LETTER

https://doi.org/10.1038/s41586-019-1464-0

Pol II phosphorylation regulates a switch between transcriptional and splicing condensates

Yang Eric Guo[1,10], John C. Manteiga[1,2,10], Jonathan E. Henninger[1], Benjamin R. Sabari[1], Alessandra Dall'Agnese[1], Nancy M. Hannett[1], Jan-Hendrik Spille[3,8], Lena K. Afeyan[1,2], Alicia V. Zamudio[1,2], Krishna Shrinivas[4,5], Brian J. Abraham[1,9], Ann Boija[1], Tim-Michael Decker[6], Jenna K. Rimel[6], Charli B. Fant[6], Tong Ihn Lee[1], Ibrahim I. Cisse[3], Phillip A. Sharp[2,7], Dylan J. Taatjes[6] & Richard A. Young[1,2]*

The synthesis of pre-mRNA by RNA polymerase II (Pol II) involves the formation of a transcription initiation complex, and a transition to an elongation complex[1–4]. The large subunit of Pol II contains an intrinsically disordered C-terminal domain that is phosphorylated by cyclin-dependent kinases during the transition from initiation to elongation, thus influencing the interaction of the C-terminal domain with different components of the initiation or the RNA-splicing apparatus[5,6]. Recent observations suggest that this model provides only a partial picture of the effects of phosphorylation of the C-terminal domain[7–12]. Both the transcription-initiation machinery and the splicing machinery can form phase-separated condensates that contain large numbers of component molecules: hundreds of molecules of Pol II and mediator are concentrated in condensates at super-enhancers[7,8], and large numbers of splicing factors are concentrated in nuclear speckles, some of which occur at highly active transcription sites[9–12]. Here we investigate whether the phosphorylation of the Pol II C-terminal domain regulates the incorporation of Pol II into phase-separated condensates that are associated with transcription initiation and splicing. We find that the hypophosphorylated C-terminal domain of Pol II is incorporated into mediator condensates and that phosphorylation by regulatory cyclin-dependent kinases reduces this incorporation. We also find that the hyperphosphorylated C-terminal domain is preferentially incorporated into condensates that are formed by splicing factors. These results suggest that phosphorylation of the Pol II C-terminal domain drives an exchange from condensates that are involved in transcription initiation to those that are involved in RNA processing, and implicates phosphorylation as a mechanism that regulates condensate preference.

Previous studies have shown that the hypophosphorylated C-terminal domain (CTD) of Pol II can interact with mediator[5,6], and that Pol II and mediator occur in condensates at super-enhancers[7,8] (Fig. 1a, b, Extended Data Fig. 1a), but have not established whether the CTD contributes to Pol II interactions with mediator condensates. To investigate whether the Pol II CTD is incorporated into mediator condensates, we purified the human mediator complex and recombinant Pol II CTD fused to GFP (full-length GFP–CTD52, and the truncated forms GFP–CTD26 and GFP–CTD10) and measured condensate formation in an in vitro droplet assay. Mediator droplets incorporated and concentrated GFP–CTD52 to a much greater extent than they did the truncated forms or control GFP (Fig. 1c). We further investigated the interaction of the CTD with mediator using MED1, which is the largest subunit of the mediator complex[13]. MED1 has proven to be a useful surrogate for mediator condensates in previous studies[8,14], has an exceptionally large intrinsically disordered region (IDR) that contributes to condensate formation[8] and has been shown to associate with Pol II in human cells[15]. Droplet assays revealed that the IDR of MED1 fused to mCherry (mCherry–MED1-IDR) formed condensates that incorporated and concentrated GFP–CTD52 to a greater extent than they did the truncated forms or GFP alone (Fig. 1d). The GFP–CTD52 and MED1-IDR condensates exhibited liquid-like fusion behaviour (Extended Data Fig. 1b), and showed evidence of dynamic internal rearrangement and internal–external exchange of molecules by fluorescence recovery after photobleaching (FRAP) (Extended Data Fig. 1c), which is consistent with liquid–liquid phase-separated condensates[16–18]. These results are consistent with the idea that the CTD of Pol II contributes to the incorporation of Pol II into mediator condensates.

We next sought to determine whether splicing-factor condensates occur at genes associated with super-enhancers because these genes are transcribed at especially high rates[19], RNA splicing can occur co-transcriptionally[20–22] (Fig. 1a), and some nuclear speckles have previously been reported to occur in the vicinity of highly transcribed genes[9–12]. We selected eight components of the splicing apparatus, and used immunofluorescence microscopy with concurrent nascent RNA fluorescence in situ hybridization (FISH) for *Nanog* and *Trim28* to determine whether the splicing apparatus occurs in puncta in the vicinity of these genes associated with super-enhancers. The results showed that all eight splicing factors occur in puncta at these two genes (Fig. 2a, Extended Data Fig. 2a). To gain additional insights into splicing-factor puncta that colocalize with Pol II, mouse embryonic stem cells (mouse ES cells) that were engineered to express endogenously tagged proteins were studied using lattice light-sheet imaging in live cells. It has previously been shown that large numbers of mediator and Pol II molecules can occur in puncta and that these puncta sometimes colocalize[7,8]. Using a similar approach, we found here that large numbers of SRSF2 molecules occur in puncta and some of these (approximately 15%) overlap with Pol II puncta (Fig. 2b, Extended Data Fig. 2b). Treatment of cells with pladienolide B (an inhibitor of splicing)—which reduced splicing as determined by a splicing reporter (Extended Data Fig. 3a)—reduced the levels of splicing factors but not of Pol II in puncta at *Trim28* DNA FISH foci (Fig. 2c, Extended Data Fig. 3b, c). This treatment also led to the incorporation of splicing factors into 'mega-speckles' at some distance from the gene (Fig. 2c), a phenomenon that has previously been observed when splicing is inhibited[23]. These results suggest that functional RNA-splicing apparatus is present in condensates at active genes associated with super-enhancers.

Actively transcribed genes may become associated with nuclear speckles or obtain splicing apparatus stored in speckles[9–12], which are thought to be phase-separated[16]. In live-cell imaging, we found that the SRSF2 puncta exhibited features of liquid-like condensates: all of these puncta showed evidence of dynamic internal rearrangement and

[1]Whitehead Institute for Biomedical Research, Cambridge, MA, USA. [2]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. [3]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA, USA. [4]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. [5]Institute of Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, USA. [6]Department of Biochemistry, University of Colorado, Boulder, CO, USA. [7]Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA, USA. [8]Present address: Department of Physics, University of Illinois at Chicago, Chicago, IL, USA. [9]Present address: Computational Biology, St Jude Children's Research Hospital, Memphis, TN, USA. [10]These authors contributed equally: Yang Eric Guo, John C. Manteiga.. *e-mail: young@wi.mit.edu

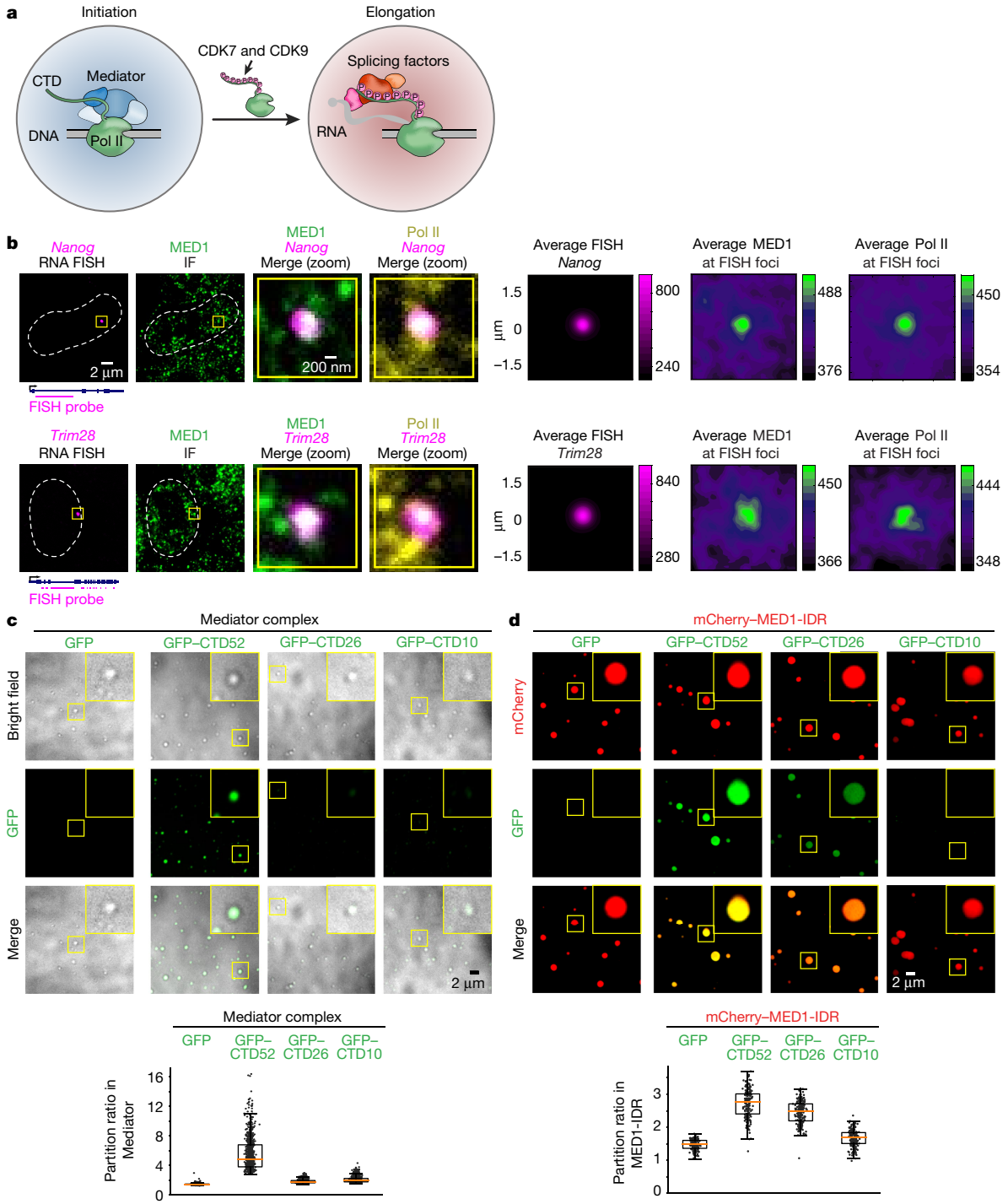


Fig. 1 | The CTD of Pol II is integrated and concentrated in mediator condensates. a, A model that depicts the transition from transcription initiation to elongation, and the role of phosphorylation of the Pol II CTD in this transition. During initiation, Pol II with a hypophosphorylated CTD interacts with mediator. CDK7 phosphorylation of the CTD leads to the formation of a paused Pol II at approximately 50–100 bp downstream of the initiation site, and subsequent CDK9 phosphorylation leads to pause release and elongation[1]. For simplicity, we show CDK7 and CDK9 phosphorylating the CTD, leading to elongation. During elongation, Pol II with a hyperphosphorylated CTD interacts with various RNA splicing factors[5,6]. The coloured compartments that surround the initiating and elongating polymerases represent initiation and splicing-factor condensates, respectively. **b**, Representative images exhibiting overlap between immunofluorescence (IF) of MED1 and Pol II with nascent RNA FISH of *Nanog* and *Trim28* in fixed mouse ES cells. The three columns on the right show the average RNA FISH signal and average MED1 or Pol II immunofluorescence signal centred on RNA FISH foci (see Methods). **c**, Representative images and quantification of partition ratios of droplet experiments that measured the incorporation of full-length or truncated CTD into droplets of human mediator complex. Purified human mediator complex (approximately 200 nM) (see Methods) was mixed with 10 µM GFP, GFP–CTD52 or truncated forms of GFP–CTD in droplet formation buffers with 140 mM monovalent salt and 16% Ficoll-400, and then visualized on a fluorescence microscope with the indicated filters. **d**, Representative images and quantification of partition ratios of droplet experiments that measured the incorporation of full-length or truncated CTD into MED1-IDR droplets. Purified human mCherry–MED1-IDR at 10 µM was mixed with 3.3 µM GFP, GFP–CTD52 or truncated forms of GFP–CTD in droplet formation buffers with 125 mM NaCl and 16% Ficoll-400.

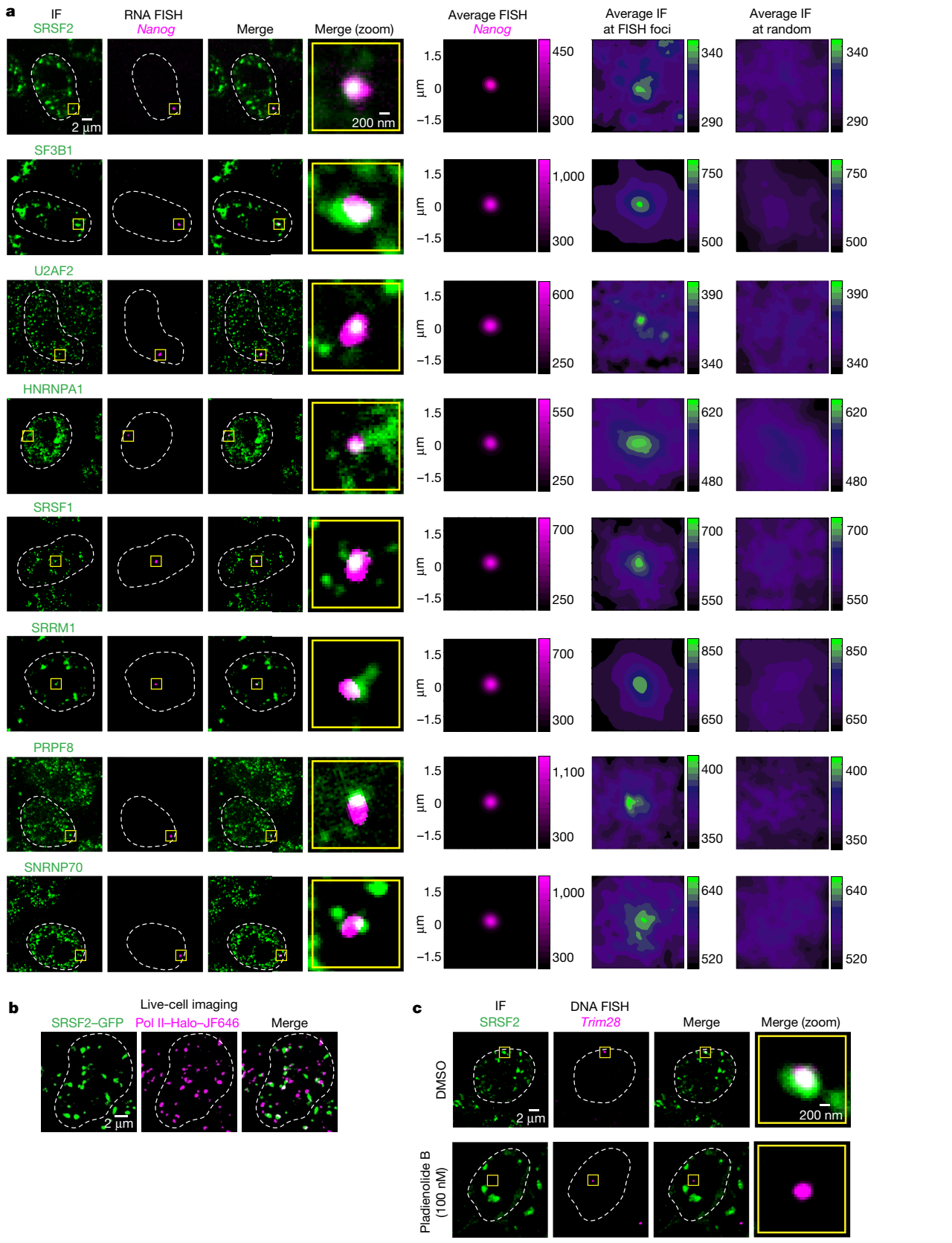


Fig. 2 | Splicing-factor condensates occur at active genes driven by super-enhancers. a, Representative images exhibiting overlap between immunofluorescence of the splicing factors SRSF2, SF3B1, U2AF2, HNRNPA1, SRSF1, SRRM1, PRPF8 or SNRNP70 with nascent RNA FISH of *Nanog* in fixed mouse ES cells. The rightmost column shows the average immunofluorescence signal for splicing factors centred on randomly selected nuclear positions (see Methods). **b**, Representative lattice light-sheet images of live mouse ES cells engineered to express GFP-tagged SRSF2 and Halo–JF646-tagged Pol II. Maximum intensity projection after background subtraction. **c**, Representative images exhibiting overlap or absence of overlap between immunofluorescence of SRSF2 and DNA FISH of *Trim28* in mouse ES cells treated with dimethylsulfoxide (DMSO) or the splicing inhibitor pladienolide B for 4 h.

