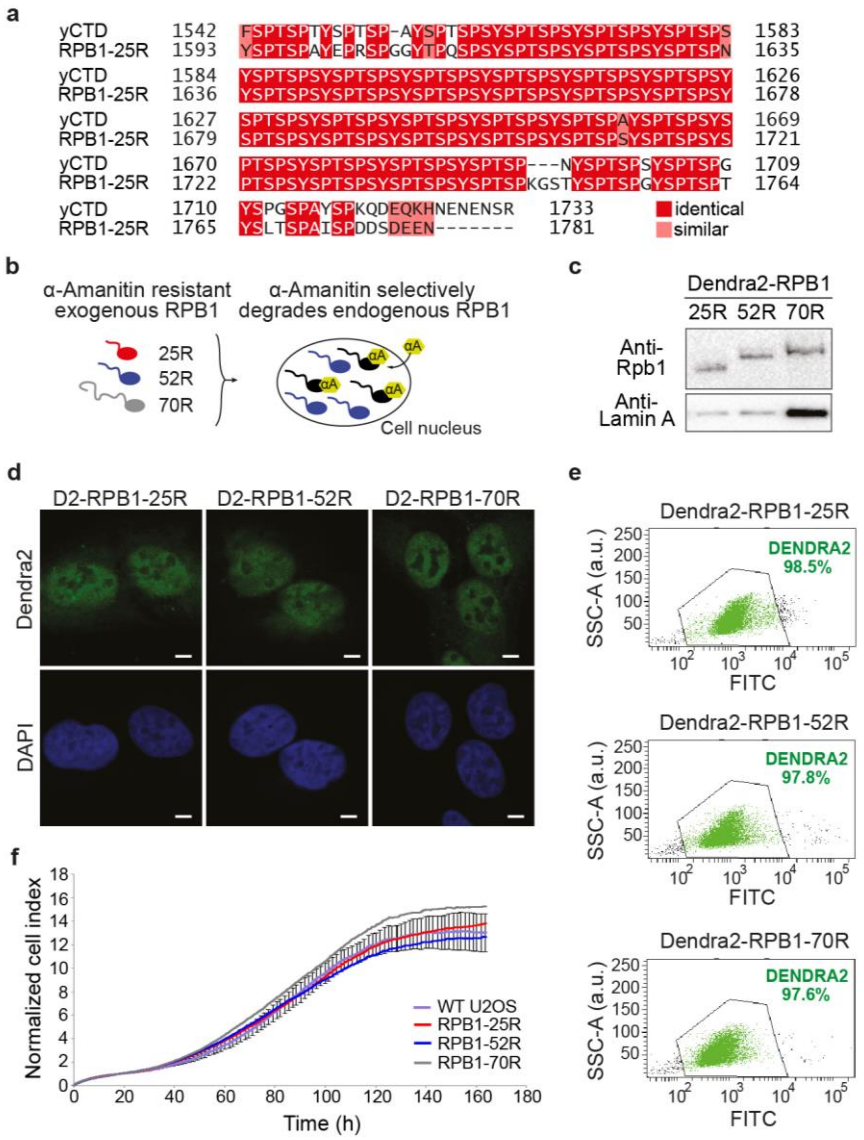

Figure 3.6 | Human Dendra2-RPB1 cell lines for imaging CTD-dependent Pol II clustering.

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 tool[306] 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. αA, 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. **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 µm. *(Figure caption continued on next page.)*

◀ **Figure 3.6 | Human Dendra2-RPB1 cell lines for imaging CTD-dependent Pol II clustering.** *(Figure caption continued from previous page.)*

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.

3.1.3 Aromatic interactions are critical for CTD phase separation

Tolerance of CTD phase separation against changes in ionic strength and sensitivity towards 1,6-hexanediol, which is known to disrupt hydrophobic contacts[246, 252, 303, 307], indicate that CTD self-association is mediated through weak hydrophobic interactions. Hydrophobic interactions between aromatic systems have been shown to be important for LLPS of various IDRs[251, 256, 308-312] and aromatic residues are strongly conserved within the CTD sequence as well as between species (Figure 1.2b). We thus speculated whether aromatic interactions between Y_1 CTD residues could drive CTD phase separation. If, on the other hand, mere hydrophobicity drives CTD self-association, Y_1 replacement with more hydrophobic amino acids should further enhance LLPS. To test this, we generated hCTD variants, in which all 52 tyrosine residues are replaced either by leucine (Y→L) or phenylalanine (Y→F) residues that are both classified to be more hydrophobic than tyrosine (Fig. 3.5a). Replacement of tyrosine by leucine fully abolished phase separation, while the phenylalanine variant underwent LLPS in the droplet assay similar to the wild-type (Fig. 3.5b). To further probe for smaller quantitative differences, the molecular mobility was assessed using FRAP. The fluorescence recovery kinetics of both variants were identical for both variants, indicating that replacement of tyrosine by phenylalanine does not detectably affect the interaction strength between the hCTD molecules (Fig. 3.5c). Together, these results suggest a critical role of aromatic interactions and imply that the lack of these interresidue contacts cannot be simply compensated by an increase in hydrophobicity.

3.1.4 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 µM. Fluorescence microscopy showed that Pol II located to CTD droplets (Fig. 3.1i).

3.1.5 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 RPB1[208]. Such cell lines are known to recapitulate the behavior of endogenous wild-type Pol II[208, 313, 314, 315]. 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. 3.6a-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. 3.6c), confocal imaging and fluorescence-activated cell sorting (FACS; Fig. 3.6d-e). The two cell lines also grew at similar rates (Fig. 3.6f).

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. 3.7)[208, 316]. Compared to cells with full-length CTD (52R), cells with the truncated, yeast-like CTD (25R) showed less Pol II clustering (Fig. 3.7a-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, Methods). This cell line was also viable and grew at a similar rate as the other two lines upon degradation of endogenous RPB1 (Fig. 3.6f), though it expressed RPB1 at

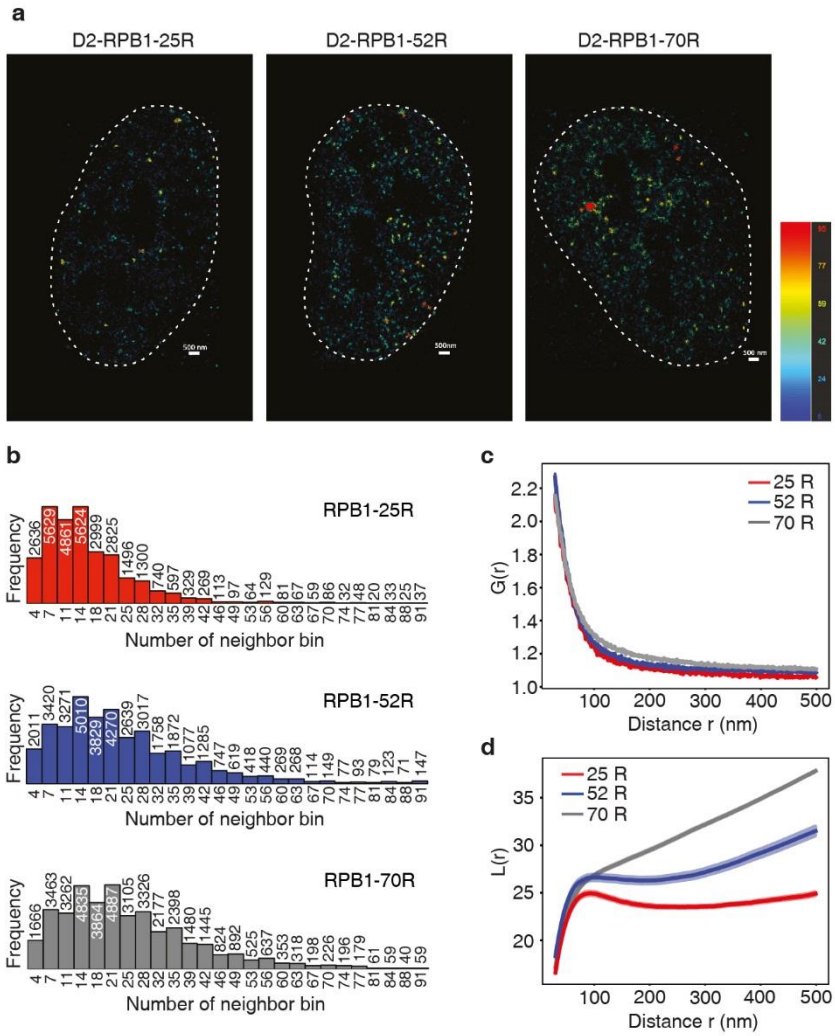

Figure 3.7 | CTD-dependent Pol II clustering in human cells.

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. 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 ($P = 1.08\ \ 10^{-21}$, 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.

a lower level (Fig. 3.6c). 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. 3.7a-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) [317]. In all cells, *L(r)* curves showed strong clustering signatures (Fig. 3.7c-d). Whereas the sharp increase observed at scales less than 100 nm can be influenced by photophysical effects, such as blinking of Dendra2[318], 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.

3.1.6 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)[293] 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. 7.1). We then tracked single molecules of Pol II in live cells as demonstrated by single-step photoactivation and photobleaching (Fig. 3.8a-b). Subsequent two-state kinetic modeling analysis assuming a free and bound state (Fig. 3.8c, Supplementary Fig. 7.2) 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 to 38.4% in RPB1-70R cells (Fig. 3.8d, Supplementary Fig. 7.2). 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 $\mu m^2/s$ were measured in RPB1-25R, RPB1-52R and RPB1-70R cells, respectively (Fig. 3.8e). These large differences in diffusion coefficients cannot be explained solely 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.

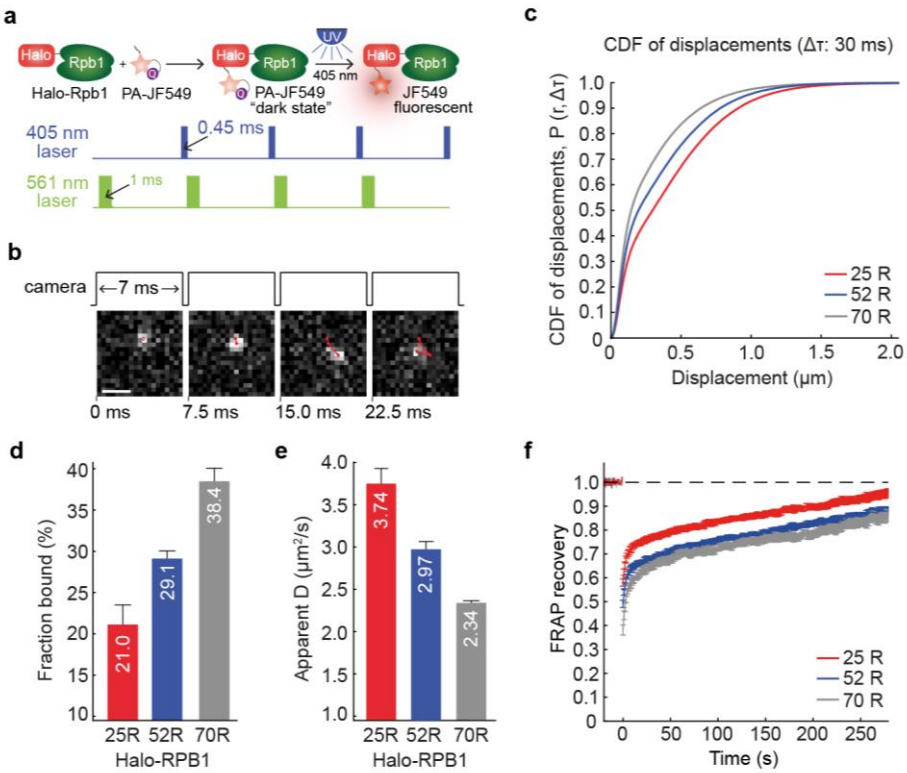

Figure 3.8 | CTD-dependent Pol II dynamics in human cells.

a, Overview of stroboscopic photoactivation SPT (spaSPT) at ~133 Hz. Halo-RPB1 labeled with photoactivatable Janelia Fluor 549 (PA-JF_{549}) 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-On[293]. 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-On[293]. 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.

These findings in cells match our observed length-dependence of CTD-CTD interactions in vitro (Fig. 3.1h; Fig. 3.5c). Indeed, FRAP recovery curves in human cells depended on CTD length (Fig. 3.8f), consistent with differences in FRAP recovery kinetics observed between hCTD and yCTD droplets in vitro (Fig. 3.1h). Analysis of these FRAP recovery curves by a reaction-dominant two-state model[214, 292] 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 to 38% in RPB1-70R cells (Supplementary Fig. 7.2). This trend is consistent with the SPT results (Fig. 3.8d), 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 flavopiridol treatment, which blocks the transition into productive elongation by targeting positive elongation factor B (P-TEFb; Supplementary Fig. 7.3). 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 chromatin association in vivo, reflecting the influence of CTD length on LLPS in vitro.

3.1.7 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 S_5 CTD residues by the cyclin-dependent kinase 7 (CDK7) in transcription factor IIH (TFIIH) stimulates the transition of Pol II into active elongation[67, 319]. We treated hCTD with recombinant human TFIIH subcomplex containing CDK7 kinase[320] and ATP, leading to S_5 phosphorylation of hCTD (Fig. 3.9a). The resulting CDK7-phosphorylated hCTD was no longer able to form droplets, whereas prior incubation with ATP alone did not inhibit LLPS (Fig. 3.9b). Phosphorylation of yCTD by the yeast TFIIH kinase subcomplex also inhibited phase separation (Fig. 3.10). In addition, phosphorylation of preformed hCTD droplets by human CDK7 caused gradual shrinking and ultimately disappearance of hCTD droplets (Fig. 3.9c). Therefore, phosphorylation at S_5 positions is incompatible with CTD phase separation and transfers the CTD from the highly concentrated state within droplets to the dispersed pool.

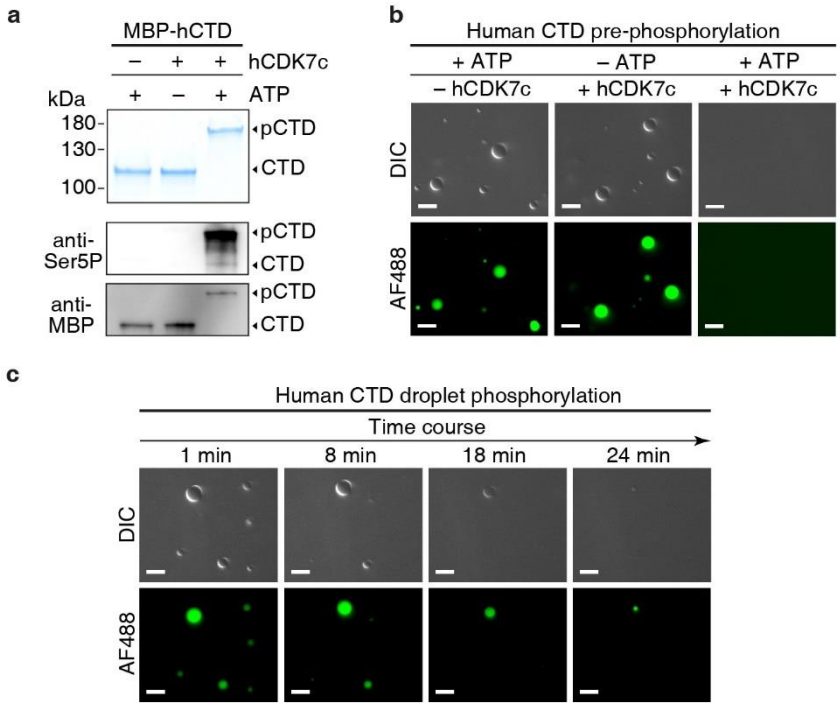

Figure 3.9 | CDK7 phosphorylation counteracts phase separation of human CTD.

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. **b**, hCTD phase separation is inhibited upon CTD phosphorylation by the human TFIIH subcomplex containing the CDK7 kinase. This effect is caused neither by hydrotropic properties of ATP[321] 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 µm.

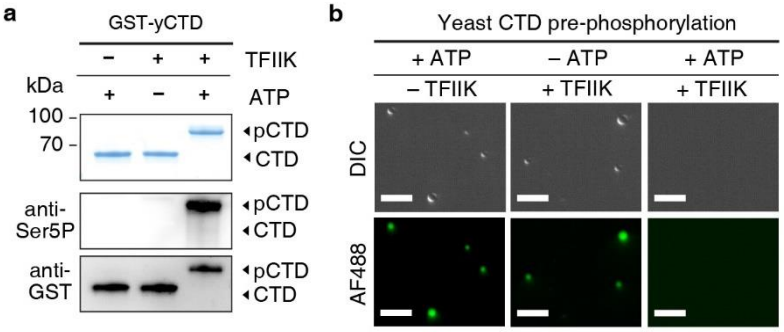

Figure 3.10 | TFIIK phosphorylation counteracts phase separation of yeast CTD.

a, 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. **b**, 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.

3.2 Discussion

Here we show that the Pol II CTD can undergo concentration-dependent liquid-liquid phase separation in vitro and that this behavior is strongly influenced by CTD length/repeat number. We further find that CTD length also controls Pol II clustering as well as its dynamics in living human cells. Our results indicate that the intermolecular interactions between CTD molecules are driven by weak hydrophobic contacts between the aromatic tyrosine residues in heptad position 1. These repeat-repeat interactions are sensitive to CTD phosphorylation by the cyclin-dependent kinase CDK7, which dissolves CTD droplets in vitro. Several implications emerge from these findings. I will first discuss direct consequences for our understanding of CTD structure and function, and then discuss broader implications for our emerging view of the spatiotemporal organization of eukaryotic gene transcription.

3.2.1 Implications for CTD structure and function

3.2.1.1 Multivalent interactions contribute to Pol II CTD function

Many previous studies suggested that CTD function is mainly based on binary interactions between short CTD regions with CTD-binding factors. These factors were found to bind 1-3 CTD heptad repeats in a transcription stage-dependent manner (reviewed in[21,22]). Together these studies led to the notion that short heptad motifs may constitute the functional unit of the Pol II CTD[322, 323]. While such binary interactions are undoubtedly an important aspect of CTD function, they do not suffice to explain the extreme overall conservation of CTD length. For example, why is a consecutive stretch of 26 consensus heptad repeats strongly conserved between yeast and human if binding sites of all known CTD binders only span three repeats or less? Instead our results indicate that CTD function is additionally based on multivalent intermolecular interactions that manifest in liquid-liquid phase separation. Phase separation is known to be strongly controlled by the interaction valency and very sensitive to small valency changes[231, 245, 251, 324]. Consistent with this, our in vitro data shows that the ability to engage in such intermolecular interactions and their strength is strongly dependent on the CTD repeat number. In live human cells, CTD truncation reduced Pol II clustering and mobility, while an artificial CTD extension beyond the wild-type length had the reverse effect. These results

suggest that weak multivalent interactions govern Pol II localization and dynamics in vivo and together emphasize their importance for CTD function.

3.2.1.2 Aromatic interactions underlie CTD phase separation

We find that replacement of all 52 tyrosine residues within the human CTD by phenylalanine still supports phase separation, while replacement by leucine abolishes it. This indicates that aromatic but not mere hydrophobic interactions drive CTD self-association. It seems likely that π-π interactions between the strongly conserved Y_1 residues play a dominant role in this context. The similar behavior of hCTD WT and Y→F variants in phase separation assays can be rationalized by results from fluorescence spectroscopy measurements, which suggest very similar binding energies for aromatic π-π interactions formed between two tyrosine and between two phenylalanine aromatic ring systems[325]. This might offer a plausible explanation for the similar behavior of hCTD WT and Y→F variants in phase separation assays. Beyond the aromatic interaction, Vernon et al.[256] pointed out that the conceptual basis of π-π interactions can furthermore be extended to all sp2 hybridized atoms that occur in the peptide bond or amino acid side chain and interact through exposed π orbitals. While not as strong π-π interactions between aromatic systems, the widespread occurrence of π-π interactions between aromatic and non-aromatic π-orbitals was shown to have profound impact on phase separation propensity[256]. Thus, the small side chain size of the CTD consensus residues serine, threonine and proline might allow for additional extensive contacts of the aromatic tyrosine residues with the π orbitals of the exposed sp2 hybridized peptide bond. Indeed, such interactions are frequently observed in crystal structures of folded proteins[256]. Taken together, our in vitro results suggest that π interactions involving the aromatic systems constitute the underlying basis for CTD phase separation.

The requirement of an aromatic residue in the first heptad position for phase separation is consistent with results from genetic in vivo studies, which found that replacement of all CTD tyrosine residues by leucine is lethal in yeast[276, 326]. Tyrosine substitution with phenylalanine was found to be tolerated in up to half of the human CTD repeats, but results in severe termination defects[327]; complete replacement of tyrosine with phenylalanine was found to be lethal in *S. cerevisae* and human cells[28, 328, 329]. As the phenolic hydroxyl group is not required

to form homotypic heptad-heptad interactions, this points toward an additional role of Y_1 in vivo: Indeed, Y_1 is well known to undergo reversible phosphorylation during the transcription cycle, which prevents termination factor recruitment to elongating Pol II[42], to contribute to CTD stability and has been implicated in upstream-antisense and enhancer RNA transcription[328, 329]. In addition, the Y_1 hydroxyl group frequently engages in favorable interactions with additional factors binding to the CTD, for example, the Y_1 hydroxyl group forms stabilizing hydrogen bonds with the capping enzyme subunit Cgt1[70] as well as with the CTD-interacting domain of Pcf11[330].

3.2.2 Implications for the organization of Pol II transcription

3.2.2.1 Unphosphorylated CTD recruits RNA Pol II to activated gene promoters

How transcription is organized inside the crowded nuclear environment has been an intense area of research[200-202]. Transcription initiation requires a complex sequence of protein-protein and protein-DNA interactions to facilitate pre-initiation complex formation[59]. However, during stimulus-activated gene transcription (for example upon heat shock), Pol II initiates at a rate that is solely limited by the elongation velocity of already engaged Pol II clearing the promoter (~40 Pol II/min)[162, 331, 332]. Recent super-resolution microscopy experiments revealed that Pol II forms transient clusters[208]. These Pol II clusters were found to precede mRNA synthesis and localize into close proximity to gene promoters[208, 211]. However, the underlying molecular mechanism of Pol II clustering could not be deciphered in these studies, hampering further functional interpretation.

Here, we have shown that the unphosphorylated RNA Pol II CTD possesses the ability to self-interact and undergo liquid-liquid phase separation in vitro. We further demonstrate that Pol II clustering strongly correlates with CTD length in live human cells. From these findings, a model emerges for Pol II recruitment to activated gene promoters (Fig. 3.10): Multivalent interactions between unphosphorylated CTDs mediate the formation of Pol II condensates inside living cells (Fig. 3.10). Super-resolution imaging techniques and computational modeling estimated these condensates to contain on average ~80 Pol II molecules[211]. Such Pol II condensates might provide a concentrated pool of initiation-competent Pol II that might help to overcome rate-limiting steps and facilitate high initiation rates upon activated gene transcription.

It is likely that transcriptional activators that bind near gene promoters or distant enhancer elements can direct the formation of Pol II condensates through their long, disordered activation domains. In fact, studies have demonstrated that activation domains of several transcription factors can self-interact through aromatic contacts to form phase-separated droplets or hydrogels in vitro[135, 245] that recruit Pol II CTD[135]. Inside the nucleus, these activation domains were also shown to form punctate condensates that co-localize with Pol II condensates[333]. Similarly, the well-known yeast activator GCN4 was recently shown to undergo LLPS through aromatic interactions[334], and CTD truncation leads to gene activation defects in yeast[30, 335] and mammalian cells[32]. Together with the strong dependence of CTD phase separation on aromatic amino acids, these results suggest a predominant role of aromatic contacts in the interaction between transcription factors and the Pol II CTD. Strikingly, results from systematic mutational screens of various different transcription factors suggest that exposed aromatic residues within the activation domain are crucial to mediate transcriptional activation[135, 229, 333, 336, 337]. It is thus an attractive possibility that aromatic interactions between the Pol II CTD and activation domains drive the co-condensation of both factors at activated gene promoters. An important consequence that follows from such interaction with DNA-bound transcription factors is that Pol II condensation could occur even below the saturation concentration if attractive interactions between the Pol II CTD and spatially constrained transcription factor activation domains increase the Pol II concentration locally above the saturation threshold. Additionally, such mechanism would constitute a simple yet effective way to ensure that Pol II condensation occurs at gene promoters only upon transcription factor binding. Consequently, the unphosphorylated CTD delivers and condenses Pol II at sites of active transcription. This idea is consistent with work from Lu et al.[338], who show that the Pol II CTD is both required and sufficient to mediate recruitment to transcriptionally active loci at *Drosophila* polytene chromosomes.

