CDK9 and PP2A regulate the link between RNA polymerase II transcription termination and RNA maturation.

Michael Tellier[1*], Justyna Zaborowska[1], Jonathan Neve[2], Takayuki Nojima[3], Svenja Hester[2], Andre Furger[2], Shona Murphy[1 *]

[1] Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK.
[2] Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK.
[3] Medical Institute of Bioregulation, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan

*For correspondence: michael.tellier@path.ox.ac.uk (MT); shona.murphy@path.ox.ac.uk (SM)

Abstract

CDK9 is a critical kinase required for the productive transcription of protein-coding genes by RNA polymerase II (pol II) in higher eukaryotes. Phosphorylation of targets including the elongation factor SPT5 and the carboxyl-terminal domain (CTD) of RNA pol II allows the polymerase to pass an early elongation checkpoint (EEC), which is encountered soon after initiation. In addition to halting RNA polymerase II at the EEC, CDK9 inhibition also causes premature termination of transcription across the last exon, loss of polyadenylation factors from chromatin, and loss of polyadenylation of nascent transcripts. Inhibition of the phosphatase PP2A abrogates the premature termination and loss of polyadenylation caused by CDK9 inhibition, suggesting that CDK9 and PP2A, working together, regulate the coupling of elongation and transcription termination to RNA maturation.

Our phosphoproteomic analyses, using either DRB or an ATP analog-sensitive CDK9 cell line confirm the splicing factor SF3B1 as an additional key target of this kinase. CDK9 inhibition causes loss of interaction of splicing and export factors with SF3B1, suggesting that CDK9 also helps to co-ordinates coupling of splicing and export to transcription.

Introduction

Transcription of a human protein-coding gene by RNA polymerase (pol) II is a complex process comprising initiation, elongation, and termination. In addition, mRNA capping, splicing, and cleavage and polyadenylation, which are required to produce a mature mRNA, are largely co-transcriptional. The dynamic phosphorylation and dephosphorylation of proteins that control transcription and pre-mRNA processing, including pol II itself, are fundamental to the regulation of gene expression. Phosphorylation of pol II mostly occurs on the carboxyl-terminal domain (CTD) of its largest subunit, RBP1. The CTD of human RPB1 contains 52 repeats of the heptapeptide sequence Tyr1,Ser2,Pro3,Thr4,Ser5,Pro6,Ser7 (YSPTSPS) (Zaborowska et al., 2016). During transcription, tyrosine, serine and threonine residues are reversibly and dynamically phosphorylated whilst the two prolines are subject to cis-trans isomerisation. The pattern of heptapeptide phosphorylation and proline isomerisation during the transcription process creates distinct CTD profiles throughout the transcription cycle to coordinate the recruitment of transcription elongation and pre-mRNA processing factors in space and time (Buratowski, 2009, Hsin and Manley, 2012, Corden, 2013, Zaborowska et al., 2016). For example, CTD Ser5 phosphorylation helps to recruit capping proteins at the 5' end of genes and Ser2 phosphorylation helps to recruit polyadenylation and termination factors at the 3' end of genes

In human cells, phosphorylation of the pol II CTD heptapeptide is mainly carried out by cyclin-dependent kinases (CDKs), including CDK7, CDK9, and CDK12 (Zaborowska et al., 2016). CDK9, together with Cyclin T1, forms the Positive Transcription Elongation Factor b complex (P-TEFb). P-TEFb phosphorylates Ser2, Thr4, and Ser5 of the CTD heptapeptide. In addition, P-TEFb phosphorylates several proteins involved in regulation of transcriptional elongation, including the negative elongation factor subunit E (NELFE), the SPT5 subunit of DSIF, and SPT6 (Yamada et al., 2006, Peterlin and Price, 2006, Vos et al., 2018b, Vos et al., 2018a). Soon after initiation of transcription of protein-coding genes, pol II stalls at an early elongation checkpoint (EEC) due to the recruitment of NELF and DSIF. Phosphorylation of these complexes by P-TEFb results in release of NELF and turns DSIF into a positive elongation factor, allowing productive elongation to proceed (Jonkers et al., 2014, Laitem et al., 2015, Vos et al., 2018b, Vos et al., 2018a). Accordingly, inhibition of CDK9 by small molecule inhibitors leads to global accumulation of pol II pausing at the EEC (Jonkers et al., 2014, Laitem et al., 2015, Vos et al., 2018b, Vos et al., 2018a). Using short-term treatment of cells with CDK9 inhibitors, we have shown that inhibition of CDK9 also disrupts transcription at the 3'end of protein-coding genes, causing premature termination of pol II close to the poly(A) site (Laitem et al., 2015). This premature termination is associated with the loss of several pol II-associated factors, including CDK9 itself, SPT5, SSU72, and Cstf64 from the 3' end of protein-coding genes (Laitem et al., 2015). This led us to hypothesize the presence of a poly(A)-associated checkpoint (PAAC), where CDK9 is required to overcome the pause and transcribe past the poly(A) site for cleavage and polyadenylation occur (Laitem et al., 2015, Tellier et al., 2016).

Termination of transcription downstream of an active polyadenylation site requires the action of the exonuclease Xrn2, which degrades the uncapped RNA left associated with pol II after cleavage of the nascent pre-mRNA at the polyA site (Proudfoot, 2016). This is thought to destabilize pol II to allow termination of transcription. Interestingly, phosphorylation of Xrn2 by CDK9 enhances its activity (Sanso et al., 2016).

While the functions of transcriptional kinases are becoming clearer, thanks in part to the development of cell lines with analog-sensitive kinases (Bishop et al., 2000), the role of

phosphatases in this process is still poorly understood. Several CDK9 targets, including SPT5 and Xrn2, are dephosphorylated by the protein phosphatases (PP)1, PP2A, and PP4 (Parua et al., 2018, Huang et al., 2020, Parua et al., 2020, Vervoort et al., 2021). These PPs are involved in numerous cellular processes (Shi, 2009) and whilst their roles in splicing are well known (Mermoud et al., 1992, Shi et al., 2006), their function in transcription and other co-transcriptional processes is only beginning to emerge. Phosphatases have recently been implicated in the regulation of poly(A)-site-dependent transcription termination. PP1 regulates this process by dephosphorylation of SPT5 (Cortazar et al., 2019, Eaton et al., 2020). Knockdown of PP1 or its associated subunit, PNUTS, or inhibition of PP1 with the small molecule inhibitor Tautomycetin, leads to hyperphosphorylation of SPT5 and the elongation rate of pol II no longer decreases downstream of the poly(A). This leads to a transcription termination defect as the exonuclease Xrn2 fails to "catch up" with the elongating pol II (Cortazar et al., 2019, Eaton et al., 2020). PP2A is known to interact with Integrator, a protein complex that cleaves RNA and regulates the amount of paused pol II at the EEC (Huang et al., 2020, Zheng et al., 2020, Vervoort et al., 2021). PP2A is also thought to dephosphorylate the pol II CTD and other factors involved in the early stages of transcription (Huang et al., 2020, Zheng et al., 2020, Vervoort et al., 2021).

Termination of transcription of intron-containing protein-coding genes in human cells is thought to be coupled to terminal exon definition, which requires the coordinated recognition of the 3'SS and the poly(A) site (Cooke et al., 1999, Tellier et al., 2020a). Accordingly, mutation of the 3'SS causes a transcription termination defect due to failure to recognise the poly(A) site (Dye and Proudfoot, 1999). Interactions between splicing factors, such as SF3B1 or U2AF65, and CPA factors, such as CPSF100 or poly(A) polymerase, PAPOLA, have been demonstrated (Niwa et al., 1990, Gunderson et al., 1994, Vagner et al., 2000, Kyburz et al., 2006, Tellier et al., 2020a). In addition, subunit CDC73 of the PAF complex, which plays a role in elongation, interacts with the CPSF73 CPA endonuclease (Rozenblatt-Rosen et al., 2009). Thus, inhibition of CDK9 may trigger premature termination by disrupting the balance of kinase and phosphatase activities required for enabling protein-protein/-DNA/-RNA interactions.

To elucidate the function of CDK9 in the PAAC we generated, using CRISPR/Cas9, a HEK293 cell line where the endogenous genes express analog-sensitive CDK9 (CDK9as). Inhibition of CDK9as or inhibition of CDK9 using small molecule inhibitors promote premature termination of pol II, impair CPA factor recruitment to chromatin, and cause loss of polyadenylation of newly-made pre-mRNA. Interestingly, bioinformatic analysis indicates that the defect in transcription caused by CDK9 inhibition starts across the last exon, rather than at the poly(A) site, implicating disruption of definition of the last exon in the premature termination.

Using phosphoproteomics, we have identified numerous proteins as targets of CDK9, most of which are involved in transcription and RNA biology. The targets include, as expected, SPT5 and NELF, and the U2 snRNP splicing factor component SF3B1, with SPT5 T806 and SF3B1 T142 confirmed as targets of PP1. In addition, CDK9 inhibition causes loss of interaction of splicing and export factors with SF3B1, suggesting that CDK9 helps to co-ordinate coupling of splicing and export to transcription. PP1 inhibition promotes a transcription termination defect, as previously shown (Eaton et al., 2020). However, inhibition of PP1 does not reverse the effect of CDK9 inhibition on transcription and pol II CTD phosphorylation.

Surprisingly, we found that inhibition of the multifunctional phosphatase PP2A reverses the effect of CDK9 inhibition on premature termination, indicating that this phosphatase has a previously unsuspected role at the 3'end of protein-coding genes. Inhibition of PP2A also restores recruitment of poly(A) factors and polyadenylation of newly-made pre-mRNA disrupted by CDK9 inhibition. PP2A inhibition alone causes an increase in the recruitment of poly(A) factors/production of polyadenylated mRNA and transcription termination closer to the poly(A) site. These findings indicate that PP2A can act as a negative regulator of mRNA CPA and that inhibition leads to more efficient cleavage, polyadenylation, and termination.
Taken together our results emphasize that several kinase-phosphatase switches regulate the coordination between pre-mRNA splicing, cleavage/polyadenylation, and termination of transcription.

Results

CDK9 inhibition causes failure of polyadenylation.

We have previously shown that CDK9 inhibition leads to loss of pol II association downstream of the poly(A) site genes (Laitem et al., 2015), indicating either that premature termination is occurring and maturation of pre-mRNA is aborted or that polyadenylation/termination are more efficient and a fully mature mRNA is still produced. We have already shown that levels of the polyadenylation factor CstF64 at the 3' end of genes is reduced after CDK9 inhibition, suggesting that polyadenylation is compromised. In order to follow the production of newly-polyadenylated mRNA, we have analysed production of transcripts from TNFα-inducible genes. HeLa cells were treated with DMSO or TNFα for 30 minutes followed by treatment with DMSO or 5,6-dichlorobenzimidazone-1-β-D-ribofuranoside (DRB) for 30 minutes and 3'READS was then carried out on the purified nuclear mRNAs to measure the production of newly polyadenylated mRNA (Figure 1A) (Neve et al., 2016). There are 307 genes where mRNA production is induced more than two-fold by TNFα in both repeats. Pol II ChIP-qPCR on two of these genes, *SLCO4A1* and *LDLR*, was carried out to analyse induction of transcription by TNFα and the effect of DRB treatment on pol II at their 3'end (Supplementary Figures 1A and B). Polyadenylation of the 307 TNFα-induced mRNAs in the nucleus is significantly decreased, both for all of the genes included and for genes longer than 40 kb, where pol II is still elongating at the 3'end (Figure 1B and Supplementary Figures 1A and B). qRT-PCR of the nuclear polyadenylated mRNAs encoded by selected TNFα-induced genes confirmed that DRB causes a reduction in polyadenylation (Supplementary Figure 1C).
We also performed western blotting to assess the amount of the polyadenylation/termination factors Xrn2, CPSF2, CPSF73, total pol II, and Ser2P in chromatin and nucleoplasm fractions after treatment of cells for 15 or 30 minutes with DRB (Figure 1C-E and Supplementary Figure 1D). As expected, Ser2P is decreased on the chromatin following CDK9 inhibition. There is also a loss of Xrn2, CPSF2 and CPSF73 from the chromatin fraction, with a stronger loss after 30 than 15 minutes. In contrast, the level of these factors in the nucleoplasm fraction increases, indicating that these factors dissociate from chromatin after CDK9 inhibition. Importantly, there is no reduction of these factors in whole cell extract, indicating that the loss from chromatin is not due to active degradation (Supplementary Figures 1E and 1F). Levels of PAPOLA, the poly(A) polymerase, Xrn2 and CPSF30, measured by ChIP-qPCR on our model gene, *KPNB1*, are also reduced after 30 minutes treatment with DRB, whether ratioed to pol II levels or not (Figures 1F and G).

This data indicates that CDK9 inhibition causes both premature termination of pol II and failure to recruit polyadenylation factors, which causes production of polyadenylated mRNA to be aborted.

CDK9 inhibition causes an elongation defect starting at the last exon of protein-coding genes.

To better understand the kinetics of premature termination of pol II caused by CDK9 inhibition, mNET-seq was carried out with a total pol II antibody after treatment of HeLa cells with DRB for 5, 10, 15, or 30 minutes (Figure 2A-C and Supplementary Figure 2A and 2B). The results are similar to those we previously obtained using GRO-seq (Laitem et al., 2015), and metagene profiles analysis indicates that DRB treatment of cells causes an increase in pol II pausing close to the TSS, a loss of pol II entering productive elongation, and premature termination of pol II close to the poly(A) site is detected on genes longer than 40 kb where pol II has not yet "run off" (Figures 2B-D, and Supplementary Figure 2A and 2B). Pol II ChIP-qPCR on our model gene, *KPNB1* gives the same result (Supplementary Figure 2C). DRB concentrations between 12.5 to 100 µM and treatment of cells with a different CDK9 inhibitor, SNS-032 (Chen et al., 2009), give the same result (Supplementary Figure 2C).
To focus on the effect of CDK9 inhibition on pol II behaviour at the 3' end of genes, a metagene analysis was carried out with the pol II signal scaled across the last exon (Figure 2E). Interestingly, the metaprofile indicates that CDK9 inhibition causes an increase in pol II signal across the last exon followed by loss of pol II signal downstream of the poly(A) site. Importantly, the increase in the pol II signal after CDK9 inhibition is specific to the last exon as the pol II signal over the penultimate exons or other internal exons is either unchanged or decreased (Figure 2F and Supplementary Figure 2D). This is particularly apparent on, for example, *ARHGAP23* where pol II levels increase specifically across the last exon (Figure 2G).

Inhibition of analog-sensitive (as) CDK9 produces similar results to small molecule CDK9 inhibitors.

Small molecule kinase inhibitors may target more than one kinase *in vivo* (Bensaude, 2011). To confirm that the results we observed with DRB are due solely to inhibition of CDK9, we used CRISPR/Cas9 to change the gatekeeper phenylalanine to an alanine in all endogenous copies of CDK9 HEK293 cells to make a CDK9 analog-sensitive (as) cell line (Supplementary Figure 3-1A). Mutation of the gatekeeper residue does not greatly affect the cell growth rate (Supplementary Figure 3-1B). Treatment of wild type HEK293 or CDK9as cells with 7.5, 10, or 15 µM of the ATP-analogue 1-NA-PP1 (NA), which should specifically inhibit CDK9as, affects the growth rate of the CDK9as cells with little effect on the wild type HEK293 cells (Supplementary Figure 3-1C). Pol II ChIP-qPCR of *KPNB1* in CDK9as cells after treatment of cells with 7.5, 10, or 15 µM NA for 15 minutes show that at all NA concentrations, pol II pausing and pol II entry into productive elongation is affected (Supplementary Figure 3-1D). Readily detectable premature termination of pol II close to the poly(A) site is caused by 15 µM NA (Supplementary Figure 3-1D). Importantly, no effect is observed on pol II in wild-type HEK293 with 15 µM NA (Supplementary Figure 3-1E). Introduction of the CDK9as mutation or treatment with 15 µM NA also do not affect the level of CDK9 or Cyclin T1 in whole cell extract (Supplementary Figures 3-1F and G).
As CDK9 phosphorylates the pol II CTD, the effect of CDK9 inhibition on pol II CTD phosphorylation was analysed by western blotting of chromatin fractions (Figures 3A and B). As expected, CDK9as inhibition causes loss of Ser2P and Ser5P with two different

phosphoantibodies to each CTD mark (ab5095 and ab5121 from Abcam, 13499S and 13523S from Cell Signaling) while NA treatment has no effect on CTD phosphorylation in wild-type HEK293 cells. ChIP-qPCR on *KPNB1* also indicates that CDK9as inhibition causes loss of Ser2P and Ser5P, whether ratioed to pol II or not (Figure 3C and Supplementary Figure 3-2), indicating that *in vivo* CDK9 activity is necessary for efficient Ser2 and Ser5 phosphorylation.

CDK9 phosphorylates several transcription and splicing factors in vivo.

Several non-CTD targets play critical roles in the functions of these kinases (Sanso et al., 2016, Zaborowska et al., 2016). To identify the *in vivo* targets of CDK9, we performed biological duplicates of SILAC phosphoproteomics in HeLa cells treated with DMSO or DRB for 30 minutes (Figure 4A and Supplementary Figure 4-1A, Supplementary Table 1). We found 100 phosphosites across 74 proteins decreased more than 1.5 fold and with a p-value < 0.1. Amongst these phosphosites, 34 have a SP or a TP motif and are located in 25 different proteins (Supplementary Figure 4-1B). In line with previous similar studies, most CDK9 targets are factors involved in transcription, RNA processing, and RNA biology. Importantly, we identified several known targets of CDK9, including MED1 (T1440), MEPCE (T213, S217, T291), NELFA (S225), SPT5 (S666, S671), and SPT6 (T1523) (Sanso et al., 2016, Vos et al., 2018b, Vos et al., 2018a, Decker et al., 2019). We also identified several residues of the splicing factors SF3B1 (T142, T227, T436) and CDC5L (T377, T396, T430, T442) as targets of CDK9. However, comparison with previously published CDK9 *in vitro* or *in vivo* phosphoproteomics only shows limited overlap (Supplementary Figure 4-1C) (Sanso et al., 2016, Decker et al., 2019).