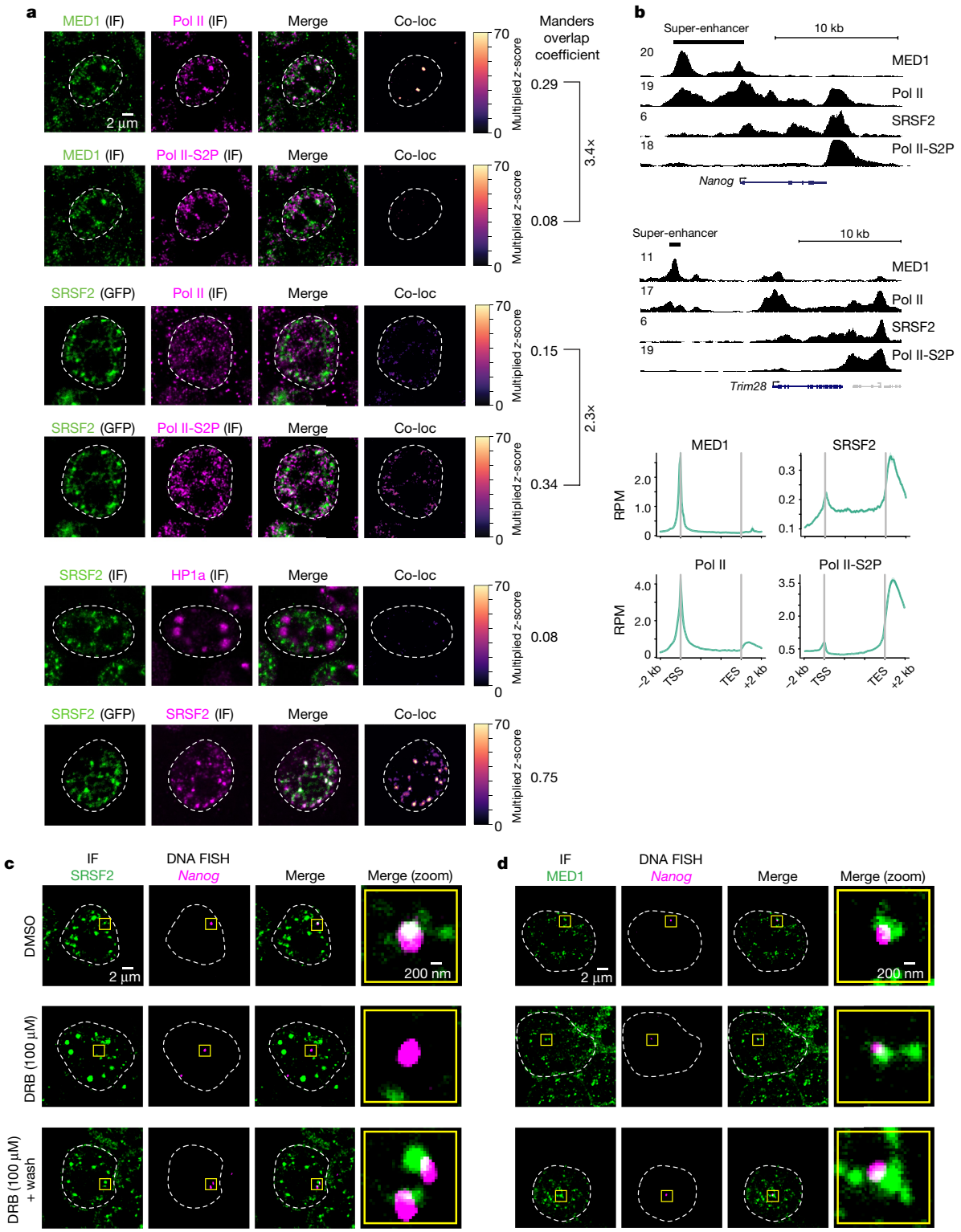


Fig. 3 | Pol II partitioning in transcriptional and splicing-factor condensates. a, Top four rows, immunofluorescence imaging using antibodies for the hypophosphorylated and serine-2-phosphorylated (S2P) Pol II CTD, coupled with immunofluorescence for MED1 or direct visualization of SRSF2 in the GFP–SRSF2 mouse ES cells. Bottom two rows, immunofluorescence for SRSF2 coupled to immunofluorescence for HP1a or direct visualization of SRSF2 in the GFP–SRSF2 mouse ES cells. The 'co-loc' column highlights overlapped pixels for each factor in an example *z*-slice, and the Manders overlap coefficient gives a relative score for the degree of overlap from multiple cells and images (see Methods). **b**, Top, representative ChIP–seq tracks of MED1, SRSF2 and the hypophosphorylated or serine-2-phosphorylated forms of Pol II in mouse ES cells. The *y* axis is in reads per million (RPM). Bottom, metagene plots of average ChIP–seq RPM for the same factors across gene bodies (see Methods). ChIP–seq was performed once for each factor with approximately 100 million cells. TES, transcription end site; TSS, transcription start site. **c**, Representative images exhibiting overlap or lack of overlap between immunofluorescence of SRSF2 and DNA FISH of *Nanog* in mouse ES cells treated with DMSO for 2 h, DRB for 2 h, or DRB for 2 h followed by a 2-h washout. **d**, Representative images exhibiting overlap between immunofluorescence of MED1 and DNA FISH of *Nanog* in mouse ES cells treated as in **c**.

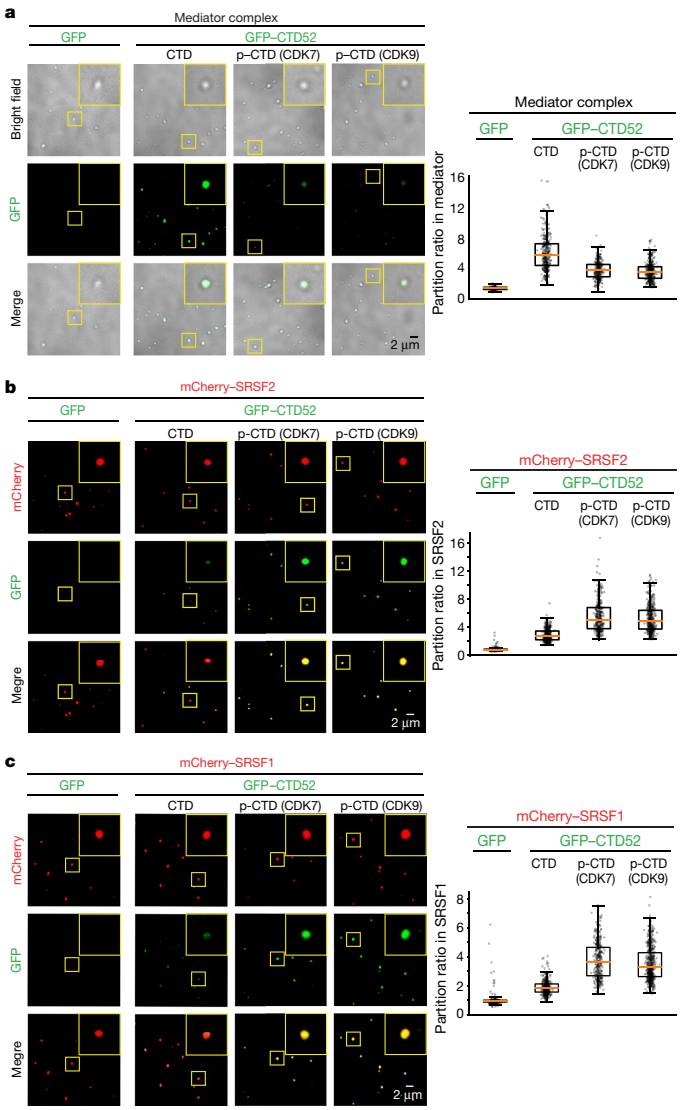


Fig. 4 | CTD phosphorylation promotes a condensate-preference switch between mediator and splicing-factor condensates. a, Representative images and quantification of partition ratios of droplet experiments that measured CTD incorporation into mediator droplets. Purified human mediator complex (approximately 200 nM) (see Methods) was mixed with 10 μM GFP, GFP–CTD52 or GFP–CTD52 phosphorylated with CDK7 or CDK9 in droplet formation buffers with 140 mM monovalent salt and 16% Ficoll-400. p-CTD, phosphorylated CTD. **b**, Representative images and quantification of partition ratios of droplet experiments that measured CTD incorporation into SRSF2 droplets. Purified human SRSF2 fused to mCherry (mCherry–SRSF2) at 2.5 μM was mixed with 3.3 μM GFP, GFP–CTD52 or GFP–CTD52 phosphorylated with CDK7 or CDK9 in droplet formation buffers with 120 mM NaCl and 10% PEG-8000. **c**, Representative images and quantification of partition ratios of droplet experiments that measured CTD incorporation into SRSF1 droplets under the same conditions as in **b**.

internal–external exchange of molecules by FRAP (Extended Data Fig. 4a–c), were sensitive to treatment with 1,6-hexanediol (Extended Data Fig. 4d), and some would occasionally fuse (Extended Data Fig. 4e). These results are consistent with previous reports on speckle behaviour[9] and suggest that the SRSF2-containing puncta that come into contact with active genes associated with super-enhancers are liquid-like condensates. At highly transcribed genes (such as those driven by super-enhancers), large numbers of Pol II molecules may be engaged in transcription elongation[7,8,17] and these might serve to recruit into condensates some portion of the apparatus that is otherwise located in speckles.

We next investigated whether hypophosphorylated Pol II tends to be associated with MED1 condensates, and whether, by contrast, hyperphosphorylated Pol II tends to be associated with SRSF2 condensates. Using immunofluorescence and antibodies against the hypophosphorylated CTD of Pol II or the CTD of Pol II phosphorylated at serine 2, we confirmed this prediction: MED1 puncta more frequently overlapped with signals for the hypophosphorylated CTD, whereas SRSF2 puncta more frequently overlapped with signals for the serine-2-phosphorylated CTD (Fig. 3a). A control experiment showed that there was essentially no overlap between SRSF2 puncta and the heterochromatin protein HP1a, and strong overlap between SRSF2 puncta visualized using independent methods (Fig. 3a). An independent experimental approach that used chromatin immunoprecipitation followed by sequencing (ChIP–seq) with antibodies against MED1, SRSF2 and the two phosphoforms of Pol II also confirmed that MED1 tends to occupy super-enhancers and promoters together with Pol II that contains a hypophosphorylated CTD, whereas SRSF2 is observed across the transcription unit and is prominent at the ends of genes together with Pol II that contains a serine-2-phosphorylated CTD (Fig. 3b).

If the formation or maintenance of splicing-factor condensates is dependent on the phosphorylation of the Pol II CTD, we would expect that inhibition of CTD phosphorylation in cells would prevent the formation of splicing-factor condensates at genes driven by super-enhancers. Indeed, inhibition of CTD phosphorylation by the cyclin-dependent kinase (CDK) inhibitor D-ribofuranosylbenzimidazole (DRB) caused a marked reduction in the occupancy of multiple components of splicing-factor condensates (SRSF2, SRSF1, SF3B1, U2AF2 and PRPF8) at *Nanog* or *Trim28* DNA FISH foci, and a washout of the drug led to a partial reestablishment of most of these splicing factors within two hours (Fig. 3c, Extended Data Fig. 5a, d). By contrast, treatment with DRB had minor effects on mediator and Pol II condensates (Fig. 3d, Extended Data Fig. 5b–d). These results suggest that phosphorylation of the Pol II CTD is necessary for the formation of splicing-factor condensates at these genes in vivo, although it is possible that altered phosphorylation of other substrates of CDKs may contribute to these observations.

The transition of Pol II from initiation to elongation is accompanied by phosphorylation of the CTD heptapeptide repeat by CDK7 and CDK9[24,25] (Fig. 1a). Phosphorylation of the CTD has previously been shown to affect its interaction with hydrogels formed by the low-complexity domains of FUS, EWS and TAF15 (FET) proteins[26], which suggests that phosphorylation may affect the condensate-interacting properties of the CTD. We investigated whether phosphorylation of the CTD by CDK7 or CDK9 would affect its incorporation into mediator condensates. We found that CTD phosphorylation by either CDK7 or CDK9 (Extended Data Fig. 6a, b) caused a reduction in CTD incorporation into mediator droplets (Fig. 4a). Similarly, CTD phosphorylation caused a reduction in the incorporation of the CTD into MED1-IDR droplets (Extended Data Fig. 6c, d). These results are consistent with a model in which phosphorylation of the Pol II CTD causes eviction from mediator condensates.

