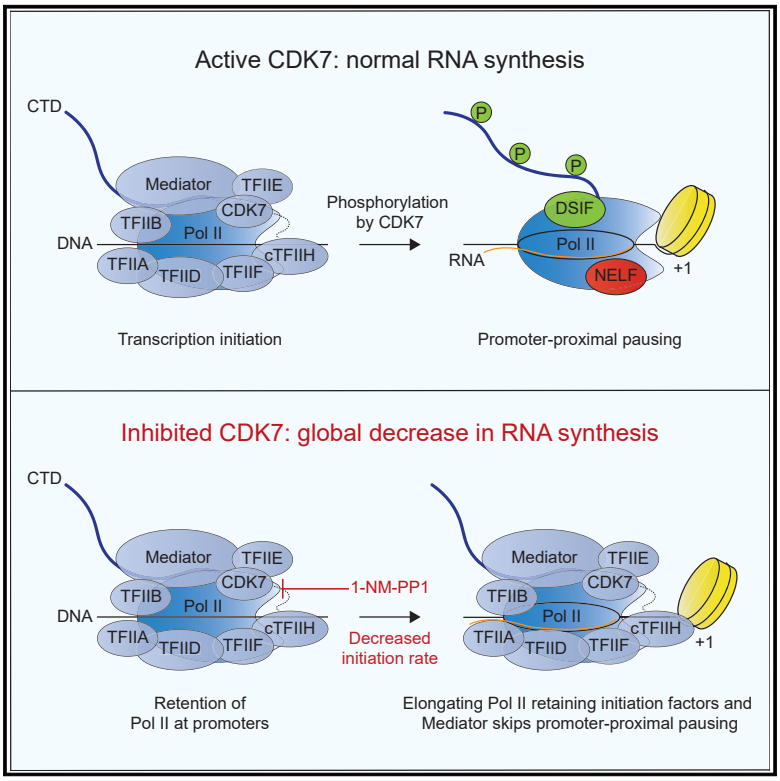
Molecular Cell

Article

CDK7 kinase activity promotes RNA polymerase II promoter escape by facilitating initiation factor release

Graphical abstract



Highlights

- CDK7 inhibition leads to an immediate global decrease in RNA Pol II transcript synthesis
- RNA Pol II, initiation factors, and Mediator accumulate at promoters
- Elongation, termination, and co-transcriptional factor recruitment remain unaffected
- CDK7 promotes RNA Pol II promoter escape by releasing initiation factors and Mediator

Authors

Taras Velychko, Eusra Mohammad, Ivan Ferrer-Vicens, ..., Shona Murphy, Patrick Cramer, Michael Lidschreiber

Correspondence

patrick.cramer@mpinat.mpg.de (P.C.), michael.lidschreiber@mpinat.mpg.de (M.L.)

In brief

Velychko et al. show that CDK7 kinase activity is required for the efficient release of initiation factors and Mediator from RNA polymerase II (RNA Pol II), facilitating RNA Pol II release from the promoter region. This step is crucial for robust RNA synthesis of all genes and thus cellular activity.

Velychko et al., 2024, Molecular Cell *84*, 1–17
June 20, 2024
https://doi.org/10.1016/j.molcel.2024.05.007

CellPress

Molecular Cell

CellPress
OPEN ACCESS

Article

CDK7 kinase activity promotes RNA polymerase II promoter escape by facilitating initiation factor release

Taras Velychko,[1] Eusra Mohammad,[1] Ivan Ferrer-Vicens,[2,5] Iwan Parfentev,[3] Marcel Werner,[1] Cecilia Studniarek,[2] Björn Schwalb,[1] Henning Urlaub,[3,4] Shona Murphy,[2] Patrick Cramer,[1,*] and Michael Lidschreiber[1,6,*]

[1]Department of Molecular Biology, Max Planck Institute for Multidisciplinary Sciences, Am Fassberg 11, 37077 Göttingen, Germany
[2]Sir William Dunn School of Pathology, University of Oxford, Oxford OX1 3RE, UK
[3]Bioanalytical Mass Spectrometry, Max Planck Institute for Multidisciplinary Sciences, Am Fassberg 11, 37077 Göttingen, Germany
[4]Institute of Clinical Chemistry, University Medical Center Göttingen, 37075 Göttingen, Germany
[5]Present address: School of Biochemistry, University of Bristol, Bristol BS8 1TD, UK
[6]Lead contact
*Correspondence: patrick.cramer@mpinat.mpg.de (P.C.), michael.lidschreiber@mpinat.mpg.de (M.L.)
https://doi.org/10.1016/j.molcel.2024.05.007

SUMMARY

Cyclin-dependent kinase 7 (CDK7), part of the general transcription factor TFIIH, promotes gene transcription by phosphorylating the C-terminal domain of RNA polymerase II (RNA Pol II). Here, we combine rapid CDK7 kinase inhibition with multi-omics analysis to unravel the direct functions of CDK7 in human cells. CDK7 inhibition causes RNA Pol II retention at promoters, leading to decreased RNA Pol II initiation and immediate global downregulation of transcript synthesis. Elongation, termination, and recruitment of co-transcriptional factors are not directly affected. Although RNA Pol II, initiation factors, and Mediator accumulate at promoters, RNA Pol II complexes can also proceed into gene bodies without promoter-proximal pausing while retaining initiation factors and Mediator. Further downstream, RNA Pol II phosphorylation increases and initiation factors and Mediator are released, allowing recruitment of elongation factors and an increase in RNA Pol II elongation velocity. Collectively, CDK7 kinase activity promotes the release of initiation factors and Mediator from RNA Pol II, facilitating RNA Pol II escape from the promoter.

INTRODUCTION

Phosphorylation of the C-terminal repeat domain (CTD) of RNA polymerase II (RNA Pol II) plays a central role in regulating gene transcription in eukaryotes.[1,2] In yeast and mammals, the CTD is highly conserved and consists of 26 and 52 repeats of the consensus heptapeptide sequence ($Y_1S_2P_3T_4S_5P_6S_7$), respectively.[3] Changes in RNA Pol II CTD phosphorylation patterns regulate the transcription cycle via recruitment of transcription and RNA processing factors.[4,5] Cyclin-dependent kinase 7 (CDK7) phosphorylates Ser5 and Ser7 of the CTD at the beginning of transcription.[6–9] This is consistent with CDK7 activity as a part of the general transcription factor TFIIH,[10] which together with TFIIA, TFIIB, TFIID, TFIIE, and TFIIF assembles into a pre-initiation complex (PIC) at active gene promoters.[11,12] After PIC assembly and promoter melting, RNA Pol II escapes from the promoter and proceeds into elongation.[13] How the initiation complex is converted into an elongation complex remains only partially understood.

In yeast, the CDK7 ortholog Kin28 facilitates the dissociation of RNA Pol II from Mediator[14,15] and allows RNA Pol II to escape from the promoter.[16,17] However, studies comparing temperature- versus analog-sensitive mutants of Kin28 revealed that the function of core-TFIIH, rather than the kinase activity, is required for global mRNA synthesis and promoter escape.[18,19] In a more recent study, irreversible inhibition of Kin28 resulted in a variable effect on promoter release but led to the accumulation of RNA Pol II at the +2 nucleosome, due to a decreased transition to productive elongation.[20] Furthermore, Kin28 regulates transcription start site (TSS) selection through RNA Pol II promoter scanning in budding yeast,[21] a process that is not observed in metazoans and is technically challenging to distinguish from promoter escape. Thus, it is unclear which steps of RNA Pol II transcription require Kin28 kinase activity in yeast.

The function of human CDK7 differs from that in budding yeast. First, the CDK-activating kinase (CAK), comprising CDK7/Cyclin H/MAT1, can dissociate from TFIIH and phosphorylate other CDKs to promote cell-cycle transition.[22] Second, promoter-proximal pausing regulates the transcription initiation rate in metazoans,[23,24] and CDK7 is involved in the regulation of this process. CDK7 imposes the promoter-proximal pause by facilitating the recruitment of pausing factors and triggers the

Please cite this article in press as: Velychko et al., CDK7 kinase activity promotes RNA polymerase II promoter escape by facilitating initiation factor release, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.05.007

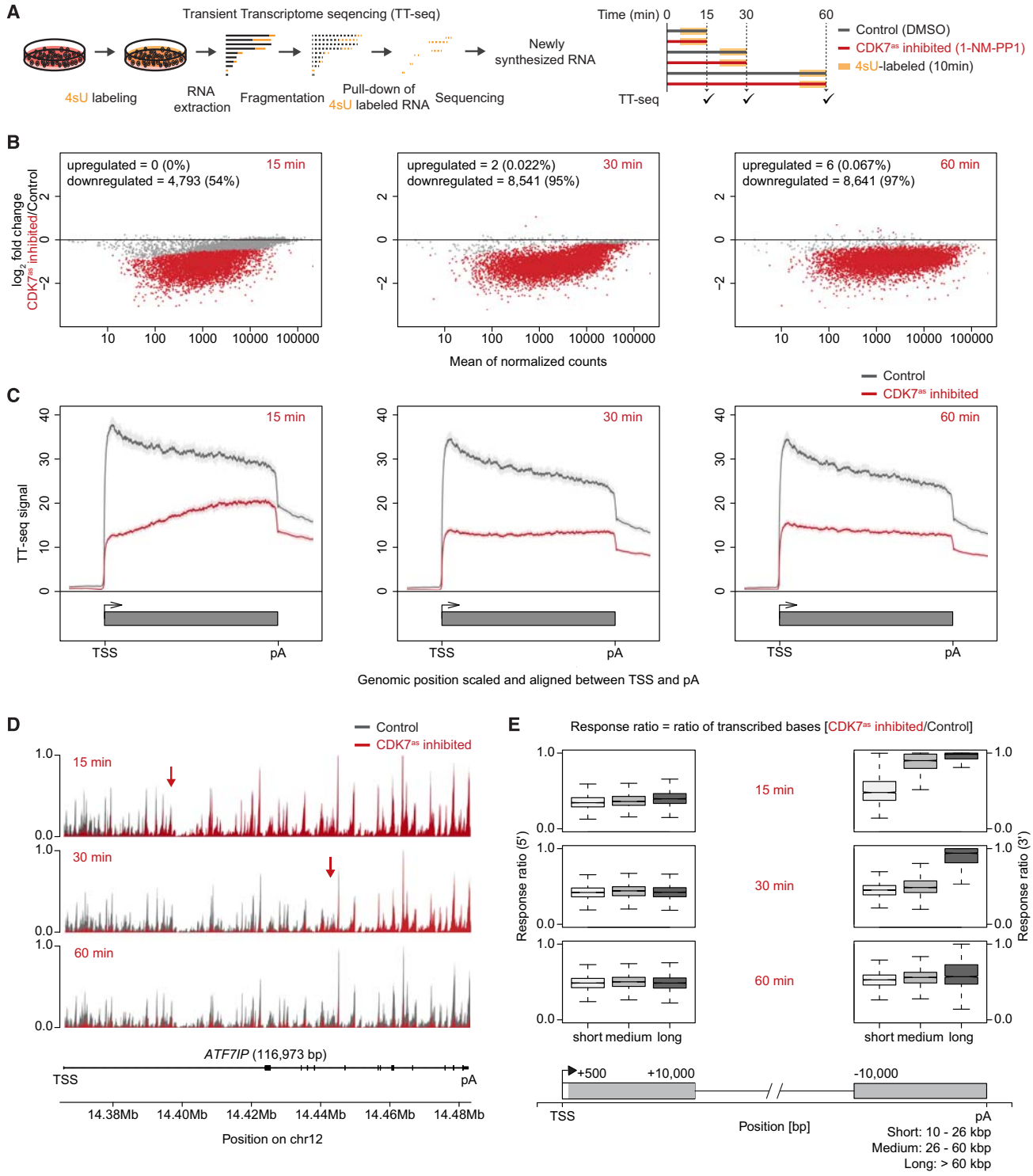


Figure 1. Rapid CDK7 kinase inhibition results in global downregulation of new RNA synthesis

(A) Schematic overview of TT-seq.

(B) MA plots showing changes in RNA synthesis detected by TT-seq upon CDK7 inhibition for 15 (left), 30 (middle), and 60 (right) min. Significantly upregulated or downregulated genes (adjusted $p < 0.05$) are marked in red.

(C) Metagene analysis of TT-seq signal of the expressed genes before (control, gray) and after (red) CDK7 inhibition for 15 (left), 30 (middle), and 60 (right) min. TT-seq signal from two biological replicates was merged and aligned at the TSS and polyadenylation sites. The shaded areas around the average signal (solid lines) indicate 95% confidence intervals of the mean.

(legend continued on next page)

subsequent release from pausing through activation of the positive transcription elongation factor b (P-TEFb: CDK9 and Cyclin T).[7,25,26] How CDK7 coordinates the two opposing functions of RNA Pol II pausing and release remains uncertain.[27]

In addition, CDK7 inhibition affects most co-transcriptional processes. For instance, CDK7 inhibition impairs the recruitment of pre-mRNA 5′ capping enzymes in yeasts[19,28–30] and humans[26,31,32]; CDK7 activity regulates splicing, and its inhibition results in alternative exon inclusion and intron retention[33]; CDK7 inhibition affects histone modifications by reducing H3K4me3 spreading and redistributing H3K36me3 toward gene 3′ ends[32]; and finally, disruption of CDK7 activity impairs downstream transcription steps, such as RNA Pol II elongation,[34] mRNA 3′ end formation, and termination.[7,25,32,35] However, rapid perturbations combined with the time-resolved information of transcription activity[36] are required to distinguish the primary function of CDK7 from the potential indirect effects that occur downstream of CDK7 activity.

To reveal the direct function of CDK7 kinase activity, we used an analog-sensitive CDK7 (CDK7^{as}) cell line that allows rapid and highly selective inhibition of the CDK7 kinase activity. We combined rapid inhibition of CDK7 with a multi-omics approach to measure the immediate changes in RNA synthesis via transient transcriptome sequencing (TT-seq),[37] changes in RNA Pol II occupancy with mammalian native elongating transcript sequencing (mNET-seq)[38,39] and high-resolution micrococcal nucleasechromatin immunoprecipitation sequencing (MNase-ChIP-seq),[40] and changes in transcription factor chromatin binding with tandem mass tag (TMT)-based quantitative proteomics. Multi-omics data from a time course of short CDK7 inhibitions (15, 30, and 60 min) allowed us to dissect the primary role of CDK7 from the secondary effects of transcription perturbation. Our findings provide insight into the mechanism by which CDK7 kinase activity promotes RNA Pol II promoter escape, particularly through facilitating the release of initiation factors and Mediator.

RESULTS

Rapid CDK7 kinase inhibition results in global downregulation of new RNA synthesis

To investigate the direct function of CDK7 kinase activity, we generated a HEK293 CDK7^{as} cell line by mutating the bulky phenylalanine residue in the ATP-binding site to a smaller glycine (Figure S1A). This mutation allows for selective CDK7 inhibition with the ATP-competitive analog 1-NM-PP1.[41] 1-NM-PP1 treatment decreased the proliferation rate of CDK7^{as} cells and had no effect on the HEK293 wild-type (WT) cell line (Figures S1B and S1C). This agrees with previously published data in which CDK7^{as} inhibition with 1-NM-PP1 reduced the proliferation and cell viability of human colon cancer cells.[42]

To study changes in cellular RNA synthesis, we performed spike-in-normalized TT-seq before and after CDK7 inhibition. CDK7^{as} cells were treated with DMSO (solvent control) or 1-NM-PP1 for either 15, 30, or 60 min with 4-thiouridine (4sU) included for the final 10 min (Figures 1A and S2A). While 1-NM-PP1 treatment did not affect HEK293 WT cells (Figure S1D), CDK7 inhibition resulted in a global downregulation of newly synthesized RNA at all time points (Figure 1B). The percentage of significantly downregulated genes increased from 54% after 15 min to 97% after 60 min of inhibition. As an orthogonal approach, we performed a 5-ethynyl uridine (EU) incorporation assay, which confirmed the global downregulation of RNA synthesis observed with spike-in-normalized TT-seq (Figure S3A). TT-seq metagene profiles showed a more substantial decrease in the signal at the 5′ end compared with the 3′ end of the genes, after inhibition of CDK7 for 15 min, but a stable decrease in TT-seq signal across the gene bodies for 30 and 60 min (Figure 1C). Furthermore, for long genes such as *ATF7IP*, downregulation of RNA synthesis was only observed within ~30 kbp of the TSS after 15 min CDK7 inhibition (Figure 1D). This downregulation extended to ~70 kbp after 30 min inhibition and was uniform across the gene body after 60 min. To further investigate this, we calculated CDK7 inhibition response ratios by comparing the changes in the TT-seq signal at the 5′ and 3′ ends of three gene quantiles, based on their length: short (10–26 kbp), medium (26–60 kbp), and long (>60 kbp) genes (Figures 1E and S2B; STAR Methods). The obtained lengths of the gene groups correspond to the potential distances transcribed by RNA Pol II in 15, 30, and 60 min, based on the previously estimated RNA Pol II elongation velocity of ~2 kbp/min.[24] The 5′ end response ratios followed a narrow distribution, with a median reduction in RNA synthesis by more than 50%, irrespective of gene length and CDK7 inhibition time, confirming defective RNA synthesis at the beginning of all protein-coding genes (Figures 1E and S2C). For 98% of genes, 5′ end RNA synthesis was reduced by at least 30%. Synthesis of different types of non-coding RNAs, including enhancer RNAs, was also downregulated (Figure S2D).

In contrast to gene 5′ ends, the 3′ ends of long genes were largely unaffected after 15 min of CDK7 inhibition (Figures 1E and S2C), as well as transcription past polyadenylation sites (Figure S2E). Thus, RNA Pol II molecules that proceeded to productive elongation prior to CDK7 inhibition continued RNA synthesis, largely without elongation impairment. In contrast to conclusions from previous studies using longer and/or non-selective inhibitions,[7,25,32,35] our results indicate that CDK7 does not directly regulate RNA Pol II transcription elongation and termination.

In conclusion, TT-seq analysis shows that CDK7 is a global regulator of RNA Pol II transcription, and its inhibition affects all RNA Pol II-transcribed genes.

(D) TT-seq signal before (gray) and after (red) CDK7 inhibition for 15 (upper), 30 (middle), and 60 (lower) min of the *ATF7IP* gene (116,937 bp length). Arrows indicate the end of the regions affected by CDK7 inhibition. Two biological replicates were merged.

(E) Response ratios of RNA synthesis for the 5′ end (left) and 3′ end (right) of the expressed genes for 15 (top), 30 (middle), and 60 (bottom) min inhibition of CDK7 compared to control. The genes were split into three (n = 2,307) quantiles according to their length. Box limits are the first and third quartiles, the band inside the box is the median. The ends of the whiskers extend the box by 1.5 times the interquartile range. Notches represent 95% confidence intervals for the median values. Outliers not shown.