We have generated new phosphoantibodies for SPT5 T806P and SF3B1 T142P, as these two proteins are known to have functions at the 3'end of protein-coding genes, either at the transcriptional level (SPT5) or through splicing and definition of the last exon (SF3B1) (Kyburz et al., 2006, Cortazar et al., 2019, Tellier et al., 2020a, Parua et al., 2020). Kinase-phosphatase switches have recently been shown to regulate transcription at the 3' end of gene and CDK9 and PP1 activities play a major role in controlling SPT5 T806 phosphorylation and pol II elongation close to poly(A) sites (Parua et al., 2018, Cortazar et al., 2019, Parua et al., 2020). In addition, PP2A is known to dephosphorylate SPT5 S666P (Parua et al., 2018, Huang et al., 2020, Parua et al., 2020). Accordingly, we have analysed the effect on transcription of low concentrations of two small molecule inhibitors, 25 nM Tautomycetin (TT) and 2.5 nM Calyculin A (CA), which inhibit PP1 and PP2A, respectively. H3S10P and Myc S62P, which are known targets of PP1 and PP2A, respectively, were used to control for specific inhibition (Supplementary Figure 4-1D). We observed that at the concentrations used, TT preferentially inhibits PP1, while CA mainly inhibits PP2A. We also tested the effect of CDK9 inhibition by DRB on phosphorylation of these two targets but an increase rather than a decrease in phosphorylation was observed.

Western blotting of the chromatin fraction of cells treated or not with CDK9 inhibitors using these antibodies confirms that CDK9 phosphorylates SPT5 T806P and SF3B1 T142P while PP1, but not PP2A, dephosphorylates these two phosphoresidues (Figures 4B and C). As SF3B1 phosphorylation is associated with splicing activity and chromatin association of the spliceosome (Wang et al., 1998, Bessonov et al., 2010, Murthy et al., 2018, Cossa et al., 2020), we analysed the effect of CDK9 inhibition on proteins interacting with SF3B1. SF3B1 was immunoprecipitated from CDK9as cells treated with DMSO or NA for 30 minutes, followed by proteomic analysis (Figure 4D and Supplementary Table 2). CDK9 inhibition does not affect the interaction between SF3B1 and other SF3B subunits. However,

interaction with SF3A3 is lost. Interaction between SF3B1 and the splicing factors SNW1 and PRPF19, and with the mRNA export factor THO4 was also reduced. To determine whether the decreased association between SF3B1 and other splicing factors could affect co-transcriptional pre-mRNA splicing, mNET-seq was performed with a CTD Ser5P antibody, with treatment of the cells for 15 minutes with either DMSO or DRB. The mNET-seq signal at the 3'end of the internal exons of genes where transcription is still occurring (located after the first 20 kb) was analysed (Figure 4E). The Ser5P peaks at 5' splice sites (5' SS) correspond to the first co-transcriptional cleavage step of pre-mRNA splicing (Nojima et al., 2018a). After 15 minutes of CDK9 inhibition, there is a decrease of the Ser5P signal at 5' SS, indicating that either the first cleavage step of splicing is affected or that interaction between pol II and the product of the first cleavage step of splicing is affected.
These results suggest that phosphorylation of SF3B1 by CDK9 plays a key role in ensuring co-transcriptional splicing and co-transcriptional loading of export factors onto spliced RNA.

PP2A counteracts CDK9 activity at the 3' end of genes.
The phosphatases PP1 and PP2A reverse CDK9-mediated phosphorylation of several residues of SPT5 and SF3B1 (Figure 4B and C) (Parua et al., 2018, Cortazar et al., 2019, Parua et al., 2020). The effect of PP1 and PP2A inhibition on pol II CTD Ser2 and Ser5 phosphorylation in the absence or presence of CDK9as inhibition was therefore investigated (Supplementary Figures 4-2B and C). As expected, NA causes loss of Ser2P and Ser5P while inhibition of PP1 with TT or PP2A with CA results in an increase of Ser2P and Ser5P, supporting the known role of PP1 and PP2A as both CTD Ser2 and Ser5 phosphatases (Washington et al., 2002, Zheng et al., 2020, Vervoort et al., 2021). The combination of NA and TT still results in a decreased Ser2P and Ser5P level while the combination of NA and CA increases Ser2P and Ser5P levels compared to NA alone, indicating that, for CTD phosphorylation, PP1 inhibition cannot overcome CDK9 inhibition whereas PP2A inhibition and CDK9 inhibition are more balanced.
The effect of inhibition of these two phosphatases on transcription of the *KPNB1* gene was also investigated by pol II ChIP-qPCR (Figure 5A). Inhibition of PP1 leads to an increase in pol II pausing at the TSS coupled to a termination defect downstream of the poly(A) site (primer TerDef), as previously reported (Eaton et al., 2020). Inhibition of PP2A also results in an increase in pol II pausing at the TSS with apparent termination of pol II closer to the poly(A) site (primer TerDef). Inhibition of CDK9 and PP1 at the same time has the same effect as CDK9 inhibition alone (Figure 5B). However, inhibition of PP2A and CDK9 at the same time mitigates the effect of CDK9 inhibition at the 3'end of the gene (see primer pA +1.4). Inhibition of CDK9 for 15 minutes followed by treatment with CA for another 15 minutes, or vice-versa, gives a similar outcome, apart from an increased pol II level in the gene body when DRB is followed by CA (Figure 5C). A similar approach with TT results in the same outcome as TT treatment alone (Figure 5D). Treatment of cells with DRB, CA, and TT together has the same effect as DRB and CA (Figure 5E).

CDK9 and PP2A regulate mRNA cleavage and polyadenylation.
As inhibition of PP2A causes a partial reversal of the effect of CDK9 inhibition on transcription at the 3'end of the gene, we tested whether PP2A inhibition also reverses the effect of CDK9 inhibition on pre-mRNA CPA. qRT-PCRs of newly-synthesized nuclear polyadenylated mRNA from TNFα-induced genes after treatment of cells for 30 minutes with DRB or DRB and CA indicate that inhibiting PP2A fully reverses the effect of CDK9

inhibition on the production of polyadenylated mRNA (Figures 6A and B). Surprisingly, inhibiting PP2A alone causes greater production of polyadenylated mRNA for most of the genes tested, indicating that PP2A can act as a negative regulator of CPA. Inhibition of PP1 instead has little effect on polyadenylation of newly-synthesized RNA (Supplementary Figure 6-1A and B).

To determine whether CPA factor recruitment is affected by PP2A inhibition, we performed western blots of Xrn2, CPSF2, and CPSF73 on the chromatin and nucleoplasm fractions following 30 minutes of treatment with DRB, CA, or DRB and CA (Figure 6C and D, and Supplementary Figure 6-1C). Whereas CDK9 inhibition causes a decrease in Xrn2, CPSF2, and CPSF73 on chromatin, CA or DRB and CA treatment cause an increase in CPSF2 and CPSF73 recruitment to chromatin and a coupled decrease of CPSF2 in the nucleoplasmic fraction. Importantly, there was no effect of DRB, CA, or DRB and CA on the total protein level of these termination factors in whole cell extract (Supplementary Figure 6-1D). As pol II CTD Ser2P is associated with the recruitment of CPA factors and CDK9 and PP2A modulate Ser2P, we next investigated by ChIP-qPCR on the *KPNB1* model gene how pol II, Ser2P, CPSF73, and CPSF2 levels are affected by 30 minutes treatment of cells with DRB, CA, or DRB and CA (Figure 6E and Supplementary Figure 6-2A). DRB and CA treatment together results in a localized increase of the Ser2P/pol II ratio at the 3'end of *KPNB1*, which is coupled with an increased ratio of CPSF73/pol II and CPSF2/pol II. Analysis of the nuclear polyadenylated mRNA level of KPNB1 shows that as expected from the 3'READS experiments, it is not induced by TNFα (Supplementary Figure 6-2B). However, we also observe an increase in *KPNB1* nuclear polyadenylated mRNA level after CA treatment while DRB+TT and DRB+CA lead to a relatively small decrease in poly(A)+ mRNA level compared to the strong loss induced by DRB treatment. In addition, we performed a total pol II immunoprecipitation to determine whether DRB, CA, or DRB and CA treatment affects the interaction of CPSF2 and Xrn2 with pol II (Figure 6F). As expected, DRB treatment reduces Ser2P while CA or DRB and CA together does not cause a decrease in Ser2P. Interestingly, DRB treatment causes a strong decrease in the interaction between pol II and CPSF2, and to a lesser level of Xrn2. In contrast, DRB and CA treatment together preserves the interaction between pol II and CPSF2 and Xrn2, indicating that CDK9 and PP2A regulate the interaction between pol II and CPA complex components.

Controlling PP2A recruitment and SPT5 phosphorylation across genes

PP2A recruitment to pol II has been found to be mediated by a specific Integrator complex, INTAC, at the 5'end of genes. We have shown that PP2A is also active at the 3'end of genes, which raise the question of how PP2A is recruited here. Re-analysis of ChIP-seq of Integrator subunits Ints2, Ints11, and Ints13 shows as expected a strong signal at the 5'end of genes with no obvious signal at the 3'end of genes (Figures 7A and 7B) (Stadelmayer et al., 2014). However, re-analysis of pol II ChIP-seq with and without Ints3 or Ints11 knockdown (Stadelmayer et al., 2014) show that in addition to the effect on pol II at the 5'end of the genes, a specific increase of pol II signal is also observed downstream of the poly(A) site of protein-coding genes when Ints3 and Ints11 are knocked-down, indicating that Integrator could also be involved in regulating transcription termination through PP2A recruitment. In addition, a recent ChIP-seq of PPP2R1A, a regulatory subunit of PP2A, indicates that this phosphatase is recruited to the 3'end of genes (Vervoort et al., 2021), supporting our finding of a novel role of PP2A in transcription termination and mRNA CPA.

The control of SPT5 phosphorylation is important for controlling pol II elongation rate and transcription termination (Parua et al., 2018, Cortazar et al., 2019, Eaton et al., 2020). To understand better how SPT5 phosphorylation is modified at the 3'end of genes, we re-analyzed SPT5, SPT5 S666P, and SPT5 T806P ChIP-seq (Figures 7C and 7D) (Parua et al., 2020). While the T806P/SPT5 ratio shows a peak at the TSS followed by a general conserved ratio across the gene body and a loss at the 3'end of the genes, S666P/SPT5 is mostly found across the gene body with a loss at the 3'end of the genes. When the SPT5-P/SPT5 ratio is normalized to the pol II level, a similar pattern is observed for both phosphorylated residues, ie a constant level across gene bodies followed by dephosphorylation at the 3'end of the genes. Interestingly, dephosphorylation of S666P and T806P starts from the last exon rather than the poly(A) site, further supporting the notion that the last exon plays an essential role in coordinating the transition from transcription elongation to transcription termination. In addition, it is clear that phosphatases begin to dominate over CDK9 activity to SPT5 across the last exon. In addition, S666P peaks just upstream and downstream of internal exons and decreases across these exons. Pol II elongation rate is known to decrease across exons, which results in an increased pol II signal and Ser5P across exons compared to introns (Figure 2F) (Jonkers et al., 2014, Nojima et al., 2015), which could facilitate co-transcriptional splicing. Modulation of the SPT5 S666P level, which affects the pol II elongation rate, around exons could therefore regulate pol II transcription across exons. In addition, the peaks of SPT5 S666P may reflect higher CDK9 activity, which could, at the same time, phosphorylate the splicing factors CDC5L and SF3B1, which are part of the U2 snRNP.
However, further investigation will be required to understand the local activities of CDK9 and phosphatases in the regulation of transcription and co-transcriptional processes around internal exons.

Discussion

We previously showed that CDK9 inhibition, in addition to affecting pol II at the EEC, also leads to premature termination of pol II close to the poly(A) site (Laitem et al., 2015). We show here that premature termination of pol II is associated with a loss of mRNA polyadenylation and a loss of recruitment of polyadenylation and termination factors to chromatin. Although we have termed this 3' end CDK9 checkpoint the poly(A)-associated checkpoint, analysis of pol II transcription at single nucleotide resolution using mNET-seq indicates that pol II slows down prematurely from the start of the last exon. Thus, CDK9 inhibition may be causing failure to properly define the last exon, which helps to ensure the correct transition of pol II between elongation and termination (Figure 8A). In support of this, we found several proteins involved in definition of the last exon and the transition between transcription elongation and termination, including SPT5, SF3B1, CDC5L, or METTL3 to be phosphorylated by CDK9 (Ke et al., 2015, Kyburz et al., 2006, Cortazar et al., 2019, Tellier et al., 2020a, Parua et al., 2020). Inhibition of CDK9 also leads to a partial decrease in co-transcriptional splicing, likely due to the loss of phosphorylation of splicing factors like CDC5L and SF3B1, caused, for example, by the loss of interaction of SF3A3, PRPF19 and SNW1 splicing factors with phosphorylated SF3B1. We show here that, in addition to phospho-SPT5, SF3B1 is a target of the phosphatase PP1, supporting a role for this phosphatase in splicing and potentially in the definition of the last exon (Mermoud et al., 1992, Shi et al., 2006).
In addition to the previously-demonstrated roles of PP1 and PP2A in transcription regulation, at the 5' and 3' ends of genes, respectively (Cortazar et al., 2019, Eaton et al.,

2020, Huang et al., 2020, Zheng et al., 2020, Vervoort et al., 2021), we found that inhibiting CDK9 and PP2A, either simultaneously or consecutively, leads to a partial reversal of the premature termination caused by CDK9 inhibition. In addition, inhibition of PP2A reverses the loss of CPA factors from chromatin caused by CDK9 inhibition and production of polyadenylated mRNA is restored, indicating that the CDK9/PP2A kinases/phosphatase pair are involved in regulating the recruitment and activity of the CPA complex (Figure 8B).

Although a low concentration of Tautomycetin or Calyculin A can specifically inhibit PP1 or PP2A, respectively, we cannot rule out that these small molecule inhibitors affect other proteins. However, PP1 and PP2A are clearly not completely redundant as PP1, but not PP2A, dephosphorylates SPT5 T806P and SF3B1 T142P. In addition, both phosphatases are active at the 5' and 3' ends of protein-coding genes but have different functions. PP1 inhibition leads to a termination defect, likely due to a high pol II elongation rate after the poly(A) site caused by hyperphosphorylation of SPT5, which impedes Xrn2-mediated transcription termination (Cortazar et al., 2019, Eaton et al., 2020) (Figure 8C). Conversely, PP2A inhibition seems to promote more efficient cleavage and polyadenylation of transcripts and consequently more efficient termination closer to the poly(A) site (Figure 8B). However, the mechanism behind this remains unclear as the Ser2P level or CPA factor recruitment to the pol II are not higher than in untreated conditions, implicating a different PP2A target.

A decrease in the pol II level downstream of the poly(A) site can therefore be associated with premature non-productive termination, where gene expression is aborted (CDK9 inhibition) (Figure 8A) or more efficient termination coupled to an increase in mature polyadenylated mRNA (CDK9 and PP2A inhibition (Figure 8B)). Premature termination of pol II caused by inhibition of CDK9 is associated with a decreased level of Ser2P and Ser5P and reduced recruitment of CPA factors (Figure 8A). In contrast, inhibition of CDK9 and PP2A together or PP2A alone promote more efficient termination of pol II with the production of de novo nuclear polyadenylated mRNA, which is associated with higher Ser2P, efficient recruitment of CPA factors, and termination closer to the poly(A) site (Figure 8B). A surprising observation is that while CDK9 and PP1 inhibition results in increased SPT5 T806 and SF3B1 T142 phosphorylation, Ser2 and Ser5 phosphorylation remains at a level similar to CDK9 inhibition alone (Figures 4B, 4C, 8C and Supplementary Figures 4-2B and 4-2C). This result reinforces the notion that dephosphorylation of Ser2P and Ser5P by PP1 is redundant with other CTD phosphatases like PP2A.

In line with previous findings, we found that CDK9 activity is needed for CTD phosphorylation on Ser2P and Ser5P *in vivo* (Czudnochowski et al., 2012, Ghamari et al., 2013, Laitem et al., 2015, Greifenberg et al., 2016). However, Ser2 does not seem to always be an *in vivo* target of CDK9 (Parua et al., 2020, Decker et al., 2019). This is reminiscent of identification of CDK12 targets in the pol II CTD, which may be Ser2P, Ser5P, and/or Ser7P (Krajewska et al., 2019, Chirackal Manavalan et al., 2019, Tellier et al., 2020b). The differing results could be explained by the usage of different methods (whole cell extract vs chromatin fraction), different treatment times (minutes vs hours), different molecules (analog sensitive cell lines vs small molecule inhibitors), and different cell lines.

Our DRB mNET-seq time course indicates that after 5 minutes of CDK9 inhibition, the effect on pol II at the EEC is more drastic than at the 3'end of the genes, where a defect becomes clear only after 10 minutes of inhibition. However, the drastic effect of CDK9 inhibition on mRNA CPA may indicate that the loss of mRNA polyadenylation occurs before premature termination of pol II. Thus, dephosphorylation of CDK9 targets including SPT5 and SF3B1 and

the loss of the CPA complex from chromatin occurs rapidly. The subsequent slowing down of pol II over the last exon and early disengagement of pol II may be caused by the loss of elongation factors, such as SPT5, SPT6, or the PAF1 complex, which are known to keep the pol II clamped on the chromatin (Bernecky et al., 2017, Vos et al., 2018a, Hou et al., 2019).