The observation that CTD phosphorylation is necessary for the formation of splicing-factor condensates at highly transcribed genes (Fig. 3c) suggests that CTD phosphorylation might enhance the partitioning of the Pol II CTD into condensates formed by splicing components. To investigate this idea, we first selected four human splicing factors (SRSF2, SRSF1, U2AF2 and HNRNPA1) as surrogates for the more complex splicing-factor condensates, and explored their condensate-forming properties. Each of the four purified human proteins fused to mCherry formed phase-separated droplets (Extended Data Fig. 7a, b). SRSF2 is one of multiple proteins involved in pre-mRNA splicing that contain serine-arginine (SR) dipeptide repeats, and it has an especially large SR-rich domain[27]; we therefore used SRSF2 as a core component to study whether it could concentrate the other three factors into heterotypic droplets. Indeed, all of these factors could form binary heterotypic droplets with SRSF2 (Extended Data Fig. 7c). We then asked whether phosphorylation of the Pol II CTD influences the incorporation of the CTD into splicing-factor condensates in vitro

using recombinant SRSF1 and SRSF2. The results showed that the unphosphorylated CTD was not efficiently incorporated into SRSF1 or SRSF2 droplets, whereas the CDK7- or CDK9-phosphorylated CTD was incorporated and concentrated in both SRSF1 and SRSF2 droplets (Fig. 4b, c, Extended Data Fig. 8a–c). The ability of SRSF2 to incorporate the phosphorylated CTD was dependent on the length of the CTD (Extended Data Fig. 8d)—as expected for a high-valency condensate interaction[16–18], and consistent with models in which the truncation of the CTD leads to splicing defects[28]. We conclude that phosphorylation of the Pol II CTD leads to a switch in the preference of the CTD for interactions between mediator condensates and condensates that contain proteins with SR-rich domains.

Our results indicate that phosphorylation of the Pol II CTD alters the condensate-partitioning behaviour of Pol II and may thus drive an exchange of Pol II from condensates that are involved in transcription initiation to condensates that are involved in RNA splicing at genes associated with super-enhancers. This model is consistent with evidence from previous studies that suggests that large clusters of Pol II can fuse with mediator condensates in cells[7], that phosphorylation dissolves CTD-mediated Pol II clusters[29], that CDK9 and cyclin T can interact with the Pol II CTD through a phase-separation mechanism[30], that Pol II is no longer associated with mediator during transcription elongation[13], and that nuclear speckles containing splicing factors can be observed at loci with high transcriptional activity[9–12]. Previous studies have shown that the Pol II CTD can interact with components of the transcription-initiation apparatus and RNA-processing machinery in a phosphoform-specific manner[5,6], but did not explore the possibility that these components occur in condensates or that phosphorylation of the Pol II CTD alters the partitioning behaviour of Pol II between these condensates. Our results reveal that mediator condensates and splicing-factor condensates occur at the same genes driven by super-enhancers, and suggest that the transition of Pol II from interactions with components involved in initiation to those involved in splicing can be mediated, in part, through a change in condensate partitioning that is regulated by phosphorylation of the CTD. These results also suggest that phosphorylation may be one of the mechanisms that regulate the condensate-partitioning of proteins in processes in which protein function involves eviction from one condensate and migration to another.

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/s41586-019-1464-0.

Received: 17 October 2018; Accepted: 10 July 2019;
Published online 7 August 2019.

1. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13**, 720–731 (2012).
2. 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).
3. Levine, M., Cattoglio, C. & Tjian, R. Looping back to leap forward: transcription enters a new era. *Cell* **157**, 13–25 (2014).
4. Sainsbury, S., Bernecky, C. & Cramer, P. Structural basis of transcription initiation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16**, 129–143 (2015).
5. Eick, D. & Geyer, M. The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem. Rev.* **113**, 8456–8490 (2013).
6. Ebmeier, C. C. et al. Human TFIIH kinase CDK7 regulates transcription-associated chromatin modifications. *Cell Reports* **20**, 1173–1186 (2017).
7. Cho, W. K. et al. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412–415 (2018).
8. Sabari, B. R. et al. Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, eaar3958 (2018).
9. Spector, D. L. & Lamond, A. I. Nuclear speckles. *Cold Spring Harb. Perspect. Biol.* **3**, a000646 (2011).
10. Chen, Y. et al. Mapping 3D genome organization relative to nuclear compartments using TSA-seq as a cytological ruler. *J. Cell Biol.* **217**, 4025–4048 (2018).
11. Quinodoz, S. A. et al. Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. *Cell* **174**, 744–757 (2018).
12. Hall, L. L., Smith, K. P., Byron, M. & Lawrence, J. B. Molecular anatomy of a speckle. *Anat. Rec. A Discov. Mol. Cell. Evol. Biol.* **288A**, 664–675 (2006).
13. Allen, B. L. & Taatjes, D. J. The Mediator complex: a central integrator of transcription. *Nat. Rev. Mol. Cell Biol.* **16**, 155–166 (2015).
14. Boija, A. et al. Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* **175**, 1842–1855 (2018).
15. Zhang, X. et al. MED1/TRAP220 exists predominantly in a TRAP/Mediator subpopulation enriched in RNA polymerase II and is required for ER-mediated transcription. *Mol. Cell* **19**, 89–100 (2005).
16. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
17. 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).
18. Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **357**, eaaf4382 (2017).
19. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
20. Braunschweig, U., Gueroussov, S., Plocik, A. M., Graveley, B. R. & Blencowe, B. J. Dynamic integration of splicing within gene regulatory pathways. *Cell* **152**, 1252–1269 (2013).
21. Herzel, L., Ottoz, D. S. M., Alpert, T. & Neugebauer, K. M. Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat. Rev. Mol. Cell Biol.* **18**, 637–650 (2017).
22. Hsin, J. P. & Manley, J. L. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev.* **26**, 2119–2137 (2012).
23. Kotake, Y. et al. Splicing factor SF3b as a target of the antitumor natural product pladienolide. *Nat. Chem. Biol.* **3**, 570–575 (2007).
24. Akhtar, M. S. et al. TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Mol. Cell* **34**, 387–393 (2009).
25. Czudnochowski, N., Bösken, C. A. & Geyer, M. Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. *Nat. Commun.* **3**, 842 (2012).
26. Kwon, I. et al. Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell* **155**, 1049–1060 (2013).
27. Long, J. C. & Caceres, J. F. The SR protein family of splicing factors: master regulators of gene expression. *Biochem. J.* **417**, 15–27 (2009).
28. McCracken, S. et al. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**, 357–361 (1997).
29. Boehning, M. et al. RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat. Struct. Mol. Biol.* **25**, 833–840 (2018).
30. Lu, H. et al. Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. *Nature* **558**, 318–323 (2018).

METHODS

Cell culture. V6.5 mouse ES cells were a gift from the Jaenisch laboratory. Cells were grown on 0.2% gelatinized (Sigma, G1890) tissue-culture plates in 2i medium, DMEM-F12 (Life Technologies, 11320082), 0.5× B27 supplement (Life Technologies, 17504044), 0.5× N2 supplement (Life Technologies, 17502048), an extra 0.5 mM L-glutamine (Gibco, 25030-081), 0.1 mM β-mercaptoethanol (Sigma, M7522), 1% penicillin–streptomycin (Life Technologies, 15140163), 1× nonessential amino acids (Gibco, 11140-050), 1,000 U/ml LIF (Chemico, ESG1107), 1 μM PD0325901 (Stemgent, 04-0006-10) and 3 μM CHIR99021 (Stemgent, 04-0004-10). Cells were grown at 37 °C with 5% CO_2 in a humidified incubator. For confocal imaging, cells were grown on glass coverslips (Carolina Biological Supply, 633029), coated with 5 μg/ml of poly-L-ornithine (Sigma Aldrich, P4957) for at least 30 min at 37 °C and with 5 μg/ml of laminin (Corning, 354232) for 2–16 h at 37 °C. For passaging, cells were washed in PBS (Life Technologies, AM9625), 1,000 U/ml LIF. TrypLE Express Enzyme (Life Technologies, 12604021) was used to detach cells from plates. TrypLE was quenched with fetal bovine serum (FBS) plus LIF medium (DMEM K/O (Gibco, 10829-018), 1× nonessential amino acids, 1% penicillin–streptomycin, 2 mM L-glutamine, 0.1 mM β-mercaptoethanol and 15% FBS (Sigma Aldrich, F4135).

Western blot. Purified phosphorylated Pol II CTD was mixed in 1× XT buffer (Bio-Rad) and run on 3–8% Criterion XT Tris-acetate Precast Gels (Bio-Rad) at 100 V, until the dye front reached the end of the gel. Protein was then wet-transferred to a 0.45-μm PVDF membrane (Millipore, IPVH00010) in ice-cold transfer buffer (25 mM Tris, 192 mM glycine and 10% methanol) at 250 mA for 2 h at 4 °C. After transfer, the membrane was blocked with 5% non-fat milk in TBS for 1 h at room temperature, with shaking. The membrane was then incubated with a 1:5,000 dilution of anti-GFP (Abcam no. ab290) antibodies in 5% non-fat milk in TBST overnight at 4 °C, with shaking. The membrane was washed three times with TBST for 10 min at room temperature, with shaking. The membrane was incubated with 1:10,000 secondary antibodies (GE Health) for 1 h at room temperature and washed three times in TBST for 5 min. Membranes were developed with Femto ECL substrate (Thermo Scientific, 34095) and imaged using a CCD camera.

Immunofluorescence with RNA FISH. Coverslips were coated at 37 °C with 5 μg/ml poly-L-ornithine (Sigma-Aldrich, P4957) for 30 min and 5 μg/ml of laminin (Corning, 354232) for 2 h. Cells were plated on the pre-coated cover slips and grown for 24 h, followed by fixation using 4% paraformaldehyde (PFA) (VWR, BT140770) in PBS for 10 min. After washing cells 3 times in PBS, the coverslips were put into a humidifying chamber or stored at 4 °C in PBS. Permeabilization of cells was performed using 0.5% Triton X100 (Sigma Aldrich, X100) in PBS for 10 min, followed by 3 PBS washes. Cells were blocked with 4% IgG-free bovine serum albumin (VWR, 102643-516) for 30 min. Cells were then incubated with the indicated primary antibody at a concentration of 1:500 in PBS for 4–16 h. Antibodies used for immunofluorescence in this study include SRSF2 (Abcam ab11826), MED1 (Abcam ab64965), Pol II CTD (Abcam ab817), Pol II CTD-S2 (Millipore 04-1571), SF3B1 (Sigma HPA050275), U2AF2 (Abcam ab37530), HNRNPA1 (Abcam ab5832), SRSF1 (Santa Cruz 33652), SRRM1 (Abcam ab221061), PRPF8 (Santa Cruz 55533), SNRNP70 (Sigma HPA043516) and HP1a (Abcam ab203432). SRSF2, MED1, Pol II CTD, U2AF2, HNRNPA1, SRSF1 and SRRM1 antibodies were validated in-house by small interfering RNA knockdown. Pol II CTD-S2 antibody was validated in-house by treatment of cells with DRB. SF3B1 and SNRNP70 antibodies were validated by The Cell Atlas and meet the 'enhanced' validation criteria. HP1a antibody was knockout-validated by Abcam. Cells were washed with PBS 3 times, followed by incubation with secondary antibody at a concentration of 1:500 in PBS for 1 h. After washing twice with PBS, cells were fixed using 4% PFA (VWR, BT140770) in PBS for 10 min. After two washes of PBS, wash buffer A (20% Stellaris RNA FISH wash buffer A (Biosearch Technologies SMF-WA1-60), 10% deionized formamide (EMD Millipore S4117)) in RNase-free water (Life Technologies, AM9932) was added to cells and incubated for 5 min. RNA probe (12.5 μM) in hybridization buffer (90% Stellaris RNA FISH hybridization buffer (Biosearch Technologies, SMF HB1-10) and 10% deionized formamide) was added to cells and incubated overnight at 37 °C. After washing with wash buffer A for 30 min at 37 °C, the nuclei were stained in 20 μm/ml Hoechst 33258 (Life Technologies, H3569) for 5 min, followed by a 5-min wash in wash buffer B (Biosearch Technologies, SMFWB1-20). Cells were washed once in water, followed by mounting the coverslip onto glass slides with Vectashield (VWR, 101098-042), and finally by sealing the cover slip with nail polish (Electron Microscopy Science Nm, 72180). Images were acquired on the RPI Spinning Disk confocal microscope with 100× objective using MetaMorph acquisition software and a Hammamatsu ORCA-ER CCD camera (W. M. Keck Microscopy Facility). Images were post-processed using Fiji Is Just ImageJ (FIJI). RNA FISH probes were custom-designed and generated by Biosearch Technologies (Stellaris RNA FISH) to target *Nanog* and *Trim28* intronic regions to visualize nascent RNA.

Cell line generation. V6.5 mouse ES cells were a gift from the Jaenisch laboratory, and have been verified using short tandem repeat analysis and tested negative for mycoplasma contamination. The other cell lines discussed in this article were generated from these V6.5 cells. CRISPR–Cas9 was used to generate mouse ES cells with SRSF2 endogenously tagged with monomeric enhanced GFP (mEGFP), and mouse ES cells with SRSF2 endogenously tagged with mEGFP and RPB1 endogenously tagged with Halo. Oligonucleotides coding for guide RNAs targeting the N terminus were cloned into a px330 vector that expresses Cas9 and mCherry (a gift from R. Jaenisch). The sequence that was targeted for *Srsf2* was 5′-CGTAGCTCATGGCTGCGAAG-3′. The sequence that was targeted for *Polr2a* was 5′-TGCCTCGCCATGCACGGGGG-3′. Repair templates were cloned into a pUC19 vector (NEB) containing mEGFP, a GS linker and 800-bp homology arms flanking the insert. Five hundred thousand mouse ES cells were transfected with 1.25 μg px330 vector and 1.25 μg repair templates using lipofectamine-3000 (Thermofisher). Cells were sorted two days after transfection for mCherry and one week after the first sort for mEGFP. Fifty thousand cells were serially diluted in a 6-well plate and colonies were picked 4 days after seeding into a 96-well plate. Two to four days after colony picking, cells were passaged into three plates. One plate was used for genotyping and the other two were frozen down at −80 °C in 10% DMSO, 10% FBS and 80% 1× DMEM. The primer pairs that were used for genotyping were the following: *mEGFP-Srsf2*: forward, 5′-TTTGGCGGGCTTTCTAACTGC-3′; reverse, 5′-CGGTAGGTCAGGTTGTCCAC-3′. *Halo-Polr2a*: forward, 5′-GAGC CCTAGCGTCAACAACT-3′; reverse, 5′-CCTCTGGTATCAGCTCCCCT-3′. A clone with heterozygous mEGFP–SRSF2 was subsequently passaged for all assays. A clone of homozygous Halo-RPB1 and heterozygous mEGFP–SRSF2 were used for live-cell lattice light-sheet imaging.

Live-cell imaging of GFP–SRSF2 cell line. Cells were grown on glass dishes (Mattek P35G-1.5-20-C) coated with 5 μg/ml of poly-L-ornithine (Sigma-Aldrich, P4957) for 30 min at 37 °C, and with 5 μg/ml of laminin (Corning, 354232) for 2–16 h at 37 °C. Before imaging the cells, culture medium was replaced with phenol red-free 2i medium, and imaged using the Andor Revolution Spinning Disk Confocal microscope. Raw Andor images were processed using FIJI or ImageJ.

FRAP in live cells. FRAP was performed on an Andor Revolution Spinning Disk Confocal microscope with 488-nm laser. Bleaching was performed using 100% laser power with 30 microseconds dwell time for 5 cycles, and images were collected every 500 ms. Fluorescence intensity at the bleached spot, a control unbleached spot and background was measured using the FIJI plugin FRAP Profiler. Background intensity was subtracted, and values are reported relative to the unbleached spot to control for photobleaching during image acquisition.