Please cite this article in press as: Velychko et al., CDK7 kinase activity promotes RNA polymerase II promoter escape by facilitating initiation factor release, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.05.007

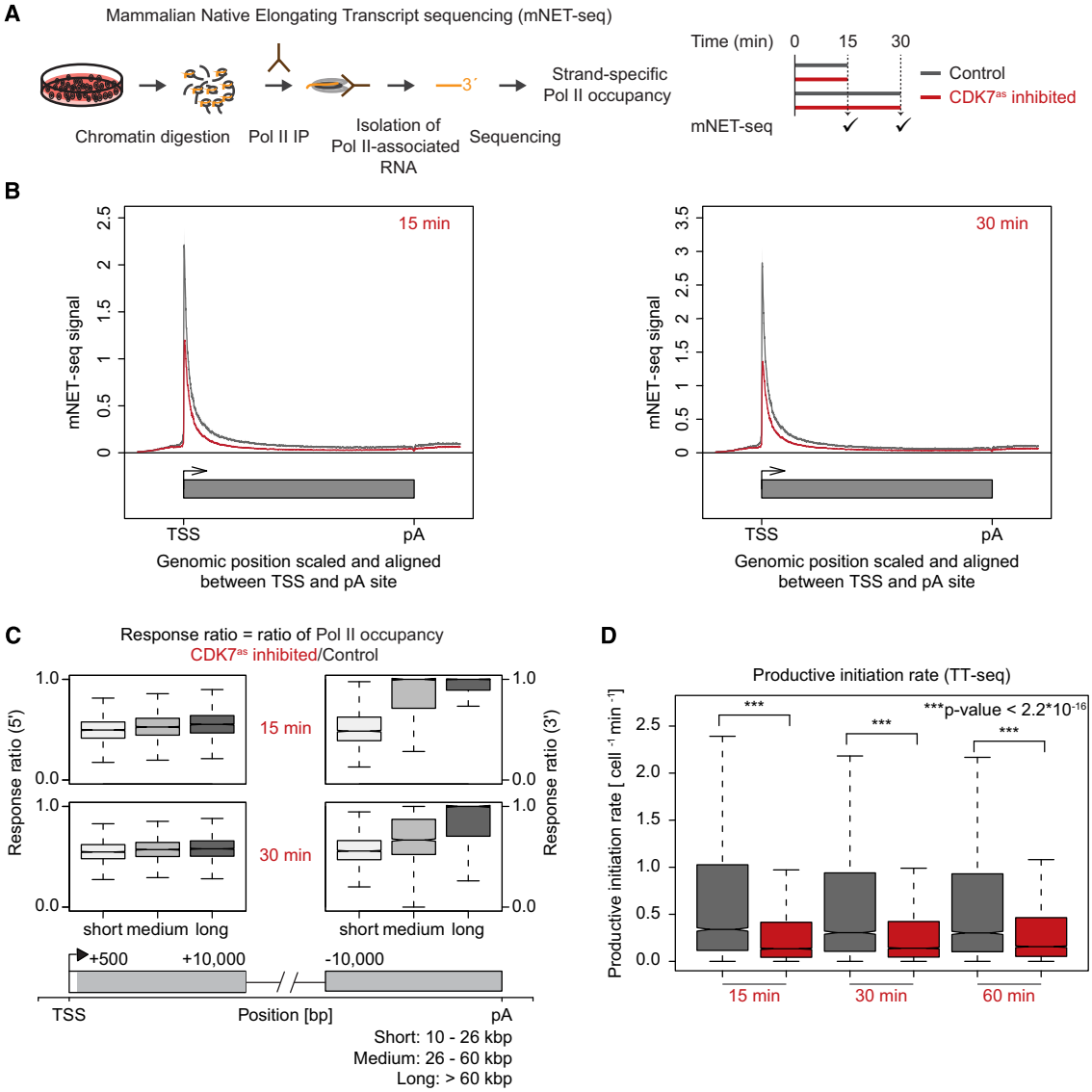


Figure 2. CDK7 inhibition results in decreased RNA Pol II gene occupancy due to decreased initiation rate

(A) Schematic overview of mNET-seq.

(B) Metagene analysis of mNET-seq signal of the expressed genes before (control, gray) and after (red) CDK7 inhibition for 15 (left) and 30 (right) min. mNET-seq signal from two biological replicates was merged and aligned at the TSS and polyadenylation sites. The shaded areas around the average signal (solid lines) indicate 95% confidence intervals of the mean.

(C) Response ratios of RNA Pol II occupancy for the 5′ end (left) and 3′ end (right) of the expressed genes for 15 (top) and 30 (bottom) min inhibition of CDK7 compared to control. Gene groups and boxplot representations as in Figure 1E.

(D) Productive initiation rates (n = 6,247, STAR Methods) before (control; gray) and after (red) CDK7 inhibition for 15 (left), 30 (middle), and 60 (right) min. Boxplot representations as in Figure 1E. p values by two-sided Mann-Whitney U test.

CDK7 inhibition results in decreased RNA Pol II gene occupancy

The global reduction observed in the TT-seq signal at the beginning of genes after CDK7 inhibition may result from a direct decrease in initiation rate and/or an increase in the duration of promoter-proximal pausing that can also lead to a decrease in initiation rate.[23,24,43] To distinguish between these two possibilities, we performed spike-in-normalized mNET-seq, which provides genome-wide occupancy of transcriptionally engaged RNA Pol II, before and after CDK7 inhibition for 15 and 30 min (Figures 2A and S3B). Metagene analysis of the mNET-seq signals showed decreased RNA Pol II occupancy in the pause region and across the gene body for both time points (Figures 2B and S3C). In contrast, inhibition of CDK9, which facilitates promoter-proximal pause release, resulted in a global decrease in the RNA synthesis rate due to RNA Pol

Please cite this article in press as: Velychko et al., CDK7 kinase activity promotes RNA polymerase II promoter escape by facilitating initiation factor release, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.05.007

CellPress
OPEN ACCESS

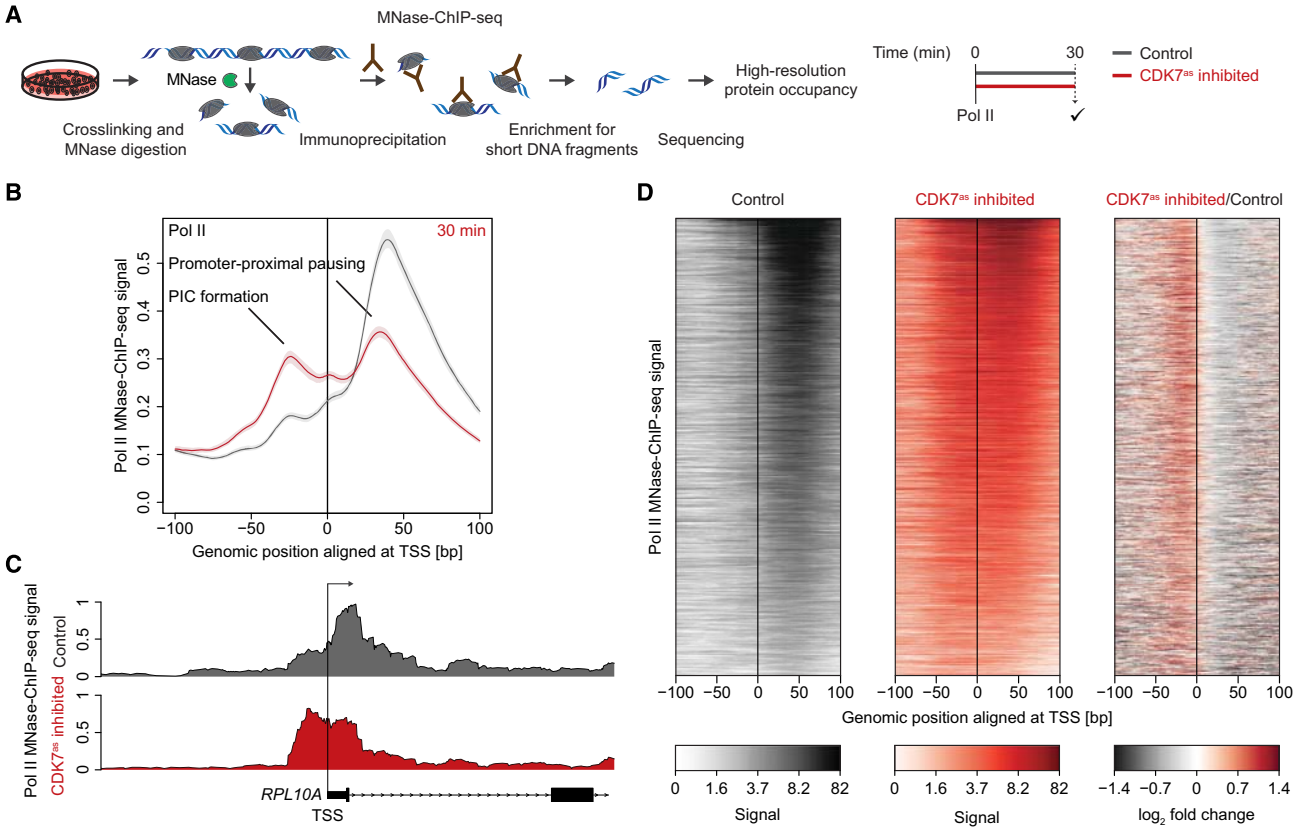


Figure 3. CDK7 inhibition leads to RNA Pol II accumulation upstream of the transcription start site

(A) Schematic overview of MNase-ChIP-seq.

(B) Metagene analysis of the RNA Pol II (RPB1) MNase-ChIP-seq signal of the expressed genes before (control; gray) and after (red) CDK7 inhibition for 30 min. MNase-ChIP-seq signal from three biological replicates was merged and aligned at the TSS. The shaded areas around the average signal (solid lines) indicate 95% confidence intervals of the mean.

(C) RNA Pol II MNase-ChIP-seq signal around the TSS of the *RPL10A* gene before (control; gray) and after (red) CDK7 inhibition for 30 min. Three biological replicates were merged.

(D) Heatmaps of RNA Pol II occupancy based on RNA Pol II MNase-ChIP-seq signal before (left) and after (middle) CDK7 inhibition for 30 min and changes in occupancy (right). The expressed genes are aligned at the TSS and sorted by the RNA Pol II signal in the control sample's pause window (TSS to TSS + 100 bp).

II accumulation in the pause region.[24,43] The response ratios of RNA Pol II occupancy derived from mNET-seq (Figure 2C) showed the same pattern as the response ratios of RNA synthesis from TT-seq (Figure 1E), with 5′ ends affected equally and 3′ ends disproportionally between subgroups of genes with different gene lengths. Similar drops in TT-seq and mNET-seq signals, as well as a clear decrease in RNA Pol II occupancy in the pause region, indicate that the major direct transcriptional defect upon CDK7 inhibition occurs upstream of the RNA Pol II promoter-proximal pause site.

Next, we calculated the productive initiation rate based on TT-seq, which corresponds to the number of polymerases that initiated and successfully transitioned into productive elongation.[24,43] We observed a decrease in productive initiation rate at all time points of CDK7 inhibition (Figure 2D). Together with the observed decrease in RNA Pol II occupancy in the pause region upon CDK7 inhibition, the decrease in productive initiation rate likely reflects the direct decrease in RNA Pol II initiation rate.

CDK7 inhibition leads to RNA Pol II accumulation upstream of the TSS

Since mNET-seq does not provide information on RNA Pol II occupancy upstream of the TSS, we performed high-resolution MNase-ChIP-seq of RNA Pol II after 30 min of CDK7 inhibition (Figures 3A and S4A). MNase-ChIP-seq was normalized to the signal at non-transcribed regions to allow for the detection of global changes (STAR Methods). Short fragment selection (<70 bp) for MNase-ChIP-seq allowed us to distinguish paused RNA Pol II (typically 20–50 bp downstream of the TSS[44]) from the RNA Pol II PIC upstream of the TSS and from upstream antisense RNA Pol II PIC (~176 bp upstream of the sense TSS[45]). Metagene profiles of the MNase-ChIP-seq signal centered around the TSS of expressed genes showed decreased RNA Pol II occupancy in the promoter-proximal pause region just downstream of the TSS (Figures 3B and S4B), consistent with our mNET-seq results (Figure 2B). In addition, we observed a strong increase in RNA Pol II occupancy at ~30 bp upstream of the TSS (Figures 3B and S4B). The upstream peak position

Please cite this article in press as: Velychko et al., CDK7 kinase activity promotes RNA polymerase II promoter escape by facilitating initiation factor release, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.05.007

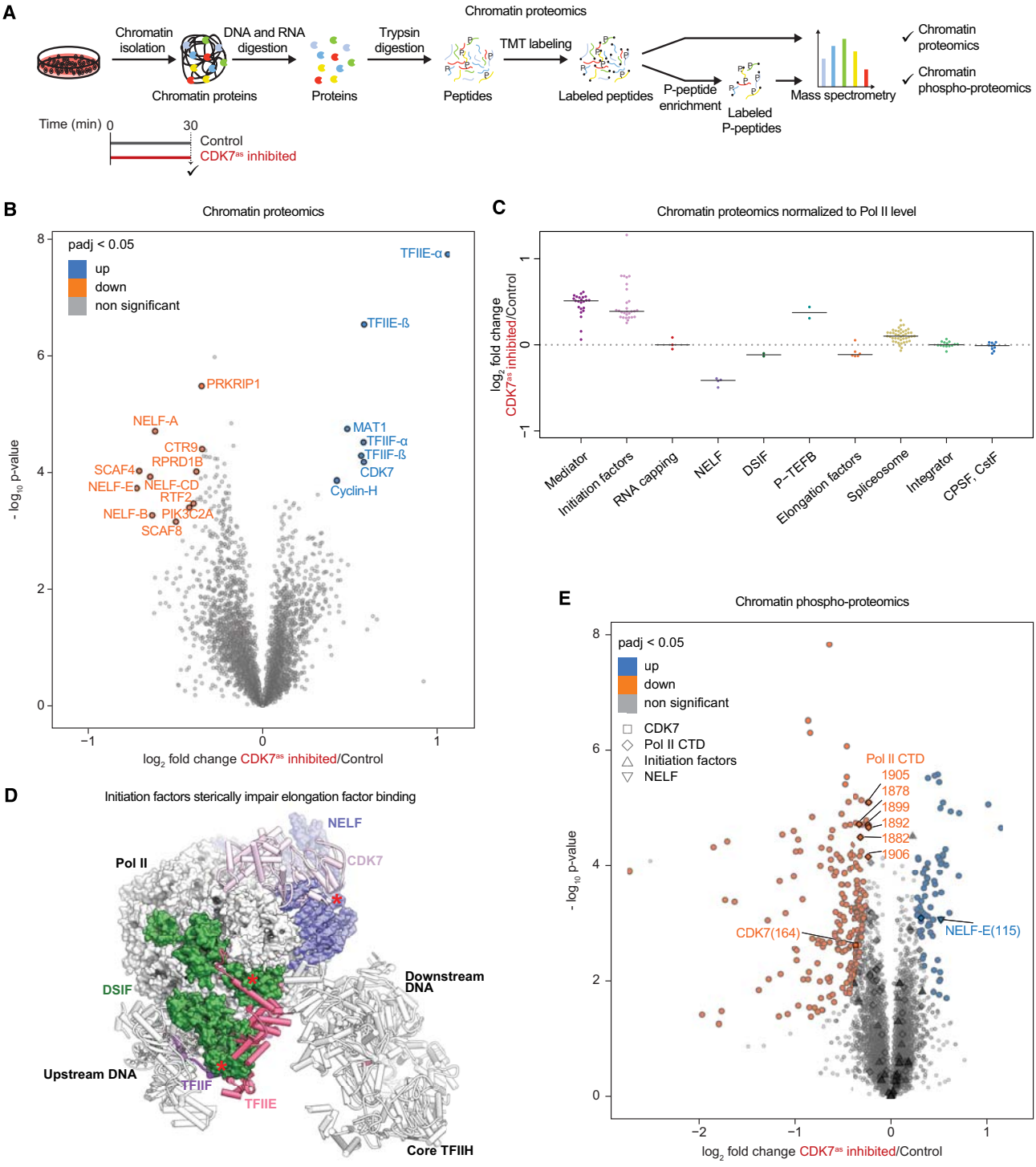


Figure 4. Efficient release of initiation factors and Mediator requires CDK7 activity

(A) Schematic overview of quantitative chromatin and phospho-proteomics.

(B) Volcano plot showing global changes in chromatin binding upon CDK7 inhibition for 30 min. 4,507 proteins were identified. The significantly enriched (n = 7) and lost (n = 11) proteins from chromatin are shown in blue and orange, respectively (q < 0.05; STAR Methods). Unchanged proteins are shown in gray. See also Table S1.

(C) Changes in chromatin binding of RNA Pol II-associated complexes upon CDK7 inhibition normalized to RNA Pol II level. Each dot indicates a subunit of a multi-subunit protein complex detected by chromatin proteomics. The dashed line indicates no change. See also Tables S2 and S3.

(legend continued on next page)

corresponds to previously defined initiation factor footprint locations along the core promoter[9,23] and thus likely reflects PIC assembly. RNA Pol II occupancy of the exemplary *RPL10A* gene showed a similar change after CDK7 inhibition (Figure 3C). Heatmaps of RNA Pol II occupancy aligned at the TSS and sorted by RNA Pol II signal in the pause region indicated that the CDK7 inhibition effect does not depend on promoter-proximal pausing and affects all genes to a similar extent (Figure 3D). These results can be explained by the stabilization of the PIC upstream of the TSS upon CDK7 kinase inhibition.

Efficient release of initiation factors and Mediator requires CDK7 activity

To gain further mechanistic insights into the decrease in the RNA Pol II initiation rate after CDK7 inhibition, we performed proteomic characterization of chromatin-associated proteins and phospho-proteomics experiments after 30 min inhibition (Figure 4A). This allowed us to identify CDK7-dependent changes in chromatin binding and phosphorylation status of proteins. Chromatin proteomics analysis showed a significant increase in chromatin binding of all subunits of the TFIIH kinase module (CDK7/Cyclin H/MAT1), TFIIE, and TFIIF (Figures 4B and S5A; Table S1). Interestingly, TFIIF, which enters the PIC together with RNA Pol II, enhances the stability of the TFIID-TFIIA-TFIIB-RNA Pol II-TFIIF promoter complex[46,47] and directly interacts with subsequently recruited TFIIE.[48,49] TFIIE further stabilizes the PIC and anchors the TFIIH kinase module within the PIC.[50] The high enrichment of these factors in chromatin upon CDK7 inhibition suggests that the accumulation of RNA Pol II observed at promoters (Figure 3B) corresponds to the complete RNA Pol II PIC.

In contrast to the elevated levels of initiation factors, we observed a significant decrease in chromatin binding of all four subunits of the negative elongation factor (NELF), which, together with the DRB sensitivity-inducing factor (DSIF), forms the paused elongation complex.[53] Additionally, we identified a reduction of two CTD-associated factors, SCAF4 and SCAF8. Both proteins bind phosphorylated RNA Pol II at the Ser2 and Ser5 positions of the CTD and suppress early mRNA transcript cleavage and polyadenylation.[54]

To investigate the relative changes in chromatin binding for the RNA Pol II-associated complexes, we normalized protein levels to RNA Pol II abundance (Figure 4C; Tables S2 and S3). Chromatin binding of mRNA capping, integrator, and termination complexes remained unchanged, indicating that the recruitment of these complexes does not depend on CDK7 activity. Levels of NELF, DSIF, and elongation factors were decreased while Mediator and initiation factors were enriched in the chromatin. This agrees with superimposed structures of human PIC[51] and paused transcription[52] complexes, showing that CDK7, as part of the TFIIH kinase module, binds to the same RNA Pol II interface as NELF and that initiation factors TFIIE and TFIIF bind at the same RNA Pol II location as DSIF (Figure 4D). Additionally, previously reported *in vitro* assays showed that TFIIE and DSIF compete for the RNA Pol II clamp during initiation and elongation,[55] and ChIP-qPCR for selected genes showed that CDK7 inhibition alters the exchange of TFIIE for DSIF in humans.[25] Interestingly, upon CDK7 inhibition, TFIIF, TFIIE, and the TFIIH kinase module showed significantly higher chromatin abundance, compared with the upstream complex (TFIID, TFIIA, TFIIB) and TFIIH core (Figure S5B). Altogether, CDK7 kinase activity is essential for the release of initiation factors and Mediator, which in turn would allow for the binding of elongation factors and the transition to productive elongation.

