ARTICLES
https://doi.org/10.1038/s41589-021-00904-5
nature chemical biology

Check for updates

Phase separation of RNA-binding protein promotes polymerase binding and transcription

Wen Shao[1], Xianju Bi[1,6], Yixuan Pan[2,6], Boyang Gao[1,6], Jun Wu[2], Yafei Yin[1], Zhimin Liu[1], Mengyuan Peng[2], Wenhao Zhang[3], Xu Jiang[3], Wenlin Ren[1], Yanhui Xu[1], Zhongyang Wu[1], Kaili Wang[1], Ge Zhan[1], J. Yuyang Lu[1], Xue Han[1], Tong Li[1], Jianlong Wang[4], Guohong Li[5], Haiteng Deng[3], Bing Li[2]✉ and Xiaohua Shen[1]✉

An RNA-involved phase-separation model has been proposed for transcription control. However, the molecular links that connect RNA to the transcription machinery remain missing. Here we find that RNA-binding proteins (RBPs) constitute half of the chromatin proteome in embryonic stem cells (ESCs), some being colocalized with RNA polymerase (Pol) II at promoters and enhancers. Biochemical analyses of representative RBPs show that the paraspeckle protein PSPC1 inhibits the RNA-induced premature release of Pol II, and makes use of RNA as multivalent molecules to enhance the formation of transcription condensates and subsequent phosphorylation and release of Pol II. This synergistic interplay enhances polymerase engagement and activity via the RNA-binding and phase-separation activities of PSPC1. In ESCs, auxin-induced acute degradation of PSPC1 leads to genome-wide defects in Pol II binding and nascent transcription. We propose that promoter-associated RNAs and their binding proteins synergize the phase separation of polymerase condensates to promote active transcription.

Intricate regulation of transcription is central for cell differentiation and development[1,2]. Transcription is thought to take place at discrete nuclear sites known as transcription 'factories', in the form of phase-separated condensates that allow compartmentalization and coupling of polymerases engaged at multiple genomic sites[3–7]. Genome-wide studies have revealed the prevalent binding of RNA Pol II in the promoter-proximal regions of most metazoan genes[8–10]. The activity and release of promoter-bound Pol II into elongation is regulated through the phosphorylation state of an intrinsically disordered C-terminal domain (CTD) of the largest subunit of Pol II[10,11]. Intriguingly, transcription of most active genes occurs in short bursts[12–16]. Live-cell imaging studies have shown transient residence and clustering at the scale of seconds for Pol II that initiates at the promoter[5,13,17,18]. It has been estimated that only one of 100 Pol II–gene interactions proceed to productive elongation[13,18]. Dynamic assembly of Pol II during initiation suggests key regulatory events that are necessary to stabilize Pol II binding for transcription elongation.

Increasing evidence indicates that RNA broadly associates with chromatin and feeds back on transcription and chromatin states[9,19–26]. A phase-separation model of RNA-mediated feedback control appears attractive in explaining the features of transcription processes[27,28]. RNA stimulates transcription factor condensates at low levels but dissolves them at high levels[28,29]. However, this hypothesis remains inconclusive because the key link that connects RNA to the transcriptional machinery with characteristic DNA-binding activity is still missing. It is widely believed that eukaryotic transcription is coupled with RNA processing[10,11,30]. RNA-binding proteins (RBPs) constitute a major family of regulators that process RNA transcripts from synthesis to decay[31]. A number of RBPs, such as WDR43, DDX21/18/5, SRSF1/2, FUS, hnRNPK/U/L, NCL and NONO, have been implicated in modulation of transcriptional, epigenetic and signaling responses in various cellular contexts[32–41]. Nevertheless, the direct involvement of RBPs and their interplay with RNA in transcription regulation remains to be proven.

In this study, proteomic profiling reveals abundant and dynamic associations of RBPs with chromatin in ESCs. Surveys of selected RBPs show that they interact with Pol II and preferentially bind regulatory hotspots across the genome, and their knockdown attenuates global transcription. Importantly, through combined biochemical and cellular analyses, we delineate the role of PSPC1, a representative RBP, in the promotion of Pol II engagement and transcriptional activity. The synergistic interplay between PSPC1 and RNA in facilitating polymerase condensate formation is critically dependent on both the phase-separation and RNA-binding activities of PSPC1—the two biochemical features shared by many chromatin-associated RBPs. The unexpected involvement of RBPs in transcription provides new insights into gene regulation beyond the canonical components of transcription machineries.

Results

RBPs comprise half of the ESC chromatin proteome. To gain a fuller understanding of transcription in the chromatin context, we sought to capture all chromatin-associated proteins by a crosslinking-based method in ESCs (Extended Data Fig. 1a and Methods). Out of 1,357 proteins (histones excluded) detected, 537 are involved in transcription and chromatin-related functions, comprising 25% of protein peptide abundance (Fig. 1a, Extended

[1]School of Medicine and School of Life Sciences, Tsinghua University; Tsinghua-Peking Joint Center for Life Sciences, Beijing, China. [2]Department of Biochemistry and Molecular Cell Biology, Shanghai Key Laboratory for Tumor Microenvironment and Inflammation, Shanghai Jiao Tong University School of Medicine, Shanghai, China. [3]Key Laboratory of Bioinformatics, School of Life Sciences, Tsinghua University, Beijing, China. [4]Department of Medicine, Columbia Center for Human Development, Columbia University Irving Medical Center, New York, NY, USA. [5]National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing, China. [6]These authors contributed equally: Xianju Bi, Yixuan Pan, Boyang Gao ✉e-mail: bingli@shsmu.edu.cn; xshen@tsinghua.edu.cn

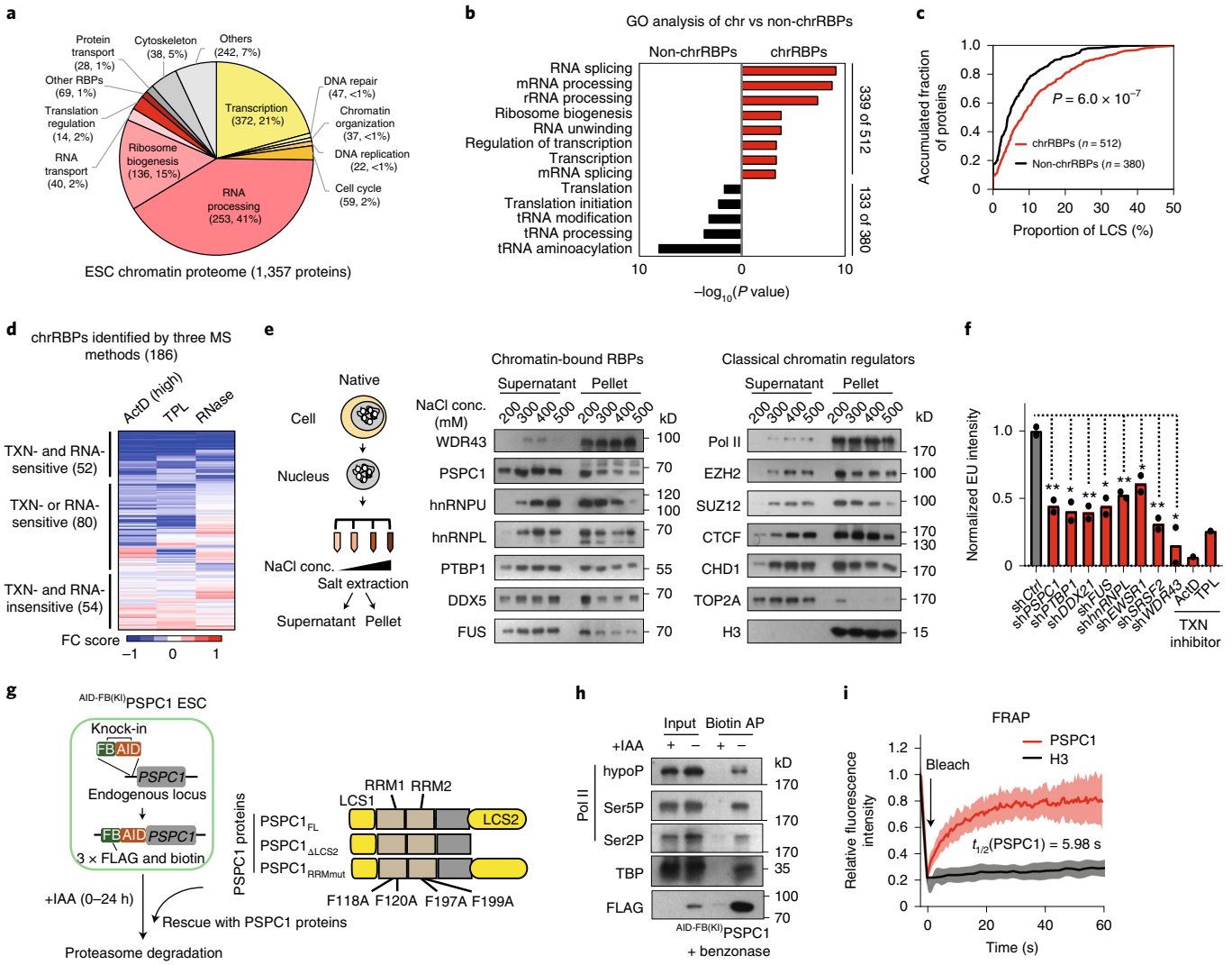


Fig. 1 | Abundant and dynamic associations of RBPs with chromatin. a, Percentages of peptide abundance of chromatin proteins. Protein number and mean peptide abundance of two independent biological replicates (indicated by iBaq ratio; Methods) are shown in parentheses. **b**, Gene Ontology (GO) analysis of chrRBPs ($n = 512$) versus non-chrRBPs ($n = 380$). Total RBPs expressed in ESCs ($n = 892$) were used as background. The numbers of RBPs associated with corresponding terms and total RBPs analyzed are indicated on the right. *P* values, two-sided Fisher's exact tests performed by DAVID. **c**, Cumulative distribution curve showing the LCS content in chrRBPs and non-chrRBPs. *P* values, two-sided Kolmogorov–Smirnov test. **d**, Heatmap showing the FC score (Methods) of chromatin abundance for 186 chrRBPs identified by three MS approaches. Data are shown as the mean of four biological replicates for ActD (1 μg ml^{-1}) and RNase and three replicates for TPL (1 μM). **e**, Protein analysis of native ESC chromatin. Left: pipeline for biochemical extraction; middle and right: immunoblot analysis. **f**, Quantitative analysis of 5-EU incorporation by fluorescence activated cell sorting (FACS) following depletion of RBPs in ESCs. The *y* axis shows the average EU intensity normalized to control cells treated with scramble short hairpin RNA (shCtrl). Data shown as mean of two independent shRNAs. $^{*}P < 0.05$, $^{**}P < 0.01$ by two-sided Student's *t*-test. $P = 0.0094$ (sh*PSPC1*), 0.0157 (sh*PTBP1*), 0.0087 (sh*DDX21*), 0.0145 (sh*FUS*), 0.0062 (sh*hnRNPL*), 0.0283 (sh*EWSR1*), 0.0056 (sh*SRSF2*), 0.0306 (sh*WDR43*). **g**, Schematic diagram of $^{AID\text{-}FB(KI)}$PSPC1 knock-in ESCs and the rescue strategy with wild-type and mutant PSPC1 proteins. **h**, Biotin-mediated affinity purification (AP) in $^{AID\text{-}FB(KI)}$PSPC1 ESCs. IAA treatment (24 h) was used as the negative control. **i**, FRAP analysis of mCherry-PSPC1 puncta in ESCs; GFP-tagged histone H3 serves as the control. The *y* axis shows fluorescence intensity normalized to the initial level. Data shown as mean ± s.d. of ten biological replicates. $t_{1/2}$, half-life. **e**,**h**, Representative results of two independent experiments are shown.

Data Fig. 1b,c and Supplementary Table 1). Congruent with previous proteomic analysis[42], RNA-binding proteins are significantly enriched (Extended Data Fig. 1d; $P < 1 \times 10^{-10}$). By intersecting with Tuschl's RBP repertoire[31], we defined the 512 overlapping proteins as chromatin-bound RBPs (chrRBPs), which accounts for 62% of protein abundance on chromatin (Fig. 1a and Extended Data Fig. 1e). These chrRBPs are enriched in nuclear processes, including messenger RNA splicing and ribosomal RNA processing, in comparison to transfer RNA and translation-related functions for non-chromatin RBPs (380 proteins) (Fig. 1b). Analysis of the proteins pulled down in vitro by the CTD of Pol II[43] revealed a large proportion (62%, 318) of chrRBPs (>twofold enrichment) in CTD interactomes, compared to 32% (123) of non-chromatin RBPs (Supplementary Table 2). In addition, chrRBPs are more positively charged with higher isoelectric points and, intriguingly, exhibit significantly higher contents of low-complexity sequences (LCS) and intrinsically disordered regions[44,45] (IDR) (Fig. 1c and Extended Data Fig. 1f).

Treatments that inhibit transcription or degrade RNA dramatically attenuated RBP-chromatin associations but had less effect on transcription factors and epigenetic enzymes (Extended Data Figs. 1a and 2a,b). Among the 186 chrRBPs consistently detected by three quantitative mass spectrometry (MS) methods, the majority (71%, 132) exhibited reduced chromatin association in response to at least one treatment (Fig. 1d, Extended Data Fig. 2c, Supplementary Table 3 and Methods). Validation of individual proteins showed that the majority (eight out of ten) fell off the chromatin following treatment with RNases or transcription inhibitors (Extended Data Fig. 2d). Detection of PSPC1 and DDX21 in the chromatin fraction may have resulted from incomplete digestion by RNases. Thus, most RBPs appear to be dynamically recruited to chromatin by RNA and/or transcription, ruling out a potential crosslinking artifact. Indeed, we tested seven chrRBPs under non-crosslinking conditions and found that they all exhibited strong chromatin binding at 200 mM salt, similar to that observed for epigenetic factors (Fig. 1e).

chrRBPs interact with Pol II and modulate transcription. To explore a potential role of chrRBPs in transcription, we knocked down a number of them, including PSPC1, PTBP1, DDX21, FUS and hnRNPL. Their depletion caused global reduction of nascent transcripts that were pulse-labeled by 5-ethynyl uridine (EU) (Fig. 1f and Extended Data Fig. 3a,b). Coimmunoprecipitation (co-IP) of Pol II captured all eight chrRBPs tested, including PSPC1, PTBP1 and hnRNPL, in ESC lysates treated with benzonase, which degrades DNA (Extended Data Fig. 3c,d). We reported previously that the paraspeckle protein PSPC1 regulates retroviral ERVL and associated genes via TET2 but is negligible regarding ESC self-renewal[37]. Paraspeckles are absent in mouse ESCs[46]. To elucidate transcriptional effects not complicated by cell survival, we chose PSPC1 as a representative RBP for in-depth characterization.

We constructed homozygous knock-in ESCs carrying an in-frame insertion of FLAG and biotin tags fused with an auxin-inducible degron (AID) epitope at the 5′ end of *PSPC1* (referred to as $^{AID\text{-}FB(KI)}$PSPC1; Fig. 1g). With this cellular platform, we could simultaneously tag and degrade the endogenous PSPC1 protein. Reciprocal co-IPs showed that $^{AID\text{-}FB(KI)}$PSPC1 interacts with the initiating Pol II (hypophosphorylated (hypoP) and phosphorylated at serine 5 (Ser5P) and with TATA-box binding protein (TBP) (Fig. 1h and Extended Data Fig. 3c,e). TBP is the first protein that binds to DNA to initiate assemblage of the preinitiation complex (PIC)[2]. Notably, these interactions were abolished in the presence of both benzonase and RNase, but not by benzonase alone when RNA was incompletely degraded (Extended Data Fig. 3e). Of note, PSPC1 captured the elongating Ser2P Pol II but not vice versa. In addition, PSPC1 exhibited punctate immunofluorescence signals that partially overlapped with Pol II and TBP in the nucleus (Extended Data Fig. 3f,g). PSPC1 puncta showed rapid fluorescence recovery (~5.98 s) after photobleaching compared to histone H3 (Fig. 1i and Extended Data Fig. 3h). Treatment with 1,6-hexanediol, which inhibits weak hydrophobic interactions, dissolved PSPC1 puncta that may have been held together by labile and dynamic interactions (Extended Data Fig. 3i).

PSPC1 promotes CTD phase behavior and phosphorylation. PSPC1 has 66% IDR content with a large LCS domain (named LCS2) at the carboxyl terminus, enriched in glycine (G) and proline (P) (Fig. 1g and Extended Data Fig. 4a). We found that recombinant full-length $PSPC1_{FL}$ protein formed spherical liquid-like droplets around its estimated nuclear concentration of 5 μM (Extended Data Fig. 4b–d, Supplementary Table 1 and Methods). In comparison, recombinant TBP protein (51% IDR) formed fiber-like irregular aggregates in the absence of dextran but was able to form liquid-like droplets in the presence of dextran at a concentration of 5 μM, which is ~15–30-fold above its estimated nuclear concentration of ~0.06–0.3 μM (Extended Data Fig. 4b,c,e). Given the well-recognized role of TBP in transcription initiation, we regarded TBP droplets as a surrogate for the more complex in vivo initiation condensates.

Recombinant CTD (with 20 heptad repeats) failed to phase separate on its own but was incorporated into PSPC1 and TBP droplets at 0.6 μM, which is around the estimated nuclear concentration of Pol II (Extended Data Fig. 3c–e). Simultaneous addition of $PSPC1_{FL}$ and TBP resulted in the formation of larger and brighter droplets that incorporated more CTD inside, compared to TBP alone (Fig. 2a and Extended Data Fig. 4f,g). These droplets exhibited liquid-like fusion behavior and were quickly dissolved by 1,6-hexanediol (Extended Data Fig. 4h,i). Droplet sedimentation analysis also confirmed that ~threefold more CTD proteins were trapped inside $PSPC1_{FL}$-TBP-CTD droplets compared with TBP-CTD droplets (Fig. 2b and Extended Data Fig. 4j,k).

To investigate how the LCS- and RNA-binding domains are involved in PSPC1 function, we generated two mutant proteins (Fig. 1g). $PSPC1_{\Delta LCS2}$ lacks LCS2 while $PSPC1_{RRMmut}$ carries four point mutations (F118A, F120A, K197A and F199A) in the RNA recognition motifs (RRMs)[37]. Both mutants failed to form droplets in the absence of dextran, and poorly phase separated in its presence (Extended Data Fig. 4b,l). $PSPC1_{\Delta LCS2}$ failed to affect TBP-CTD droplets, while $PSPC1_{RRMmut}$ had a weaker effect than $PSPC1_{FL}$ (Fig. 2a,b and Extended Data Fig. 4f,g,k). These results indicate that LCS-mediated phase separation contributes largely to the effect of PSPC1 in promotion of CTD incorporation.

Hyperphosphorylation of the Pol II CTD is required for its activity and release in cells[10]. In the presence of recombinant CTD kinases cyclin-dependent kinase 7 (CDK7) or CDK9, $PSPC1_{FL}$ protein markedly enhanced CTD phosphorylation in a PSPC1 dose-dependent manner whereas $PSPC1_{\Delta LCS2}$ and bovine serum albumin (BSA) had no effect (Fig. 2c and Extended Data Fig. 5a). In accordance with increased CTD phosphorylation, $PSPC1_{FL}$ led to a more rapid release of CTD from TBP-$PSPC1_{FL}$ droplets compared to those containing TBP alone (Fig. 2d, Extended Data Fig. 5b,c and Supplementary Video 1). $PSPC1_{FL}$ skewed the release rate curve from a peak of ~37 to ~10 min following the addition of ATP (Extended Data Fig. 5b and Methods). Droplet sedimentation analysis also confirmed an accelerated release of phosphorylated CTD to 15 min—the earliest time point analyzed (Fig. 2e and Extended Data Fig. 5d). We note that CDK9-mediated phosphorylation did not affect the phase separation of TBP and/or PSPC1 (Extended Data Fig. 5c). Thus, PSPC1 accelerates CDKs-mediated phosphorylation and release of CTD, probably through a phase-separation mechanism.

PSPC1 synergizes with RNA to promote CTD phase behavior. The addition of total RNA from ESCs promoted the formation of PSPC1 droplets in a manner dependent on PSPC1 and RNA concentrations (Extended Data Figs. 4d and 5e). We noted that high RNA levels led to smaller droplets and the appearance of irregular, fiber-like aggregates. This suggests that only within the range of balanced RNA–protein interactions does RNA act as a multivalent ligand to promote PSPC1 phase behavior. Note that $PSPC1_{RRMmut}$ and TBP with minimal or no RNA-binding activity appeared to be less sensitive to RNA (Extended Data Fig. 4e,l).