Materials and methods

Cell culture

HEK293 and HeLa cells were obtained from ATCC (ATCC® CRL-1573™ and ATCC® CCL-2™, respectively). HeLa, HEK293 parental cells, and CDK9as HEK293 cells were grown in DMEM medium supplemented with 10% foetal calf serum, 100 U/ml penicillin, 100 µg/ml streptomycin, 2 mM L-glutamine at 37°C and 5% CO_2. HEK293 and CDK9as cells were treated with 7.5, 10, or 15 µM 1-NA-PP1 (Cayman Chemical Company) for 15 and 30 minutes. HEK293, CDK9as, or HeLa cells were treated with 10 ng/ml of TNFα (PeproTech), 2.5 nM Calyculin A (Sigma), 25 nM Tautomycetin (Bio-Techne), 100 µM DRB (Sigma) for 5, 10, 15 or 30 minutes, or 0.5 or 1 µM SNS-032 (LKT labs). As a negative control, HEK293, CDK9as, and HeLa cells were treated with DMSO (the resuspension vehicle for NA, DRB, Calyculin A, and Tautomycetin). Cells were routinely checked to be free of mycoplasma contamination using Plasmo Test Mycoplasma Detection Kit (InvivoGen, rep-pt1).

Analog sensitive cell line creation

Guide RNAs were computationally designed.

Guide RNA 1: 5'-GCTCGCAGAAGTCGAACACC-3'
Guide RNA 2: 5'-CTTCTGCGAGCATGACCTTGC-3'

The modified CDK9 genomic sequence (NCBI RefSeq Accession NG_033942.1) (500 bp either site of the mutation) was cloned into pcDNA3 and used as the repair template for genome editing. The repair template contains a TTC (phenylalanine) to GCT (alanine) mutation. The guide (g)RNAs inserts were cloned into the pX462 vector (obtained from Addgene). HEK293 cells were transfected with the gRNA vectors and correction template using Lipofectamine 2000 (Life Technologies) following the manufacturer instructions. Single clones were isolated by low density plating after Puromycin and Neomycin selection. Genomic DNA from each clone was analysed using PCR and Sanger sequencing.

gDNA preparation

HEL293 and CDK9as cells were incubated in 180 µl ChIP lysis buffer (10 mM Tris–HCl ph8.0, 0.25% Triton X-100, 10 mM EDTA) for 10 min and sonicated for 3 min (30 seconds on/30 seconds off) using a Q800R2 sonicator (QSONICA). 20 µl ammonium acetate was added (final concentration of 400 mM) and samples were mixed with 200 µl phenol/chloroform by vortexing. The samples were centrifuged at 13,000 g for 5 minutes at 25°C and the upper phase was transferred to a new tube. Genomic DNA was precipitated in 70% ethanol, pelleted by centrifugation at 13,000 g for 5 minutes at 25°C and dissolved in nuclease free water. A CDK9 fragment was PCR amplified using the following primers: Forward: 5′-AAGGCTTCTGAGACAGCTGG-3′; Reverse: 5′-CAACCAGCTTCTTTCTTCCTGC-3′. DNA was purified using a QIAquick PCR purification kit (Qiagen) and sequenced by the Source Bioscience Sanger Sequencing Service, Oxford.

RNA preparation

RNA was extracted from HEK293 and CDK9as cells using a *Quick*-RNA Miniprep kit (Zymo Research) according to the manufacturer's instructions. Reverse-transcription (RT) was performed with 500 ng of RNA using random hexamers with the SuperScript III kit (Invitrogen) according to the manufacturer's instructions. Sequencing was performed on a cDNA PCR fragment generated with Forward primer: 5'-AAAGCAGTACGACTCGGTGG-3' and Reverse primer: 5'-GTAGAGGCCGTTAAGCAGCA-3', purified using a QIAquick PCR purification kit (Qiagen) and sequenced by the Source Bioscience Sanger Sequencing Service, Oxford.

Cell proliferation analysis

Cells were seeded at 500/well in 95 µl in 96-well microplates (Greiner, 655090) and measured every 12 or 24 hours by adding alamarBlue HS (Invitrogen) (1/20) and reading in a fluorimeter after 1 hour's incubation, according to the manufacturer's instructions. 1-NA-PP1 or DMSO was added to the cells as noted on the figure.

Chromatin immunoprecipitation (ChIP)

ChIP analysis were performed as previously described (Tellier et al., 2020b). HeLa, HEK293, and CDK9as cells were grown in 100 or 150 mm dishes until they reached ~80% confluency. The cells were fixed with 1% formaldehyde for 10 minutes at room temperature with shaking. Formaldehyde was quenched with 125 mM glycine for 5 minutes at room temperature with shaking. The cells were washed twice with ice-cold PBS, scraped with ice-cold PBS, and transferred into 1.5 or 15 ml Eppendorf tubes. Cells were pelleted for 10 minutes at 1,500 rpm at 4°C. The pellets were then resuspended in ChIP lysis buffer (10 mM Tris–HCl ph8.0, 0.25% Triton X-100, 10 mM EDTA, protease inhibitor cocktail, and phosphatase inhibitor) and incubated 10 minutes on ice before being centrifuged at 1,500 g for 5 minutes at 4°C. Pellets were resuspended in ChIP wash buffer (10 mM Tris–HCl pH8.0, 200 mM NaCl, 1 mM EDTA, protease inhibitor cocktail, and phosphatase inhibitor) and centrifuged at 1,500 g for 5 minutes at 4°C. Pellets were resuspended in ChIP sonication buffer (10 mM Tris–HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, protease inhibitor cocktail, and phosphatase inhibitor) and incubated 10 minutes on ice. HeLa cells were sonicated for 30 cycles, 30 seconds on/30 seconds off using a Bioruptor Pico (Diagenode). HEK293 and CDK9as cells were sonicated for one hour, 30 seconds on/30 seconds off, 40% amplitude, using a Q800R2 sonicator (QSONICA). Chromatin was pelleted at 13,000 rpm for 15 minutes at 4°C and supernatant transferred to a new Eppendorf tube.

Chromatin was pre-cleared for 30 min on a rotating wheel at 4°C with 10 µl of Protein G Dynabeads, previously washed with 100 µl of RIPA buffer (10 mM Tris–HCl pH8.0, 150 mM NaCl, 1 mM EDTA, 0.1% SDS, 1% Triton X-100, 0.1% sodium deoxycholate). Chromatin was quantified on a NanoDrop One with 60-100 µg of chromatin used per IP (antibodies described in Table 1) and incubated overnight on a rotating wheel at 4°C. 15 µl of Dynabeads per IP were washed in 100 µl RIPA buffer. The beads were saturated with 15 µl RIPA containing 4mg/ml of bovine serum albumin (BSA) and mixed overnight on a rotor at 4°C.

Dynabeads were then mixed for 1 hour on a rotating wheel at 4°C with the chromatin incubated with the antibody. Beads were then washed three times with 300 µl ice-cold RIPA buffer, three times with 300 µl High Salt Wash buffer (10 mM Tris–HCl pH8.0, 500 mM NaCl, 1 mM EDTA, 0.1% SDS, 1% Triton X-100, 0.1% sodium deoxycholate), twice with 300 µl LiCl

Wash buffer (10 mM Tris–HCl pH8.0, 250 mM LiCl, 1 mM EDTA, 1% NP-40, 1% sodium deoxycholate), and twice with 300 µl TE buffer (10 mM Tris–HCl pH 7.5, 1 mM EDTA). Each sample was eluted twice from the Dynabeads with 50 µl of Elution buffer (100 mM NaHCO3, 1% SDS, 10 mM DTT) for 15 minutes at 25°C at 1,400 rpm on a Thermomixer. For each input sample, 90 µl of Elution buffer was added to 10 µl total input. Each sample was treated with RNase A (0.6 µl of 10 mg/ml) for 30 minutes at 37°C followed by the addition of 200 mM NaCl and a five hours incubation at 65°C to reverse the crosslinks. Precipitation was performed overnight at -20°C following the addition of 2.5x volume of 100% ethanol. Ethanol was removed after a 20 minutes centrifugation at 13,000 rpm at 4°C and pellets resuspended in 100 µl TE, 25 µl 5x Proteinase K buffer (50 mM Tris–HCl pH 7.5, 25 mM EDTA, 1.25% SDS) and 1.5 µl Proteinase K (20 mg/ml). The samples were incubated two hours at 45°C to degrade the proteins. DNA was purified using Qiagen PCR Purification Kit and kept at -20°C.
ChIP samples were analysed by real-time qPCR using QuantiTect SYBR Green PCR kit (Qiagen) and Rotor-Gene RG-3000 (Corbett Research). Signals are presented as percentage of Input after removing the background signal from the IP with the IgG antibody. The sequence of primers used for ChIP-qPCR is given in Table 2. Experiments were replicated three times and each ChIP sample was measured in triplicate by qPCR.

Co-immunoprecipitation

For each sample and IgG control: 120 µl of Dynabeads M-280 Sheep anti-mouse IgG (Thermo Fisher) were pre-blocked overnight at 4°C on a wheel in 1 ml of PBS supplemented with 0.5% BSA. The next day, the beads were washed three times in IP buffer (25 mM Tris–HCl pH 8.0, 150 mM NaCl, 0.5% NP-40, 10% Glycerol, 2.5 mM MgCl2), before being incubated for two hours at 4°C on a wheel in 600 µl of IP buffer supplemented with 5 µg of total pol II antibody (MABI0601, MBL International) and protease inhibitor cocktail (cOmplete™, EDTA-free Protease Inhibitor Cocktail, Sigma-Aldrich). In the meantime, a 70–80% confluent 15 cm dish of HeLa cells was washed twice with ice-cold PBS and scrapped with ice-cold PBS supplemented with protease inhibitor cocktail. The cells were pelleted at 500 g for 5 minutes at 4°C. The pellets were re-suspended in 800 µl of Lysis buffer (50 mM Tris–HCl pH 8.0, 150 mM NaCl, 1% NP-40, 10% glycerol, 2.5 mM MgCl2, protease inhibitor cocktail, PhosSTOP (Sigma-Aldrich), 1× PMSF (Sigma-Aldrich), and 25–29 units of Benzonase (Merck Millipore)) and incubated at 4°C on a wheel at 16 rpm for 30 minutes. After centrifuging for 15 minutes at 13,000 g at 4°C, 800 µl of Dilution buffer (150 mM NaCl, 10% glycerol, 2.5 mM MgCl2, protease inhibitor cocktail, PhosSTOP, and 1× PMSF) was added to each supernatant
The beads conjugated with antibodies were washed three times with IP buffer supplemented with protease inhibitor cocktail before being incubated with 1 mg of proteins at 4°C on a wheel at 16 rpm for 2 hours. The beads were washed three times with IP buffer supplemented with protease inhibitor cocktail and three times with IP buffer without NP-40 supplemented with protease inhibitor cocktail. Proteins were eluted in 40 µl of 1× LDS plus 100 mM DTT for 10 minutes at 70°C. Western blots were performed with NuPAGE Novex 3–8% Tris-Acetate Protein Gels (Life Technologies).
For the SF3b1 immunoprecipitation, the Pierce™ MS-Compatible Magnetic IP Kit, protein A/G (ThermoFisher Scientific) was used with 5 µg of Sap155 antibody (D221-3, MBL International), according to the manufacturer protocol. Proteomics and data analysis were

performed by the Advanced Mass Spectrometry Facility at the University of Birmingham, United Kingdom.

Trypsin digestion

Trypsin digestion was performed using 10 µL of samples (up to 10 µg of protein) and added 40 µL of 100 mM ammonium bicarbonate (pH 8). After it was added 50 µL 10 mM dithiothreitol (DTT) and samples were incubated at 56°C for 30 minutes. Samples were then cooled to room temperature and cysteines alkylated by addition of 50 µl 50 mM iodoacetamide, mixed and incubated at room temperature in the dark for 30 minutes. 50 µl of trypsin gold (Promega, Southampton, Hampshire, UK, 6 ng/µl) was subsequently added to the samples, which were then incubated at 37°C overnight.

Desalt samples

The liquid samples contain the mixture of peptides are desalted using millipore C18 ZipTips. Tips are prepared by pre-wetting in 100% acetonitrile and rinsed in 2x10 µL 0.1% formic acid. Samples are repeat pipetted throughout the volume of the samples ten times. The tip is then washed with 3x10 µl 0.1% formic acid to remove excess salts before elution of peptides with 10 µL of 50% acetonitrile/water/0.1% formic acid. Samples are dried down to remove the acetonitrile, and then re-suspended in 0.1% formic acid solution in water.

LC-MS/MS Experiment

UltiMate® 3000 HPLC series (Dionex, Sunnyvale, CA USA) was used for peptide concentration and separation. Samples were trapped on precolumn, Acclaim PepMap 100 C18, 5 um, 100A 300um i.d. x 5mm (Dionex, Sunnyvale, CA USA) and separated in Nano Series™ Standard Columns 75 µm i.d. x 15 cm, packed with C18 PepMap100, 3 µm, 100Å (Dionex, Sunnyvale, CA USA). The gradient used was from 3.2% to 44% solvent B (0.1% formic acid in acetonitrile) for 30 minutes. The column was then washed with 90% mobile phase B before re-equilibrating at 3.2% mobile phase B. Peptides were eluted directly (~ 350 nL min-1) via a Triversa Nanomate nanospray source (Advion Biosciences, NY) into a QExactive HF Orbitrap mass spectrometer (ThermoFisher Scientific). The spray voltage of QE HF was set to 1.7 kV through Triversa NanoMate and heated capillary at 275°C. The mass spectrometer performed a full FT-MS scan (m/z 360–1600) and subsequent HCD MS/MS scans of the 20 most abundant ions with dynamic exclusion setting 15S. Full scan mass spectra were recorded at a resolution of 120,000 at m/z 200 and ACG target of 3×106. Precursor ions were fragmented in HCD MS/MS with resolution set up at 15,000 and a normalized collision energy of 28. ACG target for HCD MS/MS was 1x105. The width of the precursor isolation window was 1.2 m/z and only multiply-charged precursor ions were selected for MS/MS. Spectra were acquired for 56 minutes.

The MS and MS/MS scans were searched against Uniprot database using Protein Discovery 2.2 software, Sequest HT algorithm (Thermo Fisher). Variable modifications were deamidation (N and Q), oxidation (M) and phosphorylation (S, T and Y). The precursor mass tolerance was 10 ppm and the MS/MS mass tolerance was 0.02 Da. Two missed cleavage was allowed and data were filtered with a false discovery rate (FDR) of 0.01. Protein with at least two high confidence peptides are accepted as real hit.

Protein extraction and western blot

Western blot analysis was performed on chromatin and nucleoplasm extracts as previously described in the mNET-seq procedure (Nojima et al., 2015) until purification of the chromatin fraction. The chromatin pellet was digested in 100 µl of nuclease-free water supplemented with 1 µl of Benzonase (25–29 units, Merck Millipore) for 15 minutes at 37°C in a thermomixer at 1,400 rpm. 10 µg of proteins were boiled in 1× LDS plus 100 mM DTT. Western blots were performed with NuPAGE Novex 4–12% Bis–Tris Protein Gels (Life Technologies).
For whole cell extract, cells were washed in ice-cold PBS twice, collected in ice-cold PBS with a 3,000 rpm centrifugation for 5 minutes at 4°C. The pellets were re-suspended in RIPA buffer supplemented with protease inhibitor cocktail and PhosSTOP, kept on ice for 30 minutes with a vortexing step every 10 minutes. After centrifugation at 14,000 g for 15 minutes at 4°C, the supernatants were kept and quantified with the Bradford method. 20 µg of proteins were boiled in 1× LDS plus 100 mM DTT. Western blots were performed with NuPAGE Novex 4–12% Bis–Tris Protein Gels (Life Technologies). The list of primary antibodies is shown in Table 1.
Secondary antibodies were purchased from Merck Millipore (Goat Anti-Rabbit IgG Antibody, HRP-conjugate, 12-348, and Goat Anti-Mouse IgG Antibody, HRP conjugate, 12-349), the chemiluminescent substrate (SuperSignal West Pico PLUS) from Thermo Fisher, and the membranes visualized on an iBright FL1000 Imaging System (Thermo Fisher). Quantification of the western blots was performed with Image Studio Lite software.

RNA subcellular fractionation

RNA subcellular fractionation as performed as described before (Neve et al., 2016). Briefly, a ~80% confluent 10 or 15 cm dish was washed twice with ice-cold PBS, scrapped in ice-cold PBS. The cells were pelleted at 1,000 rpm for 5 minutes at 4°C and then resuspended with slow pipetting in 1 ml of Lysis Buffer B (10 mM Tris-HCl pH 8, 140 mM NaCl, 1.5 mM MgCl2, 0.5 % NP-40). Following centrifugation at 1,000 g for 3 minutes at 4°C, the pellets were resuspended in 1 ml of Lysis Buffer B and 100 µl of the Detergent Stock Solution (3.3 % (w/v) sodium deoxycholate, 6.6 % (v/v) Tween 40) was added under slow vortexing. The nuclei were then spun down at 1000 g for 3 minutes at 4°C. The nuclei pellet was then washed once more in 1 ml of Lysis Buffer B and spun down at 1000 g for 3 minutes at 4°C. The nuclei pellet was then resuspended in 1 ml of TRIzol using a 21-gauge syringe and incubated 5 minutes at room temperature. Following the addition of 200 µl of chloroform, the samples were vortexed vigorously for 15 seconds and spun at 12,000 g for 15 minutes at 4°C. The aqueous fraction was transferred to a new tube containing 580 µl of isopropanol. After a 10 minutes incubation at room temperature, the samples were spun at 12,000 g for 10 minutes at 4°C. The pellets were resuspended in 87 µl of water, 10 µl of 10 X DNase buffer, 2 µl of DNase I (Roche), and 1 µl of RNase OUT (ThermoFisher Scientific), and incubated 30 minutes at 32°C. RNA were then purified twice with phenol:chloroform extraction and finally resuspended in nuclease free water and concentrations determined using a NanoDrop One.

qRT-PCR

For each qRT-PCR reaction, 500 ng of RNA were reverse transcribed with Oligo(dT)12-18 Primer (ThermoFisher Scientific) and the SuperScript III kit (ThermoFisher Scientific), according to the manufacturer's instructions. cDNA was amplified by qPCR with a QuantiTect SYBR Green PCR kit (QIAGEN) and a Rotor-Gene RG-3000 (Corbett Research). The sequence of primers used for qRT-PCR is given in Table 2. Values are normalized to

GAPDH mRNA, used as control. Experiments were replicated at least three times to ensure reproducibility, and each RNA sample was measured in triplicate by qPCR.