Immunofluorescence with DNA FISH. Immunofluorescence was performed as described in 'Immunofluorescence with RNA FISH'. After incubating the cells with the secondary antibodies, cells were washed 3 times in PBS for 5 min at room temperature, fixed with 4% PFA in PBS for 10 min and washed 3 times in PBS. Cells were incubated in 70% ethanol, 85% ethanol and then 100% ethanol for 1 min at room temperature. Probe hybridization mixture was made by mixing 7 ml of FISH Hybridization Buffer (Agilent G9400A), 1 ml of FISH probe (see below for the region) and 2 ml of water. Five millilitres of mixture was added on a slide and the coverslip was placed on top (cell-side towards the hybridization mixture). Coverslips were sealed using rubber cement. Once the rubber cement solidified, genomic DNA and probes were denatured at 78 °C for 5 min and slides were incubated at 37 °C in the dark overnight. The coverslip was removed from the slide and incubated in pre-warmed wash buffer 1 (Agilent G9401A) at 73 °C for 2 min, and in wash buffer 2 (Agilent, G9402A) for 1 min at room temperature. Slides were air-dried and nuclei were stained with Hoechst in PBS for 5 min at room temperature. Coverslips were washed three times in PBS, mounted on slides using Vectashield and sealed with nail polish. Images were acquired on an RPI Spinning Disk confocal microscope with a 100× objective using MetaMorph acquisition software and a Hammamatsu ORCA-ER CCD camera (W.M. Keck Microscopy Facility). Images were post-processed using FIJI. DNA FISH probes were custom-designed and generated by Agilent to target *Nanog* and *Trim28* super-enhancers.

For *Nanog*, the design input region was mm9, chr6 122605249–122705248, and the design region was mm9, chr6: 122605985–122705394. For *Trim28*, the design input region was mm9, chr7:13551990–13651989, and the design region was mm9, chr7:13552132–13651971.

Drug treatments. V6.5 mouse ES cells were grown in 24-well plates on coated-glass coverslips, as described in 'Immunofluorescence with RNA FISH'. Cells were treated with drugs or vehicle (pladienolide B at 100 nM, DRB at 100 μM or DMSO at 0.1%), followed by fixation with 4% PFA in PBS. For DRB-washout experiments, cells were treated with DRB for 2 h, gently washed twice with fresh 2i medium and left to recover in 2i medium containing 0.1% DMSO for 2 h before fixation.

Co-immunofluorescence co-localization analysis. For analysis of co-localization data (Fig. 3a), custom Python scripts were written to process and analyse 3D image data gathered in immunofluorescence and DAPI channels. Nuclei were detected by Otsu thresholding, and a mask of nuclei was applied to the

immunofluorescence channels. Manual minimal thresholds were called for the immunofluorescence channels. Manders' coefficients were then calculated for masked immunofluorescence channels A and B in 3D, using with the following formulas: $M1 = I_A[I_B > 0]/\sum I_A$ and $M2 = I_B[I_A > 0]/\sum I_B$ (see ref. [31]). Manders' coefficients were calculated for at least three images per sample, and then averaged. To generate the heat maps of co-localization in Fig. 3a, a representative *z*-slice was selected for each dataset. Each channel was standardized by subtracting the mean and dividing by the s.d. of the slice to generate a per-pixel *z*-score. The *z*-scores between channels were then multiplied, and a heat map was generated from multiplied *z*-scores using the Python package matplotlib with the magma heat map. All heat maps are displayed on the same scale ($v_{max} = 70$). This method highlights areas in which both channels have overlapping pixels.

Protein purification. Human cDNA was cloned into a modified version of a T7 pET expression vector. The base vector was engineered to include a 5′ 6×His tag followed by either mEGFP or mCherry and a 14 -amino-acid linker sequence (GAPGSAGSAAGGSG). NEBuilder HiFi DNA Assembly Master Mix (NEB E2621S) was used to insert these sequences (generated by PCR) in-frame with the linker amino acids. For MED1-IDR, the inserted sequence encodes residues 948 to 1574 of the full-length MED1 protein. Vector expressing mEGFP alone contains the linker sequence followed by a stop codon. All expression constructs were sequenced to ensure sequence identity.

For protein expression, plasmids were transformed into LOBSTR cells (a gift of Cheeseman Laboratory). A fresh bacterial colony was inoculated into LB medium containing kanamycin and chloramphenicol, and grown overnight at 37 °C. Cells containing Pol II CTD constructs were diluted 1:30 in 500 ml of room-temperature LB medium, with freshly added kanamycin and chloramphenicol, and grown for 1.5 h at 16 °C. IPTG was added to 1 mM, and growth continued for 20 h. Cells were collected and stored frozen at −80 °C. Cells containing all other constructs were treated in a similar manner, except they were grown for 5 h at 37 °C after IPTG induction.

For wild-type (GFP–CTD52, full-length CTD with 52 heptapeptide repeats) and mutant (GFP–CTD26, C-terminal 26 repeats; and GFP–CTD10, C-terminal 10 repeats) CTD and HNRNPA1, pellets of 500 ml of cells were resuspended in 15 ml of buffer A (50 mM Tris pH 7.5, 500 mM NaCl) with cOmplete protease inhibitors (Roche, 11873580001), and sonicated (ten cycles of 15-s on, 60-s off). The lysates were cleared by centrifugation at 12,000*g* for 30 min and added to 1 ml of Ni-NTA agarose (Invitrogen, R901-15) that had been pre-equilibrated with 10 volumes of the same buffer, and rotated at 4 °C for 1.5 h. The slurry was centrifuged at 3,000 r.p.m. for 10 min in a Thermo Legend XTR swinging bucket rotor. The resin pellets were washed twice with 5 ml of buffer A, followed by centrifugation as above. Protein was eluted 3 times with 2 ml of buffer A, plus 250 mM imidazole. For each cycle, the elution buffer was added and rotated at least 10 min at 4 °C, and centrifuged as above. Eluates were analysed on a 12% acrylamide gel stained with Coomassie. Fractions containing protein of the expected size were pooled, diluted 1:1 with the 250 mM imidazole buffer and dialysed against two changes of buffer containing 50 mM Tris 7.5, 125 mM NaCl, 10% glycerol and 1 mM DTT at 4 °C. Protein concentration was measured by Thermo BCA Protein Assay Kit – Reducing Agent Compatible.

For SRSF2, SRSF1 and U2AF2, pellets of 500 ml of cells were resuspended in 15 ml of denaturing buffer (50 mM Tris 7.5, 300 mM NaCl, 10 mM imidazole, 8 M Urea) with cOmplete protease inhibitors, sonicated and then cleared as above. The lysates were added to 1 ml of Ni-NTA agarose that had been pre-equilibrated with 10 volumes of the same denaturing buffer. Tubes containing this agarose lysate slurry were rotated for 1.5 h at room temperature. The slurry was centrifuged and washed twice, as above. Protein was eluted with 3× 2 ml of denaturing buffer containing 250 mM imidazole. Fractions containing protein of the expected size were diluted 1:1 and dialysed versus 50 mM Tris pH 7.5, 500 mM NaCl, 1 mM DTT with 4 M urea, followed by the same buffer containing 2 M urea and, finally, 2 changes of buffer with 10% glycerol and no urea. Any precipitate after dialysis was removed by centrifugation at 3,000 r.p.m. for 10 min, and concentration was determined by BCA assay.

Purification of mediator. The mediator samples were purified as previously described[32], with modifications. Before affinity purification, the P0.5M/QFT fraction was concentrated to 12 mg/ml by ammonium sulfate precipitation (35%). The pellet was resuspended in pH 7.9 buffer containing 20 mM KCl, 20 mM HEPES, 0.1 mM EDTA, 2 mM $MgCl_2$ and 20% glycerol, and then dialysed against pH 7.9 buffer containing 0.15 M KCl, 20 mM HEPES, 0.1 mM EDTA, 20% glycerol and 0.02% NP-40 before the affinity purification step. Affinity purification was carried out as previously described[32], eluted material was loaded onto a 2.2-ml centrifuge tube containing 2 ml 0.15 M KCl HEMG (20 mM HEPES, 0.1 mM EDTA, 2 mM $MgCl_2$ and 10% glycerol) and centrifuged at 50,000 r.p.m. for 4 h at 4 °C. This served to remove excess free GST–VP16 and to concentrate the mediator in the final fraction. Before droplet assays, purified mediator was further concentrated using a Microcon-30-kDa Centrifugal Filter Unit with Ultracel-30 membrane (Millipore MRCF0R030) to reach about 650 nM of mediator complex. Concentrated mediator was added to the droplet assay to a final concentration of about 200 nM with 10 μM indicated GFP-tagged protein. Droplet reactions contained 16% Ficoll-400 and 140 mM salt.

ChIP–seq. Mouse ES cells were grown to 80% confluence in 2i medium. About 1% formaldehyde in PBS was used for crosslinking of cells for 15 min, followed by quenching with glycine at a final concentration of 125 mM on ice. Cells were washed with cold PBS and collected by scraping cells in cold PBS. Collected cells were pelleted at 1,000*g* for 3 min at 4 °C, flash-frozen in liquid nitrogen and stored at −80 °C. All buffers contained freshly prepared cOmplete protease inhibitors (Roche, 11873580001). For ChIP using phospho-specific antibodies, all buffers contained freshly prepared PhosSTOP phosphatase inhibitor cocktail (Roche, 4906837001). Frozen crosslinked cells were thawed on ice and then resuspended in LB1 (50 mM HEPES–KOH, pH 7.9, 140 mM NaCl, 1 mM EDTA 0.5 ml 0.5 M, 10% glycerol, 0.5% NP-40, 1% Triton X-100, 1× protease inhibitor) and incubated for 20 min rotating at 4 °C. Cells were pelleted for 5 min at 1,350*g*, resuspended in LB2 (10 mM Tris pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1× protease inhibitor) and incubated for 5 min rotating at 4 °C. Pellets were resuspended in LB3 (10 mM Tris pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% sodium deoxycholate, 0.5% sodium lauroyl sarcosinate, 1% Triton X-100, 1× protease inhibitor) at a concentration of 30–50 million cells per millilitre. Cells were sonicated using Covaris S220 for 12 min (duty cycle: 5%, intensity: 4, cycles per burst: 200). Sonicated material was clarified by spinning at 20,000*g* for 30 min at 4 °C. The supernatant is the soluble chromatin used for the ChIP. Dynabeads, pre-blocked with 0.5% bovine serum albumin, were incubated with indicated antibodies for 2 h. Chromatin was added to antibody–bead complex and incubated rotating overnight at 4 °C. Beads were washed 3 times each with wash buffer 1 (50 mM HEPES pH 7.5, 500 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Triton, 0.1% NaDoc, 0.1% SDS) and wash buffer 2 (20 mM Tris pH 8, 1 mM EDTA, 250 mM LiCl, 0.5% NP-40, 0.5% NaDoc) at 4 °C, followed by washing once with TE at room temperature. Chromatin was eluted by adding elution buffer (50 mM, Tris pH 8.0, 10 mM EDTA, 1% sodium dodecyl sulfate) to the beads and incubated with shaking at 60 °C for 30 min. Reversal of crosslinking was performed overnight at 58 °C. RNaseA was added and incubated for 1 h at 50 °C for RNA removal. Proteinase K was added and incubated for 1 h at 60 °C for protein removal. DNA was purified using Qiagen PCR purification kit, as per the manufacturer's instructions, and eluted in 50 μl 10 mM Tris-HCl, pH 8.5, which was used for quantification and ChIP library preparation. ChIP libraries were prepared with the Swift Biosciences Accel-NGS 2S Plus DNA Library Kit, according to the kit instructions with an additional size-selection step on the PippinHT system from Sage Science. Following library preparation, ChIP libraries were run on a 2% gel on the PippinHT with a size-collection window of 200–600 bases. Final libraries were quantified by qPCR with the KAPA Library Quantification kit from Roche, and sequenced in single-read mode for 40 bases on an Illumina HiSeq 2500.

ChIP–seq data were aligned to the mm9 version of the mouse reference genome using bowtie, with parameters –k 1 –m 1 –best and –l set to read length. Wiggle files for the display of read coverage in bins were created using MACS, with parameters –w –S – space = 50 –nomodel –shiftsize = 200, and read counts per bin were normalized to the millions of mapped reads used to make the wiggle file. Reads-per-million-normalized wiggle files were displayed in the UCSC genome browser. Metagene plots were made using ngs.plot[33] (v.2.61) using default parameters. The top 20% of expressed genes were calculated from a published RNA-seq dataset (GSE112807)[8].

SRSF2 and serine-2-phosphorylated Pol II ChIP–seq were generated in this study using antibodies against SRSF2 (Abcam ab11826) and Pol II Ser2-phospho-CTD (Millipore 04-1571). MED1 and total Pol II ChIP–seq have previously been published (GSE112808)[8].

RNA FISH average image analysis. For analysis of RNA FISH with immunofluorescence, custom Python scripts were written to process and analyse 3D image data gathered in FISH and immunofluorescence channels. Nuclear stains were blurred with a Gaussian filter ($\sigma = 2.0$), maximally projected in the *z* plane and clustered into two clusters (nuclei and background) by *K*-means clustering. FISH foci were either manually called with ImageJ or automatically called using the scipy ndimage package. For automatic detection, an intensity threshold (mean + 3 × s.d.) was applied to the FISH channel. The ndimage find_objects function was then used to call contiguous FISH foci in 3D. These FISH foci were then filtered by various criteria, including size (minimum 100 voxels), circularity of a maximum *z*-projection $\left(\text{circularity} = 4\pi \cdot \frac{\text{area}}{\text{perimeter}^2}\,;\, 0.7\right)$, and being present in a nucleus (determined by nuclear mask, described in 'Co-immunofluorescence co-localization analysis'). For manual calling, FISH foci were identified in maximum *z*-projections of the FISH channel, and the *x* and *y* coordinates were used as reference points to guide the automatic detection described above. The FISH foci were then centred in a 3D box (length size (ℓ) = 3.0 μm). The immunofluorescence

signals centred at FISH foci for each FISH and immunofluorescence pair were then combined, and an average intensity projection was calculated, providing averaged data for immunofluorescence signal intensity within a $l \times l$ square centred at FISH foci. As a control, this same process was carried out for immunofluorescence signals centred at an equal number of randomly selected nuclear positions. These average-intensity projections were then used to generate 2D contour maps of the signal intensity. Contour plots were generated using the matplotlib Python package. For the contour plots, the intensity-colour ranges presented were customized across a linear range of colours ($n! = 15$). For the FISH channel, black to magenta was used. For the immunofluorescence channel, we used chroma.js (an online colour generator) to generate colours across 15 bins, with the key transition colours chosen as black, blue–violet, medium blue and lime. This was done to ensure that the reader's eye could more-readily detect the contrast in signal. The generated colour map was used in 15 evenly spaced intensity bins for all immunofluorescence plots. The averaged immunofluorescence, centred at FISH or at randomly selected nuclear locations, is plotted using the same colour scale, set to include the minimum and maximum signal from each plot.