Next, we tested the effects of RNA on the condensate-interacting behaviors of the CTD. Interestingly, in the absence of PSPC1, RNA led to a gradual loss of CTD fluorescence from TBP droplets in an RNA dosage-dependent manner, while TBP droplets were not affected (Fig. 2f (left),g and Extended Data Fig. 5f,g (gray)). We postulate that negative charges on RNA may evict CTD, mimicking phosphorylation-induced release of CTD. Strikingly, the addition of $PSPC1_{FL}$ not only completely blocked RNA-induced eviction of the CTD, but also dramatically increased CTD incorporation

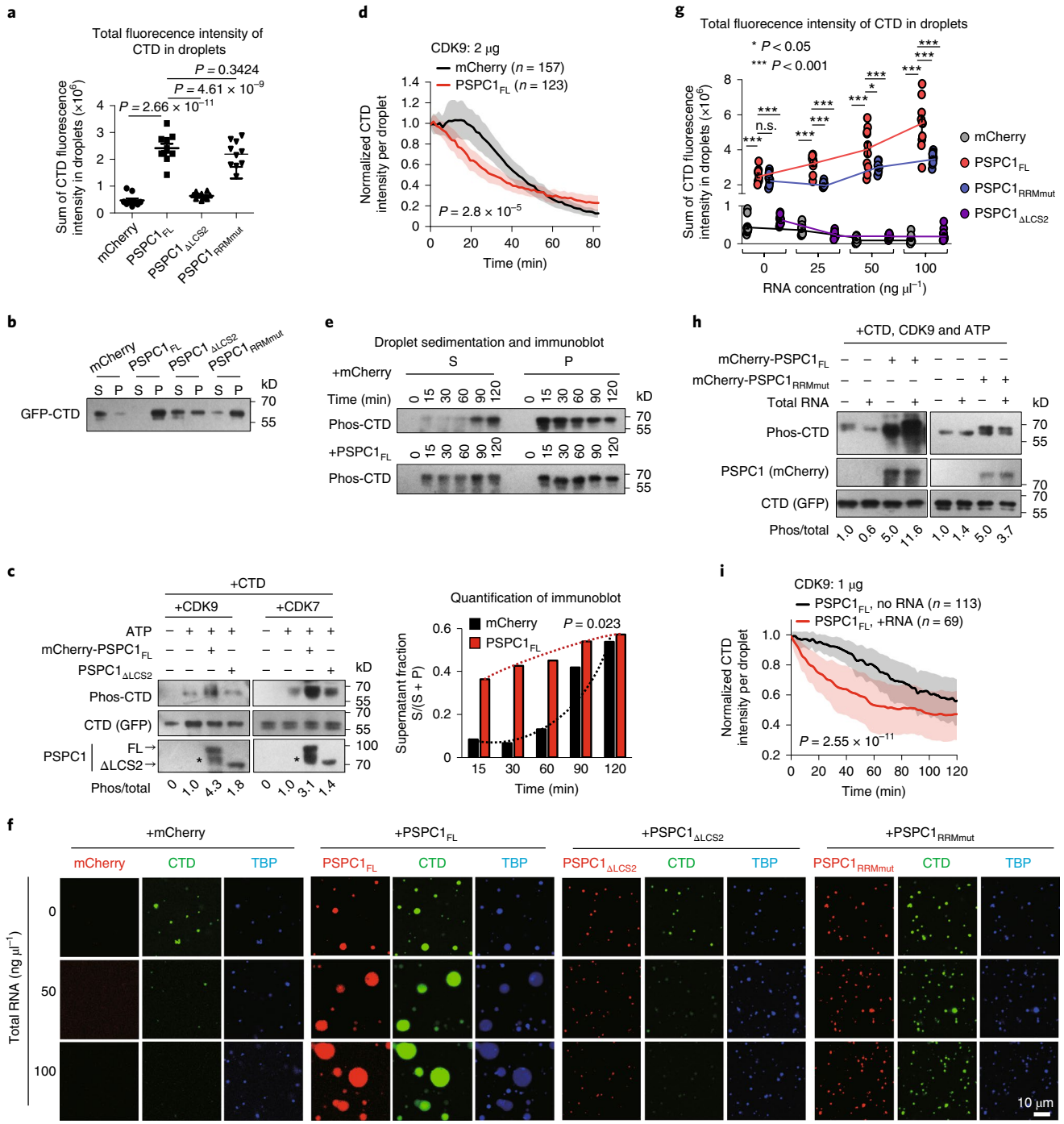


Fig. 2 | RNA synergizes with PSPC1 in promotion of CTD incorporation, phosphorylation and release. a,**b**, Droplet assay by imaging quantification (**a**) and sedimentation and immunoblot analysis (**b**). TBP, 5 μM; CTD, 0.6 μM; PSPC1, 5 μM; mCherry, 5 μM. Representative images are shown in Extended Data Fig. 4f. *P* values indicated as shown. S, supernatant; P, pellet. Quantification with additional replicates is shown in Extended Data Fig. 4k. **c**, Kinase assay. mCherry, mCherry-tagged $PSPC1_{FL}$ and $PSPC1_{\Delta LCS2}$, 5 μg protein; CTD and CDK9/CDK7, 0.2 μg; ATP, 0.1 mM. Asterisks indicate a truncated PSPC1 protein generated during purification. An additional replicate is shown in Extended Data Fig. 5a. **d**,**e**, Time-lapse analysis of CTD release by imaging (**d**) and by sedimentation and immunoblot assay (**e**). **d**, Data shown as mean ± s.d. of droplets as indicated. CDK9, 2 μg. Representative images are shown in Extended Data Fig. 5c. **e**, Quantification of supernatant fractions (bottom) based on immunoblot (top). *P* values, two-tailed Student's paired *t*-test ($n = 5$ time points). **f**,**g**, Effects of RNA and PSPC1 proteins on CTD incorporation into TBP droplets. **f**, Representative images; **g**, quantification. *P* values shown in the sequence 0, 25, 50 and 100 ng μl^{-1}. mCherry versus $PSPC1_{FL}$, 2.66×10^{-11}, 1.43×10^{-12}, 5.26×10^{-13}, 4.77×10^{-10}; $PSPC1_{FL}$ versus $PSPC1_{RRMmut}$, 0.1919, 1.22×10^{-6}, 0.02, 5.32×10^{-4}; $PSPC1_{FL}$ versus $PSPC1_{\Delta LCS2}$, 2.46×10^{-11}, 4.02×10^{-13}, 1.00×10^{-8}, 7.31×10^{-10}. **h**, RNA and PSPC1 proteins following CTD phosphorylation. A representative result of two independent experiments is shown. **i**, RNA following CTD release by time-lapse imaging. RNA, 50 ng μl^{-1}; CDK9, 1 μg. Data shown as mean ± s.d. of a number (*n*) of droplets as indicated. **a**,**g**, *y* axis shows the sum of fluorescence intensity of CTD within droplets in each field ($n = 10$). *$P < 0.05$, ***$P < 0.001$ by two-sided Student's *t*-test. NS, not significant. **d**,**i**, *y* axis shows CTD intensity normalized to TBP intensity within individual droplets (Methods). *P*-values, two-sided Kolmogorov–Smirnov test. **a**,**b**,**d**–**g**,**i**, 10% dextran was added.

into TBP droplets (Fig. 2f (second from left),g and Extended Data Fig. 5f,g (red)). In contrast, PSPC1$_{\Delta LCS2}$ failed to block CTD eviction by RNA (Fig. 2f (left),g and Extended Data Fig. 5f,g (purple). Although PSPC1$_{RRMmut}$ prevented RNA-induced eviction of the CTD, TBP-PSPC1$_{RRMmut}$ droplets remained small in size and the levels of incorporated CTD did not scale with RNA concentrations (Fig. 2f(right),g and Extended Data Fig. 5f,g (blue).

Interestingly, CDK9-mediated phosphorylation of the CTD was dramatically enhanced ~12–19-fold following the simultaneous addition of PSPC1$_{FL}$ and RNA, compared to a five- to eight-fold increase with PSPC1$_{FL}$ alone, whereas RNA alone had no effect (Fig. 2h). By comparison, PSPC1$_{RRMmut}$ led to a moderate increase (~2.6–5.0-fold) regardless of RNA. Moreover, the release of phosphorylated CTD was synergistically enhanced by the presence of both PSPC1$_{FL}$ and RNA (Fig. 2i, Extended Data Fig. 6a,b and Supplementary Videos 2 and 3). Taking these results together, PSPC1 not only neutralizes the charge effect of RNA that may expel CTD but also makes use of RNA to promote phase separation, thereby efficiently compartmentalizing the CTD for enhanced phosphorylation and release in the presence of CDKs. The synergistic interplay between PSPC1 and RNA is critically dependent on the LCS and RRM domains of PSPC1 (Extended Data Fig. 6c).

PSPC1 promotes Pol II binding and transcription in vitro. Next, we examined the effects of PSPC1 in a fully defined in vitro transcription system (Fig. 3a and Extended Data Fig. 7a,b). We utilized a DNA template containing a heteroduplex bubble, which has been widely used as a nucleic acid scaffold in structural studies[47]. Pol II can bind to single-stranded DNA within the bubble without the help of general transcription factors. As expected, autoradiography detected a 278-nt, full-length RNA transcript in the presence of a full set of nucleoside triphosphates (+NTP) (Extended Data Fig. 7c). When GTP was omitted (+NTP-GTP), a 33-nt, G-less transcript was detected due to Pol II stalling at the triple-C site. Intriguingly, prominent signals of short discrete transcripts in the range 15–25 nt were detected, indicating high-level abortive transcription in vitro. This observation is consistent with inefficient transcription onset observed in the nucleus[13,18,48]. Importantly, the addition of PSPC1$_{FL}$ led to moderate but consistent increases in both abortive and full-length transcripts in the presence of NTPs (Fig. 3b,c and Extended Data Fig. 7d).

To investigate how PSPC1 promotes in vitro transcription, we performed EMSA to measure Pol II binding to the template by quantifying the supershifted Pol II:DNA signal. We optimized the binding assay by titrating heparin and double-stranded DNA competitors to minimize loosely docked Pol II, and to prevent nonspecific binding of PSPC1$_{FL}$ to the bubble template (Extended Data Fig. 7e,f). PSPC1 did not form a stable complex with Pol II, suggesting weak interactions between them. Transcription led to gradually decreased signals of the supershifted Pol II:DNA band, from the docking Pol II to the stalled (+NTP, –green fluorescent protein (GFP)) and engaged Pol II (+ NTP) (Fig. 3d (lanes 3–5) and Extended Data Fig. 8a (lanes 3–5 and 9–11). This pattern of decrease is in line with the frequent fall-off of Pol II observed during nuclear transcription[13,18], and is also consistent with a report that RNAs nonspecifically inhibit Pol II binding to the DNA template in vitro[49].

Interestingly, the addition of PSPC1$_{FL}$ consistently enhanced Pol II:DNA signals for all three states of Pol II (Fig. 3d,e and Extended Data Fig. 8a,b). This enhancement was more obvious for stalled and engaged Pol II (~twofold) than docking Pol II (~1.2-fold), implying a cooperative effect of PSPC1 with RNA to engage Pol II to the template. In contrast, PSPC1$_{\Delta LCS2}$ and PSPC1$_{RRMmut}$ had negligible effects (Fig. 3e and Extended Data Fig. 8b,c). Together, these results demonstrate that PSPC1 directly promotes Pol II–DNA engagement during initial loading, stalling and elongation to enhance RNA production.

Notably, Pol II:DNA supershifted signals were also increased by the addition of PTBP1 (ref. [50]) that binds pre-mRNA introns (Fig. 3f and Extended Data Fig. 8d). Recombinant PTBP1 (41% IDR) incorporated the CTD within its phase-separated droplets in an RNA dose-dependent manner (Extended Data Fig. 8e). In comparison, recombinant proteins that do not simultaneously possess both RNA-binding and phase-separation activities failed to enhance Pol II binding to the template. Besides PSPC1$_{RRMmut}$ and PSPC1$_{\Delta LCS2}$, this list includes mCherry, the LCS domains of hnRNPL (hnRNPL$_{LCS}$) and DDX21 (DDX21$_{LCS}$[39]) and isocitrate dehydrogenase IDH1 (14% IDR), which binds to GA- or AU-rich RNA[24]. Based on these results, we postulate that many chrRBPs might act similarly to PSPC1 and PTBP1 to promote Pol II engagement and transcription via their RNA-binding and phase-separation activities.

LCS- and RNA-binding activity target PSPC1 to promoters. To explore the in vivo function of PSPC1 in transcription, we first mapped its chromatin-binding sites by chromatin immunoprecipitation followed by sequencing (ChIP–seq). The overall targets of endogenously and ectopically tagged PSPC1 are highly similar ($P < 2.2 \times 10^{-16}$ by Fisher's exact test; Extended Data Fig. 9a, Supplementary Table 4 and Methods). Among a total of 11,589 overlapping peaks, 53% are localized in the promoters of 5,262 genes and 6.1% are in enhancers (Fig. 4a,b). PSPC1 targets overlap extensively with those of initiating (hypoP and Ser5P) Pol II (Extended Data Fig. 9b–e). Enriched binding of PSPC1 at the transcriptional start site (TSS) mimics that of hypoP Pol II, and is positively correlated with active histone marks and gene expression (Fig. 4c–e).

To profile the RNA targets of PSPC1, we performed FLAG- and biotin-tandem purification-mediated crosslinking–immunoprecipitation (CLIP–seq)[39] (crosslinking and immunoprecipitation followed by sequencing). PSPC1 CLIP–seq reads were strongly enriched at the TSS and around PSPC1 ChIP–seq signals (Fig. 4d–h and Extended Data Fig. 9e,f). Heatmap analysis showed a genome-wide coincidence of ChIP and CLIP read distributions of PSPC1 (Fig. 4e). In support of RNA-directed targeting of PSPC1, the RNA-binding mutant PSPC1$_{RRMmut}$ showed abolished chromatin binding by ChIP–seq (Fig. 4d,i and Extended Data Fig. 9g) and exhibited attenuated associations with Pol II and TBP by co-IP (Fig. 4j).

In comparison, PSPC1$_{\Delta LCS2}$ exhibited genome-wide reduction at TSSs, albeit it still interacted with Pol II and TBP. In addition, treatment of ESCs with 1,6-hexanediol abolished PSPC1 binding to its targets (Extended Data Fig. 9h). Both PSPC1 mutants showed diffuse nuclear distribution in contrast to punctate staining of PSPC1$_{FL}$ (Extended Data Fig. 9i). Taken together, these results led us to conclude that efficient targeting of PSPC1 to chromatin requires both LCS- and RNA-binding domains, the latter of which appears to be essential for its binding to gene promoters.

Acute degradation of PSPC1 impairs transcription. Next, we used $^{AID\text{-}FB(KI)}$PSPC1 ESCs to examine the primary effects of PSPC1 degradation. Addition of the auxin analog indole-3-acetic acid (IAA) induced rapid degradation of PSPC1 protein, which was reduced to <40% at 2 h and became barely detectable at 4 h (Fig. 5a). The protein levels of phosphorylated Pol II, but not of total Pol II, were dramatically decreased to 20–30% at 4 h. Levels of Pol II phosphorylation recovered after prolonged IAA treatment, which suggests compensatory mechanism(s) that safeguard steady-state Pol II activity.

Consistently, ChIP–seq showed reduced binding of Ser5P Pol II at the TSS and elongatation of Pol II (Ser2P) across the gene body and downstream regions after 3 and 6 h of IAA treatment (Figs. 4d and 5b,c and Extended Data Fig. 9j). The degree of downregulation in Pol II ChIP–seq signals was positively correlated with that of PSPC1 (Fig. 5d). Importantly, transient expression of full-length PSPC1$_{FL}$, but not of mutants PSPC1$_{\Delta LCS2}$ or PSPC1$_{RRMmut}$, rescued

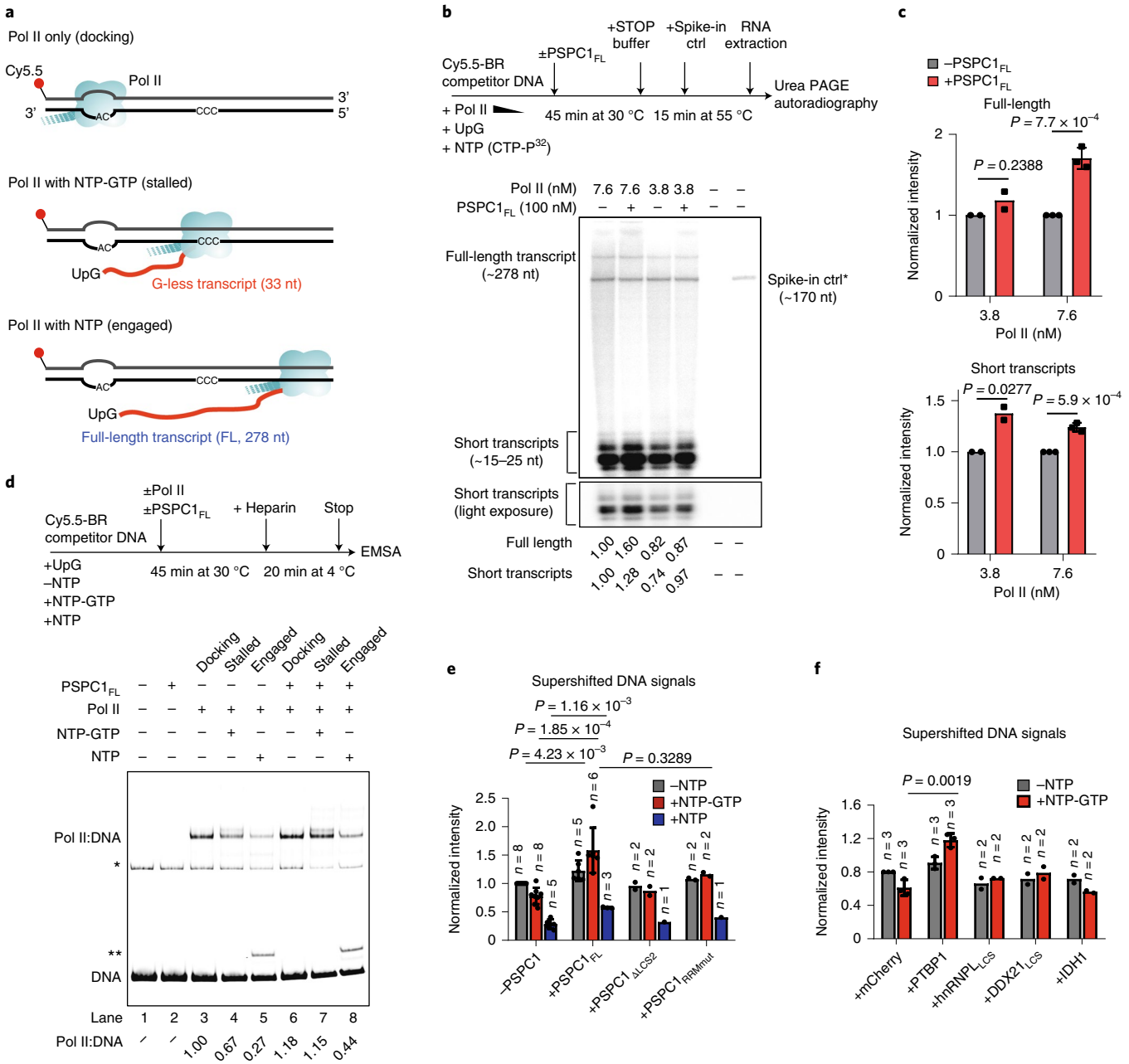


Fig. 3 | PSPC1 promotes Pol II engagement and activity during transcription in vitro. a, Schematic diagram of the in vitro transcription assay. Top: Pol II loosely binds to the bubble in the absence of NTPs; middle: Pol II initiates transcription and pauses at the CCC site, producing a 33-nt transcript when NTPs, but not GTP, were added (+NTP-GTP); bottom: Pol II runs off the template and produces a ~278-nt transcript in the presence of a full set of NTPs. **b**, Autoradiography of in vitro transcription. Relative intensities of full-length/short transcripts normalized to the spike-in control (ctrl) (*) are indicated at the bottom. **c**, Summary of relative intensities of in vitro transcribed RNA. The *y* axis shows intensity normalized to the spike-in control and to the parallel reaction without $PSPC1_{FL}$. Data shown are mean ± s.d. *P* values by two-sided Student's *t*-test indicated as shown. For 3.8 and 7.6 nM of Pol II, two and three biological replicates were performed, respectively. Each dot indicates an independent experiment. **d**, EMSA showing Pol II binding to the BR template during in vitro transcription. The free template (DNA) and the supershifted Pol II:DNA bands are indicated. Bands marked by a single asterisk are a nonspecific byproduct of gel purification; bands marked by a double asterisk are likely to be DNA:RNA hybrids (R-loop). Relative Pol II:DNA-binding intensities are indicated at the bottom. **e**,**f**, Quantification summary of supershifted Pol II:DNA signals in independent biological replicates (+ heparin). **e** shows effects of PSPC1 proteins on Pol II binding while **f** shows effects of other RBPs. The *y* axis shows band intensity normalized to either mCherry control or the reaction without PSPC1. Data shown as mean ± s.d. of a number of biological replicates as indicated. *P* values, two-sided Student's *t*-test, indicated as shown..

the genome-wide reduction in Ser2P binding (Figs. 4d and 5e and Extended Data Fig. 9e). In accordance with the in vitro experiments (Fig. 3), these results indicate that PSPC1 utilizes its phase-separation and RNA-binding activities to stabilize Pol II binding in vivo.

Transient transcriptome sequencing (TT–seq) of nascent transcripts labeled with 4-thiouridine (4sU) revealed downregulated transcription that occurred at the early time point of 3 h following the addition of IAA (Figs. 4d and 5f and Extended Data Fig. 9e).