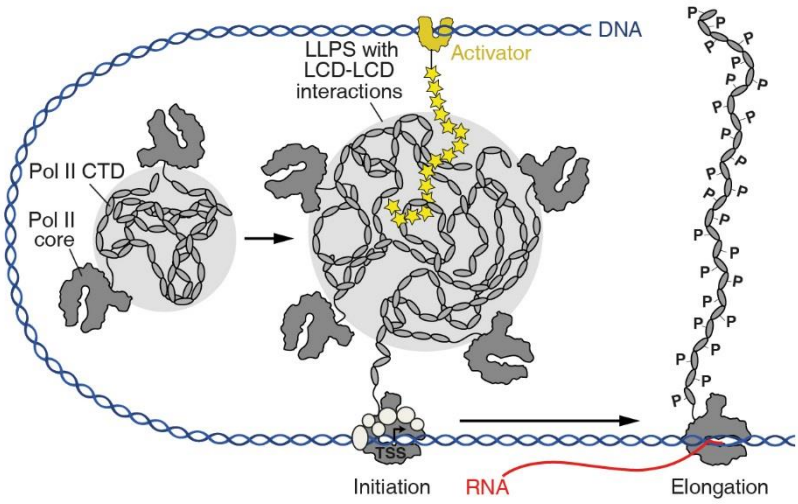

Figure 3.11 | Model for the role of CTD-driven phase separation in activated transcription.
Pol II clusters in nuclear condensates or 'hubs' through multivalent intermolecular interactions between unphosphorylated CTDs. Transcriptional activators that bind to regulatory sites such as enhancers might promote the formation of Pol II clusters through their disordered transactivation domains. High Pol II concentrations at gene promoters might facilitate efficient transcription initiation. Initiation-coupled CTD phosphorylation releases single Pol II enzymes into active early elongation. During elongation the phosphorylated CTD can form novel multivalent interactions with selected factors. Figure was adapted from Boehning et al. (2018)[25].

While our study was completed and under peer review several other factors that bind to super-enhancers at high density were described to form phase-separated condensates[275, 339]. Sabari et al.[340] provided experimental evidence that the surrogate subunit MED1 of the Mediator complex and the bromodomain-containing co-activator BRD4 form condensates at super-enhancers and that their IDRs can undergo liquid-liquid phase separation in vitro. Cho et al.[212] showed that Pol II and Mediator form large stable condensates in mouse embryonic stem cells that co-localize in a transcription-dependent manner, in addition to the small transient Pol II condensates in differentiated cells described here and elsewhere[208, 211]. Boija et al.[334] showed that various activation domains of transcription factors form phase-separated condensates that interact with Mediator. Together with previous findings[131, 132, 339], these results are consistent with the model that enhancers serve as scaffolds, on which transcription factors and transcriptional co-activators such as BRD4 and Mediator accumulate.

Ultimately, transcriptional activation requires enhancer-promoter communication[116]. Interesting results from recent live-cell microscopy studies suggest that this does not involve direct physical enhancer-promoter interaction, but that mere proximity (~100-300 nm) of enhancer and promoter suffices to trigger gene transcription[116, 121, 341]. This *action at a distance*-behavior can be rationalized in light of a promoter condensate or hub model in which Pol II (this study, and[212]) and Mediator co-activator[340], promoted by transcription factors[334], condense through weak multivalent IDR-IDR interactions between enhancer and promoter elements. The formation of a liquid-like phase-separated condensate at the gene promoter results in the local accumulation of these factors that might simultaneously act as functional and structural bridge[342]. The all-or-nothing nature of phase separation might explain rapid (dis-)assembly kinetics of the promoter condensate and can provide an underlying mechanistic basis for the enigmatic phenomenon of transcriptional bursting[343]. In such a model the condensate lifetime might then dictate the lifetime of the enhancer-promoter interaction and correlate with the bursting frequency. Whether the condensate only forms after sustained enhancer-promoter proximity[121, 341] or whether it results from coalescence of distinct condensates that might be pre-formed at enhancers (containing e.g. Mediator) and promoters (containing i.e. Pol II)[212], is still part of active research. Interestingly, the surface tension resulting from (thermodynamically favorable) coalescence of such phase-separated chromatin domains can in principle provide sufficient force to pull two genomic loci into close spatial proximity[344].

3.2.2.2 CTD phosphorylation releases Pol II from promoter condensates

Incorporation of Pol II into the pre-initiation complex results in CTD phosphorylation of S_5 residues through the TFIIH kinase CDK7 (Fig. 3.11). CDK7 phosphorylation disrupts the weak hydrophobic interactions that underlie CTD self-association and concentration of Pol II within the promoter condensate. This liberates the Pol II enzyme after transcription initiation from the promoter and allows the transition into active transcription elongation (Fig. 3.11).

During elongation, the phosphorylated CTD allows Pol II subsequently form novel multivalent interactions with selected factors. The disordered histidine-rich C-terminal region of cyclin T1, a subunit of P-TEFb, can form phase separated droplets that incorporate the CDK7-phosphorylated CTD[345]. In vivo, P-TEFb foci have long been known to co-localize with

S_5-phosphorylated Pol II, but not S_2-phosphorylated Pol II[204]. Such a mechanism ensures efficient CTD hyperphosphorylation and might enable the rapid activation of the paused Pol II elongation complex through P-TEFb[87, 345]. Additionally, hyperphosphorylated CTD was shown to be recruited to phase-separated droplets formed by splicing factors in vitro[346, 347]. Splicing factors often possess low-complexity RS-rich domains and are condensed in splicing speckles[231, 348, 349]. Splicing occurs co-transcriptionally[90, 350] and speckles locate to genomic loci with high transcriptional activity[351]. Indeed, speckles are well known to co-localize with phosphorylated Pol II[350, 352]. Thus, the co-condensation between the Phospho-CTD and SR proteins through weak multivalent interactions might underlie the coordinated coupling of transcription with mRNA splicing[346].

In summary, our results together with recently published findings from other groups suggest that the CTD can undergo phase separation through two different mechanisms. During the pre-initiation phase, homotypic intermolecular interactions with other CTD molecules (Fig. 3.1) or heterotypic interactions with transcriptional activators or coactivators promote CTD phase separation. These interactions are largely based on weak-hydrophobic (i.e. aromatic) contacts (Fig. 3.5). After promoter release and concomitant CTD phosphorylation, the weak-hydrophobic interaction network is disrupted and the CTD does not engage in homotypic intermolecular interactions anymore (Fig. 3.9-10). Upon elongation into the gene body, the phosphorylated (negatively charged) CTD rather engages in multivalent heterotypic interactions with elongation and RNA processing factors. These interactions are likely electrostatic in nature and thus physiochemically distinct from the interactions within the promoter condensate as implied by the positively charged histine-rich cyclin T1 domain or RS domains of splicing factors. Thus, CTD phosphorylation may serve as a molecular switch that regulates partitioning between ‘promoter condensates’ and ‘gene-body condensates’ (Fig. 3.9-10)[346, 347]. Upon CTD dephosphorylation during transcription termination, the Pol II enzyme is released from interactions with elongation and RNA processing factors and relocates to the promoter condensate.

4 NELF condensation accompanies stress-induced transcriptional downregulation

4.1 Results

The results presented in this section were obtained in collaboration with Prashant Rawat from the laboratory of Dr. Ritwick Sawarkar (Max Planck Institute for Immunology and Epigenetics, Freiburg) and are currently prepared for publication. Experiments that were not performed by the author of this dissertation, but are included in this section for a coherent presentation of the obtained findings, are marked in the figure legends. Detailed author contributions can be found on Page VI.

The cellular response to proteotoxic stress such as heat shock involves the downregulation of thousands of genes involved in metabolism, cell cycle and protein synthesis[162, 173, 175]. In human cells, stress-induced transcriptional downregulation is accompanied by enhanced recruitment of the negative transcription elongation factor NELF to downregulated genes and is thus thought to be regulated at the step of promoter-proximal pausing[175, 176, 189, 353]. However, the underlying molecular basis of the increased residence time of NELF at chromatin and stable Pol II pausing upon stress has remained unclear.

4.1.1 NELF concentrates in punctate structures upon stress

In order to visualize the nuclear redistribution of NELF upon heat shock, we transfected HeLa cells transiently with a plasmid encoding the mCherry-tagged NELFA subunit. Under normal conditions, NELFA was fairly homogenously distributed throughout the nucleus and only a few

nuclear puncta were visible, consistent with a previous study that revealed co-localization with histone gene loci[354]. In contrast, heat shock stress caused the rapid concentration of NELF into droplet-like structures that formed throughout the entire nucleoplasm (Fig. 4.1a). To confirm these findings, we additionally constructed a stable human cell line that expresses a GFP-tagged version of the NELFA subunit at similar levels to the endogenous protein. As in transfected cells, NELF formed numerous clusters inside the nucleus upon heat stress (Fig. 4.1b). After the heat stress ceased, NELF puncta readily dispersed within a 1 h recovery period (Fig. 4.1c). Similar to heat shock, arsenic treatment induced puncta formation, suggesting that NELF condensation is a general consequence of proteotoxic stress rather than caused by the mere temperature difference (Fig. 4.1d). The ability of NELF to concentrate in punctate structures raises the possibility that sequestration within these puncta could cause its increased residence time at chromatin. To characterize the properties of stress-induced NELF condensates, we conducted fluorescence recovery after photobleaching (FRAP) experiments in live cells. While GFP-NELFA showed rapid and complete recovery within few seconds in the absence of heat stress, its mobility decreased considerably upon heat shock (Fig. 4.1e). Such stable immobilization within puncta resembles the concentration of proteins within membraneless organelles[225, 231]. These protein condensates form through interactions between intrinsically disordered protein domains that lead to liquid-liquid phase separation into a protein-rich phase with liquid-like properties. Consistent with the liquid-like behavior, we observed puncta with shapes that were reminiscent of fusion events (Fig. 4.1f). Many intracellular condensates are stabilized by weak interactions that are disrupted by the aliphatic alcohol 1,6-hexanediol[307, 355]. To further explore whether NELF puncta possess properties consistent with phase-separated condensates, we treated the cells with 1,6-hexanediol after exposure to heat stress. Similar to other membraneless organelles, 1,6-hexanediol treatment strongly reduced the number of puncta, further supporting the idea that the same mechanisms may underlie their formation.

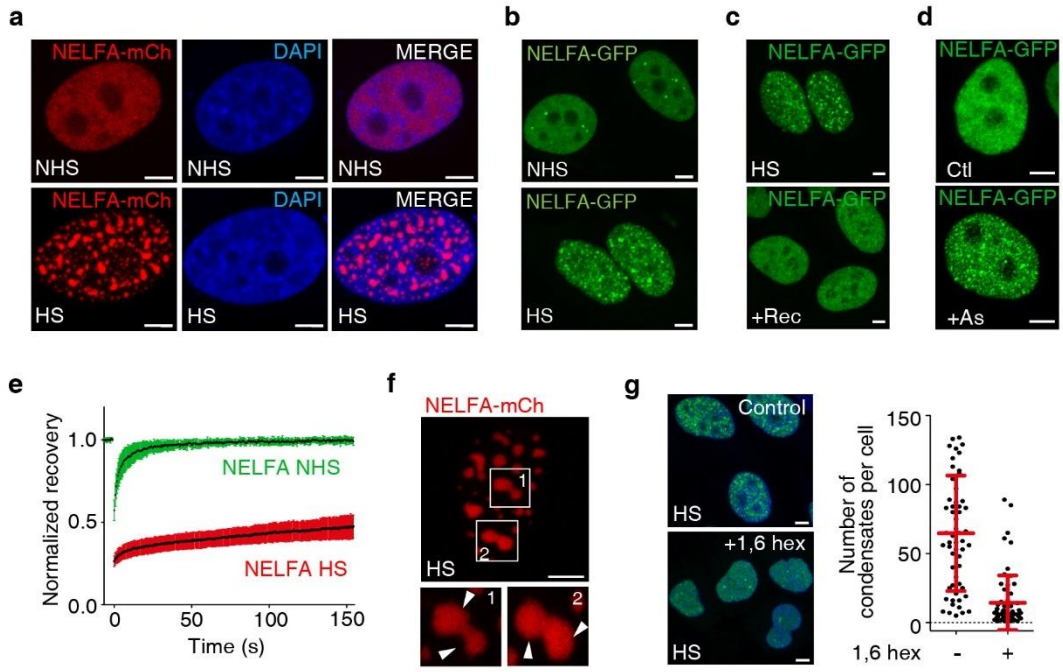

Figure 4.1 | NELF forms puncta upon stress that resemble phase-separated condensates.

a, Confocal image of human HeLa cells that were transfected with NELFA-mCherry fusion protein. While NELFA-mCherry is homogenously distributed in the absence of stress (NHS), it condenses into numerous puncta upon heat shock (HS). **b**, Confocal image of HeLa cells that stably express NELFA-GFP as single copy gene under a tetracycline-inducible promoter. The number of NELFA puncta drastically increases upon heat shock. **c,** Heat stress-induced NELFA puncta reversed (+Rec) upon 60 min incubation at 37°C. **d,** Similar to heat shock, arsenic treatment (+As) induces the formation of NELF-GFP puncta. **e,** Photobleaching-corrected FRAP recovery curves. While NELFA-GFP readily recovers in the absence of heat stress, its mobility is drastically reduced upon heat shock. Curves are representative for 10 (NHS) and 12 (HS) FRAP experiments. Error bars show the standard deviation. **f,** Stress-induced nuclear puncta formed by transiently transfected NELFA-mCherry undergo fusion. **g,** NELF puncta are sensitive to 1,6-hexanediol. HeLa cells that stably express NELFA-GFP were heat shocked (43°C) in the presence or absence of 10% 1,6-hexanediol. Representative confocal images are shown. Analysis based on 57 cells (HS – 1,6 hex) and 61 cells (HS + 1,6 hex), respectively. Boxes indicate mean and standard deviation. Scale bars, 5 µm in all panels. All experiments shown in this figure were conducted by Prashant Rawat (MPI for Immunology and Epigenetics, Freiburg).

4.1.2 NELF is capable of liquid-liquid phase separation in vitro

Resident proteins of various membraneless organelles have been shown to undergo phase separation in vitro. Prompted by the results we obtained in vivo, we wanted to investigate whether liquid-liquid phase separation could underlie NELF immobilization at chromatin. To test whether NELF is capable of undergoing liquid-liquid phase separation in vitro, we recombinantly expressed the human full-length four-subunit NELF complex in insect cells.

Previous biochemical studies provided strong evidence that binding of unphosphorylated NELF induces Pol II pausing, while NELF phosphorylation causes its displacement from Pol II leading to pause release[76, 87]. To mimic the paused state, we dephosphorylated the NELF complex during its purification to homogeneity with Lambda phosphatase and labeled the complex with an amine-reactive Alexa Fluor 488 dye at a low labeling density of ~1 fluorophore per molecule (Fig. 4.2a-b). We then used fluorescence microscopy to assess the ability of the purified NELF complex to undergo phase separation using an in vitro phase separation assay.

At a sodium chloride concentration of 50 mM which is commonly used to examine LLPS in vitro [262, 356], a 5 µM NELF solution underwent phase separation to form numerous spherical micron-sized droplets (Fig. 4.2c). Droplets formed in the bulk solution and settled onto the coverslip surface due to gravity. With increasing salt concentration, the number and size of the droplets decreased (Fig. 4.2c). This strong dependence of phase separation on ionic strength suggests that electrostatic interactions are required for NELF self-association. While NELF underwent phase separation in the absence of any crowding agents at low ionic strength, the inclusion of low amounts of the polysaccharide dextran promoted NELF phase separation under near-physiological salt conditions (Fig. 4.2d). At a salt concentration of 50 mM, NELF formed droplets with a critical concentration of ~0.5 µM (Fig. 4.2e). Upon contact, droplets coalesced into larger droplets that readily relaxed to spherical shape (Fig. 4.2f). This underpins the liquid-like nature of NELF droplets that form in vitro and is reminiscent of the properties of NELF condensates that are observed in living human cells (Fig. 4.1f). We then further examined the dynamics of molecules within the condensed phase using FRAP. Rapid recovery of fluorescence within the bleached region demonstrated that NELF molecules can freely diffuse within the condensed phase (Fig. 4.2g). It has been shown for a few phase-separating systems that the molecular mobility within the condensed phase can decrease over time, an irreversible process that is referred to as ‘maturation’ and implicated in several neurodegenerative diseases[303, 310, 357, 358]. However, identical recovery kinetics even after prolonged incubation for 20 h suggest that NELF droplets maintain their liquid-like nature over longer time scales in vitro (data not shown). Taken together, our results indicate that NELF molecules can self-associate to form liquid-like droplets in vitro, providing a potential mechanistic basis for stress-induced NELF condensation in vivo.

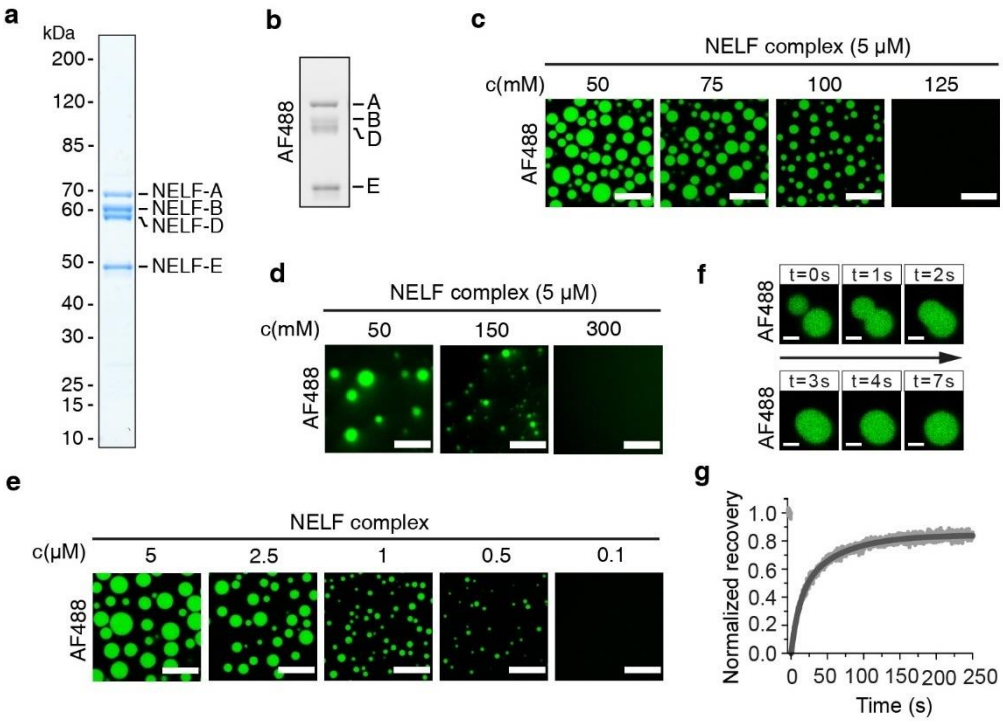

Figure 4.2 | NELF undergoes liquid-liquid phase separation in vitro.

a, Coomassie-stained SDS-PAGE of purified NELF complex (~1 µg). The four subunits of the complex NELFA (57.3 kDa), NELFB (65.7 kDa), NELFD (isoform of NELFC that lacks the first nine amino acids; 66.2 kDa), NELFE (43.2 kDa) are indicated. **b**, SDS-PAGE fluorescence scan of NELF complex after chemical labeling using an amine-reactive Alexa Fluor 488 (AF488) dye. NELFA and NELFE subunits become preferentially labelled. **c,** Confocal microscopy reveals the formation of spherical droplets in dependence on the sodium chloride concentration. Concentration of NELF complex was 5 µM. **d,** Epifluorescence microscopic images of NELF phase separation in the presence of low amounts of molecular crowding agents. Addition of dextran to a final concentration of 3% promotes phase separation of a 5 µM NELF solution at near-physiological ionic strength. **e,** Confocal microscopy images of NELF phase separation at 50 mM NaCl. NELF forms spherical droplets with a critical concentration of 0.5 µM. **f,** Droplet fusion. Upon contact two NELF droplets coalescence into one larger droplet consistent with liquid-like behavior. **g,** Photobleaching-corrected and normalized FRAP recovery curves. Bleached fluorescence within a 1 µm circular spot distant from the droplet boundary recovers within seconds. The points show mean and standard error of three independent replicates and were fit with a double-exponential recovery curve. Scale bars, 20 µm in **c**, **d**, **e**, and 2 µm in **f**.

4.1.3 NELF tentacles drive phase separation in vitro

NELF consists of the two heterodimeric subcomplexes NELFA/C and NELFB/E, which associate via contacts between NELFB and NELFC to form a three-lobed structure[76, 156]. While the NELFB and NELFC subunits are largely composed of structured domains, NELFA and

NELFE possess large flexible C-terminal regions that were termed 'tentacles'[74, 87] (Fig. 4.3). Multivalent interactions between intrinsically-disordered protein domains are well-known to promote LLPS of various proteins. To elucidate whether the flexible tentacles drive phase separation of NELF, we expressed both regions as GFP fusion proteins in *Escherichia coli* and purified them to homogeneity (Fig. 4.4a). We then used the pre-established in vitro droplet assay to test whether the GFP-tentacle fusion proteins can individually undergo phase separation under the same conditions that promote LLPS of the NELF complex. Both fusion proteins alone did not undergo phase separation, not even at very high concentrations (Fig. 4.4b). This result indicates that the tentacles are not able to self-interact. However, when NELFA and NELFE tentacle fusion proteins were mixed at equimolar concentrations, they readily formed droplets at much lower concentrations (Fig. 4.4b). Instead of mixing both tentacles *in trans* we then aimed to combine them *in cis* to mimic their arrangement within the complex. For this we generated another GFP fusion protein that contains the NELFE tentacle at its N-terminus and the NELFA tentacle at its C-terminus (Fig. 4.4a). Genetic fusion of both tentacles to the same polypeptide reduced the critical concentration for phase separation 8-fold to about 5 µM (Fig. 4.4c). These results indicate that both NELF tentacles can form multivalent interactions with each other that lead to phase separation with increasing protein concentration. Since both tentacles are sufficient to promote phase separation of a GFP fusion tag independent of the NELF core, it further implies that the synergistic interaction between the flexible tentacles drives NELF phase separation. Although both tentacles are in principle sufficient to drive phase separation, the saturation concentration difference to the wild-type NELF complex suggests that the remaining parts of the complex ('NELF core') might also support phase separation, consistent with other examples in the literature[134, 241, 245]. In addition, the transfer of the C-terminal NELFE tentacle to the GFP N-terminus might compromise its ability to engage in productive intermolecular interactions.

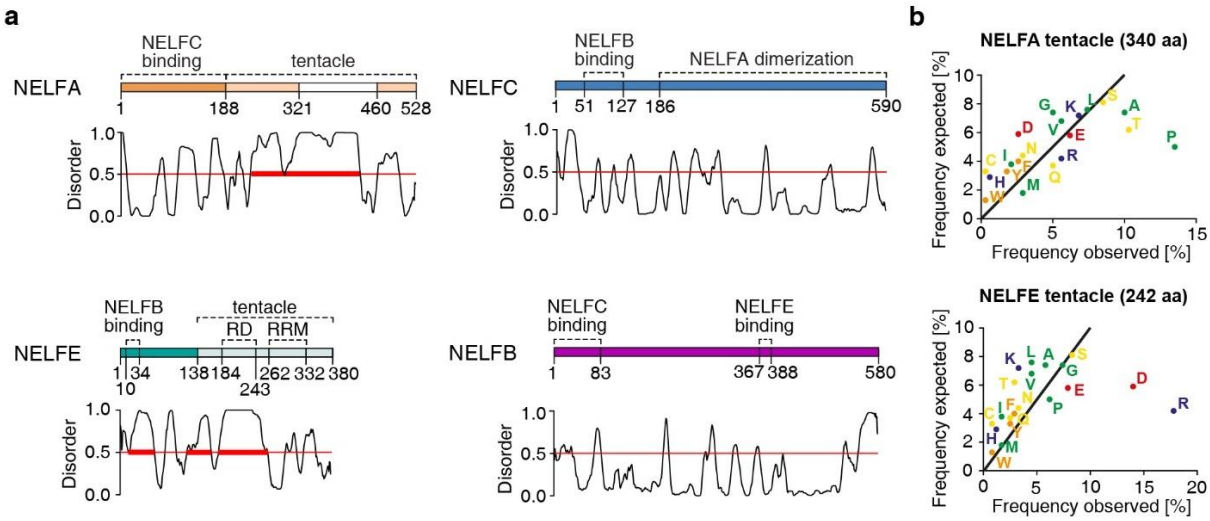

Figure 4.3 | NELFA and NELFE subunits possess disordered tentacles at the C-terminus.

a, Domain architecture and disorder prediction of all four NELF subunits using the PONDR tool[24]. Subunits NELFA and NELFE possess long C-terminal regions with high disorder propensity, while NELFB and NELFC subunits are composed of structured protein domains. The disordered NELFA region comprising residues 189-528 and the disordered NELFE region with residues 139-380 are referred to as 'tentacles'[76]. The definition of the protein domains is taken from Vos et al. (2018)[76]. RD, arginine-aspartate repeat domain, RRM, RNA recognition motif. **b,** Sequence analysis of both NELF tentacles. The relative abundance of each amino acid was plotted against the relative abundance within vertebrate proteins[359]. Amino acids are color-coded based on their physicochemical properties: Negatively charged, red; positively charged, blue; polar, yellow; aromatic, orange; aliphatic, green. Proline (P) is particularly enriched within the NELFA tentacle, while arginine (R) and aspartate (D) are enriched within the NELFE tentacle sequence.