CDK7 exclusively phosphorylates the RNA Pol II CTD within the PIC

To understand how CDK7 kinase activity regulates the release of initiation factors, we analyzed the phospho-proteome of chromatin-bound proteins after CDK7 inhibition (Figures 4E and S5C; Table S4). We detected a significant reduction in RNA Pol II CTD phosphorylation, while the phosphorylation status of initiation factors remained largely unchanged. These data agree with previous *in vitro* kinase assays that showed that CDK7, as a part of TFIIH, is restricted to exclusive phosphorylation of the RNA Pol II CTD.[33] We detected an increase in NELF-E phosphorylation after CDK7 inhibition, suggesting that P-TEFb kinase activity, which phosphorylates NELF-E in multiple positions,[56] may not depend on CDK7 *in vivo* (Figure 4C). We also observed a decrease in CDK7 T-loop phosphorylation at the Ser164 position,[57] suggesting the possibility that CDK7 phosphorylates itself at this site. The motif analysis of all significantly decreased phospho-peptides showed a strong preference for serine (S) and threonine (T) phosphorylation (Figure S5D), consistent with CDK7 being a serine-threonine kinase. A moderate preference for proline (P) in the +1 position aligns with the Ser5 position in the $Y_1S_2P_3T_4S_5P_6S_7$ CTD consensus sequence.

Altogether, phospho-proteomics revealed that only the RNA Pol II CTD phosphorylation level decreased within the PIC upon CDK7 inhibition, suggesting that CDK7 phosphorylation of the CTD is likely required for efficient release of initiation factors.

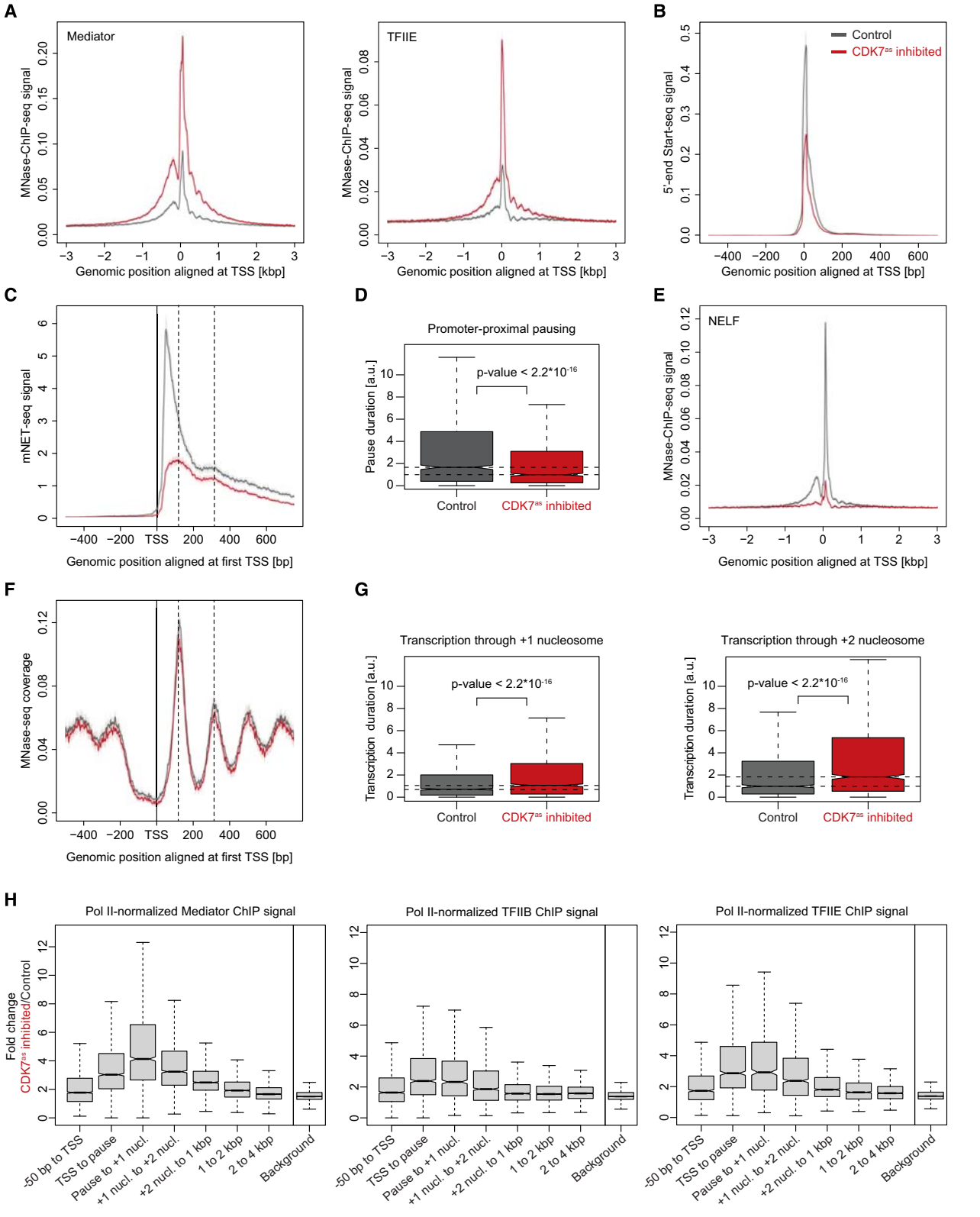
Initiation factors and Mediator can be retained during early elongation

To confirm our proteomics data, we performed MNase-ChIP-seq for Mediator, TFIIB, TFIIE, and NELF. Unexpectedly, both initiation factors and Mediator showed increased occupancies not only at promoters but throughout the first 1–2 kbp downstream of the TSS (Figures 5A, S6A, and S6B). Considering the possibility of alternative TSS selection upon CDK7 inhibition, as observed in yeasts,[21] we performed spike-in-normalized Start-seq[58] to show that TSS selection is not affected by CDK7 kinase activity (Figure 5B). The observed median decrease of

(D) Initiation factors sterically impair elongation factor binding. Structures of human PIC (PDB: 7EGC)[51] and paused elongation (PDB: 6GML)[52] complexes were superimposed by matching the 12 subunits of RNA Pol II. Clashes are visualized with red asterisks.

(E) Volcano plot showing changes in the phospho-proteome after CDK7 inhibition for 30 min. 13,030 phospho-peptides were identified in total. Significantly upregulated (92 phospho-peptides of 55 proteins) and downregulated (281 phospho-peptides of 100 proteins) phospho-peptides are shown in blue and orange, respectively ($q < 0.05$; STAR Methods). Unchanged phospho-peptides are shown in gray. See also Table S4.

Please cite this article in press as: Velychko et al., CDK7 kinase activity promotes RNA polymerase II promoter escape by facilitating initiation factor release, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.05.007



(legend on next page)

43% in capped RNA 5′ end levels (Figures 5B and S6C) corresponded to the decreases seen in both mNET-seq (of 43%; Figure 2C) and TT-seq (of 56%; Figure 1E). This was consistent with unchanged RNA Pol II-normalized levels of the mRNA capping complex in our chromatin proteomics data (Figure 4C), indicating that the newly synthesized RNA was efficiently capped.

To ensure that the increase in Mediator and initiation factors downstream of the TSS was not due to peak signal increase leading to peak broadening, we analyzed TFIIB signal across different RNA synthesis levels. We did not observe any TFIIB signal downstream of the TSS associated with increased signal at the PIC assembly region (Figure S6F). Additionally, we compared Mediator, TFIIB, and TFIIE signals between control and CDK7-inhibited samples for genes with similar signal strength at the PIC assembly region (Figure S6G). There was a clear increase in Mediator, TFIIB, and TFIIE signals downstream of the TSS in CDK7-inhibited samples. Altogether, this supports that the increase in signal downstream of the TSS is specific to CDK7 inhibition.

To further investigate what happens to RNA Pol II transcription downstream of the TSS, we analyzed the RNA Pol II mNET-seq signal in this region. CDK7 inhibition resulted in attenuation of the promoter-proximal peak and improved visibility of nucleosome-like peaks downstream of the pause site (Figure 5C). MNase-seq confirmed that these peaks correspond to +1 and +2 nucleosomes and also showed that CDK7 inhibition does not affect nucleosome positioning (Figure 5F).

Next, we estimated the relative time it took for RNA Pol II to transcribe through the promoter-proximal pause region and the +1 and +2 nucleosomes (STAR Methods). The pause duration strongly decreased upon CDK7 inhibition (Figure 5D), consistent with an almost complete loss of NELF MNase-ChIP-seq signal (Figure 5E), which was also observed for selected genes using ChIP-qPCR[25] and for nuclear extracts using an *in vitro* system.[26] Since a decrease in pause duration can result in increased RNA Pol II initiation,[24,43] we wanted to confirm that the accumulation of RNA Pol II upstream of the TSS (Figure 3B) represents RNA Pol II retention, rather than an increase in RNA Pol II initiation, followed by destabilization of early elongation complexes due to decreased association of pausing factors. We hypothesized that under such a scenario, genes with the most significant decrease in RNA Pol II pause duration would show the greatest increase in RNA Pol II signal at the promoter. However, we did not observe such a correlation (Figure S6D), consistent with the heatmaps in Figure 3D. Together with the observed similar reduction in 3′ end Start-seq signal (Figure S6E) and RNA synthesis (Figure 1C), this indicates RNA Pol II retention rather than an increase in RNA Pol II initiation at promoters.

Although the position of maximum RNA Pol II occupancy within the +1 and +2 nucleosomes remained unchanged (Figure S6H), the relative time that it took for RNA Pol II to transcribe through the +1 and +2 nucleosomes (transcription duration) was significantly longer upon CDK7 inhibition (Figure 5G). RNA Pol II-normalized ChIP signals for Mediator, TFIIB, and TFIIE showed a clear increase in the signal of all three factors for the regions between the TSS and the +2 nucleosome and a slight increase for Mediator and TFIIE within 2 kbp downstream of the TSS (Figure 5H). This is consistent with chromatin proteomics, where we observed that Mediator, TFIIF, TFIIE, and the TFIIH kinase module were more enriched in chromatin compared to the upstream complex and the TFIIH core (Figures 4B, 4C, and S5B). RNA Pol II levels on chromatin decreased by an average of 14% upon CDK7 inhibition (Table S1), consistent with the MNase-ChIP-seq data showing a reduction of RNA Pol II across gene bodies (Figures S4B, S7A, and S7B).

Collectively, these data suggest that upon CDK7 inhibition, hypophosphorylated RNA Pol II can proceed to elongation without the release of initiation factors and Mediator, skipping promoter-proximal pausing and taking more time to transcribe through +1 and +2 nucleosomes.

Figure 5. Initiation factors and Mediator can be retained during early elongation

(A) Metagene analysis of Mediator (MED26) and TFIIE (TFIIE-α) MNase-ChIP-seq signals before (control; gray) and after (red) CDK7 inhibition for 30 min. MNase-ChIP-seq signal from three biological replicates was merged and aligned at the TSS. The shaded areas around the average signal (solid lines) indicate 95% confidence intervals of the mean.

(B) Metagene analysis of 5′ end Start-seq signal before (control; gray) and after (red) CDK7 inhibition for 30 min. Start-seq signal from two biological replicates was merged and aligned at the TSS. The shaded areas around the average signal (solid lines) indicate 95% confidence intervals of the mean.

(C) Metagene analysis of mNET-seq signal before (control; gray) and after (red) CDK7 inhibition for 30 min. mNET-seq signal from two biological replicates was merged and aligned at the TSS. The shaded areas around the average signal (solid lines) indicate 95% confidence intervals of the mean. Dashed lines indicate +1 and +2 nucleosome dyads.

(D) RNA Pol II pause durations before (control; gray) and after (red) CDK7 inhibition for 30 min at promoter-proximal pause positions (n = 5,985). Pause duration was calculated by combining TT-seq and mNET-seq data and defined as the time RNA Pol II needs to transcribe through the pause region (STAR Methods). Boxplot representations as in Figure 1E. p values by two-sided Mann-Whitney U test.

(E) Metagene analysis of the NELF (NELF-E) MNase-ChIP-seq signal before (control; gray) and after (red) CDK7 inhibition for 30 min. MNase-ChIP-seq signal from three biological replicates was merged and aligned at the TSS. The shaded areas around the average signal (solid lines) indicate 95% confidence intervals of the mean.

(F) Metagene analysis of MNase-seq signal of the expressed genes before (control; gray) and after (red) CDK7 inhibition for 15 min. MNase-seq signal from two biological replicates was merged and aligned at the TSS. The shaded areas around the average signal (solid lines) indicate 95% confidence intervals of the mean. Dashed lines indicate +1 and +2 nucleosome dyads.

(G) RNA Pol II transcription durations before (control; gray) and after (red) CDK7 inhibition for 30 min at +1 (left; n = 1,808) and +2 (right; n = 1,758) nucleosome positions. Transcription duration was calculated by combining TT-seq and NET-seq data and defined as the time RNA Pol II needs to transcribe through the +1 or +2 nucleosome regions (STAR Methods).Boxplot representations as in Figure 1E. p values by two-sided Mann-Whitney U test.

(H) RNA Pol II-normalized Mediator (MED26), TFIIB, and TFIIE (TFIIE-α) MNase-ChIP-seq signal ratios (Mediator n_{mean} = 2,034, TFIIB n_{mean} = 1,536, TFIIE n_{mean} = 1,136) before and after CDK7 inhibition for 30 min. The background was calculated based on RNA Pol II signals for a window from +2 to +4 kbp downstream of the TSS before and after CDK7 inhibition. Boxplot representations as in Figure 1E.

Please cite this article in press as: Velychko et al., CDK7 kinase activity promotes RNA polymerase II promoter escape by facilitating initiation factor release, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.05.007

Increase in CTD phosphorylation coincides with release of initiation factors and Mediator and recruitment of elongation factors

Next, we asked what happens to elongating RNA Pol II with retained initiation factors and Mediator ~1–2 kbp downstream of the TSS where the signals for Mediator and initiation factors drop to background levels (Figures 5A, 5H, and S6A). We performed MNase-ChIP-seq for the phosphorylated Ser5 (Ser5P) RNA Pol II CTD, which is the main CDK7 target. The metagene profile aligned at the TSS showed decreased Ser5P signal all over the gene body (Figures 6A and S7A). Since total RNA Pol II also decreased in the gene body (Figures S7A and S7B), we normalized the Ser5P signal to the RNA Pol II level and compared it before and after CDK7 inhibition for the regions that range from the promoter to 8 kbp downstream of the TSS. The Ser5P level relative to RNA Pol II decreased most at the promoter, where CDK7 phosphorylates the CTD as part of TFIIH, and gradually recovered with increasing distance from the TSS until it reached the control level ~2 kbp downstream (Figure 6B). This pattern aligns with previously observed significant decreases in RNA Pol II-normalized Ser5P ChIP-qPCR signals nearer to the 5′ ends of selected genes upon CDK7 inhibition.[25] At the same distance, the RNA Pol II-normalized MNase-ChIP-seq signals for initiation factors and Mediator reached background levels (Figures 5A, 5H, and S6A). To investigate whether the downstream increase in Ser5P level and release of initiation factors and Mediator from RNA Pol II facilitated the recruitment of elongation factors, we performed MNase-ChIP-seq for SPT5, a DSIF subunit. Since the decrease in DSIF chromatin binding upon CDK7 inhibition was the same as for other elongation factors (Figure 4C, except RTF1, which showed no changes), and DSIF travels with RNA Pol II as part of the activated elongation complex,[59] we assume that changes in the SPT5 signal represent changes of the elongation complex. The RNA Pol II-normalized level of SPT5 (Figure 6D) decreased most in the pause region, where DSIF is normally recruited and recovered after ~2 kbp, similarly to Ser5P (Figure 6B).

RNA Pol II-normalized chromatin proteomics upon CDK7 inhibition showed no significant change in chromatin localization for SSU72, which is responsible for Ser5 dephosphorylation during elongation,[60] but significant enrichment of all three transcriptional CDKs, CDK9/12/13, that act downstream of CDK7 (Figure 6E). These CDKs can phosphorylate the CTD at the Ser5 position,[1,61–63] and CDK12/13 kinase activities were shown to promote elongation factor association with RNA Pol II,[63,64] suggesting that they may be involved in the increase in Ser5P and SPT5 levels during late elongation (Figures 6B and 6D).

The apparent exchange of initiation factors for elongation factors ~1–2 kbp downstream of the TSS coincided with an increase in RNA Pol II elongation velocity, which we estimated by combining the MNase-ChIP-seq (Figure 6F) or the mNET-seq (Figure S7C) RNA Pol II occupancy measurements with the TT-seq-derived productive initiation rates as described.[65] A similar drop in RNA Pol II elongation velocity, calculated with DNA-based (MNase-ChIP-seq) and RNA-based (mNET-seq) methods that detect RNA Pol II occupancy, indicated that elongating RNA Pol II with retained initiation factors and Mediator is active and synthesizes RNA.

Altogether, our MNase-ChIP-seq experiments show that a gradual increase in Ser5P level with a simultaneous release of initiation factors and Mediator allows for gradual recruitment of elongation factors and an increase in RNA Pol II elongation velocity.

DISCUSSION

In all eukaryotic organisms, RNA Pol II CTD phosphorylation by transcription-related CDKs is critical for the regulation of the transcription cycle.[66] Over the years, several studies have investigated the roles of CDK7 kinase activity in transcription, yet the use of various inhibition strategies, inhibition times, and methods to monitor transcriptional changes has resulted in several inconsistent findings so that the direct function of CDK7 kinase activity remains to be fully established. To address this, we rapidly and selectively inhibited CDK7 kinase in human cells and monitored genome-wide changes in RNA synthesis, RNA Pol II occupancy, and transcription factor binding to chromatin.

Our results show that CDK7 inhibition for 15 min leads to global downregulation of RNA synthesis due to a decrease in initiation rate of all RNA Pol II-transcribed genes. High-resolution RNA Pol II MNase-ChIP-seq allowed us to separate the PIC formation peak from the paused RNA Pol II peak and showed that upon CDK7 inhibition, RNA Pol II, as a part of Mediator-PIC, accumulates at promoters. Short inhibition time points and careful consideration of gene lengths revealed that RNA Pol II molecules that proceeded to productive elongation before the start of inhibition were not impaired by CDK7 inhibition, indicating that elongation and termination are not directly regulated by CDK7 kinase activity.

Despite the observed reduction in CTD phosphorylation levels upon CDK7 inhibition, our data indicate that the newly synthesized RNA was efficiently capped. A previous study using an *in vitro* RNA Pol II transcription system showed that robust activation of co-transcriptional capping can occur even in the absence of the RNA Pol II CTD or TFIIH, indicating the presence of a CTD-independent, but RNA Pol II-mediated, mechanism for capping activation.[31] Recent structural studies showed that human capping enzymes stably bind to the RNA Pol II stalk at the RNA exit tunnel.[67] Together, this suggests that the direct interaction between RNA Pol II and capping enzymes may be sufficient for their recruitment upon CDK7 inhibition.