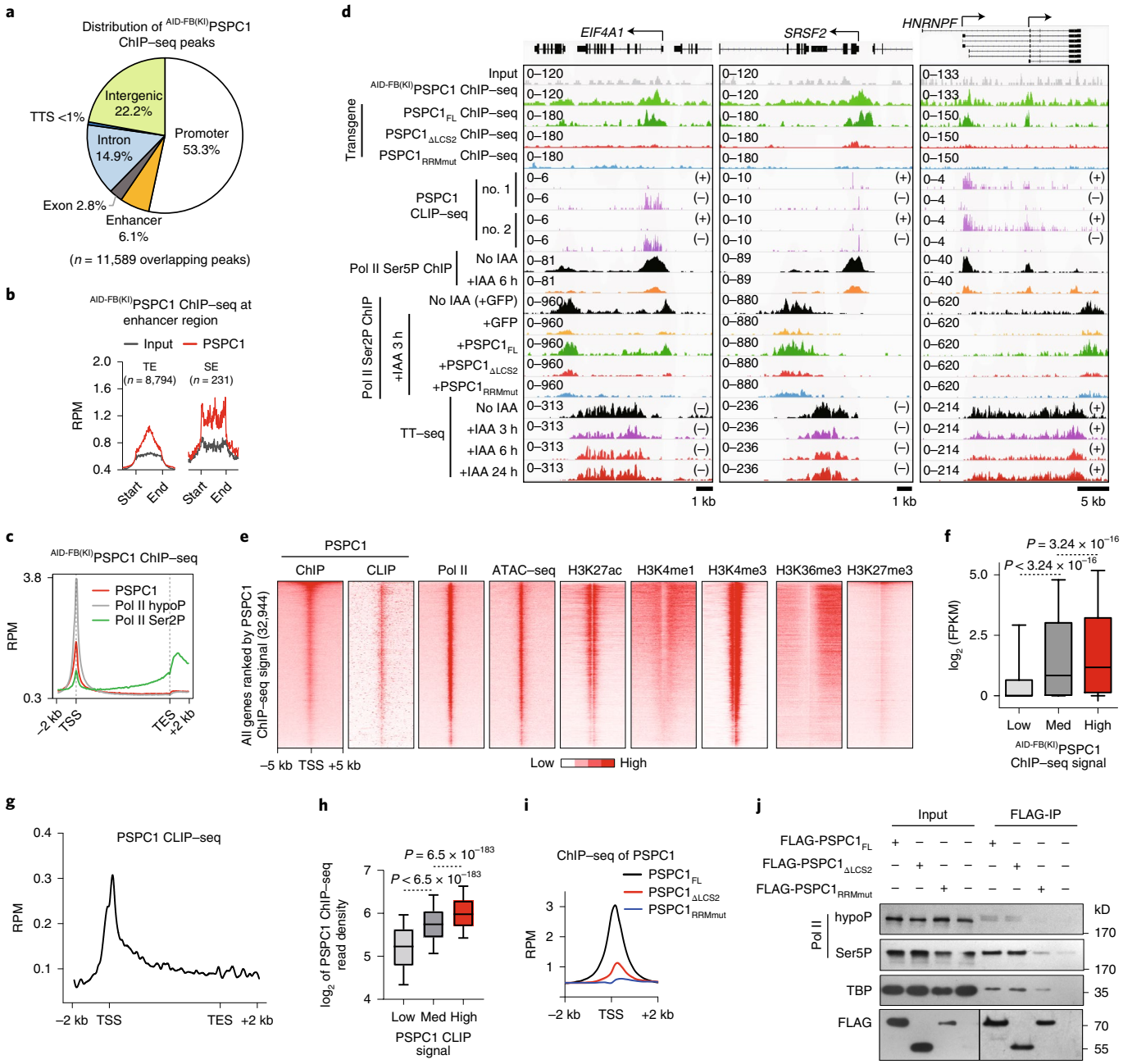


Fig. 4 | Genome-wide colocalization of PSPC1 with Pol II is dependent on the RNA-binding and phase-separation activities of PSPC1. a, Distribution of AID-FB(KI)PSPC1 ChIP-seq peaks. About 11,589 overlapping peaks were detected in two biological replicates. **b**, Metagene analysis of ChIP-seq signals for AID-FB(KI)PSPC1 across enhancers. TE, typical enhancers ($n = 8,704$); SE, super-enhancers ($n = 231$). **c**, Metagene analysis showing AID-FB(KI)PSPC1 and Pol II ChIP-seq signals across all mouse genes ($n = 32,944$). **d**, University of California at Santa Cruz (UCSC) genome browser view of ChIP-seq, CLIP-seq and TT-seq at representative loci. **e**, Heatmap analysis around TSS (±5 kb) across all mouse genes. Heatmaps were sorted by PSPC1 ChIP-seq signals. **f**, Correlation of PSPC1 ChIP-seq signals with gene expression. The *y* axis is $\log_2$(fragments per kilobase exon per million mapped reads). **g**, Metagene analysis of PSPC1 CLIP-seq signals across all mouse genes ($n = 32,944$). **h**, Correlation analysis between PSPC1 CLIP-seq and ChIP-seq signals. The *y* axis is $\log_2$ of PSPC1 ChIP-seq read density. **i**, ChIP-seq and metagene analysis of PSPC1 proteins transiently expressed in ESCs around the TSS of all mouse genes ($n = 32,944$). **j**, Co-IP analysis of PSPC1 proteins with Pol II and TBP. This experiment was repeated twice, with similar results. **b**,**c**,**g**,**i**, The *y* axis shows reads per million reads (RPM). **g**,**i**, Data shown as mean of two biological replicates. *P* values, two-sided Kolmogorov–Smirnov test. **f**,**h**, All genes ($n = 32,944$) classified equally into three groups ($n = 10,981$ for each group) according to PSPC1 ChIP-seq (**f**) or CLIP-seq signals (**h**). Whiskers are drawn within the 10th to 90th percentile; points below and above the whiskers are not shown. *P* values, two-sided Student's *t*-test. ATAC-seq, assay for transposase-accessible chromatin using sequencing.

TT–seq signals were lowest at 6 h and remained lower than the level recorded before IAA treatment, despite a slight recovery at 24 h. Sequencing analysis of EU-labeled, newly synthesized RNA (EU–seq) also revealed significant decreases in nascent transcripts at 3–9 h after the addition of IAA (Fig. 5g and Extended Data Fig. 9e). Thus, globally attenuated transcription corresponds to early defects in phosphorylation and chromatin binding of Pol II following PSPC1 degradation, demonstrating a direct role for PSPC1 in the regulation

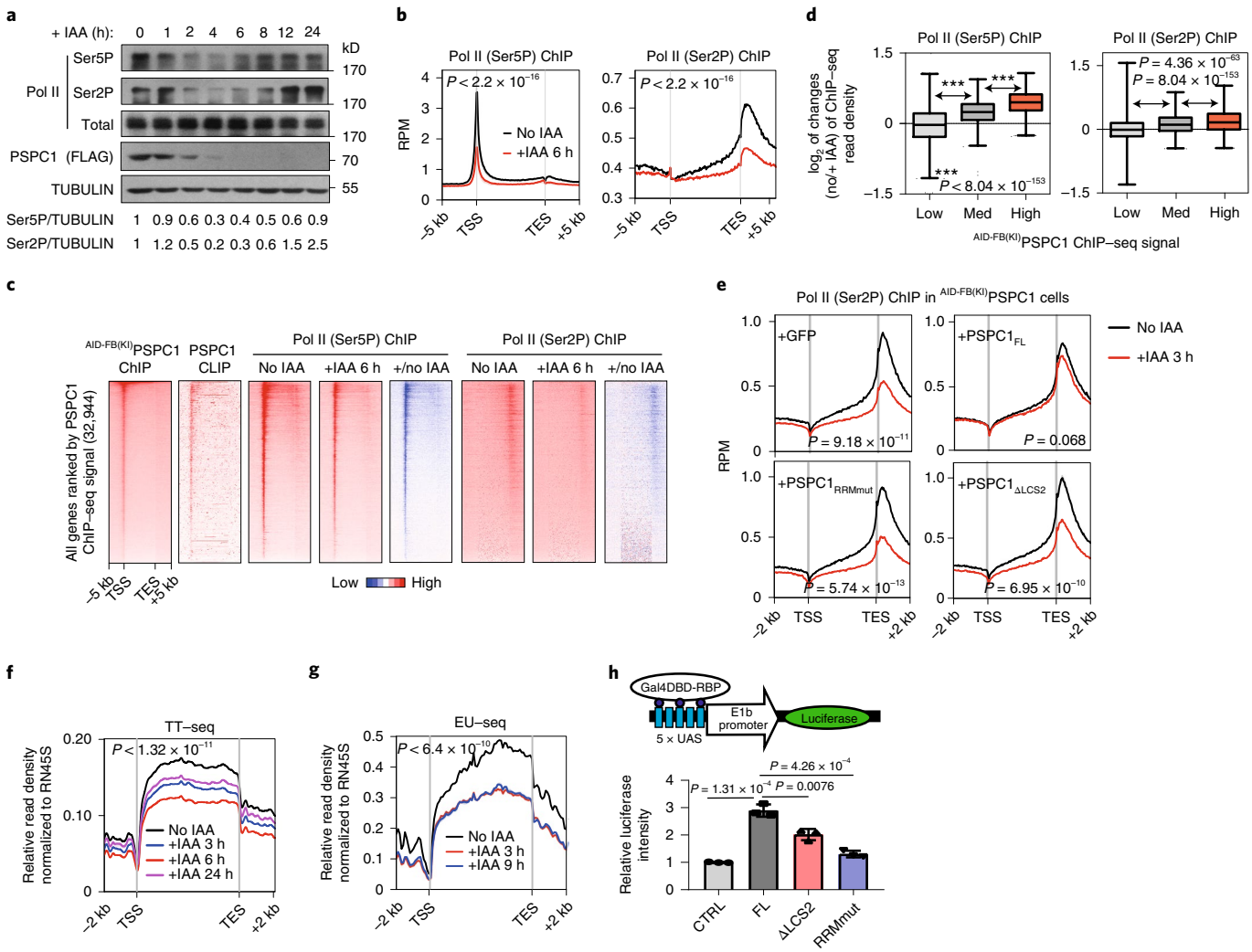


Fig. 5 | PSPC1 promotes Pol II binding and nascent transcription in ESCs. a, Time-course immunoblot analysis of $^{AID\text{-}FB(KI)}$PSPC1 cells treated with IAA. The relative levels of Ser5P and Ser2P normalized to TUBULIN are shown at the bottom. The experiment was repeated twice, with similar results. **b**, Metagene analysis of Ser5P and Ser2P Pol II across all genes following 6-h treatment with IAA. *P* values as indicated. **c**, Heatmap analysis of Pol II ChIP-seq signals following PSPC1 degradation. +/no IAA indicates changes in Pol II binding ratio after IAA treatment. All heatmaps were sorted by PSPC1 ChIP-seq signals. **d**, Correlation analysis of changes in Pol II and PSPC1 ChIP-seq signals following PSPC1 degradation. All genes ($n = 32{,}944$) are classified into three equal groups ($n = 10{,}981$ for each group) according to PSPC1 ChIP-seq signals. The *y* axis shows $\log_2$ of the average change in Pol II ChIP-seq signals. Whiskers are drawn within the 10th to 90th percentile. Points below and above whiskers are not shown. *P* values, two-sided Student's *t*-test, indicated as shown. **e**, Rescue of Pol II Ser2P ChIP-seq signals by various PSPC1 proteins during 3-h IAA treatment. Transient expression of GFP serves as the negative control. **f**,**g**, Time-course metagene analysis of nascent transcripts by TT-seq (**f**) and EU-seq (**g**) following PSPC1 degradation. *P* values are shown in the sequence: no IAA versus +IAA 3h, +IAA 6h, +IAA 24h (**f**) and +IAA 3h, +IAA 9h (**g**): 5.78×10^{-13}, 6.55×10^{-14}, 1.31×10^{-11}, 4.72×10^{-12}, 6.35×10^{-10}. **h**, Promoter tethering and reporter assay. The *y* axis shows relative luciferase intensity normalized to control (Gal4 DBD/Gal4 DNA-binding domain). Data shown as mean ± s.d. of three biological replicates. *P* values by two-sided Student's *t*-test indicated as shown. **b**,**e**,**f**,**g**, Data shown as mean of two biological replicates. *P* values, two-sided Kolmogorov–Smirnov test. **f**,**g**, The *y* axis shows the relative read density of nascent transcripts across all protein-coding genes ($n = 20{,}516$) normalized to the *RN45S* rRNA gene.

of Pol II transcription in vivo. Consistently, tethering $PSPC1_{FL}$, but not $PSPC1_{RRMmut}$, to the promoter of a reporter gene significantly enhanced luciferase expression in 293T cells, whereas $PSPC1_{\Delta LCS2}$ exhibited weaker transactivation activities (Fig. 5h). These in vivo results highlight the biological significance of the RNA-binding and phase-separation mechanisms in nuclear transcription.

Genome-wide colocalization of chrRBPs and Pol II. To explore a general role for chrRBPs in transcription, we checked where they bind in the genome. We performed ChIP–seq in ESCs for RNA chaperone hnRNPU, the nuclear matrix proteins SAFB1 and SAFB2 and the proteins UTP3, UTP6 and CIRH1A—known as components of the small-subunit processome. We also reanalyzed seven published ChIP–seq datasets in ESCs (WDR43, hnRNPK, SRSF2, NONO, DDX21, LIN28A and METTL3; Supplementary Table 4). Similar to what we have observed for PSPC1, all analyzed chrRBPs bind strongly to regulatory DNA elements, including TSSs, enhancers and super-enhancers (Fig. 6a,b and Extended Data Fig. 10a), in line with a previous observation in human HepG2 and K562 cells[38].

This set of 14 chrRBPs co-occupy a total of 15,317 promoters and 231 super-enhancers, of which 77% (11,730) and 92% (212), respectively, are also targeted by RNA Pol II (Fig. 6c and Supplementary

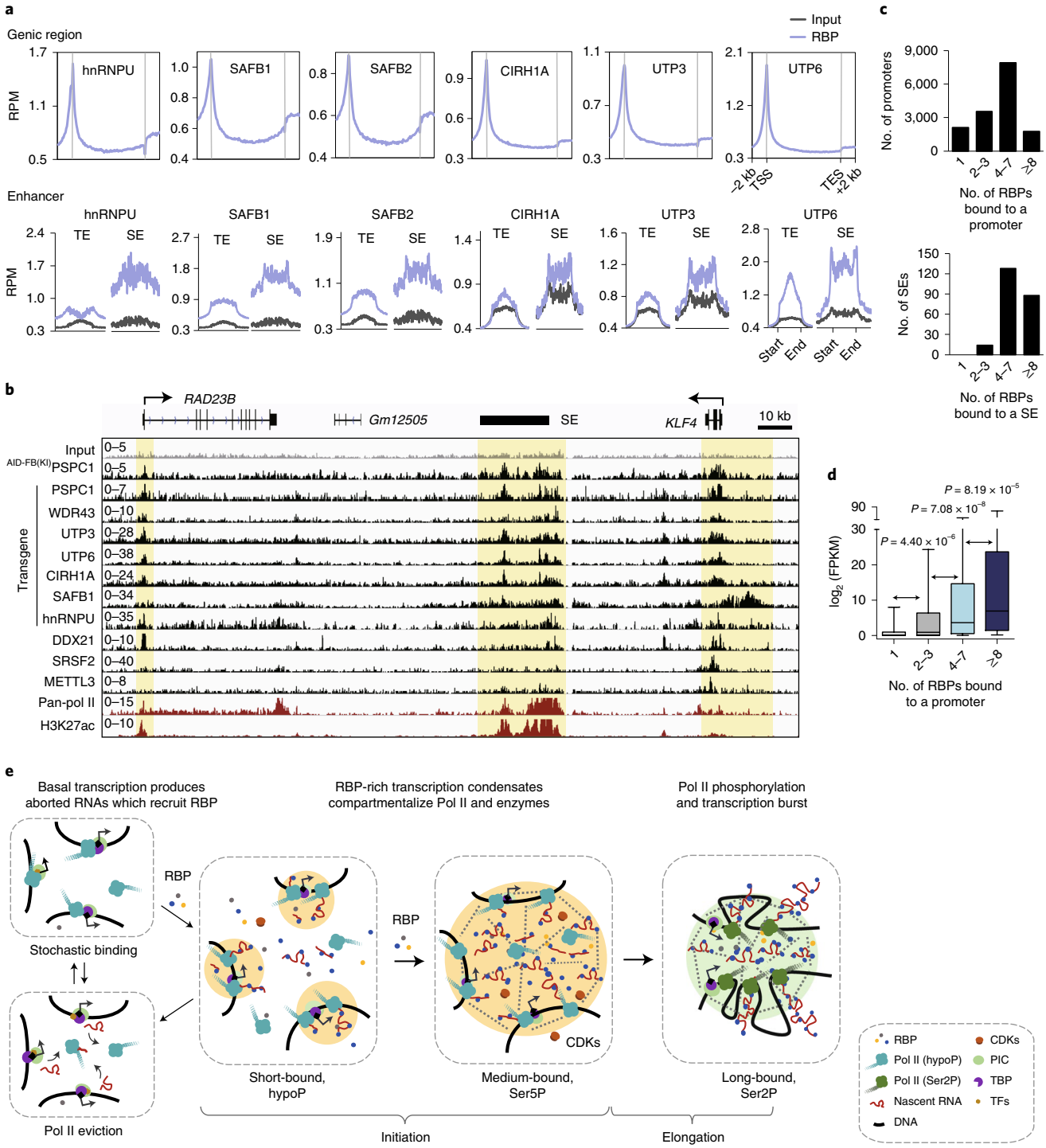


Fig. 6 | Colocalization and the model of RBPs at promoters and enhancers in transcription control. a, Metagene analysis of ChIP–seq signals of various RBPs across all mouse genes (top) and enhancers (bottom). TE, typical enhancers ($n = 8,704$); SE, super-enhancers ($n = 231$). **b**, UCSC genome browser view of ChIP–seq tracks at representative loci. **c**, Extensive colocalization of RBPs at enhancers and promoters. The *y* axis shows the number of promoters (top) and super-enhancers (bottom) bound by different numbers of RBPs. **d**, Positive correlation between the number of cobound RBPs (*x* axis) and gene expression (*y* axis). Whiskers are drawn within the 10th to 90th percentile. *P* values by two-sided Student's *t*-test indicated as shown. **e**, Model showing how RBPs harness RNA-binding and phase-separation activities in the promotion of Pol II transcription. Stochastic binding of Pol II to chromatin initiates a basal level of transcription (top left). Nascent RNAs recruit their binding proteins and phase separate with RBPs at the proximity of the transcription sites (top left). When RBPs are scarce or not readily recruited, RNA carrying negative charges may dissolve the transcription condensates and evict Pol II from the chromatin template (bottom left). Through iterative cycles of Pol II fall-off and rebound, increased local crowding of aborted transcripts and RBPs leads to the formation of large transcription condensates, which efficiently concentrates Pol II and necessary enzymes such as CDKs (middle left and right). Pol II hyperphosphorylation and the production of longer RNA transcripts generate a massive amount of negative charge, which consequently leads to a burst of RBP-rich transcription condensates and the release of Pol II into the gene body for transcription elongation (right). We propose that weak multivalent RNA–protein and protein–protein interactions recruit RBPs to the proximity of transcription sites and also empower RBPs to modulate transcription via the phase-separated mechanism.

Table 5). Remarkably, ~13,191 (86%) promoters are cobound by two or more chrRBPs, ~8,234 by four to seven and ~1,376 by eight or more (Fig. 6c). Over 98% of ESC super-enhancers (~226) are cobound by three or more chrRBPs. The degree of cobinding positively correlates with the level of mRNA expression (Fig. 6d). Unsupervised clustering also revealed a strong positive correlation with Pol II, active histone marks and transcription regulators such as MED1, OCT4 and NANOG, but relatively poor correlation with repressive marks (Extended Data Fig. 10b).

Consistent with the genome-wide colocalization of multiple RBPs, the simultaneous addition of PSPC1, PTBP1 and hnRNPL$_{LCS}$ produced larger droplets and incorporated more CTD than single RBPs (Extended Data Fig. 10c,d). In addition, tethering of PSPC1 alone or together with PTBP1, hnRNPL$_{LCS}$, FUS and WDR43 to a synthetic promoter led to 2.5–5-fold incremental increases in luciferase activity, a functional correspondence to the number of proteins cotethered (Extended Data Fig. 10e). We posit that the prevalent cobinding of RBPs at genome hotspots may provide an opportunity for diverse RBPs to act collaboratively in the promotion of transcription condensate formation, thereby enhancing polymerase incorporation and activity at transcription sites.

Discussion

Here we reveal that hundreds of RBPs are dynamically present on chromatin, with their numbers and abundance surpassing those of classic epigenetic and transcription factors. Surveys of selected RBPs show that they tend to interact and colocalize with Pol II at the genome-wide level, and that their knockdown attenuates, and coexpression enhances, transcription. By focusing on a representative RBP, we delineate the biochemical mechanism by which PSPC1 promotes Pol II transcription in sequential steps. RNA directs PSPC1 to the proximity of transcription initiation sites, where it not only prevents the RNA-induced eviction of unphosphorylated CTD but also synergizes with RNA to promote CTD incorporation and subsequent phosphorylation and release by CDKs. In addition, PSPC1 stabilizes the binding of the Pol II holoenzyme to template during in vitro transcription. Accordingly, auxin-induced degradation of PSPC1 leads to global downregulation of Pol II occupancy and nascent transcription in ESCs. The rescue of defective Pol II binding was not observed in PSPC1 mutants lacking either the major LCS or RRM domain. Thus, complementary lines of evidence corroborate a direct, functional role of PSPC1 in promoting Pol II engagement and transcription through its phase-separation and RNA-binding activities. These two intrinsic properties, which are shared by many chrRBPs, endow PSPC1 with the ability to modulate Pol II binding and activity on chromatin.

Based on these findings we extrapolate that, in cells, RBPs stabilize Pol II engagement to transcription sites via RNA and phase separation (Fig. 6e). Pol II with basal transcription activity produces short RNAs when it binds to chromatin stochastically. These highly negatively charged RNAs evict Pol II before the CTD is properly phosphorylated. On the other hand, nascent RNAs recruit their binding proteins—RBPs—to counteract the abortive effect of RNA. In addition, nascent RNAs serve as multivalent molecules to facilitate the phase separation of RBPs, which confines Pol II to the proximity of transcription sites. Thus, insufficient recruitment of RBPs coincides with the precocious eviction of Pol II by RNA. During numerous rounds of fall-off and rebinding, Pol II continues to produce abortive transcripts and more RBPs are recruited, which eventually leads to the formation of RBP-rich transcription condensates when a threshold of RNA levels is reached. Subsequently, Pol II is rapidly hyperphosphorylated and transcribes into the gene body. In this regard, we propose that the recruitment of RBPs to gene promoters critically contributes to the rate-limiting step of transcription condensate formation. This model suggests an important role for the interplay of chromatin-bound RBPs with abortive and noncoding transcripts around transcription sites in the regulation of dynamic and transient assembly of polymerase clusters in cells[5,17].

RNA-binding proteins are abundantly present in the proximity of chromatin and are capable of polymerizing and binding RNA. These features favor RBPs as the major nuclear components driving the phase separation of transcription condensates under physiological conditions. In addition, cotranscriptional RNA processing deploys multifunctional RBPs to reside in the proximity of transcription sites, which offers a convenient means for their moonlighting in the assemblage of transcription condensates. By sensing levels of nascent transcripts, RBPs may leverage transcription output to balance cellular activities. Some RBPs, like PSPC1, directly contribute to the formation of transcription condensates via their intrinsic capability to polymerize while others, like WDR43, modulate the activity of associated enzymes[39] and yet others merely increase molecular crowding. Nevertheless, these RBPs are actively recruited and play collaborative roles in both forming and running transcription factories[3]. We propose that RBP–RNA interplay represents a key layer of gene regulation, expanding the horizon in our understanding of the intricate regulation of transcription and expression heterogeneity in multicellular organisms.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41589-021-00904-5.

Received: 25 March 2021; Accepted: 22 September 2021;
Published online: 16 December 2021

References

1. Roeder, R. G. & Rutter, W. J. Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature* **224**, 234–237 (1969).
2. Cramer, P. Organization and regulation of gene transcription. *Nature* **573**, 45–54 (2019).
3. Iborra, F. J., Pombo, A., Jackson, D. A. & Cook, P. R. Active RNA polymerases are localized within discrete transcription 'factories' in human nuclei. *J. Cell Sci.* **109**, 1427–1436 (1996).
4. 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. *Mol. Cell* **40**, 965–975 (2010).
5. Cisse, I. I. et al. Real-time dynamics of RNA polymerase II clustering in live human cells. *Science* **341**, 664–667 (2013).
6. Chong, S. et al. Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* **361**, eaar2555 (2018).
7. Cook, P. R. Predicting three-dimensional genome structure from transcriptional activity. *Nat. Genet.* **32**, 347–352 (2002).
8. 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).
9. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848 (2008).
10. Harlen, K. M. & Churchman, L. S. The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat. Rev. Mol. Cell Biol.* **18**, 263–273 (2017).
11. McCracken, S. et al. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**, 357–361 (1997).
12. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
13. Darzacq, X. et al. In vivo dynamics of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **14**, 796–806 (2007).
14. Boettiger, A. N. & Levine, M. Synchronous and stochastic patterns of gene activation in the *Drosophila* embryo. *Science* **325**, 471–473 (2009).
15. Gebhardt, J. C. et al. Single-molecule imaging of transcription factor binding to DNA in live mammalian cells. *Nat. Methods* **10**, 421–426 (2013).
16. Rodriguez, J. et al. Intrinsic dynamics of a human gene reveal the basis of expression heterogeneity. *Cell* **176**, 213–226 (2019).