3'READS Protocol

The 3'READS protocol was originally described in (Hoque et al., 2013). Briefly, 25-30 µg of RNA was subjected to one round of poly(A) selection using the Poly(A)PuristTM MAG kit (Ambion) according to the manufacturer's protocol, followed by fragmentation using Ambion's RNA fragmentation kit at 70°C for 5 minutes. Poly(A)-containing RNA fragments were isolated using the CU5T45 oligo (a chimeric oligo containing 5 Us and 45 Ts, Sigma) which were bound to the MyOne streptavidin C1 beads (Invitrogen) through biotin at its 5' end. Binding of RNA with CU5T45 oligo-coated beads was carried out at room temperature for 1 hour in 1x binding buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA), followed by washing with a low salt buffer (10 mM Tris-HCl pH 7.5, 1 mM NaCl, 1 mM EDTA, 10% formamide). RNA bound to the CU5T45 oligo was digested with RNase H (5U in 50 µl reaction volume) at 37°C for 1 hour, which also eluted RNA from the beads. Eluted RNA fragments were purified by phenol:chloroform extraction and ethanol precipitation, followed by phosphorylation of the 5' end with T4 kinase (NEB). Phosphorylated RNA was then purified by the RNeasy kit (Qiagen) and was sequentially ligated to a 5'-adenylated 3'-adapter (5'-rApp/NNNNGATCGTCGGACTGTAGAACTCTGAAC/3ddC) with the truncated T4 RNA ligase II (Bioo Scientific) and to a 5' adapter (5'- GUUCAGAGUUCUACAGUCCGACGAUC) with T4 RNA ligase I (NEB). The resultant RNA was reverse-transcribed to cDNA with Superscript III (Invitrogen) followed by a library preparation with the NEBNext Fast DNA Library Prep Set for Ion Torrent (NEB). cDNA libraries were sequenced on an Ion Torrent Proton.

mNET-seq and library preparation

mNET-seq was carried out as previously described (Nojima et al., 2015) with minor changes. In brief, the chromatin fraction was isolated from four ~80% confluent 15 cm dish cells treated with DMSO or DRB (5, 10, 15, or 30 minutes). Chromatin was digested in 100 µl of MNase (40 units/µl) reaction buffer for 2 minutes at 1,400 rpm at 37°C in a Thermomixer. MNase was inactivated by the addition of 10 µl EGTA (25mM). The soluble digested chromatin was collected after centrifugation at 13,000 rpm for 5 minutes at 4°C. The supernatant was diluted with 400 µl of NET-2 buffer and antibody-conjugated Dynabeads M-280 Sheep anti-mouse IgG (ThermoFisher Scientific) beads were added. Antibodies used: Pol II (MABI0601, MBL International) and Ser5P (MABI0603, MBL International). Immunoprecipitation was performed at 4°C for one hour. The beads were washed in the cold room six times with 1 ml of NET-2 buffer, and once with 100 µl of 1xPNKT (1xPNK buffer and 0.05% Triton X-100) buffer. Washed beads were incubated in 200 µl PNK reaction mix at 1,400 rpm at 37°C in a Thermomixer for 6 minutes. After the reaction, beads were washed once with 1 ml of NET-2 buffer and RNA was extracted with Trizol reagent. RNA was suspended in urea Dye and resolved on 6% TBU gel (ThermoFisher Scientific) at 200 V for 5 minutes. In order to size select 35–100 nt RNAs, a gel fragment was cut between BPB and XC dye markers. A 0.5 ml tube was prepared with 3–4 small holes made with 25G needle and placed in a 1.5 ml tube. Gel fragments were placed in the layered tube and broken down by centrifugation at 12,000 rpm for 1 minute at room temperature. The small RNAs were eluted from the gel using RNA elution buffer (1 M NaOAc and 1 mM EDTA) at 25°C for one hour on a rotating wheel at 16 rpm at room temperature. Eluted RNA was purified with

SpinX column (Coster) with two glass filters (Millipore) and the flow-through RNA was ethanol precipitated. RNA libraries were prepared according to manual of TruSeq Small RNA Library Preparation Kit (Illumina). 12–14 cycles of PCR were used to amplify the library. Libraries were resolved on a 6% TBE polyacrylamide gel (ThermoFisher Scientific), size-selected to remove primer-primer ligated DNA, and eluted from the gel with the RNA elution buffer. Deep sequencing (Hiseq4000, Illumina) was conducted by the high throughput genomics team of the Wellcome Trust Centre for Human Genetics (WTCHG), Oxford.

SILAC phosphoproteomics

SILAC phosphoproteomics was performed as previously described (Poss et al., 2016). For stable isotope labelling with amino acids in cell culture (SILAC), Hela cells were grown in DMEM media for SILAC (minus L-Lysine and L-Arginine, Fisher Scientific) and with SILAC dialysed Foetal Bovine Serum (Dundee Cell Products). The medium was supplemented with either Arg10 (33.6 mg/ml) and Lys8 (73 mg/ml) or Arg0 and Lys0 for heavy and light treatment, respectively. After six passages at 1:3 ratio, SILAC incorporation test in HeLa cells was validated by mass spectrometry analysis.

Cells were passaged 7-8 times in SILAC media on 15 cm dishes. For each replicate, approximately 20 mg total protein was harvested for analysis after treatment with either DMSO or DRB for 30 minutes (first replicate: heavy cells DRB; light cells: DMSO; second replicate: heavy cells DMSO; light cells: DRB). After removing the media, each dish was scraped in 750 µl 95°C SDT (4% SDS, 100 mM Tris pH 7.9, 10 mM TCEP) buffer with subsequent heating at 95°C for 10 minutes. Lysates were sonicated for two minutes each. Protein concentrations were determined using a Bradford assay and samples were mixed 1:1 based on total protein concentrations. FASP was carried out in two 10 kDa MWCO filters with a 50 mM iodoacetamide alkylation step and proteins were digested in 2M urea with 2% wt/wt Lys-C (Wako) for 6 h and 2% modified trypsin (Promega) for 12 h at 37°C. FASP eluates were acidified and desalted on Oasis HLB extraction cartridges.

TiO2 Phosphopeptide Enrichment, ERLIC Chromatography, and LC-MS/MS

Protocols were carried out as described (Stuart et al., 2015). An Orbitrap Velos (Thermo Fisher) was used for quantitative proteome analysis while an Orbitrap LTQ (Thermo Fisher) was used for phosphoproteomics. The samples were run on a 60 mins gradient / 10 HCD method.

Proteomics data analysis

All raw mass spectrometry files for phosphoproteomics and quantitative proteomics were searched using the MaxQuant (v1.5.0.35) software package. Duplicate proteomic and phosphoproteomic were searched individually against the Uniprot human proteome database (downloaded on 16/01/2013) using the following MaxQuant parameters: multiplicity was set to 2 (heavy/light) with Arg10 and Lys8 selected, LysC/P was selected as an additional enzyme, “re-quantify” was unchecked, and Phospho (STY) was selected as a variable modification in both runs.

For phosphosite analysis, the Phospho (STY) table was processed with Perseus (v1.6.2.3) using the following workflow: reverse and contaminant reads were removed, the site table was expanded to accommodate differentially phosphorylated peptides, and rows without any quantification were removed after site table expansion. Normalized heavy to light ratios

were log2 transformed for statistical analyses. Differential abundance of peptides following DRB treatment was estimated by t-tests with Welch correction, two sided, unpaired. The volcano plot was prepared with GraphPad Prism 9.1. Sequence motif plots were prepared with WebLogo 3 (Crooks et al., 2004).

Generation of phosphoantibodies

Rabbit phosphoantibodies against SPT5 T806P and SF3B1 T142P were generated by Eurogentec based on the peptide sequences: H - CYG SQT(PO3H2) PLH DGS R - NH2 for SPT5 and H - CAD GGK T(PO3H2)PD PKM N - NH2 for SF3B1. Phosphoantibodies specificity was ensured by selecting against recognition of the unphosphorylated peptide.

QUANTIFICATION AND STATISTICAL ANALYSIS

Gene annotation

The Gencode V35 annotation, based on the hg38 version of the human genome, was used to extract the list of protein-coding genes. A list of 9,883 expressed protein-coding genes was obtained by keeping only the genes longer than 2 kb and with their highest transcript isoform expressed in two nuclear RNA-seq in HeLa cells (Nojima et al., 2018b) at more than 0.1 transcript per million (TPM), following quantification of transcript expression with Salmon version 0.14.1 (Patro et al., 2017). The list of used exons was obtained by extracting the location of exons from Gencode V35 from the highest transcribed nuclear poly(A)+ RNA of each of the 9,883 protein-coding genes.

mNET-seq data processing

Adapters were trimmed with Cutadapt version 1.18 (Martin, 2011) in paired-end mode with the following options: --minimum-length 10 -q 15,10 -j 16 – A GATCGTCGGACTGTAGAACTCTGAAC – a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC. Trimmed reads were mapped to the human GRCh38.p13 reference sequence with STAR version 2.7.3a (Dobin et al., 2013) and the parameters: --runThreadN 16 --readFilesCommand gunzip -c -k --limitBAMsortRAM 20000000000 --outSAMtype BAM SortedByCoordinate. SAMtools version 1.9 (Li et al., 2009) was used to retain the properly paired and mapped reads (-f 3). A custom python script (Nojima et al., 2015) was used to obtain the 3′ nucleotide of the second read and the strandedness of the first read. Strand-specific bam files were generated with SAMtools. Samples normalization was checked against the termination region of the *RNU2* snRNA, which is known to be insensitive to DRB (Medlin et al., 2003), and also verified for consistency against a previously published GRO-seq performed by our group in HeLa cells treated for 30 minutes with DMSO or DRB (Laitem et al., 2015). FPKM-normalized bigwig files were created with deepTools version 3.4.2 (Ramirez et al., 2016) bamCoverage tool with the parameters -bs 1 -p max –normalizeUsing RPKM.

ChIP-seq data processing

Adapters were trimmed with Cutadapt with the following options: --minimum-length 10 -q 15, 10 -j 16 -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCA. Trimmed reads were mapped to the human GRCh38.p13 reference genome with STAR and the parameters: --runThreadN 16 --readFilesCommand gunzip -c -k –alignIntronMax 1 --limitBAMsortRAM 20000000000 --

outSAMtype BAM SortedByCoordinate. SAMtools was used to retain the properly mapped reads and to remove PCR duplicates. Reads mapping to the DAC Exclusion List Regions (accession: ENCSR636HFF) were removed with BEDtools version 2.29.2 (Quinlan and Hall, 2010). FPKM-normalized bigwig files were created with deepTools bamCoverage tool with the parameters -bs 10 -p max –e --normalizeUsing RPKM.

Analysis of 3'READS data

Raw reads were mapped to the human genome (hg19) with Bowtie 2 (Langmead and Salzberg, 2012) using the option "-M 8 --local". This mode allows the ends of the sequence not to be part of the alignment (soft clipped). This is useful for reads containing sequence corresponding to the poly(A) tail which are not present in the genomic sequence. Reads that were shorter than 15 nt, were non-uniquely mapped to genome (MAPQ < 10), or contained more than 2 mismatches in alignment were discarded. FPKM-normalized bigwig files were created with deepTools bamCoverage tool with the parameters -bs 10 -p max --normalizeUsing RPKM.

Metagene profiles

Metagene profiles of genes scaled to the same length were then generated with Deeptools2 computeMatrix tool with a bin size of 10 bp and the plotting data obtained with plotProfile –outFileNameData tool. Graphs representing the (IP – Input) signal (ChIP-seq) or the mNET-seq signal were then created with GraphPad Prism 9.1. Metagene profiles are shown as the average of two biological replicates.

P-values and significance tests

P-values were computed with an unpaired two-tailed Student's t test. Statistical tests were performed in GraphPad Prism 9.1.

Acknowledgments

We thank Chris Norbury for his helpful comments on the manuscript. We thank Shabaz Mohammed for help with the SILAC phosphoproteomics. We thank the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics for sequencing. We would also like to thank the University of Birmingham Advanced Mass Spectrometry Facility for the SF3B1 mass spectrometry analysis. This work was supported by a Wellcome Trust Investigator Award (WT106134AIA and WT210641/Z/18/Z) and a Biotechnology and Biological Sciences Research Council grant (BB/R016836/1) to S.M. A.F.'s research is funded by the BBSRC (BB/N001184/1) and J.N. is funded by the MRC

Competing interests

The authors declare that no competing interests exist.

Data availability

Sequencing data have been deposited in GEO under accession code GSE176541. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al., 2019) partner repository with the dataset identifier PXD026662 (SF3B1 IP) and PXD026720 (SILAC). All data generated or analysed during this study are included in the manuscript and supporting files. We include full excel spreadsheets representing original mass spectrometry data.

Author contributions

Michael Tellier, Conceptualization, Formal analysis, Validation, Investigation, Visualization, Writing - original draft, Writing - review and editing; Justyna Zaborowska, Investigation; Jonathan Neve, Investigation, Formal analysis; Takayuki Nojima, Investigation; Svenja Hester, Investigation, Resources; Andre Furger, Supervision, Resources; Shona Murphy, Conceptualization, Supervision, Investigation, Funding acquisition, Methodology, Project administration, Writing - original draft, Writing - review and editing.

References

BENSAUDE, O. 2011. Inhibiting eukaryotic transcription: Which compound to choose? How to evaluate its activity? *Transcription,* 2**,** 103-108.

BERNECKY, C., PLITZKO, J. M. & CRAMER, P. 2017. Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. *Nat Struct Mol Biol,* 24**,** 809-815.

BESSONOV, S., ANOKHINA, M., KRASAUSKAS, A., GOLAS, M. M., SANDER, B., WILL, C. L., URLAUB, H., STARK, H. & LUHRMANN, R. 2010. Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. *RNA,* 16**,** 2384-403.

BISHOP, A. C., UBERSAX, J. A., PETSCH, D. T., MATHEOS, D. P., GRAY, N. S., BLETHROW, J., SHIMIZU, E., TSIEN, J. Z., SCHULTZ, P. G., ROSE, M. D., WOOD, J. L., MORGAN, D. O. & SHOKAT, K. M. 2000. A chemical switch for inhibitor-sensitive alleles of any protein kinase. *Nature,* 407**,** 395-401.

BURATOWSKI, S. 2009. Progression through the RNA polymerase II CTD cycle. *Mol Cell,* 36**,** 541-6.

CHEN, R., WIERDA, W. G., CHUBB, S., HAWTIN, R. E., FOX, J. A., KEATING, M. J., GANDHI, V. & PLUNKETT, W. 2009. Mechanism of action of SNS-032, a novel cyclin-dependent kinase inhibitor, in chronic lymphocytic leukemia. *Blood,* 113**,** 4637-45.

CHIRACKAL MANAVALAN, A. P., PILAROVA, K., KLUGE, M., BARTHOLOMEEUSEN, K., RAJECKY, M., OPPELT, J., KHIRSARIYA, P., PARUCH, K., KREJCI, L., FRIEDEL, C. C. & BLAZEK, D. 2019. CDK12 controls G1/S progression by regulating RNAPII processivity at core DNA replication genes. *EMBO Rep,* 20**,** e47592.

COOKE, C., HANS, H. & ALWINE, J. C. 1999. Utilization of splicing elements and polyadenylation signal elements in the coupling of polyadenylation and last-intron removal. *Mol Cell Biol,* 19**,** 4971-9.

CORDEN, J. L. 2013. RNA polymerase II C-terminal domain: Tethering transcription to transcript and template. *Chem Rev,* 113**,** 8423-55.

CORTAZAR, M. A., SHERIDAN, R. M., ERICKSON, B., FONG, N., GLOVER-CUTTER, K., BRANNAN, K. & BENTLEY, D. L. 2019. Control of RNA Pol II Speed by PNUTS-PP1 and Spt5 Dephosphorylation Facilitates Termination by a "Sitting Duck Torpedo" Mechanism. *Mol Cell,* 76**,** 896-908 e4.

COSSA, G., ROESCHERT, I., PRINZ, F., BALUAPURI, A., SILVEIRA VIDAL, R., SCHULEIN-VOLK, C., CHANG, Y. C., ADE, C. P., MASTROBUONI, G., GIRARD, C., WORTMANN, L., WALZ, S., LUHRMANN, R., KEMPA, S., KUSTER, B., WOLF, E., MUMBERG, D. & EILERS, M. 2020. Localized Inhibition of Protein Phosphatase 1 by NUAK1 Promotes Spliceosome Activity and Reveals a MYC-Sensitive Feedback Control of Transcription. *Mol Cell,* 77**,** 1322-1339 e11.

CROOKS, G. E., HON, G., CHANDONIA, J. M. & BRENNER, S. E. 2004. WebLogo: a sequence logo generator. *Genome Res,* 14**,** 1188-90.

CZUDNOCHOWSKI, N., BOSKEN, C. A. & GEYER, M. 2012. Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. *Nat Commun,* 3**,** 842.

DECKER, T. M., FORNE, I., STRAUB, T., ELSAMAN, H., MA, G., SHAH, N., IMHOF, A. & EICK, D. 2019. Analog-sensitive cell line identifies cellular substrates of CDK9. *Oncotarget,* 10**,** 6934-6943.

DOBIN, A., DAVIS, C. A., SCHLESINGER, F., DRENKOW, J., ZALESKI, C., JHA, S., BATUT, P., CHAISSON, M. & GINGERAS, T. R. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics,* 29**,** 15-21.

DYE, M. J. & PROUDFOOT, N. J. 1999. Terminal exon definition occurs cotranscriptionally and promotes termination of RNA polymerase II. *Mol Cell,* 3**,** 371-8.

EATON, J. D., FRANCIS, L., DAVIDSON, L. & WEST, S. 2020. A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes. *Genes Dev,* 34**,** 132-145.

GHAMARI, A., VAN DE CORPUT, M. P., THONGJUEA, S., VAN CAPPELLEN, W. A., VAN IJCKEN, W., VAN HAREN, J., SOLER, E., EICK, D., LENHARD, B. & GROSVELD, F. G. 2013. In vivo live imaging of RNA polymerase II transcription factories in primary cells. *Genes Dev,* 27**,** 767-77.