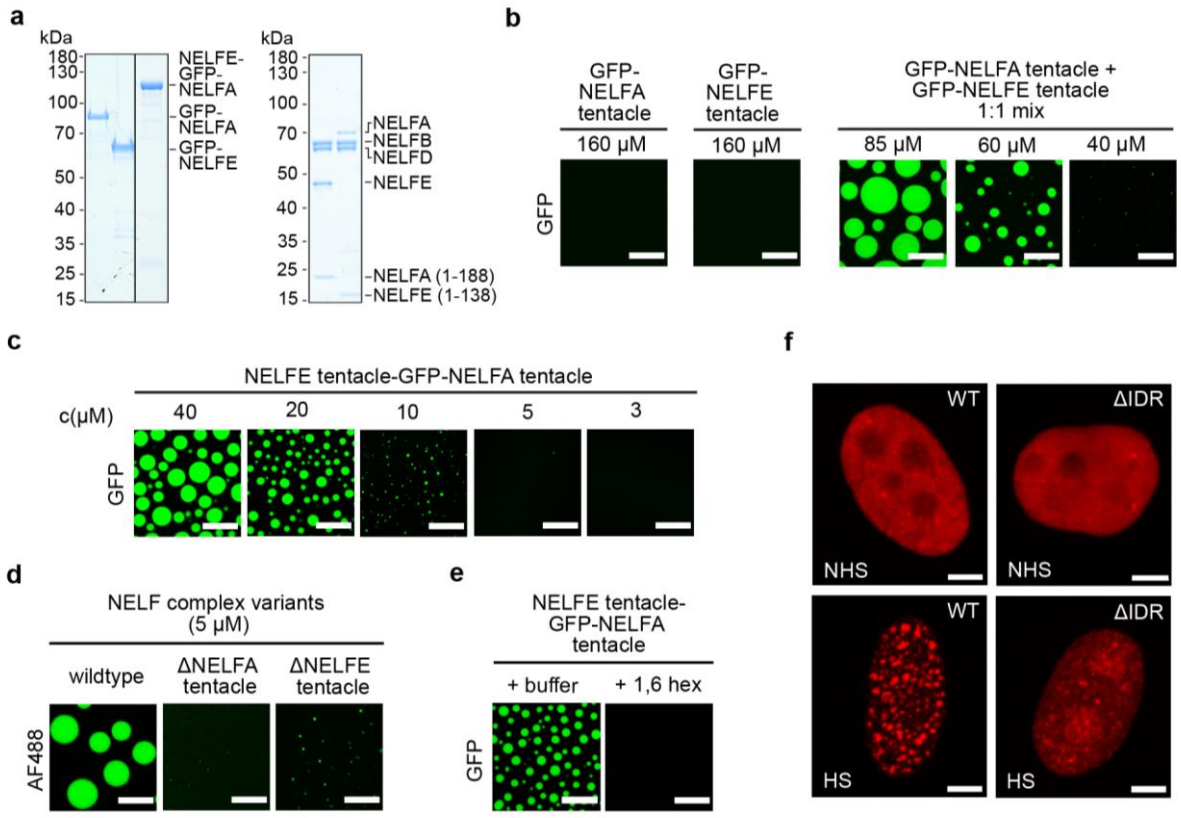

Figure 4.4 | Flexible tentacles drive NELF phase separation.

a, Coomassie-stained SDS-PAGE of GFP-NELF tentacle fusion proteins and truncated NELF complex variants. For single tentacle variants, sequences encoding the NELFA tentacle (NELFA residues 189-528) or the NELFE tentacle (NELFE residues 139-380) were fused individually to the GFP C-terminus. For double tentacle fusion proteins, the sequence encoding the NELFE tentacle was fused to the GFP N-terminus and the NELFA tentacle to the C-terminus. NELF complex truncation variants lack either the NELFA or NELFE tentacle region. About 1 µg of each protein was loaded for SDS-PAGE analysis. **b,** Phase separation of GFP-NELF tentacle fusion proteins. At a concentration of 160 µM GFP-NELFA tentacle or GFP-NELFE tentacle individually did not undergo LLPS in buffer containing 50 mM sodium chloride. If both fusion proteins were mixed at equimolar concentrations, phase separation was observed with a critical concentration of 40 µM. **c,** Phase separation of a double tentacle-GFP fusion protein. The fusion protein contains the NELFE tentacle at its N-terminus and the NELFE tentacle at its C-terminus. Fusion of both tentacles to the same polypeptide further decreased the critical concentration to ~5 µM. **d,** Effect of tentacle deletion on NELF complex phase separation. Deletion of either tentacle strongly attenuates phase separation of a 5 µM NELF solution. **e,** 1,6-hexanediol sensitivity of tentacle-driven phase separation. The addition of 10% 1,6-hexanediol inhibits phase separation of the double tentacle GFP fusion protein. **f,** Effect of NELFA IDR deletion on stress-induced condensation in vivo. Stress-induced NELF condensation into nuclear puncta is inhibited upon deletion of a highly disordered segment (amino acids 321-460, Fig. 4.3a) within the NELFA tentacle in transiently transfected HeLa cells. This experiment was conducted by Prashant Rawat (MPI for Immunology and Epigenetics, Freiburg). Scale bars, 20 µm in **b**, **c**, **d**, **e**, and 5 µm in **f**.

The interdependence of both tentacles implies that the removal of either tentacle should heavily impair phase separation of the NELF complex. In order to test this, we produced two complex variants that lack either the NELFA tentacle or the NELFE tentacle (Fig. 4.4a). Compared to the wild-type complex, phase separation of both variants was heavily attenuated (Fig. 4.4d), strongly supporting our previous findings.

We could not test whether phase separation of the full-length NELF complex is sensitive towards 1,6-hexanediol because the complex precipitated upon addition of the aliphatic alcohol in vitro. Our previous findings suggest that tentacle-tentacle interactions primarily drive NELF phase separation. We thus used the double tentacle GFP fusion protein as surrogate to probe whether the multivalent interactions formed between the NELF tentacles are sensitive to 1,6-hexanediol. Indeed, addition of 10% 1,6-hexanediol fully abolished phase separation of the GFP double tentacle fusion protein (Fig. 4.4e), consistent with our observations in vivo (Fig. 4.1g).

4.1.4 NELF tentacles drive condensation in vivo

Based on our findings in vitro, we wanted to investigate next whether multivalent interactions between both tentacles may also underlie NELF condensation in vivo. Following an identical rationale as in the in vitro droplet assay, we aimed to inhibit the formation of intermolecular interactions through deletion of the entire NELFA tentacle. This was however not feasible because the NELFA tentacle harbors a nuclear localization signal (residues 268-277) that is essential for nuclear import (data not shown). We therefore deleted a smaller segment of 140 amino acids that is predicted to be highly disordered (residues 321-460; Fig. 4.3a) and named this variant ‘ΔIDR’. Deletion of this region did not impair the nuclear localization of the ΔIDR NELFA-mCherry fusion protein under normal growth conditions; however, it strongly attenuated NELF condensation upon heat shock (Fig. 4.4f). This is consistent with the interdependent interaction of both tentacles that drives phase separation in vitro (Fig. 4.4d). Thus, our combined results indicate that the NELFA- and NELFE-tentacles can engage in multivalent interactions with each other, which underlie NELF phase separation in vitro as well as condensation in vivo.

4.1.5 NELF dephosphorylation promotes phase separation

We then searched for the mechanism that could cause NELF condensation under heat shock conditions. It is well established that binding of unphosphorylated NELF to Pol II induces pausing and subsequent NELF phosphorylation by P-TEFb leads to its displacement and pause release[74, 76, 87]. We therefore wanted to test the effect of P-TEFb phosphorylation on NELF phase separation. For this we produced recombinant NELF complex that was treated with active P-TEFb during purification. We then analyzed untreated and P-TEFb-treated NELF complex using phosphate-affinity SDS-PAGE, which reduces the electrophoretic mobility of phosphorylated proteins due to the interaction with gel-embedded Phos-tag molecules[360]. Mobility shifts were observed for NELFA and NELFE, but not for the NELFB and NELFC/D subunits, indicating that primarily NELFA and NELFE are phosphorylated by P-TEFb (Fig. 4.5a). Using mass spectrometry-based phosphorylation site mapping, we found 21 different P-TEFb phosphorylation sites that were exclusively located on the NELFA and NELFE subunits (Fig. 4.5b). In addition, the detected sites were located predominantly in disordered regions (compare to Fig. 4.3a), consistent with previous analyses[87]. Post-translational modifications that alter the physicochemical properties of the underlying amino acids sequence are emerging as key mechanism to regulate IDR-IDR interactions. We thus reasoned that P-TEFb phosphorylation within the flexible NELFA and NELFE tentacle regions might influence their ability to interact in a multivalent manner with each other.

In order to test the effect of P-TEFb phosphorylation on NELF phase separation in vitro, we formed NELF droplets at low ionic strength and in the presence of magnesium chloride and ATP. After the droplets had settled on the coverslip surface, we added either active wild-type P-TEFb or a catalytically inactive P-TEFb variant[87] to the preformed NELF droplets and tracked changes over time using fluorescence microscopy. Time-resolved imaging revealed no change in droplet diameter over a period of 2 h when inactive (kinase-dead) P-TEFb was added. In contrast, addition of active wild-type P-TEFb resulted in significant shrinkage of NELF droplets within the same time frame (Fig. 4.5c). This result implies that P-TEFb phosphorylation inhibits NELF phase separation. This hypothesis is further supported by FRAP analyses of droplets formed either with dephosphorylated or P-TEFb-treated NELF. The faster fluorescence recovery of P-TEFb-treated NELF suggests that phosphorylation weakens the interaction strength between the tentacles and renders the molecules more mobile within the

condensed phase (Fig. 4.5d-f). While the molecules are generally mobile within the condensed phase (Fig. 4.5d), their exchange with the bulk solution over the phase boundary is much slower. While about 80% of phosphorylated NELF molecules exchanged with the surrounding solution after 20 min, the recovered fraction was lower than 40% for an equally-sized droplet formed by dephosphorylated NELF (Fig. 4.5e-f). This slower exchange illustrates the stronger intermolecular interactions that deplete non-phosphorylated NELF from the bulk solution and concentrate it inside droplets. Consistent with this, pre-phosphorylation of the double tentacle fusion protein with active P-TEFb resulted in smaller and fewer droplets that often deviated from the shape of an ideal sphere (Fig. 4.5g). Together, the combined in vitro data thus demonstrates that P-TEFb phosphorylation counteracts NELF phase separation.

Since NELF puncta emerge upon heat shock and NELF dephosphorylation promotes its self-association, the phosphorylation level of P-TEFb target sites on NELF should decrease upon heat shock in vivo. We affinity-purified NELF from heat-stressed and non heat-stressed cells and analyzed NELF phosphorylation sites using mass spectrometry. Three different phosphorylation sites on NELFA were detected in these experiments. They were mostly P-TEFb target sites (Fig. 4.5b) and their phosphorylation level decreased upon heat shock (Fig. 4.5h). The result suggests that unphosphorylated NELF accumulates upon heat shock in vivo and is consistent with a model in which this increasing NELF fraction is involved in pause stabilization that underlies the genome-wide transcriptional downregulation upon heat shock[76, 175].

4.1.6 Heat shock stress causes P-TEFb inactivation in the 7SK snRNP complex

To investigate the regulatory mechanism that prevents NELF phosphorylation during heat shock, we focused on the pause release kinase P-TEFb. In cells, P-TEFb exists in a dynamic equilibrium of a free active form that becomes inactivated upon incorporation into the 7SK snRNP complex[168, 169]. This raises the possibility that enhanced sequestration of P-TEFb within the inactivating 7SK snRNP complex upon heat shock could lead to reduction of NELF phosphorylation and accompanying Pol II pause-release. To test whether the ratio between active and inactive P-TEFb might change upon heat shock, we conducted a CDK9 pulldown in

the presence or absence of heat stress, respectively, and examined CDK9-interaction partners using a SILAC-based quantitative mass spectrometry approach (Fig. 4.5i). Association with (co-)chaperone complexes HSP90/Cdc37 and HSP70 that bind a significant proportion of CDK9 inside cells[361] did not change. Contrary to this, the fraction of CDK9 that was associated with components of the 7SK snRNP complex was ~6 times higher upon heat shock (Fig. 4.5i). These findings indicate that heat shock stress causes increased P-TEFb sequestration in the inactivating 7SK snRNP complex.

4.1.7 NELF sumoylation enhances condensation

The previous results suggested that P-TEFb inactivation could be sufficient to induce NELF condensation in cells. To explore this possibility, we treated cells with the chemical CDK9 inhibitor DRB. However, DRB treatment did not induce the formation of NELF puncta in the absence of heat stress (not shown), indicating that the accumulation of dephosphorylated NELF is a required prerequisite, but not sufficient for puncta formation. Heat shock stress is known to cause extensive post-translational modification of nuclear proteins[362-364]. Prominently, modification of nuclear proteins with SUMO2/3 (Small ubiquitin-like modifier 2/3) upon heat shock was shown to be required for cellular survival[363]. Reanalysis of published proteome-wide datasets of stress-triggered sumoylation revealed that all four NELF subunits become modified with SUMO2/3 upon heat shock[363, 364] (Fig. 4.6a).

We thus wanted to explore whether sumoylation could be required for stress-induced NELF condensation. SUMO is attached through an enzymatic cascade involving an E1 activating enzyme, an E2 conjugating enzyme and an E3 ligase that ultimately modifies the target protein. In order to block the first step of the enzyme cascade, we treated cells with the E1 inhibitor ML-792. Considerably less NELF puncta formed upon heat shock (Fig. 4.6b), indicating that sumoylation is required for puncta formation in vivo. Consistent with this finding, published data suggests that the sumoylation machinery becomes enriched at chromatin upon heat shock (Aprile-Garcia et al. (2019)[175], Supplementary Dataset 4).

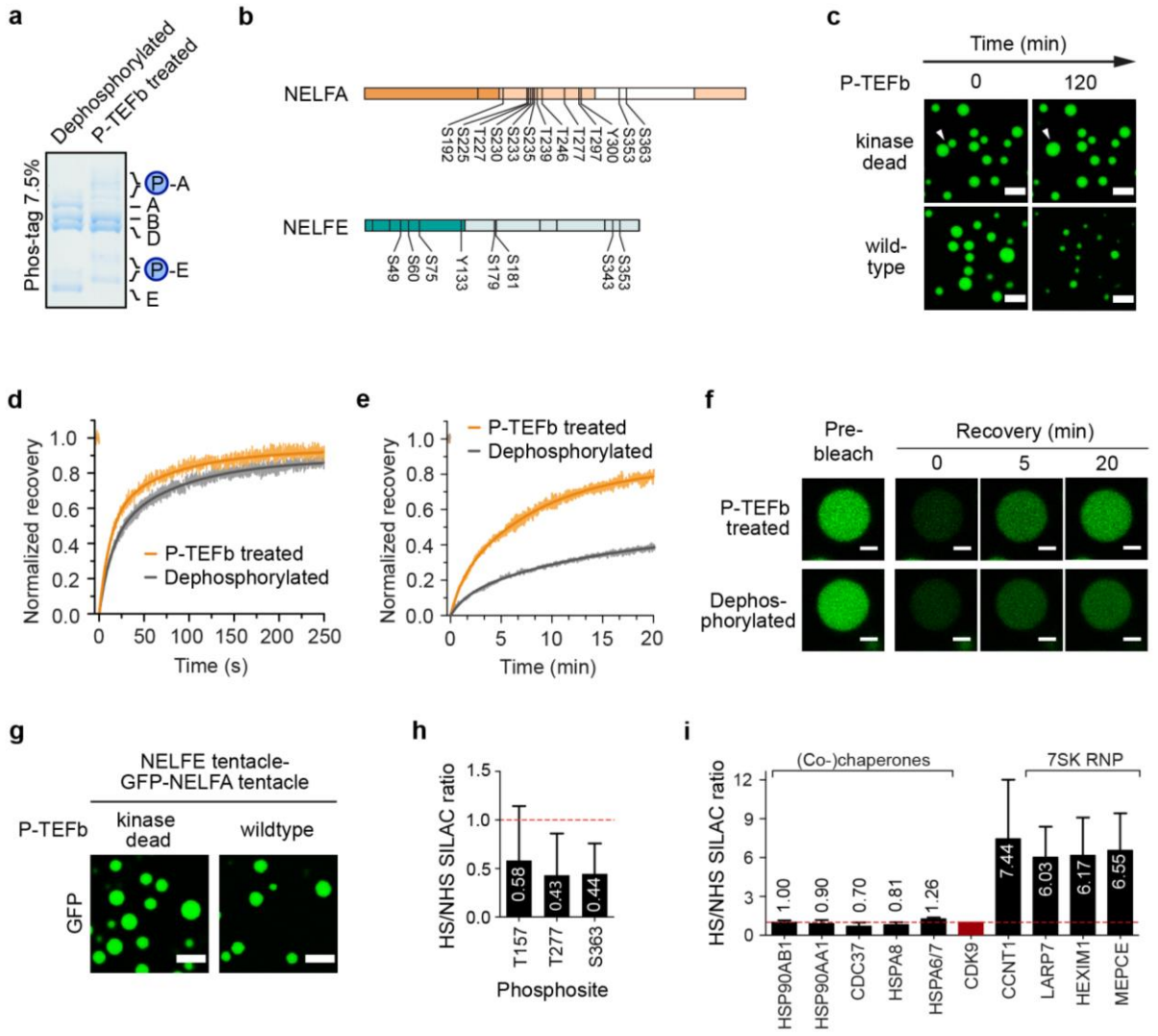

Figure 4.5 | Effect of P-TEFb phosphorylation on NELF phase separation.

a, Phosphate-affinity SDS-PAGE analysis of NELF complex treated with P-TEFb in vitro. Gel-embedded Phos-tag molecules cause retardation of phosphorylated proteins. P-TEFb phosphorylates the NELFA and NELFE subunits. **b,** Mass spectrometry-based phosphorylation site mapping. In total, 21 different phosphorylation sites were detected after titanium dioxide-based phosphopeptide enrichment exclusively on the NELFA and NELFE subunits. The majority of P-TEFb phosphorylation sites were localized in flexible NELF regions. **c,** Time-resolved effect of P-TEFb phosphorylation on NELF phase separation. NELF droplets were either incubated with active wild-type P-TEFb or a catalytically inactive P-TEFb variant and then imaged in regular intervals. After incubation with active P-TEFb for 120 min the size of NELF droplets decreased considerably. The arrow indicates a droplet fusion event. Scale bar, 5 µm. **d,** Photobleaching-corrected and normalized FRAP recovery curves for partial droplet bleaching experiments. A 1 µm circular region in the droplet interior was bleached. Lines represent mean and standard error of five bleached droplets and the recovery was fit with a double exponential function. *(Figure caption continued on next page.)*

◀ **Figure 4.5 | Effect of P-TEFb phosphorylation on NELF phase separation.** *(Figure caption continued from previous page.)*

e, Photobleaching-corrected and normalized FRAP recovery curves for full droplet bleaching experiments. Full droplets with equivalent diameter were photobleached. Lines represent mean and standard error of three bleached droplets. The recovery was fit to a double exponential model. **f,** Exemplary confocal images of the recovery of unphosphorylated and P-TEFb-treated NELF droplets in a full FRAP experiment. Scale bar, 2 µm. **g**, Effect of P-TEFb phosphorylation on phase separation of GFP-fusion protein containing NELFA and NELFE tentacle regions. Pre-phosphorylation of the GFP fusion protein with active P-TEFb decreased the number and size of formed droplets. Droplets that formed from phosphorylated protein often possessed non-spherical, 'crumbled' shapes. Scale bar, 10 µm. **h**, Mass spectrometric quantification of NELFA phosphorylation sites. Heat shock caused reduction of the phosphorylation level of three NELFA sites, two of them being P-TEFb target sites (compare Fig. 4.5b). Bars show mean and standard deviation for three independent replicates. **i,** Mass spectrometric quantification of CDK9 interactors in the presence or absence of heat shock. Bars show mean and standard deviation for three independent replicates. Experiments shown in panels **h** and **i** were conducted by Prashant Rawat (MPI for Immunology and Epigenetics, Freiburg).

Prominently, the known E3 ligase zinc-finger protein 451 (ZNF451) exhibits an almost 6-fold higher chromatin association compared to steady state conditions (Aprile-Garcia et al.[175], Supplementary Dataset 4). We thus asked whether ZNF451 could sumoylate NELF in vitro. To reconstitute the sumoylation reaction in vitro, we incubated E1, E2 and ZNF451 together with NELF, ATP and SUMO2. Indeed, western blot analysis revealed the modification of the NELF subunits NELFA, NELFC and NELFE with SUMO2 (Fig. 4.6c).

How might sumoylation promote NELF phase separation? The tendency of a protein to undergo phase separation strongly depends on the valency of the intermolecular interaction. The covalent modification with (poly-)SUMO can provide additional interaction interfaces that bind short SUMO-interacting motifs (SIMs) to increase the effective interaction valency. We analyzed the NELF sequence for the presence of potential SIMs. Interestingly, the NELFE tentacle harbors a predicted SIM with the sequence ^{286}IIDL^{289} that was shown before in other proteins to mediate the interaction with SUMO2/3[365, 366]. To investigate the possibility that sumoylation enhances NELF phase separation in vitro, we formed NELF droplets in the presence of ATP and added the E1/E2/E3 SUMOylation machinery. Due to the low reaction efficiency at conditions that support phase separation in vitro we could not observe any effect on NELF phase separation (not shown). In an orthogonal approach, we thus tested whether NELF can interact with SUMO2/3. Using fluorescence anisotropy assays, no significant binding could be detected between 1xSUMO and NELF up to a concentration of 10 µM, indicating that potential NELF-SUMO interactions must be weak-affine. To specifically probe

for the presence of such weak interactions, we formed NELF droplets and added substoichiometric amounts of Alexa Fluor 647-labeled 4xSUMO, 1xSUMO or MBP. Indeed, at high protein concentrations that exist within a condensed phase, NELF incorporated 4xSUMO and to a lesser extent 1xSUMO, while MBP was excluded (Fig. 4.6d). This shows that NELF, at high protein concentrations, is able to specifically interact *in trans* with SUMO2/3 in a chain length-dependent manner. Analogous to results we obtained for the tentacles (Fig. 4.4b-c), covalent attachment of (poly-)SUMO2/3 should thus further enhance the interaction strength and promote self-interaction of the NELF complex.

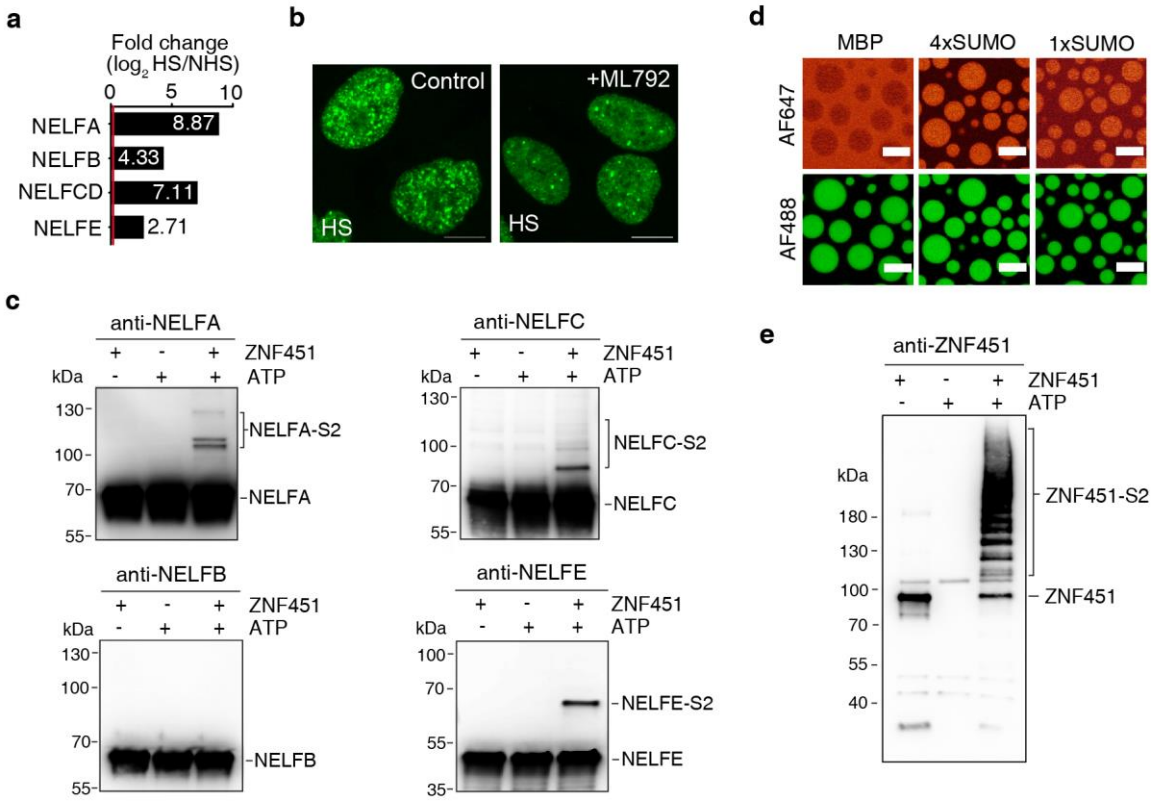

Figure 4.6 | NELF sumoylation enhances condensation.

a, Mass spectrometric analysis of the human SUMO proteome in HeLa cells revealed that the NELF complex undergoes sumoylation upon heat shock. Plot is based on data published by Hendriks et al. (2017) (Supplementary Table 3)[364]. **b,** Effect of inhibition of the SUMO E1 activating enzyme. Treatment of HeLa cells with the SUMO E1 inhibitor ML-792 reduced the number of NELF puncta that formed upon heat shock. This experiment was conducted by Prashant Rawat (MPI for Immunology and Epigenetics, Freiburg). **c**, In vitro reconstituted sumoylation reaction with NELF and the SUMO E3 ligase ZNF451. Western blotting using subunit-specific antibodies revealed that ZNF451 can sumoylate the NELFA, NELFC, and NELFE subunits. **d,** NELF droplet partitioning analysis. Substoichiometric amounts of Alexa Fluor 647-labelled H_6-MBP-N_{10}-TEV (44.8 kDa), HA-$(SUMO3)_4$ (43.9 kDa), or SUMO2 (10.7 kDa) were added to NELF prior to induction of LLPS. SUMO2 and SUMO3 are almost identical isoforms (sequence identity: 94.6% (88/93 residues), sequence similarity: 96.8% (90/93 residues), calculation is based on the recombinant protein). While tetra-SUMO3 became enriched in NELF droplets, similar-sized MBP was excluded. Compared to tetrameric SUMO3, monomeric SUMO2 showed weaker enrichment, suggesting that valency is the key factor that determines partitioning. **e**, Autosumoylation of ZNF451. Western blotting revealed that ZNF451 polysumoylates itself in reconstituted in vitro assays, consistent with published results[295].