17. Cho, W. K. et al. RNA Polymerase II cluster dynamics predict mRNA output in living cells. *eLife* **5**, e13617 (2016).
18. Steurer, B. et al. Live-cell analysis of endogenous GFP-RPB1 uncovers rapid turnover of initiating and promoter-paused RNA Polymerase II. *Proc. Natl Acad. Sci. USA* **115**, E4368–E4376 (2018).
19. Seila, A. C. et al. Divergent transcription from active promoters. *Science* **322**, 1849–1851 (2008).
20. Preker, P. et al. RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322**, 1851–1854 (2008).
21. Yin, Y. et al. Opposing roles for the lncRNA haunt and its genomic locus in regulating HOXA gene activation during embryonic stem cell differentiation. *Cell Stem Cell* **16**, 504–516 (2015).
22. Luo, S. et al. Divergent lncRNAs regulate gene expression and lineage differentiation in pluripotent cells. *Cell Stem Cell* **18**, 637–652 (2016).
23. Li, X. et al. GRID-seq reveals the global RNA-chromatin interactome. *Nat. Biotechnol.* **35**, 940–950 (2017).
24. Liu, L. C. et al. Insight into novel RNA-binding activities via large-scale analysis of lncRNA-bound proteome and IDH1-bound transcriptome. *Nucleic Acids Res.* **47**, 2244–2262 (2019).
25. Yin, Y. F. et al. U1 snRNP regulates chromatin retention of noncoding RNAs. Nature 580, 147–150 (2020).
26. Skalska, L., Beltran-Nebot, M., Ule, J. & Jenner, R. G. Regulatory feedback from nascent RNA to chromatin and transcription. *Nat. Rev. Mol. Cell Biol.* **18**, 331–337 (2017).
27. 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).
28. Henninger, J. E. et al. RNA-mediated feedback control of transcriptional condensates. *Cell* **184**, 207–225 (2021).
29. Maharana, S. et al. RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science* **360**, 918–921 (2018).
30. Swinburne, I. A., Meyer, C. A., Liu, X. S., Silver, P. A. & Brodsky, A. S. Genomic localization of RNA binding proteins reveals links between pre-mRNA processing and transcription. *Genome Res.* **16**, 912–921 (2006).
31. Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat. Rev. Genet.* **15**, 829–845 (2014).
32. Schwartz, J. C. et al. FUS binds the CTD of RNA polymerase II and regulates its phosphorylation at Ser2. *Genes Dev.* **26**, 2690–2695 (2012).
33. Ji, X. et al. SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. *Cell* **153**, 855–868 (2013).
34. Calo, E. et al. RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. *Nature* **518**, 249–253 (2015).
35. Zeng, Y. et al. Lin28A binds active promoters and recruits Tet1 to regulate gene expression. *Mol. Cell* **61**, 153–160 (2016).
36. Nozawa, R. S. et al. SAF-A regulates interphase chromosome structure through oligomerization with chromatin-associated RNAs. *Cell* **169**, 1214–1227 (2017).
37. Guallar, D. et al. RNA-dependent chromatin targeting of TET2 for endogenous retrovirus control in pluripotent stem cells. *Nat. Genet.* **50**, 443–451 (2018).
38. Xiao, R. et al. Pervasive chromatin-RNA binding protein interactions enable RNA-based regulation of transcription. *Cell* **178**, 107–121 (2019).
39. Bi, X. et al. RNA targets ribogenesis factor WDR43 to chromatin for transcription and pluripotency control. *Mol. Cell* **75**, 102–116 (2019).
40. Lu, J. Y. et al. Genomic repeats categorize genes with distinct functions for orchestrated regulation. *Cell Rep.* **30**, 3296–3311 (2020).
41. Zhang, H. et al. DEAD-box helicase 18 counteracts PRC2 to safeguard ribosomal DNA in pluripotency regulation. *Cell Rep.* **30**, 81–97 (2020).
42. Graumann, J. et al. Stable isotope labeling by amino acids in cell culture (SILAC) and proteome quantitation of mouse embryonic stem cells to a depth of 5,111 proteins. *Mol. Cell. Proteomics* **7**, 672–683 (2008).
43. Ebmeier, C. C. et al. Human TFIIH kinase CDK7 regulates transcription-associated chromatin modifications. *Cell Rep.* **20**, 1173–1186 (2017).
44. Kato, M. et al. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767 (2012).
45. Lin, Y., Protter, D. S., Rosen, M. K. & Parker, R. Formation and maturation of phase-separated liquid droplets by RNA-binding proteins. *Mol. Cell* **60**, 208–219 (2015).
46. Ghosal, S., Das, S. & Chakrabarti, J. Long noncoding RNAs: new players in the molecular mechanism for maintenance and differentiation of pluripotent stem cells. *Stem. Cells Dev.* **22**, 2240–2253 (2013).
47. Vos, S. M. et al. Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. *Nature* **560**, 607–612 (2018).
48. Goldman, S. R., Ebright, R. H. & Nickels, B. E. Direct detection of abortive RNA transcripts in vivo. *Science* **324**, 927–928 (2009).
49. Pai, D. A. et al. RNAs nonspecifically inhibit RNA polymerase II by preventing binding to the DNA template. *RNA* **20**, 644–655 (2014).
50. Ghetti, A., Pinol-Roma, S., Michael, W. M., Morandi, C. & Dreyfuss, G. hnRNP I, the polypyrimidine tract-binding protein: distinct nuclear localization and association with hnRNAs. *Nucleic Acids Res.* **20**, 3671–3678 (1992).

Methods

Experimental model and subject details. Mouse ESCs (CJ9, 46C lines and cells expressing endogenous or exogenous 3× FLAG- and biotin-tagged RBPs) were cultured in complete ESC medium as previously described[39]. HEK 293T cells were cultured in medium containing DMEM, 10% FCS and 1× penicillin/streptomycin solution.

Construction of ESC lines and treatments. Usage of the AID system was based on a previous report[51]. We note that cells with integrated TIR1 (ref. [51]) became adapted to IAA after repeated passage. For all experiments involving AID knock-in ESCs, we used early-passage ESCs and freshly infected them with TIR1 virus immediately before the addition of IAA (1 mM; Sigma, no. I5148). For rescue experiments, $^{FB(KI)}$*PSPC1* ESCs following fresh infection with TIR1 were transiently transfected with PSPC1 proteins, treated with IAA 48 h after transfection and collected for ChIP–seq analysis. ESCs stably expressing FLAG/biotin-tagged RBPs were constructed as previously described[39].

Cells were treated with actinomycin D (ActD, 1 μg ml^{-1} to inhibit both Pol I and Pol II, 10 ng ml^{-1} to inhibit only Pol I, abcam, no. ab141058), triptolide (TPL, 1 μM, abcam, no. ab120720) or DMSO for 2 h before chromatin fractionation. RNase A treatment was performed as previously described[52].

Lentivirus-mediated RNA interference (pLKO) and EU incorporation assay were performed as previously described[39]. ESCs were infected and selected by puromycin for 36 h at 24 h post infection, then incubated for 20 min with 5-EU (Jena Bioscience, no. CLK-N002, final concentration 1 mM). EU-incorporated ESCs were labeled with Alexa 647 using a Click-iT Cell Reaction Buffer Kit (Life Technologies, no. C10269). The intensity of each sample was first compared to the nonlabeled blank control; sh*RBP* samples were then normalized to the control sample transfected with sh*Ctrl*.

Chromatin fractionation. We adapted a crosslinking-based chromatin fractionation method reported previously[53]. Briefly, ESC nuclei[39] were crosslinked by 1% formaldehyde for 10 min then resuspended with 2× pellet volumes of nuclear lysis buffer (50 mM Tris-HCl pH 8.1, 10 mM EDTA, 1% SDS) and incubated on ice for 10 min. DNA–protein complexes were precipitated by the addition of 0.5 volume of ethanol at −20 °C for 1 h and spun down at 5,000*g* at 4 °C for 20 min. The pellet was washed with 75% ethanol and resuspended in 50 mM Tris-HCl buffer pH 7.4. Urea (final, 8 M) and SDS (final, 2%) were added to the suspension and the mixture was incubated at 37 °C for 30 min. An equal volume of 5 M NaCl was added followed by a further 30-min incubation. The complexes were precipitated by the addition of 0.1 volume of 3 M sodium acetate and three volumes of ethanol, centrifuged at 5,000*g* at 4 °C for 5 min and washed twice with 75% ethanol. The pellet was resuspended in DNase digestion buffer (20 mM HEPES pH 7.5, 15 mM NaCl, 6 mM $MgCl_2$, 1 mM $CaCl_2$, 10% glycerol) containing DNase I (10 U, Takara) and incubated at 37 °C for 1 h. EDTA was added and the pellet was spun down at 13,000 r.p.m. for 20 min at 4 °C. The supernatant was used for SDS–polyacrylamide gel electrophoresis (PAGE). Proteins >20 kD (to exclude histones) were collected for MS sequencing.

For salt extraction of native chromatin, non-crosslinked nuclei[39] were incubated with four volumes of extraction buffer (20 mM HEPES pH 7.5, 10 mM KCl, 1.5 mM $MgCl_2$, 1 mM EDTA, 0.1 mM Na_3VO_4, 25% glycerol, 1 mM phenylmethanesulfonyl fluoride, 1/200 Proteinase Inhibitor cocktail) with different concentrations of NaCl at 4 °C for 30 min, and spun down at 4 °C. The same percentage (5%) of supernatant and pellet was used for immunoblots.

Quantitative MS. Protein identification was performed by the MaxQuant platform, and protein abundance was evaluated by the intensity of the sum of all peptide intensities compared to the number of observable peptides of a protein (iBaq intensity)[54]. We took the proteins identified by two replicates with iBaq intensity >500 and molecular weight >20 kD as chromatin proteins. To measure the relative abundance of different proteins and compare different batches of experiments, we defined an iBaq ratio by normalizing the iBaq intensity of each protein to the sum of all protein iBaq intensities (Supplementary Table 1).

For the label-free quantification (LBQ) method, we performed MS analysis under four experimental conditions (DMSO versus ActD, mock versus RNase) with one replicate for each, and quantified the relative abundance of each protein by iBaq ratio as described above. For the tandem mass tag (TMT) method, five experimental conditions were analyzed (DMSO versus ActD or TPL, mock versus RNase) with one replicate for each. We performed the experiment as previously published[55]. After chromatin fractionation, we used the same amount of chromatin protein for different conditions, labeled with different amine-reactive TMT 6-plex reagents (ThermoFisher). We then mixed these samples and carried out MS analysis.

For stable isotope labeling with amino acids (SILAC), cells cultured in heavy SILAC medium were first treated with transcription inhibitor or RNase and mixed with an equal number of mock cells cultured in light medium. We exchanged the medium for different treatments for an additional biological replicate, to exclude medium bias. Mixed cells were used for chromatin fractionation and MS analysis.

To compare cross-experiments with different MS methods, we first calculated fold-change (FC) scores by normalization of the experimental (exp) sample to the corresponding mock treatment, as below. LBQ: $\log_8$(exp/mock + 0.001); TMT: $\log_3$(exp/mock + 0.001); SILAC: $\log_2$(exp/mock + 0.001). For each treatment, we then calculated the mean of normalized FC scores by three methods and used a cutoff <−0.2 to select chrRBPs dynamically regulated by transcription/RNA.

The majority (>91%) of the defined set of 512 chrRBPs were identified by the LBQ method. For TMT and SILAC, only 50–90% of chrRBP hits were detected with quantitative information among nine samples. It is possible that not all 512 chrRBPs were detected by TMT and SILAC because these two MS methods involve isotope labeling. In addition, we noted that transcription inhibition or RNase treatment resulted in decreased abundance of certain proteins. We reasoned that the combined effects of labeling efficiency and decreased protein abundance might have contributed to the limited protein detection by TMT and SILAC. Nevertheless, all 512 chrRBPs were identified using at least one method.

Analysis of biochemical features of RBPs. We used the following websites for the analysis of various biochemical features. LCS: http://repeat.biol.ucy.ac.cy/fgb2/gbrowse/swissprot/; IDR: http://www.pondr.com/, https://iupred2a.elte.hu/ and https://github.com/zhanzhan90/distribution-of-amino-acid.git (https://doi.org/10.5281/zenodo.3874019); isoelectric point: https://web.expasy.org/compute_pi/.

co-IP, affinity purification, immunofluorescence, fluorescence recovery after photobleaching and antibodies used. Co-IP and affinity purification were performed as previously described[39]. Benzonase was added during lysis for all co-IP analysis. Although Benzonase degrades DNA and RNA, its digestion of RNA was incomplete under our experimental conditions. In addition to Benzonase, we added RNase A (Takara) to remove RNA during co-IP washes.

Immunofluorescence was performed as previously described[39]. Dilution was based on the manufacturer's instructions: PSPC1 (1:100), pan-Pol II (1:100), TBP (1:100), FLAG (1:500). Images were taken with a Nikon A1R-HD-Multiphoton microscope. Fluorescence recovery after photobleaching (FRAP) analysis was performed as previously described[56] in ESCs transfected with mCherry-fused PSPC1 or GFP-fused H3. A Nikon A1R-HD-Multiphoton microscope was used for photobleaching.

Antibodies used included WDR43 (Abclonal, no. Q659), hnRNPU (abcam, no. ab180952), hnRNPL (Santa Cruz, no. sc-32317), PTBP1 (Abclonal, no. A6107), DDX5 (abcam, no. ab126730), FUS (abcam, no. ab70381), PSPC1 (abcam, no. ab104238), pan-Pol II (abcam, no. ab52202), EZH2 (CST, no. 5246S), SUZ12 (CST, no. 3737S), CTCF (abcam, no. ab128873), CHD1 (CST, no. 4351S), TOP2A (abcam, no. ab52934), H3 (Easybio, no. BE3015), DDX21 (Novus Biologicals, no. NBP1-83310), FUBP1 (abcam, no. ab181111), FUBP3 (abcam, no. ab181025), LIN28A (abcam, no. ab155542), NCL (abcam, no. ab134164), TUBULIN (CWBIO, no. CW0098), SNRNP70 (Santa Cruz, no. sc-390988), hypophosphorylated Pol II (8WG16; Covance, no. MMS-126R), Pol II Ser5P (CST, no. 13523), Pol II Ser2P (CST, no. 13499), TBP (Santa Cruz, no. sc-421), FLAG (Sigma-Aldrich, no. F3165), GFP (Santa Cruz, no. sc-9996) and mCherry (CST, no. 43590).

Protein purification. Protein purification in bacteria (His-tagged PSPC1, mCherry-PSPC1, mCherry-PSPC1$_{\Delta LCS2}$, mCherry-PSPC1$_{RRMmut}$, GFP-CTD and SNAP-TBP) was performed as previously described[39]. For the purification of CDK7, ~200 ml of 293F cells was transfected by 200 μg of pMlink-StrepII-FLAG-CDK7 plasmids and 600 μg of Polyethyleneimine. After 48 h, cells were lysed in 20 ml of lysis buffer (50 mM Tris pH 7.4, 150 mM NaCl, 0.5% Triton X-100 and 10% glycerol) at 4 °C for 30 min. Supernatant was incubated with 500 μl of Strep Tactin beads (pre-equilibrated) for 1 h. The beads were washed five times with high-salt wash buffer (50 mM Tris pH 7.4, 350 mM NaCl, 0.5% Triton X-100 and 5% glycerol) and eluted by 2.5 mM desthiobiotin in 50 mM Tris pH 7.4, 150 mM NaCl and 10% glycerol. Purified proteins were concentrated in Millipore concentration tubes (<10 kDa) in 20 mM Tris pH 8.0 and 150–500 mM NaCl.

Baculoviruses expressing 6× His-tagged human CDK9 and CyclinT1 were a gift from G. Li's laboratory. Purification was performed as previously described[39]. A Ni^{2+} beads column was used for affinity purification.

Estimation of nuclear protein concentrations. The nuclear concentration of a protein was calculated by dividing the protein molecule number by the volume of the nucleus (~220 fl)[29]. Molecule number and nuclear concentration of each protein are listed in Supplementary Table 1.

Droplet formation assay. In droplet formation assays, protein samples were added to a buffer containing 20 mM Tris pH 7.5 and 150 mM NaCl supplemented with or without 10% dextran as indicated, and either incubated on ice for 30 min or analyzed immediately for imaging or sedimentation. For imaging analysis, we mixed fluorescence-labeled proteins with the corresponding unlabeled proteins in the following ratios. mCherry:PSPC1, 1:3; Cy5.5:NHS ester-labeled PSPC1, 1:10; Alexa647:SNAP-TBP, 1:10. RNase inhibitor was included for assays involving total RNA from ESCs. All images were taken with a Nikon confocal microscope concurrently and analyzed by ImageJ. Total CTD fluorescence intensity was obtained by calculating the sum of CTD fluorescence intensity in droplets for each field of view. We captured between three and ten images in different fields of view for each condition, then used these for statistical analysis.

In droplet sedimentation experiments, samples were centrifuged for 10 min at 14,000 r.p.m., 4 °C. The same fraction of supernatant and pellet was used for immunoblot analysis. Anti-GFP antibody (Santa Cruz, no. sc-9996) was used for detection of GFP-fused CTD.

Kinase assays. In kinase assays[32], we preincubated 0.2 μg of GFP-CTD with 1–5 μg of mCherry, BSA, mCherry-PSPC1, mCherry-PSPC1$_{\Delta LCS2}$ or mCherry-PSPC1$_{RRMmut}$ for 20 min at room temperature in buffer (20 mM Tris-HCl pH 7.0, 150 mM KCl, 2 mM $MgCl_2$, 2 mM DTT, 0.15 mg ml^{-1} BSA). We added RNA (50 ng μl^{-1}) as indicated in the presence of RNase inhibitor, followed by 0.2 μg of CDK9 or CDK7 and ATP (final, 0.1 mM) for 10 min at room temperature. The reaction was stopped by the addition of SDS loading buffer. Pol II Ser5P antibody (CST, no. 13523) was used for detection of phos-CTD, and mCherry antibody (CST, no. 43590) for detection of mCherry-fused PSPC1.

CTD release assay. TATA-box binding protein (5 μM) and CTD (0.6 μM) were first mixed together with PSPC1 (5 μM) or mCherry (5 μM) in a buffer containing 20 mM Tris pH 7.5, 150 mM NaCl and 10% dextran. For imaging experiments we added ATP (0.1 mM) and CDK9 (2 μg) and immediately placed the plate under the microscope for recording. For every sample, at least two views were recorded. Because the addition of RNA greatly accelerated the release of CTD, we used a lower amount of CDK9 (1 μg). Time-lapse analysis of droplet intensity was performed using Nikon NIS-element AR software. We first normalized CTD fluorescence intensity to TBP in individual droplets and then normalized it to the initial level to obtain the normalized CTD intensity curve, i(t), for each droplet. We did nonlinear fitting of the intensity curve using GraphPad Prism and assumed that the droplet was a homogenous sphere. For each droplet we drew an intensity curve and then derived the CTD release rate based on two equations: $v(t) = \Delta I(t)/(4\pi R^2)$ and $I(t) = 4\pi R^3/3 \times i(t)$, where v, t, I, i, ΔI and R, respectively, represent release rate, time, total fluorescence intensity, mean fluorescence intensity, change in total fluorescence intensity and droplet radius. Thus we derive the rate curve, $v(t) = R/3 \times \Delta i(t)$.

For droplet sedimentation, we fractionated droplets at each indicated time point. The corresponding fractions of the supernatant and pellet were used for immunoblot analysis.

In vitro transcription and EMSA assays. Template DNA was produced by annealing the forward primer (AGGCAGGCCTTAGCTCCGTTCGCCGTGTCC TACCTATCCTCTCCTCACCACTCCCGGGGCCATTC) with or without a 5′ Cy5.5 label with the 5′ phosphorylated reverse primer (5′(phos)TGGCCCCG GGAGTGGTGAGGAGAGGATAGGTAATCAGTTACGCCCGGAGCTAAGG CCTGCCTAGT), and then ligated to the 601-R fragments (generated by BglI and Dra III double digestion of plasmid pJW013[57]). Yeast Pol II complexes were purified by tandem-affinity purification (Rpb9-TAP) as described for yeast strain YBL360 (ref. [58]).

In vitro transcription reaction was carried out in 20 μl of transcription buffer (25 mM HEPES pH 7.5, 50 mM KCl, 10% glycerol, 5 mM $MgCl_2$ (Sigma), 1 mM DTT and 0.05 mg ml^{-1} BSA). Each reaction contained 0.4 μl of Low-C NTP mix (25 mM ATP, 25 mM UTP, 25 mM GTP and 0.25 mM CTP), 0.5 μl of α-^{32}P-CTP (3,000 Ci mmol^{-1}, 10 mCi ml^{-1}; Perkin Elmer), 1 ng of the Bubble-601-R (BR-DNA) template and 100 ng of competitor DNA, along with or without additional proteins to be tested. Reactions were incubated at 30 °C for 45 min and then stopped by the addition of 120 μl of STOP buffer (0.3 M NaAC, 5 mM EDTA, 0.1% SDS, 40 μg ml^{-1} linear acrylamide (Ambion, no. AM9520) and 1 μl of Proteinase K (20 mg ml^{-1})). Mixtures were kept at 55 °C for 15 min and then supplemented with a ^{32}P-CTP-labeled spike-in RNA control. Spike-in RNA was transcribed using MEGAscript T7 (Ambion, no. AM1330) in the presence of α-^{32}P-CTP. The reaction mix with spike-in RNA was subjected to ethanol precipitation for RNA extraction. RNA pellets were resuspended in 12 μl of 90% formamide-Tris-borate EDTA (TBE) loading buffer and denatured before loading onto an 8% polyacrylamide gel (19:1) containing 7 M urea. Dried gel was exposed to phosphoimagimg and scanned with a GE Typhoon Scanner.

