nature structural & molecular biology

Article

https://doi.org/10.1038/s41594-025-01707-1

Kinetic control of mammalian transcription elongation

Received: 5 March 2025

Accepted: 3 October 2025

Published online: 27 November 2025

Check for updates

Yukun Wang[1,6], Xizi Chen[2,6], Maximilian Kümmecke[1], John W. Watters[1], Joel E. Cohen[3,4,5], Yanhui Xu[2] & Shixin Liu[1]

Transcription elongation by RNA polymerase II (Pol II) is an integral step in eukaryotic gene expression. The speed of Pol II is controlled by a multitude of elongation factors, but the exact regulatory mechanisms remain incompletely understood, especially for higher eukaryotes. Here we develop a single-molecule platform to visualize the dynamics of individual mammalian transcription elongation complexes (ECs) reconstituted from purified proteins. This platform allows us to follow the elongation and pausing behavior of EC in real time and unambiguously determine the role of each elongation factor in the kinetic control of Pol II. We find that the mammalian EC harbors multiple speed gears dictated by its associated factors and phosphorylation status. Moreover, the elongation factors are not functionally redundant but act hierarchically and synergistically to achieve optimal elongation activity. We propose that such elaborate kinetic regulation underlies the major speed-changing events during the transcription cycle and enables cells to adapt to a changing environment.

Transcription elongation by RNA polymerase II (Pol II) is increasingly recognized to be a focal point of regulation in eukaryotic gene expression[1,2], and its dysregulation has been linked to disease and aging[3,4]. In metazoans, after Pol II assembles the initiation complex at the promoter and transcribes 30–60 nucleotides (nt) of RNA, it enters a paused state known as promoter–proximal pausing[5]. Several elongation factors, including P-TEFb, DSIF and PAF1C, control the release of Pol II into productive elongation[6]. At this stage, Pol II speeds up to several kilobases (kb) per minute, reads through exons and introns, and is kinetically coupled with other cotranscriptional processes such as messenger RNA (mRNA) splicing[7,8]. At the 3′ end of genes, Pol II slows down to facilitate cotranscriptional mRNA cleavage and polyadenylation, events critical for the proper termination of transcription[9].

The activity of the mammalian elongation complex (EC) is regulated by myriad elongation factors and complexes, which decorate a major fraction of the Pol II surface as revealed by structural studies (Fig. 1a)[10,11]. Much of the knowledge regarding the functions of these regulatory factors came from in vivo studies, particularly those using genetic or chemical perturbations[12–14]. In vitro biochemical studies offer a complementary approach to determine whether a factor stimulates or represses EC activity[15–17]. However, the readouts of these assays are either stable genomic positions of EC components or endpoint RNA products from many transcription events. It remains challenging to distinguish direct versus indirect effects on the elongation activity by specific factors from often subtle phenotypes. Filling this knowledge gap requires assays capable of following the elongation kinetics in real time and directly relating factor binding to EC activity.

Single-molecule techniques circumvent the limitations associated with ensemble averaging, and are thus well suited for studying dynamic biomolecular processes[18,19]. Indeed, single-molecule studies have provided a wealth of insights into the mechanism and regulation of transcription elongation in model systems such as *Escherichia coli* and *Saccharomyces cerevisiae*[20,21]. But a similar level of mechanistic understanding has not been achieved for higher eukaryotic systems,

[1]Laboratory of Nanoscale Biophysics and Biochemistry, The Rockefeller University, New York, NY, USA. [2]Fudan University Shanghai Cancer Center, Institutes of Biomedical Sciences, New Cornerstone Science Laboratory, Department of Biochemistry and Biophysics, School of Life Sciences, Shanghai Medical College of Fudan University, Shanghai, China. [3]Laboratory of Populations, The Rockefeller University, New York, NY, USA. [4]Climate School, School of International and Public Affairs, Department of Statistics, Columbia University, New York, NY, USA. [5]Department of Statistics, University of Chicago, Chicago, IL, USA. [6]These authors contributed equally: Yukun Wang, Xizi Chen. e-mail: xuyh@fudan.edu.cn; shixinliu@rockefeller.edu