GREIFENBERG, A. K., HONIG, D., PILAROVA, K., DUSTER, R., BARTHOLOMEEUSEN, K., BOSKEN, C. A., ANAND, K., BLAZEK, D. & GEYER, M. 2016. Structural and Functional Analysis of the Cdk13/Cyclin K Complex. *Cell Rep,* 14**,** 320-31.

GUNDERSON, S. I., BEYER, K., MARTIN, G., KELLER, W., BOELENS, W. C. & MATTAJ, L. W. 1994. The human U1A snRNP protein regulates polyadenylation via a direct interaction with poly(A) polymerase. *Cell,* 76**,** 531-41.

HOQUE, M., JI, Z., ZHENG, D., LUO, W., LI, W., YOU, B., PARK, J. Y., YEHIA, G. & TIAN, B. 2013. Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nat Methods,* 10**,** 133-9.

HOU, L., WANG, Y., LIU, Y., ZHANG, N., SHAMOVSKY, I., NUDLER, E., TIAN, B. & DYNLACHT, B. D. 2019. Paf1C regulates RNA polymerase II progression by modulating elongation rate. *Proc Natl Acad Sci U S A,* 116**,** 14583-14592.

HSIN, J. P. & MANLEY, J. L. 2012. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev,* 26**,** 2119-37.

HUANG, K. L., JEE, D., STEIN, C. B., ELROD, N. D., HENRIQUES, T., MASCIBRODA, L. G., BAILLAT, D., RUSSELL, W. K., ADELMAN, K. & WAGNER, E. J. 2020. Integrator Recruits Protein Phosphatase 2A to Prevent Pause Release and Facilitate Transcription Termination. *Mol Cell,* 80**,** 345-358 e9.

JONKERS, I., KWAK, H. & LIS, J. T. 2014. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *Elife,* 3**,** e02407.

KE, S., ALEMU, E. A., MERTENS, C., GANTMAN, E. C., FAK, J. J., MELE, A., HARIPAL, B., ZUCKER-SCHARFF, I., MOORE, M. J., PARK, C. Y., VAGBO, C. B., KUSSNIERCZYK, A., KLUNGLAND, A., DARNELL, J. E., JR. & DARNELL, R. B. 2015. A majority of m6A residues are in the last exons, allowing the potential for 3' UTR regulation. *Genes Dev,* 29**,** 2037-53.

KRAJEWSKA, M., DRIES, R., GRASSETTI, A. V., DUST, S., GAO, Y., HUANG, H., SHARMA, B., DAY, D. S., KWIATKOWSKI, N., POMAVILLE, M., DODD, O., CHIPUMURO, E., ZHANG,

T., GREENLEAF, A. L., YUAN, G. C., GRAY, N. S., YOUNG, R. A., GEYER, M., GERBER, S. A. & GEORGE, R. E. 2019. CDK12 loss in cancer cells affects DNA damage response genes through premature cleavage and polyadenylation. *Nat Commun,* 10**,** 1757.

KYBURZ, A., FRIEDLEIN, A., LANGEN, H. & KELLER, W. 2006. Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3' end processing and splicing. *Mol Cell,* 23**,** 195-205.

LAITEM, C., ZABOROWSKA, J., ISA, N. F., KUFS, J., DIENSTBIER, M. & MURPHY, S. 2015. CDK9 inhibitors define elongation checkpoints at both ends of RNA polymerase II-transcribed genes. *Nat Struct Mol Biol,* 22**,** 396-403.

LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods,* 9**,** 357-9.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA PROCESSING, S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.

MARTIN, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal,* 17**,** 10-12.

MEDLIN, J. E., UGUEN, P., TAYLOR, A., BENTLEY, D. L. & MURPHY, S. 2003. The C-terminal domain of pol II and a DRB-sensitive kinase are required for 3' processing of U2 snRNA. *EMBO J,* 22**,** 925-34.

MERMOUD, J. E., COHEN, P. & LAMOND, A. I. 1992. Ser/Thr-specific protein phosphatases are required for both catalytic steps of pre-mRNA splicing. *Nucleic Acids Res,* 20**,** 5263-9.

MURTHY, T., BLUEMN, T., GUPTA, A. K., REIMER, M., JR., RAO, S., PILLAI, M. M. & MINELLA, A. C. 2018. Cyclin-dependent kinase 1 (CDK1) and CDK2 have opposing roles in regulating interactions of splicing factor 3B1 with chromatin. *J Biol Chem,* 293**,** 10220-10234.

NEVE, J., BURGER, K., LI, W., HOQUE, M., PATEL, R., TIAN, B., GULLEROVA, M. & FURGER, A. 2016. Subcellular RNA profiling links splicing and nuclear DICER1 to alternative cleavage and polyadenylation. *Genome Res,* 26**,** 24-35.

NIWA, M., ROSE, S. D. & BERGET, S. M. 1990. In vitro polyadenylation is stimulated by the presence of an upstream intron. *Genes Dev,* 4**,** 1552-9.

NOJIMA, T., GOMES, T., GROSSO, A. R. F., KIMURA, H., DYE, M. J., DHIR, S., CARMO-FONSECA, M. & PROUDFOOT, N. J. 2015. Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell,* 161**,** 526-540.

NOJIMA, T., REBELO, K., GOMES, T., GROSSO, A. R., PROUDFOOT, N. J. & CARMO-FONSECA, M. 2018a. RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-transcriptional Splicing. *Mol Cell,* 72**,** 369-379 e4.

NOJIMA, T., TELLIER, M., FOXWELL, J., RIBEIRO DE ALMEIDA, C., TAN-WONG, S. M., DHIR, S., DUJARDIN, G., DHIR, A., MURPHY, S. & PROUDFOOT, N. J. 2018b. Deregulated Expression of Mammalian lncRNA through Loss of SPT6 Induces R-Loop Formation, Replication Stress, and Cellular Senescence. *Mol Cell,* 72**,** 970-984 e7.

PARUA, P. K., BOOTH, G. T., SANSO, M., BENJAMIN, B., TANNY, J. C., LIS, J. T. & FISHER, R. P. 2018. A Cdk9-PP1 switch regulates the elongation-termination transition of RNA polymerase II. *Nature,* 558**,** 460-464.

PARUA, P. K., KALAN, S., BENJAMIN, B., SANSO, M. & FISHER, R. P. 2020. Distinct Cdk9-phosphatase switches act at the beginning and end of elongation by RNA polymerase II. *Nat Commun,* 11**,** 4338.

PATRO, R., DUGGAL, G., LOVE, M. I., IRIZARRY, R. A. & KINGSFORD, C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods,* 14**,** 417-419.

PEREZ-RIVEROL, Y., CSORDAS, A., BAI, J., BERNAL-LLINARES, M., HEWAPATHIRANA, S., KUNDU, D. J., INUGANTI, A., GRISS, J., MAYER, G., EISENACHER, M., PEREZ, E., USZKOREIT, J., PFEUFFER, J., SACHSENBERG, T., YILMAZ, S., TIWARY, S., COX, J., AUDAIN, E., WALZER, M., JARNUCZAK, A. F., TERNENT, T., BRAZMA, A. & VIZCAINO, J. A. 2019. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res,* 47**,** D442-D450.

PETERLIN, B. M. & PRICE, D. H. 2006. Controlling the elongation phase of transcription with P-TEFb. *Mol Cell,* 23**,** 297-305.

POSS, Z. C., EBMEIER, C. C., ODELL, A. T., TANGPEERACHAIKUL, A., LEE, T., PELISH, H. E., SHAIR, M. D., DOWELL, R. D., OLD, W. M. & TAATJES, D. J. 2016. Identification of Mediator Kinase Substrates in Human Cells using Cortistatin A and Quantitative Phosphoproteomics. *Cell Rep,* 15**,** 436-50.

PROUDFOOT, N. J. 2016. Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science,* 352**,** aad9926.

QUINLAN, A. R. & HALL, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics,* 26**,** 841-2.

RAMIREZ, F., RYAN, D. P., GRUNING, B., BHARDWAJ, V., KILPERT, F., RICHTER, A. S., HEYNE, S., DUNDAR, F. & MANKE, T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res,* 44**,** W160-5.

ROZENBLATT-ROSEN, O., NAGAIKE, T., FRANCIS, J. M., KANEKO, S., GLATT, K. A., HUGHES, C. M., LAFRAMBOISE, T., MANLEY, J. L. & MEYERSON, M. 2009. The tumor suppressor Cdc73 functionally associates with CPSF and CstF 3' mRNA processing factors. *Proc Natl Acad Sci U S A,* 106**,** 755-60.

SANSO, M., LEVIN, R. S., LIPP, J. J., WANG, V. Y., GREIFENBERG, A. K., QUEZADA, E. M., ALI, A., GHOSH, A., LAROCHELLE, S., RANA, T. M., GEYER, M., TONG, L., SHOKAT, K. M. & FISHER, R. P. 2016. P-TEFb regulation of transcription termination factor Xrn2 revealed by a chemical genetic screen for Cdk9 substrates. *Genes Dev,* 30**,** 117-31.

SHI, Y. 2009. Serine/threonine phosphatases: mechanism through structure. *Cell,* 139**,** 468-84.

SHI, Y., REDDY, B. & MANLEY, J. L. 2006. PP1/PP2A phosphatases are required for the second step of Pre-mRNA splicing and target specific snRNP proteins. *Mol Cell,* 23**,** 819-29.

STADELMAYER, B., MICAS, G., GAMOT, A., MARTIN, P., MALIRAT, N., KOVAL, S., RAFFEL, R., SOBHIAN, B., SEVERAC, D., RIALLE, S., PARRINELLO, H., CUVIER, O. & BENKIRANE, M. 2014. Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. *Nat Commun,* 5**,** 5531.

STUART, S. A., HOUEL, S., LEE, T., WANG, N., OLD, W. M. & AHN, N. G. 2015. A Phosphoproteomic Comparison of B-RAFV600E and MKK1/2 Inhibitors in Melanoma Cells. *Mol Cell Proteomics,* 14**,** 1599-615.

TELLIER, M., FERRER-VICENS, I. & MURPHY, S. 2016. The point of no return: The poly(A)-associated elongation checkpoint. *RNA Biol,* 13**,** 265-71.

TELLIER, M., MAUDLIN, I. & MURPHY, S. 2020a. Transcription and splicing: A two-way street. *Wiley Interdiscip Rev RNA,* 11**,** e1593.

TELLIER, M., ZABOROWSKA, J., CAIZZI, L., MOHAMMAD, E., VELYCHKO, T., SCHWALB, B., FERRER-VICENS, I., BLEARS, D., NOJIMA, T., CRAMER, P. & MURPHY, S. 2020b. CDK12

globally stimulates RNA polymerase II transcription elongation and carboxyl-terminal domain phosphorylation. *Nucleic Acids Res,* 48**,** 7712-7727.

VAGNER, S., VAGNER, C. & MATTAJ, I. W. 2000. The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing. *Genes Dev,* 14**,** 403-13.

VERVOORT, S. J., WELSH, S. A., DEVLIN, J. R., BARBIERI, E., KNIGHT, D. A., OFFLEY, S., BJELOSEVIC, S., COSTACURTA, M., TODOROVSKI, I., KEARNEY, C. J., SANDOW, J. J., FAN, Z., BLYTH, B., MCLEOD, V., VISSERS, J. H. A., PAVIC, K., MARTIN, B. P., GREGORY, G., DEMOSTHENOUS, E., ZETHOVEN, M., KONG, I. Y., HAWKINS, E. D., HOGG, S. J., KELLY, M. J., NEWBOLD, A., SIMPSON, K. J., KAUKO, O., HARVEY, K. F., OHLMEYER, M., WESTERMARCK, J., GRAY, N., GARDINI, A. & JOHNSTONE, R. W. 2021. The PP2A-Integrator-CDK9 axis fine-tunes transcription and can be targeted therapeutically in cancer. *Cell*.

VOS, S. M., FARNUNG, L., BOEHNING, M., WIGGE, C., LINDEN, A., URLAUB, H. & CRAMER, P. 2018a. Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. *Nature,* 560**,** 607-612.

VOS, S. M., FARNUNG, L., URLAUB, H. & CRAMER, P. 2018b. Structure of paused transcription complex Pol II-DSIF-NELF. *Nature,* 560**,** 601-606.

WANG, C., CHUA, K., SEGHEZZI, W., LEES, E., GOZANI, O. & REED, R. 1998. Phosphorylation of spliceosomal protein SAP 155 coupled with splicing catalysis. *Genes Dev,* 12**,** 1409-14.

WASHINGTON, K., AMMOSOVA, T., BEULLENS, M., JEREBTSOVA, M., KUMAR, A., BOLLEN, M. & NEKHAI, S. 2002. Protein phosphatase-1 dephosphorylates the C-terminal domain of RNA polymerase-II. *J Biol Chem,* 277**,** 40442-8.

YAMADA, T., YAMAGUCHI, Y., INUKAI, N., OKAMOTO, S., MURA, T. & HANDA, H. 2006. P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Mol Cell,* 21**,** 227-37.

ZABOROWSKA, J., EGLOFF, S. & MURPHY, S. 2016. The pol II CTD: new twists in the tail. *Nat Struct Mol Biol,* 23**,** 771-7.

ZHENG, H., QI, Y., HU, S., CAO, X., XU, C., YIN, Z., CHEN, X., LI, Y., LIU, W., LI, J., WANG, J., WEI, G., LIANG, K., CHEN, F. X. & XU, Y. 2020. Identification of Integrator-PP2A complex (INTAC), an RNA polymerase II phosphatase. *Science,* 370.

Legends

Figure 1. CDK9 inhibition causes failure of polyadenylation.

A. Schematic of the nuclear 3'READS experiments. **B.** Violin plots of the 3'READS signal on all the TNFα induced genes (n=307) or the set of TNFα induced genes longer than 40 kb (n=111). **** $p < 0.0001$. Statistical test: Friedman test with Dunn's multiple test correction. **C.** Western blot of total pol II, Ser2P, Xrn2, CPSF2, CPSF73, and histone H3 as a loading control, on the chromatin fraction. **D.** Western blot of Xrn2, CPSF2, and Nucleolin as a loading control, on the nucleoplasm fraction. The CPSF73 antibody does not provide reliable results on the nucleoplasm fraction. **E.** Quantification of the western blots shown in **C** and **D**. n=2 biological replicates, p-value: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Statistical test: two-tailed unpaired t test. **F.** ChIP-qPCR of total pol II, CPSF73, PAPOLA, and Xrn2 on the model gene, KPNB1. n=3 biological replicates, p-value: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Statistical test: two-tailed unpaired t test. **G.** Ratio of the different CPA and termination

factors to total pol II. n=3 biological replicates, p-value: * $p < 0.05$, ** $p < 0.01$. Statistical test: two-tailed unpaired t test.

Figure 2. CDK9 inhibition causes an elongation defect starting at the last exon of protein-coding genes.

A. Schematic of the total pol II mNET-seq experiments. **B.** Metagene profile of total pol II after 30 minutes treatment with DMSO (black) or DRB (green) around the TSS of expressed protein-coding genes (n=6,965). **C.** Metagene profile of total pol II after 30 minutes treatment with DMSO (black) or DRB (green) around the used poly(A) site of expressed protein-coding genes longer than 40 kb (n=2,816). **D.** Screenshot of the genome browser mNET-seq tracks across the protein-coding gene *KPNB1*. Red arrows indicate increased pol II pausing, loss of pol II entering productive elongation, and premature termination. **E.** Metagene profile of total pol II after 30 minutes treatment with DMSO (black) or DRB (green) around the used last exon of expressed protein-coding genes (n=2,526). **F.** Metagene profile of total pol II after 30 minutes treatment with DMSO (black) or DRB (green) around the penultimate exon of expressed protein-coding genes (n=2,426). **G.** Screenshot of the genome browser mNET-seq tracks at the 3'end of protein-coding gene *ARHGAP23*. The red arrow indicates increased pol II signal across the last exon.

Figure 3. Inhibition of analog-sensitive (as) CDK9 produces similar results to small molecule CDK9 inhibitors.

A. Western blot of total pol II, Ser2P (Abcam or Cell Signaling), Ser5P (Abcam or Cell Signaling), and histone H3 as a loading control, on the chromatin fraction of CDK9as or wild-type HEK293 cells treated with DMSO or 15 μM 1-NA-PP1 for 15 or 30 minutes. **B.** Quantification of the western blots shown in A. n=3 biological replicates, p-value: * $p < 0.05$. Statistical test: two-tailed unpaired t test. **C.** ChIP-qPCR of total pol II and of Ser2P (Abcam or Cell Signaling), and Ser5P (Abcam or Cell Signaling) ratioed to total pol II on the model gene, *KPNB1*. n=3 biological replicates, p-value: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Statistical test: two-tailed unpaired t test.

Figure 4. CDK9 phosphorylates several transcriptional and splicing factors *in vivo*.

A. Distribution of the proteins found to be phosphorylated by CDK9 in SILAC phosphoproteomics in HeLa cells treated or not with 100 μM DRB for 30 minutes (fold change > 1.5 in both biological duplicates, p-value < 0.1, number of phosphopeptides decreased shown in brackets). **B.** Western blot of total SPT5, SPT5 T806P, total SF3B1, SF3B1 T142P, and histone H3 as a loading control, on the chromatin fraction of CDK9as cells treated for 30 minutes with DMSO, NA, tautomycetin (TT), calyculin A (CA), NA+TT, or NA+CA. **C.** Quantification of the western blots shown in B. n=2 biological replicates, p-value: ** $p < 0.01$, *** $p < 0.001$. Statistical test: two-tailed unpaired t test. **D.** Number of unique peptides found in proteomics after SF3b1 pull-down from CDK9as cells treated with DMSO or 1-NA-PP1 for 30 minutes. **E.** Metagene profile of mNET-seq Ser5P after 15 minutes treatment with DMSO (blue) or DRB (red) around the last nucleotide of used exons with a positive Ser5P signal of expressed protein-coding genes (n=2,001). Boxplot of mNET-seq Ser5P signal at the last nucleotide. P-value: **** $p < 0.0001$, statistical test: Wilcoxon matched-pairs signed rank test.