For EMSA assays, 50 ng of Cy5.5-labeled BR template and 25 mM NTP mix were used in a transcription condition similar to that described above. We used 2.3 μl of 250 ng μl^{-1} heparin to reduce nonspecific, loosely bound Pol II or mock for each reaction, with incubation at 4 °C for 20 min. Samples were loaded onto a 3.5% native polyacrylamide gel (37.5:1) in 0.3× TBE. Electrophoresis was carried out at 4 °C for 3.5 h, and gels were scanned with a Li-Cor CLX scanner.

Quantitative ChIP–PCR, ChIP–seq and CLIP–seq analysis. ChIP assays for Pol II Ser2P (CST, no. 13499) and Pol II Ser5P (CST, no. 13523) were performed as previously described[39]. For ChIP analysis of $^{AID\text{-}FB(KI)}$PSPC1, exogenous FLAG-HA-tagged PSPC1$_{FL}$, PSPC1$_{RRMmut}$, PSPC1$_{\Delta LCS2}$, FLAG-biotin-tagged UTP3, UTP6 and CIRH1A, cells were crosslinked by 3% formaldehyde for 10 min. For FLAG-biotin-tagged hnRNPU, SAFB and SAFB2, cells were crosslinked by 2 mM dithiobis succinimidyl propionate for 30 min followed by a 10-min 1% formaldehyde crosslinking. Crosslinked cells were partially fragmented by 12 U ml^{-1} DNase I at 37 °C for 10 min then sonicated at 25% amplitude for 30 s. Following anti-FLAG IP, samples were subjected to a second purification step using 30 μl of M-280 Streptavidin Dynabeads (Invitrogen, no. 11205D). The remaining steps were performed as previously described[39]. The ChIP–seq library was constructed using either a NEBNext ChIP–seq Library Prep Reagent Set or Tn5, and sequenced on an Illumina Hiseq 2500 or X10 platform. ChIP–seq peaks were called using the MACS program ($P < 1 \times 10^{-3}$). ChIP–seq peaks located within 5 kb around transcription start sites were defined as promoter peaks. For clustering analysis of RBP ChIP–seq, read counts in a region encompassing 5 kb around the peak center of all sites bound by RBPs were calculated and used for Pearson correlation analysis between ChIP–seqs. Unsupervised clustering was analyzed with R software.

We performed PSPC1 CLIP–seq with an improved tandem-affinity capture and an optimized library construction procedure as previously described[39,59]. Whole-cell lysates of ESCs carrying a stably integrated transgene of PSPC1 (tagged with FLAG and biotin moieties) were used for CLIP–seq.

Nascent RNA analysis by TT–seq and EU–seq. Transient transcriptome sequencing of nascent transcripts (TT–seq) was performed as described previously, with modifications[25]. ESCs were labeled for 10 min with 500 μM 4sU (Sigma-Aldrich), then 4sU-labeled RNAs were biotinylated, fragmented and subjected to two rounds of biotin-affinity purification by M-280 dynabeads (Invitrogen). The beads were washed twice at 45 °C with 0.5 ml of 100 mM Tris pH 7.4, 10 mM EDTA, 1 M NaCl and 0.1% Tween-20, followed by two washes with SDS washing buffer (0.5 ml of 50 mM Tri-Cl pH 8.1, 10 mM EDTA, 1% SDS) at room temperature (5–10 min for each wash). RNA was eluted by 50 μl of SDS washing buffer at 95 °C for 5 min. The eluate was subjected to a second immunoprecipitation followed by washes. Libraries were constructed using NEBNext Ultra II directional RNA library prep kits (NEB).

EU–seq was performed as described previously, with modifications[60]. ESCs were labeled with 1 mM EU for 10 min. EU-labeled RNA was biotinylated, fragmented and subjected to two rounds of biotin-affinity purification, similar to TT–seq. We used the ssDNA-seq Lib Prep Kit (Abconal, no. RK20222) to construct the library. RNA-seq analysis was performed as previously described[25]. For data processing, adapters and low-quality reads were removed using Trim_galore and the Bedtools package was used to remove polyC tracts. To compare different samples we calculated read density by normalization to the reads mapped to rRNA gene *RN45S*.

Luciferase assay. Various RBPs were fused with Gal4 DNA-binding domain (Gal4DBD) in pcDNA3.1 vector, as previously described[39]. A 5× upstream activation sequence (UAS) was inserted upstream of the E1b minimal promoter in the vector psiCHECK-2. For each assay, 100 ng of psiCHECK-2 and 200 ng of each pcDNA3.1-GAL4-RBP vector were cotransfected into HEK293T cells (one 24-well). To exclude dosage effects, we cotransfected pcDNA3.1-GAL4-empty vectors so that every well was treated with a total of 1 μg of GAL4 expression construct. We performed luciferase assays at 36 h post transfection (Dual-Luciferase Reporter, Promega). Renilla luciferase activity was normalized to firefly luciferase.

Published datasets used in this study. We used the following published datasets in our analysis. GSM2988821 WDR43 ChIP–seq; GSM2988831 Pol II 8WG16 ChIP–seq; GSM2988824 Pol II Ser2P ChIP–seq; GSM2988827 Pol II Ser5P ChIP–seq; GSM1941467 pan-Pol II ChIP–seq; GSM3713432 hnRNPK ChIP–seq; GSM3407052 SRSF2 ChIP–seq; GSM1893472 NONO ChIP–seq; GSM1693793 DDX21 ChIP–seq; GSM1915715 LIN28A ChIP–seq; GSM2424700 METTL3 ChIP–seq; GSM560347 MED1 ChIP–seq; GSM1082340 OCT4 ChIP–seq; GSM288356 c-MYC ChIP–seq; GSM1082341 SOX2 ChIP–seq; GSM611197 SIN3A ChIP–seq; GSM1023124 TET2 ChIP–seq; GSM918750 P300 ChIP–seq; GSM480162 SUZ12 ChIP–seq; GSM480161 EZH2 ChIP–seq; GSM769008 H3K4me3 ChIP–seq; GSM1000089 H3K27me3 ChIP–seq; GSM1000099 H3K27ac ChIP–seq; GSM769009 H3K4me1 ChIP–seq; and GSM1000109 H3K36me3 ChIP–seq.

Quantification and statistical analysis. Statistical analyses were carried out using Excel 2019, R v.3.3.0 or R v.4.0.2. The statistical tests used are stated in the relevant figure legends.

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

Data availability

The main data supporting the findings of this study are available within the article and in Supplementary information. Sequencing data have been deposited in the GEO database under the accession number GSE150399. Source data are provided with this paper.

References

51. Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T. & Kanemaki, M. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat. Methods* **6**, 917–922 (2009).

52. Beltran, M. et al. The interaction of PRC2 with RNA or chromatin is mutually antagonistic. *Genome Res.* **26**, 896–907 (2016).
53. Qin, H. & Wang, Y. Exploring DNA-binding proteins with in vivo chemical cross-linking and mass spectrometry. *J. Proteome Res.* **8**, 1983–1991 (2009).
54. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
55. Christoforou, A. et al. A draft map of the mouse pluripotent stem cell spatial proteome. *Nat. Commun.* **7**, 8992 (2016).
56. Meshorer, E. et al. Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev. Cell* **10**, 105–116 (2006).
57. Lee, C. H., Wu, J. & Li, B. Chromatin remodelers fine-tune H3K36me-directed deacetylation of neighbor nucleosomes by Rpd3S. *Mol. Cell* **52**, 255–263 (2013).
58. Li, B. et al. Combined action of PHD and chromo domains directs the Rpd3S HDAC to transcribed chromatin. *Science* **316**, 1050–1054 (2007).
59. Bi, X., Zhang, X. & Shen, X. Transcriptome-wide profiling of protein-RNA interactions by cross-linking and immunoprecipitation mediated by FLAG-biotin tandem purification. *J. Vis. Exp.* https://doi.org/10.3791/60730 (2020).
60. Eisen, T. J., Eichhorn, S. W., Subtelny, A. O. & Bartel, D. P. MicroRNAs cause accelerated decay of short-tailed target mRNAs. *Mol. Cell* **77**, 775–785 (2020).

Acknowledgements

We thank the members of the Shen Laboratory for insightful discussions. This work was supported in part by the National Basic Research Program of China (nos. 2018YFA0107604 and 2017YFA0504204 to X.S.); the National Key R&D Program of China (no. 2018YFC1004500 to B.L.); the National Natural Science Foundation of China (nos. 31925015 and 31829003 to X.S. and 32030019 and 31872817 to B.L.); and Beijing Advanced Innovation Center for Structural Biology at Tsinghua University (to X.S.).

Author contributions

X.S. supervised the study. X.S. and W.S. conceived of and designed the experiments. W.S. performed the majority of experiments, with help from X.B., B.G., Z.L. and W.R., and conducted bioinformatics analysis with assistance from Y.X., X.B. and J.Y.L. X.B. performed ChIP–seq of $^{FB(EXO)}$PSPC1, UTP3, UTP6 and CIRH1A and CLIP–seq of PSPC1. The in vitro transcription system was designed and set up by B.L. and J.W. W.S. and Y.P. performed in vitro transcription assays with the help of M.P. Y.Y. performed ChIP–seq of SAFB1, SAFB2 and hnRNPU. W.Z., X.J. and H.D. provided technical assistance/suggestions for MS analysis. Z.W., K.W., G.Z., T.L. and J.W. contributed with assistance/suggestions for experiments. X.H. performed ESC total proteome analysis. X.S. and W.S. wrote the manuscript with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

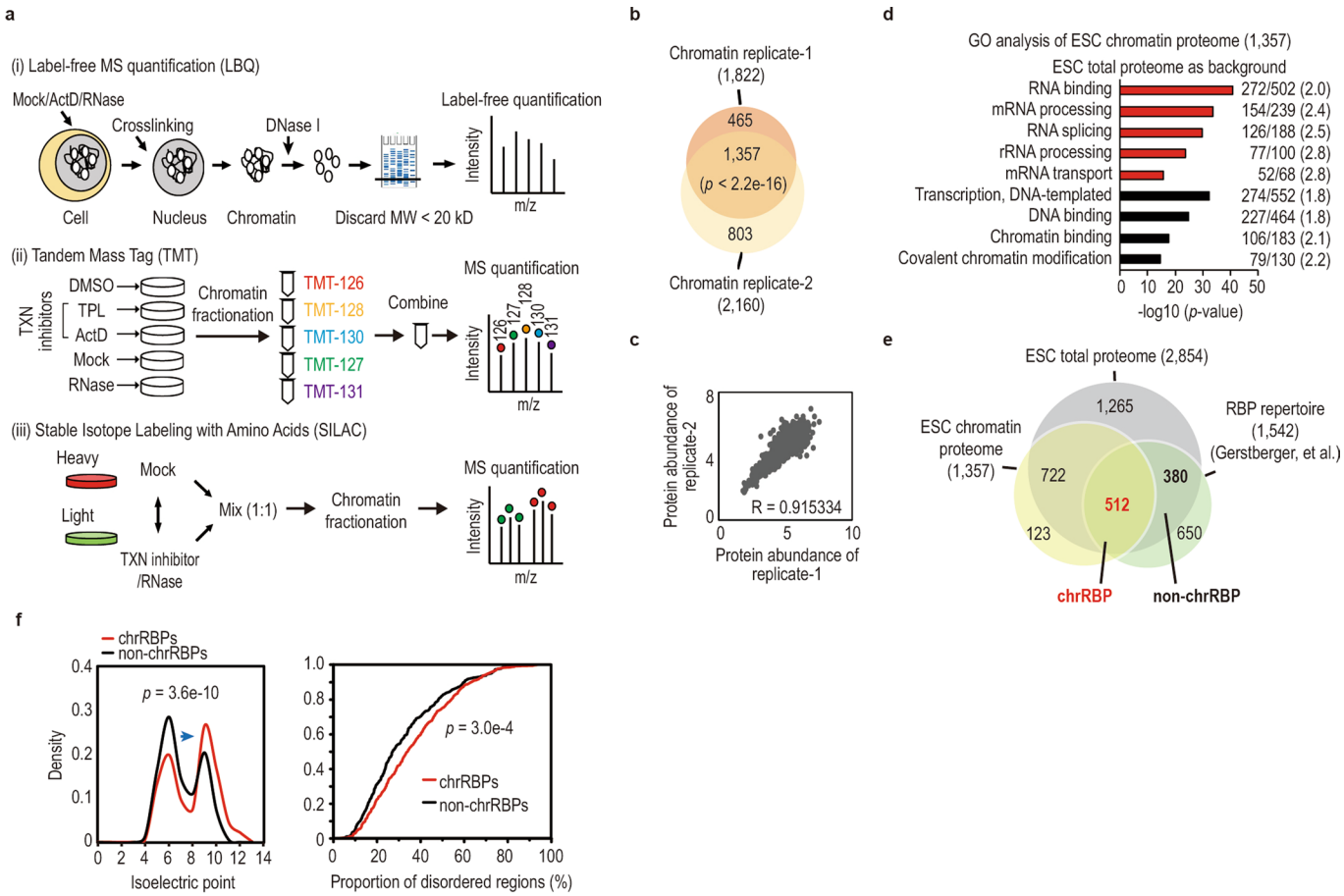
Extended data is available for this paper at https://doi.org/10.1038/s41589-021-00904-5.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41589-021-00904-5.

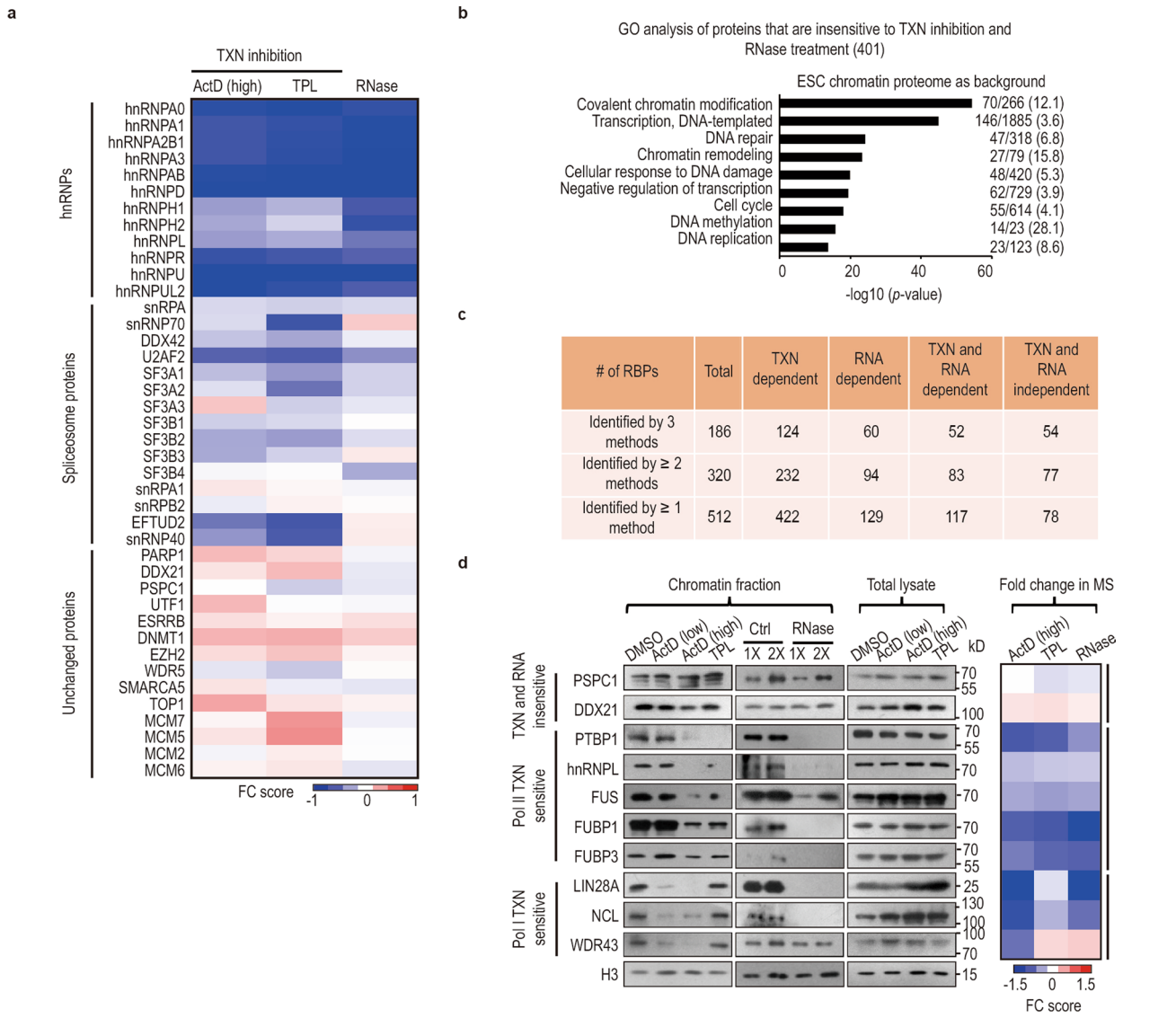
Correspondence and requests for materials should be addressed to Bing Li or Xiaohua Shen.

Peer review information *Nature Chemical Biology* thanks Huasong Lu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

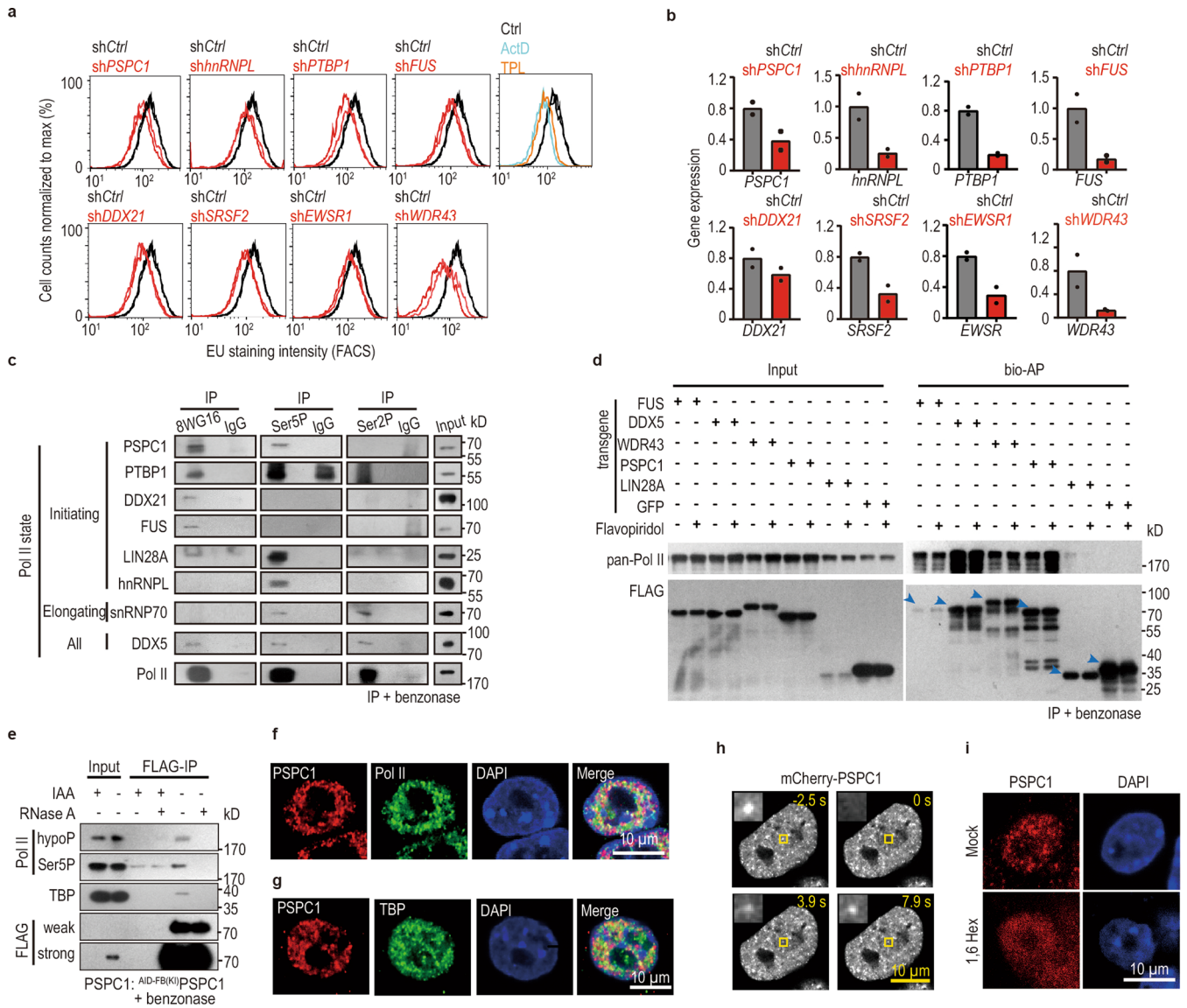
Reprints and permissions information is available at www.nature.com/reprints.