4.2 Discussion

Previous studies indicated that stress-induced transcriptional downregulation is mediated by the enhanced recruitment of the negative elongation factor NELF to chromatin[175]. NELF accumulation near gene promoters is thought to stabilize promoter-proximal pausing of Pol II[175]. However, the molecular basis for the increased residence time of NELF at chromatin has remained unclear. Here we provide evidence that a phase separation mechanism can explain the increased dwell time of NELF at chromatin upon stress. We show that multivalent interactions between the disordered NELF tentacles lead to the concentration of NELF in phase-separated, liquid-like droplets in vitro and are essential for stress-induced NELF condensation in vivo. P-TEFb phosphorylation counteracts NELF phase separation in vitro and is prevented through the inactivation of P-TEFb upon heat shock in vivo. Our results imply that sumoylation is further required for stress-induced NELF condensation. NELF itself can be sumoylated in vitro and interacts with SUMO2/3 in a chain length-dependent manner, representing a potential mechanism how sumoylation might enhance condensation. Although our data cannot yet demonstrate that paused Pol II is present within stress-induced NELF condensates, the strong increase of NELF and SUMO2/3 occupancy near downregulated gene promoters upon heat shock, which both correlate with enhanced Pol II pausing at these genes suggest a mechanistic link[175, 353, 367, 368]. Together with these published findings[175, 353, 367, 368], our data suggest a model that involves the stress-induced sequestration of promoter-proximally paused Pol II through NELF condensates at downregulated genes.

Model for stress-induced transcriptional downregulation

After promoter release, NELF binds together with DSIF to the early Pol II elongation complex within the promoter-proximal region. In cooperation with DSIF, NELF mediates Pol II pausing through the stabilization of a nonproductive DNA-RNA hybrid conformation and prevents escape from the paused state[76, 158]. Under conditions of cellular homeostasis at steady state, paused Pol II is readily released into productive elongation. During proteotoxic stress, however, Pol II becomes stably paused at downregulated gene promoters through the accumulation of NELF[175]. We suggest that a phase separation mechanism can account for the increased residence time of NELF at downregulated gene promoters. Our data implies that two

mechanisms promote the formation of NELF condensates. In the following I present a model, which proposes that the accumulation of dephosphorylated NELF under heat stress might inevitably trigger polysumoylation.

During stress, the available pool of active P-TEFb is reduced through an increased sequestration within the 7SK snRNP complex (Fig. 4.5i). Consequently, this results in the accumulation of dephosphorylated NELF (Fig. 4.5h). In the non-phosphorylated state, NELF has a stronger tendency to self-interact (Fig. 4.5c-g) so that the fractional increase of the dephosphorylated form might be sufficient to overcome the saturation threshold. Alternatively, it is an intriguing possibility that other disordered domains of the paused elongation complex such as the Pol II CTD could provide a localized scaffold for the subcritical condensation of dephosphorylated NELF (see also Section 3.2.2). Such a model assumes that weak NELF-scaffold interactions can locally increase the NELF concentration over saturation and implies that condensation is spatially limited to the range of this interaction. Independent of the precise mechanism, both possibilities could lead to the accumulation of NELF around the paused elongation complex. Similar to other transcriptional condensates[211], such clusters might initially consist only of a limited number (≤100) of molecules, representing a potential reason why they are not visible using diffraction-limited microscopy techniques.

Growth of stress-induced NELF condensates over the optical detection limit additionally requires the functional sumoylation machinery (Fig. 4.6). The NELF subunits NELFA, NELFC and NELFE become (poly-)sumoylated upon stress (Fig. 4.6a), mediated by the SUMO E3-ligase ZNF451 (Fig. 4.6c). How might the E3-ligase reach its substrate NELF within the nucleus? Under normal conditions ZNF451 resides in PML bodies[369], which partially disassemble upon heat stress[370, 371]. ZNF451 itself is polysumoylated[295, 369] (Fig. 4.6e) and phase-separated NELF droplets interact with SUMO2/3 in a chain length-dependent manner, presumably through a putative SIM motif within the NELFE tentacle (Fig. 4.6d). Thus, a plausible mechanism could be that small clusters of dephosphorylated NELF that formed due to P-TEFb inactivation can trap ZNF451 through SUMO-SIM interactions. Within this condensate, large NELF concentrations might promote high catalytic efficiency[372]. Based on our previous results (Fig. 4.4b-c), it is likely that covalent polysumoylation further enhances NELF self-association through the presence of additional SUMO-SIM interactions. This would in turn lead to increased recruitment of the E3-ligase, creating a positive feedback loop that

guarantees a rapid response. Growing condensates on the same and neighboring genes might coalesce to form larger assemblies. High NELF concentrations might assure stable Pol II pausing under stress conditions. At the same time, the phase boundary of the condensate increases the residence time of NELF within the promoter-proximal region and might constitute a selective barrier that restricts the access of elongation factors (Fig. 4.6c).

The model highlights the functional synergism between P-TEFb sequestration and ZNF451 mediated sumoylation for NELF condensation and stress-induced stabilization of Pol II pausing. According to the model, P-TEFb inactivation on its own could lead to formation of small clusters with few molecules that cannot grow further, while available ZNF451 can only be trapped efficiently through high local NELF concentrations that do not build up in the presence of active P-TEFb. This could allow signal integration from multiple orthogonal stress-triggered pathways. Such a co-incidence detection mechanism would increase the robustness against intracellular fluctuations and noise and could ensure that a pervasive transcriptional response, resulting in the downregulation at the majority of genes becomes only elicited upon receiving multiple cues.

The combination of heat shock experiments in human cells with orthogonal in vitro phase separation assays in this study has expanded our understanding about the mechanisms that might underlie the genome-wide transcriptional downregulation upon stress. It revealed the mechanistic basis of stress-induced condensation of NELF through an intricate interplay between P-TEFb sequestration and sumoylation. These findings can guide future investigations that may provide corroborative evidence. Imaging-based approaches shall be used in order to confirm that Pol II is present within NELF condensates and that condensates are associated with downregulated genes. Furthermore, it will be important to understand to which extent NELF condensation promotes stable Pol II pausing. For this, integrative multiomic approaches will be required that facilitate the accurate estimation of Pol II pause duration on a genome-wide scale[162]. However, this question is inherently difficult to answer as the deletion of protein regions that affect condensation might at the same time also influence NELF association with Pol II. In vitro pausing assays[76] in combination with in vivo imaging approaches used in this study might help to elucidate appropriate regions. One such region could be the putative SUMO-interacting motif within the NELFE tentacle. If this motif is required for E3-ligase

recruitment, its removal should attenuate stress-induced NELF condensation in vivo. It is located in a flexible loop region within the RRM domain, which refolds and might become inaccessible upon RNA binding[373, 374]. RNA binding is not required for Pol II pausing and pause release[76, 87], and it is currently unclear in which conditions NELFE associates with RNA in vivo[152, 156]. Future PAR-CLIP experiments, which allow the identification of RNA binding sites transcriptome-wide[375], may provide insights on how RNA binding contributes to NELF function.

The interplay of the NELF tentacles with other disordered domains within the paused elongation complex, such as the Pol II CTD, remains an interesting aspect that should be explored in the future. Early biochemical studies showed that Pol II with CTD that was pre-phosphorylated by P-TEFb is resistant to NELF-DSIF-induced pausing in an in vitro transcription assay, because CTD phosphorylation impairs the association of NELF with Pol II[148]. These results suggest that the Pol II CTD can interact in a phosphorylation state-specific manner with NELF[148]. In this regard, it is plausible that multivalent interactions between the disordered NELF tentacles and the Pol II CTD contribute to NELF-Pol II association. In vitro droplet assays that probe co-recruitment of differentially phosphorylated Pol II CTD constitute an ideal experimental set-up to help to answer this question in the future.

Although we find that heat stress leads to increased P-TEFb sequestration in the inactive 7SK complex, the underlying signaling mechanism is not yet clear. It could potentially involve the kinase p38α that becomes activated upon heat stress[376] and translocates from the cytoplasm to down-regulated gene promoters in the nucleus[175], post-translational modification of CDK9[377] and/or other complex members[378], or post-transcriptional modification of the 7SK snRNA[379]. Similarly, the signaling pathways that regulate the availability of the SUMO E3-ligase ZNF451 are not known and should be investigated in the future.

Finally, it will be important to understand how NELF-dependent downregulation is coordinated with HSF1-driven transcriptional upregulation of pro-survival genes[176, 177]. Stress-induced transcriptional activation at the *Drosophila* HSP70 locus leads to the formation of a ‘transcription compartment’ that concentrates Pol II together with other positive elongation factors, retains them for multiple rounds of transcription and is dependent on poly(ADP-ribose) polymerase activity[185, 188]. In retrospect, these data suggest the control of stress-induced transcriptional activation at the *Drosophila* HSP70 locus through an analogous phase separation

mechanism. Interestingly, heat shock factor HSF1 is known to directly recruit P-TEFb to these compartments[182], representing a potential way as to how the active kinase can still be delivered efficiently to activated loci despite globally decreasing levels (Fig. 4.5i). The coordination of stress-induced activation and inactivation as well as the differential allocation of factors should be examined using multi-color imaging of surrogate proteins combined with assays that probe their partitioning into reconstituted phases in vitro.

Together with the results presented here, these experimental endeavors will allow us to gain a holistic understanding about the fundamental mechanisms that govern the coordinated redistribution of the transcriptional machinery upon stress in the near future.

5 Conclusion and outlook

Research over the last few years has brought a sea change in our understanding of the spatiotemporal organization of eukaryotic gene transcription (reviewed in[191, 347]). Classical textbook models describe for example gene activation as the sequential recruitment of single Pol II enzymes to the target gene promoter. In contrast, recent experiments provided compelling evidence that gene activation rather involves the transient formation of multiprotein assemblies, which may contain about 80 Pol II enzymes[211]. These findings could be rationalized through an increasing molecular mechanistic understanding of the intermolecular interactions that may stabilize such extremely large multiprotein assemblies. Liquid-liquid phase separation that is based on weak multivalent interactions between disordered protein domains has emerged as a fundamental organizational principle to concentrate proteins in living cells[225, 231, 235].

In this thesis, I provide evidence that phase separation mechanisms control key aspects of transcriptional regulation in eukaryotic cells. In all eukaryotes, gene transcription is highly regulated at the step of initiation, mediated through the recruitment of the Pol II machinery[107]. Based on the results presented in Chapter 3, it appears that the ability of the disordered Pol II CTD to engage in weak multivalent interactions in a phosphorylation state-specific manner underlies Pol II accumulation at activated gene promoters through a phase separation mechanism. In higher eukaryotes, gene transcription is additionally regulated during early elongation by promoter-proximal pausing[73]. Because accumulation of paused Pol II simultaneously blocks new initiation, this allows for rapid gene downregulation during cellular stress responses. Heat shock leads to the recruitment of the pausing factor NELF to downregulated gene promoters[175] and is accompanied by its IDR-dependent concentration in phase-separated condensates that may sequester paused Pol II (Chapter 4). Together with other recently published studies[135, 212, 333, 334, 340, 345, 346, 380], and in contrast to current textbook models, the findings presented here indicate that many factors involved in gene transcription may function in condensates, and that these large multiprotein assemblies are stabilized by dynamic, weak and multivalent rather than stable lock-and-key interactions.

Future research should strive to provide a more complete picture of the interactions that underlie condensate formation, examine their constituent proteins, and characterize the dynamic interplay of different transcriptional condensates. Below I highlight certain aspects, which might in particular merit further investigation.

5.1 Uncovering the molecular details of CTD-activator interactions in promoter condensates

Upon gene activation, initiation-competent Pol II with an unphosphorylated CTD assembles together with transcriptional (co-)activators in 'promoter condensates'[347]. So far, all experimental efforts were focused on the interaction of CTD with only very few different transactivation domains[134, 135, 343, 381]. The human proteome, however, comprises over 1,600 different transcription factors[113], with many of them containing disordered transactivation domains[227]. Future research should thus strive for a more comprehensive picture of CTD-activator interactions: It is likely that not all activators interact equally well with unphosphorylated CTD[382]; rather, different types of activation domains might exist that preferentially recruit distinct promoter condensate components. Based on activation domain sequence motifs together with their binding preference at enhancer or promoter regions[112, 339], it may be possible to obtain different transcription factor classes that can then be tested experimentally in phase separation assays for their ability to interact with the Pol II CTD.

Although disordered domains such as the CTD and activation domains continuously sample a heterogeneous population of different conformations, it is possible that transient secondary structure elements exist[383-385] and may become stabilized upon activator-CTD interaction[386, 387]. To which extent residual secondary structure elements form within the condensed phase of liquid droplets and may contribute to the interaction specificity is debated[139, 311, 388-391]. NMR spectroscopy can be used to investigate the structure of the CTD in the dispersed phase, and probe for potential structural differences in the condensed phase. This may then provide the starting point to analyze the molecular basis of heterotypic interactions between the CTD and different activation domain classes in the future.

5.2 Refining the components of gene-body condensates

Upon CTD phosphorylation, Pol II escapes from promoter condensates and a launched convoy of polymerases enters the gene body[220]. Nuclear condensates that associate with transcribing Pol II with a hyperphosphorylated CTD and nascent RNA should be better investigated in the future. Such condensates are collectively termed 'gene-body condensates'[347]. Among these are splicing speckles – phase-separated condensates that contain high concentrations of splicing factors with arginine-serine (RS)-rich low-complexity domains – that are known to associate with sites of active transcription[348, 351]. It was suggested[347] and shown[346] that hyperphosphorylated CTD can engage in multivalent interactions with such splicing factor condensates. Beyond splicing speckles, other such nuclear domains exist: Similar phase-separated protein assemblies called paraspeckles form in a transcription-dependent manner at active loci[392-394]. Paraspeckles are mostly composed of RNA-binding proteins and their formation is strongly dependent on RNA, which is known to promote phase separation by enhancing their intermolecular interactions[303, 358]. Although the function of paraspeckles is not entirely understood, they have been suggested to be important for 3'-end processing of the nascent RNA[395]. Indeed, cleavage factor I (CFI), a key component of the 3'-end processing machinery composed of CPSF5 and CPSF6 (or the shorter isoform CPSF7)[396, 397], is known to be concentrated within paraspeckles[394, 398, 399]. Analogous to splicing speckles, paraspeckles might thus correspond to condensates that provide a concentrated pool of 3'-end processing factors.

Although the precise molecular basis is not well understood, it is clear that an intricate network of multiple binary and multivalent protein-protein and protein-RNA interactions underlies mRNA 3'-end processing in eukaryotes[330, 400, 401]. Intriguingly, CPSF6 possesses a long disordered C-terminus that encompasses a proline-rich region and a RS-like domain[397], similar to splicing factors. Whether CPSF6 condensates can potentially associate with hyperphosphorylated CTD to couple 3'-end processing and transcription, analogous to splicing factors, merits further investigation. Support for such a model comes from the analysis of a prion protein that compartmentalizes 3'-end processing factors in liquid-like condensates to promote polyadenylation efficiency in *Arabidopsis thaliana*[402].

5.3 Understanding the dynamics of transcriptional condensates

The organization of transcription is dynamic and transcriptional condensates can be extremely transient structures[191, 347]. For example, promoter condensates with Pol II and Mediator form and disassemble within few seconds at most genes[208, 211, 212]. The all-or-nothing nature of cooperative liquid-liquid phase separation can provide the molecular basis for such fast (dis-) assembly kinetics. Whether condensation of Pol II at gene promoters underlies the hitherto enigmatic phenomenon of transcriptional bursting[217-219, 343], remains to be explored in future experiments. It will also be important to elucidate the events that precede condensation: Whether condensates form only after sustained enhancer-promoter proximity in between[121, 341] or whether they are pre-formed at enhancers that only transiently contact the target promoter[212], warrants further investigation.

The cellular response to stress intriguingly demonstrates the dynamics that can underlie genome-wide transcriptional reorganization. To mediate broad transcriptional downregulation upon heat shock, the rapid condensation of negative elongation factor NELF near repressed gene promoters may sequester and stabilize promoter-proximal Pol II in ‘pausing condensates’. It has been proposed that paused Pol II may correspond to a transition intermediate between promoter and gene-body condensates[347]. Stabilization of the transition state through high NELF concentrations could thus assure stable pausing during stress conditions. NELF binding to Pol II is abolished upon P-TEFb mediated CTD hyperphosphorylation[148], suggesting that only promoter-proximal Pol II is amendable to NELF sequestration. Which additional factors and whether also promoter condensate components are present within pausing condensates necessitates thorough future investigation. Future work should also explore how the formation of pausing condensates at downregulated genes influences the localization of gene-body condensates[351]. How such transcriptional re-organization is triggered on a molecular level and how it affects or is affected by the three-dimensional organization of the genome[403, 404] will require additional dedicated research.

Together, these endeavors will bring us yet one step closer to deciphering the ‘molecular grammar’ that underlies condensate formation, and will ultimately allow us to attain an improved understanding of the organization of eukaryotic gene transcription, the fundamental process that shapes cellular identity and, therefore, life itself.

6 Bibliography

1. Wilson, E.B. THE STRUCTURE OF PROTOPLASM. *Science (New York, N.Y.)* **10**, 33-45 (1899).
2. Watson, J.D. & Crick, F.H.C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737-738 (1953).
3. Gurdon, J.B., Laskey, R.A. & Reeves, O.R. The developmental capacity of nuclei transplanted from keratinized skin cells of adult frogs. *Journal of embryology and experimental morphology* **34**, 93-112 (1975).
4. Crick, F. Central Dogma of Molecular Biology. *Nature* **227**, 561-563 (1970).
5. Young, R.A. Control of the embryonic stem cell state. *Cell* **144**, 940-954 (2011).
6. Steitz, T.A. A mechanism for all polymerases. *Nature* **391**, 231-232 (1998).
7. Cramer, P., Armache, K.J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G.E., Dengl, S., Geiger, S.R., Jasiak, A.J., Jawhari, A., Jennebach, S., Kamenski, T., Kettenberger, H., Kuhn, C.D., Lehmann, E., Leike, K., Sydow, J.F. & Vannini, A. Structure of Eukaryotic RNA Polymerases. *Annual review of biophysics* **37**, 337-352 (2008).
8. Chambon, P. Eukaryotic nuclear RNA polymerases. *Annual review of biochemistry* **44**, 613-638 (1975).
9. Sharifi, S. & Bierhoff, H. Regulation of RNA Polymerase I Transcription in Development, Disease, and Aging. *Annual review of biochemistry* **87**, 51-73 (2018).
10. Willis, I.M. & Moir, R.D. Signaling to and from the RNA Polymerase III Transcription and Processing Machinery. *Annual review of biochemistry* **87**, 75-100 (2018).
11. Paule, M.R. & White, R.J. Survey and summary: transcription by RNA polymerases I and III. *Nucleic acids research* **28**, 1283-1298 (2000).
12. Haag, J.R. & Pikaard, C.S. Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nature reviews. Molecular cell biology* **12**, 483-492 (2011).
13. Cramer, P., Bushnell, D.A. & Kornberg, R.D. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science (New York, N.Y.)* **292**, 1863-1876 (2001).
14. Kornberg, R.D. Eukaryotic transcriptional control. *Trends in cell biology* **9**, M46-49 (1999).
15. Reed, R. & Hurt, E. A Conserved mRNA Export Machinery Coupled to pre-mRNA Splicing. *Cell* **108**, 523-531 (2002).
16. Jackson, R.J., Hellen, C.U. & Pestova, T.V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature reviews. Molecular cell biology* **11**, 113-127 (2010).
17. Hantsche, M. & Cramer, P. The Structural Basis of Transcription: 10 Years After the Nobel Prize in Chemistry. *Angewandte Chemie International Edition* **55**, 15972-15981 (2016).
18. Corden, J.L., Cadena, D.L., Ahearn, J.M., Jr. & Dahmus, M.E. A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America* **82**, 7934-7938 (1985).

19. Allison, L.A., Moyle, M., Shales, M. & Ingles, C.J. Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell* **42**, 599-610 (1985).
20. Buratowski, S. Progression through the RNA polymerase II CTD cycle. *Molecular cell* **36**, 541-546 (2009).
21. Meinhart, A., Kamenski, T., Hoeppner, S., Baumli, S. & Cramer, P. A structural perspective of CTD function. *Genes & development* **19**, 1401-1415 (2005).
22. Eick, D. & Geyer, M. The RNA polymerase II carboxy-terminal domain (CTD) code. *Chemical reviews* **113**, 8456-8490 (2013).
23. Stiller, J.W. & Hall, B.D. Evolution of the RNA polymerase II C-terminal domain. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 6091-6096 (2002).
24. Romero, P., Obradovic, Z., Li, X., Garner, E.C., Brown, C.J. & Dunker, A.K. Sequence complexity of disordered protein. *Proteins: Structure, Function, and Bioinformatics* **42**, 38-48 (2001).
25. Boehning, M., Dugast-Darzacq, C., Rankovic, M., Hansen, A.S., Yu, T., Marie-Nelly, H., McSwiggen, D.T., Kokic, G., Dailey, G.M., Cramer, P., Darzacq, X. & Zweckstetter, M. RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nature structural & molecular biology* **25**, 833-840 (2018).
26. Zehring, W.A., Lee, J.M., Weeks, J.R., Jokerst, R.S. & Greenleaf, A.L. The C-terminal repeat domain of RNA polymerase II largest subunit is essential in vivo but is not required for accurate transcription initiation in vitro. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 3698-3702 (1988).
27. Kim, W.Y. & Dahmus, M.E. The major late promoter of adenovirus-2 is accurately transcribed by RNA polymerases IIO, IIA, and IIB. *The Journal of biological chemistry* **264**, 3169-3176 (1989).
28. West, M.L. & Corden, J.L. Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics* **140**, 1223-1233 (1995).
29. Scafe, C., Chao, D., Lopes, J., Hirsch, J.P., Henry, S. & Young, R.A. RNA polymerase II C-terminal repeat influences response to transcriptional enhancer signals. *Nature* **347**, 491-494 (1990).
30. Allison, L.A. & Ingles, C.J. Mutations in RNA polymerase II enhance or suppress mutations in GAL4. *Proceedings of the National Academy of Sciences* **86**, 2794-2798 (1989).
31. Bartolomei, M.S., Halden, N.F., Cullen, C.R. & Corden, J.L. Genetic analysis of the repetitive carboxyl-terminal domain of the largest subunit of mouse RNA polymerase II. *Molecular and cellular biology* **8**, 330-339 (1988).
32. Gerber, H.P., Hagmann, M., Seipel, K., Georgiev, O., West, M.A., Litingtung, Y., Schaffner, W. & Corden, J.L. RNA polymerase II C-terminal domain required for enhancer-driven transcription. *Nature* **374**, 660-662 (1995).
33. Smith, L.J., Fiebig, K.M., Schwalbe, H. & Dobson, C.M. The concept of a random coil: Residual structure in peptides and denatured proteins. *Folding and Design* **1**, R95-R106 (1996).
34. Pappu, R.V., Wang, X., Vitalis, A. & Crick, S.L. A polymer physics perspective on driving forces and mechanisms for protein aggregation. *Archives of Biochemistry and Biophysics* **469**, 132-141 (2008).