Figure 5. PP2A counteracts CDK9 activity at the 3' end of genes.

A. ChIP-qPCR of total pol II after 30 minutes treatment with DMSO, DRB, TT, or CA on the model gene, *KPNB1*. **B.** ChIP-qPCR of total pol II after 30 minutes treatment with DMSO, DRB, DRB+TT, or DRB+CA on the model gene, *KPNB1*. **C.** ChIP-qPCR of total pol II after 30 minutes treatment with DMSO or DRB, or 15 minutes treatment with DRB or CA followed by another 15 minutes with CA (DRB->CA) or DRB (CA->DRB), respectively, on the model gene, *KPNB1*. **D.** ChIP-qPCR of total pol II after 30 minutes treatment with DMSO or DRB, or 15 minutes treatment with DRB or TT followed by another 15 minutes with TT (DRB->TT) or DRB (TT->DRB), respectively, on the model gene, *KPNB1*. **E.** ChIP-qPCR of total pol II after 30 minutes treatment with DMSO or DRB+CA+TT on the model gene, *KPNB1*. n=3 biological replicates, p-value: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Statistical test: two-tailed unpaired t test.

Figure 6. CDK9 and PP2A regulate mRNA cleavage and polyadenylation.
A. Schematic of the nuclear qRT-PCR experiments. **B.** qRT-PCR of nuclear polyadenylated mRNAs of several TNFα induced or non-induced genes with a 30 minutes DMSO, DRB, CA, or DRB+CA treatment. n=4 biological replicates, p-value: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Statistical test: two-tailed unpaired t test. **C.** Western blot of Xrn2, CPSF2, CPSF73, and histone H3 as a loading control, on the chromatin fraction of HeLa cells after 30 minutes DMSO, DRB, CA, or DRB+CA treatment. **D.** Quantification of the western blots shown in C. n=3 biological replicates, p-value: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Statistical test: two-tailed unpaired t test. **E.** ChIP-qPCR of Ser2P, CPSF73, or CPSF2 ratioed to total pol II after 30 minutes treatment with DMSO, DRB, CA, or DRB+CA on the model gene, *KPNB1*. n=3 biological replicates, p-value: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Statistical test: two-tailed unpaired t test. **F.** Co-immunoprecipitation of total pol II from HeLa cells treated for 30 minutes with DMSO, DRB, CA, or DRB+CA followed by western blot with total pol II, Ser2P, CPSF2, Xrn2, and GAPDH antibodies.

Figure 7. Controlling PP2A recruitment and SPT5 phosphorylation across genes
A. Metagene profile of Integrator subunits Ints3, Ints11, and Ints13 across expressed protein-coding genes. ChIP-seq data re-analysed from (Stadelmayer et al., 2014) (n=6,965). **B.** Metagene profile of pol II following knockdown of Ints3 or across expressed protein-coding genes. ChIP-seq data re-analysed from (Stadelmayer et al., 2014) (n=6,965). **C.** Left: Metagene profile of S666P or T806P ratioed to total SPT5 across expressed protein-coding genes. Right: Data of the left side ratioed to the pol II signal. ChIP-seq data re-analysed from (Parua et al., 2020) (n=6,965). **D.** Metagene profile of S666P or T806P ratioed to total SPT5 around the first exon, internal exons, or last exons of expressed protein-coding genes (n=6,041, 62,559, 6,041, respectively).

Figure 8. Model of CDK9, PP1, and PP2A function in transcription.
During the transcription cycle, CDK9 phosphorylates in addition to the pol II CTD on Ser2 and Ser5, proteins found in the elongation complex and the U2 snRNP complex, such as SPT5 and SF3B1. The phosphatase PP1 dephosphorylates several CDK9 targets, including the pol II CTD on Ser2P and Ser5P, SPT5 T806P, and SF3B1 T142P. The phosphatase PP2A dephosphorylates the pol II CTD on Ser2P and Ser5P. **A.** Following CDK9 inhibition, a general decrease in pol II CTD phosphorylation and of proteins found in the elongation complex and the U2 snRNP is observed, coupled to a loss of recruitment of the CPA complex, resulting in the production of poly(A)- mRNAs. **B.** Inhibition of CDK9 and PP2A together results in more

phosphorylation of the pol II CTD, but not of SPT5 T806P, and SF3B1 T142P. However, CDK9 and PP2A inhibition is sufficient to counteract the effect of CDK9 inhibition on transcription and on CPA. **C.** Inhibition of CDK9 and PP1 together results in higher phosphorylation of SPT5 T806P, and SF3B1 T142P but is not sufficient to counteract the effect of CDK9 inhibition on transcription and on CPA factor recruitment.

Antibody	Supplier	Catalogue number
Mouse monoclonal anti-Pol II CTD, Total	MBL international	MABI0601
Mouse monoclonal anti-Pol II CTD, phospho Ser5	MBL international	MABI0603
RNA Polymerase II/POLR2A Antibody	Novus Biologicals	NBP2-32080
RNA pol II antibody (mAb)	Active Motif	39097
RNA pol II CTD phospho Ser2 antibody (mAb)	Active Motif	61083
RNA pol II CTD phospho Ser5 antibody (mAb)	Active Motif	61085
Anti-RNA polymerase II CTD repeat YSPTSPS (phospho S2) antibody	Abcam	ab5095
Anti-RNA polymerase II CTD repeat YSPTSPS (phospho S5) antibody	Abcam	ab5131
Rabbit anti-SF3b155/SAP155 Antibody, Affinity Purified	Bethyl Laboratories	A300-996A
Anti-SAP155 Monoclonal Antibody (SF3b1)	MBL international	D221-3
SF3b1 T142P	This study	/
Rabbit anti-SUPT5H Antibody, Affinity Purified	Bethyl Laboratories	A300-868A
SPT5 T806P	This study	/
Myc	Santa Cruz Biotechnology	sc-764X
Myc S62P	Abcam	ab51156
Anti-Cdk9 antibody [EPR3119Y]	Abcam	ab76320
Cyclin T1 (H-245)	Santa Cruz Biotechnology	sc-10750
Rabbit anti-CPSF30 Antibody, Affinity Purified	Bethyl Laboratories	A301-584A
Rabbit anti-CPSF73 Antibody, Affinity Purified	Bethyl Laboratories	A301-091A
Rabbit anti-CPSF100 Antibody, Affinity Purified	Bethyl Laboratories	A301-581A
Rabbit anti-PAPOLA Antibody, Affinity Purified	Bethyl Laboratories	A301-008A
Rabbit anti-XRN2 Antibody, Affinity Purified	Bethyl Laboratories	A301-103A
Histone H3	Abcam	ab1791
Histone H3 S10P	Life Technologies	PA517869
Nucleolin	Abcam	ab22758
Beta-actin	Cell Signaling	4967S
Anti-beta Tubulin antibody - Loading Control	Abcam	ab6046
Mouse monoclonal anti-Pol II CTD, Total	MBL international	MABI0601
Mouse monoclonal anti-Pol II CTD, phospho Ser5	MBL international	MABI0603
RNA Polymerase II/POLR2A Antibody	Novus Biologicals	NBP2-32080

Table 1: List of antibodies used in mNET-seq, ChIP, co-immunoprecipitation, and western blots.

Name	Forward primer	Reverse primer
KPNB1 TSS	TTACTTCCTCCCTCCAAATGGG	ACAGCCTCCCTTCCTTCTTTC
KPNB1 TSS+3.2	GCCCAGAGAACAAGAAATCG	GGAATGGACAAGCTGTGTTG
KPNB1 TSS+20.2	TGCAAGAGCCAGTGGGAACACTT	CCTCTACTCAGCAATGATACTTC
KPNB1 pA-4.8	CTGAGGAAACTGAAGAACCAAG	GAAGGCAGTGCTTGCCAGAAT
KPNB1 pA-2.9	GAGGAGTGTGCACGGATGCTGAA	CCAAGATGGCCGATGTTATGG
KPNB1 pA-0.4	TAGTTACCGTCTGCTTGGGAAGATG	CCTCTGACAGCAAGTCCAACATT
KPNB1 pA+1.4	GACTCATCACACCAAGGTCAC	GATAGTGCTGGGAAGGAAATGG
KPNB1 pA+2.6	GTACATCTCAGCTTTGGCATATG	GCCCAGAACATAGCAGGCATTGC
KPNB1 pA+4.1	GTTTCACCGTGTTAGCCAGGATGG	CCACAGCCATGTTCATTTCTGC
KPNB1 TD	AGGAGCATGGCTTTTCTCTG	TCATGCTGGAACTGGTTGAG
SLCO4A1 TSS	CCAGCGGATGAATGAAGCG	ACTCACCGTCCCCGTCTC
SLCO4A1 TSS +1.3	TCAGCGACGTTTTCATTCCC	ACATTCCAGCACTCAACACC
SLCO4A1 TSS +16	ACAGTGCACTTGGCTTTGAG	AGATGAGTCTGTGCAGTTGC
SLCO4A1 pA -0.8	ATTTGGGGCCCACCTTAGATC	TGTCATGGGTTGAACAGTGC
SLCO4A1 pA	TGAGGAAGGCTTGTGTGTCC	GACCACCAAACTCTGCAAACTC
SLCO4A1 pA +1.5	TGGACTGCAGACACATCAGAG	TTTTCGCAGCACACGTTGAC
SLCO4A1 pA +3.7	GAGATTTTCCACGGCTGCTG	TTTGACGTTTCCAGCTGCTC
Neg	TGGTACAACCACAGCTCAGTG	AAGCTGGACATGGTTGTGTG
LDLR TSS	AATCACCCCACTGCAAACTC	TAGCTGGAAACCCTGGCTTC
LDLR TSS +1.9	TGGGATTGCCTGATGAACAC	AAGGCAGTTCAAAGCTCTCC
LDLR TSS +25	TTCAGTGTGGTGCTGACAAC	TTCTCTGCTGGAAACCCAAC
LDLR pA -1.8	CAGAGAAGACCAAAGCATTGCC	AATCCCAACCCAAGCCATTG
LDLR pA	AATCGCCGTGTTACTGTTGC	TGCCAATCCCTTGTGACATC
LDLR pA +1.6	GTGATTGTGTTCTCTGCTGTCG	TCGCACTTAGCACTCAACAG
ATF3	AACCACAGTCAGTGGAGAGATG	TTCTCACAGCTGCAAACACC
LDLR	TTTGACGGGACTTCAGGTTC	TCCCTTGTGACATCTTCACG
NR4A3	AAGCCACCAGCTGTTAATGG	GCAATGCTGTTAGAGGAGCAG
PHLDB2	GAAACGACTTCAGGCAAGTCTC	GCCATGTTTTAGGAAAGAGCAC
SLCO4A1	TCCTCTTCTTTGCCATAGCC	ACAAGTTTCCAGGCCATCTG
ATP5F1B	ACCCATTGAAGAAGCTGTGG	CAATCAAGGCTCTTGTGCAG
TARS1	CATGGAAAAGGAGGAACAGC	CTTTGCCAAACTCCTCCAAG
TUBB	ATATGTTCCTCGTGCCATCC	TTTGGCCCAGTTGTTACCTG
GAPDH	CAACGACCACTTTGTCAAGC	TTCCTCTTGTGCTCTTGCTG
KPNB1	CAGTCTGTGATGGCATTTAAG	TCTGACTTTCAAACCATTTCACCTCC

Table 2: List of primers used for ChIP-qPCR and qRT-PCR

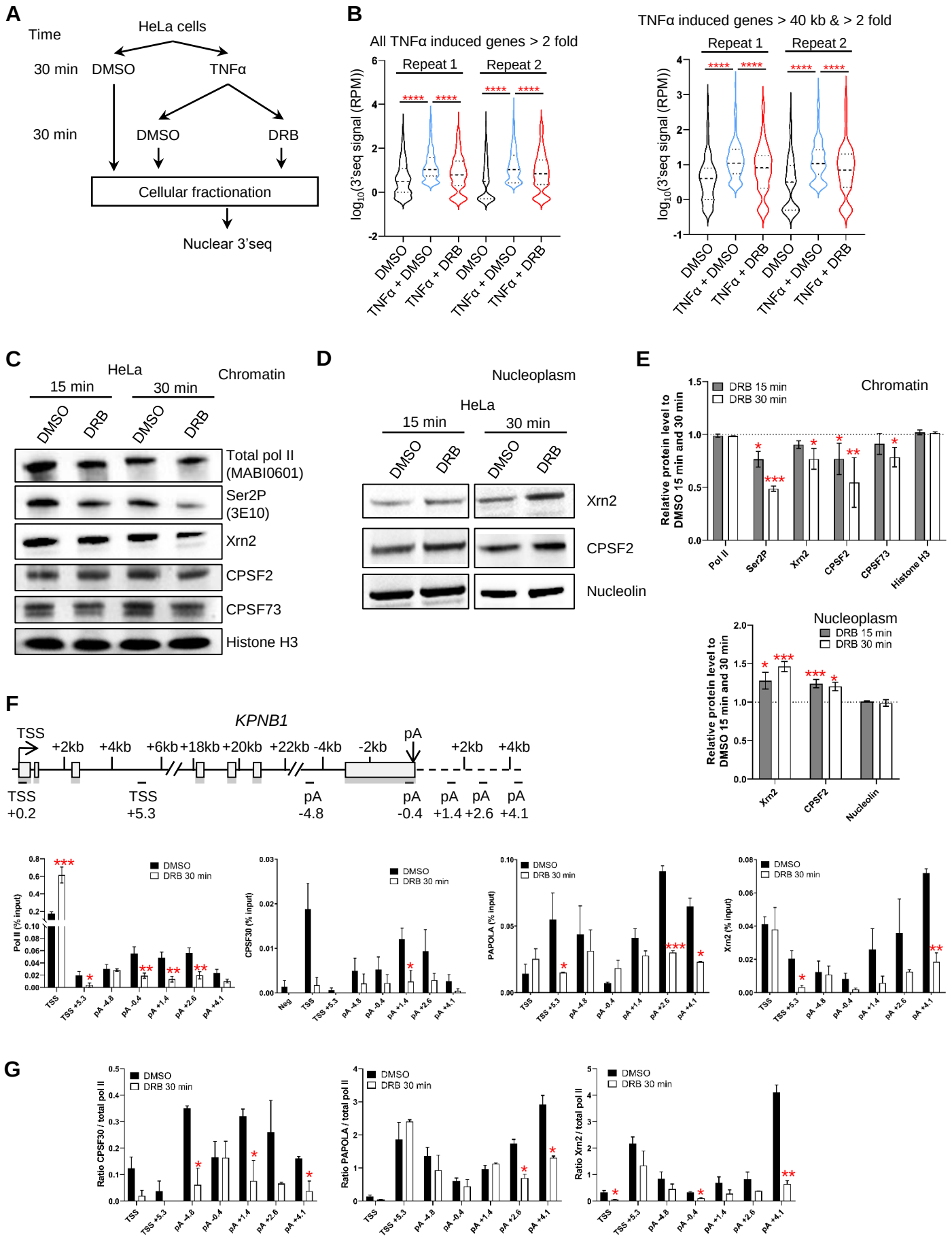

A
Time
HeLa cells
30 min
DMSO
TNFα
30 min
DMSO
DRB
Cellular fractionation
Nuclear 3'seq
B
All TNFα induced genes > 2 fold
Repeat 1
Repeat 2
log10(3'seq signal (RPM))
DMSO
TNFα + DMSO
TNFα + DRB
TNFα induced genes > 40 kb & > 2 fold
C
HeLa
Chromatin
15 min
30 min
DMSO
DRB
Total pol II (MABI0601)
Ser2P (3E10)
Xrn2
CPSF2
CPSF73
Histone H3
D
Nucleoplasm
HeLa
15 min
30 min
Xrn2
CPSF2
Nucleolin
E
Chromatin
DRB 15 min
DRB 30 min
Relative protein level to DMSO 15 min and 30 min
Pol II
Ser2P
Xrn2
CPSF2
CPSF73
Histone H3
Nucleoplasm
F
KPNB1
TSS
+2kb
+4kb
+6kb
+18kb
+20kb
+22kb
-4kb
-2kb
pA
TSS +0.2
TSS +5.3
pA -4.8
pA -0.4
pA +1.4
pA +2.6
pA +4.1
Pol II (% input)
CPSF30 (% input)
PAPOLA (% input)
Xrn2 (% input)
Neg
G
Ratio CPSF30 / total pol II
Ratio PAPOLA / total pol II
Ratio Xrn2 / total pol II