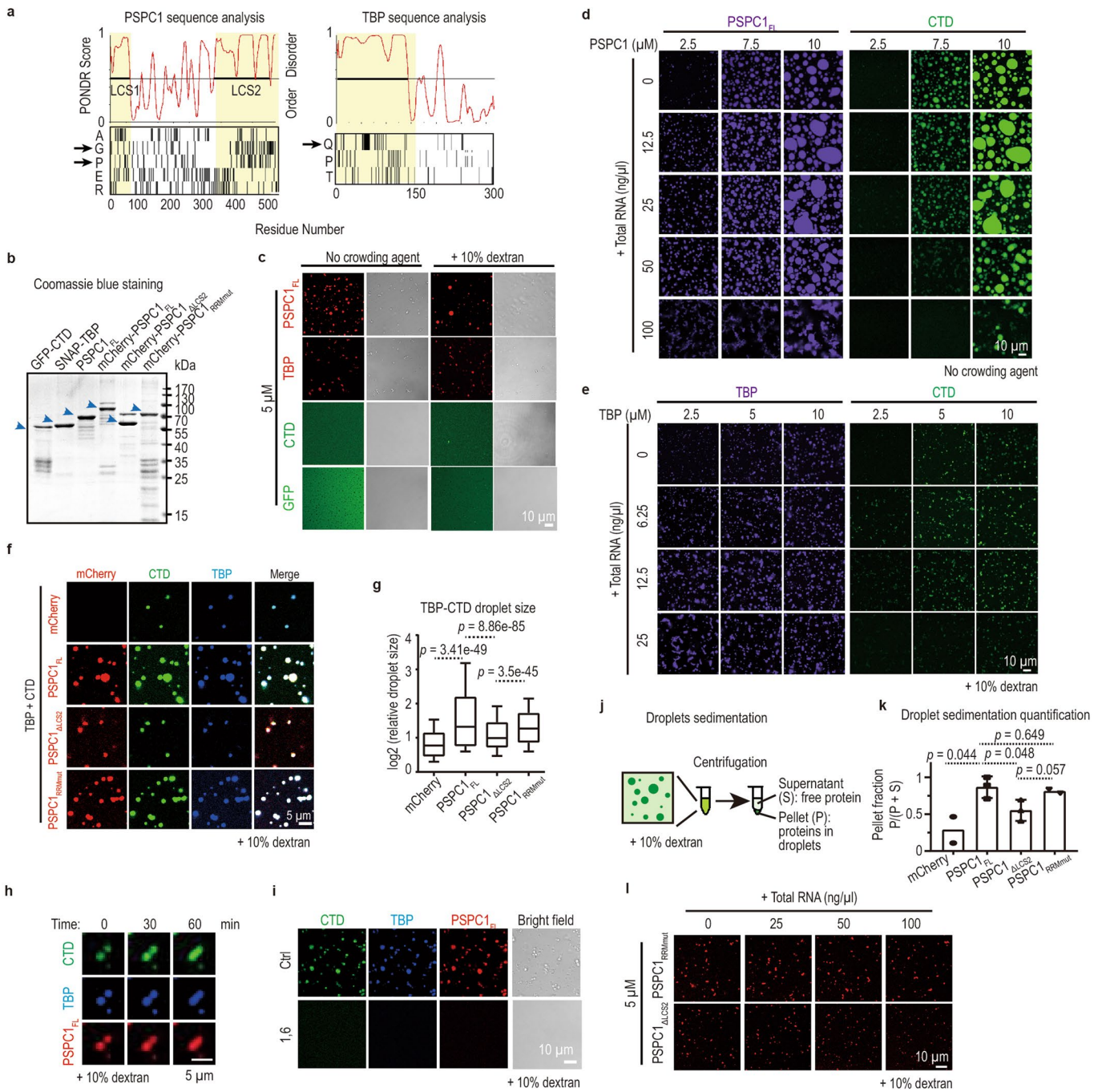
Extended Data Fig. 1 | Abundant associations of RBPs with chromatin in ESCs. a, Schemes showing quantitative analysis of chromatin proteomes under various treatments by three mass spec (MS) methods. (i) Label-free MS quantification (LBQ); (ii) Tandem mass tag (TMT); (iii) Stable isotope labeling with amino acids (SILAC). Transcription (TXN) inhibition: ActD (actinomycin D, 1 μg/ml) or TPL (triptolide, 1 μM). RNase: RNase A (1 mg/ml). DMSO/Mock: mock treatment for transcription inhibition or RNase treatment. **b**, Overlap between two biological replicates of the chromatin proteome. $P < 2.2e\text{-}16$, two-sided Fisher's exact test. **c**, Correlation analysis of two biological replicates of the chromatin proteome. The x-axis and y-axis represent the abundance of each protein identified in the two replicates, indicated by -log10 (iBaq ratio) (Methods; Supplementary Table 1). **d**, Gene ontology (GO) analysis of chromatin proteins (n = 1,357). The total proteome from ESCs (n = 2,854) was used as background. Selected GO terms ($p < 1.0e\text{-}10$) are shown on the y-axis. The x-axis shows enrichment significance by -log10 (*p*-value). Red bars represent terms related to RNA processes; black bars indicate terms associated with transcription and chromatin functions. For each GO term, the number of functionally associated genes identified from analysis of the chromatin proteome and the total number of functionally associated genes expressed in ESCs are indicated sequentially. The numbers in the brackets indicate the fold enrichment. *P*-values, two-sided Fisher-exact tests performed by DAVID. **e**, Comparison of the chromatin proteome (n = 1,357) with the RBP repertoire (n = 1,542) and the ESC total proteome (n = 2,854). The numbers of chrRBPs (red, 512) and non-chrRBPs (black, 380) are indicated in bold. See also Supplementary Table 2. **f**, Biochemical characterization of chrRBPs and non-chrRBPs. (i) Density distribution curve of the isoelectric points of chrRBPs and non-chrRBPs. The blue arrow indicates a shift in the distribution of isoelectric point. (ii) Cumulative distribution curve showing the content of intrinsically disordered regions (IDR) in chrRBPs or non-chrRBPs. *P*-values, two-sided Kolmogorov-Smirnov test. See also Supplementary Table 2.



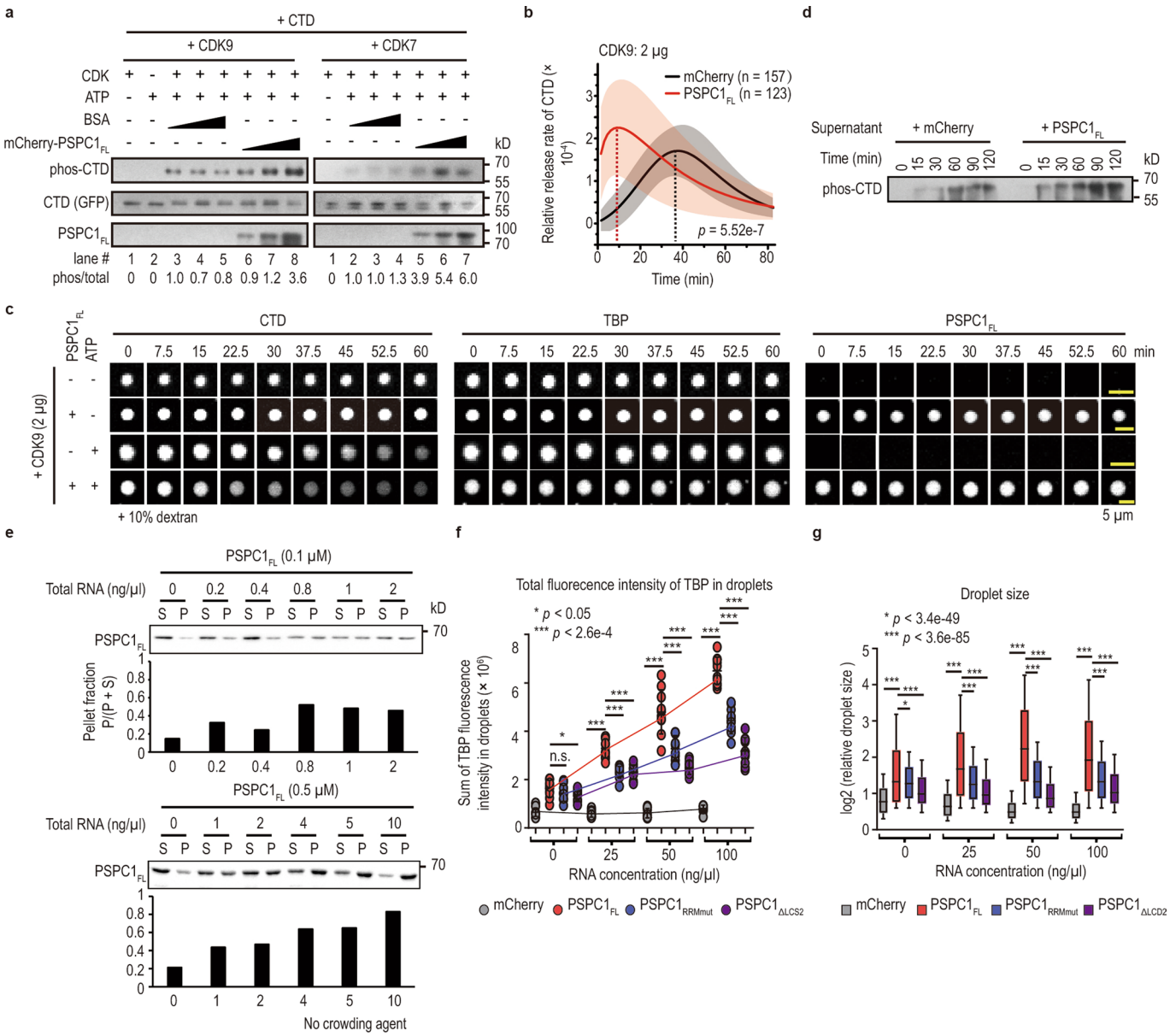
Extended Data Fig. 2 | Dynamic chromatin associations of RBPs mediated by transcription and RNA. a, Heatmap showing the average fold-change (FC) score of chromatin abundance for representative proteins including hnRNPs, spliceosome proteins and unchanged proteins. Numerous hnRNPs and splicing factors are dependent on both RNA and transcription for their chromatin binding. In comparison, the chromatin-binding activities of transcription factors (such as UTF1 and ESRRB) and epigenetic enzymes (such as DNMT1, EZH2, WDR5, topoisomerases, and DNA helicases) were less likely to be affected. The ratio calculation is described in Methods. Data are shown as the mean of 4 biological replicates for ActD and RNase, and 3 replicates for TPL. See also Supplementary Table 3. **b**, GO analysis of chromatin proteins that are insensitive to transcription inhibition and RNase treatment (n = 401). The ESC chromatin proteome was used as the background. The x-axis shows enrichment significance by -log10 (*p*-value). The top enriched terms are shown on the y-axis. For each GO term, the number of functionally associated genes identified from the input list and the background list are indicated sequentially. The numbers in the brackets indicate the enrichment fold. *P*-values, two-sided Fisher-exact tests performed by DAVID. **c**, Summary of the effects of transcription (TXN) and RNA on chromatin-RBP associations. See also Supplementary Table 3. **d**, Chromatin fraction and western-blot analysis of selected chrRBPs upon treatments with transcription inhibitors or RNase A. The corresponding FC score in the mass spec data is shown in heatmap (right). ActD: low (10 ng/ml) or high (1 μg/ml). RBPs are classified into 3 groups based on their sensitivity to inhibition of Pol I or Pol II transcription (TXN) or RNA. Because nascent transcripts are loaded and protected by a battery of RBPs once they emerge from Pol II, we cannot rule out incomplete degradation of RNA by treatment with RNase A. Thus, despite an overall decrease of chrRBP associations with chromatin, the role of RNA in recruiting and mediating chrRBPs to chromatin could still be underestimated based on the observed effects of RNase treatment. Experiment was repeated twice with similar results.



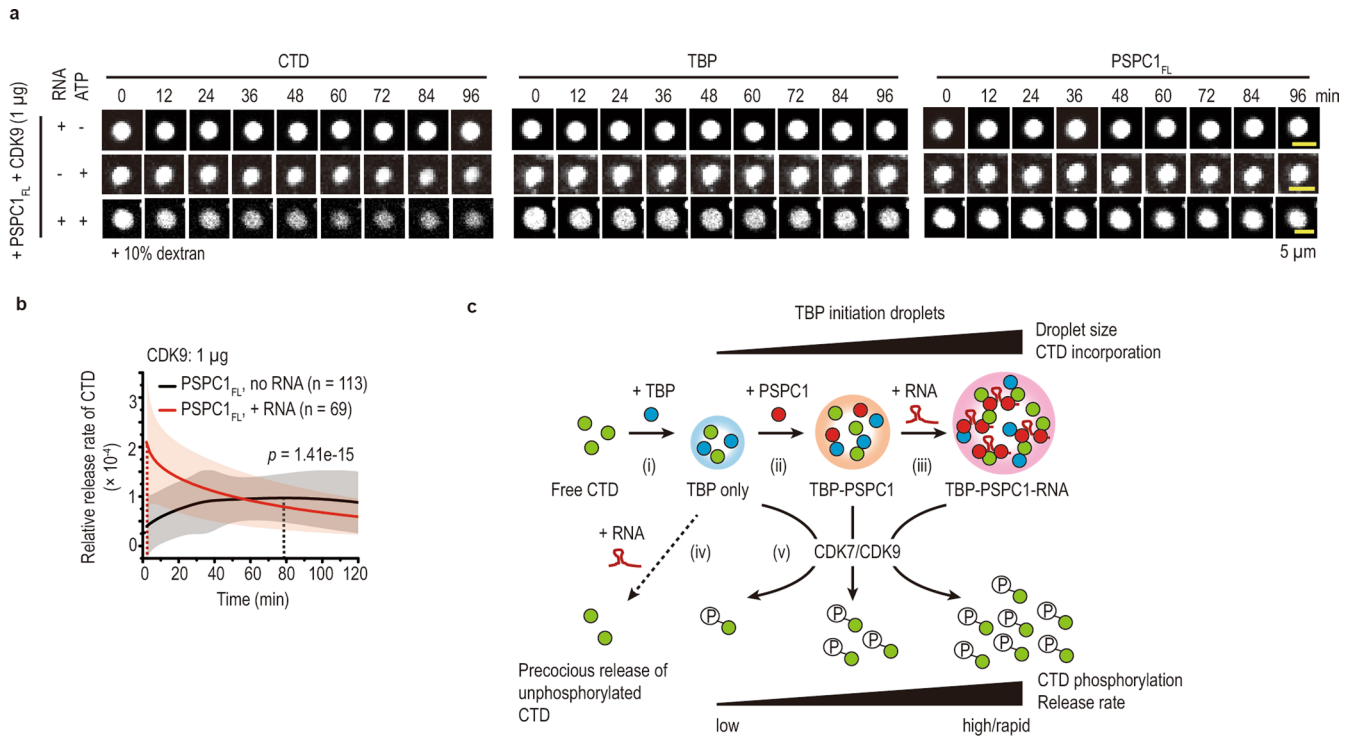
Extended Data Fig. 3 | chrRBPs tend to interact with Pol II and modulate transcription. a, Fluorescence activated cell sorting of EU incorporation after RBP knockdown for 60 hours. The x-axis shows EU intensity. The y-axis shows cell numbers normalized to max (%). Two biological replicates are shown for controls (grey) and RBP knockdown (red). **b**, RT-qPCR analysis of the relative expression of RBPs upon knockdown for 60 hours. Data are shown as mean of 2 biological replicates with different shRNAs. **c**, Co-immunoprecipitation (co-IP) and western blot analysis of endogenous proteins. Pol II in different phosphorylation states captured all 8 tested chrRBPs. Capture of various phosphorylation states of Pol II was confirmed by western blots using the corresponding antibody in each IP (bottom). Benzonase was added during cell lysis and co-IP. 8WG16 and Ser5P, hypo-phosphorylated (hypoP) and serine 5-phosphorylated Pol II that represent initiating Pol II. Ser2P, serine 2-phosphorylated Pol II that represents elongating Pol II. **d**, Reciprocal co-IP of various chrRBPs captured Pol II. Biotin-mediated affinity purification (bio-AP) was performed (+ benzonase) in ESCs that stably express individual FLAG-biotin tagged chrRBPs. ESCs expressing FLAG-biotin tagged GFP were used as the negative control. Of note, flavopiridol can reduce elongating transcripts by inhibiting Pol II pause-release, but results in higher promoter-associated transcripts[39]. Blue arrows indicate exogenously expressed FLAG-tagged proteins. **e**, FLAG-mediated co-immunoprecipitation (FLAG-IP) of $^{AID\text{-}FB(KI)}$PSPC1 ESCs and western-blot analysis. ESCs treated with IAA for 24 hours were used as a negative control. Benzonase was present during cell lysis and FLAG-IP procedures. RNase A was added during washes. **f**, Co-immunostaining analysis of PSPC1 with pan-Pol II. The experiment was performed in $^{AID\text{-}FB(KI)}$*PSPC1* cells and anti-FLAG antibody was used to image endogenously tagged PSPC1 proteins. **g**, Co-immunostaining analysis of PSPC1 with TBP. This experiment was performed in wild-type ESCs and PSPC1 antibody was used to detect endogenous proteins. **h**, FRAP analysis showing the fast recovery of mCherry-PSPC1 puncta. Related to Fig. 1i. **i**, Immunofluorescence analysis of PSPC1 upon treatment with 10% 1,6-hexanediol for 2 minutes. Mock: PBS. Anti-PSPC1 antibody was used for imaging. For panels **e-i**, experiments were repeated twice with similar results.



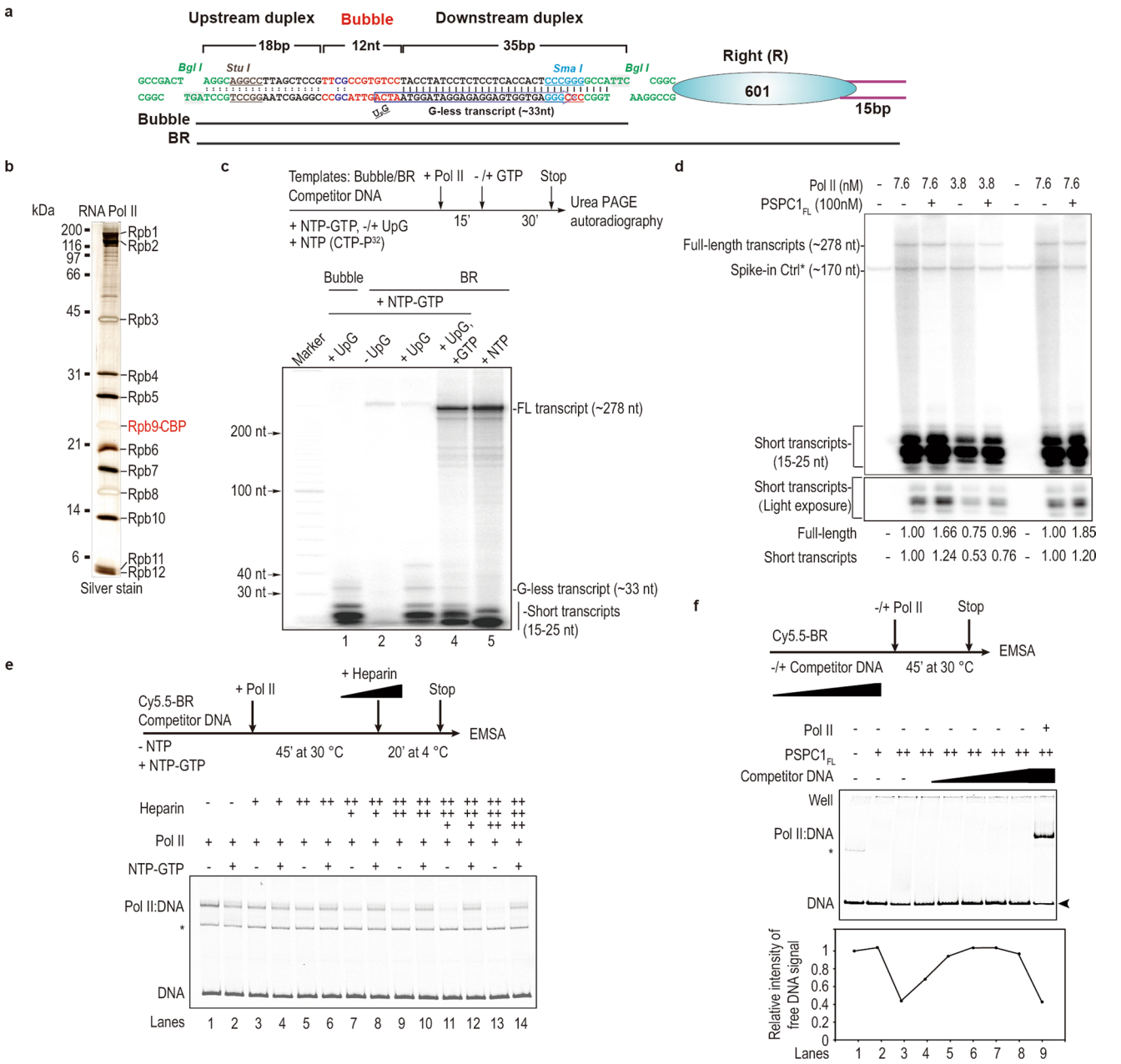
Extended Data Fig. 4 | PSPC1 promotes the incorporation of unphosphorylated CTD into TBP condensates. a, Analysis of IDR and LCS in PSPC1 and TBP. The PONDR score indicates the probability of a region being disordered. The distributions of representative amino acids are shown at the bottom. The regions in yellow indicate the disordered regions which contain the hydrophobic G/P-rich or Q-rich sequences respectively (arrow). **b**, Coomassie blue staining of purified proteins. The blue arrows indicate the main band of the corresponding protein. **c**, *In vitro* droplet formation of recombinant proteins. **d-e**, Phase diagram of PSPC1-CTD (**d**) or TBP-CTD (**e**) droplets in the presence of different concentrations of total RNA isolated from ESCs. We used 0.6 μM of CTD in all assays. **f-g**, Droplet formation assays of TBP (5 μM) and CTD (0.6 μM) with full-length (FL) or mutant PSPC1 (5 μM) or mCherry (5 μM) in the presence of 10% dextran. Representative pictures and quantification are shown in panel **f** and **g** respectively. In panel **g**, the y-axis is log2 (relative droplet size). The whiskers are drawn within 10-90 percentile. Points below and above the whiskers are not shown. The median droplet size is 0.71 μm^2 for mCherry (n = 2,641), 1.50 μm^2 for $PSPC1_{FL}$ (n = 2,932), 0.99 μm^2 for $PSPC1_{\Delta LCS}$ (n = 2,097), and 1.42 μm^2 for $PSPC1_{RRMmut}$ (n = 4,758). *P*-values, two-sided Student's t-test, are shown as indicated. **h**, Fusion of TBP-CTD-PSPC1 droplets. **i**, Phase-separated droplets composed of TBP, CTD and PSPC1 with or without 10% 1,6-hexanediol. Ctrl: mock treatment. **j-k**, Droplets sedimentation and western-blot assays. The schematic diagram and quantification are shown in panel **j** and **k**, respectively. Representative western-blot result is shown in Fig. 2b. The pellet fraction ratio P/(S + P) was shown as mean ± s.d. *P*-values, two-sided student's t-tests, are shown as indicated. N = 2 independent experiments for 'mCherry' and '$PSPC1_{RRMmut}$', N = 3 for '$PSPC1_{FL}$' and '$PSPC1_{\Delta LCS2}$'. **l**, *In vitro* droplet formation of recombinant $PSPC1_{RRMmut}$ and $PSPC1_{\Delta LCS2}$ proteins. For panels **b-e, g-h**, experiments were repeated twice with similar results. Dextran was used as indicated.