Lattice light-sheet microscopy. For lattice light-sheet microscopy cells were plated on a coated coverslip 24 h before imaging. Before imaging, cells were incubated for 20 min with 250 nM Halo–JF646 ligand[34] and washed in growth medium for 20 min. Dual-colour stacks were acquired with 100-ms exposure time and 340-nm effective z-spacing. Light sheet data were processed (deskewed, deconvolved, and corrected for chromatic aberration) using LLSpy[35] and analysed using custom MATLAB scripts. Foci were localized in 3D following a two-step procedure. First, background was subtracted by subtracting a median filtered image from each slice in a z-stack, and intensity peaks were detected using the MTT algorithm[36]. Foci were then identified as peaks that were found in at least 4 subsequent z-slices within a 100-nm radius in x–y. A 3D Gaussian peak function was fitted to the intensity distribution to obtain 3D centre coordinates for foci of SRSF2 and Pol II. Detection of foci was performed on deconvolved, background-subtracted data, whereas subsequent quantitative analysis of foci brightness was performed on deskewed, background-subtracted data. To estimate the number of fluorescently tagged proteins in foci, cells were fixed in 4% PFA for 10 min, washed 3× in PBS, and imaged on the lattice light-sheet microscope until they were almost entirely bleached. A single plane was then imaged continuously for 1,000 frames to detect single emitter signals. The apparent brightness of Halo–JF646 ($n = 204$) or GFP ($n = 236$) single emitters was determined by calculating the integrated intensity above background[37] (Extended Data Fig. 2b). We note that signals are close to the noise floor of the camera, and the integrated intensity measure can sometimes (in about 10% of the cases) yield negative values when nearby emitters lead to overestimation of the local background intensity. For the subsequent analysis, we excluded emitters with negative integrated intensity measures. We used the mean estimates obtained for the remaining single emitters to normalize values obtained by the same metric for foci of SRSF2–GFP and Halo–JF646–Pol II in the z slice closest to the 3D centre position, taking into account the higher laser-power densities that were used to detect single emitters. To assess co-localization, we paired Pol II–Halo foci with their nearest neighbour in 3D space. We note that axial resolution of the imaging method is considerably lower than x–y resolution. We therefore call co-localized only those foci with centre coordinates separated by less than the optical resolution of 300 nm laterally (x–y) and 900 nm axially (z).

Splicing reporter assay. The splicing reporter assay was performed as previously described[38]. In brief, mouse ES cells were transfected with a plasmid encoding luciferase with an intervening intron (Addgene 62858) or a plasmid encoding luciferase with no intervening intron (Addgene 62857). Cells were then treated with DMSO or 100 nM pladienolide B for 4 h, at which point they were lysed and assayed for luciferase activity. Relative splicing levels in each condition were determined by normalizing the luciferase activity detected in cells transfected with the intron-containing plasmid to the luciferase activity detected in cells transfected with the intronless plasmid.

FISH–immunofluorescence overlap analysis. DNA FISH spots were identified as described in 'RNA FISH average image analysis'. Images of the spots with the corresponding immunofluorescence channel for all conditions were randomized and blindly scored for FISH–immunofluorescence overlap (at least 25% of the FISH spot overlapping with an immunofluorescence puncta) or no overlap. Overlap scores for each condition were then tallied and compared. For presentation, the FISH foci overlap with immunofluorescence was indexed with the DMSO condition set to 1.

In vitro droplet assay. Recombinant GFP or mCherry fusion proteins were concentrated and desalted to an appropriate protein concentration and 125–500 mM NaCl using Amicon Ultra centrifugal filters (30K MWCO, Millipore). Recombinant proteins were added to solutions at varying concentrations with 120–125 mM final salt and 16% Ficoll-400 or 10% PEG-8000 as crowding agent in droplet-formation buffer (50 mM Tris-HCl pH 7.5, 10% glycerol, 1 mM DTT), as described in the figure legends. The protein solution was incubated for 1 h and loaded onto a chamber made in house, comprising a glass slide with a coverslip attached by two parallel strips of double-sided tape. Slides were then imaged with the Andor confocal microscope with a 150× objective. Unless indicated, the images presented are of droplets settled on the glass coverslip. For FRAP of in vitro droplets, 2 pulses of laser (20% power) at a 20-μs dwell time were applied to the droplet, and recovery was imaged on the Andor microscope every 1 s for the indicated time periods. For CDK7- or CDK9-mediated phosphorylation of the Pol II CTD, commercially available active CDK7–MAT1–CCNH (CAK complex; Millipore 14-476) or CDK9 and cyclin T1 (Millipore 14-685) was used to phosphorylate GFP–CTD52 in kinase reaction buffer (20 mM MOPs–NaOH pH 7.0, 1 mM EDTA, 0.001% NP-40, 2.5% glycerol, 0.05% β-mercaptoethanol, 10 mM MgAc, 10 μM ATP) at room temperature for 3 h. The CTD to enzyme ratio is about 1 μM CTD to about 5 ng/μl CDK7 or CDK9.

In vitro droplet quantification. To analyse in vitro droplet experiments, custom Python scripts using the scikit-image package were written to identify droplets and characterize their size, shape and intensity. Droplets were segmented from average images of captured channels on various criteria: (1) an intensity threshold that was three s.d. above the mean of the image; (2) size thresholds (20 pixel minimum droplet size); and (3) a minimum circularity $\left(\text{circularity} = 4\pi \cdot \frac{\text{area}}{\text{perimeter}^2}\right)$ of 0.8 (1 being a perfect circle). After segmentation, mean intensity for each droplet was calculated while excluding pixels near the phase interface, and background-corrected by subtracting intensity of dark images of droplet formation buffer only (see ref. [39]). Hundreds of droplets identified in (typically) ten independent fields of view were quantified. The mean intensity within the droplets (C-in) and in the bulk (C-out) were calculated for each channel. The partition ratio was computed as (C-in)/(C-out). The box plots show the distributions of all droplets. Each dot represents an individual droplet. The measured datasets for partition ratio versus the protein concentration in Extended Data Fig. 7b were fitted by the logistic equation (see ref. [40]):

$$f = \frac{a}{1 + e^{-\frac{(x - x_0)}{b}}}$$

in which f is the partition ratio, and x is the corresponding protein concentration.

Statistics and reproducibility. For all immunofluorescence and FISH experiments, one coverslip of cells was stained for the indicated factors and at least eight independent imaging fields were acquired, which typically contained 50–200 FISH foci. The exact number of FISH foci analysed, and the fraction of these foci that overlapped with immunofluorescence puncta, and relevant comparative statistics for experiments for which these comparisons were made, are as follows. For Fig. 1b and Extended Data Fig. 1a, 86 *Nanog* foci and 131 *Trim28* foci; for Fig. 2a, *Nanog* FISH foci counts for each immunofluorescence experiment were 97 (SRSF2), 122 (SF3B1), 74 (U2AF2), 88 (HNRNPA1), 109 (SRSF1), 137 (SRRM1), 103 (PRPF8) and 119 (SNRNP70).

For Fig. 2c and Extended Data Fig. 3b, c, the numbers of overlapped foci out of the total number of foci for each factor in the DMSO and pladienolide B condition, and the P values associated with each DMSO versus pladienolide B comparison, are as follows. For SRSF2, 31/61 (solidus denotes number of overlapped foci/total number of foci throughout), 19/125, $P < 0.0001$; for SF3B1, 29/61, 30/126, $P = 0.0014$; and Pol II, 16/71, 15/65, $P > 0.9999$.

For 'co-loc' analysis in Fig. 3a, one coverslip of cells was stained for the indicated factors and five independent fields were imaged and analysed for each experimental comparison.

For Fig. 3c, d and Extended Data Fig. 5a–d, the numbers of overlapped foci out of the total number of foci for each factor in the DMSO, DRB and washout condition, and the P values associated with the DMSO versus DRB and DRB versus washout comparisons, are as follows. For SRSF2 and *Nanog*, 40/91 (solidus denotes number of overlapped foci/total number of foci throughout), 11/146 and 19/78, $P < 0.0001$ and $P = 0.0008$; for Pol II and *Nanog*, 33/114, 23/92 and 36/160, $P = 0.6368$ and 0.6467; for MED1 and *Nanog*, 28/89, 32/133 and 27/84, $P = 0.2804$ and 0.2122; for SRSF2 and *Trim28*, 26/71, 5/111 and 12/92, $P < 0.0001$ and $P = 0.0403$; for SRSF1 and *Trim28*, 19/36, 12/55 and 22/73, $P = 0.0347$ and 0.3189; for SF3B1 and *Trim28*, 48/91, 14/63 and 27/99, $P = 0.0002$ and 0.5788; for U2AF2 and *Trim28*, 21/42, 17/50 and 27/78, $P = 0.1406$ and $P > 0.9999$; for PRPF8 and *Trim28*, 15/80, 6/69 and 17/102, $P = 0.0996$ and 0.1721; for MED1 and *Trim28*, 19/73, 30/96 and 21/91, $P = 0.4971$ and 0.251; and for Pol II and *Trim28*, 25/102, 22/93 and 21/78, $P > 0.9999$ and $P = 0.7238$).

For Extended Data Fig. 2a, *Trim28* FISH foci counts for each immunofluorescence experiment were 115 (SRSF2), 151 (SF3B1), 104 (U2AF2), 90 (HNRNPA1), 145 (SRSF1), 127 (SRRM1), 175 (PRPF8) and 157 (SNRNP70).

For lattice light-sheet imaging, the number of images acquired and data points plotted are as follows. For Fig. 2b, the full image dataset comprises 102 cells from 10 independent fields of view.

For Extended Data Fig. 2b, 288 colocalized condensates were plotted in the scatter plot.

For all in vitro droplet experiments, one slide of droplet mix was imaged and at least 7 independent fields of view were acquired, which typically contained about 100–1,000 droplets. In all cases in which the droplet data are quantified and displayed as a box plot, the box in the boxplot extends from the 25th to 75th percentiles, the line in the middle of the box is plotted at the median, the whiskers represent the range within 1.5× interquartile, and each dot represents an individual droplet.

The exact number of fields and droplets analysed are as follows: in Fig. 1c, 10/540 (GFP; solidus denotes number of fields/droplets throughout), 7/842 (GFP–CTD52), 10/879 (GFP–CTD26) and 10/1,293 (GFP–CTD10); in Fig. 1d, 11/159 (GFP), 10/175 (GFP–CTD52), 11/207 (GFP–CTD26) and 10/206 (GFP–CTD10); in Fig. 4a, 10/114 (GFP), 7/420 (CTD), 8/342 (p-CTD(CDK7)) and 8/464 (p-CTD(CDK9)); in Fig. 4b, 10/385 (GFP), 10/328 (CTD), 10/294 (p-CTD(CDK7)) and 12/361 (p-CTD(CDK9)); in Fig. 4c, 10/400 (GFP), 10/369 (CTD), 10/314 (p-CTD(CDK7)) and 15/513 (p-CTD(CDK9)); in Extended Data Fig. 1b, one fusion event was captured from one field; in Extended Data Fig. 6c, d, with 16% Ficoll, 21/231 (GFP), 21/289 (CTD), 10/134 (p-CTD(CDK7)), 11/144 (p-CTD(CDK9)); with 10% PEG, 24/147 (GFP), 21/227 (CTD), 10/106 (p-CTD(CDK7)) and 10/83 (p-CTD(CDK9)); in Extended Data Fig. 7a, b, the number of fields acquired are as follows: for SRSF2, 10 (5 μM), 10 (2.5 μM), 10 (1.25 μM), 10 (0.625 μM), 10 (0.313 μM), 11 (0.156 μM), 12 (0.078 μM), 12 (0.039 μM), 10 (0.0195 μM), 10 (0.0098 μM), 10 (0.0049 μM), 10 (0.0024 μM), 10 (0.0012 μM) and 10 (0.0006 μM); for SRSF1, 10 (5 μM), 10 (2.5 μM), 10 (1.25 μM), 10 (0.625 μM), 10 (0.313 μM), 10 (0.156 μM), 10 (0.078 μM), 10 (0.039 μM), 10 (0.0195 μM), 10 (0.0098 μM), 10 (0.0049 μM), 11 (0.0024 μM), 11 (0.0012 μM) and 10 (0.0006 μM); for U2AF2, 10 for each sample; for HNRNPA1, 10 for each sample; and for mCherry, 10 for each sample; for Extended Data Fig. 7c, 10 independent fields of view were acquired for each sample; for Extended Data Fig. 8a, b, same as Fig. 4b, c; for Extended Data Fig. 8c, 10/365 (GFP), 10/321 (CTD), 10/325 (p-CTD(CDK7)) and 10/313 (p-CTD(CDK9)); and for Extended Data Fig. 8d, 10/423 (GFP), 11/437 (CTD), 10/412 (p-CTD(CDK7)) and 10/381 (p-CTD(CDK9)).

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

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

Data availability

Datasets generated in this study have been deposited in the Gene Expression Omnibus under accession number GSE120656. Uncropped gel images can be found in Supplementary Fig. 1.

Code availability

All custom code used in this study is available upon request.

31. Bolte, S. & Cordelières, F. P. A guided tour into subcellular colocalization analysis in light microscopy. *J. Microsc.* **224**, 213–232 (2006).
32. Meyer, K. D. et al. Cooperative activity of cdk8 and GCN5L within Mediator directs tandem phosphoacetylation of histone H3. *EMBO J.* **27**, 1447–1457 (2008).
33. Shen, L., Shao, N., Liu, X. & Nestler, E. ngs.plot: quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* **15**, 284 (2014).
34. Grimm, J. B. et al. A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nat. Methods* **12**, 244–250 (2015).
35. Lambert, T. & Shao, L. tlambert03/LLSpy v.0.3.8. https://zenodo.org/record/1204615#.XTB-miJKhaQ (2018).
36. Sergé, A., Bertaux, N., Rigneault, H. & Marguet, D. Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes. *Nat. Methods* **5**, 687–694 (2008).
37. Verdaasdonk, J. S., Lawrimore, J. & Bloom, K. Determining absolute protein numbers by quantitative fluorescence microscopy. *Methods Cell Biol.* **123**, 347–365 (2014).
38. Younis, I. et al. Rapid-response splicing reporter screens identify differential regulators of constitutive and alternative splicing. *Mol. Cell. Biol.* **30**, 1718–1728 (2010).
39. Banani, S. F. et al. Compositional control of phase-separated cellular bodies. *Cell* **166**, 651–663 (2016).
40. Wang, J. et al. A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* **174**, 688–699 (2018).