Although RNA Pol II retention at promoters is the most prominent consequence of CDK7 inhibition, hypophosphorylated RNA Pol II can also proceed to elongation without the release of initiation factors and Mediator (Figure 7). Previous ChIP experiments have similarly shown that upon depletion of the yeast CDK12/13 ortholog, Ctk1, initiation factors can remain associated with an elongating RNA Pol II.[68] Our data together with a previous study[25] indicate that prolonged association of initiation factors and Mediator during early elongation prevents the recruitment of pausing and elongation factors, resulting in bypassing promoter-proximal pausing. This suggests that CDK7 kinase activity is also required to establish pausing. The elongation velocity of RNA Pol II retaining initiation factors and Mediator was lower than that of normal elongating RNA Pol II, especially at the +1 and +2 nucleosomes. This decreased elongation velocity

Please cite this article in press as: Velychko et al., CDK7 kinase activity promotes RNA polymerase II promoter escape by facilitating initiation factor release, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.05.007

CellPress
OPEN ACCESS

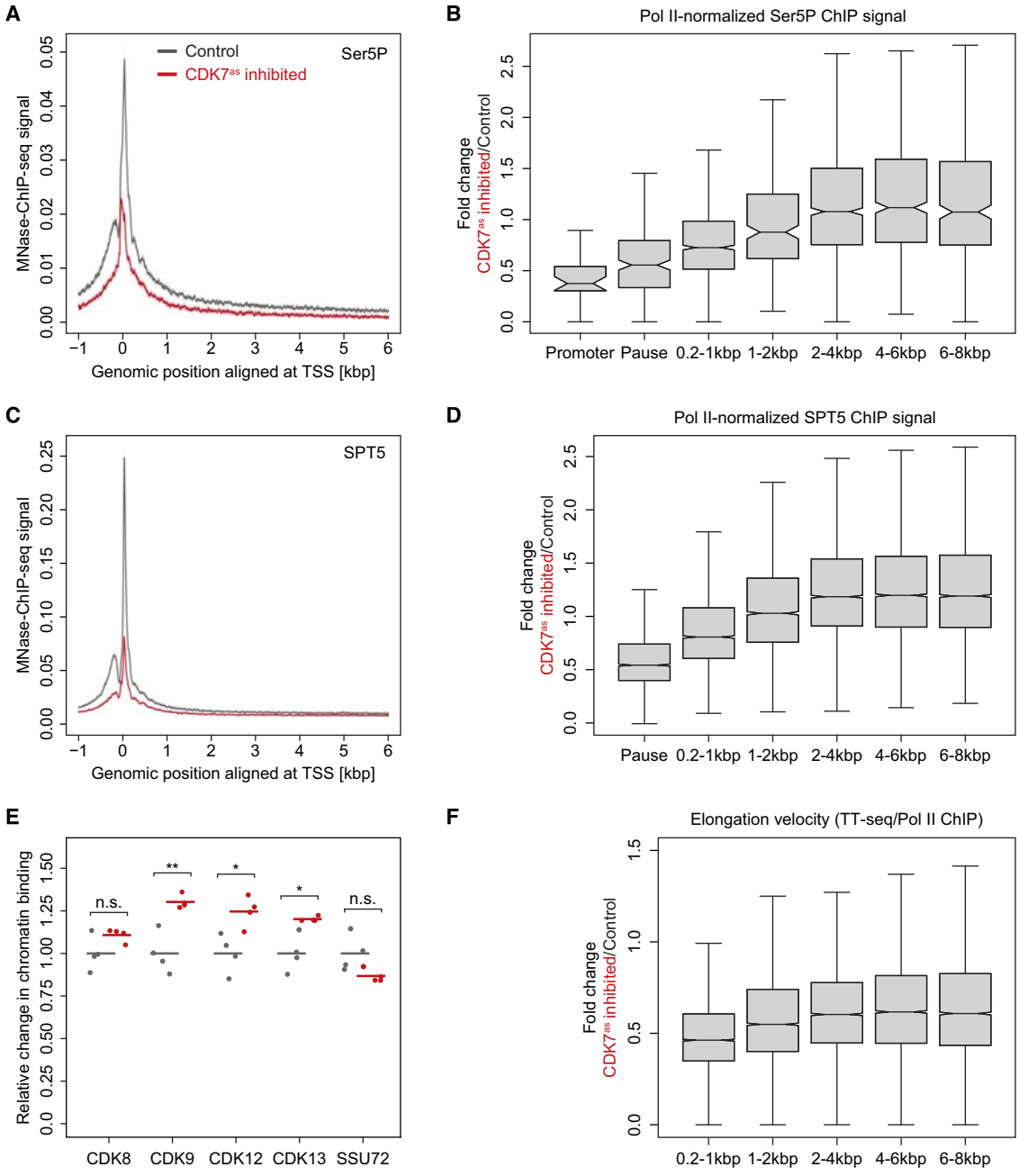


Figure 6. Increase in CTD phosphorylation coincides with release of initiation factors and Mediator and recruitment of elongation factors

(A) Metagene analysis of Ser5P MNase-ChIP-seq signal of the expressed genes before (control; gray) and after (red) CDK7 inhibition for 30 min. MNase-ChIP-seq signal from three biological replicates was merged and aligned at the TSS. The shaded areas around the average signal (solid lines) indicate 95% confidence intervals of the mean.

(B) Ratios of RNA Pol II-normalized Ser5P MNase-ChIP-seq signal (n_{mean} = 508) before and after CDK7 inhibition for 30 min. Boxplot representations as in Figure 1E.

(C) Metagene analysis of the SPT5 MNase-ChIP-seq signal of the expressed genes before (control; gray) and after (red) CDK7 inhibition for 30 min. MNase-ChIP-seq signal from three biological replicates was merged and aligned at the TSS. The shaded areas around the average signal (solid lines) indicate 95% confidence intervals of the mean.

(D) Ratios of RNA Pol II-normalized SPT5 MNase-ChIP-seq signal (n_{mean} = 4,981) before and after CDK7 inhibition for 30 min. Boxplot representations as in Figure 1E.

(E) Changes in chromatin binding of CDK8/9/12/13 and SSU72 upon CDK7 inhibition, based on chromatin proteomics normalized to RNA Pol II level. Each dot represents a biological replicate. Horizontal lines indicate mean values of four biological replicates. CDK8, which acts upstream of CDK7, serves as a control. **q < 0.01, *q < 0.05, n.s. q > 0.05 (STAR Methods). See also Table S3.

(F) Ratios of RNA Pol II elongation velocity (STAR Methods) before and after CDK7 inhibition for 30 min based on TT-seq and RNA Pol II MNase-ChIP-seq signals (n_{mean} = 6,972). Boxplot representations as in Figure 1E.

Please cite this article in press as: Velychko et al., CDK7 kinase activity promotes RNA polymerase II promoter escape by facilitating initiation factor release, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.05.007

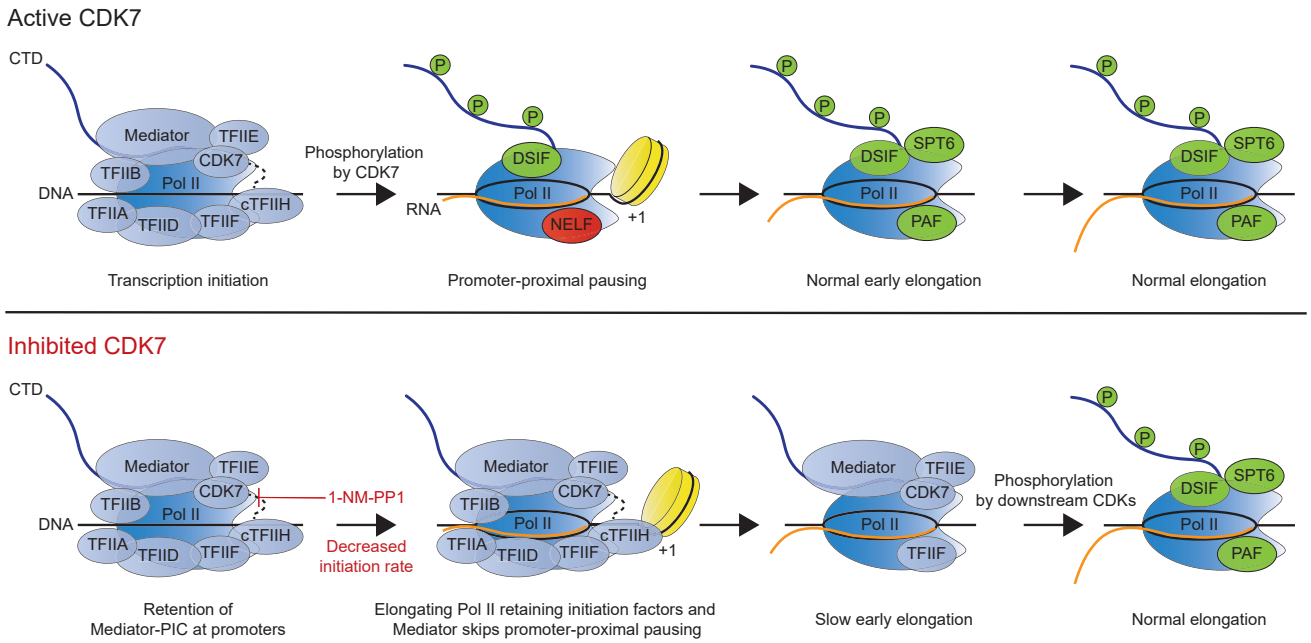


Figure 7. Proposed model of CDK7 function in transcription

Top: CDK7 phosphorylates RNA Pol II CTD repeats located at the RNA Pol II-Mediator interface. The introduced negatively charged phosphoryl groups on the CTD may repel Mediator from RNA Pol II, destabilizing the PIC and promoting the dissociation of initiation factors from RNA Pol II. The release of initiation factors facilitates the release of RNA Pol II from the promoter region and allows for the recruitment of elongation factors.

Bottom: in the absence of CDK7 activity, initiation factors and Mediator fail to dissociate from RNA Pol II, leading to Mediator-PIC retention at active promoters. This is the main effect of CDK7 inhibition, leading to a global decrease in RNA synthesis and impaired cell growth. With lower efficiency, RNA Pol II complexes escape from promoter regions, retaining initiation factors and Mediator, and proceed into gene bodies without promoter-proximal pausing and taking more time to transcribe through +1 and +2 nucleosomes. During early elongation, the upstream complex (TFIID, TFIIA, TFIIB) and the TFIIH core dissociate from RNA Pol II earlier than Mediator, TFIIF, TFIIE, and the TFIIH kinase module (CDK7/Cyclin H/MAT1). The prolonged association of Mediator, TFIIF, TFIIE, and the TFIIH kinase module with RNA Pol II acts as a limiting factor for the recruitment of pausing and elongation factors. The increase in the RNA Pol II CTD phosphorylation level, possibly by downstream CDKs, results in the late release of initiation factors and Mediator from RNA Pol II and allows for the recruitment of elongation factors.

is likely due to inefficient transcription by RNA Pol II associated with initiation factors, which is prone to backtracking, as demonstrated *in vitro*,[69] and to the absence of elongation factors that typically enhance the elongation rate.[70]

Our findings also suggest that RNA Pol II release from the promoter region does not require RNA Pol II release from initiation factors. This agrees with recent biochemical and structural studies of promoter escape showing that initiation factors, in particular TFIIE, TFIIF, and TFIIH, can remain associated with RNA Pol II during early elongation *in vitro*, even in the absence of Mediator.[71,72] Here, upon CDK7 inhibition *in vivo*, Mediator remains associated with RNA Pol II, likely stabilizing the association of initiation factors with elongating RNA Pol II and leading to the observed prolonged retention. RNA Pol II release from the promoter region seems to be a stochastic process, which can only be facilitated by CDK7-dependent CTD phosphorylation, as shown in our study or previously,[14,16,17] by capping[69,73] or by evicting the +1 nucleosome downstream of the promoter.[74] However, CTD phosphorylation seems to be required for the release of initiation factors and Mediator from RNA Pol II *in vivo*, and we show that CDK7 regulates this process. Previous *in vitro* biochemical assays revealed that RNA Pol II CTD hyperphosphorylation leads to RNA Pol II dissociation from Mediator,[14] and recent structural studies of the human Mediator-PIC complex showed that part of the CTD peptide is located at the interface of the Mediator-RNA Pol II complex.[51,75] Thus, negatively charged phosphoryl groups on the CTD introduced by CDK7 could repel Mediator from RNA Pol II. Since Mediator stabilizes the initiation complex,[76] its release would also promote the dissociation of initiation factors, thereby facilitating promoter escape. Elongation factor competition for the RNA Pol II surface could further promote the dissociation of initiation factors, as was shown *in vitro*.[72] Although we did not observe a decrease in the phosphorylation level of initiation factors by phospho-proteomics, it remains plausible that CDK7 phosphorylates some of these factors in addition to the RNA Pol II CTD, potentially further facilitating PIC destabilization and initiation factor dissociation. Upon CDK7 inhibition, the Ser5P level of the CTD increased during late elongation, presumably due to downstream CDK kinase activities, coinciding with the dissociation of initiation factors and the recruitment of elongation factors. At the TSS, however, the CTD remained hypophosphorylated, indicating that downstream CDK kinases are unable to compensate for the absence of CDK7 activity within the Mediator-PIC complex.

Our proteomics and MNase-ChIP-seq data, upon CDK7 inhibition, suggest that during early elongation, the upstream complex (TFIID, TFIIA, TFIIB) and the TFIIH core dissociate from RNA Pol II earlier than Mediator, TFIIF, TFIIE, and the

TFIIH kinase module. This earlier dissociation may be attributed to multiple contacts between the upstream complex and the TFIIH core with DNA.[11,12] The constant competition to remain associated with RNA Pol II or to bind to DNA may increase the likelihood of their dissociation from RNA Pol II during early elongation. In contrast, Mediator, TFIIF, TFIIE, and the TFIIH kinase module have limited interactions with DNA,[11,12] and therefore are more likely to remain associated with RNA Pol II. Since only Mediator, TFIIF, TFIIE, and the TFIIH kinase module occupy the same interface as pausing and elongation factors, their prolonged association with RNA Pol II can act as a limiting factor for the recruitment of the pausing and elongation factors.

CDK7 has long attracted attention as a potential drug target for cancer chemotherapy,[77–79] with three CDK7 inhibitors currently in phase 1 or 2 clinical trials: SY-5609 (NCT04247126), Q901 (NCT05394103), and XL102 (NCT04726332). Initial studies with the covalent CDK7 inhibitor THZ1, which also inhibits CDK12/13 at slightly higher concentrations,[80] demonstrated that tumor-specific super-enhancer-associated genes are disproportionally sensitive to CDK7 inhibition,[80–82] which is likely due to their high expression levels, and thus agrees with our observation that genes with higher initiation rates respond stronger to CDK7 inhibition than genes with lower initiation rates (Figure S3D). Several studies also showed that CDK7 is not essential for global transcription and regulates only subsets of genes.[19,32–34,83–87] However, these studies were based on changes in total RNA level and/or lacked external RNA normalization (e.g., spike-ins), making it difficult to detect global changes in RNA synthesis. Our results show that selective and rapid CDK7 inhibition downregulates the synthesis of all genes, consistent with irreversible Kin28 inhibition in yeast.[20] Therefore, cancer treatment with CDK7 inhibitors would not be restricted to tumor-specific gene transcription but could also target essential processes in normal cells, potentially causing side effects. For instance, CDK7 is required for neuronal activity-induced gene expression, and CDK7 inhibition with THZ1 caused memory deficits in mice.[88] Taken together, CDK7 inhibitors show promise in cancer treatment, but caution should be taken regarding the genome-wide role of CDK7 in RNA synthesis.

In conclusion, we show that CDK7 kinase activity promotes the release of initiation factors and Mediator from RNA Pol II, facilitating promoter escape.

Limitations of the study

In our study, we used HEK293 cells, immortalized human embryonic kidney cells whose gene expression profile may differ from that of normal cells. 1-NM-PP1 is a competitive, reversible inhibitor, so CDK7 inhibition may not be complete. 1-NM-PP1 inhibition requires genomic alteration, making this technology inapplicable for cancer treatment. The repetitiveness of the CTD makes it challenging to distinguish Ser5P from Ser7P by mass spectrometry (MS) in our chromatin phospho-proteomics. Thus, we discuss only the overall decrease in CTD phosphorylation in Figure 4E. Due to the limitations of MS, our analysis might not have captured all significant phosphorylation site changes in the PIC upon CDK7 inhibition, especially in regions with poor sequence coverage. Since phosphorylation at the Ser7 position is less abundant than at Ser5 and is only required for the expression of small nuclear RNA (snRNA),[89] we focus on the decrease in Ser5P level upon CDK7 inhibition. Because the MED26 subunit showed a similar reduction in chromatin binding to the majority of other Mediator subunits, we assume that the changes in the MED26 signal represent the changes in the whole Mediator complex. Additionally, since MNase-ChIP-seq involves crosslinking, it is possible that some initiation factors are only flexibly tethered to elongating RNA Pol II or through other initiation factors upon CDK7 inhibition, yet their association seems to be strong enough to prevent elongation factor binding.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Cell lines and cell culture
- METHOD DETAILS
  - Genome editing with CRISPR/Cas9
  - Cell viability and proliferation assays
  - EU incorporation assay
  - TT-seq and RNA-seq
  - mNET-seq
  - MNase-ChIP-seq
  - Start-seq
  - MNase-seq
  - Chromatin and phospho-proteomics
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - MS database search
  - MS data analysis
  - Annotation
  - Major isoform annotation
  - GenoSTAN annotation of transcribed units and their classification for HEK293 CDK7^{as}
  - Analysis of high-throughput sequencing data
  - RNA-seq and TT-seq data preprocessing and normalization
  - RNA amount per cell
  - Expressed gene set for analysis
  - Response ratio
  - Productive initiation frequency
  - mNET-seq data preprocessing and normalization
  - MNase-ChIP-seq data preprocessing and normalization
  - Start-seq data preprocessing and normalization
  - MNase-seq data preprocessing and normalization
  - Detection of +1 and +2 nucleosome positions
  - Detection of promoter-proximal pause sites
  - Nucleosome arrest sites
  - Pause duration and transcription duration through the +1 and +2 nucleosomes
  - Elongation velocity estimation
  - Differential expression analysis
  - Replicates/statistics

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2024.05.007.

ACKNOWLEDGMENTS

We thank Rebecca Hobrecht for her help with MNase-ChIP-seq. We thank Kerstin Maier, Petra Rus, and Shyam Sundar Ramasamy for their help with sequencing. We thank Kseniia Lysakovskaia for advice on mNET-seq and Anna Sawicka for providing spike-ins for mNET-seq. We also thank Monika Raabe for assistance in MS and Aleksandar Chernev for submitting MS data to the PRIDE database. We are grateful to James Walshe, Elisa Oberbeckmann, and Caitlin MacCarthy for their feedback on the manuscript and to Shintaro Aibara for making the figure of superimposed structures. We thank all past and current members of the Cramer lab for productive discussions, particularly Kristina Zumer, Christian Dienemann, and Livia Caizzi. T.V. and E.M. were supported by the International Max Planck Research School for Genome Science, part of the Göttingen Graduate Center for Neurosciences, Biophysics and Molecular Biosciences. P.C. was supported by the Deutsche Forschungsgemeinschaft (SFB860, SPP1935, EXC 2067/1-390729940) and the European Research Council (advanced investigator grant CHROMATRANS, grant agreement no. 882357).