35. Meredith, G.D., Chang, W.H., Li, Y., Bushnell, D.A., Darst, S.A. & Kornberg, R.D. The C-terminal domain revealed in the structure of RNA polymerase II. *Journal of molecular biology* **258**, 413-419 (1996).
36. Tsai, K.L., Sato, S., Tomomori-Sato, C., Conaway, R.C., Conaway, J.W. & Asturias, F.J. A conserved Mediator-CDK8 kinase module association regulates Mediator-RNA polymerase II interaction. *Nature structural & molecular biology* **20**, 611-619 (2013).
37. Schilbach, S., Hantsche, M., Tegunov, D., Dienemann, C., Wigge, C., Urlaub, H. & Cramer, P. Structures of transcription pre-initiation complex with TFIIH and Mediator. *Nature* **551**, 204-209 (2017).
38. Portz, B., Lu, F., Gibbs, E.B., Mayfield, J.E., Rachel Mehaffey, M., Zhang, Y.J., Brodbelt, J.S., Showalter, S.A. & Gilmour, D.S. Structural heterogeneity in the intrinsically disordered RNA polymerase II C-terminal domain. *Nature communications* **8**, 15231 (2017).
39. Chapman, R.D., Heidemann, M., Albert, T.K., Mailhammer, R., Flatley, A., Meisterernst, M., Kremmer, E. & Eick, D. Transcribing RNA Polymerase II Is Phosphorylated at CTD Residue Serine-7. *Science (New York, N.Y.)* **318**, 1780-1782 (2007).
40. Hsin, J.P., Sheth, A. & Manley, J.L. RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing. *Science (New York, N.Y.)* **334**, 683-686 (2011).
41. Komarnitsky, P., Cho, E.J. & Buratowski, S. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes & development* **14**, 2452-2460 (2000).
42. Mayer, A., Heidemann, M., Lidschreiber, M., Schreieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D. & Cramer, P. CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science (New York, N.Y.)* **336**, 1723-1725 (2012).
43. Zhang, M., Wang, X.J., Chen, X., Bowman, M.E., Luo, Y., Noel, J.P., Ellington, A.D., Etzkorn, F.A. & Zhang, Y. Structural and kinetic analysis of prolyl-isomerization/phosphorylation cross-talk in the CTD code. *ACS chemical biology* **7**, 1462-1470 (2012).
44. Voss, K., Forne, I., Descostes, N., Hintermair, C., Schuller, R., Maqbool, M.A., Heidemann, M., Flatley, A., Imhof, A., Gut, M., Gut, I., Kremmer, E., Andrau, J.C. & Eick, D. Site-specific methylation and acetylation of lysine residues in the C-terminal domain (CTD) of RNA polymerase II. *Transcription* **6**, 91-101 (2015).
45. Sims, R.J., 3rd, Rojas, L.A., Beck, D.B., Bonasio, R., Schuller, R., Drury, W.J., 3rd, Eick, D. & Reinberg, D. The C-terminal domain of RNA polymerase II is modified by site-specific methylation. *Science (New York, N.Y.)* **332**, 99-103 (2011).
46. Sharma, P., Lioutas, A., Fernandez-Fuentes, N., Quilez, J., Carbonell-Caballero, J., Wright, R.H.G., Di Vona, C., Le Dily, F., Schuller, R., Eick, D., Oliva, B. & Beato, M. Arginine Citrullination at the C-Terminal Domain Controls RNA Polymerase II Transcription. *Molecular cell* **73**, 84-96.e87 (2019).
47. Buratowski, S. The CTD code. *Nature structural biology* **10**, 679-680 (2003).
48. Schuller, R., Forne, I., Straub, T., Schreieck, A., Texier, Y., Shah, N., Decker, T.M., Cramer, P., Imhof, A. & Eick, D. Heptad-Specific Phosphorylation of RNA Polymerase II CTD. *Molecular cell* **61**, 305-314 (2016).

49. Suh, H., Ficarro, S.B., Kang, U.B., Chun, Y., Marto, J.A. & Buratowski, S. Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II. *Molecular cell* **61**, 297-304 (2016).
50. Corden, J.L. Pol II CTD Code Light. *Molecular cell* **61**, 183-184 (2016).
51. Zhang, J. & Corden, J.L. Phosphorylation causes a conformational change in the carboxyl-terminal domain of the mouse RNA polymerase II largest subunit. *The Journal of biological chemistry* **266**, 2297-2302 (1991).
52. Laybourn, P.J. & Dahmus, M.E. Transcription-dependent structural changes in the C-terminal domain of mammalian RNA polymerase subunit IIa/o. *The Journal of biological chemistry* **264**, 6693-6698 (1989).
53. Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. & Richmond, T.J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-260 (1997).
54. Struhl, K. Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell* **98**, 1-4 (1999).
55. Lorch, Y., LaPointe, J.W. & Kornberg, R.D. Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell* **49**, 203-210 (1987).
56. Lorch, Y. & Kornberg, R.D. Chromatin-remodeling for transcription. *Quarterly reviews of biophysics* **50**, e5 (2017).
57. Peterson, C.L. & Workman, J.L. Promoter targeting and chromatin remodeling by the SWI/SNF complex. *Current opinion in genetics & development* **10**, 187-192 (2000).
58. Lee, T.I. & Young, R.A. Transcription of eukaryotic protein-coding genes. *Annual review of genetics* **34**, 77-137 (2000).
59. Sainsbury, S., Bernecky, C. & Cramer, P. Structural basis of transcription initiation by RNA polymerase II. *Nature reviews. Molecular cell biology* **16**, 129-143 (2015).
60. Steitz, T.A. & Steitz, J.A. A general two-metal-ion mechanism for catalytic RNA. *Proceedings of the National Academy of Sciences* **90**, 6498-6502 (1993).
61. Brueckner, F., Ortiz, J. & Cramer, P. A movie of the RNA polymerase nucleotide addition cycle. *Current Opinion in Structural Biology* **19**, 294-299 (2009).
62. Vassylyev, D.G., Vassylyeva, M.N., Perederina, A., Tahirov, T.H. & Artsimovitch, I. Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* **448**, 157-162 (2007).
63. Dvir, A. Promoter escape by RNA polymerase II. *Biochimica et biophysica acta* **1577**, 208-223 (2002).
64. Holstege, F.C., Fiedler, U. & Timmers, H.T. Three transitions in the RNA polymerase II transcription complex during initiation. *The EMBO journal* **16**, 7468-7480 (1997).
65. Sainsbury, S., Niesser, J. & Cramer, P. Structure and function of the initially transcribing RNA polymerase II–TFIIB complex. *Nature* **493**, 437 (2012).
66. Feaver, W.J., Gileadi, O., Li, Y. & Kornberg, R.D. CTD kinase associated with yeast RNA polymerase II initiation factor b. *Cell* **67**, 1223-1230 (1991).
67. Feaver, W.J., Svejstrup, J.Q., Henry, N.L. & Kornberg, R.D. Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIH/TFIIK. *Cell* **79**, 1103-1109 (1994).
68. Akhtar, M.S., Heidemann, M., Tietjen, J.R., Zhang, D.W., Chapman, R.D., Eick, D. & Ansari, A.Z. TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Molecular cell* **34**, 387-393 (2009).

69. Luse, D.S. Promoter clearance by RNA polymerase II. *Biochimica et biophysica acta* **1829**, 63-68 (2013).
70. Fabrega, C., Shen, V., Shuman, S. & Lima, C.D. Structure of an mRNA Capping Enzyme Bound to the Phosphorylated Carboxy-Terminal Domain of RNA Polymerase II. *Molecular cell* **11**, 1549-1561 (2003).
71. Martinez-Rucobo, F.W., Kohler, R., van de Waterbeemd, M., Heck, A.J., Hemann, M., Herzog, F., Stark, H. & Cramer, P. Molecular Basis of Transcription-Coupled Pre-mRNA Capping. *Molecular cell* **58**, 1079-1089 (2015).
72. Coppola, J.A., Field, A.S. & Luse, D.S. Promoter-proximal pausing by RNA polymerase II in vitro: transcripts shorter than 20 nucleotides are not capped. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 1251-1255 (1983).
73. Core, L. & Adelman, K. Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes & development* **33**, 960-982 (2019).
74. Narita, T., Yamaguchi, Y., Yano, K., Sugimoto, S., Chanarat, S., Wada, T., Kim, D.K., Hasegawa, J., Omori, M., Inukai, N., Endoh, M., Yamada, T. & Handa, H. Human transcription elongation factor NELF: identification of novel subunits and reconstitution of the functionally active complex. *Molecular and cellular biology* **23**, 1863-1873 (2003).
75. Marshall, N.F., Peng, J., Xie, Z. & Price, D.H. Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *The Journal of biological chemistry* **271**, 27176-27183 (1996).
76. Vos, S.M., Farnung, L., Urlaub, H. & Cramer, P. Structure of paused transcription complex Pol II-DSIF-NELF. *Nature* **560**, 601-606 (2018).
77. Sdano, M.A., Fulcher, J.M., Palani, S., Chandrasekharan, M.B., Parnell, T.J., Whitby, F.G., Formosa, T. & Hill, C.P. A novel SH2 recognition mechanism recruits Spt6 to the doubly phosphorylated RNA polymerase II linker at sites of transcription. *eLife* **6** (2017).
78. Zeitlinger, J., Stark, A., Kellis, M., Hong, J.W., Nechaev, S., Adelman, K., Levine, M. & Young, R.A. RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. *Nat Genet* **39**, 1512-1516 (2007).
79. Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., Tomsho, L.P., Qi, J., Glaser, R.L., Schuster, S.C., Gilmour, D.S., Albert, I. & Pugh, B.F. Nucleosome organization in the Drosophila genome. *Nature* **453**, 358-362 (2008).
80. Farnung, L., Vos, S.M. & Cramer, P. Structure of transcribing RNA polymerase II-nucleosome complex. *Nature communications* **9**, 5432 (2018).
81. Kujirai, T., Ehara, H., Fujino, Y., Shirouzu, M., Sekine, S.-i. & Kurumizaka, H. Structural basis of the nucleosome transition during RNA polymerase II passage. *Science (New York, N.Y.)* **362**, 595-598 (2018).
82. Orphanides, G., LeRoy, G., Chang, C.H., Luse, D.S. & Reinberg, D. FACT, a factor that facilitates transcript elongation through nucleosomes. *Cell* **92**, 105-116 (1998).
83. Saunders, A., Werner, J., Andrulis, E.D., Nakayama, T., Hirose, S., Reinberg, D. & Lis, J.T. Tracking FACT and the RNA polymerase II elongation complex through chromatin in vivo. *Science (New York, N.Y.)* **301**, 1094-1096 (2003).
84. Ng, H.H., Robert, F., Young, R.A. & Struhl, K. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Molecular cell* **11**, 709-719 (2003).

85. Bortvin, A. & Winston, F. Evidence that Spt6p controls chromatin structure by a direct interaction with histones. *Science (New York, N.Y.)* **272**, 1473-1476 (1996).
86. Kizer, K.O., Phatnani, H.P., Shibata, Y., Hall, H., Greenleaf, A.L. & Strahl, B.D. A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. *Molecular and cellular biology* **25**, 3305-3316 (2005).
87. Vos, S.M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H. & Cramer, P. Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. *Nature* **560**, 607-612 (2018).
88. Jonkers, I., Kwak, H. & Lis, J.T. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *eLife* **3**, e02407 (2014).
89. Gressel, S., Schwalb, B., Decker, T.M., Qin, W., Leonhardt, H., Eick, D. & Cramer, P. CDK9-dependent RNA polymerase II pausing controls transcription initiation. *eLife* **6**, e29736 (2017).
90. Herzel, L., Ottoz, D.S.M., Alpert, T. & Neugebauer, K.M. Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nature reviews. Molecular cell biology* **18**, 637-650 (2017).
91. Hirose, Y., Tacke, R. & Manley, J.L. Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. *Genes & development* **13**, 1234-1239 (1999).
92. Misteli, T. & Spector, D.L. RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. *Molecular cell* **3**, 697-705 (1999).
93. Mortillaro, M.J., Blencowe, B.J., Wei, X., Nakayasu, H., Du, L., Warren, S.L., Sharp, P.A. & Berezney, R. A hyperphosphorylated form of the large subunit of RNA polymerase II is associated with splicing complexes and the nuclear matrix. *Proceedings of the National Academy of Sciences* **93**, 8253-8257 (1996).
94. Hirose, Y. & Manley, J.L. RNA polymerase II is an essential mRNA polyadenylation factor. *Nature* **395**, 93-96 (1998).
95. McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M. & Bentley, D.L. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**, 357-361 (1997).
96. Connelly, S. & Manley, J.L. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes & development* **2**, 440-452 (1988).
97. Baejen, C., Andreani, J., Torkler, P., Battaglia, S., Schwalb, B., Lidschreiber, M., Maier, K.C., Boltendahl, A., Rus, P., Esslinger, S., Soding, J. & Cramer, P. Genome-wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes. *Molecular cell* **66**, 38-49.e36 (2017).
98. Proudfoot, N.J. Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science (New York, N.Y.)* **352**, aad9926 (2016).
99. Kim, M., Krogan, N.J., Vasiljeva, L., Rando, O.J., Nedea, E., Greenblatt, J.F. & Buratowski, S. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* **432**, 517-522 (2004).
100. West, S., Gromak, N. & Proudfoot, N.J. Human 5′ → 3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* **432**, 522-525 (2004).

101. Schwalb, B., Michel, M., Zacher, B., Frühauf, K., Demel, C., Tresch, A., Gagneur, J. & Cramer, P. TT-seq maps the human transient transcriptome. *Science (New York, N.Y.)* **352**, 1225-1228 (2016).
102. Logan, J., Falck-Pedersen, E., Darnell, J.E., Jr. & Shenk, T. A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 8306-8310 (1987).
103. Ansari, A. & Hampsey, M. A role for the CPF 3′-end processing machinery in RNAP II-dependent gene looping. *Genes & development* **19**, 2969-2978 (2005).
104. O'Sullivan, J.M., Tan-Wong, S.M., Morillon, A., Lee, B., Coles, J., Mellor, J. & Proudfoot, N.J. Gene loops juxtapose promoters and terminators in yeast. *Nat Genet* **36**, 1014-1018 (2004).
105. Levine, M. & Tjian, R. Transcription regulation and animal diversity. *Nature* **424**, 147-151 (2003).
106. Levine, M., Cattoglio, C. & Tjian, R. Looping back to leap forward: transcription enters a new era. *Cell* **157**, 13-25 (2014).
107. Ptashne, M. & Gann, A. Transcriptional activation by recruitment. *Nature* **386**, 569-577 (1997).
108. Buratowski, S., Hahn, S., Guarente, L. & Sharp, P.A. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56**, 549-561 (1989).
109. Sikorski, T.W. & Buratowski, S. The basal initiation machinery: beyond the general transcription factors. *Current opinion in cell biology* **21**, 344-351 (2009).
110. Dynan, W.S. & Tjian, R. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell* **35**, 79-87 (1983).
111. Ptashne, M. How eukaryotic transcriptional activators work. *Nature* **335**, 683-689 (1988).
112. Frietze, S. & Farnham, P.J. Transcription factor effector domains. *Sub-cellular biochemistry* **52**, 261-277 (2011).
113. Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R. & Weirauch, M.T. The Human Transcription Factors. *Cell* **175**, 598-599 (2018).
114. Struhl, K., Kadosh, D., Keaveney, M., Kuras, L. & Moqtaderi, Z. Activation and repression mechanisms in yeast. *Cold Spring Harbor symposia on quantitative biology* **63**, 413-421 (1998).
115. Hahn, S. & Young, E.T. Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics* **189**, 705-736 (2011).
116. Furlong, E.E.M. & Levine, M. Developmental enhancers and chromosome topology. *Science (New York, N.Y.)* **361**, 1341-1345 (2018).
117. Amano, T., Sagai, T., Tanabe, H., Mizushina, Y., Nakazawa, H. & Shiroishi, T. Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. *Developmental cell* **16**, 47-57 (2009).
118. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nature reviews. Genetics* **15**, 272-286 (2014).
119. Reiter, F., Wienerroither, S. & Stark, A. Combinatorial function of transcription factors and cofactors. *Current opinion in genetics & development* **43**, 73-81 (2017).

120. Dunn, T.M., Hahn, S., Ogden, S. & Schleif, R.F. An operator at -280 base pairs that is required for repression of araBAD operon promoter: addition of DNA helical turns between the operator and promoter cyclically hinders repression. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 5017-5020 (1984).
121. Chen, H., Levo, M., Barinov, L., Fujioka, M., Jaynes, J.B. & Gregor, T. Dynamic interplay between enhancer-promoter topology and gene activity. *Nat Genet* **50**, 1296-1303 (2018).
122. Fukaya, T., Lim, B. & Levine, M. Enhancer Control of Transcriptional Bursting. *Cell* **166**, 358-368 (2016).
123. Ghavi-Helm, Y., Klein, F.A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W. & Furlong, E.E. Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**, 96-100 (2014).
124. Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S. & Aiden, E.L. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680 (2014).
125. Sanyal, A., Lajoie, B.R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489**, 109-113 (2012).
126. Larsson, A.J.M., Johnsson, P., Hagemann-Jensen, M., Hartmanis, L., Faridani, O.R., Reinius, B., Segerstolpe, Å., Rivera, C.M., Ren, B. & Sandberg, R. Genomic encoding of transcriptional burst kinetics. *Nature* **565**, 251-254 (2019).
127. Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., Gribnau, J., Barillot, E., Bluthgen, N., Dekker, J. & Heard, E. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-385 (2012).
128. Naar, A.M., Lemon, B.D. & Tjian, R. Transcriptional coactivator complexes. *Annual review of biochemistry* **70**, 475-501 (2001).
129. Malik, S. & Roeder, R.G. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nature Reviews Genetics* **11**, 761 (2010).
130. El Khattabi, L., Zhao, H., Kalchschmidt, J., Young, N., Jung, S., Van Blerkom, P., Kieffer-Kwon, P., Kieffer-Kwon, K.R., Park, S., Wang, X., Krebs, J., Tripathi, S., Sakabe, N., Sobreira, D.R., Huang, S.C., Rao, S.S.P., Pruett, N., Chauss, D., Sadler, E., Lopez, A., Nobrega, M.A., Aiden, E.L., Asturias, F.J. & Casellas, R. A Pliable Mediator Acts as a Functional Rather Than an Architectural Bridge between Promoters and Enhancers. *Cell* **178**, 1145-1158.e1120 (2019).
131. Jeronimo, C., Langelier, M.F., Bataille, A.R., Pascal, J.M., Pugh, B.F. & Robert, F. Tail and Kinase Modules Differently Regulate Core Mediator Recruitment and Function In Vivo. *Molecular cell* **64**, 455-466 (2016).
132. Petrenko, N., Jin, Y., Wong, K.H. & Struhl, K. Evidence that Mediator is essential for Pol II transcription, but is not a required component of the preinitiation complex in vivo. *eLife* **6**, e28447 (2017).
133. Lu, H., Flores, O., Weinmann, R. & Reinberg, D. The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 10004-10008 (1991).

134. 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. *Molecular cell* **60**, 231-241 (2015).
135. Kwon, I., Kato, M., Xiang, S., Wu, L., Theodoropoulos, P., Mirzaei, H., Han, T., Xie, S., Corden, J.L. & McKnight, S.L. Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell* **155**, 1049-1060 (2013).
136. Myers, L.C., Gustafsson, C.M., Bushnell, D.A., Lui, M., Erdjument-Bromage, H., Tempst, P. & Kornberg, R.D. The Med proteins of yeast and their function through the RNA polymerase II carboxy-terminal domain. *Genes & development* **12**, 45-54 (1998).
137. Naar, A.M., Taatjes, D.J., Zhai, W., Nogales, E. & Tjian, R. Human CRSP interacts with RNA polymerase II CTD and adopts a specific CTD-bound conformation. *Genes & development* **16**, 1339-1344 (2002).
138. Usheva, A., Maldonado, E., Goldring, A., Lu, H., Houbavi, C., Reinberg, D. & Aloni, Y. Specific interaction between the nonphosphorylated form of RNA polymerase II and the TATA-binding protein. *Cell* **69**, 871-881 (1992).
139. Kato, M. & McKnight, S.L. A Solid-State Conceptualization of Information Transfer from Gene to Message to Protein. *Annual review of biochemistry* (2017).
140. Sigler, P.B. Transcriptional activation. Acid blobs and negative noodles. *Nature* **333**, 210-212 (1988).
141. Gilmour, D.S. & Lis, J.T. RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells. *Molecular and cellular biology* **6**, 3984-3989 (1986).
142. Rougvie, A.E. & Lis, J.T. The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. *Cell* **54**, 795-804 (1988).
143. Core, L.J., Waterfall, J.J. & Lis, J.T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science (New York, N.Y.)* **322**, 1845-1848 (2008).
144. Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grissom, S.F., Zeitlinger, J. & Adelman, K. RNA polymerase is poised for activation across the genome. *Nat Genet* **39**, 1507-1511 (2007).
145. Hendrix, D.A., Hong, J.-W., Zeitlinger, J., Rokhsar, D.S. & Levine, M.S. Promoter elements associated with RNA Pol II stalling in the Drosophila embryo. *Proceedings of the National Academy of Sciences* **105**, 7762-7767 (2008).
146. Nechaev, S., Fargo, D.C., dos Santos, G., Liu, L., Gao, Y. & Adelman, K. Global Analysis of Short RNAs Reveals Widespread Promoter-Proximal Stalling and Arrest of Pol II in Drosophila. *Science (New York, N.Y.)* **327**, 335-338 (2010).
147. Yamaguchi, Y., Shibata, H. & Handa, H. Transcription elongation factors DSIF and NELF: promoter-proximal pausing and beyond. *Biochimica et biophysica acta* **1829**, 98-104 (2013).
148. Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J. & Handa, H. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* **97**, 41-51 (1999).
149. Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G.A., Winston, F., Buratowski, S. & Handa, H. DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is

composed of human Spt4 and Spt5 homologs. *Genes & development* **12**, 343-356 (1998).
150. Hartzog, G.A. & Fu, J. The Spt4-Spt5 complex: a multi-faceted regulator of transcription elongation. *Biochimica et biophysica acta* **1829**, 105-115 (2013).
151. Bernecky, C., Plitzko, J.M. & Cramer, P. Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. *Nature structural & molecular biology* **24**, 809-815 (2017).
152. Missra, A. & Gilmour, D.S. Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the Drosophila RNA polymerase II transcription elongation complex. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 11301-11306 (2010).
153. Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T. & Handa, H. P-TEFb-Mediated Phosphorylation of hSpt5 C-Terminal Repeats Is Critical for Processive Transcription Elongation. *Molecular cell* **21**, 227-237 (2006).
154. Liu, Y., Warfield, L., Zhang, C., Luo, J., Allen, J., Lang, W.H., Ranish, J., Shokat, K.M. & Hahn, S. Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex. *Molecular and cellular biology* **29**, 4852-4863 (2009).
155. Baugh, L.R., Demodena, J. & Sternberg, P.W. RNA Pol II accumulates at promoters of growth genes during developmental arrest. *Science (New York, N.Y.)* **324**, 92-94 (2009).
156. Vos, S.M., Pollmann, D., Caizzi, L., Hofmann, K.B., Rombaut, P., Zimniak, T., Herzog, F. & Cramer, P. Architecture and RNA binding of the human negative elongation factor. *eLife* **5** (2016).
157. Kaplan, C.D., Larsson, K.M. & Kornberg, R.D. The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin. *Molecular cell* **30**, 547-556 (2008).
158. Cheung, A.C.M. & Cramer, P. Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature* **471**, 249 (2011).
159. Henriques, T., Gilchrist, D.A., Nechaev, S., Bern, M., Muse, G.W., Burkholder, A., Fargo, D.C. & Adelman, K. Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals. *Molecular cell* **52**, 517-528 (2013).
160. Henriques, T., Scruggs, B.S., Inouye, M.O., Muse, G.W., Williams, L.H., Burkholder, A.B., Lavender, C.A., Fargo, D.C. & Adelman, K. Widespread transcriptional pausing and elongation control at enhancers. *Genes & development* **32**, 26-41 (2018).
161. Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R. & Young, R.A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77-88 (2007).
162. Gressel, S., Schwalb, B. & Cramer, P. The pause-initiation limit restricts transcription activation in human cells. *Nature communications* **10**, 3603 (2019).
163. Nilson, K.A., Lawson, C.K., Mullen, N.J., Ball, C.B., Spector, B.M., Meier, J.L. & Price, D.H. Oxidative stress rapidly stabilizes promoter-proximal paused Pol II across the human genome. *Nucleic acids research* **45**, 11088-11105 (2017).
164. Marshall, N.F. & Price, D.H. Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *The Journal of biological chemistry* **270**, 12335-12338 (1995).