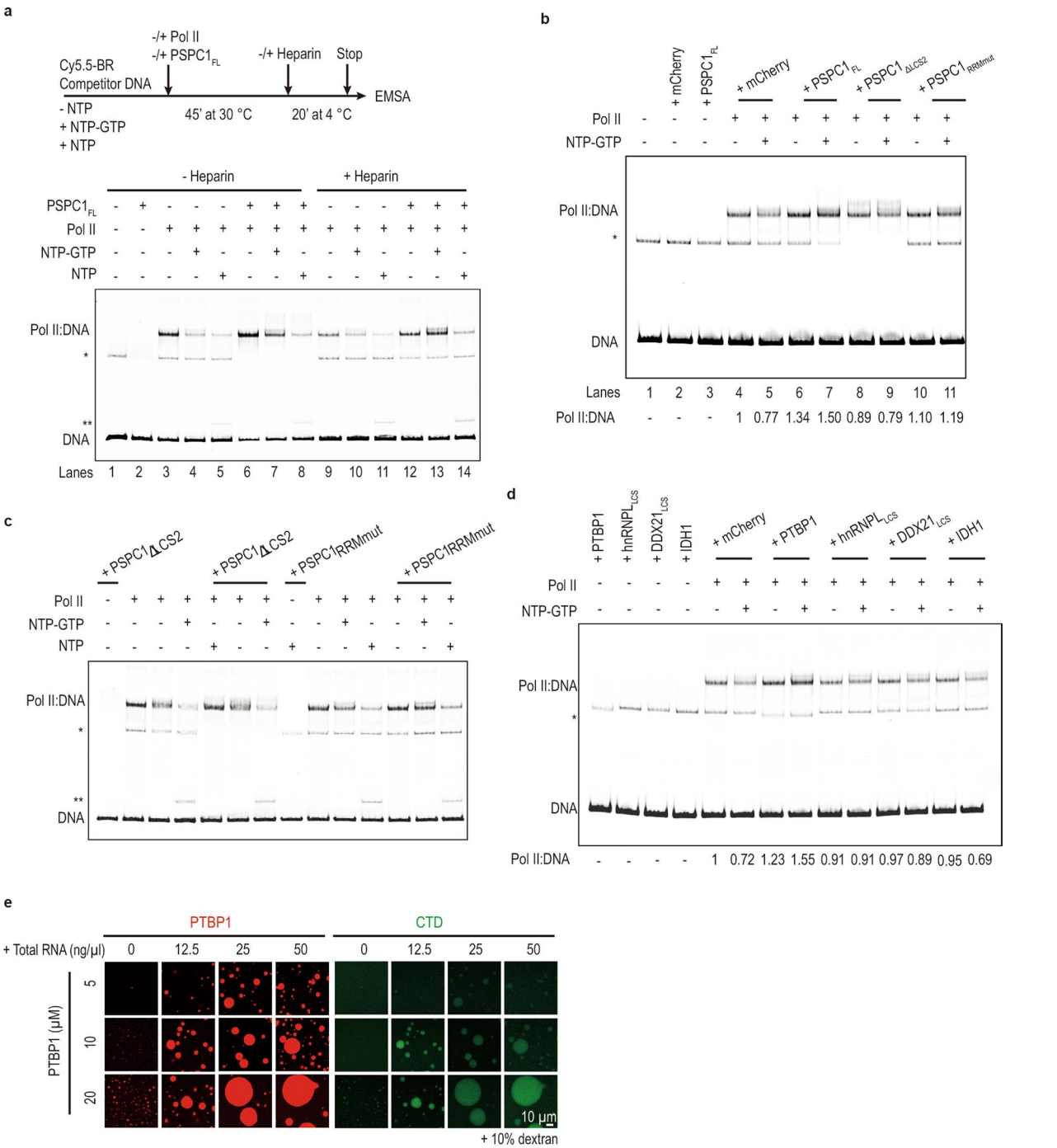
Extended Data Fig. 5 | PSPC1 promotes CTD phosphorylation and release. a, Kinase assays with CDK9 (left) and CDK7 (right). Quantification of the ratio of phosphorylated CTD versus total CTD is shown at the bottom. Additional replicate was shown in Fig. 2c. **b-c**, Time-lapse imaging analysis of CTD release. Panel **b** shows the relative release rate of CTD (y-axis) (Methods). Data are shown as mean ± s.d of droplets. *P*-value, two-tailed Student's t-test for the comparison of max rate between the two groups. Panel **c** shows representative images. Related to Fig. 2d. **d**, Time-lapse sedimentation analysis of released CTD. The samples are the same with Fig. 2e. **e**, RNA promotes PSPC1 phase separation by droplets sedimentation. Quantification is shown at the bottom of each panel. **f-g**, Quantifications of total fluorescence intensity of TBP (**f**) and droplet size (**g**) for data shown in Fig. 2f. *P*-values, two-sided Student's t-test (n.s., not significant). **f**, Data are shown as mean ± s.d. of 10 fields. *P*-values are shown in the sequence of 'no RNA', '25', '50', and '100'. 'mCherry' vs 'PSPC1$_{FL}$': 2.28e-7, 7.75e-14, 3.43e-11, and 5.89e-17. 'PSPC1$_{FL}$' vs 'PSPC1$_{RRMmut}$': 0.2122, 2.43e-6, 2.59e-4 and 1.66e-7. 'PSPC1$_{FL}$' vs 'PSPC1$_{\Delta LCS2}$': 0.0379, 5.77e-6, 4.89e-7 and 1.26e-10. **g**, Droplet sizes are presented as log2 (relative droplet size). The whiskers are drawn within 10-90 percentile. Points beyond the whiskers are not shown. The median droplet sizes and *p*-values are shown in the sequence of 'no RNA', '25', '50', and '100'. mCherry: 0.71 (n = 2,641), 0.56 (n = 2,555), 0.40 (n = 1,698), and 0.40 (n = 1,961). PSPC1$_{FL}$: 1.50 (n = 2,932), 2.19 (n = 2,732), 3.69 (n = 1,838), and 2.79 (n = 3,663). PSPC1$_{RRMmut}$: 1.42 (n = 4,758), 1.37 (n = 5,029), 1.50 (n = 7,073), and 1.50 (n = 8,567). PSPC1$_{\Delta LCS2}$: 0.99 (n = 2,097), 0.94 (n = 1,313), 0.82 (n = 1,436), and 1.03 (n = 1,429). *P*-values: 'mCherry' vs 'PSPC1$_{FL}$': 5.60e-241, < 5.60e-241, 3.00e-236 and 1.65e-143. 'PSPC1$_{FL}$' vs 'PSPC1$_{RRMmut}$': 3.41e-49, 6.88e-137, 1.20e-231, and 3.56e-128. 'PSPC1$_{FL}$' vs 'PSPC1$_{\Delta LCS2}$': 8.86e-85, 3.0e-101, 1.18e-170 and 5.48e-70. Dextran (10%) was used in panels **b-d** and **f-g**.



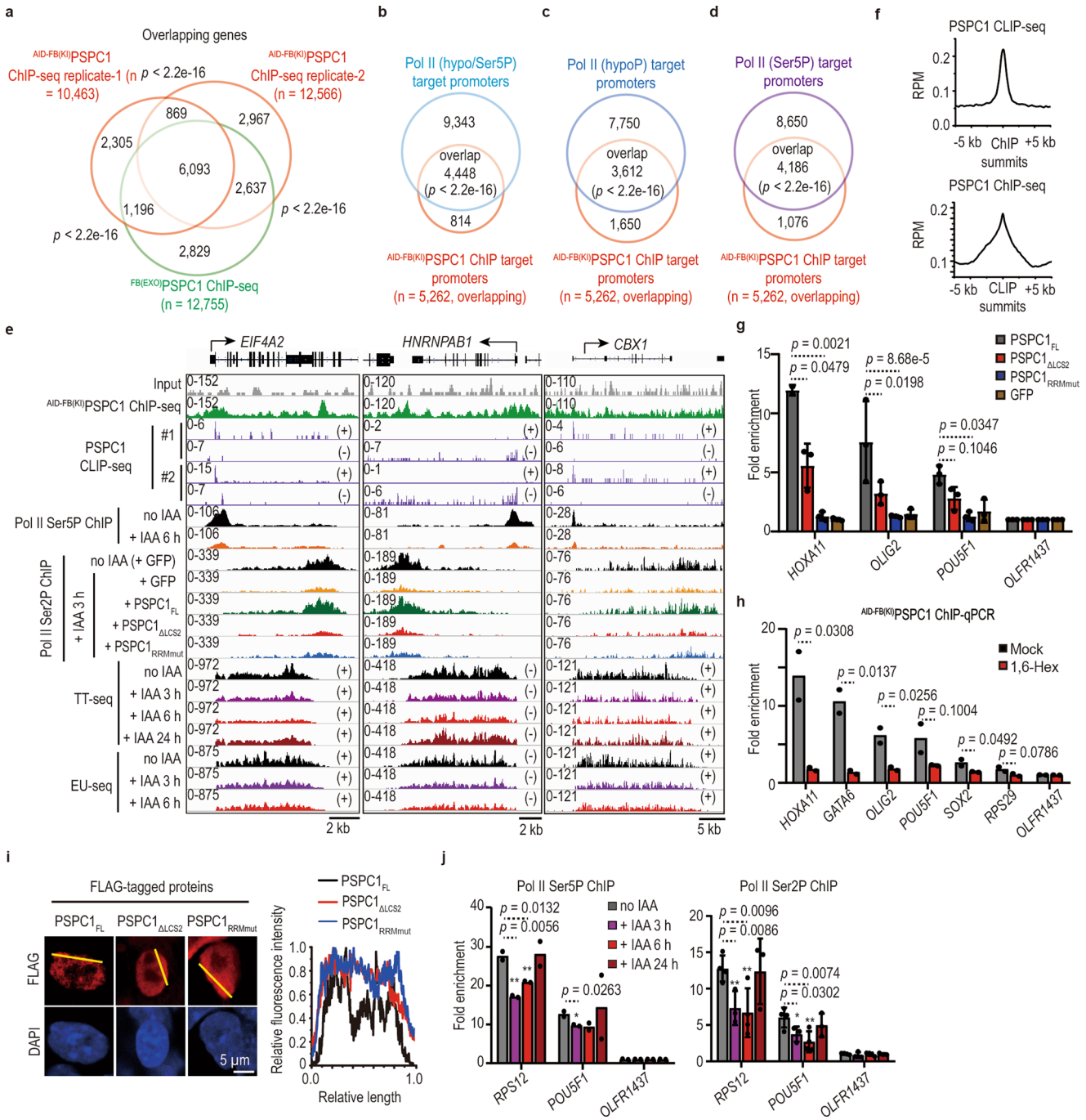
Extended Data Fig. 6 | RNA synergizes with PSPC1 in promoting CTD incorporation, phosphorylation and release. a-b, Effects of RNA on CTD release by time-lapse imaging analysis. Panel **a** shows images taken of representative droplets. Related to Fig. 2i. This assay included 10% dextran. Panel **b** shows the quantification of CTD release rate (Methods). Data are shown as mean ± s.d. of 113 droplets for the '$PSPC1_{FL}$, no RNA' group and 69 droplets for the '$PSPC1_{FL}$, + RNA' group. *P*-value, two-tailed Student's t-test for the comparison of max rate between the two groups. **c**, Schematic diagram showing the interplay of PSPC1 and RNA in promoting CTD incorporation and subsequent phosphorylation and release. TBP alone has weak ability to phase separate and trap CTD within its droplets in the presence of dextran (i). Addition of PSPC1 enhances this phase separation and produces larger droplets that concentrate more CTD inside (ii). RNA further synergizes with PSPC1 to drastically promote phase separation and CTD incorporation (iii). By contrast, in the absence of PSPC1, RNA evicts CTD from TBP droplets (iv). Upon activation by CTD kinases (v), efficient compartmentalization and concentration of CTD inside TBP-PSPC1-RNA droplets lead to stronger phosphorylation and faster release of CTD compared to TBP and TBP-PSPC1 droplets. Note that RNA synergizes with PSPC1 in a manner that critically depends on the phase-separation and RNA-binding activities of PSPC1. Thus, *in vitro* assays with defined components allow us to biochemically dissect the more complex processes of Pol II engagement and release in cells.



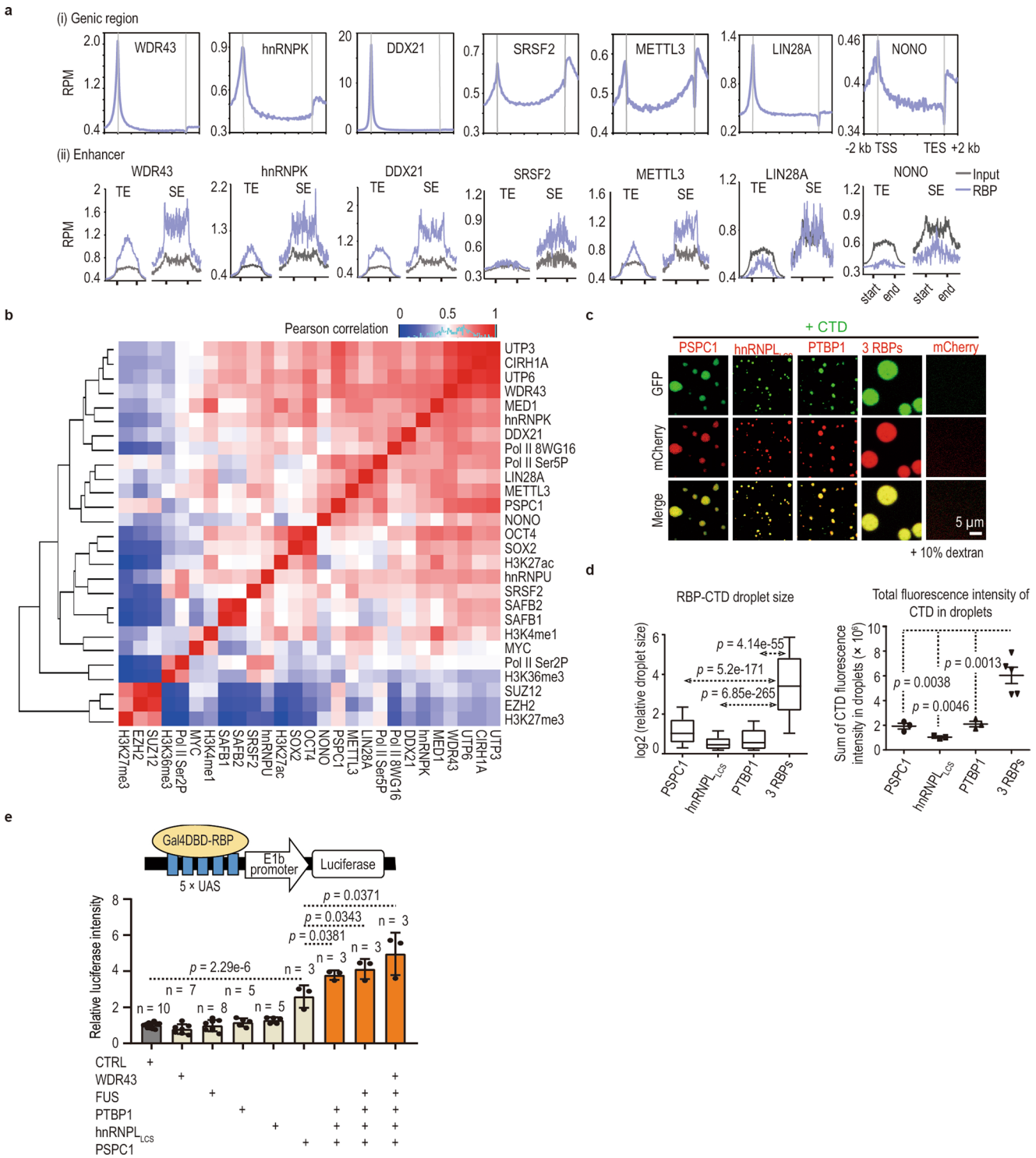
Extended Data Fig. 7 | PSPC1 promotes Pol II transcription in vitro. a, Schematic diagram of the template DNA. 'BR' stands for bubble right (the 601 sequence). The template was labeled with either biotin (panel **c**) or Cy5.5 (panels **d-f**). **b**, Silver staining of purified Pol II. **c**, Schematic diagram (upper) and autoradiography (lower) of *in vitro* transcription with various DNA templates and nucleotides as indicated. **d**, Autoradiography of *in vitro* transcription with the template in the presence of NTPs and ^{32}P-CTP. Additional replicate was shown in Fig. 3b. Besides the full-length (~278 nt) transcript, we detected high levels of abortive short transcripts (15-25 nt), which is consistent with published results[48]. It may result from the frequent fall-off of Pol II from the template due to dsDNA torsion during transcription. The asterisk indicates spike-in control (~170 nt). The relative intensities of full-length/short transcripts normalized to spike-in control were quantified and indicated at the bottom. **e**, Titration of heparin to reduce loosely docked Pol II on the template by ESMA. The concentration of heparin used in lanes 3-4 was chosen for *in vitro* transcription and EMSA assays. **f**, Titration of a competitor DNA to reduce nonspecific binding of PSPC1 to the BR template. The bottom panel shows quantification of the unbound Cy5.5-labeled BR template (indicated as 'DNA' by an arrowhead). In the absence of a competitor DNA (lanes 2-3 vs lane 1), the addition of $PSPC1_{FL}$ led to decreased signals of unbound templates in the bottom and increased BR signals (likely bound by PSPC1) that were stuck in the well on the top. When a 100:1 ratio of the competitor to BR template was used (lanes 8-9), the binding of PSPC1 to the BR template was minimized, while Pol II binding was not affected. We thus used the 100:1 ratio for *in vitro* transcription and EMSA assays.



Extended Data Fig. 8 | PSPC1 stabilizes Pol II during in vitro transcription. a, Effects of $PSPC1_{FL}$ on the binding of Pol II to BR template during *in vitro* transcription in the absence or presence of heparin by ESMA. The free template ('DNA') and the supershifted 'Pol II:DNA' bands are indicated on the left. The bands marked by single asterisk is likely to be a non-specified byproduct during BR template assembly and gel purification. The bands marked by double asterisks are likely to be a R-loop, given its sensitivity to RNase H. Heparin effectively removed the docking Pol II from the template in the absence of NTPs (comparing lane 9 to lane 3), but had negligible effects on the stalled or elongating Pol II (comparing lane 10 to 4 and lane 11 to 5). Importantly, addition of recombinant $PSPC1_{FL}$ consistently enhanced the Pol II:DNA signals in both the absence (lanes 6-8 vs 3-5) and the presence of heparin (lanes 12-14 vs 9-11). $PSPC1_{FL}$ did not form a stable complex with Pol II, suggesting weak interactions. **b-c**, Effects of PSPC1 mutants on the binding of Pol II to BR template during *in vitro* transcription (heparin included). $PSPC1_{RRMmut}$ and $PSPC1_{\Delta LCS2}$ had negligible effects on Pol II binding (panel **b**, lanes 8-11). Representative results of independent experiments are shown in panels **b** and **c**. **d**, Effects of recombinant proteins on the binding of Pol II to the BR template during *in vitro* transcription by ESMA. For panels **b** and **d**, the relative intensity of supershifted Pol II:DNA signals was indicated at the bottom. **e**, Phase diagram of PTBP1-CTD droplets in the presence of ESC RNA. We used 0.6 μM of CTD in all assays in the presence of 10% dextran. PTBP1 poorly phase separated in the absence of dextran. For panel **a**, **d** and **e**, experiments were repeated twice with similar results.



Extended Data Fig. 9 | PSPC1 co-localizes with Pol II and its acute degradation impairs transcription. a-e, ChIP-seq analysis of PSPC1. Panel **a** shows overlap between targets identified by two biological replicates of $^{AID\text{-}FB(KI)}$PSPC1 ChIP-seq and one ectopically tagged PSPC1 ($^{FB(EXO)}$PSPC1) ChIP-seq. Panels **b-d** show the overlap of $^{AID\text{-}FB(KI)}$PSPC1-targeted promoters (n = 5,262, overlapping promoters of two-biological replicates) and Pol II (hypo/Ser5P)-targeted promoters (13,791) (**b**), hypoP **(c)** or Ser5P ChIP-seq **(d)** respectively. *P*-values by two-sided Fisher's exact test are shown as indicated. Panel **e** shows UCSC genome browser view of ChIP-seq, CLIP-seq, TT-seq and EU-seq at representative loci. **f**, Average read density of PSPC1 CLIP-seq of two biological replicates on PSPC1 ChIP-seq summits (i) and vice versa (ii). **g**, Anti-FLAG ChIP-qPCR of PSPC1 and GFP proteins that were transiently expressed in ESCs. Data are shown as mean ± s.d. of 3 biological replicates. **h**, Effect of 1,6-hexanediol on $^{AID\text{-}FB(KI)}$PSPC1 binding to its targets. Data are shown as mean of 2 independent biological replicates. **i**, Anti-FLAG immunofluorescence analysis. Left, representative images; right, quantification of relative fluorescence intensities along the yellow lines. **j**, Time-course ChIP-qPCR analysis of Pol II Ser5P (left) and Ser2P (right) upon PSPC1 degradation. Data are shown as mean ± s.d. of biological replicates. Ser5P ChIP: 2 replicates; Ser2P ChIP: 4 replicates for 'no IAA' and '+ IAA 6 h'; 3 replicates for '+ IAA 3 h' and '+ IAA 24 h'. *P*-values by one-sided Student's t-test are shown as indicated. For panels **i**, and **j**, experiments were repeated twice with similar results. For panels **g** and **h**, the relative fold of enrichment at each target was normalized to *OLFR1437*. *P*-values by one-sided (**h** and **j**) or two-sided (**g**) Student's t-test are shown as indicated.