AUTHOR CONTRIBUTIONS

T.V. conceived and planned the study and conducted experiments. E.M. and M.L. designed and carried out bioinformatics analysis. B.S. advised on bioinformatics analysis. T.V., E.M., and M.L. interpreted sequencing data. I.P. conducted and analyzed MS/MS experiments under supervision of H.U. I.F.-V. and S.M. generated the CDK7^{as} cell line, and C.S. performed EU incorporation assay under supervision of S.M. M.W. helped with chromatin proteomics and cell line characterization. T.V. wrote the original draft of the manuscript with input from all authors. P.C. and M.L. reviewed and edited the manuscript. T.V., E.M., I.P., M.W., C.S., and M.L. visualized the data. P.C. and M.L. supervised the project. P.C. conceived the project and acquired funding.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 2, 2022
Revised: February 1, 2024
Accepted: May 8, 2024
Published: May 30, 2024

REFERENCES

1. Zaborowska, J., Egloff, S., and Murphy, S. (2016). The Pol II CTD: new twists in the tail. Nat. Struct. Mol. Biol. *23*, 771–777. https://doi.org/10.1038/nsmb.3285.
2. Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. Mol. Cell *36*, 541–546. https://doi.org/10.1016/J.MOLCEL.2009.10.019.
3. Venkat Ramani, M.K., Yang, W., Irani, S., and Zhang, Y. (2021). Simplicity is the ultimate sophistication—crosstalk of post-translational modifications on the RNA polymerase II. J. Mol. Biol. *433*, 166912. https://doi.org/10.1016/j.jmb.2021.166912.
4. Dieci, G. (2021). Removing quote marks from the RNA polymerase II CTD 'code.'. Biosystems. *207*, 104468. https://doi.org/10.1016/j.biosystems.2021.104468.
5. Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. Nat. Rev. Genet. *15*, 163–175. https://doi.org/10.1038/nrg3662.
6. Akhtar, M.S., Heidemann, M., Tietjen, J.R., Zhang, D.W., Chapman, R.D., Eick, D., and Ansari, A.Z. (2009). TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. Mol. Cell *34*, 387–393. https://doi.org/10.1016/J.MOLCEL.2009.04.016.
7. Glover-Cutter, K., Larochelle, S., Erickson, B., Zhang, C., Shokat, K., Fisher, R.P., and Bentley, D.L. (2009). TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. Mol. Cell. Biol. *29*, 5455–5464. https://doi.org/10.1128/MCB.00637-09.
8. Kim, M., Suh, H., Cho, E.J., and Buratowski, S. (2009). Phosphorylation of the yeast Rpb1 C-terminal domain at serines 2, 5, and 7. J. Biol. Chem. *284*, 26421–26426. https://doi.org/10.1074/jbc.M109.028993.
9. Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Söding, J., and Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. Nat. Struct. Mol. Biol. *17*, 1272–1278. https://doi.org/10.1038/nsmb.1903.
10. Rimel, J.K., and Taatjes, D.J. (2018). The essential and multifunctional TFIIH complex. Protein Sci. *27*, 1018–1037. https://doi.org/10.1002/pro.3424.
11. Sainsbury, S., Bernecky, C., and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. *16*, 129–143. https://doi.org/10.1038/nrm3952.
12. Farnung, L., and Vos, S.M. (2022). Assembly of RNA polymerase II transcription initiation complexes. Curr. Opin. Struct. Biol. *73*, 102335. https://doi.org/10.1016/J.SBI.2022.102335.
13. Luse, D.S. (2013). Promoter clearance by RNA polymerase II. Biochim. Biophys. Acta *1829*, 63–68. https://doi.org/10.1016/J.BBAGRM.2012.08.010.
14. Søgaard, T.M.M., and Svejstrup, J.Q. (2007). Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. J. Biol. Chem. *282*, 14113–14120. https://doi.org/10.1074/jbc.M701345200.
15. Svejstrup, J.Q., Li, Y., Fellows, J., Gnatt, A., Bjorklund, S., and Kornberg, R.D. (1997). Evidence for a mediator cycle at the initiation of transcription. Proc. Natl. Acad. Sci. USA *94*, 6075–6078. https://doi.org/10.1073/pnas.94.12.6075.
16. Jeronimo, C., and Robert, F. (2014). Kin28 regulates the transient association of Mediator with core promoters. Nat. Struct. Mol. Biol. *21*, 449–455. https://doi.org/10.1038/nsmb.2810.
17. Wong, K.H., Jin, Y., and Struhl, K. (2014). TFIIH phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape. Mol. Cell *54*, 601–612. https://doi.org/10.1016/J.MOLCEL.2014.03.024.
18. Liu, Y., Kung, C., Fishburn, J., Ansari, A.Z., Shokat, K.M., and Hahn, S. (2004). Two cyclin-dependent kinases promote RNA polymerase II transcription and formation of the scaffold complex. Mol. Cell. Biol. *24*, 1721–1735. https://doi.org/10.1128/MCB.24.4.1721-1735.2004.
19. Kanin, E.I., Kipp, R.T., Kung, C., Slattery, M., Viale, A., Hahn, S., Shokat, K.M., and Ansari, A.Z. (2007). Chemical inhibition of the TFIIH-associated kinase Cdk7/Kin28 does not impair global mRNA synthesis. Proc. Natl. Acad. Sci. USA *104*, 5812–5817. https://doi.org/10.1073/PNAS.0611505104.
20. Rodríguez-Molina, J.B., Tseng, S.C., Simonett, S.P., Taunton, J., and Ansari, A.Z. (2016). Engineered covalent inactivation of TFIIH-kinase reveals an elongation checkpoint and results in widespread mRNA stabilization. Mol. Cell *63*, 433–444. https://doi.org/10.1016/j.molcel.2016.06.036.
21. Murakami, K., Mattei, P.J., Davis, R.E., Jin, H., Kaplan, C.D., and Kornberg, R.D. (2015). Uncoupling promoter opening from start-site scanning. Mol. Cell *59*, 133–138. https://doi.org/10.1016/J.MOLCEL.2015.05.021.
22. Fisher, R.P. (2005). Secrets of a double agent: CDK7 in cell-cycle control and transcription. J. Cell Sci. *118*, 5171–5180. https://doi.org/10.1242/JCS.02718.
23. Shao, W., and Zeitlinger, J. (2017). Paused RNA polymerase II inhibits new transcriptional initiation. Nat. Genet. *49*, 1045–1051. https://doi.org/10.1038/ng.3867.
24. Gressel, S., Schwalb, B., Decker, T.M., Qin, W., Leonhardt, H., Eick, D., and Cramer, P. (2017). CDK9-dependent RNA polymerase II pausing

controls transcription initiation. eLife *6*, e29736. https://doi.org/10.7554/eLife.29736.

25. Larochelle, S., Amat, R., Glover-Cutter, K., Sansó, M., Zhang, C., Allen, J.J., Shokat, K.M., Bentley, D.L., and Fisher, R.P. (2012). Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II. Nat. Struct. Mol. Biol. *19*, 1108–1115. https://doi.org/10.1038/nsmb.2399.
26. Nilson, K.A., Guo, J., Turek, M.E., Brogie, J.E., Delaney, E., Luse, D.S., and Price, D.H. (2015). THZ1 reveals roles for Cdk7 in co-transcriptional capping and pausing. Mol. Cell *59*, 576–587. https://doi.org/10.1016/J.MOLCEL.2015.06.032.
27. Fisher, R.P. (2018). Cdk7: a kinase at the core of transcription and in the crosshairs of cancer drug discovery. Transcription *10*, 47–56. https://doi.org/10.1080/21541264.2018.1553483.
28. Schroeder, S.C., Schwer, B., Shuman, S., and Bentley, D. (2000). Dynamic association of capping enzymes with transcribing RNA polymerase II. Genes Dev. *14*, 2435–2440. https://doi.org/10.1101/GAD.836300.
29. Viladevall, L., v.St Amour, C.V., Rosebrock, A., Schneider, S., Zhang, C., Allen, J.J., Shokat, K.M., Schwer, B., Leatherwood, J.K., and Fisher, R.P. (2009). TFIIH and P-TEFb coordinate transcription with capping enzyme recruitment at specific genes in fission yeast. Mol. Cell *33*, 738–751. https://doi.org/10.1016/J.MOLCEL.2009.01.029.
30. Komarnitsky, P., Cho, E.-J., and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes Dev. *14*, 2452–2460. https://doi.org/10.1101/gad.824700.
31. Noe Gonzalez, M., Sato, S., Tomomori-Sato, C., Conaway, J.W., and Conaway, R.C. (2018). CTD-dependent and -independent mechanisms govern co-transcriptional capping of Pol II transcripts. Nat. Commun. *9*, 3392. https://doi.org/10.1038/S41467-018-05923-W.
32. Ebmeier, C.C., Erickson, B., Allen, B.L., Allen, M.A., Kim, H., Fong, N., Jacobsen, J.R., Liang, K., Shilatifard, A., Dowell, R.D., et al. (2017). Human TFIIH kinase CDK7 regulates transcription-associated chromatin modifications. Cell Rep. *20*, 1173–1186. https://doi.org/10.1016/j.celrep.2017.07.021.
33. Rimel, J.K., Poss, Z.C., Erickson, B., Maas, Z.L., Ebmeier, C.C., Johnson, J.L., Decker, T.-M., Yaron, T.M., Bradley, M.J., Hamman, K.B., et al. (2020). Selective inhibition of CDK7 reveals high-confidence targets and new models for TFIIH function in transcription. Genes Dev. *34*, 1452–1473. https://doi.org/10.1101/gad.341545.120.
34. Sampathi, S., Acharya, P., Zhao, Y., Wang, J., Stengel, K.R., Liu, Q., Savona, M.R., and Hiebert, S.W. (2019). The CDK7 inhibitor THZ1 alters RNA polymerase dynamics at the 5′ and 3′ ends of genes. Nucleic Acids Res. *47*, 3921–3936. https://doi.org/10.1093/nar/gkz127.
35. Medler, S., and Ansari, A. (2015). Gene looping facilitates TFIIH kinase-mediated termination of transcription. Sci. Rep. *5*, 12586. https://doi.org/10.1038/srep12586.
36. Jaeger, M.G., and Winter, G.E. (2021). Fast-acting chemical tools to delineate causality in transcriptional control. Mol. Cell *81*, 1617–1630. https://doi.org/10.1016/j.molcel.2021.02.015.
37. Schwalb, B., Michel, M., Zacher, B., Frühauf, K., Demel, C., Tresch, A., Gagneur, J., and Cramer, P. (2016). TT-seq maps the human transient transcriptome. Science *352*, 1225–1228. https://doi.org/10.1126/science.aad9841.
38. Nojima, T., Gomes, T., Grosso, A.R.F., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, N.J. (2015). Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. Cell *161*, 526–540. https://doi.org/10.1016/j.cell.2015.03.027.
39. Nojima, T., Gomes, T., Carmo-Fonseca, M., and Proudfoot, N.J. (2016). Mammalian NET-seq analysis defines nascent RNA profiles and associated RNA processing genome-wide. Nat. Protoc. *11*, 413–428. https://doi.org/10.1038/nprot.2016.012.
40. Skene, P.J., and Henikoff, S. (2015). A simple method for generating highresolution maps of genome-wide protein binding. eLife *4*, e09225. https://doi.org/10.7554/eLife.09225.
41. Lopez, M.S., Kliegman, J.I., and Shokat, K.M. (2014). The logic and design of analog-sensitive kinases and their small molecule inhibitors. Methods Enzymol. *548*, 189–213. https://doi.org/10.1016/B978-0-12-397918-6.00008-2.
42. Larochelle, S., Merrick, K.A., Terret, M.-E., Wohlbold, L., Barboza, N.M., Zhang, C., Shokat, K.M., Jallepalli, P.V., and Fisher, R.P. (2007). Requirements for Cdk7 in the assembly of Cdk1/cyclin B and activation of Cdk2 revealed by chemical genetics in human cells. Mol. Cell *25*, 839–850. https://doi.org/10.1016/j.molcel.2007.02.003.
43. Gressel, S., Schwalb, B., and Cramer, P. (2019). The pause-initiation limit restricts transcription activation in human cells. Nat. Commun. *10*, 3603. https://doi.org/10.1038/s41467-019-11536-8.
44. Core, L., and Adelman, K. (2019). Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. Genes Dev. *33*, 960–982. https://doi.org/10.1101/gad.325142.119.
45. Scruggs, B.S., Gilchrist, D.A., Nechaev, S., Muse, G.W., Burkholder, A., Fargo, D.C., and Adelman, K. (2015). Upstream anti-sense promoters are hubs of transcription factor binding and active histone modifications. Mol. Cell *58*, 1101–1112. https://doi.org/10.1016/J.MOLCEL.2015.04.006.
46. Tan, S., Aso, T., Conaway, R.C., and Conaway, J.W. (1994). Roles for both the RAP30 and RAP74 subunits of transcription factor IIF in transcription initiation and elongation by RNA polymerase II. J. Biol. Chem. *269*, 25684–25691. https://doi.org/10.1016/S0021-9258(18)47303-6.
47. He, Y., Fang, J., Taatjes, D.J., and Nogales, E. (2013). Structural visualization of key steps in human transcription initiation. Nature *495*, 481–486. https://doi.org/10.1038/NATURE11991.
48. Chen, H.T., Warfield, L., and Hahn, S. (2007). The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. Nat. Struct. Mol. Biol. *14*, 696–703. https://doi.org/10.1038/nsmb1272.
49. Plaschka, C., Hantsche, M., Dienemann, C., Burzinski, C., Plitzko, J., and Cramer, P. (2016). Transcription initiation complex structures elucidate DNA opening. Nature *533*, 353–358. https://doi.org/10.1038/nature17990.
50. Compe, E., Genes, C.M., Braun, C., Coin, F., and Egly, J.M. (2019). TFIIE orchestrates the recruitment of the TFIIH kinase module at promoter before release during transcription. Nat. Commun. *10*, 2084. https://doi.org/10.1038/s41467-019-10131-1.
51. Chen, X., Yin, X., Li, J., Wu, Z., Qi, Y., Wang, X., Liu, W., and Xu, Y. (2021). Structures of the human Mediator and Mediator-bound preinitiation complex. Science *372*, eabg0635. https://doi.org/10.1126/science.abg0635.
52. Vos, S.M., Farnung, L., Urlaub, H., and Cramer, P. (2018). Structure of paused transcription complex Pol II–DSIF–NELF. Nature *560*, 601–606. https://doi.org/10.1038/s41586-018-0442-2.
53. Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H. (1999). NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell *97*, 41–51. https://doi.org/10.1016/S0092-8674(00)80713-8.
54. Gregersen, L.H., Mitter, R., Ugalde, A.P., Nojima, T., Proudfoot, N.J., Agami, R., Stewart, A., and Svejstrup, J.Q. (2019). SCAF4 and SCAF8, mRNA anti-terminator proteins. Cell *177*, 1797–1813.e18. https://doi.org/10.1016/j.cell.2019.04.038.
55. Grohmann, D., Nagy, J., Chakraborty, A., Klose, D., Fielden, D., Ebright, R.H., Michaelis, J., and Werner, F. (2011). The initiation factor TFE and the elongation factor Spt4/5 compete for the RNAP clamp during transcription initiation and elongation. Mol. Cell *43*, 263–274. https://doi.org/10.1016/j.molcel.2011.05.030.
56. Fujinaga, K., Irwin, D., Huang, Y., Taube, R., Kurosu, T., and Peterlin, B.M. (2004). Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the

transactivation response element. Mol. Cell. Biol. *24*, 787–795. https://doi.org/10.1128/MCB.24.2.787-795.2004.

57. Larochelle, S., Chen, J., Knights, R., Pandur, J., Morcillo, P., Erdjument-Bromage, H., Tempst, P., Suter, B., and Fisher, R.P. (2001). T-loop phosphorylation stabilizes the CDK7-cyclin H-MAT1 complex in vivo and regulates its CTD kinase activity. EMBO J. *20*, 3749–3759. https://doi.org/10.1093/emboj/20.14.3749.

58. Nechaev, S., Fargo, D.C., dos Santos, G., Liu, L., Gao, Y., and Adelman, K. (2010). Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in *Drosophila*. Science *327*, 335–338. https://doi.org/10.1126/science.1181421.

59. Vos, S.M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H., and Cramer, P. (2018). Structure of activated transcription complex Pol II–DSIF–PAF–SPT6. Nature *560*, 607–612. https://doi.org/10.1038/s41586-018-0440-4.

60. Krishnamurthy, S., He, X., Reyes-Reyes, M., Moore, C., and Hampsey, M. (2004). Ssu72 is an RNA polymerase II CTD phosphatase. Mol. Cell *14*, 387–394. https://doi.org/10.1016/S1097-2765(04)00235-7.

61. Greifenberg, A.K., Hönig, D., Pilarova, K., Düster, R., Bartholomeeusen, K., Bösken, C.A., Anand, K., Blazek, D., and Geyer, M. (2016). Structural and functional analysis of the Cdk13/Cyclin K complex. Cell Rep. *14*, 320–331. https://doi.org/10.1016/J.CELREP.2015.12.025.

62. Czudnochowski, N., Bösken, C.A., and Geyer, M. (2012). Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. Nat. Commun. *3*, 842. https://doi.org/10.1038/ncomms1846.

63. Tellier, M., Zaborowska, J., Caizzi, L., Mohammad, E., Velychko, T., Schwalb, B., Ferrer-Vicens, I., Blears, D., Nojima, T., Cramer, P., et al. (2020). CDK12 globally stimulates RNA polymerase II transcription elongation and carboxyl-terminal domain phosphorylation. Nucleic Acids Res. *48*, 7712–7727. https://doi.org/10.1093/nar/gkaa514.

64. Fan, Z., Devlin, J.R., Hogg, S.J., Doyle, M.A., Harrison, P.F., Todorovski, I., Cluse, L.A., Knight, D.A., Sandow, J.J., Gregory, G., et al. (2020). CDK13 cooperates with CDK12 to control global RNA polymerase II processivity. Sci. Adv. *6*, eaaz5041. https://doi.org/10.1126/sciadv.aaz5041.

65. Sawicka, A., Villamil, G., Lidschreiber, M., Darzacq, X., Dugast-Darzacq, C., Schwalb, B., and Cramer, P. (2021). Transcription activation depends on the length of the RNA polymerase II C-terminal domain. EMBO J. *40*, e107015. https://doi.org/10.15252/EMBJ.2020107015.

66. Zheng, Z.L. (2022). Cyclin-dependent kinases and CTD phosphatases in cell cycle transcriptional control: conservation across eukaryotic kingdoms and uniqueness to plants. Cells *11*, 279. https://doi.org/10.3390/CELLS11020279.

67. Garg, G., Dienemann, C., Farnung, L., Schwarz, J., Linden, A., Urlaub, H., and Cramer, P. (2023). Structural insights into human co-transcriptional capping. Mol. Cell *83*, 2464–2477.e5. https://doi.org/10.1016/j.molcel.2023.06.002.

68. Ahn, S.H., Keogh, M.C., and Buratowski, S. (2009). Ctk1 promotes dissociation of basal transcription factors from elongating RNA polymerase II. EMBO J. *28*, 205–212. https://doi.org/10.1038/EMBOJ.2008.280.