Acknowledgements We thank I. A. Klein, J. Schuijers, C. H. Li, E. L. Coffey and other members in the Young laboratory for helpful discussions, W. Salmon of the W. M. Keck Microscopy Facility, T. Volkert, J. Love and S. Gupta of the Whitehead Genomics Core facility for technical assistance, M. Stubna for help with droplet and biochemical assays and S. L. McKnight, I. Kwon and M. Kato for CTD constructs. The work was supported by NIH grant GM123511 (R.A.Y.), NSF grant PHY1743900 (R.A.Y. and P.A.S.), NIH GM117370 and GM110064 (D.J.T.), NIH grant R01-GM034277 (P.A.S.), Cancer Research Institute Irvington Fellowship (Y.E.G.), Damon Runyon Cancer Research Foundation Fellowship (2309-17) (B.R.S.), Hope Funds for Cancer Research fellowship (B.J.A.), Swedish Research Council Postdoctoral Fellowship (VR 2017-00372) (A.B.), the German Research Foundation DFG Postdoctoral Fellowship SP 1680/1-1 (J.-H.S.), DE 3069/1-1 (T.-M.D.), NIH T32 GM008759 (C.B.F.) and funds from Novo Nordisk (R.A.Y. and P.A.S.).

Author contributions Y.E.G. and R.A.Y. conceived the project. Y.E.G., J.C.M. and R.A.Y. organized the studies and wrote the manuscript. Y.E.G. and B.R.S. performed in vitro droplet formation assays. J.C.M., A.D. and A.B. performed immunofluorescence experiments. J.E.H. and K.S. developed and performed computational analyses. B.R.S., Y.E.G, B.J.A. and J.C.M. performed ChIP and analysed data. N.M.H. purified recombinant proteins. Y.E.G., J.E.H. and L.K.A. generated cell lines. J.-H.S. and A.V.Z. performed lattice light-sheet microscopy and analysis. T.-M.D., J.K.R., C.B.F. and D.J.T. purified human mediator. Y.E.G., J.C.M. and J.E.H. generated constructs. J.E.H. performed live cell imaging. J.E.H. and T.I.L. contributed to writing the manuscript. P.A.S. and I.I.C. provided input into experimental design and interpretation. R.A.Y. supervised the project with the help from T.I.L. All authors contributed to editing the manuscript.

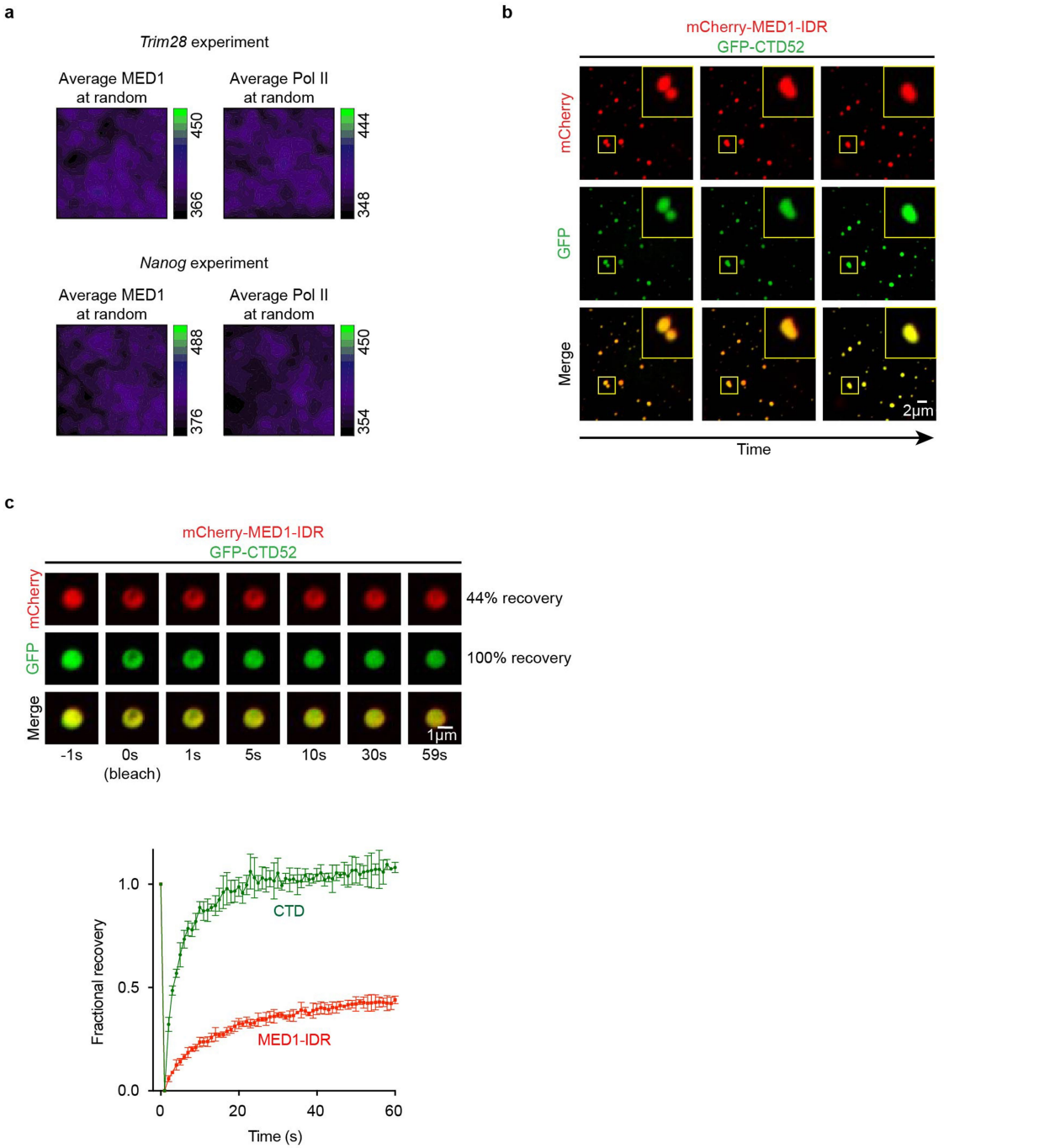
Competing interests R.A.Y. is a founder and shareholder of Syros Pharmaceuticals, Camp4 Therapeutics, Omega Therapeutics and Dewpoint Therapeutics. P.A.S. is a member of the board and shareholder in Syros and a member of the Scientific Advisory Board of Dewpoint. B.J.A. and T.I.L. are shareholders of Syros Pharmaceuticals. T.I.L. is a consultant to Camp4 Therapeutics and I.I.C. is a consultant to Dewpoint Therapeutics. All other authors declare no competing interests.

Additional information

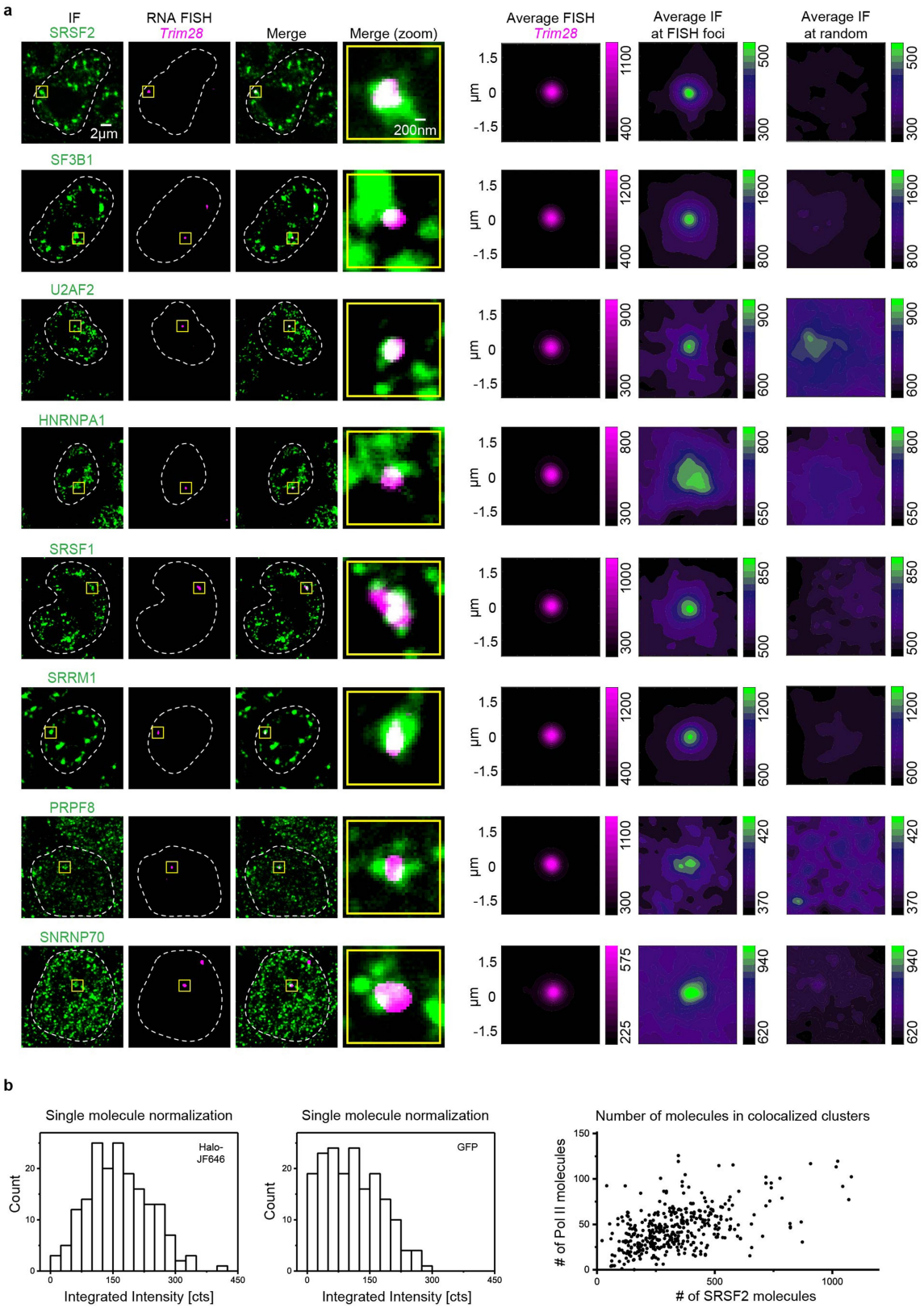
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1464-0.

Correspondence and requests for materials should be addressed to R.A.Y.

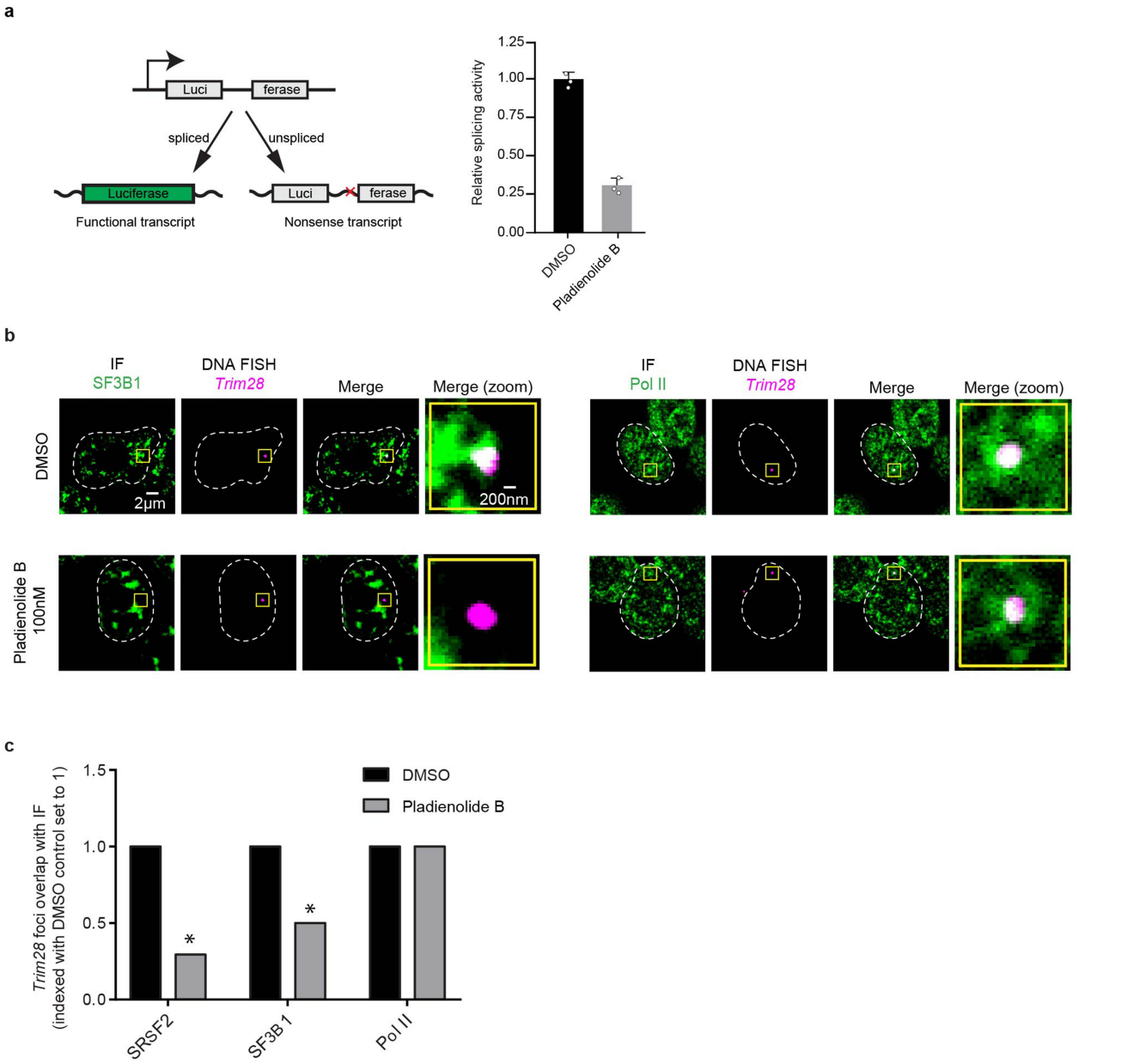
Peer review information *Nature* thanks Robert Fisher, Richard Kriwacki and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



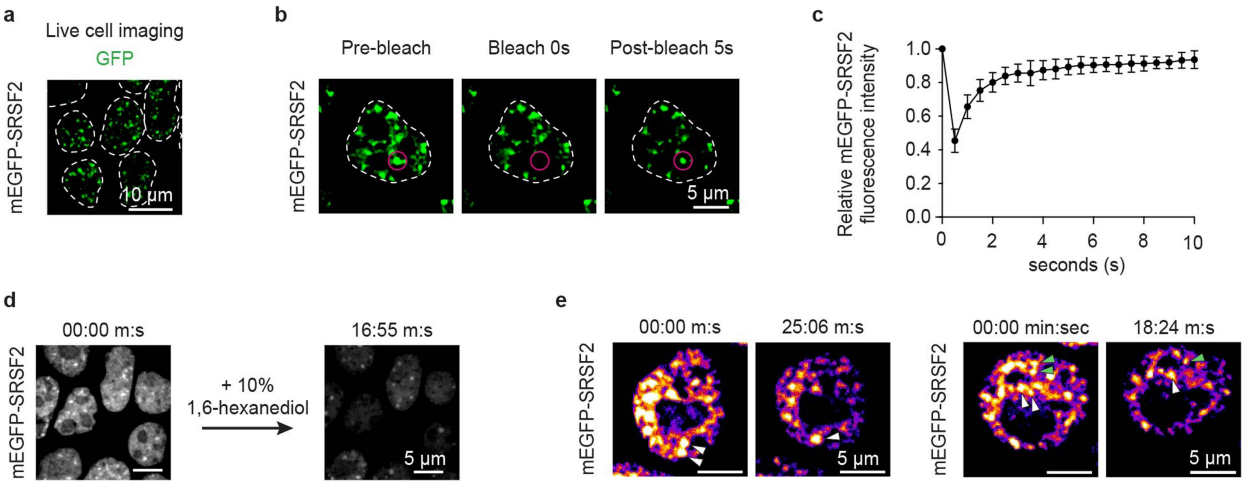
Extended Data Fig. 1 | CTD partitioning in mediator condensates. **a**, Average immunofluorescence signal for MED1 and Pol II centred on randomly selected nuclear positions. **b**, Images of a fusion event between two full-length CTD–MED1-IDR droplets. GFP–CTD52 at 10 μM was mixed with 10μM mCherry–MED1-IDR in droplet-formation buffer with 125 mM NaCl and 16% Ficoll-400. **c**, Top, representative images of FRAP of heterotypic droplets of mCherry–MED1-IDR and GFP–CTD52. Droplet-formation conditions are the same as in **b**. Bottom, quantification of the fraction of FRAP of mCherry–MED1-IDR and GFP–CTD52. Data are mean ± s.d. ($n = 3$ experiments).