Figure 1

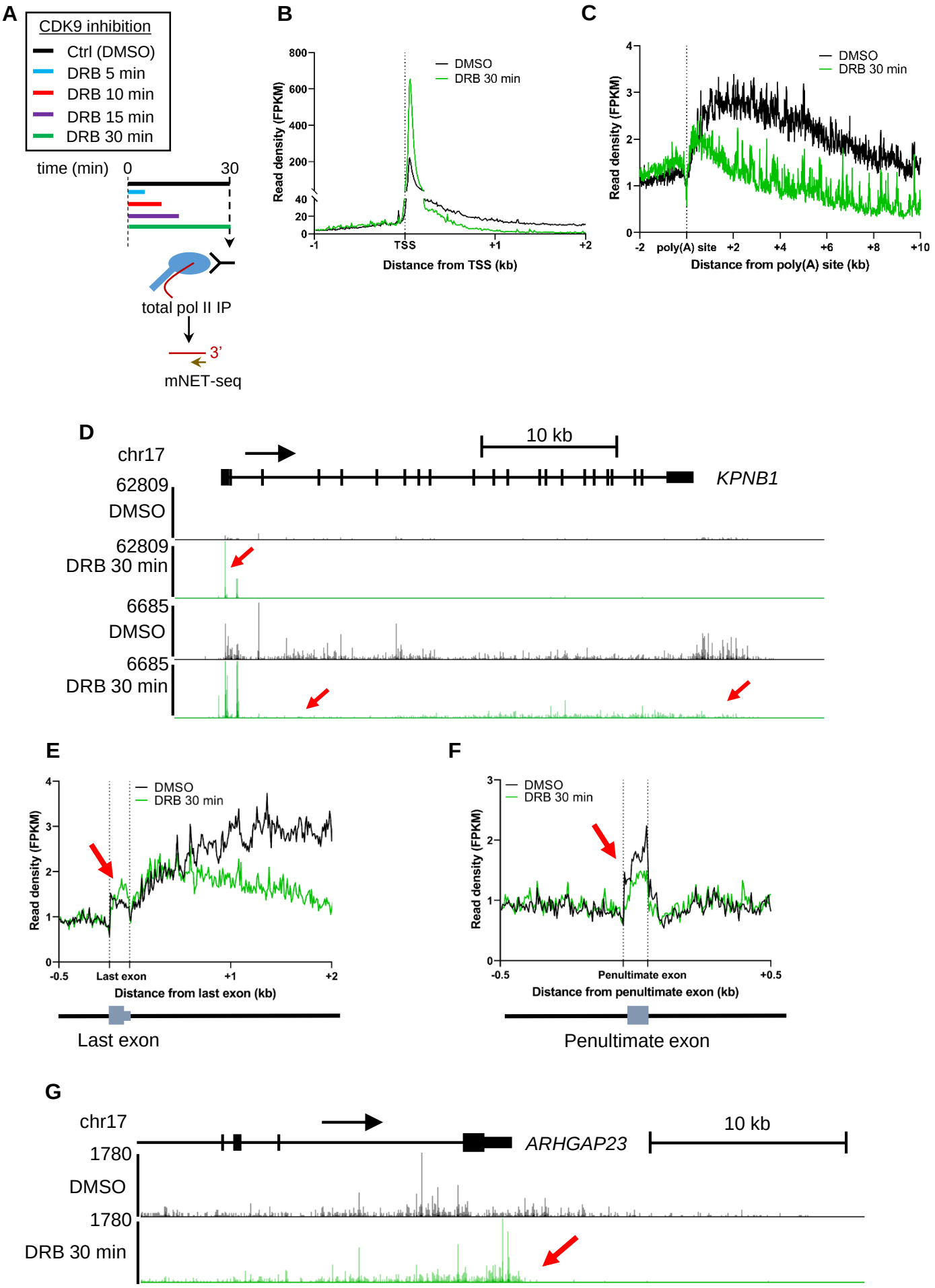

A
CDK9 inhibition
Ctrl (DMSO)
DRB 5 min
DRB 10 min
DRB 15 min
DRB 30 min
time (min)
0
30
total pol II IP
3'
mNET-seq
B
Read density (FPKM)
DMSO
DRB 30 min
TSS
Distance from TSS (kb)
C
Read density (FPKM)
DMSO
DRB 30 min
poly(A) site
Distance from poly(A) site (kb)
D
chr17
10 kb
KPNB1
62809
DMSO
62809
DRB 30 min
6685
DMSO
6685
DRB 30 min
E
Read density (FPKM)
DMSO
DRB 30 min
Last exon
Distance from last exon (kb)
Last exon
F
Read density (FPKM)
DMSO
DRB 30 min
Penultimate exon
Distance from penultimate exon (kb)
Penultimate exon
G
chr17
ARHGAP23
10 kb
1780
DMSO
1780
DRB 30 min

Figure 2

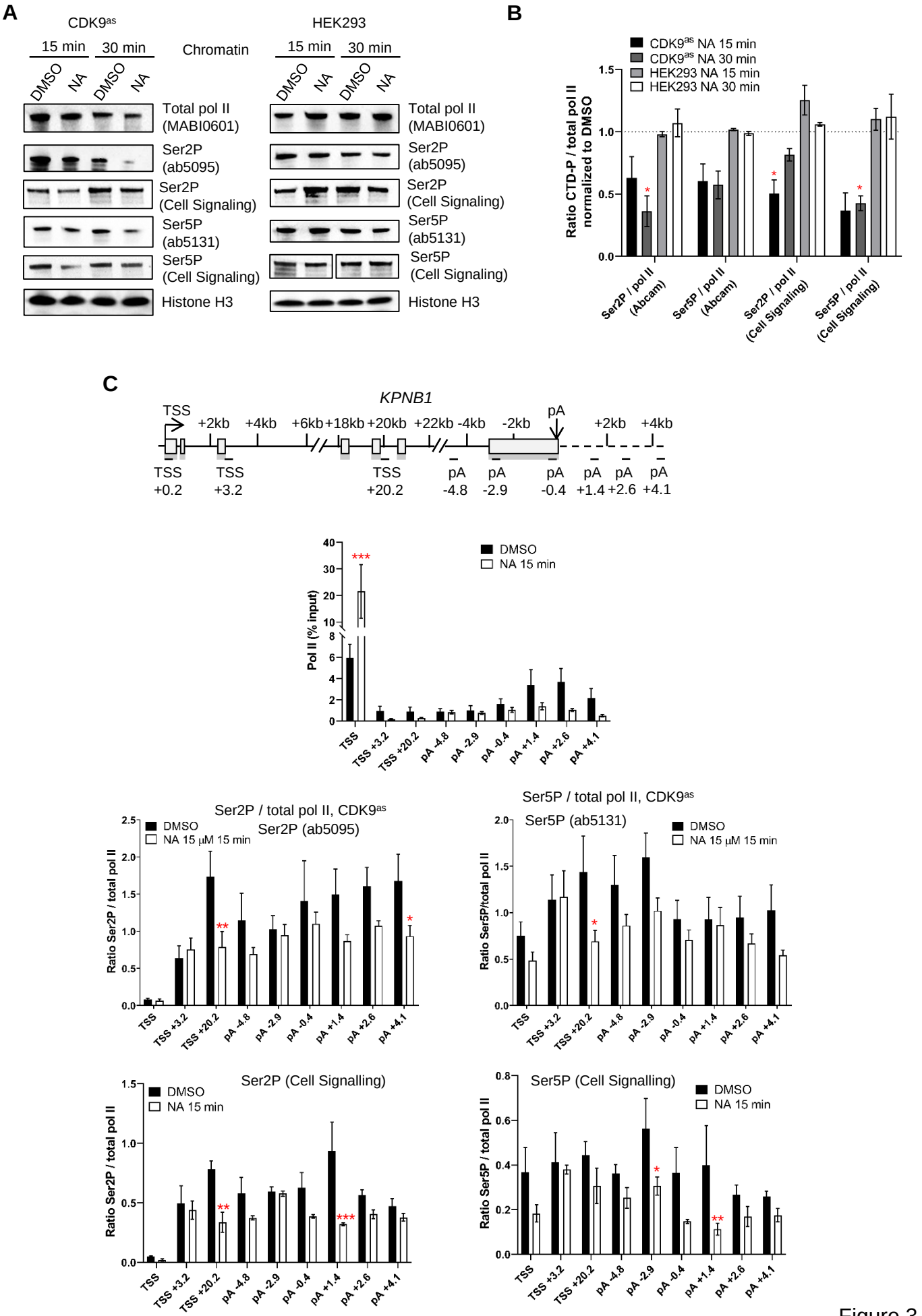
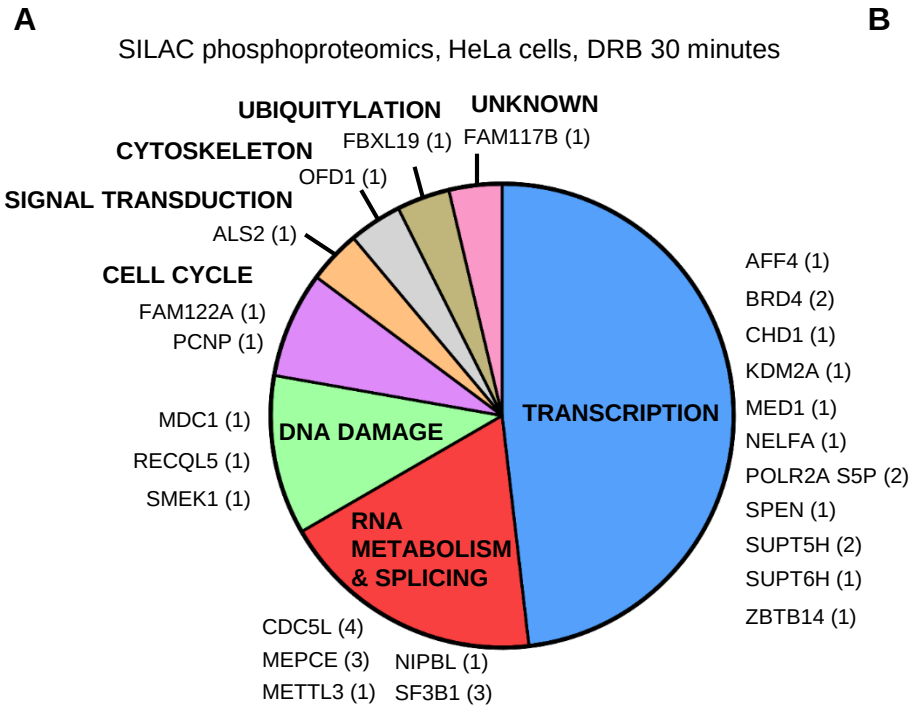
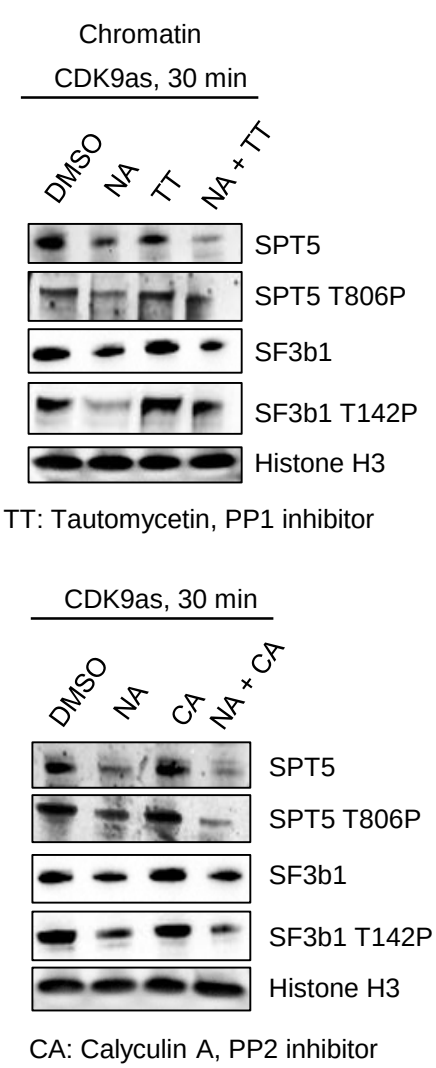
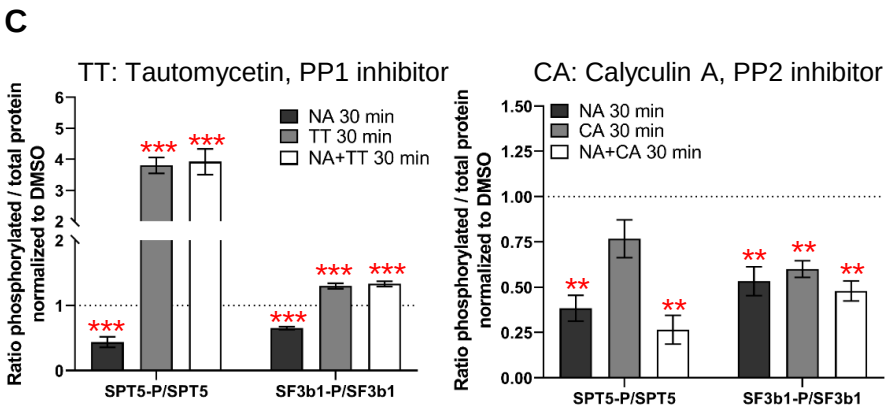

A
CDK9as
15 min
30 min
Chromatin
HEK293
DMSO
NA
Total pol II (MABI0601)
Ser2P (ab5095)
Ser2P (Cell Signaling)
Ser5P (ab5131)
Ser5P (Cell Signaling)
Histone H3
B
CDK9as NA 15 min
CDK9as NA 30 min
HEK293 NA 15 min
HEK293 NA 30 min
Ratio CTD-P / total pol II normalized to DMSO
Ser2P / pol II (Abcam)
Ser5P / pol II (Abcam)
Ser2P / pol II (Cell Signaling)
Ser5P / pol II (Cell Signaling)
C
KPNB1
TSS
pA
+2kb
+4kb
+6kb
+18kb
+20kb
+22kb
-4kb
-2kb
TSS +0.2
TSS +3.2
TSS +20.2
pA -4.8
pA -2.9
pA -0.4
pA +1.4
pA +2.6
pA +4.1
Pol II (% input)
DMSO
NA 15 min
Ser2P / total pol II, CDK9as
Ser2P (ab5095)
NA 15 μM 15 min
Ratio Ser2P / total pol II
Ser5P / total pol II, CDK9as
Ser5P (ab5131)
Ratio Ser5P/total pol II
Ser2P (Cell Signalling)
Ser5P (Cell Signalling)
Ratio Ser5P / total pol II

Figure 3







D

Protein name	Number of unique peptides DMSO (R1, R2)	Number of unique peptides NA (R1, R2)
SF3B1	6, 4	4, 3
SF3B2	10, 15	14, 13
SF3B5	4, 2	3, 3
SF3A1	6, 10	3, 6
SF3A2	2, 3	1, 2
SF3A3	3, 4	0, 1
SNW1	3, 3	0, 0
PRPF19	1, 3	0, 0
THO4	2, 4	1, 2