165. Price, D.H. P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. *Molecular and cellular biology* **20**, 2629-2634 (2000).
166. Wei, P., Garber, M.E., Fang, S.M., Fischer, W.H. & Jones, K.A. A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell* **92**, 451-462 (1998).
167. Cheng, B. & Price, D.H. Properties of RNA polymerase II elongation complexes before and after the P-TEFb-mediated transition into productive elongation. *The Journal of biological chemistry* **282**, 21901-21912 (2007).
168. Nguyen, V.T., Kiss, T., Michels, A.A. & Bensaude, O. 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. *Nature* **414**, 322-325 (2001).
169. Yang, Z., Zhu, Q., Luo, K. & Zhou, Q. The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. *Nature* **414**, 317-322 (2001).
170. Paparidis, N.F.d.S., Durvale, M.C. & Canduri, F. The emerging picture of CDK9/P-TEFb: more than 20 years of advances since PITALRE. *Molecular BioSystems* **13**, 246-276 (2017).
171. Peterlin, B.M., Brogie, J.E. & Price, D.H. 7SK snRNA: a noncoding RNA that plays a major role in regulating eukaryotic transcription. *Wiley Interdisciplinary Reviews: RNA* **3**, 92-103 (2012).
172. Amente, S., Gargano, B., Napolitano, G., Lania, L. & Majello, B. Camptothecin releases P-TEFb from the inactive 7SK snRNP complex. *Cell cycle (Georgetown, Tex.)* **8**, 1249-1255 (2009).
173. Lindquist, S. The heat-shock response. *Annual review of biochemistry* **55**, 1151-1191 (1986).
174. Zanton, S.J. & Pugh, B.F. Full and partial genome-wide assembly and disassembly of the yeast transcription machinery in response to heat shock. *Genes & development* **20**, 2250-2265 (2006).
175. Aprile-Garcia, F., Tomar, P., Hummel, B., Khavaran, A. & Sawarkar, R. Nascent-protein ubiquitination is required for heat shock–induced gene downregulation in human cells. *Nature structural & molecular biology* **26**, 137-146 (2019).
176. Vihervaara, A., Duarte, F.M. & Lis, J.T. Molecular mechanisms driving transcriptional stress responses. *Nature reviews. Genetics* **19**, 385-397 (2018).
177. Gomez-Pastor, R., Burchfiel, E.T. & Thiele, D.J. Regulation of heat shock transcription factors and their roles in physiology and disease. *Nature reviews. Molecular cell biology* **19**, 4-19 (2018).
178. Vihervaara, A. & Sistonen, L. HSF1 at a glance. *Journal of cell science* **127**, 261-266 (2014).
179. Amin, J., Ananthan, J. & Voellmy, R. Key features of heat shock regulatory elements. *Molecular and cellular biology* **8**, 3761-3769 (1988).
180. Pelham, H.R. A regulatory upstream promoter element in the Drosophila hsp 70 heat-shock gene. *Cell* **30**, 517-528 (1982).
181. Sorger, P.K. & Pelham, H.R. Yeast heat shock factor is an essential DNA-binding protein that exhibits temperature-dependent phosphorylation. *Cell* **54**, 855-864 (1988).
182. Lis, J.T., Mason, P., Peng, J., Price, D.H. & Werner, J. P-TEFb kinase recruitment and function at heat shock loci. *Genes & development* **14**, 792-803 (2000).
183. Park, J.M., Werner, J., Kim, J.M., Lis, J.T. & Kim, Y.J. Mediator, not holoenzyme, is directly recruited to the heat shock promoter by HSF upon heat shock. *Molecular cell* **8**, 9-19 (2001).

184. Fuda, N.J., Ardehali, M.B. & Lis, J.T. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* **461**, 186-192 (2009).
185. Zobeck, K.L., Buckley, M.S., Zipfel, W.R. & Lis, J.T. Recruitment timing and dynamics of transcription factors at the Hsp70 loci in living cells. *Molecular cell* **40**, 965-975 (2010).
186. Kaplan, C.D., Morris, J.R., Wu, C. & Winston, F. Spt5 and spt6 are associated with active transcription and have characteristics of general elongation factors in D. melanogaster. *Genes & development* **14**, 2623-2634 (2000).
187. Teves, S.S. & Henikoff, S. Heat shock reduces stalled RNA polymerase II and nucleosome turnover genome-wide. *Genes & development* **25**, 2387-2397 (2011).
188. Yao, J., Ardehali, M.B., Fecko, C.J., Webb, W.W. & Lis, J.T. Intranuclear distribution and local dynamics of RNA polymerase II during transcription activation. *Molecular cell* **28**, 978-990 (2007).
189. Mahat, D.B., Salamanca, H.H., Duarte, F.M., Danko, C.G. & Lis, J.T. Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation. *Molecular cell* **62**, 63-78 (2016).
190. Shao, W. & Zeitlinger, J. Paused RNA polymerase II inhibits new transcriptional initiation. *Nat Genet* **49**, 1045-1051 (2017).
191. Liu, Z. & Tjian, R. Visualizing transcription factor dynamics in living cells. *The Journal of cell biology* **217**, 1181-1191 (2018).
192. Jackson, D.A., Hassan, A.B., Errington, R.J. & Cook, P.R. Visualization of focal sites of transcription within human nuclei. *The EMBO journal* **12**, 1059-1065 (1993).
193. Iborra, F.J., Pombo, A., Jackson, D.A. & Cook, P.R. Active RNA polymerases are localized within discrete transcription "factories' in human nuclei. *Journal of cell science* **109 (Pt 6)**, 1427-1436 (1996).
194. Wansink, D.G., Schul, W., van der Kraan, I., van Steensel, B., van Driel, R. & de Jong, L. Fluorescent labeling of nascent RNA reveals transcription by RNA polymerase II in domains scattered throughout the nucleus. *The Journal of cell biology* **122**, 283-293 (1993).
195. Jackson, D.A., Iborra, F.J., Manders, E.M. & Cook, P.R. Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. *Molecular biology of the cell* **9**, 1523-1536 (1998).
196. Osborne, C.S., Chakalova, L., Brown, K.E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J.A., Lopes, S., Reik, W. & Fraser, P. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nature Genetics* **36**, 1065-1071 (2004).
197. Brown, J.M., Green, J., das Neves, R.P., Wallace, H.A., Smith, A.J., Hughes, J., Gray, N., Taylor, S., Wood, W.G., Higgs, D.R., Iborra, F.J. & Buckle, V.J. Association between active genes occurs at nuclear speckles and is modulated by chromatin environment. *The Journal of cell biology* **182**, 1083-1097 (2008).
198. Morey, C., Kress, C. & Bickmore, W.A. Lack of bystander activation shows that localization exterior to chromosome territories is not sufficient to up-regulate gene expression. *Genome research* **19**, 1184-1194 (2009).
199. Pombo, A. & Cook, P.R. The localization of sites containing nascent RNA and splicing factors. *Experimental cell research* **229**, 201-203 (1996).
200. Buckley, M.S. & Lis, J.T. Imaging RNA Polymerase II transcription sites in living cells. *Current opinion in genetics & development* **25**, 126-130 (2014).

201. Cook, P.R. The organization of replication and transcription. *Science (New York, N.Y.)* **284**, 1790-1795 (1999).
202. Papantonis, A. & Cook, P.R. Transcription factories: genome organization and gene regulation. *Chemical reviews* **113**, 8683-8705 (2013).
203. Darzacq, X., Yao, J., Larson, D.R., Causse, S.Z., Bosanac, L., de Turris, V., Ruda, V.M., Lionnet, T., Zenklusen, D., Guglielmi, B., Tjian, R. & Singer, R.H. Imaging transcription in living cells. *Annual review of biophysics* **38**, 173-196 (2009).
204. Ghamari, A., van de Corput, M.P., Thongjuea, S., van Cappellen, W.A., van Ijcken, W., van Haren, J., Soler, E., Eick, D., Lenhard, B. & Grosveld, F.G. In vivo live imaging of RNA polymerase II transcription factories in primary cells. *Genes & development* **27**, 767-777 (2013).
205. Zhao, Z.W., Roy, R., Gebhardt, J.C.M., Suter, D.M., Chapman, A.R. & Xie, X.S. Spatial organization of RNA polymerase II inside a mammalian cell nucleus revealed by reflected light-sheet superresolution microscopy. *Proceedings of the National Academy of Sciences* **111**, 681 (2014).
206. Hieda, M., Winstanley, H., Maini, P., Iborra, F.J. & Cook, P.R. Different populations of RNA polymerase II in living mammalian cells. *Chromosome Research* **13**, 135-144 (2005).
207. Kimura, H., Sugaya, K. & Cook, P.R. The transcription cycle of RNA polymerase II in living cells. *The Journal of cell biology* **159**, 777-782 (2002).
208. Cisse, II, Izeddin, I., Causse, S.Z., Boudarene, L., Senecal, A., Muresan, L., Dugast-Darzacq, C., Hajj, B., Dahan, M. & Darzacq, X. Real-time dynamics of RNA polymerase II clustering in live human cells. *Science (New York, N.Y.)* **341**, 664-667 (2013).
209. Sengupta, P., Jovanovic-Talisman, T. & Lippincott-Schwartz, J. Quantifying spatial organization in point-localization superresolution images using pair correlation analysis. *Nature protocols* **8**, 345-354 (2013).
210. Sengupta, P., Jovanovic-Talisman, T., Skoko, D., Renz, M., Veatch, S.L. & Lippincott-Schwartz, J. Probing protein heterogeneity in the plasma membrane using PALM and pair correlation analysis. *Nature methods* **8**, 969-975 (2011).
211. Cho, W.-K., Jayanth, N., English, B.P., Inoue, T., Andrews, J.O., Conway, W., Grimm, J.B., Spille, J.-H., Lavis, L.D., Lionnet, T. & Cisse, I.I. RNA Polymerase II cluster dynamics predict mRNA output in living cells. *eLife* **5**, e13617 (2016).
212. Cho, W.K., Spille, J.H., Hecht, M., Lee, C., Li, C., Grube, V. & Cisse, II Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science (New York, N.Y.)* **361**, 412-415 (2018).
213. McNally, J.G., Muller, W.G., Walker, D., Wolford, R. & Hager, G.L. The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. *Science (New York, N.Y.)* **287**, 1262-1265 (2000).
214. Sprague, B.L., Pego, R.L., Stavreva, D.A. & McNally, J.G. Analysis of binding reactions by fluorescence recovery after photobleaching. *Biophysical journal* **86**, 3473-3495 (2004).
215. Stenoien, D.L., Patel, K., Mancini, M.G., Dutertre, M., Smith, C.L., O'Malley, B.W. & Mancini, M.A. FRAP reveals that mobility of oestrogen receptor-alpha is ligand- and proteasome-dependent. *Nature cell biology* **3**, 15-23 (2001).

216. Chen, J., Zhang, Z., Li, L., Chen, B.C., Revyakin, A., Hajj, B., Legant, W., Dahan, M., Lionnet, T., Betzig, E., Tjian, R. & Liu, Z. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* **156**, 1274-1285 (2014).
217. Chubb, J.R., Trcek, T., Shenoy, S.M. & Singer, R.H. Transcriptional pulsing of a developmental gene. *Current biology : CB* **16**, 1018-1025 (2006).
218. Deng, Q., Ramsköld, D., Reinius, B. & Sandberg, R. Single-Cell RNA-Seq Reveals Dynamic, Random Monoallelic Gene Expression in Mammalian Cells. *Science (New York, N.Y.)* **343**, 193 (2014).
219. Suter, D.M., Molina, N., Gatfield, D., Schneider, K., Schibler, U. & Naef, F. Mammalian genes are transcribed with widely different bursting kinetics. *Science (New York, N.Y.)* **332**, 472-474 (2011).
220. Tantale, K., Mueller, F., Kozulic-Pirher, A., Lesne, A., Victor, J.-M., Robert, M.-C., Capozi, S., Chouaib, R., Bäcker, V., Mateos-Langerak, J., Darzacq, X., Zimmer, C., Basyuk, E. & Bertrand, E. A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting. *Nature communications* **7**, 12248 (2016).
221. Lis, J.T. Imaging Drosophila gene activation and polymerase pausing in vivo. *Nature* **450**, 198-202 (2007).
222. Yao, J., Zobeck, K.L., Lis, J.T. & Webb, W.W. Imaging transcription dynamics at endogenous genes in living Drosophila tissues. *Methods (San Diego, Calif.)* **45**, 233-241 (2008).
223. Petesch, S.J. & Lis, J.T. Activator-induced spread of poly(ADP-ribose) polymerase promotes nucleosome loss at Hsp70. *Molecular cell* **45**, 64-74 (2012).
224. Holehouse, A.S. in Intrinsically Disordered Proteins. (ed. N. Salvi) 209-255 (Academic Press, 2019).
225. Shin, Y. & Brangwynne, C.P. Liquid phase condensation in cell physiology and disease. *Science (New York, N.Y.)* **357** (2017).
226. Martin, E.W. & Mittag, T. Relationship of Sequence and Phase Separation in Protein Low-Complexity Regions. *Biochemistry* **57**, 2478-2487 (2018).
227. Liu, J., Perumal, N.B., Oldfield, C.J., Su, E.W., Uversky, V.N. & Dunker, A.K. Intrinsic disorder in transcription factors. *Biochemistry* **45**, 6873-6888 (2006).
228. Brzovic, P.S., Heikaus, C.C., Kisselev, L., Vernon, R., Herbig, E., Pacheco, D., Warfield, L., Littlefield, P., Baker, D., Klevit, R.E. & Hahn, S. The acidic transcription activator Gcn4 binds the mediator subunit Gal11/Med15 using a simple protein interface forming a fuzzy complex. *Molecular cell* **44**, 942-953 (2011).
229. Warfield, L., Tuttle, L.M., Pacheco, D., Klevit, R.E. & Hahn, S. A sequence-specific transcription activator motif and powerful synthetic variants that bind Mediator using a fuzzy protein interface. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E3506-3513 (2014).
230. Nagulapalli, M., Maji, S., Dwivedi, N., Dahiya, P. & Thakur, J.K. Evolution of disorder in Mediator complex and its functional relevance. *Nucleic acids research* **44**, 1591-1612 (2015).
231. Banani, S.F., Lee, H.O., Hyman, A.A. & Rosen, M.K. Biomolecular condensates: organizers of cellular biochemistry. *Nature reviews. Molecular cell biology* **18**, 285-298 (2017).
232. Flory, P.J. Thermodynamics of High Polymer Solutions. *The Journal of Chemical Physics* **10**, 51-61 (1942).

233. Huggins, M.L. Some Properties of Solutions of Long-chain Compounds. *The Journal of Physical Chemistry* **46**, 151-158 (1942).
234. Brangwynne, C.P., Eckmann, C.R., Courson, D.S., Rybarska, A., Hoege, C., Gharakhani, J., Julicher, F. & Hyman, A.A. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science (New York, N.Y.)* **324**, 1729-1732 (2009).
235. Hyman, A.A., Weber, C.A. & Julicher, F. Liquid-liquid phase separation in biology. *Annual review of cell and developmental biology* **30**, 39-58 (2014).
236. Söding, J., Zwicker, D., Sohrabi-Jahromi, S., Boehning, M. & Kirschbaum, J. Mechanisms of active regulation of biomolecular condensates. *bioRxiv* (2019).
237. Weber, C.A., Zwicker, D., Julicher, F. & Lee, C.F. Physics of active emulsions. *Reports on progress in physics. Physical Society (Great Britain)* **82**, 064601 (2019).
238. Asherie, N., Lomakin, A. & Benedek, G.B. Phase Diagram of Colloidal Solutions. *Physical review letters* **77**, 4832-4835 (1996).
239. Mitrea, D.M., Chandra, B., Ferrolino, M.C., Gibbs, E.B., Tolbert, M., White, M.R. & Kriwacki, R.W. Methods for Physical Characterization of Phase-Separated Bodies and Membrane-less Organelles. *Journal of molecular biology* **430**, 4773-4805 (2018).
240. Li, P., Banjade, S., Cheng, H.C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J.V., King, D.S., Banani, S.F., Russo, P.S., Jiang, Q.X., Nixon, B.T. & Rosen, M.K. Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336-340 (2012).
241. Alberti, S., Gladfelter, A. & Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**, 419-434 (2019).
242. Rubinstein, M. & Dobrynin, A. Solutions of Associative Polymers. *Trends in Polymer Science* **5**, 181-186 (1997).
243. Semenov, A.N. & Rubinstein, M. Thermoreversible Gelation in Solutions of Associative Polymers. 1. Statics. *Macromolecules* **31**, 1373-1385 (1998).
244. Harmon, T.S., Holehouse, A.S., Rosen, M.K. & Pappu, R.V. Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *eLife* **6** (2017).
245. Wang, J., Choi, J.M., Holehouse, A.S., Lee, H.O., Zhang, X., Jahnel, M., Maharana, S., Lemaitre, R., Pozniakovsky, A., Drechsel, D., Poser, I., Pappu, R.V., Alberti, S. & Hyman, A.A. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **174**, 688-699.e616 (2018).
246. Brangwynne, Clifford P., Tompa, P. & Pappu, Rohit V. Polymer physics of intracellular phase transitions. *Nature Physics* **11**, 899 (2015).
247. Ambadipudi, S., Biernat, J., Riedel, D., Mandelkow, E. & Zweckstetter, M. Liquid-liquid phase separation of the microtubule-binding repeats of the Alzheimer-related protein Tau. *Nature communications* **8**, 275 (2017).
248. Boyko, S., Qi, X., Chen, T.H., Surewicz, K. & Surewicz, W.K. Liquid-liquid phase separation of tau protein: The crucial role of electrostatic interactions. *The Journal of biological chemistry* **294**, 11054-11059 (2019).
249. Elbaum-Garfinkle, S., Kim, Y., Szczepaniak, K., Chen, C.C., Eckmann, C.R., Myong, S. & Brangwynne, C.P. The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 7189-7194 (2015).

250. Franzmann, T.M., Jahnel, M., Pozniakovsky, A., Mahamid, J., Holehouse, A.S., Nuske, E., Richter, D., Baumeister, W., Grill, S.W., Pappu, R.V., Hyman, A.A. & Alberti, S. Phase separation of a yeast prion protein promotes cellular fitness. *Science (New York, N.Y.)* **359** (2018).
251. Nott, T.J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T.D., Bazett-Jones, D.P., Pawson, T., Forman-Kay, J.D. & Baldwin, A.J. Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Molecular cell* **57**, 936-947 (2015).
252. Pak, C.W., Kosno, M., Holehouse, A.S., Padrick, S.B., Mittal, A., Ali, R., Yunus, A.A., Liu, D.R., Pappu, R.V. & Rosen, M.K. Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein. *Mol Cell* **63**, 72-85 (2016).
253. Schmidt, H.B. & Gorlich, D. Transport Selectivity of Nuclear Pores, Phase Separation, and Membraneless Organelles. *Trends in biochemical sciences* **41**, 46-61 (2016).
254. Dao, T.P., Kolaitis, R.M., Kim, H.J., O'Donovan, K., Martyniak, B., Colicino, E., Hehnly, H., Taylor, J.P. & Castaneda, C.A. Ubiquitin Modulates Liquid-Liquid Phase Separation of UBQLN2 via Disruption of Multivalent Interactions. *Molecular cell* **69**, 965-978.e966 (2018).
255. Riback, J.A., Katanski, C.D., Kear-Scott, J.L., Pilipenko, E.V., Rojek, A.E., Sosnick, T.R. & Drummond, D.A. Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response. *Cell* **168**, 1028-1040.e1019 (2017).
256. Vernon, R.M., Chong, P.A., Tsang, B., Kim, T.H., Bah, A., Farber, P., Lin, H. & Forman-Kay, J.D. Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *eLife* **7** (2018).
257. Bergeron-Sandoval, L.P., Safaee, N. & Michnick, S.W. Mechanisms and Consequences of Macromolecular Phase Separation. *Cell* **165**, 1067-1079 (2016).
258. Crick, S.L., Ruff, K.M., Garai, K., Frieden, C. & Pappu, R.V. Unmasking the roles of N- and C-terminal flanking sequences from exon 1 of huntingtin as modulators of polyglutamine aggregation. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 20075-20080 (2013).
259. Bah, A. & Forman-Kay, J.D. Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications. *The Journal of biological chemistry* **291**, 6696-6705 (2016).
260. Hofweber, M. & Dormann, D. Friend or foe-Post-translational modifications as regulators of phase separation and RNP granule dynamics. *The Journal of biological chemistry* **294**, 7137-7150 (2019).
261. Hofweber, M., Hutten, S., Bourgeois, B., Spreitzer, E., Niedner-Boblenz, A., Schifferer, M., Ruepp, M.D., Simons, M., Niessing, D., Madl, T. & Dormann, D. Phase Separation of FUS Is Suppressed by Its Nuclear Import Receptor and Arginine Methylation. *Cell* **173**, 706-719.e713 (2018).
262. Qamar, S., Wang, G., Randle, S.J., Ruggeri, F.S., Varela, J.A., Lin, J.Q., Phillips, E.C., Miyashita, A., Williams, D., Strohl, F., Meadows, W., Ferry, R., Dardov, V.J., Tartaglia, G.G., Farrer, L.A., Kaminski Schierle, G.S., Kaminski, C.F., Holt, C.E., Fraser, P.E., Schmitt-Ulms, G., Klenerman, D., Knowles, T., Vendruscolo, M. & St George-Hyslop, P. FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation-pi Interactions. *Cell* **173**, 720-734.e715 (2018).

263. Banani, S.F., Rice, A.M., Peeples, W.B., Lin, Y., Jain, S., Parker, R. & Rosen, M.K. Compositional Control of Phase-Separated Cellular Bodies. *Cell* **166**, 651-663 (2016).
264. Altmeyer, M., Neelsen, K.J., Teloni, F., Pozdnyakova, I., Pellegrino, S., Grøfte, M., Rask, M.-B.D., Streicher, W., Jungmichel, S., Nielsen, M.L. & Lukas, J. Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose). *Nature communications* **6**, 8088 (2015).
265. Monahan, Z., Ryan, V.H., Janke, A.M., Burke, K.A., Rhoads, S.N., Zerze, G.H., O'Meally, R., Dignon, G.L., Conicella, A.E., Zheng, W., Best, R.B., Cole, R.N., Mittal, J., Shewmaker, F. & Fawzi, N.L. Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity. *The EMBO journal* **36**, 2951-2967 (2017).
266. Vlastaridis, P., Kyriakidou, P., Chaliotis, A., Van de Peer, Y., Oliver, S.G. & Amoutzias, G.D. Estimating the total number of phosphoproteins and phosphorylation sites in eukaryotic proteomes. *GigaScience* **6**, 1-11 (2017).
267. Iakoucheva, L.M., Radivojac, P., Brown, C.J., O'Connor, T.R., Sikes, J.G., Obradovic, Z. & Dunker, A.K. The importance of intrinsic disorder for protein phosphorylation. *Nucleic acids research* **32**, 1037-1049 (2004).
268. Khoury, G.A., Baliban, R.C. & Floudas, C.A. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Scientific reports* **1**, 90 (2011).
269. Aumiller, W.M., Jr. & Keating, C.D. Phosphorylation-mediated RNA/peptide complex coacervation as a model for intracellular liquid organelles. *Nature chemistry* **8**, 129-137 (2016).
270. Misteli, T., Caceres, J.F., Clement, J.Q., Krainer, A.R., Wilkinson, M.F. & Spector, D.L. Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo. *The Journal of cell biology* **143**, 297-307 (1998).
271. Wippich, F., Bodenmiller, B., Trajkovska, M.G., Wanka, S., Aebersold, R. & Pelkmans, L. Dual specificity kinase DYRK3 couples stress granule condensation/dissolution to mTORC1 signaling. *Cell* **152**, 791-805 (2013).
272. Gibson, B.A., Doolittle, L.K., Jensen, L.E., Gamarra, N., Redding, S. & Rosen, M.K. Organization and Regulation of Chromatin by Liquid-Liquid Phase Separation. *bioRxiv*, 523662 (2019).
273. Shen, T.H., Lin, H.K., Scaglioni, P.P., Yung, T.M. & Pandolfi, P.P. The mechanisms of PML-nuclear body formation. *Molecular cell* **24**, 331-339 (2006).
274. McGurk, L., Gomes, E., Guo, L., Mojsilovic-Petrovic, J., Tran, V., Kalb, R.G., Shorter, J. & Bonini, N.M. Poly(ADP-Ribose) Prevents Pathological Phase Separation of TDP-43 by Promoting Liquid Demixing and Stress Granule Localization. *Molecular cell* **71**, 703-717.e709 (2018).
275. 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).
276. Corden, J.L. RNA polymerase II C-terminal domain: Tethering transcription to transcript and template. *Chemical reviews* **113**, 8423-8455 (2013).
277. Kireeva, M.L., Lubkowska, L., Komissarova, N. & Kashlev, M. Assays and affinity purification of biotinylated and nonbiotinylated forms of double-tagged core RNA polymerase II from Saccharomyces cerevisiae. *Methods in enzymology* **370**, 138-155 (2003).