69. Fujiwara, R., Damodaren, N., Wilusz, J.E., and Murakami, K. (2019). The capping enzyme facilitates promoter escape and assembly of a follow-on preinitiation complex for reinitiation. Proc. Natl. Acad. Sci. USA *116*, 22573–22582. https://doi.org/10.1073/pnas.1905449116.

70. Aoi, Y., and Shilatifard, A. (2023). Transcriptional elongation control in developmental gene expression, aging, and disease. Mol. Cell *83*, 3972–3999. https://doi.org/10.1016/j.molcel.2023.10.004.

71. Chen, X., Liu, W., Wang, Q., Wang, X., Ren, Y., Qu, X., Li, W., and Xu, Y. (2023). Structural visualization of transcription initiation in action. Science *382*, eadi5120. https://doi.org/10.1126/science.adi5120.

72. Zhan, Y., Grabbe, F., Oberbeckmann, E., Dienemann, C., and Cramer, P. (2024). Three-step mechanism of promoter escape by RNA polymerase II. Mol. Cell *84*, 1699–1710.e6. https://doi.org/10.1016/j.molcel.2024.03.016.

73. Mandal, S.S., Chu, C., Wada, T., Handa, H., Shatkin, A.J., and Reinberg, D. (2004). Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. Proc. Natl. Acad. Sci. USA *101*, 7572–7577. https://doi.org/10.1073/pnas.0401493101.

74. Skene, P.J., Hernandez, A.E., Groudine, M., and Henikoff, S. (2014). The nucleosomal barrier to promoter escape by RNA polymerase II is overcome by the chromatin remodeler Chd1. eLife *3*, e02042. https://doi.org/10.7554/eLife.02042.

75. Rengachari, S., Schilbach, S., Aibara, S., Dienemann, C., and Cramer, P. (2021). Structure of the human Mediator–RNA polymerase II pre-initiation complex. Nature *594*, 129–133. https://doi.org/10.1038/s41586-021-03555-7.

76. Plaschka, C., Larivière, L., Wenzeck, L., Seizl, M., Hemann, M., Tegunov, D., v.Petrotchenko, E.V., Borchers, C.H., Baumeister, W., Herzog, F., et al. (2015). Architecture of the RNA polymerase II–Mediator core initiation complex. Nature *518*, 376–380. https://doi.org/10.1038/nature14229.

77. Diab, S., Yu, M., and Wang, S. (2020). CDK7 inhibitors in cancer therapy: the sweet smell of success? J. Med. Chem. *63*, 7458–7474. https://doi.org/10.1021/acs.jmedchem.9b01985.

78. Vervoort, S.J., Devlin, J.R., Kwiatkowski, N., Teng, M., Gray, N.S., and Johnstone, R.W. (2021). Targeting transcription cycles in cancer. Nat. Rev. Cancer *22*, 5–24. https://doi.org/10.1038/s41568-021-00411-8.

79. Parua, P.K., and Fisher, R.P. (2020). Dissecting the Pol II transcription cycle and derailing cancer with CDK inhibitors. Nat. Chem. Biol. *16*, 716–724. https://doi.org/10.1038/s41589-020-0563-4.

80. Kwiatkowski, N., Zhang, T., Rahl, P.B., Abraham, B.J., Reddy, J., Ficarro, S.B., Dastur, A., Amzallag, A., Ramaswamy, S., Tesar, B., et al. (2014). Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. Nature *511*, 616–620. https://doi.org/10.1038/NATURE13393.

81. Chipumuro, E., Marco, E., Christensen, C.L., Kwiatkowski, N., Zhang, T., Hatheway, C.M., Abraham, B.J., Sharma, B., Yeung, C., Altabef, A., et al. (2014). CDK7 inhibition suppresses super-enhancer-linked oncogenic transcription in MYCN-driven cancer. Cell *159*, 1126–1139. https://doi.org/10.1016/J.CELL.2014.10.024.

82. Christensen, C.L., Kwiatkowski, N., Abraham, B.J., Carretero, J., Al-Shahrour, F., Zhang, T., Chipumuro, E., Herter-Sprie, G.S., Akbay, E.A., Altabef, A., et al. (2014). Targeting transcriptional addictions in small cell lung cancer with a covalent CDK7 inhibitor. Cancer Cell *26*, 909–922. https://doi.org/10.1016/J.CCELL.2014.10.019.

83. Veo, B., Danis, E., Pierce, A., Wang, D., Fosmire, S., Sullivan, K.D., Joshi, M., Khanal, S., Dahl, N., Karam, S., et al. (2021). Transcriptional control of DNA repair networks by CDK7 regulates sensitivity to radiation in MYC-driven medulloblastoma. Cell Rep. *35*, 109013. https://doi.org/10.1016/J.CELREP.2021.109013.

84. Choi, Y.J., Kim, D.H., Yoon, D.H., Suh, C., Choi, C.M., Lee, J.C., Hong, J.Y., and Rho, J.K. (2019). Efficacy of the novel CDK7 inhibitor QS1189 in mantle cell lymphoma. Sci. Rep. *9*, 7193. https://doi.org/10.1038/s41598-019-43760-z.

85. Hu, S., Marineau, J.J., Rajagopal, N., Hamman, K.B., Choi, Y.J., Schmidt, D.R., Ke, N., Johannessen, L., Bradley, M.J., Orlando, D.A., et al. (2019). Discovery and characterization of SY-1365, a selective, covalent inhibitor of CDK7. Cancer Res. *79*, 3479–3491. https://doi.org/10.1158/0008-5472.CAN-19-0119.

86. Ganuza, M., Sáiz-Ladera, C., Cañamero, M., Gómez, G., Schneider, R., Blasco, M.A., Pisano, D., Paramio, J.M., Santamaría, D., and Barbacid, M. (2012). Genetic inactivation of Cdk7 leads to cell cycle arrest and induces premature aging due to adult stem cell exhaustion. EMBO J. *31*, 2498–2510. https://doi.org/10.1038/EMBOJ.2012.94.

87. Kelso, T.W.R., Baumgart, K., Eickhoff, J., Albert, T., Antrecht, C., Lemcke, S., Klebl, B., and Meisterernst, M. (2014). Cyclin-dependent

kinase 7 controls mRNA synthesis by affecting stability of preinitiation complexes, leading to altered gene expression, cell cycle progression, and survival of tumor cells. Mol. Cell. Biol. *34*, 3675–3688. https://doi.org/10.1128/MCB.00595-14.

88. He, G., Yang, X., Wang, G., Qi, J., Mao, R., Wu, Z., and Zhou, Z. (2017). Cdk7 is required for Activity-dependent neuronal gene expression, long-lasting synaptic plasticity and long-term memory. Front. Mol. Neurosci. *10*, 365. https://doi.org/10.3389/fnmol.2017.00365.
89. Egloff, S. (2012). Role of Ser7 phosphorylation of the CTD during transcription of snRNA genes. RNA Biol. *9*, 1033–1038. https://doi.org/10.4161/rna.21166.
90. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/BIOINFORMATICS/BTS635.
91. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
92. Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169. https://doi.org/10.1093/BIOINFORMATICS/BTU638.
93. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.
94. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods *14*, 417–419. https://doi.org/10.1038/nmeth.4197.
95. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. *17*, 10–12. https://doi.org/10.14806/ej.17.1.200.
96. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.
97. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/BIOINFORMATICS/BTQ033.
98. Chen, K., Chen, Z., Wu, D., Zhang, L., Lin, X., Su, J., Rodriguez, B., Xi, Y., Xia, Z., Chen, X., et al. (2015). Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes. Nat. Genet. *47*, 1149–1157. https://doi.org/10.1038/ng.3385.
99. Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372. https://doi.org/10.1038/nbt.1511.
100. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat. Methods *13*, 731–740. https://doi.org/10.1038/nmeth.3901.
101. Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., Kundu, D.J., Prakash, A., Frericks-Zipper, A., Eisenacher, M., et al. (2022). The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res. *50*, D543–D552. https://doi.org/10.1093/NAR/GKAB1038.
102. Gressel, S., Lidschreiber, K., and Cramer, P. (2019). Transient transcriptome sequencing: experimental protocol to monitor genome-wide RNA synthesis including enhancer transcription. https://www.protocols.io/view/transient-transcriptome-sequencing-experimental-pr-3byl42y22vo5/v1.
103. Caizzi, L., Monteiro-Martins, S., Schwalb, B., Lysakovskaia, K., Schmitzova, J., Sawicka, A., Chen, Y., Lidschreiber, M., and Cramer, P. (2021). Efficient RNA polymerase II pause release requires U2 snRNP function. Mol. Cell *81*, 1920–1934.e9. https://doi.org/10.1016/j.molcel.2021.02.016.
104. Schlackow, M., Nojima, T., Gomes, T., Dhir, A., Carmo-Fonseca, M., and Proudfoot, N.J. (2017). Distinctive patterns of transcription and RNA processing for human lincRNAs. Mol. Cell *65*, 25–38. https://doi.org/10.1016/J.MOLCEL.2016.11.029.
105. Conrad, T., and Ørom, U.A. (2017). Cellular fractionation and isolation of chromatin-associated RNA. Methods Mol. Biol. *1468*, 1–9. https://doi.org/10.1007/978-1-4939-4035-6_1.
106. Hughes, C.S., Moggridge, S., Müller, T., Sorensen, P.H., Morin, G.B., and Krijgsveld, J. (2018). Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. Nat. Protoc. *14*, 68–85. https://doi.org/10.1038/s41596-018-0082-x.
107. Humphrey, S.J., Karayel, O., James, D.E., and Mann, M. (2018). High throughput and high-sensitivity phosphoproteomics with the EasyPhos platform. Nat. Protoc. *13*, 1897–1916. https://doi.org/10.1038/S41596-018-0014-9.
108. Lidschreiber, K., Jung, L.A., von der Emde, H., Dave, K., Taipale, J., Cramer, P., and Lidschreiber, M. (2021). Transcriptionally active enhancers in human cancer cells. Mol. Syst. Biol. *17*, e9873. https://doi.org/10.15252/MSB.20209873.
109. Zacher, B., Michel, M., Schwalb, B., Cramer, P., Tresch, A., and Gagneur, J. (2017). Accurate promoter and enhancer identification in 127 ENCODE and roadmap epigenomics cell types and tissues by GenoSTAN. PLoS One *12*, e0169249. https://doi.org/10.1371/JOURNAL.PONE.0169249.
110. Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
111. Żylicz, J.J., Bousard, A., Žumer, K., Dossin, F., Mohammad, E., da Rocha, S.T., Schwalb, B., Syx, L., Dingli, F., Loew, D., et al. (2019). The implication of early chromatin changes in X chromosome inactivation. Cell *176*, 182–197.e23. https://doi.org/10.1016/J.CELL.2018.11.041.
112. Zumer, K., Maier, K.C., Farnung, L., Jaeger, M.G., Rus, P., Winter, G., Maier, K.C., Farnung, L., Jaeger, M.G., Rus, P., et al. (2021). Two distinct mechanisms of RNA polymerase II elongation stimulation in vivo Two distinct mechanisms of RNA polymerase II elongation stimulation in vivo. Mol. Cell *81*, 3096–3109.e8. https://doi.org/10.1016/j.molcel.2021.05.028.

Please cite this article in press as: Velychko et al., CDK7 kinase activity promotes RNA polymerase II promoter escape by facilitating initiation factor release, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.05.007

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-Pol II CTD antibody	Diagenode	C15200004; RRID:AB_2728744
anti-phospho-CTD Ser-5 of Pol II	Millipore	04-1572; RRID:AB_10615822
anti-MED26	Cell Signaling	14950; RRID:AB_2798656
anti-TFIIB	Santa Cruz	sc-271736; RRID:AB_10709889
anti-TFIIE-alpha	Abcam	ab28177; RRID:AB_778322
anti-NELF-E	Bethyl Laboratories	A301-913A; RRID:AB_1524076
anti-SPT5	Proteintech	16511-1-AP; RRID:AB_2878268
Spike-in Antibody	Active Motif	61686; RRID:AB_2737370
Chemicals, peptides, and recombinant proteins		
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich	D2438
1-NM-PP1	Sigma-Aldrich	529581
4-thiouridine	Carbosynth	NT06186
5-ethynyl uridine	Invitrogen	C10329
Empigen	Sigma-Aldrich	30326
Formaldehyde 16% concentrate stock methanol-free	Thermo Fisher Scientific	28908
MNase	NEB	M0247S
Benzonase	Sigma-Aldrich	E1014-25KU
RNAse A	Thermo Fisher Scientific	EN0531
Proteinase K	Life Technologies	AM2546
T4 Polynucleotide Kinase (PNK) (3' phosphatase minus)	NEB	M0236L
Agencourt AMPure XP beads	Beckman Coulter	A63881
Agencourt AMPure RNAClean XP	Beckman Coulter	A63987
Dynabeads M-280 Streptavidin	Thermo Fisher Scientific	11205D
Lipofectamine 2000	Invitrogen	11668019
Phenol:chloroform:isoamyl alcohol	Sigma-Aldrich	P2069
TRIzol	Invitrogen	15596026
Terminator 5' phosphate-dependent exonuclease	Biozym	162370
rSAP	NEB	M0371
Cap-Clip	Biozym	C-CC15011H
Critical commercial assays		
Plasmo Test Mycoplasma Detection Kit	InvivoGen	rep-pt1
CellTiter 96 AQueous One Solution Cell Proliferation Assay System	Promega	G3582
Click-iT® RNA Imaging Kit	Invitrogen	C10329
μMACS Streptavidin Kit	Miltenyi	130-074-101
miRNeasy Micro Kit	Qiagen	217084
Ovation Universal RNA-seq System	NuGEN	0343-32

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
TruSeq Small RNA Library Kit	Illumina	RS-200-0048
NEBNext® Ultra™ II DNA Library Prep Kit	NEB	E7645S
TMT10plex Mass Tag Labeling Kit	Thermo Fisher Scientific	90113
Deposited data		
Raw and processed sequencing data	This study	GEO: GSE218269
Raw and processed sequencing data	Tellier et al.[63]	GEO: GSE133156
Chromatin proteomics data	This study	PRIDE: PXD037842
Experimental models: Cell lines		
Human: HEK293	ATCC	ACC-10
Human: HEK293 CDK7^{as}	This study	N/A
D. melanogaster: Schneider-2 (S2) cell line	DSMZ	CRL-1963
Yeast:*S. cerevisiae* wt strain BY4741	Euroscarf	ACC-Y00000
Oligonucleotides		
Guide RNA 1: TAGCCTTGTCTTTGATTTTA	This study	N/A
Guide RNA 2: ATTTATGTCCAAAAGCATCA	This study	N/A
Genotyping primer CDK7 Fw: TAGTGGCTGAGAGTGTCAACCTG	This study	N/A
Genotyping primer CDK7 Rv: TCCTTTTGCCTCAGAGTACATCC	This study	N/A
Software and algorithms		
STAR (2.5.2b)	Dobin et al.[90]	https://github.com/alexdobin/STAR/ RRID:SCR_015899
SAMtools (1.3.1)	Li et al.[91]	https://sourceforge.net/projects/samtools/files/samtools/1.3.1/ RRID:SCR_002105
HTSeq (0.6.1.p1)	Anders et al.[92]	https://htseq.readthedocs.io/en/master/ RRID:SCR_005514
DESeq2	Love et al.[93]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html RRID:SCR_015687
Salmon (1.5.2)	Patro et al.[94]	https://github.com/COMBINElab/salmon RRID:SCR_017036
Cutadapt (1.9.1)	Martin1[95]	https://cutadapt.readthedocs.io/en/stable/ RRID:SCR_011841
Bowtie 2 (2.3.4.1)	Langmead and Salzberg[96]	https://bowtie-bio.sourceforge.net/bowtie2/index.shtml RRID:SCR_005476
BEDtools (2.29.1)	Quinlan et al.[97]	https://github.com/arq5x/bedtools2 RRID:SCR_006646
DANPOS3	Chen et al.[98]	https://github.com/sklasfeld/DANPOS3
MaxQuant (1.6.2.10)	Cox and Mann[99]	https://www.maxquant.org/ RRID:SCR_014485
Perseus (1.6.10.43)	Tyanova et al.[100]	https://maxquant.net/perseus/ RRID:SCR_015753

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Michael Lidschreiber (michael.lidschreiber@mpinat.mpg.de).

Please cite this article in press as: Velychko et al., CDK7 kinase activity promotes RNA polymerase II promoter escape by facilitating initiation factor release, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.05.007

Materials availability

Reagents generated in this study will be made available upon request.

Data and code availability

- Next-generation sequencing datasets generated in this study are available for download from GEO: GSE218269. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[101] partner repository with the dataset identifier PXD037842.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell lines and cell culture

Human HEK293 cells were obtained from the American Tissue Culture Collection (ATCC, CRL-1573) and cultured in Dulbecco's modified Eagle's medium (DMEM, Gibco, 11960085) supplemented with 10% fetal bovine serum (Gibco, 10500–064) and 1% GlutaMAX (Thermo Fisher Scientific, 35050087) in a humidified incubator at 37°C and 5% CO_2. Generated HEK293 CDK7^{as} cell line was cultured in the same condition. Cells were passaged every 2-3 days using 0.25% Trypsin-EDTA (Gibco, 25200056). Biological replicates were cultured separately. All cell lines were verified to be free of mycoplasma contamination with PlasmoTest Mycoplasma Detection Kit (InvivoGen).

Drosophila Schneider-2 (S2) cells, used for MNase-ChIP-seq spike-ins, were obtained from DSMZ (ATCC Cat# CRL-1963, RRID: CVCL_Z232) and cultured in Schneider's *Drosophila*-Medium (Gibco, 21720024) supplemented with 10% fetal bovine serum (Gibco, 10500–064) in a non-humidified incubator at 27°C, protected from light.

Yeast *S. cerevisiae* cells (BY4741 strain), used for mNET-seq spike-ins, were obtained from Euroscarf (ACC-Y00000) and cultured in yeast medium (YP + 2% glucose) at 30°C until OD595 of 0.8 was reached.

METHOD DETAILS

Genome editing with CRISPR/Cas9

The modified CDK7 genomic sequence (GenBank: AY130859.1) (500 bp on either side of the mutation) was cloned into pcDNA3 and used as the repair template for genome editing. The repair template contains a TTT (phenylalanine) to GGC (glycine) mutation. The following guide (g) RNA inserts were computationally designed and cloned into the pX462 vector (obtained from Addgene):

Guide RNA 1: TAGCCTTGTCTTTGATTTTA

Guide RNA 2: ATTTATGTCCAAAAGCATCA

HEK293 cells were transfected with gRNA vectors and a correction template using Lipofectamine 2000 (Invitrogen, 11668019) according to the manufacturer's instructions. Single clones were isolated by low-density plating after Puromycin and Neomycin selection for 48 hours. The mutation was validated by Sanger sequencing of PCR-amplified genomic DNA using the following primers:

Forward: TAGTGGCTGAGAGTGTCAACCTG

Reverse: TCCTTTTGCCTCAGAGTACATCC

Cell viability and proliferation assays

Cell proliferation analysis was measured in three biological replicates for 72 h with measurements every 12 h using the CellTiter 96 AQueous One Solution Cell Proliferation Assay System (Promega, G3582). 3,500 cells (HEK293 WT or CDK7^{as}) were seeded into each well of a 96-well plate and supplemented with 7.5 μM 1-NM-PP1 (Sigma-Aldrich, 529581) or an equal DMSO (solvent control; Sigma-Aldrich, D2438) volume. For each time point, the MTS tetrazolium compound was added to each sample for 90 min before quantifying relative absorbance with the Infinite M1000 Pro Plate Reader (TECAN, Männedorf, Switzerland) at 490 nm and 37°C. The averaged signals were corrected for culture medium-based background absorption. Statistical significance was determined using the two-sided Student's t-test with a p-value of $p < 0.05$.