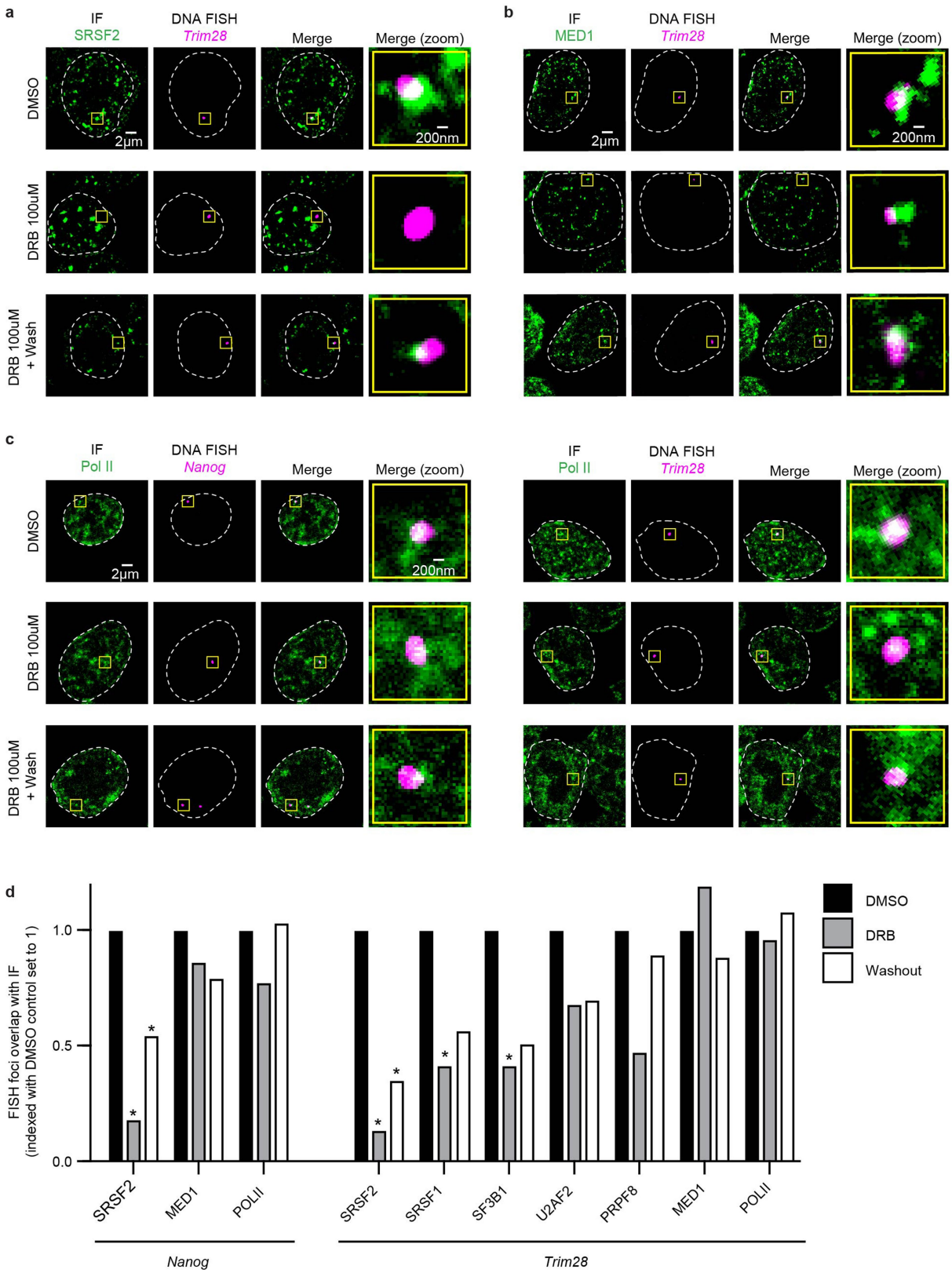
Extended Data Fig. 2 | Splicing factors at *Trim28*, and quantification of SRSF2 and Pol II molecules. a, Representative images exhibiting overlap between immunofluorescence of splicing factors SRSF2, SF3B1, U2AF2, HNRNPA1, SRSF1, SRRM1, PRPF8, or SNRNP70 with nascent RNA FISH of *Trim28* in fixed mouse ES cells. **b**, Left, histogram of the integrated intensity of single Halo–JF646 ($n = 178$) and single GFP emitters ($n = 177$). Mean values of 164.8 ± 5.6 counts (mean ± s.e.m.) and 108.6 ± 5.1 counts (mean ± s.e.m.) were used to normalize the integrated intensity of Pol II–Halo–JF646 and SRSF2–GFP, respectively. Right, scatter plot depicting the estimated numbers of Pol II and SRSF2 molecules in colocalizing Pol II and SRSF2 puncta (see Methods).



Extended Data Fig. 3 | Splicing inhibition and splicing-factor condensates. a, Left, diagram depicting the splicing reporter used to measure splicing inhibition after treatment with pladienolide B. Right, relative levels of splicing in cells treated with DMSO versus cells treated with pladienolide B for 4 h. The mean of three biological replicates (each replicate shown as dot) with s.d. is plotted (see Methods). **b**, Representative images exhibiting overlap or absence of overlap between immunofluorescence of SF3B1 or Pol II and DNA FISH of *Trim28* in mouse ES cells treated with either DMSO or pladienolide B for 4 h. **c**, Fraction of overlap of FISH foci with immunofluorescence puncta in cells treated with pladienolide B relative to cells treated with DMSO. The asterisk above the drug-treated bar indicates $*P < 0.05$ from a two-tailed χ^2 test comparing the number of overlapping and non-overlapping FISH spots in the DMSO versus the drug condition (see Methods).

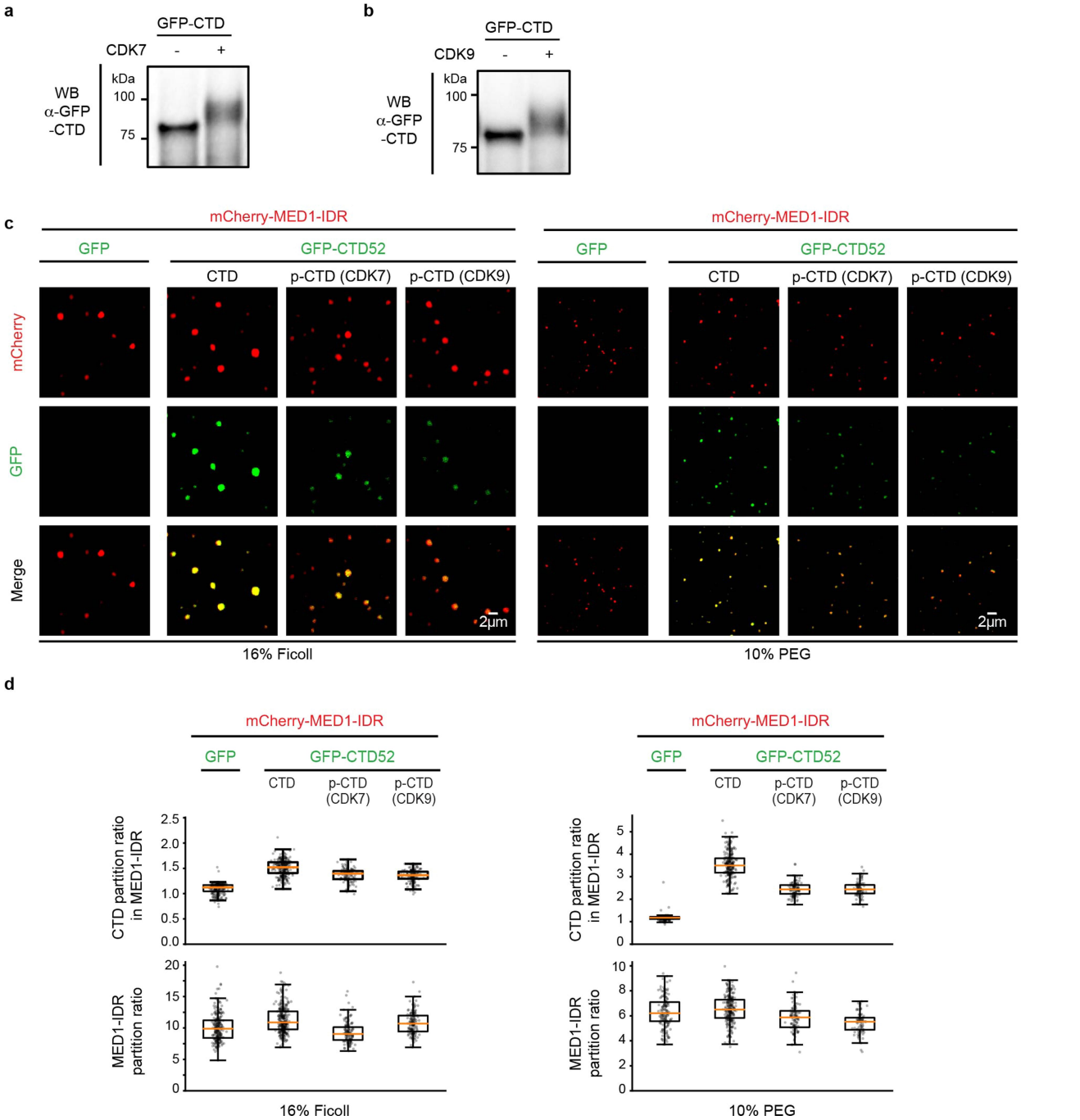


Extended Data Fig. 4 | Liquid-like properties of SRSF2 condensates. **a**, Live-cell imaging of GFP–SRSF2 mouse ES cells. Ten independent fields from one plate of cells were imaged. **b**, Representative images of FRAP experiments performed on the GFP–SRSF2 mouse ES cell line. **c**, Quantification of experiment depicted in **b**. Data are mean ± s.d. ($n = 9$ experiments). **d**, Representative images of live cells before and after treatment with 10% 1,6-hexanediol for approximately 17 min. Five independent fields from one plate of cells were imaged before and after treatment. **e**, Examples of fusion events occurring between SRSF2 puncta in the GFP–SRSF2 cell line. Two fields from two independent plates of cells were imaged over a two-hour time course, and each showed at least one fusion event.

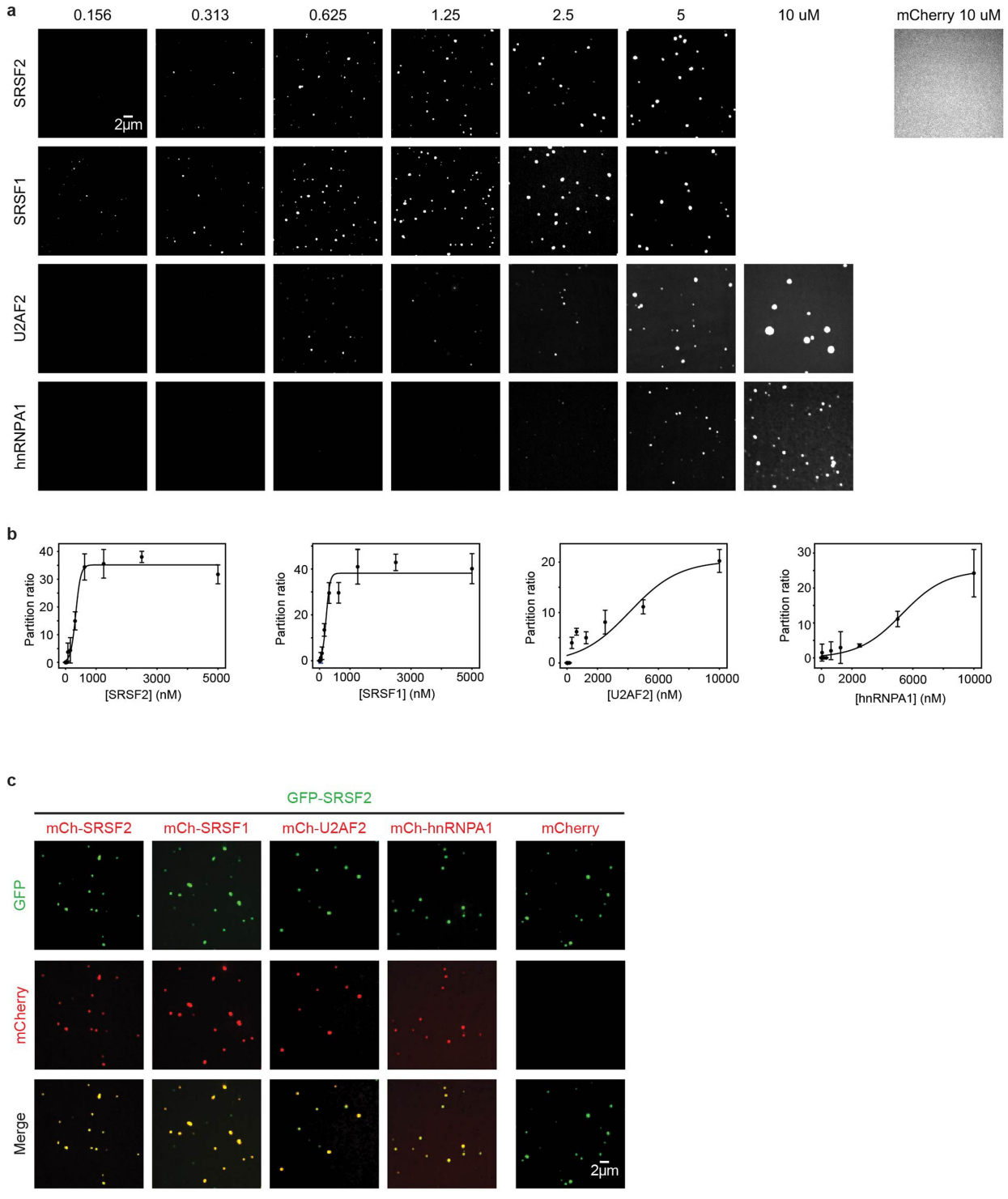


Extended Data Fig. 5 | See next page for caption.