Extended Data Fig. 10 | Genome-wide co-occupancy of chrRBPs with Pol II at promoters and enhancers. a, Metagene analysis of ChIP-seq signals of various RBPs across all mouse genes (n = 32,944) (i) and enhancers (ii). The y-axis is reads per million reads (RPM). Related to Fig. 6a. **b**, Heatmap showing hierarchical clustering of ChIP-seq signals of various RBPs, Pol II, transcription and chromatin regulators, and histone marks in ESCs. The color indicates the Pearson correlation value. **c-d**, Phase-separation assay of various RBPs with the CTD. The concentrations of GFP-tagged CTD, mCherry, and mCherry-tagged PSPC1, PTBP1 and LCS domain of hnRNPL ($hnRNPL_{LCS}$) used here are 0.6 µM, 45 µM, 5 µM, 10 µM and 20 µM respectively. Representative pictures are shown in panel **c**. Quantification for RBP-CTD droplet size (left) and the total CTD fluorescence intensity (right) in the droplet are shown in panel **d**. In the left panel, the y-axis is log2 (relative droplet size). The whiskers are drawn within 10-90 percentile. Points below and above the whiskers are not shown. The median droplet size is 1.04 μm^2 for PSPC1 (n = 782), 0.37 μm^2 for $hnRNPL_{LCS}$ (n = 6,158), 0.48 μm^2 for PTBP1 (n = 2,374), and 9.60 μm^2 for all 3 RBPs together (n = 554). In the right panel, the y-axis shows the sum of fluorescence intensity of CTD in droplets in each field of view. Data are shown as mean ± s.d. of 3 fields for 'PSPC1', '$hnRNPL_{LCS}$' and 'PTBP1', and 5 fields for '3RBPs'. **e**, Promoter tethering assay in 293T cells. The y-axis shows the relative luciferase intensity normalized to the Gal4DBD control. Data are shown as mean ± s.d. For panels **d** and **e**, *p*-values by two-sided Student's t-test are shown as indicated.

nature research

Corresponding author(s): Xiaohua Shen

Last updated by author(s): Sep 6, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Bio-Rad CFX384 Real Time System was used for RT-qPCR data collection. Illumina HiSeq 2500 and HiSeq X TEN were used for sequencing data collection. Nikon A1RMP microscopy and related softwares were used for collecting and analyzing imaging and video data.
Data analysis	For TT-seq, EU-seq, CLIP-seq and ChIP-seq data analysis, Bowtie v1.0.0, Bowtie v2.2.3, Tophat v2.0.10, BEDTools v2.17.0, MACS v1.4.2, TrimGalore v0.6.6 , Novoalign V3.09.00, Hisat2-2.1.0 and Piranha-1.2.1 were used. Integrative Genomics Viewer v2.8.13 were used for signal visualization of sequencing data. R 4.0.2 and R 3.3.0, Excel 2016, GraphPad Prism v5.0, Origin9 were used for statistical analysis. FlowJo v7.6.1 was used for FACS analysis. Image J v1.8.0 and Nikon NIS-element AR v5.2 software were used for quantification of Western-blot and imaging data. For mass spec quantification, Maxquant v1.5.8.3 and Proteome Discoverer v1.4 were used.

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

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

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

The main data supporting the findings of this study are available within the article and Supplementary Information files. Sequencing data have been deposited in the GEO database under the accession number GEO: GSE150399.

Field-specific reporting

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

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

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size	Sample sizes were not predetermined and were chosen based on our prior experience and common standards for detecting statistically significant differences between conditions (Yin et al. 2020; Bi et al. 2019). For mass spec analysis, we used 3 different quantitative mass spec methods and each condition includes at least 2 independent biological replicates. For TT-seq, CLIP-seq and ChIP-seq analysis of Pol II, PSPC1, and mutants, we included 2 independent biological replicates, which are sufficient for sequencing analysis. For ChIP-seq analysis of other 6 RBPs, one biological replicate was sent for sequencing. Additional independent experiments were repeated and ChIP-qPCRs were performed to validate the key results of sequencing. As these 6 RBPs are not the focus of the paper and the published work have performed RBP ChIP-seq with similar results (Xiao et al. 2019), we think it's uninformative to perform additional sequencing considering the vast cost and tremendous work to process. Sample sizes of the rest experiments are all noted in figure legends.
Data exclusions	No data were excluded.
Replication	All experiments were repeated (biological replicates) at least twice. All attempts of replication were successful.
Randomization	There is no bias when harvesting samples or collecting data. All samples were randomly allocated into different experiment groups and performed in parallel.
Blinding	Experiments in this study was not done blinded. The samples harvest process made it impossible to be blinded. Data were collected in parallel to minimize bias

Behavioural & social sciences study design

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

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*
Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

Ecological, evolutionary & environmental sciences study design

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

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*
Research sample	*Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.*
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data collection	*Describe the data collection procedure, including who recorded the data and how.*
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access & import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

WDR43 (Abclonal, Q659), hnRNPU (Abcam, ab180952), hnRNPL (Santa Cruz, sc-32317), PTBP1 (Abclonal, A6107), DDX5 (Abcam, ab126730), FUS (Abcam, ab70381), PSPC1 (Abcam, ab104238), pan-Pol II (Abcam, ab52202), EZH2 (CST, 5246S), SUZ12 (CST, 3737S), CTCF (Abcam, ab128873), CHD1 (CST, 4351S), TOP2A (Abcam, ab52934), H3 (Easybio, BE3015), DDX21 (Novus Biologicals, NBP1-83310), FUBP1 (Abcam, ab181111), FUBP3 (Abcam, ab181025), LIN28A (Abcam, ab155542), NCL (Abcam, ab134164), TUBULIN (CWBIO, CW0098), GFP (Santa Cruz, sc-9996), SNRNP70 (Santa Cruz, sc-390988), hypo-phosphorylated Pol II (8WG16) (Covance, MMS-126R), Pol II Ser5P (CST, 13523), Pol II Ser2P (CST, 13499), TBP (Santa Cruz, sc-421), FLAG (Sigma-Aldrich, F3165), GFP (Santa Cruz, sc-9996) and mCherry (CST, 43590).

Validation

WDR43 (Abclonal, Q659): this antibody has been claimed to react with mouse WDR43 by the manufacturer. We have demonstrated in prior work (Bi et al. 2019) that this antibody can be used for western-blot.

hnRNPU (Abcam, ab180952): this antibody has been predicted to react with mouse hnRNPU by the manufacturer. It can be used for immunoblot in prior work (Kishikawa et al. 2016).

hnRNPL (Santa Cruz, sc-32317): this antibody has been claimed to react with mouse hnRNPL by the manufacturer. Validation data for the antibodies used can be found as follows: https://www.scbt.com/p/hnrnp-l-antibody-4d11?requestFrom=search.

PTBP1 (Abclonal, A6107): this antibody has been claimed to react with mouse hnRNPL by the manufacturer. Validation data for the antibodies used can be found as follows: https://abclonal.com.cn/catalog/A6107.

DDX5 (Abcam, ab126730): this antibody has been claimed to react with mouse DDX5 by the manufacturer. Validation data for the antibodies used can be found as follows: https://www.abcam.cn/ddx5-antibody-epr7239-ab126730.html.

FUS (Abcam, ab70381): this antibody has been claimed to react with mouse FUS by the manufacturer. Validation data for the antibodies used can be found as follows: https://www.abcam.cn/tlsfus-antibody-ab70381.html.

PSPC1 (Abcam, ab104238): this antibody has been claimed to react with mouse PSPC1 by the manufacturer. Validation data for the antibodies used can be found as follows: https://www.abcam.cn/pspc1-antibody-ab104238.html

pan-Pol II (Abcam, ab52202): this antibody has been claimed to react with mouse Pol II by the manufacturer. Validation data for the antibodies used can be found as follows: https://www.abcam.cn/rna-polymerase-ii-ctd-repeat-ysptsps-antibody-ab52202.html

EZH2 (CST, 5246S): this antibody has been claimed to react with mouse EZH2 by the manufacturer. Validation data for the antibodies used can be found as follows: https://www.cellsignal.cn/products/primary-antibodies/ezh2-d2c9-xp-rabbit-mab/5246?site-search-type=Products&N=4294956287&Ntt=5246s&fromPage=plp&_requestid=1011054.

SUZ12 (CST, 3737S): this antibody has been claimed to react with mouse SUZ12 by the manufacturer. Validation data for the antibodies used can be found as follows: https://www.cellsignal.cn/products/primary-antibodies/suz12-d39f6-xp-rabbit-mab/3737?site-search-type=Products&N=4294956287&Ntt=3737s&fromPage=plp&_requestid=1011350

CTCF (Abcam, ab128873): this antibody has been claimed to react with mouse CTCF by the manufacturer. Validation data for the antibodies used can be found as follows: https://www.abcam.cn/ctcf-antibody-epr7314b-chip-grade-ab128873.html.

CHD1 (CST, 4351S): this antibody has been claimed to react with mouse CHD1 by the manufacturer. Validation data for the antibodies used can be found as follows: https://www.cellsignal.cn/products/primary-antibodies/chd1-d8c2-rabbit-mab/4351?site-search-type=Products&N=4294956287&Ntt=4351s&fromPage=plp&_requestid=1011633.

TOP2A (Abcam, ab52934): this antibody has been claimed to react with mouse TOP2A by the manufacturer. Validation data for the antibodies used can be found as follows: https://www.abcam.cn/topoisomerase-ii-alpha-antibody-ep1102y-ab52934.html.

H3 (Easybio, BE3015): this antibody has been claimed to react with mouse H3 by the manufacturer. Validation data for the antibodies used can be found as follows: http://www.bioeasytech.com/home/product/article/id/10/sear/H3.html

DDX21 (Novus Biologicals, NBP1-83310): this antibody has been claimed to react with mouse DDX21 by the manufacturer. Validation data for the antibodies used can be found as follows: https://www.novusbio.com/products/ddx21-antibody_nbp1-83310.

FUBP1 (Abcam, ab181111): this antibody has been claimed to react with mouse FUBP1 by the manufacturer. Validation data for the antibodies used can be found as follows: https://www.abcam.cn/fubp1fbp-antibody-epr12327-ab181111.html.

FUBP3 (Abcam, ab181025): this antibody has been claimed to react with mouse FUBP3 by the manufacturer. Validation data for the antibodies used can be found as follows: https://www.abcam.cn/fubp3-antibody-epr13174-ab181025.html

LIN28A (Abcam, ab155542): this antibody has been claimed to react with mouse LIN28A by the manufacturer. Validation data for the antibodies used can be found as follows: https://www.abcam.cn/lin28a-antibody-ab155542.html

NCL (Abcam, ab134164): this antibody has been claimed to react with mouse NCL by the manufacturer. Validation data for the antibodies used can be found as follows: https://www.abcam.cn/nucleolin-antibody-epr7951-ab134164.html

TUBULIN (CWBIO, CW0098): this antibody has been claimed to react with mouse TUBULIN by the manufacturer. Validation data for the antibodies used can be found as follows: https://www.cwbiotech.com/goods/index/id/10115

GFP (Santa Cruz, sc-9996): this antibody has been claimed to react with mouse GFP by the manufacturer. Validation data for the antibodies used can be found as follows: https://www.scbt.com/p/gfp-antibody-b-2?requestFrom=search

SNRNP70 (Santa Cruz, sc-390988): this antibody has been claimed to react with mouse SNRNP70 by the manufacturer. Validation data for the antibodies used can be found as follows: https://www.scbt.com/p/u1-snrnp-70-antibody-e-4?requestFrom=search

hypo-phosphorylated Pol II (8WG16) (Covance, MMS-126R): this antibody has been claimed to react with mouse hypo-Pol II by the manufacturer. Validation data for the antibodies used can be found as follows: https://www.biolegend.com/en-us/products/purified-anti-rna-polymerase-ii-rpb1-antibody-11666.

Pol II Ser5P (CST, 13523): this antibody has been claimed to react with mouse Ser5P by the manufacturer. Validation data for the antibodies used can be found as follows:https://www.biolegend.com/en-us/products/purified-anti-rna-polymerase-ii-rpb1-antibody-11666.

Pol II Ser2P (CST, 13499): this antibody has been claimed to react with mouse Ser2P by the manufacturer. Validation data for the antibodies used can be found as follows:https://www.cellsignal.cn/products/primary-antibodies/phospho-rpb1-ctd-ser2-e1z3g-rabbit-mab/13499?site-search-type=Products&N=4294956287&Ntt=13499&fromPage=plp&_requestid=1013250

TBP (Santa Cruz, sc-421): this antibody has been claimed to react with mouse TBP by the manufacturer. Validation data for the antibodies used can be found as follows: https://www.scbt.com/p/tfiid-antibody-58c9?requestFrom=search

FLAG (Sigma-Aldrich, F3165): this antibody has been claimed to react with mouse FLAG by the manufacturer. Validation data for the antibodies used can be found as follows: https://www.sigmaaldrich.com/catalog/product/sigma/f3165?lang=zh&region=CN

GFP (Santa Cruz, sc-9996): this antibody has been claimed to react with mouse GFP by the manufacturer. Validation data for the antibodies used can be found as follows: https://www.scbt.com/p/gfp-antibody-b-2?requestFrom=search
mCherry (CST, 43590): this antibody has been claimed to react with mouse mCherry by the manufacturer. Validation data for the antibodies used can be found as follows: https://www.cellsignal.cn/products/primary-antibodies/mcherry-e5d8f-rabbit-mab/43590?site-search-type=Products&N=4294956287&Ntt=43590&fromPage=plp&_requestid=1013728

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK293T cells were obtained from ATCC (CRL-3216), mouse ESCs (46C) was a gift from Austin Smith's lab. Mouse ESCs (CJ9) were stored in the Shen Laboratory and have been described previously (Luo et al., 2016; Shen et al., 2008, 2009)
Authentication	The cell lines have been used in the lab for over 4 years, so authentications were not performed. The PSPC1-AID cell line and other RBP cell-lines constructed in this study were confirmed by PCR and Western blot.
Mycoplasma contamination	All cell lines have been tested for mycoplasma contamination free by PCR.
Commonly misidentified lines (See ICLAC register)	No cell lines used in this study were found in the database of commonly misidentified cell lines that is maintained by ICLAC and NCB BioSample.

Palaeontology and Archaeology

Specimen provenance	*Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).*
Specimen deposition	*Indicate where the specimens have been deposited to permit free access by other researchers.*
Dating methods	*If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	*For laboratory animals, report species, strain, sex and age OR state that the study did not involve laboratory animals.*
Wild animals	*Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*
Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*
Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, gender, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*

Ethics oversight | *Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*
Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*
Outcomes	*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
☐	☐	Public health
☐	☐	National security
☐	☐	Crops and/or livestock
☐	☐	Ecosystems
☐	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☐	☐	Demonstrate how to render a vaccine ineffective
☐	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☐	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☐	☐	Increase transmissibility of a pathogen
☐	☐	Alter the host range of a pathogen
☐	☐	Enable evasion of diagnostic/detection modalities
☐	☐	Enable the weaponization of a biological agent or toxin
☐	☐	Any other potentially harmful combination of experiments and agents

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE150399
Files in database submission	GSM4548413 PSPC1-KI-ChIP-1 GSM4548414 PSPC1-KI-ChIP-2 GSM4548415 PSPC1-0h-input GSM4548416 PSPC1-6h-input GSM4548417 PSPC1-0h-Ser2P-1 GSM4548418 PSPC1-0h-Ser2P-2 GSM4548419 PSPC1-6h-Ser2P-1

GSM4548420 PSPC1-6h-Ser2P-2
GSM4548421 PSPC1-0h-Ser5P-1
GSM4548422 PSPC1-0h-Ser5P-2
GSM4548423 PSPC1-6h-Ser5P-1
GSM4548424 PSPC1-6h-Ser5P-2
GSM4548425 PSPC1-0h-TTseq-1
GSM4548426 PSPC1-0h-TTseq-2
GSM4548427 PSPC1-3h-TTseq-1
GSM4548428 PSPC1-3h-TTseq-2
GSM4548429 PSPC1-6h-TTseq-1
GSM4548430 PSPC1-6h-TTseq-2
GSM4548431 PSPC1-24h-TTseq-1
GSM4548432 PSPC1-24h-TTseq-2
GSM4548447 hnRNPU-OE-ChIP
GSM4548448 SAFB-OE-ChIP
GSM4548449 SAFB2-OE-ChIP
GSM4548450 CIRH1A-OE-ChIP
GSM4548451 UTP3-OE-ChIP
GSM4548452 UTP6-OE-ChIP
GSM4548453 PSPC1-OE-ChIP
GSM4548454 SAFB-OE-ChIP-input
GSM5257774 PSPC1-WT-ChIP-1
GSM5257775 PSPC1-WT-ChIP-2
GSM5257776 PSPC1-trun-ChIP-1
GSM5257777 PSPC1-trun-ChIP-2
GSM5257778 PSPC1-RRMmut-ChIP-1
GSM5257779 PSPC1-RRMmut-ChIP-2
GSM5257780 PSPC1-0h-WT-Ser2P-1
GSM5257781 PSPC1-0h-WT-Ser2P-2
GSM5257782 PSPC1-3h-WT-Ser2P-1
GSM5257783 PSPC1-3h-WT-Ser2P-2
GSM5257784 PSPC1-0h-trun-Ser2P-1
GSM5257785 PSPC1-0h-trun-Ser2P-2
GSM5257786 PSPC1-3h-trun-Ser2P-1
GSM5257787 PSPC1-3h-trun-Ser2P-2
GSM5257788 PSPC1-0h-RRMmut-Ser2P-1
GSM5257789 PSPC1-0h-RRMmut-Ser2P-2
GSM5257790 PSPC1-3h-RRMmut-Ser2P-1
GSM5257791 PSPC1-3h-RRMmut-Ser2P-2
GSM5257792 PSPC1-0h-GFP-Ser2P-1
GSM5257793 PSPC1-0h-GFP-Ser2P-2
GSM5257794 PSPC1-3h-GFP-Ser2P-1
GSM5257795 PSPC1-3h-GFP-Ser2P-2
GSM5478393 PSPC1-CLIP-1
GSM5478394 PSPC1-CLIP-2
GSM5478395 PSPC1-0h-EUseq-1
GSM5478396 PSPC1-0h-EUseq-2
GSM5478397 PSPC1-3h-EUseq-1
GSM5478398 PSPC1-3h-EUseq-2
GSM5478399 PSPC1-9h-EUseq-1
GSM5478400 PSPC1-9h-EUseq-2

Genome browser session
(e.g. UCSC)

IGV

Methodology

Replicates

Two biological replicates for Pol II (Ser2P or Ser5P), PSPC1 wildtype and mutant ChIP-seq. One replicate for other RBPs ChIP-seq (including hnRNPU, SAFB, SAFB2, UTP3, UTP6, CIRH1A).

Sequencing depth

PSPC1-KI-ChIP-1; total reads 42005496
PSPC1-KI-ChIP-2; total reads 34201312
PSPC1-0h-input; total reads 45817613
PSPC1-6h-input; total reads 53041096
PSPC1-0h-Ser2P-1; total reads 15748676

PSPC1-0h-Ser2P-2; total reads 18158686
PSPC1-6h-Ser2P-1; total reads 18558424
PSPC1-6h-Ser2P-2; total reads 18135656
PSPC1-0h-Ser5P-1; total reads 18096189
PSPC1-0h-Ser5P-2; total reads 24437005
PSPC1-6h-Ser5P-1; total reads 18945649
PSPC1-6h-Ser5P-2; total reads 15216634
hnRNPU-OE-ChIP; total reads 14486013
SAFB-OE-ChIP; total reads 12998233
SAFB2-OE-ChIP; total reads 12998233
CIRH1A-OE-ChIP; total reads 64041770
UTP3-OE-ChIP; total reads 74511504
UTP6-OE-ChIP; total reads 63657261
PSPC1-OE-ChIP; total reads 11332793
SAFB-OE-ChIP-input; total reads 16506455
PSPC1-WT-ChIP-1; total reads 22064130
PSPC1-WT-ChIP-2; total reads 24764342
PSPC1-trun-ChIP-1; total reads 49236074
PSPC1-trun-ChIP-2; total reads 47558706
PSPC1-RRMmut-ChIP-1; total reads 32801106
PSPC1-RRMmut-ChIP-2; total reads 38863624
PSPC1-0h-WT-Ser2P-1; total reads 17441892
PSPC1-0h-WT-Ser2P-2; total reads 15510552
PSPC1-3h-WT-Ser2P-1; total reads 14692830
PSPC1-3h-WT-Ser2P-2; total reads 16377350
PSPC1-0h-trun-Ser2P-1; total reads 25904800
PSPC1-0h-trun-Ser2P-2; total reads 23572834
PSPC1-3h-trun-Ser2P-1; total reads 17672194
PSPC1-3h-trun-Ser2P-2; total reads 15147506
PSPC1-0h-RRMmut-Ser2P-1; total reads 15723984
PSPC1-0h-RRMmut-Ser2P-2; total reads 19833704
PSPC1-3h-RRMmut-Ser2P-1; total reads 11753886
PSPC1-3h-RRMmut-Ser2P-2; total reads 16935708
PSPC1-0h-GFP-Ser2P-1; total reads 24743470
PSPC1-0h-GFP-Ser2P-2; total reads 22063828
PSPC1-3h-GFP-Ser2P-1; total reads 33521248
PSPC1-3h-GFP-Ser2P-2; total reads 25826454

Antibodies

Pol II Ser5P (CST, 13523)
Pol II Ser2P (CST, 13499)
FLAG (Sigma-Aldrich, F3165)

Peak calling parameters

macs14 -t #.bed -f BED -g mm -n # -p 1e-3 -B -S

Data quality

FASTQC is run to check the sequencing quality.

Software

Bowtie2, Bedtools, macs, ngs.plot;

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Cells were labelled for 20 minutes with EU (5-ethynyl uridine, Jena Bioscience CLK-N002, final concentration at 1 mM). Harvested cells were firstly labeled with Zombie AquaTM (BioLegend) to mark dead cells. Cells were then fixed with 4%

	formaldehyde for 15 minutes at room temperature and permeabilized for 5 minutes with PBS supplemented with 0.5% TritonX-100. Next, the cells were labeled with Alexa 647 using a Click-iT Cell Reaction Buffer Kit (Life Technologies, C10269) following the manufacturer's instructions. Labelled cells were subjected to fluorescence-activated cell sorting (FACS) analysis.
Instrument	BDCalibur
Software	Data collection: BD FACSDiva 8.0; Data analysis: FlowJo 7.6.1
Cell population abundance	No sorting was conducted.
Gating strategy	FSC/SSC were used to discern single cells from doublets/multiple cells. Zombie AquaTM was used to discern live cells from dead cells. Samples without EU labeling were used to establish boundaries between negative and positive cells

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

Magnetic resonance imaging

Experimental design

Design type	*Indicate task or resting state; event-related or block design.*
Design specifications	*Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*
Behavioral performance measures	*State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s)	*Specify: functional, structural, diffusion, perfusion.*
Field strength	*Specify in Tesla*
Sequence & imaging parameters	*Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*
Area of acquisition	*State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software	*Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*
Normalization	*If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*
Normalization template	*Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*
Noise and artifact removal	*Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*
Volume censoring	*Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.*

Statistical modeling & inference

Model type and settings	*Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).*
Effect(s) tested	*Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.*

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016)	*Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.*
Correction	*Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).*

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity	*Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).*
Graph analysis	*Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).*
Multivariate modeling and predictive analysis	*Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.*