278. Kokic, G., Chernev, A., Tegunov, D., Dienemann, C., Urlaub, H. & Cramer, P. Structural basis of TFIIH activation for nucleotide excision repair. *Nature communications* **10**, 2885 (2019).
279. Gibson, D.G. Enzymatic assembly of overlapping DNA fragments. *Methods in enzymology* **498**, 349-361 (2011).
280. Laemmli, U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685 (1970).
281. Gradia, S.D., Ishida, J.P., Tsai, M.-S., Jeans, C., Tainer, J.A. & Fuss, J.O. in Methods in Enzymology, Vol. 592. (ed. B.F. Eichman) 1-26 (Academic Press, 2017).
282. Battaglia, S., Lidschreiber, M., Baejen, C., Torkler, P., Vos, S.M. & Cramer, P. RNA-dependent chromatin association of transcription elongation factors and Pol II CTD kinases. *eLife* **6** (2017).
283. Sydow, J.F., Brueckner, F., Cheung, A.C., Damsma, G.E., Dengl, S., Lehmann, E., Vassylyev, D. & Cramer, P. Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Molecular cell* **34**, 710-721 (2009).
284. Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J.M. & Cramer, P. Structure of transcribing mammalian RNA polymerase II. *Nature* **529**, 551 (2016).
285. Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., Brown, C.J. & Dunker, A.K. Predicting intrinsic disorder from amino acid sequence. *Proteins* **53 Suppl 6**, 566-572 (2003).
286. Goreaud, F. & Pélissier, R. On explicit formulas of edge effect correction for Ripley's K-function. *Journal of Vegetation Science* **10**, 433-438 (1999).
287. Ripley, B.D. Modelling Spatial Patterns. *Journal of the Royal Statistical Society. Series B (Methodological)* **39**, 172-212 (1977).
288. Dixon, P., Vol. 3 (2006).
289. Pélissier, 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. *2015* **63**, 18 (2015).
290. Grimm, J.B., English, B.P., Choi, H., Muthusamy, A.K., Mehl, B.P., Dong, P., Brown, T.A., Lippincott-Schwartz, J., Liu, Z., Lionnet, T. & Lavis, L.D. Bright photoactivatable fluorophores for single-molecule imaging. *Nature methods* **13**, 985 (2016).
291. Tokunaga, M., Imamoto, N. & Sakata-Sogawa, K. Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nature methods* **5**, 159 (2008).
292. Hansen, A.S., Pustova, I., Cattoglio, C., Tjian, R. & Darzacq, X. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *eLife* **6** (2017).
293. Hansen, A.S., Woringer, M., Grimm, J.B., Lavis, L.D., Tjian, R. & Darzacq, X. Robust model-based analysis of single-particle tracking experiments with Spot-On. *eLife* **7** (2018).
294. Serge, A., Bertaux, N., Rigneault, H. & Marguet, D. Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes. *Nature methods* **5**, 687-694 (2008).
295. Eisenhardt, N., Chaugule, V.K., Koidl, S., Droescher, M., Dogan, E., Rettich, J., Sutinen, P., Imanishi, S.Y., Hofmann, K., Palvimo, J.J. & Pichler, A. A new vertebrate SUMO enzyme family reveals insights into SUMO-chain assembly. *Nature Structural &Amp; Molecular Biology* **22**, 959 (2015).

296. Oellerich, T., Grønborg, M., Neumann, K., Hsiao, H.-H., Urlaub, H. & Wienands, J. SLP-65 Phosphorylation Dynamics Reveals a Functional Basis for Signal Integration by Receptor-proximal Adaptor Proteins. *Molecular & Cellular Proteomics* **8**, 1738-1750 (2009).
297. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* **26**, 1367-1372 (2008).
298. Koulouras, G., Panagopoulos, A., Rapsomaniki, M.A., Giakoumakis, N.N., Taraviras, S. & Lygerou, Z. EasyFRAP-web: a web-based tool for the analysis of fluorescence recovery after photobleaching data. *Nucleic acids research* **46**, W467-w472 (2018).
299. Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A. & Mann, M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & cellular proteomics : MCP* **1**, 376-386 (2002).
300. Maticzka, D., Ilik, I.A., Aktas, T., Backofen, R. & Akhtar, A. uvCLAP is a fast and non-radioactive method to identify in vivo targets of RNA-binding proteins. *Nature communications* **9**, 1142 (2018).
301. Park, S.S., Wu, W.W., Zhou, Y., Shen, R.F., Martin, B. & Maudsley, S. Effective correction of experimental errors in quantitative proteomics using stable isotope labeling by amino acids in cell culture (SILAC). *Journal of proteomics* **75**, 3720-3732 (2012).
302. Brangwynne, C.P., Tompa, P. & Pappu, R.V. Polymer physics of intracellular phase transitions. *Nat Phys* **11**, 899-904 (2015).
303. Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A.P., Kim, H.J., Mittag, T. & Taylor, J.P. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**, 123-133 (2015).
304. 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).
305. Kyte, J. & Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. *Journal of molecular biology* **157**, 105-132 (1982).
306. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends in genetics : TIG* **16**, 276-277 (2000).
307. Ribbeck, K. & Gorlich, D. The permeability barrier of nuclear pore complexes appears to operate via hydrophobic exclusion. *The EMBO journal* **21**, 2664-2671 (2002).
308. Lin, Y., Currie, S.L. & Rosen, M.K. Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs. *The Journal of biological chemistry* **292**, 19110-19120 (2017).
309. Jiang, H., Wang, S., Huang, Y., He, X., Cui, H., Zhu, X. & Zheng, Y. Phase transition of spindle-associated protein regulate spindle apparatus assembly. *Cell* **163**, 108-122 (2015).
310. Kato, M., Han, T.W., Xie, S., Shi, K., Du, X., Wu, L.C., Mirzaei, H., Goldsmith, E.J., Longgood, J., Pei, J., Grishin, N.V., Frantz, D.E., Schneider, J.W., Chen, S., Li, L., Sawaya, M.R., Eisenberg, D., Tycko, R. & McKnight, S.L. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753-767 (2012).

311. Xiang, S., Kato, M., Wu, L.C., Lin, Y., Ding, M., Zhang, Y., Yu, Y. & McKnight, S.L. The LC Domain of hnRNPA2 Adopts Similar Conformations in Hydrogel Polymers, Liquid-like Droplets, and Nuclei. *Cell* **163**, 829-839 (2015).
312. Schmidt, H.B. & Görlich, D. Nup98 FG domains from diverse species spontaneously phase-separate into particles with nuclear pore-like permselectivity. *eLife* **4**, e04251 (2015).
313. Darzacq, X., Shav-Tal, Y., de Turris, V., Brody, Y., Shenoy, S.M., Phair, R.D. & Singer, R.H. In vivo dynamics of RNA polymerase II transcription. *Nat Struct Mol Biol* **14**, 796-806 (2007).
314. Becker, M., Baumann, C., John, S., Walker, D.A., Vigneron, M., McNally, J.G. & Hager, G.L. Dynamic behavior of transcription factors on a natural promoter in living cells. *EMBO reports* **3**, 1188-1194 (2002).
315. Cho, W.K., Jayanth, N., Mullen, S., Tan, T.H., Jung, Y.J. & Cisse, II Super-resolution imaging of fluorescently labeled, endogenous RNA Polymerase II in living cells with CRISPR/Cas9-mediated gene editing. *Scientific reports* **6**, 35949 (2016).
316. Huang, B., Wang, W., Bates, M. & Zhuang, X. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science* **319**, 810-813 (2008).
317. Nicovich, P.R., Owen, D.M. & Gaus, K. Turning single-molecule localization microscopy into a quantitative bioanalytical tool. *Nat Protoc* **12**, 453-460 (2017).
318. Annibale, P., Vanni, S., Scarselli, M., Rothlisberger, U. & Radenovic, A. Identification of clustering artifacts in photoactivated localization microscopy. *Nature methods* **8**, 527-528 (2011).
319. Morgan, D.O. Principles of CDK regulation. *Nature* **374**, 131-134 (1995).
320. Conaway, J.W., Shilatifard, A., Dvir, A. & Conaway, R.C. Control of elongation by RNA polymerase II. *Trends in biochemical sciences* **25**, 375-380 (2000).
321. Patel, A., Malinovska, L., Saha, S., Wang, J., Alberti, S., Krishnan, Y. & Hyman, A.A. ATP as a biological hydrotrope. *Science (New York, N.Y.)* **356**, 753-756 (2017).
322. Liu, P., Kenney, J.M., Stiller, J.W. & Greenleaf, A.L. Genetic organization, length conservation, and evolution of RNA polymerase II carboxyl-terminal domain. *Molecular biology and evolution* **27**, 2628-2641 (2010).
323. Stiller, J.W. & Cook, M.S. Functional unit of the RNA polymerase II C-terminal domain lies within heptapeptide pairs. *Eukaryotic cell* **3**, 735-740 (2004).
324. Li, P., Banjade, S., Cheng, H.-C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J.V., King, D.S., Banani, S.F., Russo, P.S., Jiang, Q.-X., Nixon, B.T. & Rosen, M.K. Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336 (2012).
325. Serrano, L., Bycroft, M. & Fersht, A.R. Aromatic-aromatic interactions and protein stability. Investigation by double-mutant cycles. *Journal of molecular biology* **218**, 465-475 (1991).
326. Schwer, B. & Shuman, S. Deciphering the RNA polymerase II CTD code in fission yeast. *Molecular cell* **43**, 311-318 (2011).
327. Shah, N., Maqbool, M.A., Yahia, Y., El Aabidine, A.Z., Esnault, C., Forne, I., Decker, T.M., Martin, D., Schuller, R., Krebs, S., Blum, H., Imhof, A., Eick, D. & Andrau, J.C. Tyrosine-1 of RNA Polymerase II CTD Controls Global Termination of Gene Transcription in Mammals. *Molecular cell* **69**, 48-61.e46 (2018).
328. Descostes, N., Heidemann, M., Spinelli, L., Schuller, R., Maqbool, M.A., Fenouil, R., Koch, F., Innocenti, C., Gut, M., Gut, I., Eick, D. & Andrau, J.C. Tyrosine

phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. *eLife* **3**, e02105 (2014).

329. Hsin, J.P., Li, W., Hoque, M., Tian, B. & Manley, J.L. RNAP II CTD tyrosine 1 performs diverse functions in vertebrate cells. *eLife* **3**, e02112 (2014).
330. Meinhart, A. & Cramer, P. Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature* **430**, 223-226 (2004).
331. Ehrensberger, Andreas H., Kelly, Gavin P. & Svejstrup, Jesper Q. Mechanistic Interpretation of Promoter-Proximal Peaks and RNAPII Density Maps. *Cell* **154**, 713-715 (2013).
332. Lengyel, J.A. & Graham, M.L. Transcription, export and turnover of Hsp70 and alpha beta, two Drosophila heat shock genes sharing a 400 nucleotide 5' upstream region. *Nucleic acids research* **12**, 5719-5735 (1984).
333. Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G., Cattoglio, C., Banala, S., Lavis, L., Darzacq, X. & Tjian, R. Dynamic and Selective Low-Complexity Domain Interactions Revealed by Live-Cell Single-Molecule Imaging. *bioRxiv* (2018).
334. Boija, A., Klein, I.A., Sabari, B.R., Dall'Agnese, A., Coffey, E.L., Zamudio, A.V., Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M., Abraham, B.J., Afeyan, L.K., Guo, Y.E., Rimel, J.K., Fant, C.B., Schuijers, J., Lee, T.I., Taatjes, D.J. & Young, R.A. Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* **175**, 1842-1855.e1816 (2018).
335. Scafe, C., Chao, D., Lopes, J., Hirsch, J.P., Henry, S. & Young, R.A. RNA polymerase II C-terminal repeat influences response to transcriptional enhancer signals. *Nature* **347**, 491-494 (1990).
336. Staller, M.V., Holehouse, A.S., Swain-Lenz, D., Das, R.K., Pappu, R.V. & Cohen, B.A. A High-Throughput Mutational Scan of an Intrinsically Disordered Acidic Transcriptional Activation Domain. *Cell systems* **6**, 444-455.e446 (2018).
337. Drysdale, C.M., Duenas, E., Jackson, B.M., Reusser, U., Braus, G.H. & Hinnebusch, A.G. The transcriptional activator GCN4 contains multiple activation domains that are critically dependent on hydrophobic amino acids. *Molecular and cellular biology* **15**, 1220-1233 (1995).
338. Lu, F., Portz, B. & Gilmour, D.S. The C-Terminal Domain of RNA Polymerase II Is a Multivalent Targeting Sequence that Supports Drosophila Development with Only Consensus Heptads. *Molecular cell* **73**, 1232-1242.e1234 (2019).
339. Hnisz, D., Abraham, Brian J., Lee, Tong I., Lau, A., Saint-André, V., Sigova, Alla A., Hoke, Heather A. & Young, Richard A. Super-Enhancers in the Control of Cell Identity and Disease. *Cell* **155**, 934-947 (2013).
340. Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A.V., Manteiga, J.C., Li, C.H., Guo, Y.E., Day, D.S., Schuijers, J., Vasile, E., Malik, S., Hnisz, D., Lee, T.I., Cisse, II, Roeder, R.G., Sharp, P.A., Chakraborty, A.K. & Young, R.A. Coactivator condensation at super-enhancers links phase separation and gene control. *Science (New York, N.Y.)* **361** (2018).
341. Heist, T., Fukaya, T. & Levine, M. Large distances separate coregulated genes in living Drosophila embryos. *Proceedings of the National Academy of Sciences* **116**, 15062 (2019).
342. Malik, S. & Roeder, R.G. Mediator: A Drawbridge across the Enhancer-Promoter Divide. *Molecular cell* **64**, 433-434 (2016).

343. Wei, M.-T., Chang, Y.-C., Shimobayashi, S.F., Shin, Y. & Brangwynne, C.P. Nucleated transcriptional condensates amplify gene expression. *bioRxiv*, 737387 (2019).
344. Shin, Y., Chang, Y.C., Lee, D.S.W., Berry, J., Sanders, D.W., Ronceray, P., Wingreen, N.S., Haataja, M. & Brangwynne, C.P. Liquid Nuclear Condensates Mechanically Sense and Restructure the Genome. *Cell* **175**, 1481-1491.e1413 (2018).
345. Lu, H., Yu, D., Hansen, A.S., Ganguly, S., Liu, R., Heckert, A., Darzacq, X. & Zhou, Q. Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. *Nature* **558**, 318-323 (2018).
346. Guo, Y.E., Manteiga, J.C., Henninger, J.E., Sabari, B.R., Dall'Agnese, A., Hannett, N.M., Spille, J.-H., Afeyan, L.K., Zamudio, A.V., Shrinivas, K., Abraham, B.J., Boija, A., Decker, T.-M., Rimel, J.K., Fant, C.B., Lee, T.I., Cisse, I.I., Sharp, P.A., Taatjes, D.J. & Young, R.A. Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* **572**, 543-548 (2019).
347. Cramer, P. Organization and regulation of gene transcription. *Nature* **573**, 45-54 (2019).
348. Spector, D.L. & Lamond, A.I. Nuclear speckles. *Cold Spring Harbor perspectives in biology* **3** (2011).
349. Gomes, E. & Shorter, J. The molecular language of membraneless organelles. *The Journal of biological chemistry* **294**, 7115-7127 (2019).
350. Misteli, T., Caceres, J.F. & Spector, D.L. The dynamics of a pre-mRNA splicing factor in living cells. *Nature* **387**, 523-527 (1997).
351. Chen, Y., Zhang, Y., Wang, Y., Zhang, L., Brinkman, E.K., Adam, S.A., Goldman, R., van Steensel, B., Ma, J. & Belmont, A.S. Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *The Journal of cell biology* **217**, 4025-4048 (2018).
352. Mortillaro, M.J., Blencowe, B.J., Wei, X., Nakayasu, H., Du, L., Warren, S.L., Sharp, P.A. & Berezney, R. A hyperphosphorylated form of the large subunit of RNA polymerase II is associated with splicing complexes and the nuclear matrix. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 8253-8257 (1996).
353. Niskanen, E.A. & Palvimo, J.J. Chromatin SUMOylation in heat stress: To protect, pause and organise? *BioEssays* **39**, 1600263 (2017).
354. Narita, T., Yung, T.M., Yamamoto, J., Tsuboi, Y., Tanabe, H., Tanaka, K., Yamaguchi, Y. & Handa, H. NELF interacts with CBC and participates in 3' end processing of replication-dependent histone mRNAs. *Molecular cell* **26**, 349-365 (2007).
355. Kroschwald, S., Maharana, S. & Alberti, S. Hexanediol: a chemical probe to investigate the material properties of membrane-less compartments. *Matters* (2017).
356. Langdon, E.M. & Gladfelter, A.S. in Methods in Enzymology, Vol. 611. (ed. E. Rhoades) 67-79 (Academic Press, 2018).
357. Patel, A., Lee, Hyun O., Jawerth, L., Maharana, S., Jahnel, M., Hein, Marco Y., Stoynov, S., Mahamid, J., Saha, S., Franzmann, Titus M., Pozniakovski, A., Poser, I., Maghelli, N., Royer, Loic A., Weigert, M., Myers, Eugene W., Grill, S., Drechsel, D., Hyman, Anthony A. & Alberti, S. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **162**, 1066-1077 (2015).
358. Lin, Y., Protter, D.S., Rosen, M.K. & Parker, R. Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Molecular cell* **60**, 208-219 (2015).

359. King, J.L. & Jukes, T.H. Non-Darwinian evolution. *Science (New York, N.Y.)* **164**, 788-798 (1969).
360. Kinoshita, E., Kinoshita-Kikuta, E., Takiyama, K. & Koike, T. Phosphate-binding tag, a new tool to visualize phosphorylated proteins. *Molecular & cellular proteomics : MCP* **5**, 749-757 (2006).
361. O'Keeffe, B., Fong, Y., Chen, D., Zhou, S. & Zhou, Q. Requirement for a kinase-specific chaperone pathway in the production of a Cdk9/cyclin T1 heterodimer responsible for P-TEFb-mediated tat stimulation of HIV-1 transcription. *The Journal of biological chemistry* **275**, 279-287 (2000).
362. Flotho, A. & Melchior, F. Sumoylation: A Regulatory Protein Modification in Health and Disease. *Annual review of biochemistry* **82**, 357-385 (2013).
363. Golebiowski, F., Matic, I., Tatham, M.H., Cole, C., Yin, Y., Nakamura, A., Cox, J., Barton, G.J., Mann, M. & Hay, R.T. System-wide changes to SUMO modifications in response to heat shock. *Science signaling* **2**, ra24 (2009).
364. Hendriks, I.A., Lyon, D., Young, C., Jensen, L.J., Vertegaal, A.C.O. & Nielsen, M.L. Site-specific mapping of the human SUMO proteome reveals co-modification with phosphorylation. *Nature Structural &Amp; Molecular Biology* **24**, 325 (2017).
365. Uwada, J., Tanaka, N., Yamaguchi, Y., Uchimura, Y., Shibahara, K.-i., Nakao, M. & Saitoh, H. The p150 subunit of CAF-1 causes association of SUMO2/3 with the DNA replication foci. *Biochemical and Biophysical Research Communications* **391**, 407-413 (2010).
366. Prudden, J., Pebernard, S., Raffa, G., Slavin, D.A., Perry, J.J., Tainer, J.A., McGowan, C.H. & Boddy, M.N. SUMO-targeted ubiquitin ligases in genome stability. *The EMBO journal* **26**, 4089-4101 (2007).
367. Niskanen, E.A., Malinen, M., Sutinen, P., Toropainen, S., Paakinaho, V., Vihervaara, A., Joutsen, J., Kaikkonen, M.U., Sistonen, L. & Palvimo, J.J. Global SUMOylation on active chromatin is an acute heat stress response restricting transcription. *Genome Biology* **16**, 153 (2015).
368. Seifert, A., Schofield, P., Barton, G.J. & Hay, R.T. Proteotoxic stress reprograms the chromatin landscape of SUMO modification. *Science signaling* **8**, rs7 (2015).
369. Karvonen, U., Jaaskelainen, T., Rytinki, M., Kaikkonen, S. & Palvimo, J.J. ZNF451 is a novel PML body- and SUMO-associated transcriptional coregulator. *Journal of molecular biology* **382**, 585-600 (2008).
370. Maul, G.G., Yu, E., Ishov, A.M. & Epstein, A.L. Nuclear domain 10 (ND10) associated proteins are also present in nuclear bodies and redistribute to hundreds of nuclear sites after stress. *Journal of cellular biochemistry* **59**, 498-513 (1995).
371. Eskiw, C.H., Dellaire, G., Mymryk, J.S. & Bazett-Jones, D.P. Size, position and dynamic behavior of PML nuclear bodies following cell stress as a paradigm for supramolecular trafficking and assembly. *Journal of cell science* **116**, 4455-4466 (2003).
372. Itakura, A.K., Futia, R.A. & Jarosz, D.F. It Pays To Be in Phase. *Biochemistry* **57**, 2520-2529 (2018).
373. Rao, J.N., Neumann, L., Wenzel, S., Schweimer, K., Rosch, P. & Wohrl, B.M. Structural studies on the RNA-recognition motif of NELF E, a cellular negative transcription elongation factor involved in the regulation of HIV transcription. *The Biochemical journal* **400**, 449-456 (2006).

374. Rao, J.N., Schweimer, K., Wenzel, S., Wohrl, B.M. & Rosch, P. NELF-E RRM undergoes major structural changes in flexible protein regions on target RNA binding. *Biochemistry* **47**, 3756-3761 (2008).
375. Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M., Ulrich, A., Wardle, G.S., Dewell, S., Zavolan, M. & Tuschl, T. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129-141 (2010).
376. de Nadal, E., Ammerer, G. & Posas, F. Controlling gene expression in response to stress. *Nature reviews. Genetics* **12**, 833-845 (2011).
377. Chen, R., Liu, M., Li, H., Xue, Y., Ramey, W.N., He, N., Ai, N., Luo, H., Zhu, Y., Zhou, N. & Zhou, Q. PP2B and PP1alpha cooperatively disrupt 7SK snRNP to release P-TEFb for transcription in response to Ca2+ signaling. *Genes & development* **22**, 1356-1368 (2008).
378. Contreras, X., Barboric, M., Lenasi, T. & Peterlin, B.M. HMBA releases P-TEFb from HEXIM1 and 7SK snRNA via PI3K/Akt and activates HIV transcription. *PLoS pathogens* **3**, 1459-1469 (2007).
379. Zhao, Y., Karijolich, J., Glaunsinger, B. & Zhou, Q. Pseudouridylation of 7SK snRNA promotes 7SK snRNP formation to suppress HIV-1 transcription and escape from latency. *EMBO reports* **17**, 1441-1451 (2016).
380. Nair, S.J., Yang, L., Meluzzi, D., Oh, S., Yang, F., Friedman, M.J., Wang, S., Suter, T., Alshareedah, I., Gamliel, A., Ma, Q., Zhang, J., Hu, Y., Tan, Y., Ohgi, K.A., Jayani, R.S., Banerjee, P.R., Aggarwal, A.K. & Rosenfeld, M.G. Phase separation of ligand-activated enhancers licenses cooperative chromosomal enhancer assembly. *Nature structural & molecular biology* **26**, 193-203 (2019).
381. Janke, A.M., Seo, D.H., Rahmanian, V., Conicella, A.E., Mathews, K.L., Burke, K.A., Mittal, J. & Fawzi, N.L. Lysines in the RNA Polymerase II C-Terminal Domain Contribute to TAF15 Fibril Recruitment. *Biochemistry* **57**, 2549-2563 (2018).
382. Tasset, D., Tora, L., Fromental, C., Scheer, E. & Chambon, P. Distinct classes of transcriptional activating domains function by different mechanisms. *Cell* **62**, 1177-1187 (1990).
383. Bowler, B.E. Residual structure in unfolded proteins. *Curr Opin Struct Biol* **22**, 4-13 (2012).
384. Lee, H., Mok, K.H., Muhandiram, R., Park, K.H., Suk, J.E., Kim, D.H., Chang, J., Sung, Y.C., Choi, K.Y. & Han, K.H. Local structural elements in the mostly unstructured transcriptional activation domain of human p53. *The Journal of biological chemistry* **275**, 29426-29432 (2000).
385. Mittag, T., Orlicky, S., Choy, W.Y., Tang, X., Lin, H., Sicheri, F., Kay, L.E., Tyers, M. & Forman-Kay, J.D. Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 17772-17777 (2008).
386. Radhakrishnan, I., Pérez-Alvarado, G.C., Parker, D., Dyson, H.J., Montminy, M.R. & Wright, P.E. Solution Structure of the KIX Domain of CBP Bound to the Transactivation Domain of CREB: A Model for Activator:Coactivator Interactions. *Cell* **91**, 741-752 (1997).
387. Uesugi, M., Nyanguile, O., Lu, H., Levine, A.J. & Verdine, G.L. Induced α Helix in the VP16 Activation Domain upon Binding to a Human TAF. *Science (New York, N.Y.)* **277**, 1310-1313 (1997).