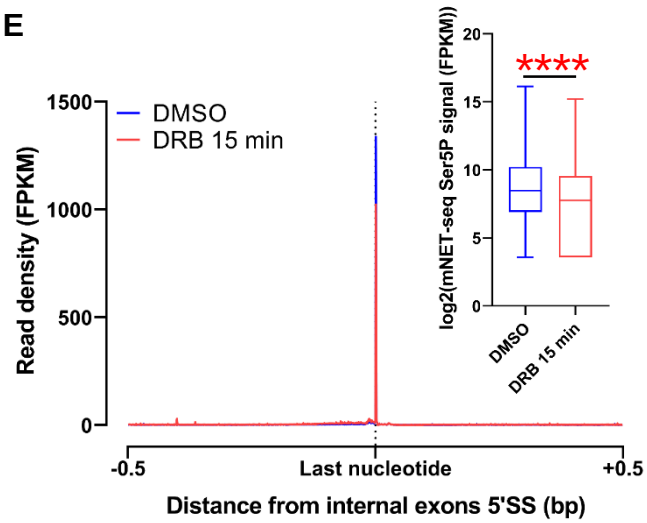


Figure 4

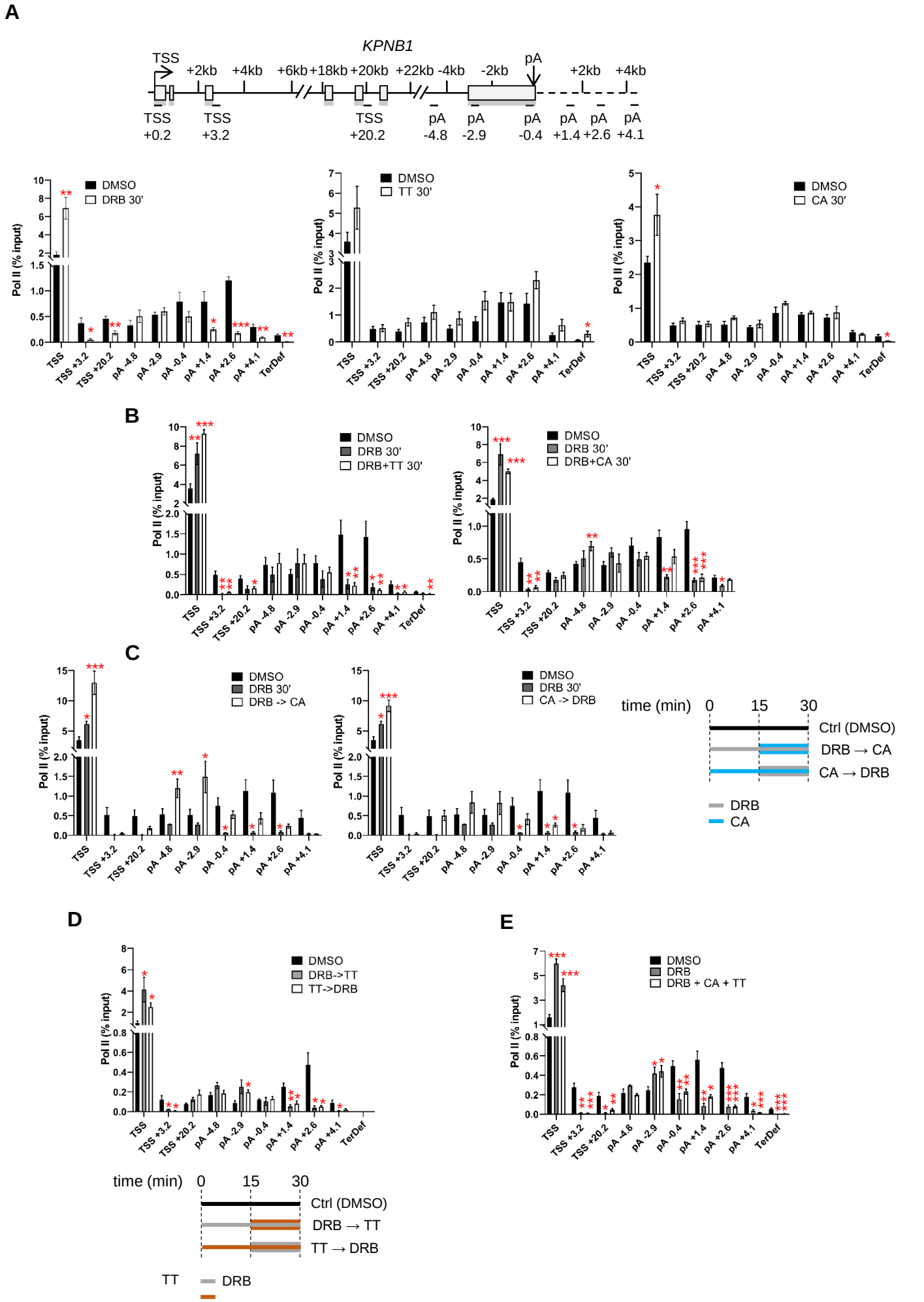
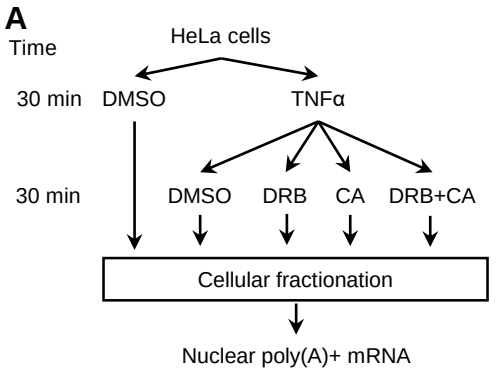
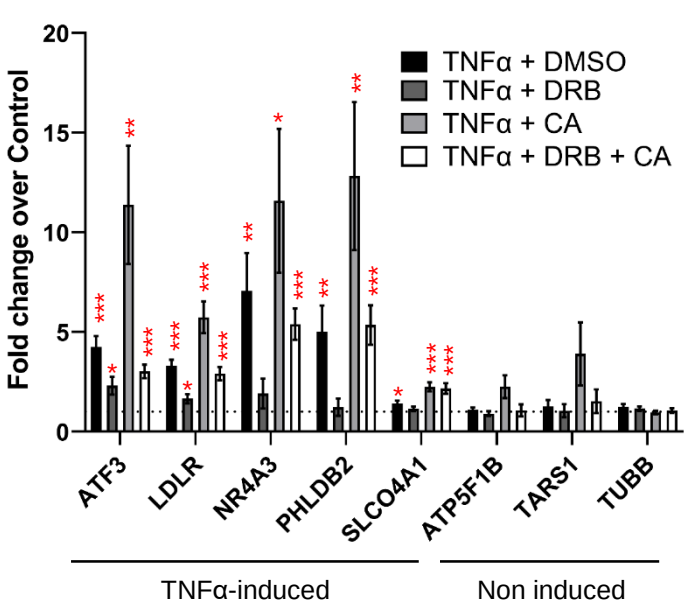
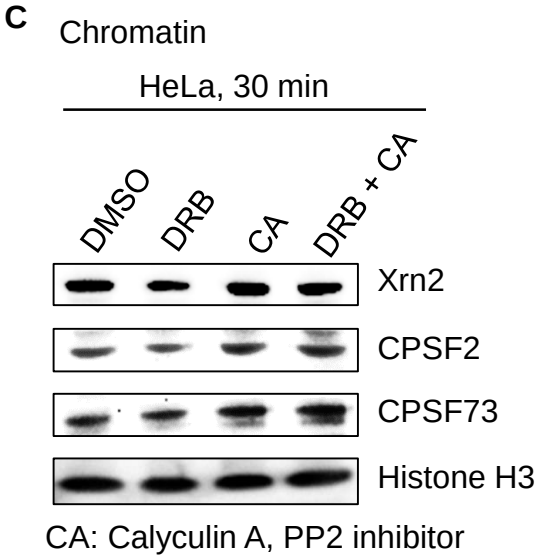
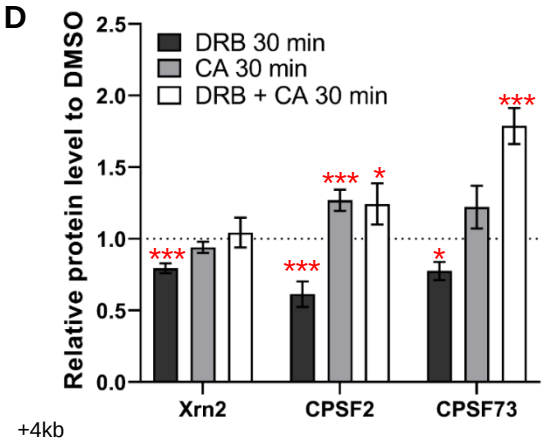

A
KPNB1
TSS
+2kb
+4kb
+6kb
+18kb
+20kb
+22kb
-4kb
-2kb
pA
+2kb
+4kb
TSS +0.2
TSS +3.2
TSS +20.2
pA -4.8
pA -2.9
pA -0.4
pA +1.4
pA +2.6
pA +4.1
Pol II (% input)
DMSO
DRB 30'
TT 30'
CA 30'
TSS
TSS +3.2
TSS +20.2
pA -4.8
pA -2.9
pA -0.4
pA +1.4
pA +2.6
pA +4.1
TerDef
B
DRB+TT 30'
DRB+CA 30'
C
DRB -> CA
CA -> DRB
time (min)
0
15
30
Ctrl (DMSO)
DRB → CA
CA → DRB
DRB
CA
D
DRB->TT
TT->DRB
TT → DRB
TT
E
DRB + CA + TT

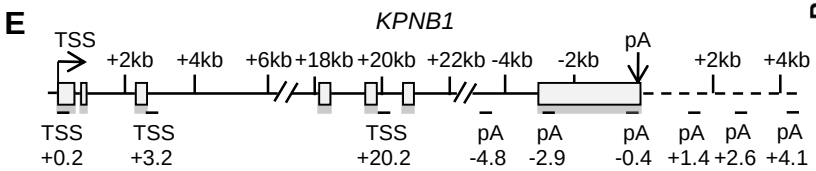
Figure 5

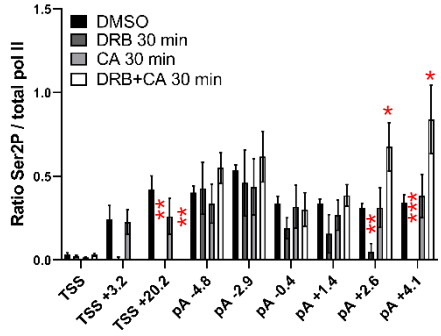


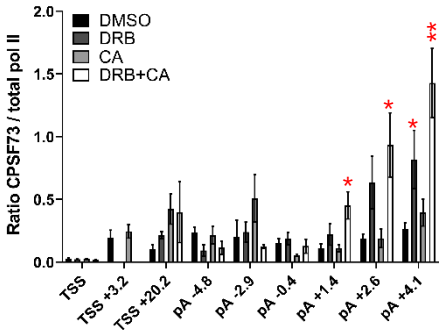


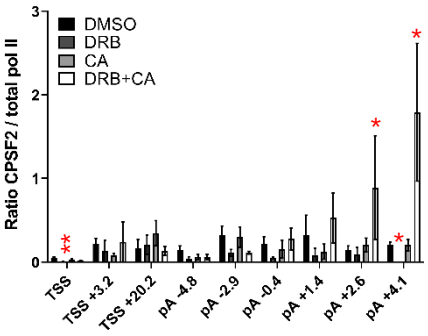


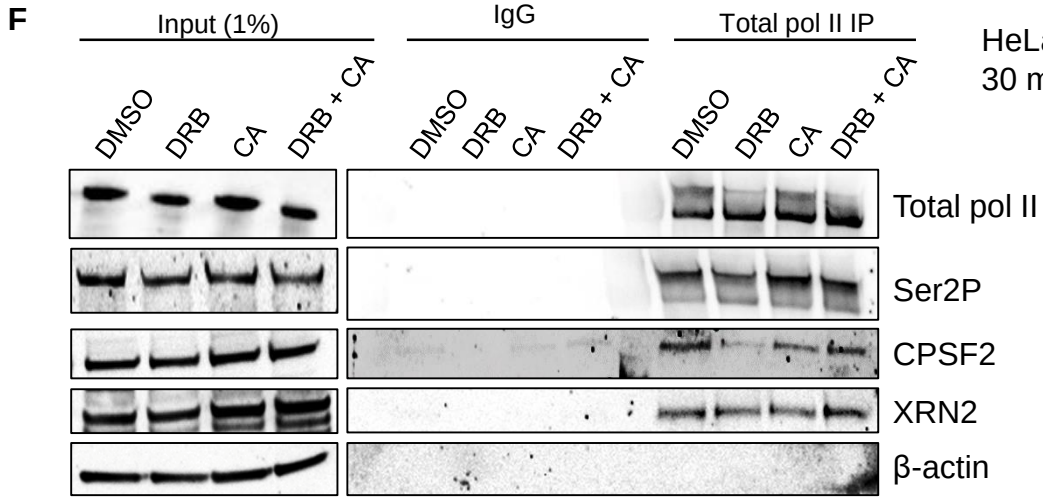












HeLa, co-IP, 30 minutes

Figure 6

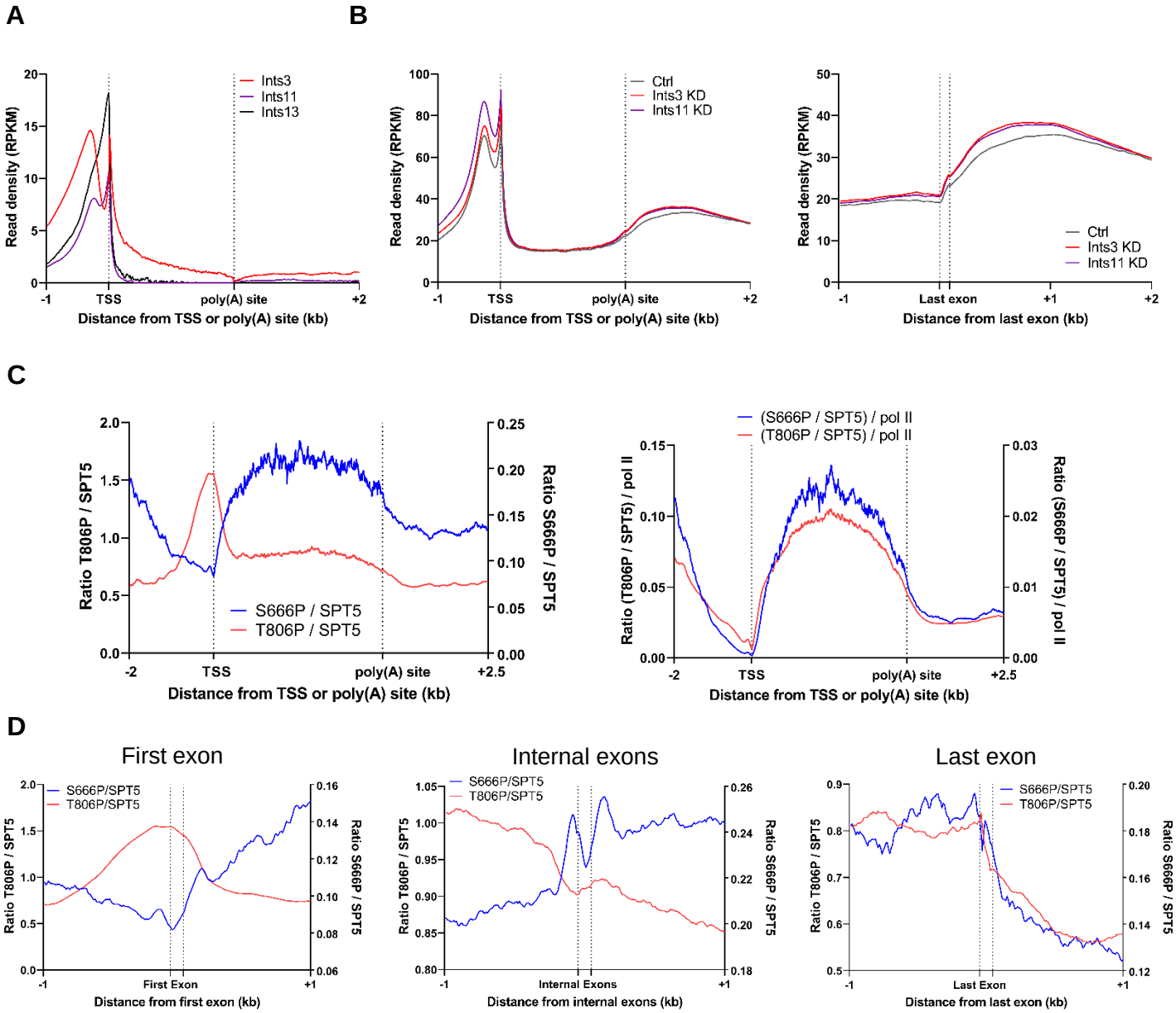

A
B
Ints3
Ints11
Ints13
Read density (RPKM)
TSS
poly(A) site
Distance from TSS or poly(A) site (kb)
Ctrl
Ints3 KD
Ints11 KD
Last exon
Distance from last exon (kb)
C
Ratio T806P / SPT5
Ratio S666P / SPT5
S666P / SPT5
T806P / SPT5
(S666P / SPT5) / pol II
(T806P / SPT5) / pol II
Ratio (T806P / SPT5) / pol II
Ratio (S666P / SPT5) / pol II
D
First exon
Internal exons
Last exon
S666P/SPT5
T806P/SPT5
First Exon
Distance from first exon (kb)
Internal Exons
Distance from internal exons (kb)
Last Exon
Distance from last exon (kb)

Figure 7

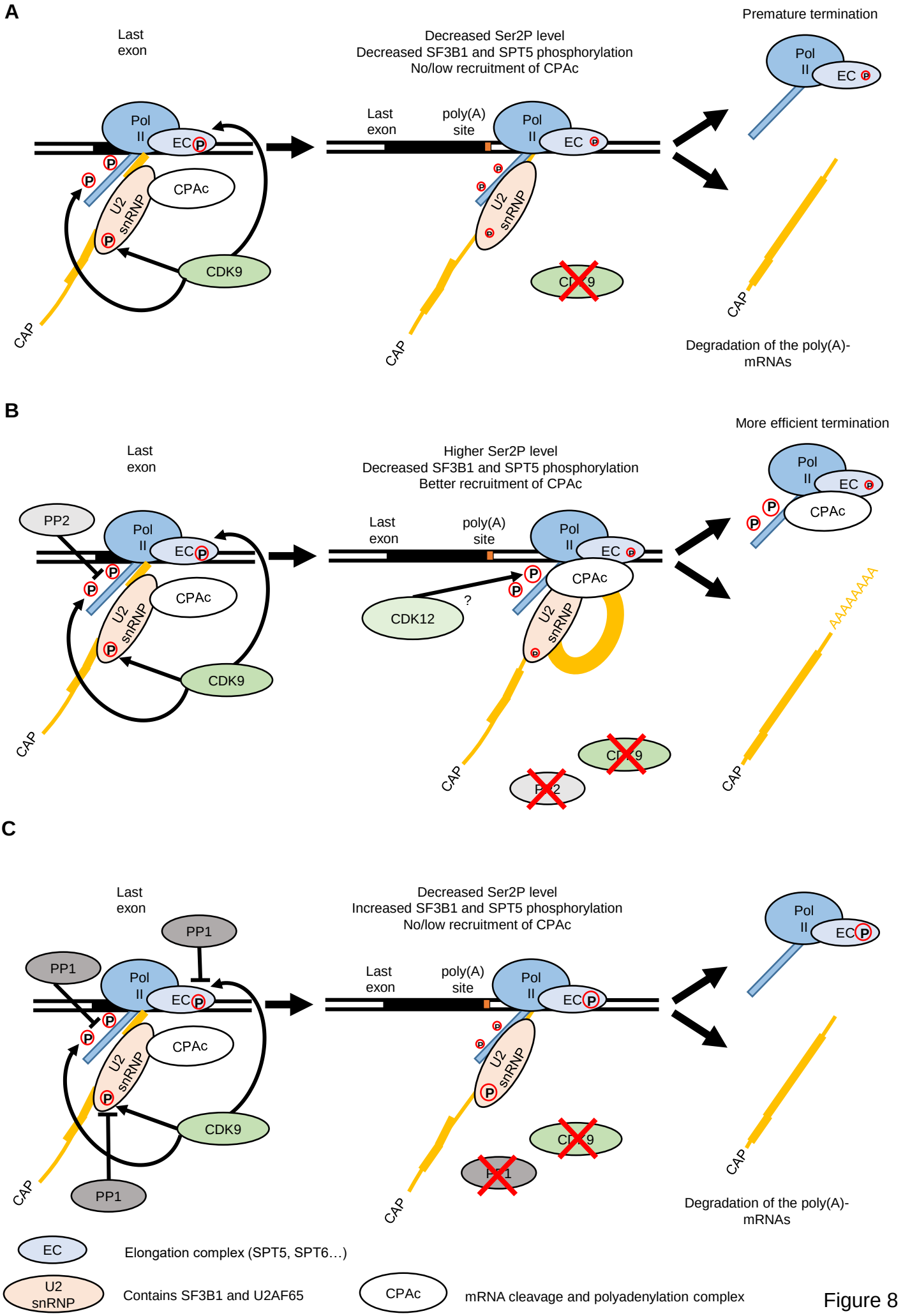

A
Last exon
Decreased Ser2P level
Decreased SF3B1 and SPT5 phosphorylation
No/low recruitment of CPAc
Premature termination
Pol II
EC
CPAc
U2 snRNP
CDK9
CAP
Last exon
poly(A) site
Degradation of the poly(A)-mRNAs
B
Last exon
Higher Ser2P level
Decreased SF3B1 and SPT5 phosphorylation
Better recruitment of CPAc
More efficient termination
PP2
CDK12
?
AAAAAAAA
C
Last exon
Decreased Ser2P level
Increased SF3B1 and SPT5 phosphorylation
No/low recruitment of CPAc
PP1
Degradation of the poly(A)-mRNAs
EC
Elongation complex (SPT5, SPT6…)
U2 snRNP
Contains SF3B1 and U2AF65
CPAc
mRNA cleavage and polyadenylation complex

Figure 8