Increasing concentrations of 1-NM-PP1 (0.12-15 μM) or DMSO (solvent control) were added to the culture medium with HEK293 WT or CDK7^{as} cells for the cell viability assay. After 72 h of incubation, the MTS tetrazolium compound was added to each sample for 90 min before quantifying relative absorbance with the Infinite M1000 Pro Plate Reader. Relative signals for each inhibitor concentration were calculated by subtracting culture medium-based background absorption from all values and dividing the signals of DMSO or 1-NM-PP1 treated cells by the corresponding signals of the no-treatment controls.

EU incorporation assay

$0.3x10^6$ cells grown on glass coverslips were treated for 30 min with 7.5 μM of 1-NM-PP1 (Cayman, 13330) or DMSO (solvent control) and with 0.5 mM 5-ethynyl uridine (EU, Invitrogen, C10329). After treatment, cells were washed with ice-cold PBS, fixed with 3.7% formaldehyde, and permeabilized with 0.5% Triton 100X. EU incorporation was revealed with the Click-iT® RNA Imaging Kit

(Invitrogen, C10329) using Alexa Fluor® 488 and Hoechst 33342 dyes according to the manufacturer's instructions. Coverslips were mounted in Vectashield medium and imaged by fluorescence microscopy using EVOS M7000 (Thermo Fisher Scientific). The mean EU fluorescence intensity per cell was obtained using ImageJ software (US National Institutes of Health) by averaging the mean gray value of the EU signal measured in each nucleus.

TT-seq and RNA-seq

TT-seq was performed in two biological replicates as previously described[102] with minor changes. 4x10^7 CDK7^{as} cells were treated for 15, 30, or 60 min with 7.5 μM of 1-NM-PP1 or DMSO (solvent control) with 4-thiouridine (4sU, Carbosynth, NT06186) included for the final 10 min. Cells were lysed with TRIzol reagent (Invitrogen, 15596026), and 120 ng of RNA spike-ins were added to each sample. Isolated total RNA was sonicated to obtain fragments of <10 kb using microTUBE AFA Fiber (Covatis, 520045) in a Covaris S220 Focused-ultrasonicator with the following parameters: 1% duty cycle, 100 W peak incident power, 200 cycles per burst, 10 sec processing time, 6°C bath temperature, continuous degassing mode, water level 12. 4sU-labeled RNA was biotinylated and purified with the μMACS streptavidin kit (Miltenyi Biotec, 130-074-101) from 240 μg of total fragmented RNA, which was split into two reactions (2x120 μg) and pooled after purification. Next, RNA was cleaned up with the miRNeasy Micro Kit (Qiagen, 217084). Libraries were prepared from 100 ng of RNA with the strand-specific Ovation Universal RNA-Seq library kit (NuGEN, 0343-32). Fragment sizes were analyzed with the Fragment Analyzer (HS NGS Fragment Kit). Pooled libraries were sequenced on the Illumina NEXTseq550 with a 2x75 base paired-end mode.

mNET-seq

mNET-seq was performed in two biological replicates as previously described[24,39,103,104] with minor changes. 5x10^7 CDK7^{as} cells were treated for 15 or 30 min with 7.5 μM of 1-NM-PP1 or DMSO (solvent control). After the treatment, chromatin isolation was performed as described in the protocol for cellular fractionation[105] using 2.5x10^7 cells per reaction. All buffers were supplemented with 7.5 μM 1-NM-PP1 or respective amounts of DMSO, protease (Sigma-Aldrich, P8340), and phosphatase (Roche, 4906837001) inhibitors. After the fractionation, the chromatin pellet was digested in 100 μl MNase digest buffer (1x MNase buffer, 1x BSA, 100 mM NaCl, 50 units/μl MNase (NEB, M0247S)) at 37°C and 1,400 rpm for 2 min. The digestion was stopped by adding 11 μl of 250 mM EGTA, and samples were centrifuged at 13,000 g for 5 min at 4°C. Supernatants were pooled together for each sample and 8-fold diluted with IP buffer (50 mM Tris HCl pH 7.4, 150 mM NaCl, 0.05% NP-40, 0.3% empigen (Sigma-Aldrich, 30326)). 15 μg of anti-Pol II CTD antibody (Diagenode, C15200004) conjugated with 125 μl of Dynabeads M-280 Sheep Anti-Mouse IgG (Invitrogen, 11201D) were added to each sample and incubated on a rotating wheel at 4°C and 12 rpm for 1 h. After IP, beads were washed seven times with 1 ml of IP buffer and once with 300 μl of PNKT buffer (1x T4 PNK buffer (NEB, M0236L), 0.1% Tween-20). After washes, 5′ RNA ends were phosphorylated with T4 Polynucleotide Kinase (PNK) (3' phosphatase minus; NEB, M0236L) at 800 rpm for 10 min at 37°C. Next, beads were washed once with IP buffer, and 0.5 ml of TRIzol reagent (Invitrogen, 15596026) were added to the beads. 2.5 ng of *S. cerevisiae* spike-ins were added to each sample in TRIzol. RNA spike-ins for mNET-seq were obtained from *S. cerevisiae* cells as previously described.[103] RNA was extracted from beads, followed by RNA precipitation with 100% ethanol and GlycoBlue coprecipitant (Invitrogen, AM9515) at -80°C overnight. RNA pellets were washed twice with 70% ethanol, air-dried, and dissolved in 7 M urea. In order to size select 25-110 nt RNA, samples were resolved on 6% polyacrylamide gel containing 7 M urea. RNA of corresponding sizes was extracted from the gel and precipitated with 100% ethanol and GlycoBlue coprecipitant. 8.4 ng of RNA for replicate 1 and 12.3 ng of RNA for replicate 2 were used for library preparation with TruSeq Small RNA Library Prep Kit (RS-200-0048) according to the manufacturer's instructions. Fragment sizes were analyzed with the Fragment Analyzer (HS NGS Fragment Kit). Pooled libraries were sequenced on the Illumina NEXTseq550 with a 2x42 base paired-end mode.

MNase-ChIP-seq

MNase-ChIP-seq was performed in three biological replicates as previously described[40] with the following changes. 4x10^7 CDK7^{as} cells were treated for 30 min with 7.5 μM of 1-NM-PP1 or DMSO (solvent control), following 8 min crosslinking with 1% formaldehyde (16% formaldehyde (w/v), methanol-free, Thermo Fisher Scientific, 28908). The crosslinking was quenched for 5 min with 125 mM glycine, and cells were washed with 10 ml PBS at room temperature. 2 ml of ice-cold PBS was added to plates for cell scraping. Scraped cells were transferred to falcons, washed twice with 5 ml ice-cold PBS, and centrifuged at 200 g for 5 min at 4°C. All buffers were supplemented with protease (Sigma-Aldrich, P8340) and phosphatase (Roche, 4906837001) inhibitors. Cell pellets were lysed with 5 ml of Farnham Lysis buffer (5 mM Pipes pH 8, 85 mM KCl, 0.5% NP-40) for 10 min on ice. Nuclei were pelleted by centrifugation at 1,700 g for 5 min at 4°C and washed with ice-cold PBS. Nuclei pellets were resuspended in 200 μl of 1% SDS Lysis buffer (50 mM Tris HCl pH 8.0, 10 mM EDTA, 1% SDS) and incubated for 10 min at room temperature to permeabilize nuclei. The reaction was quenched with 1.8 ml of IP buffer (20 mM Tris HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100) supplemented with 5 mM $CaCl_2$ and incubated for 2 min at 37°C before the addition of 16,000U of MNase (NEB, M0247S) for 20 min at 37°C. MNase digestion was quenched with 40 μl of EDTA (500 mM) and 20 μl EGTA (500 mM). 1 ml of sample was transferred to AFA milliTubes (Covaris, 520135) and sonicated with a Covaris S220 Focused-ultrasonicator with the following parameters: 5% duty cycle, 140 W peak incident power, 200 cycles per burst, 600 sec processing time, 6°C bath temperature, continuous degassing mode, water level 8. Sonicated samples were centrifuged at 10,000 g for 15 min at 4°C, and supernatants were snap-frozen in liquid nitrogen and stored at −80°C before use. For each IP, 30 μg of sample chromatin and 200 ng of *Drosophila* S2 MNase-digested crosslinked chromatin were diluted in IP buffer to obtain a final concentration of 0.05% SDS and a final volume of 1.6 ml. Pre-clearing was carried out by adding 20 μl of Dynabeads Protein G (Thermo Fisher Scientific, 10009D) diluted in IP

buffer, and incubating for 30 min at 4°C on the rotation wheel. Pre-cleared samples were incubated with 10 μg of Pol II (RPB1, Diagenode, C15200004), 10 μg of Ser5P (phospho-CTD Ser-5 of RPB1, Merck Millipore, 04-1572), 8 μl of Mediator (MED26, Cell Signaling, 14950), 6 μg of TFIIB (Santa Cruz, sc-271736), 5 μg of TFIIE (TFIIE-α, abcam, ab28177), 6 μg of NELF (NELF-E, Bethyl Laboratories, A301-913A) or 6 μg of SPT5 (Proteintech, 16511-1-AP) together with 1 μg of Spike-in (Active Motif, 61686) antibodies at 4°C overnight. The next day samples were centrifuged at 16,000g for 2 min at 4°C and inputs were taken for each sample. Samples were incubated for 2 h at 4°C with Dynabeads Protein G. Beads were washed once with Buffer 1 (20 mM Tris HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS), four times with Buffer 2 (20 mM Tris HCl pH 8.0, 500 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS), once with Buffer 3 (10 mM Tris HCl pH 8.0, 250 mM LiCl, 1 mM EDTA, 1% NP-40, 1% sodium-deoxycholate), and three times with TE buffer (10 mM Tris HCl pH 8.0, 50 mM NaCl, 1 mM EDTA). After the last wash, 200 μl of elution buffer (0.1 M $NaHCO_3$, 160 mM NaCl, 1% SDS) was added to the beads, and samples were de-crosslinked overnight at 65°C. Next, samples were diluted with 200 μl of PBS and treated with 0.02 μg/μl RNAse A (Thermo Scientific, EN0531) for 1.5 h at 37°C and 0.2 μg/μl Proteinase K (Life Technologies, AM2546) for 2 h at 45°C. DNA was extracted with an equal volume (400 μl) of phenol/chloroform/isoamyl alcohol (Sigma-Aldrich, P2069), followed by DNA precipitation with 100% ethanol and 200 mM NaCl for 1 h at -80°C. DNA pellets were washed twice with 70% ethanol, air-dried, and dissolved in water. Sample concentrations were measured with a Qubit dsDNA HS Assay Kit. Libraries were prepared with the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, E7645S) according to the manufacturer's instructions adjusted for short fragments end repair, 5′ phosphorylation, and dA-tailing thermocycler program (30 min at 20°C, 1 h at 50°C; the heated lit set to 60°C). After amplification, libraries were double-sided (1.0 – 1.6x) size-selected with Agencourt AMPure XP beads (Beckman Coulter, A63881). Fragment sizes were analyzed with the Fragment Analyzer (HS NGS Fragment Kit). Pooled libraries were sequenced on the Illumina NEXTseq550 with a 2x42 base paired-end mode.

Start-seq

Start-seq was performed in two biological replicates as previously described[58] with minor changes. $3.5x10^7$ $CDK7^{as}$ cells were treated for 30 min with 7.5 μM of 1-NM-PP1 or DMSO (solvent control). For spike-in normalization, 2.5 $x10^7$ *Drosophila* S2 cells were added to each sample. After the treatment, nuclei isolation was performed as described in the protocol for cellular fractionation.[105] All buffers were supplemented with 7.5 μM 1-NM-PP1 or respective amounts of DMSO, protease (Sigma-Aldrich, P8340), and phosphatase (Roche, 4906837001) inhibitors. RNA was extracted from the isolated nuclei using 5 ml of TRIzol reagent (Invitrogen, 15596026), precipitated with 100% ethanol and GlycoBlue coprecipitant (Invitrogen, AM9515) at -20°C for 30 minutes, then washed twice with 70% ethanol. The RNA pellets were air-dried and dissolved in 7 M urea. To size select 16-65 nt RNA, samples were resolved on 9% polyacrylamide gel containing 7 M urea. RNA of corresponding sizes was extracted from the gel and precipitated with 100% ethanol and GlycoBlue coprecipitant. After short RNA isolation, uncapped pRNAs were degraded using Terminator 5′ phosphate-dependent exonuclease (Biozym, 162370) at 30°C for 60 min. After the reaction, short RNA was cleaned up using 1.8x of AMPure RNAClean XP (Beckman Coulter, A63987) beads and 270 μl of isopropanol. Next uncapped pppRNAs and pRNAs were dephosphorylated using rSAP (NEB, M0371) at 37°C for 30 min. RNA was cleaned up using the AMPure RNAClean XP beads, and RNA was uncapped with Cap-Clip (Biozym, C-CC15011H) at 37°C for 60 min. RNA was cleaned up using the AMPure RNAClean XP beads. 44 ng of RNA was used for library preparation with TruSeq Small RNA Library Prep Kit (RS-200-0048) according to the manufacturer's instructions. Fragment sizes were analyzed with the Fragment Analyzer (HS NGS Fragment Kit). Pooled libraries were sequenced on the Illumina NEXTseq550 with a 2x42 base paired-end mode.

MNase-seq

MNase-seq was performed in two biological replicates. $5x10^6$ $CDK7^{as}$ cells were treated for 15 min with 7.5 μM of 1-NM-PP1 or DMSO (solvent control) and centrifuged at 500 g for 3 min at 4°C. Cell pellets were resuspended in 50 μl of Buffer I (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 2.5 mM $MgCl_2$, 0.5% NP-40, 10 % sucrose, 0.5 mM $CaCl_2$) and incubated at 37°C for 2 min. Next, we added 500 μl of Buffer II (50 mM Tris-HCl pH 7.5, 50 mM NaCl, 2.5 mM $MgCl_2$, 0.5% NP-40, 10 % sucrose, 2 mM $CaCl_2$) supplemented with 3,000 U of MNase (NEB, M0247S) and incubated at 37°C for 9 min. MNase digestion was quenched with 11 μl of EGTA (500 mM). 550 μl of SDS Lysis Buffer (50 mM Tris-HCl pH 8.0, 1% SDS, 10 mM EDTA) was added, and samples were incubated on ice for 10 min. Samples were then treated with 0.1 μg/μl RNAse A (Thermo Fisher Scientific, EN0531) for 1.5 h at 37°C and 0.05 μg/μl Proteinase K (Life Technologies, AM2546) for 4 h at 65°C. DNA was extracted with an equal volume of phenol/chloroform/isoamyl alcohol (Sigma-Aldrich, P2069), followed by DNA precipitation with an equal volume of 100% ethanol and 200 mM NaCl for 2 h at -20°C. DNA pellets were washed twice with 1 ml of 70% ethanol, air-dried, and dissolved in water. Sample concentrations were measured with a Qubit dsDNA HS Assay Kit. Libraries were prepared with the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, E7645S) according to the manufacturer's instructions. After amplification, libraries were double-sided (0.6 – 1.0x) size-selected with Agencourt AMPure XP beads (Beckman Coulter, A63881). Fragment sizes were analyzed with the Fragment Analyzer (HS NGS Fragment Kit). Pooled libraries were sequenced on the Illumina NEXTseq550 with a 2x42 base paired-end mode.

Chromatin and phospho-proteomics

Chromatin proteomics was performed in four biological replicates. $7x10^6$ $CDK7^{as}$ cells were seeded on 15 cm plates 20-24 h prior to the experiment following 30 min of 7.5 μM 1-NM-PP1 or DMSO (solvent control) treatment. After the treatment, chromatin isolation was performed as described in the protocol for cellular fractionation.[105] All buffers were supplemented with 7.5 μM 1-NM-PP1 or respective amounts of DMSO, protease (Sigma-Aldrich, P8340), and phosphatase (Roche, 4906837001) inhibitors. After the

Please cite this article in press as: Velychko et al., CDK7 kinase activity promotes RNA polymerase II promoter escape by facilitating initiation factor release, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.05.007

fractionation, the chromatin pellet was digested in 100 μl chromatin digest buffer (50 mM Tris HCl pH 7.4, 2.5 mM $MgCl_2$, 2.5 mM $CaCl_2$, 8 units/μl MNase (NEB, M0247S), 2.5 units/μl Benzonase (Sigma-Aldrich, E1014-25KU) at 4°C for 90 min. The digestion was stopped by the addition of 100 μl 2X SDS buffer (100 mM Tris HCl pH 7.4, 4% SDS), aiding the solubilization and homogenization of the protein solution. The sample was boiled at 99°C for 5 min to denature proteins. The remaining insoluble material was pelleted by centrifugation at 10,000 g for 10 min at RT. 400 μg of protein per sample were subjected to reduction with 5 mM tris-(2-carboxyethyl)-phosphine (TCEP) and alkylation with 10 mM 2-Iodoacetamide (IAA) for 40 to 50 min at room temperature. Single-pot solid-phase-enhanced sample preparation (SP3) on beads was conducted to remove contaminants from the protein solution.[106] In brief, proteins were bound to magnetic beads via hydrophilic interactions. Bound beads were trapped in a solvation layer and were washed for contaminant removal. Purified proteins were cleaved into peptides via trypsin digestion, thereby dissociated from the beads and eluted in aqueous conditions without peptide cleanup.

TMT-labelling with TMT10plex Mass Tag Labeling Kit (ThermoFisher Scientific, 90113) was performed according to the manufacturer's instructions. Importantly, 400 μg of protein instead of 100 μg have been labeled within a total volume of 200 μl instead of 100 μl. Based on the increased total volume of 200 μl, TMT was redissolved in 82 μl acetonitrile to keep the volume ratio constant. Following labeling, equal amounts of each sample were combined for further processing. After completion of TMT labeling, peptide samples were desalted using Sep-Pak C18 1 cc Vac Cartridges.

Phospho-peptide enrichment with TiO_2 beads was performed according to the published EasyPhos protocol.[107] Subsequent desalting was done utilizing waters sep-pak for non-phospho-peptides and Harvard Apparatus C18 tips for phospho-peptides (Micro SpinColumns, C18, Qty 96). Peptides were fractionated by high pH reversed-phase chromatography. 48 fractions were collected and pooled into 12 fractions.

The fractions obtained from high pH reversed-phase chromatography were analyzed by liquid chromatography coupled to mass spectrometry (LC-MS) on a Dionex UltiMate 3000 RSLCnano system (Thermo Fisher Scientific) connected to a Q Exactive HF-X hybrid quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific). After drying, peptides were resuspended in 2.5% acetonitrile (ACN), 0.1% trifluoroacetic acid (TFA) (v/v) and injected in technical triplicates onto a C18 PepMap100 μ-Precolumn (0.3x5 mm, 5 μm, Thermo Fisher Scientific). Chromatographic separation was performed on an in-house packed main column (75 μm x 30 cm, reprosil-Pur 120C18-AQ, 1.9 μm, Dr. Maisch GmbH, Ammerbuch, Germany) at a 300 nl/min flow rate. The chromatographic gradient was composed as follows: 2% to 5% buffer B (80% ACN, 0.08% formic acid, v/v) from 0 to 5 min, 5% to 36% buffer B from 5 to 100 min, 36% to 45% buffer B from 100 to 106 min, followed by an equilibration sequence with 90% and 5% buffer B. The total duration of the method was 118 min. MS1 spectra were acquired with 120,000 resolution (full width at half maximum, FWHM), $1x10^6$ automatic gain control (AGC) target, and 50 ms maximum injection time from 350 to 1,600 m/z scan range. The 25 most abundant precursor ions per duty cycle with a charge state between +2 and +6 were selected individually with a 0.8 m/z isolation window and were fragmented with a stepped normalized collision energy of 30 and 38. MS2 spectra were acquired with 45,000 resolution (FWHM), $4x10^3$ minimum AGC and $2 x 10^5$ target AGC, 86 ms maximum injection time, and a fixed first mass of 110 m/z. Once selected, precursors were excluded from another fragmentation event for 60 s.