388. Ambadipudi, S., Reddy, J.G., Biernat, J., Mandelkow, E. & Zweckstetter, M. Residue-specific identification of phase separation hot spots of Alzheimer's-related protein tau. *Chemical Science* **10**, 6503-6507 (2019).
389. Brady, J.P., Farber, P.J., Sekhar, A., Lin, Y.-H., Huang, R., Bah, A., Nott, T.J., Chan, H.S., Baldwin, A.J., Forman-Kay, J.D. & Kay, L.E. Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proceedings of the National Academy of Sciences* **114**, E8194 (2017).
390. Murray, D.T., Kato, M., Lin, Y., Thurber, K.R., Hung, I., McKnight, S.L. & Tycko, R. Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. *Cell* **171**, 615-627.e616 (2017).
391. Murthy, A.C., Dignon, G.L., Kan, Y., Zerze, G.H., Parekh, S.H., Mittal, J. & Fawzi, N.L. Molecular interactions underlying liquid-liquid phase separation of the FUS low-complexity domain. *Nature structural & molecular biology* **26**, 637-648 (2019).
392. Bond, C.S. & Fox, A.H. Paraspeckles: nuclear bodies built on long noncoding RNA. *The Journal of cell biology* **186**, 637-644 (2009).
393. Fox, A.H., Lam, Y.W., Leung, A.K., Lyon, C.E., Andersen, J., Mann, M. & Lamond, A.I. Paraspeckles: a novel nuclear domain. *Current biology : CB* **12**, 13-25 (2002).
394. Fox, A.H., Nakagawa, S., Hirose, T. & Bond, C.S. Paraspeckles: Where Long Noncoding RNA Meets Phase Separation. *Trends in biochemical sciences* **43**, 124-135 (2018).
395. Naganuma, T., Nakagawa, S., Tanigawa, A., Sasaki, Y.F., Goshima, N. & Hirose, T. Alternative 3′-end processing of long noncoding RNA initiates construction of nuclear paraspeckles. *The EMBO journal* **31**, 4020-4034 (2012).
396. Tian, B. & Manley, J.L. Alternative polyadenylation of mRNA precursors. *Nature reviews. Molecular cell biology* **18**, 18-30 (2017).
397. Xiang, K., Tong, L. & Manley, J.L. Delineating the structural blueprint of the pre-mRNA 3'-end processing machinery. *Molecular and cellular biology* **34**, 1894-1910 (2014).
398. Cardinale, S., Cisterna, B., Bonetti, P., Aringhieri, C., Biggiogera, M. & Barabino, S.M. Subnuclear localization and dynamics of the Pre-mRNA 3' end processing factor mammalian cleavage factor I 68-kDa subunit. *Molecular biology of the cell* **18**, 1282-1292 (2007).
399. Dettwiler, S., Aringhieri, C., Cardinale, S., Keller, W. & Barabino, S.M. Distinct sequence motifs within the 68-kDa subunit of cleavage factor Im mediate RNA binding, protein-protein interactions, and subcellular localization. *The Journal of biological chemistry* **279**, 35788-35797 (2004).
400. Zhu, Y., Wang, X., Forouzmand, E., Jeong, J., Qiao, F., Sowd, G.A., Engelman, A.N., Xie, X., Hertel, K.J. & Shi, Y. Molecular Mechanisms for CFIm-Mediated Regulation of mRNA Alternative Polyadenylation. *Molecular cell* **69**, 62-74.e64 (2018).
401. Shi, Y. & Manley, J.L. The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site. *Genes & development* **29**, 889-897 (2015).
402. Fang, X., Wang, L., Ishikawa, R., Li, Y., Fiedler, M., Liu, F., Calder, G., Rowan, B., Weigel, D., Li, P. & Dean, C. Arabidopsis FLL2 promotes liquid-liquid phase separation of polyadenylation complexes. *Nature* **569**, 265-269 (2019).
403. van Steensel, B. & Furlong, E.E.M. The role of transcription in shaping the spatial organization of the genome. *Nature Reviews Molecular Cell Biology* **20**, 327-337 (2019).

404. Rada-Iglesias, A., Grosveld, F.G. & Papantonis, A. Forces driving the three-dimensional folding of eukaryotic genomes. *Molecular Systems Biology* **14**, e8214 (2018).

7 Supplementary Information

Items presented in this section have been published.

RNA polymerase II clustering through carboxy-terminal domain phase separation

M. Boehning*, C. Dugast-Darzacq*, M. Rankovic*, A. S. Hansen, T. Yu, H. Marie-Nelly, D. T. McSwiggen, G. Kokic, G. M. Dailey, P. Cramer, X. Darzacq, M. Zweckstetter

Nature Structural and Molecular Biology **25**, 833–840 (2018)

The published text was adapted to match the style of this thesis. Numbering and references to figures as well as references to the literature deviate from the published version. A detailed list of published items can be found in the Appendix (‘List of items from publications’, Page 149). Co-author contributions are stated on Page VI.

7.1 Supplementary Note

Calculation of diffusion coefficients

The observed free diffusion coefficients obtained from fitting the spaSPT data with the Spot-On model (Brownian motion) were 3.74 +/- 0.178 $\mu m^2/s$, 2.97 +/- 0.0912 $\mu m^2/s$ and 2.34 +/- 0.049 $\mu m^2/s$ for the 25R, 52R and 70R versions of Halo-Rpb1, respectively (mean +/- standard error). Given that the molecular weight of e.g. 25R is lower, one would expect the diffusion coefficient to be higher. To estimate whether this large difference could be explained by size alone or whether it might be due to reduced multivalent interactions, we consider the Stokes-Einstein relation according to which the diffusion coefficient is given by:

$$D = \frac{k_B T}{6\pi\eta r}$$

where k_B is Boltzmann’s constant, T is the absolute temperature, η is the viscosity of the liquid (the nucleoplasm here; assumed to be the same for 25R, 52R and 70R) and r is the radius.

The Stokes-Einstein equation assumes the particle to be a sphere and accordingly the radius is given by the volume, V:

$$r = \sqrt[3]{\frac{3V}{4\pi}}$$

In turn, the volume is related to the mass, m, and density, ρ:

$$V = \frac{m}{\rho} = \frac{MW}{\rho N_A}$$

where N_A is Avogadro's constant and MW is the molecular weight in atomic mass units (Daltons). Thus, the diffusion coefficient is related to the molecular weight by:

$$D = \frac{k_B T}{6\pi\eta\sqrt[3]{\frac{3MW}{4\pi\rho N_A}}}$$

Thus using 25R and 52R as the example, the ratio between the diffusion coefficients of 25R and 52R Halo-Rpb1 (assuming that the density is the same) is:

$$\frac{D_{\text{52R-Rpb1}}}{D_{\text{25R-Rpb1}}} = \frac{\frac{k_B T}{6\pi\eta\sqrt[3]{\frac{3MW_{\text{52R-Rpb1}}}{4\pi\rho N_A}}}}{\frac{k_B T}{6\pi\eta\sqrt[3]{\frac{3MW_{\text{25R-Rpb1}}}{4\pi\rho N_A}}}} = \sqrt[3]{\frac{MW_{\text{25R-Rpb1}}}{MW_{\text{52R-Rpb1}}}}$$

According to UniProt (P24928) the molecular weight of wild-type Rpb1 is 217.2 kDa (52R). The molecular weight of the HaloTag is 33.6 kDa. Thus, the molecular weight of Halo-Rpb1-52R is ~250.8 kDa, the molecular weight of Halo-Rpb1-25R is ~230.9 kDa and the molecular weight of Halo-Rpb1-70R is ~258.1 kDa.

Thus, the expected difference in diffusion coefficients is:

$$\frac{D_{\text{52R-Rpb1}}}{D_{\text{25R-Rpb1}}}_{\text{EXPECTED}} = \sqrt[3]{\frac{MW_{\text{25R-Rpb1}}}{MW_{\text{52R-Rpb1}}}} = \sqrt[3]{\frac{230.9\ \text{kDa}}{250.8\ \text{kDa}}} = 0.973$$

We can compare this to the experimentally observed ratio:

$$\frac{D_{\text{52R-Rpb1}}}{D_{\text{25R-Rpb1}}}_{\text{OBSERVED}} = \frac{2.97\,\frac{\mu\text{m}^2}{\text{s}}}{3.74\,\frac{\mu\text{m}^2}{\text{s}}} = 0.794$$

It becomes clear that size/mass difference alone cannot explain the large difference between the diffusion coefficients that we observe in cells. To be comprehensive, below we list the Stokes-Einstein expected and observed diffusion coefficient ratios for all the combinations:

Comparison	**Stokes-Einstein expectation**	**Observed ratio**
25R vs. 52R	0.973	0.794
25R vs. 70R	0.964	0.626
52R vs. 70R	0.991	0.789

For all three combinations, the observed ratio cannot be explained by the change in size/mass. Instead this indicates a higher propensity of the full-length CTD to engage in intermolecular interactions. Moreover, in the above calculations we have just considered the change in the mass of Rpb1. In reality, Rpb1 is likely diffusing as part of the Pol II holocomplex, thus the relative difference due to the smaller CTD (e.g. ~20 kDa between 25R and 52R) is actually much smaller than the calculations using only Rpb1 would suggest and thus the expected difference in diffusion coefficients due to mass/size would be even much closer to 1. We conclude that the mass/size difference between the 25R, 52R and 70R Pol II enzymes cannot explain their observed differences in diffusion coefficients.

7.2 Supplementary Figures

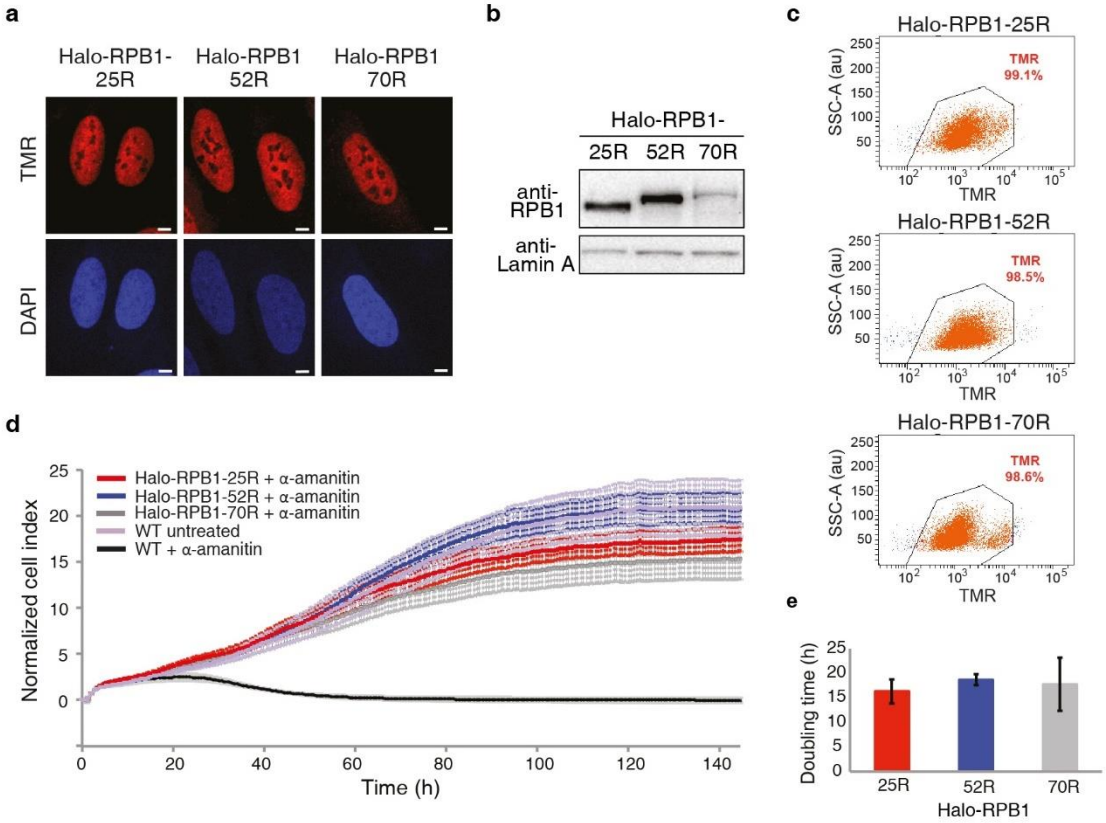

Figure 7.1 | Characterization of Halo-RPB1 cell lines.

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.

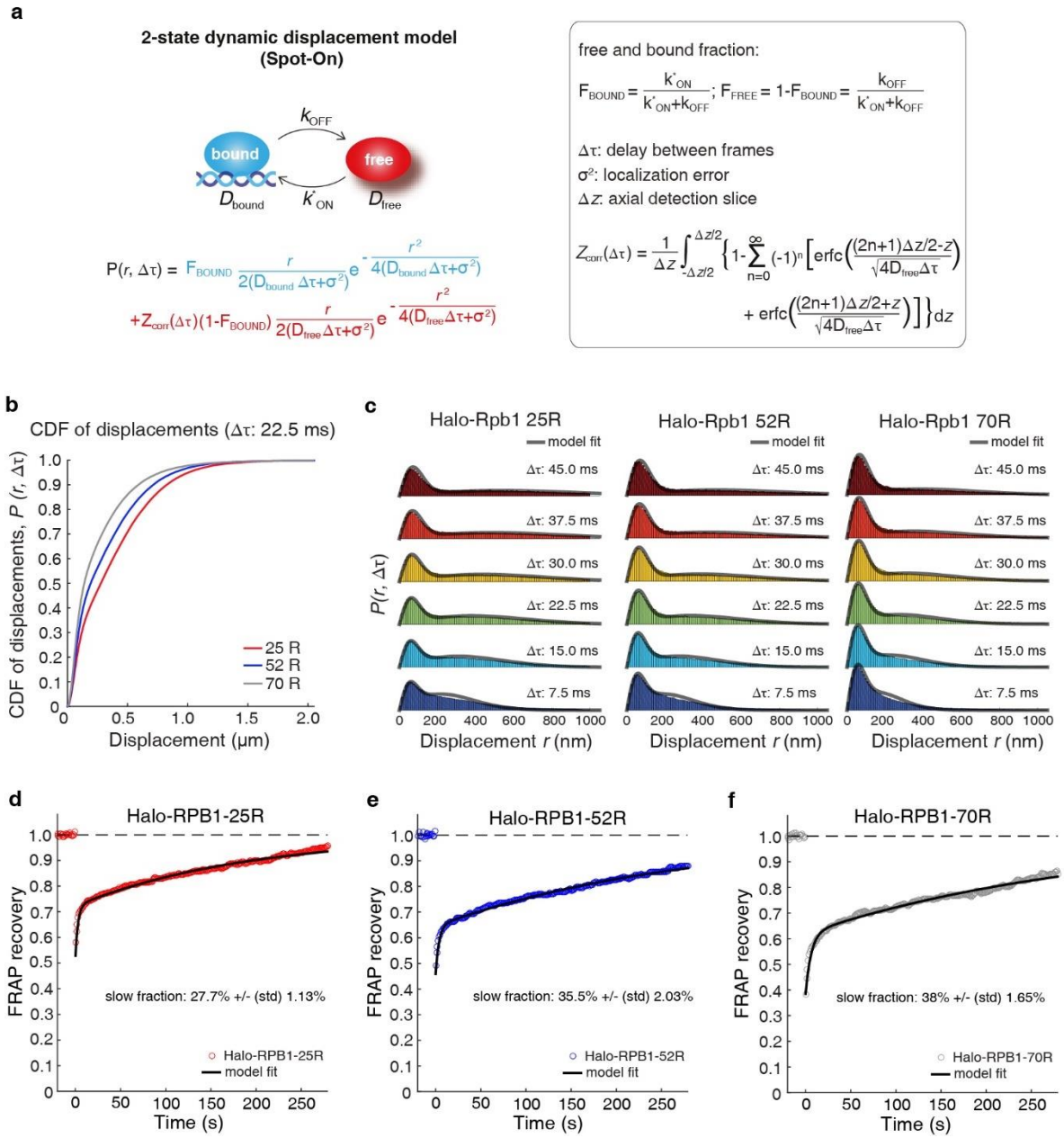

Figure 7.2 | Pol II dynamics determined by spaSPT and FRAP.

a, Overview of the 2-state model used in fitting the displacement data from spaSPT (Spot-On)[293]. 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). *(Figure caption continued on next page.)*

◀ **Figure 7.2 | Pol II dynamics determined by spaSPT and FRAP.** *(Figure caption continued from previous page.)* **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[214, 292]. We performed 50 iterations using 50% of the data in each to estimate the error (standard deviation of the subsampling) on the bound fraction.

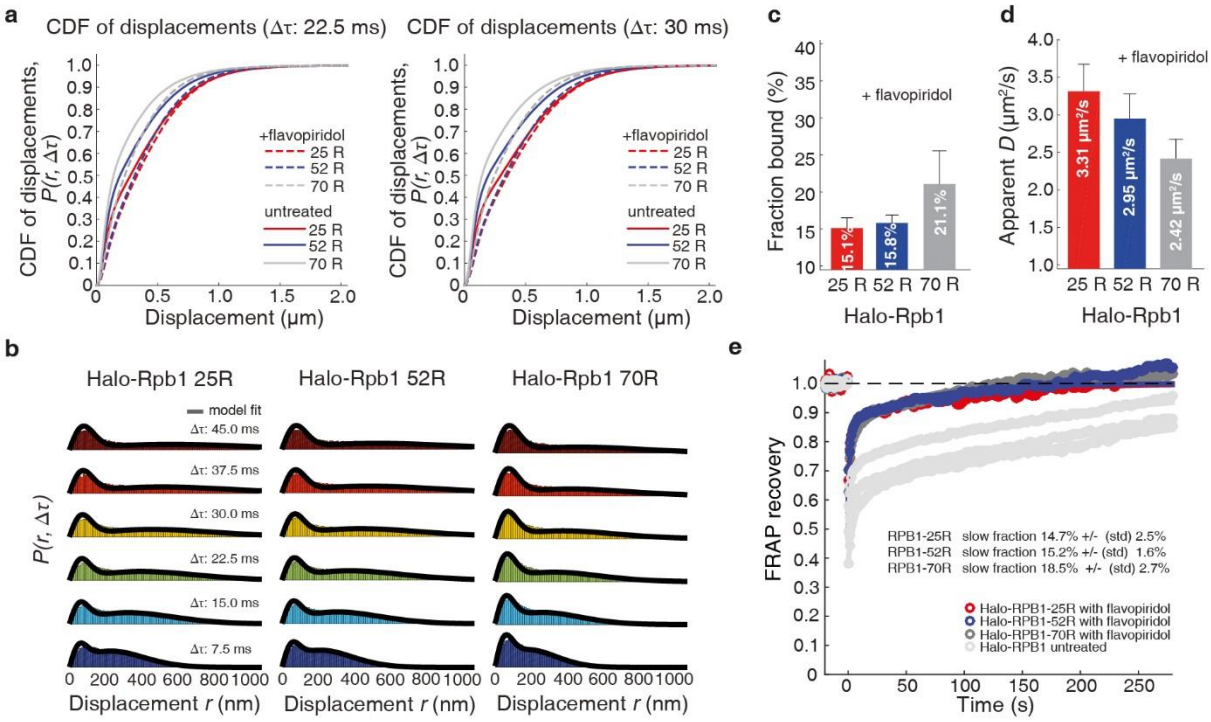

Figure 7.3 | Effect of flavopiridol treatment on RNA Pol II dynamics.

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[293]. 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[293]. 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[214, 292]. 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. 3.8) are overlaid (light grey).

Appendix

List of figures

List of tables

List of items from publications

Parts of this work have been published. A list of excerpted sections and figures can be found below.

M. Boehning*, C. Dugast-Darzacq*, M. Rankovic* et al. (2018). RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nature Structural and Molecular Biology* **25**, 833–840.

Sections

The following sections were taken from Boehning *et al.:*

Methods

2.3.1.3 Purification of TFIIH kinase modules

2.3.1.4 Pol II preparation and fluorescent labeling

2.3.1.5 CTD phosphorylation

2.3.1.6 Kinase activity assay

2.3.1.7 Disorder prediction

2.3.1.8 Differential interference contrast (DIC) and fluorescence microscopy

2.3.1.9 Pol II co-recruitment experiments

2.3.1.10 In vitro FRAP experiments

2.3.1.12 Cell line establishment and characterization (concatenated section containing the published sections ‘Cell line establishment and characterization’, ‘Western blot’, ‘FACS analysis’, ‘xCELLigence analysis’, ‘Doubling time analysis’)

2.3.1.13 Cell imaging conditions

2.3.1.14 PALM imaging

2.3.1.15 Single-molecule imaging (spaSPT)

2.3.1.16 FRAP in cells

Results

3.1.1 CTD of Pol II phase separates into liquid-like droplets

3.1.2 CTD length influences CTD phase separation in vitro

3.1.4 CTD droplet recruit intact Pol II

3.1.5 CTD length controls Pol II clustering in human cells

3.1.6 CTD length influences Pol II dynamics in cells

3.1.7 CTD phosphorylation dissolves droplets

Supplementary Information

7.1 Supplementary Note (Calculation of diffusion coefficients)

Sections that were taken from published manuscripts were adapted to match the style of this thesis. Numbering and references to figures as well as references to the literature thus deviate from Boehning *et al.*

Figures and figure legends

The following figures were taken from Boehning *et al.:*

Figure 1.2 contains items from Supplementary Figure 1a-b (panel a-b) and Figure 1a (panel c)

Figure 3.1 contains items from Figure 1b-e and g-i

Figure 3.2 corresponds to Figure 2

Figure 3.3 corresponds to Supplementary Figure 2

Figure 3.4 corresponds to Supplementary Figure 3

Figure 3.6 corresponds to Figure 3

Figure 3.7 corresponds to Figure 4

Figure 3.8 corresponds to Figure 5

Figure 3.9 corresponds to Figure 6

Figure 3.10 corresponds to Supplementary Figure 7

Figure 3.11 corresponds to Figure 7

Figure 7.1 corresponds to Supplementary Figure 4

Figure 7.2 corresponds to Supplementary Figure 5

Figure 7.3 corresponds to Supplementary Figure 6

All figures and figure legends that were taken from published manuscripts were adapted to match the style of this thesis (e.g. in regard to format, fonts, labeling etc.) and may thus deviate from Boehning *et al.*

List of abbreviations

2xYT	yeast extract tryptone
LB	lysogeny broth
µF	microfarad
µg	microgram
µL	microliter
3D	three dimensional
Å	Ångström
aa	amino acids
ADP	adenosine diphosphate
AF488	Alexa Fluor 488
AF647	Alexa Fluor 647
Asn	asparagine
ATP	adenosine triphosphate
Bis-Tris	bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methan
bp	base pair(s)
Bromo-UTP	5-bromouridine triphosphate
CCD	charge-coupled device
CDF	cummulative distribution function
CDK	cyclin-dependent kinase
cDNA	complementary DNA
cm	centimeter
cryo-EM	cryogenic electron microscopy
c_{sat}	saturation concentration
CSR	complete spatial randomness

CTD	carboxy-terminal domain
C-terminus	carboxy-terminus
CV	column volumes
Cys	cysteine
dCTP	deoxycytosine triphosphate
dGTP	deoxyguanosine triphosphate
DIC	differential interference contrast
DMEM	Dulbecco's Modified Eagle Medium
DMSO	dimethylsulfoxid
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
FBS	fetal bovine serum
fmol	femtomol
FRAP	fluorescence recovery after photobleaching
g	gram
g	standard gravity
GFP	green fluorescent protein
GO	gene ontology
GST	glutathion S-transferase
h	hour(s)
hCTD	human (*Homo sapiens*) CTD
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HRP	horseradish peroxidase

HS	heat-stressed
Hz	hertz
IDR	intrinsically disordered region
IgG	immunoglobulin G
IPTG	isopropyl-β-D-thiogalactopyranosid
kb	kilo base pairs
kDa	kilo dalton
kV	kilovolt
L	liter
lacZ	gene encoding beta-galactosidase
LCD	low complexity domain
LDS	lithium dodecyl sulfate
LLPS	liquid-liquid phase separation
m	meter(s)
MBP	maltose-binding protein
mCh	mCherry
MES	2-(N-Morpholino) ethansulfonsäure
mg	milligram
mL	milliliter
mM	millimolar
MOPS	3-(N-Morpholino) propansulfonsäure
mPEG	methoxy-polyethylene glycol
mRNA	messenger ribonucleic acid
MWCO	molecular weight cut-off
NA	not applicable

NHS	non heat-stressed
nm	nanometer(s)
NP-40	Polyethylene glycol nonylphenyl ether
nt	nucleotide(s)
N-terminus	amino-terminus
OD_{600}	optical density at 600 nm
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PA-JF549	photoactivatable Janelia Fluor 549
PALM	photoactivated localization microscopy
PBS	phosphate buffered saline
PBST	phosphate buffered saline with 0.1% Tween-20
PCR	polymerase chain reaction
PEG	polyethylene glycol
PIC	pre-initiation complex
pmol	picomol
PMSF	phenylmethylsulfonyl fluoride
PMT	photomuliplier tube
Pol II	RNA polymerase II
PolH	polyhedrin
PONDR	Predictor Of Natural Disordered Regions
PSF	point spread function
PTM	post-translational modification
PVDF	polyvinylidene difluoride
r	radius

RIPA	radioimmunoprecipitation assay
RNA	ribonucleic acid
rpm	round per minute
RT	room temperature
s	second(s)
S. cerevisiae	*Sacharomyces cerevisiae*
SDS	sodium dodecylsulfate
SILAC	stable isotope labeling in cell culture
SOC	super optimal broth with catabolite repression
SV40	simian vacuolating virus 40
TAE	tris-acetate-EDTA
TBST	Tris buffered saline with 0.1% Tween
TCEP	tris(2-carboxyethyl)phosphine
TEV	tobacco etch virus

TF	transcription factor
TIRF	total internal reflection fluorescence
TMR	tetramethylrodamine
Tris	tris(hydroxymethyl)-aminomethan
TSS	transcription start site
U	unit(s)
v	volume
V	volt
w	weight
WT	wild-type
X-gal	5-Brom-4-chlor-3-indoxyl-β-D-galactopyranosid
yCTD	yeast (*S. cerevisiae*) CTD
YFP	yellow fluorescent protein
YPD	yeast extract-peptone-dextrose