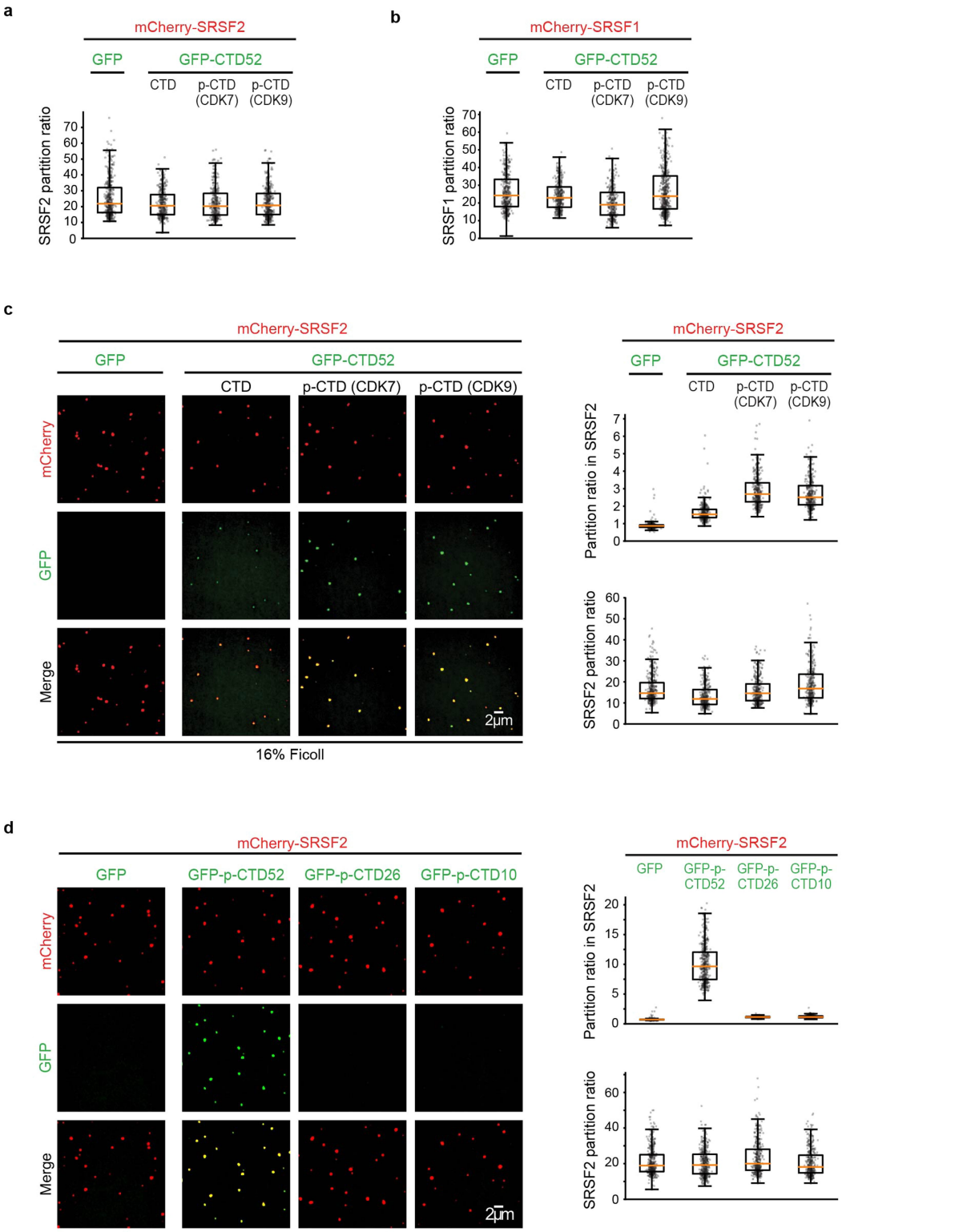
Extended Data Fig. 5 | Effects of treatment with DRB on splicing-factor and transcriptional condensates. a, Representative images exhibiting overlap or lack of overlap between immunofluorescence of SRSF2 and DNA FISH of *Trim28* in mouse ES cells treated with DMSO for 2 h, DRB for 2 h, or DRB for 2 h followed by a 2-h DRB washout. **b**, Representative images exhibiting overlap between immunofluorescence of MED1 and DNA FISH of *Trim28* in mouse ES cells treated with DMSO for 2 h, DRB for 2 h, or DRB for 2 h followed by a 2-h DRB washout. **c**, Representative images exhibiting overlap between immunofluorescence of Pol II and DNA FISH of *Nanog* or *Trim28* in mouse ES cells treated with DMSO for 2 h, DRB for 2 h, or DRB for 2 h followed by a 2-h DRB washout. **d**, Fraction of overlap of FISH foci with immunofluorescence puncta in cells treated with DRB or DRB followed by washout, relative to cells treated with DMSO. The asterisk above the DRB-treated bar indicates $^{*}P < 0.05$ from a two-tailed χ^2 test comparing the number of overlapping and non-overlapping FISH spots in the DMSO versus DRB condition. The asterisk over the washout bar indicates $^{*}P < 0.05$ from a two-tailed χ^2 test comparing the DRB versus washout condition. The splicing factors that we tested showed a significantly decreased, or trended towards a decreased, overlap with FISH spots in the DRB condition compared to DMSO, whereas MED1 and Pol II exhibited limited changes (see Methods).



Extended Data Fig. 6 | Phosphorylation of Pol II CTD in vitro, and CTD partitioning in MED1-IDR droplets. a, Western blot showing phosphorylation of Pol II CTD by CDK7. Equal amounts of non-phosphorylated and phosphorylated GFP–CTD52 (see Methods) were run on a 3–8% SDS–PAGE gel, and analysed by western blot using anti-GFP antibodies. Similar results were obtained in two biological replicates. For gel source data, see Supplementary Fig. 1. **b**, Western blot showing phosphorylation of Pol II CTD by CDK9. Similar results were obtained in two biological replicates. For gel source data, see Supplementary Fig. 1. **c**, Representative images of droplet experiments that measured incorporation of Pol II CTD into MED1-IDR droplets. Purified mCherry–MED1-IDR at 10 μM was mixed with 3.3 μM GFP, GFP–CTD52 or GFP–CTD52 phosphorylated with CDK7 or CDK9 in droplet-formation buffers with 125 mM NaCl and 16% Ficoll-400 or 10% PEG-8000. **d**, Partition ratios of GFP and GFP–CTD in MED1-IDR droplets, and partition ratios of mCherry–MED1-IDR from experiments depicted in **c**.



Extended Data Fig. 7 | Splicing factors form droplets in vitro. **a**, Representative images of droplet formation by mCherry (mCh)–SRSF2, mCherry–SRSF1, mCherry–U2AF2 and mCherry–HNRNPA1 with increasing protein concentrations. All assays were performed in the presence of 120 mM NaCl and 10% PEG-8000. **b**, Quantification of the partition ratios from the experiments depicted in **a**. **c**, Representative images of heterotypic droplets formed between SRSF2 and other splicing factors, including SRSF1, U2AF2 and HNRNPA1. Purified human SRSF2 fused to GFP (GFP–SRSF2) at 2.5 μM was mixed with 2.5 μM purified mCherry–SRSF2, mCherry–SRSF1, mCherry–U2AF2 or mCherry–HNRNPA1 in droplet-formation buffers with 120 mM NaCl and 10% PEG-8000.



Extended Data Fig. 8 | CTD partitioning in SR-protein droplets. **a**, Quantification of the partition ratios of SRSF2 from experiments depicted in Fig. 4b. **b**, Quantification of the partition ratios of SRSF1 from experiments depicted in Fig. 4c. **c**, Representative images and quantification of partition ratios of droplet experiments that measured the incorporation of Pol II CTD into SRSF2 droplets. The same reagents and conditions were used as in Fig. 4b, except that 16% Ficoll-400 was used as a crowding agent. **d**, Representative images and quantification of partition ratios of droplet experiments that measured the incorporation of phosphorylated full-length or truncated Pol II CTD into SRSF2 droplets. The droplet-formation conditions are the same as described in Fig. 4b.

nature research

Corresponding author(s): Richard Young

Last updated by author(s): Jun 15, 2019

Reporting Summary

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

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	MetaMorph acquisition software (V7.8.13.0) was used for collecting confocal images. Image Lab (V6.0.0) was used to collect western blot images. Fiji/ImageJ software (V2.0.0-rc-69) was used to make visualizations.
Data analysis	For analysis of RNA FISH with immunofluorescence custom in-house MATLAB™ scripts were written to process and analyze 3D image data gathered in FISH (RNA/DNA) and IF channels. To analyze in-vitro droplet imaging experiments, custom MATLAB™ and/or Python scripts were written to identify droplets and characterize their size, shape and intensity. ChIP-Seq data were aligned to the mm9 version of the mouse reference genome using bowtie. Wiggle files for display of read coverage in bins were created using MACS. Reads-per-million-normalized wiggle files were displayed in the UCSC genome browser. Metagene plots were made using ngs.plot tool (Shen L et al., 2014). Fiji/ImageJ software (V2.0.0-rc-69) was used to make visualizations. Python 2.7.6 was used to write custom scripts for analysis.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. 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

Datasets generated in this study have been deposited in the Gene Expression Omnibus under accession number GSE120656. Related to figure 3b. Uncropped western blot images are provided in Supplementary Figure 1.

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	For droplet experiments we imaged at least 7 independent fields of view (a few hundred droplets in total) for each experimental condition based on current practices in the field (Lin et al. Mol Cell. 2015; Langdon et al. Science. 2018; Sabari et al. Science 2018; Boija et al. Cell 2018). For IF/FISH data we imaged ~10 fields of view (typically greater than 50 FISH foci total) for each condition based on current practices in the field (Sabari et al. Science 2018; Boija et al. Cell 2018).
Data exclusions	In rare instances, out of focus images were excluded.
Replication	For microscopy experiments, we image multiple fields of view for each condition to ensure that the data we present is replicated across multiple images (at least 7 fields for droplet experiments, 10 fields for FISH/IF experiments, 5 fields for co-IF experiments, 5 fields for Hexanediol experiments). For the splicing inhibition experiment, we confirmed splicing was indeed inhibited upon drug treatment using biological triplicates. For SRSF2 FRAP experiments, we bleached 9 independent SRSF2 puncta. For SRSF2 fusion experiments, we observed multiple instances of fusion. For all experiments, replicates showed the same trends.
Randomization	Samples were not randomized. Appropriate controls were included in the figures.
Blinding	No blinding.

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
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

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

Antibodies

Antibodies used	Primary antibodies: SRSF2 (IF, ChIP: Abcam ab11826 lot GR3248373-1), MED1 (IF, ChIP: Abcam ab64965 lot GR3231334-1), Pol II-CTD (IF: Abcam ab817 lot GR3216482-7), Pol II-CTD-S2P (IF, ChIP: Millipore 04-1571-I lot 2736366), SF3B1 (IF: Sigma HPA050275 lot A115886), U2AF2 (IF: Abcam ab37530 lot 380598), HNRNPA1 (IF: Abcam ab5832 lot GR255077-20), SRSF1 (IF: Santa Cruz 33652 lot E3018), SRRM1 (IF: Abcam ab221061 lot GR322797-1), PRPF8 (IF: Santa Cruz 55533 lot D0114), SNRNP70 (IF: Sigma HPA043516 lot R40804), HP1a (IF: Abcam ab203432), GFP (Western, ChIP: Abcam ab290 lot GR3251545-1), and RNAPII-hypo (ChIP: Abcam ab817 lot GR3216482-7) Secondary antibodies: Goat anti-Rabbit- Alexa 488 (Thermo A11008 lot 1885240), Goat anti-Mouse-Alexa 488 (Thermo A32723 lot SG251135), Goat anti-Rabbit-Alexa 586 (Thermo A11011 lot 1942295), Goat anti-Mouse-Alexa 586 (Thermo A11031 lot 1946341), Goat anti-Rabbit-Alexa 647 (Thermo A21244 lot 913922), Goat anti-Mouse-Alexa 647 (Thermo A21235 lot 1890864), Goat anti-Rat-Alexa 647 (Thermo A21247 lot 1939630), Goat anti-Rat-HRP (GE NA935 batch194920).
Validation	SRSF2, MED1, Pol II-CTD, U2AF2, HNRNPA1, SRSF1, and SRRM1 antibodies were validated in house by siRNA knockdown (each exhibited reduced signal by immunofluorescence when their targets were knocked down). Pol II-CTD-S2P antibody was validated in house by treatment of cells with DRB (the antibody exhibited reduced signal by Immunofluorescence after treatment). SF3B1 and SNRNP70 antibodies were validated by The Cell Atlas and meet the “enhanced” validation criteria by virtue of showing the same staining patterns as other independent antibodies against their targets. Abcam ab64965 and ab290 are “ChIP-grade” validated by Abcam with >200 and >1500 references, respectively.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	V6.5 murine embryonic stem cells (mESCs) were a gift from the Jaenisch lab. GFP-SRSF2 cells were generated from the V6.5 line.
Authentication	Our V6.5 mESC have been verified using short tandom repeat (STR) analysis. GFP-SRSF2 cells were validated by PCR genotyping.
Mycoplasma contamination	Our V6.5 mESC have tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used.

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.*	Datasets generated in this study have been deposited in the Gene Expression Omnibus under accession number GSE120656
Files in database submission	PROCESSED DATA FILES file name 20180803_7415.9944.rpm.WIG.gz 20180803_7416.9950.rpm.WIG.gz 20180831_7546.10064.rpm.WIG.gz 20180831_7540.10070.rpm.WIG.gz RAW FILES file name 20180803_7415.ACAGTG-s_7_1_sequence.txt.gz 20180803_7416.TGACCA-s_7_1_sequence.txt.gz 20180831_7546.TAGCTT-s_4_1_sequence.txt.gz 20180831_7540.CCGTCC-s_5_1_sequence.txt.gz
Genome browser session (e.g. UCSC)	*Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

Methodology

Replicates	1
Sequencing depth	SRSF2 ChIP 43691879 reads Pol II Ser2-P ChIP 70285993 reads
Antibodies	SRSF2 Abcam ab11826 RNA Pol II (Ser2 phospho CTD) Millipore 04-1571
Peak calling parameters	ChIP-Seq data were aligned to the mm9 version of the mouse reference genome using bowtie with parameters –k 1 –m 1 –best and –l set to read length. Wiggle files for display of read coverage in bins were created using MACS with parameters –w –S – space=50 –nomodel –shiftsize=200
Data quality	Aligned Percent SRSF2 0.537831 RNA Pol II (Ser2 phospho CTD) 0.7444
Software	bowtie MACS UCSC genome browser ngs.plot tool (Shen L et al., 2014)