QUANTIFICATION AND STATISTICAL ANALYSIS

MS database search

Raw acquisition files of the proteome and phospho-proteome samples were analyzed together with Maxquant version 1.6.2.10[99] with default parameters unless stated below. The isotopic distributions of the TMT10plex reporter ions were adapted as determined for the TMT lot used in the present study. The proteome and phospho-proteome samples were defined as separate parameter groups. Phospho-proteome samples were set as True in the PTM column to indicate a PTM enrichment and exclude identifications derived from these files for protein quantification. The quantification type was set to Reporter ion MS2 using 10plex TMT on lysine and the N-terminus peptide, but excluding the channel 126C since it was not used. A precursor ion fraction of a minimum of 0.75 was applied as a filter. Searches were performed against a reviewed database of human protein sequences derived from Swiss-prot/UniProtKB (20,362 sequences, downloaded on 03/03/2020). Carbamidomethylation on cysteine was set as fixed modification, while oxidation on methionine and acetylation on the protein N-terminus were set as variable modification. A variable phosphorylation on serine, threonine, and tyrosine was also considered in the separate parameter group of the phospho-proteome sample. Enzymatic activity was set to trypsin with a maximum of three missed cleavages. The match between runs was enabled, while the second peptide option was disabled.

MS data analysis

Database search results were analyzed with Perseus version 1.6.10.43.[100] The Phospho (STY)Sites.txt and the proteinGroups.txt result tables were loaded into the software and identifications only identified by site, contaminant, and reverse hits were removed. Quantitative data from the TMT channels were log2 transformed, and phospho-peptides with a localization probability lower than 0.75 were removed. Phospho-peptide abundances were normalized against the protein abundances to adjust for proteins that change their association with the chromatin fraction in response to inhibition. Both quantitative data types were normalized against the pooled reference sample and were median-normalized to adjust for different starting amounts. A two-sample t-test was performed with 0.1 artificial within groups variance (s0) and 5% permutation-based false-discovery rate for multiple hypothesis testing

correction. Identifications with significantly changed abundance were Z-score normalized over the rows of the matrix and hierarchically clustered with default parameters (k-means clustering, Euclidean distances, Average linkages).

Annotation

For all the NGS analyses, the human genome, transcriptome, and annotation were obtained for GRCh38.p13 (RefSeq assembly accession GCF_000001405.39) from NCBI.

Major isoform annotation

The major isoform annotation was done with Salmon (1.5.2).[94] Salmon calculated transcript abundances in transcripts per million (TPM) units by quantifying RNA-seq data for the human transcriptome. For curated RefSeq annotated transcript isoforms, the major isoform was determined as the one with the maximum mean transcripts per million (TPM) value across all RNA-seq samples. The final major isoform annotation excludes overlapping genes as well as isoforms located on chromosomes X, Y, and M [n =12,585].

GenoSTAN annotation of transcribed units and their classification for HEK293 CDK7^{as}

Annotation of different transcript classes, in particular enhancer RNAs (eRNAs), was done as in Lidschreiber et al.[108] with few modifications. In brief, genome-wide coverage was calculated from all TT-seq fragment midpoints in consecutive 200 bp bins throughout the genome. A two-state hidden Markov model with a Poisson-Log-Normal emission distribution was learned in order to segment the genome into 'transcribed' and 'non-transcribed' states. Consecutive 'transcribed' states were joined if their gaps were less than 200 bp, within an annotated mRNA or lincRNA, or showed uninterrupted coverage supported by all TT-seq samples. The resulting transcribed units (TUs) were further filtered using a minimal expression threshold that was defined based on overlap with genes annotated in GENCODE (v21). The threshold was optimized using the Jaccard index criterion and resulted in 32,789 TUs.

TUs that overlapped at least 25% of an annotated protein-coding gene and overlapped with an annotated exon of the corresponding gene were classified as mRNAs. Remaining TUs were annotated as non-coding (nc) RNAs and further classified according to their genomic location relative to protein-coding genes: upstream antisense RNA (uaRNA), convergent RNA (conRNA), antisense RNA (asRNA), and intergenic RNA. ncRNAs located on the opposite strand of an mRNA were classified as asRNA if the TSS was located > 1 kbp downstream of the sense TSS, as uaRNA if the TSS was located < 1 kbp upstream of the sense TSS, and as conRNA if the TSS was located < 1 kbp downstream of the sense TSS. All remaining ncRNAs were classified as intergenic. We further classified intergenic and asRNAs as eRNAs, if their TSS ± 500 bp overlapped with an enhancer state annotated by GenoSTAN.[109] Additionally, we restricted our set of antisense eRNAs to only those that overlap with publicly available DHS and H3K27ac peaks (ENCODE ENCFF680DCW and ENCFF451UZW), and intergenic eRNAs to only those that originate from regions with TT-seq detected transcription on both strands.

Analysis of high-throughput sequencing data

All sequencing data were demultiplexed with Illumina bcl2fastq demultiplexing software, provided the barcodes for each of the samples, and subsequently quality checked with FastQC (Babraham Institute).[110]

RNA-seq and TT-seq data preprocessing and normalization

For all samples, paired-end 75 base reads were mapped using STAR (v 2.6.1b)[90] to the GRCh38.p13 genome assembly merged with synthetic RNA spike-in sequences[37,111] with a maximum 2% mismatches [mapping parameters: –outFilterMismatchNoverLmax 0.02—outFilterMultimapScoreRange 0 –alignIntronMax 500000]. SAMtools (v 1.3.1)[91] was used to quality filter SAM files, whereby alignments with MAPQ smaller than 7 (-q 7) were skipped, and only proper pairs (-f2) were selected.

A spike-in (RNA) normalization strategy was used essentially as described.[37] This spike-in normalization allows for calculating sequencing depth σ_j, cross-contamination rate ϵ_j, and antisense bias ratio c_j. Sequencing depth σ_j allows to observe global shifts. The read counts (k_{ij}) for spike-ins were calculated using HTSeq.[92] Calculations for each parameter are described in more detail below.

Antisense bias ratio c_j, the ratio of spurious reads originating from the opposite strand introduced by the reverse transcription reaction, was calculated for each sample j according to

$$c_j = \underset{i}{\text{median}}\left(\frac{k_{ij}^{\text{antisense}}}{k_{ij}^{\text{sense}}}\right)$$

for all available spike-ins i. The antisense bias corrected read counts or coverage s_{ij} for transcribed unit i in sample j was calculated as

$$s_{ij} = \frac{S_{ij} - c_j A_{ij}}{1 - c_j^2}$$

where S_{ij} and A_{ij} are the read counts observed or coverage on the sense and antisense strand.

Please cite this article in press as: Velychko et al., CDK7 kinase activity promotes RNA polymerase II promoter escape by facilitating initiation factor release, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.05.007

Cross-contamination rate ϵ_j. The cross-contamination rate ϵ_j, proportion of unlabeled reads purified in the TT-seq samples, was calculated for each sample j as

$$\epsilon_j = \underset{i}{\text{median}}\left(\frac{k_{ij}}{l_i}\right) \Big/ \sigma_j$$

using the unlabeled spike-ins i for TT-seq samples.

RNA amount per cell

A conversion factor RNA amount per cell [cell^{-1}] is calculated given the sequence and mixture of the RNA spike-ins as described.[24,43] The sequence of spike-ins allows to calculate their molecular weight M (assuming perfect RNA extraction) as

$$M = A_n \times 329.2 + (1 - \tau) \times U_n \times 306.2 + C_n \times 305.2 + G_n \times 345.2 + \tau \times 4sU_n \times 322.26 + 159$$

where A_n,U_n, C_n, G_n and $4sU_n$ are the number of each respective nucleotide within each spike-in polynucleotide. For labeled spike-ins τ is set to 0.1 and otherwise 0. The addition of 159 corresponds to the molecular weight of a 5′ triphosphate.

The number of spike-in molecules per cell N [cell^{-1}] was calculated as

$$N = \frac{m}{M} N_A$$

where m is the number of spike-ins 164 $\times$ 10^{-9} g in 7.04 $\times$ 10^6 number of cells, M is the molecular weight of the spike-ins, and N_A is the Avogadro constant.

The conversion factor to RNA amount per cell k [cell^{-1}] was then calculated for all labeled spike-ins i with length L_i as

$$k = \text{mean}\left(\text{median}_i\left(\frac{tb_i}{L_i \cdot N}\right)\right)$$

Read counts per kilobase (RPK) were calculated using antisense bias corrected read counts (k_j) falling into the region of a major isoform transcript divided by its length in kilobases.

Expressed gene set for analysis

For downstream analyses, the expressed gene set is defined as follows: based on antisense bias corrected RPKs, a group of expressed major transcript isoforms (n= 8,950) was defined to comprise all transcript isoforms with a median RPK of 20 or greater in all TT-seq control samples. An RPK of 20 corresponds to approximately a coverage of 4 per sample due to an average fragment size of 200. For MNase-ChIP-seq analysis only protein-coding genes were used (n = 7,075).

Response ratio

For each condition j (control or CDK7^{as} inhibited), the antisense bias corrected number of transcribed bases tb_{ji} was calculated for all expressed major transcript isoforms i exceeding 10 kbp in length. The 5' response ratio was calculated for a window from TSS + 500 bp to TSS + 10 kbp downstream and 3′ response ratio was calculated for a window pA - 10 kbp to pA for each i as

$$r_i = \frac{tb_i^{\text{CDK7as inhibited}}}{tb_i^{\text{Control}}}$$

Productive initiation frequency

The productive initiation frequency was calculated as described[24,43] for all major transcript isoforms i exceeding 10 kbp in length. The productive initiation frequency l_i was calculated for each condition j (control or CDK7^{as} inhibited) using sequencing depth, and antisense bias corrected TT-seq coverage on exons spanning a window from the TSS + 500 bp to TSS + 10 kbp downstream, excluding the first exon, as

$$l_i = \frac{1}{k} \cdot \frac{tb_i^{\text{Control}}}{t \cdot L_i}$$

with labeling duration t = 10 and length L_i.

mNET-seq data preprocessing and normalization

For all samples, paired-end 42 base reads were trimmed for adapters (-a TGGAATTCTCGGGTGCCAAGG -A GATCGTCGGACT), and low-quality bases (< Q20) were removed with Cutadapt (v 1.9.1).[95] The reads were then mapped to the GRCh38.p13 genome assembly merged with the yeast genome version SacCer3 using STAR (v 2.5.2b)[90] with a maximum of 2% mismatches [mapping parameters: –outFilterMultimapScoreRange 0 –outFilterMismatchNoverLmax 0.02 –outFilterMatchNmin 16 –outFilterScoreMinOverLread 0 –outFilterMatchNminOverLread 0 –alignIntronMax 500000]. Samtools (v 1.3.1)[91] was used to quality filter SAM files, where alignments with MAPQ smaller than 7 (-q 7) were skipped, and only proper pairs (-f 2) were selected.

Please cite this article in press as: Velychko et al., CDK7 kinase activity promotes RNA polymerase II promoter escape by facilitating initiation factor release, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.05.007

Read counts (k_{ij}) for yeast genes for each condition j (control or CDK7^{as} inhibited) were calculated using HTSeq.[92] The counts data was used to generate normalization factors with DESeq2[93] to normalize the mNET-seq data. Coverage tracks for further analysis were restricted to the last nucleotide incorporated by Pol II in the aligned mNET-seq reads.

MNase-ChIP-seq data preprocessing and normalization

For all samples, paired-end 42 base reads were mapped using Bowtie 2 (v 2.3.4.1)[96] to the merged human hg38 (GRCh38.p13) and Drosophila (BDGP6.28) genome assembly [mapping parameters: –no-discordant –no-mixed –very-sensitive]. SAMtools (v 1.3.1)[91] was used to quality filter SAM files, whereby alignments with MAPQ smaller than 7 (-q 7) were skipped, and only proper pairs (-f2) were selected. Further data processing was carried out using the R/Bioconductor environment. Duplicate fragments and fragments longer than 100 bp (70 bp for Pol II) were excluded from further analysis. MNase-ChIP-seq signals were obtained from piled-up fragment midpoint counts for every genomic position. Pol II, Ser5P, and SPT5 signals were normalized to the signal at non-transcribed regions to allow for the detection of global changes. To this end, genome-wide coverage was calculated from all MNase-ChIP-seq fragment midpoints in consecutive 2 kbp bins throughout the non-transcribed human genome (i.e. excluding the 32,789 TUs we annotated using GenoSTAN[109] and everything annotated in GENCODE v31) and used to generate normalization factors with DESeq2.[93] For all other factors, that exhibited a peak-like binding behavior, normalization factors were obtained from total human fragment counts. To obtain normalized MNase-ChIP-seq coverages for subsequent analyses, the coverages were divided by the respective normalization factors and then summed over replicates. MNase-ChIP-seq samples were also cross-validated using a spike-in normalization strategy to ensure consistency, before applying the more robust normalization procedures described above.

Start-seq data preprocessing and normalization

For all samples, paired-end 42 base reads were trimmed for adapters and then mapped to the merged human hg38 (GRCh38.p13) and Drosophila (BDGP6.28) genome assembly using STAR (v 2.6.0c)[90] with default parameters. Samtools (v 1.3.1)[91] was used to quality filter SAM files, where alignments with MAPQ smaller than 7 (-q 7) were skipped, and only proper pairs (-f 2) were selected. Total read counts mapping to the Drosophila genome were used to generate normalization factors to normalize the Start-seq data. Coverage tracks for further analysis were restricted to the nucleotides at the 5′ or 3′ end of the sequenced fragments.

MNase-seq data preprocessing and normalization

For all samples, paired-end 42 base reads were trimmed for low-quality bases, random matches, and for a minimum length [trimming parameters: -q 20,20 -O 12 -m 25] with Cutadapt (v 1.9.1).[95] Processed reads were then mapped to the GRCh38.p13 genome assembly using STAR (v 2.6.1b)[90] with maximum 2 percent mismatches [mapping parameters: –outFilterMismatchNoverLmax 0.02 –outFilterMultimapScoreRange 0 –alignIntronMax 1]. Samtools (v 1.3.1)[91] was used to quality filter SAM files, where alignments with MAPQ smaller than 7 (-q 7) were skipped, and only proper pairs (-f 2) were selected. Read counts for expressed major transcript isoforms for each condition j (control or CDK7^{as} inhibited) were calculated using HTSeq.[92] The counts data was used to generate normalization factors with DESeq2[93] to normalize the MNase-seq data.

Detection of +1 and +2 nucleosome positions

BAM files for each condition were merged and converted to BED files with bedtools (v 2.29.1).[97] The BED files were processed using the DANPOS3[98] algorithm (dpos) to determine nucleosome positions for each condition. Further data processing was carried out using the R/Bioconductor environment. For the annotation of the +1 nucleosome, the closest nucleosome that was found downstream of the TSS and within the window TSS + 200 bp for each gene was chosen. Similarly, the succeeding nucleosome downstream of the annotated +1 nucleosome and within the window TSS + 400 bp for each gene was annotated as the +2 nucleosome. The midpoint of the annotated nucleosome positions was determined as the positions of the +1 and +2 nucleosome dyads. The final expressed gene set includes 7,938 major transcript isoforms with annotated +1 and +2 nucleosome positions for further analysis.

Detection of promoter-proximal pause sites

The major transcript isoform set for this analysis includes expressed protein-coding isoforms with a first exon greater than 30 bp. Promoter proximal pause site m* for Pol II was calculated as described[24,43] by calling the maximum peak of mNET-seq signal ρ_i in a defined window m downstream of the TSS (exceeding three times the median signal p_{im}) for the sense strand of the control samples.

$$\rho_i = \max_m p_{im}$$

For major transcript isoforms with +1 nucleosome position downstream of the first exon, m was defined from the TSS + 30 bp to the end of the first exon (excluding the last 5 bases), and for major transcript isoforms with +1 nucleosome within the first exon, m was defined from the TSS + 30 bp to the +1 nucleosome dyad position.

Nucleosome arrest sites

For the analysis of nucleosome arrest sites, the expressed major isoform set with annotated +1 and +2 nucleosome positions was used. Nucleosome arrest sites were calculated with the same principle as described for promoter-proximal pause sites. For +1 nucleosome arrest sites, the defined window m was from 50 bp downstream of the annotated promoter-proximal pause site to the +1 nucleosome dyad position. The +2 nucleosome arrest sites were calculated within a window m from 100 bp upstream of the annotated +2 nucleosome dyad position to the +2 nucleosome dyad position for the sense strand of the control samples.

Pause duration and transcription duration through the +1 and +2 nucleosomes

The pause duration d was calculated as described[24] with minor modifications. The expressed major isoform set with annotated promoter-proximal pause sites and nucleosome arrest sites were used. Pause duration is defined as the time a polymerase needs to pass through a 40 bp 'pause window' located ±20 bp around the pause site or nucleosome arrest site. For each condition j (control or CDK7^{as} inhibited), the pause duration d was calculated in the pause window as

$$d_i = s \cdot \frac{\sum_{\pm 20} P_i}{I_i}$$

where P_i is the mNET-seq coverage values in the pause window, and I_i is the productive initiation frequency. In Figure 5, we report the pause durations in arbitrary units [a.u.] (setting the calibration factor s to 600) and focus on the relative changes in the pause durations rather than the absolute values.

Elongation velocity estimation

The elongation velocity was calculated as described.[103,112] Elongation velocity is defined as the RNA synthesis per Pol II as a function of time. As TT-seq measures the new RNA synthesis profile of Pol II during the labeling time and mNET-seq (or Pol II MNase-ChIP-seq) measures the Pol II occupancy profile, the TT-seq/mNET-seq (or TT-seq/MNase-ChIP-seq) signal ratio can be quantified as a proxy for elongation velocity.[24] For each expressed major transcript isoform i exceeding 10 kbp in length, the elongation velocity was calculated for different windows from the TSS + 200 bp to 8 kbp downstream as

$$v_i = \frac{s}{t \cdot k} \cdot \frac{\sum tb_i}{\sum P_i}$$

where tb_i is the sequencing depth and antisense bias corrected TT-seq coverage and P_i is the sequencing depth normalized mNET-seq (or MNase-ChIP-seq) coverage values in the defined window.

Differential expression analysis

Differential gene expression analysis was performed with the DESeq2 package[93] to observe changes in RNA synthesis (TT-seq) after inhibition. The size factor parameter was set to the spike-in normalization strategy derived sequencing depth. Significant changes were denoted for the padj cutoff set to 0.05.

Replicates/statistics

All treatments were performed in two biological replicates to generate sample NGS libraries, except MNase-ChIP-seq, for which three replicates were collected. The numbers of genes analyzed and statistical tests are indicated in the corresponding figure legends. The error bars and representations are described in the respective figure legends.