https://doi.org/10.1038/s41594-025-01707-1

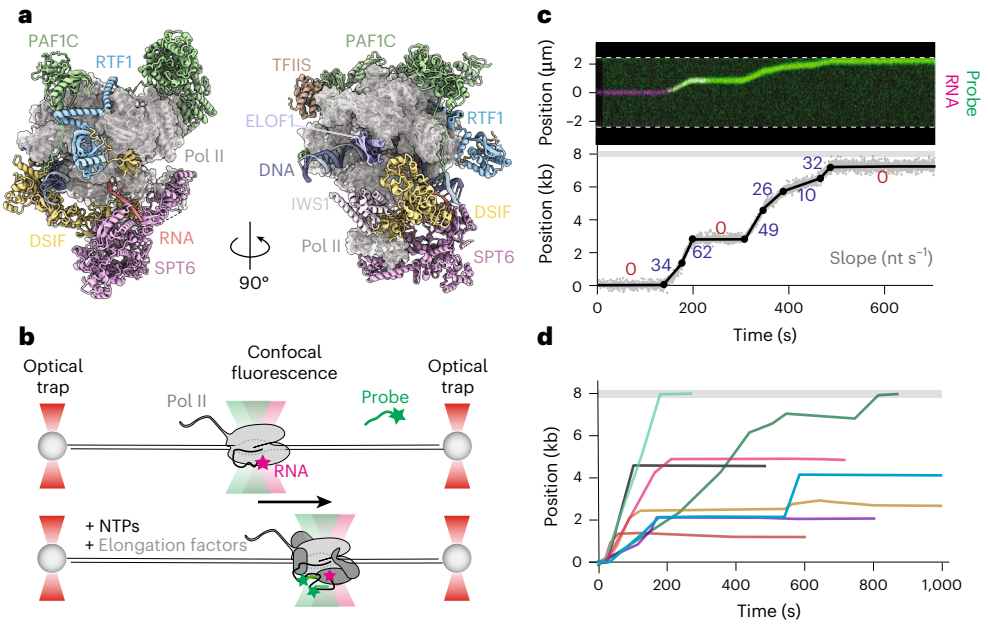


Fig. 1 | Real-time observation of mammalian transcription elongation. **a**, Composite structural model of the human transcription EC with a set of elongation factors, as supplied in our system, decorating the Pol II surface. The model incorporates published structures for the mammalian EC (PDB 6TED), TFIIS (PDB 8A40), ELOF1 (PDB 8B3F) and a fitted AlphaFold prediction of IWS1 based on the position of Spn1 in the yeast EC (PDB 7XN7). **b**, Schematic of the single-molecule experimental setup. The starting location of the EC is marked by the ATTO647N-labeled RNA, while Cy3-labeled DNA probes that hybridize to the nascent RNA indicate active elongation. **c**, An example of transcription elongation by a single mammalian EC observed in real time. Top: kymograph showing the EC position on the DNA template, indicated by RNA (magenta) and probe (green) signals, as a function of time. Bottom: elongation trajectory extracted from the kymograph. Raw data (gray dots) were fitted to discrete linear segments (black line). Change points are marked as filled circles. The slopes for each segment are indicated, differentiating active elongation (blue) from pausing or stalling events (red). **d**, Eight additional examples of fitted elongation trajectories by individual mammalian ECs aligned by their starting positions. NTP, nucleoside triphosphate.

largely due to the increased difficulty in reconstituting active ECs at the single-molecule level. A previous study reported a low basal activity of the mammalian Pol II compared to the yeast counterpart[22], probably due to the omission of certain elongation factors.

In this work, we reconstituted the 26-subunit mammalian EC with a total molecular weight of 1.6 MDa using purified proteins at the single-molecule level. This allowed us to measure the real-time elongation and pausing kinetics of the mammalian transcription machinery. Our results provide insights into the hierarchy of several key elongation factors—P-TEFb, DSIF, PAF1C, SPT6 and RTF1—in the kinetic control of Pol II and offer a quantitative framework for understanding the intricate regulation of mammalian transcription elongation.

Results

Single-molecule observation of mammalian transcription elongation

We purified the mammalian Pol II (Extended Data Fig. 1) and assembled EC using a synthetic nucleic acid scaffold containing an RNA primer with its 5′ end labeled with an ATTO647N fluorophore and a DNA template that was biotinylated at either end (Extended Data Fig. 2). The construct with an EC positioned at the center was tethered between a pair of streptavidin-coated beads and visualized on a C-Trap instrument that combines dual-trap optical tweezers and single-molecule fluorescence microscopy (Fig. 1b)[23]. For enhanced fluorescence detection, the downstream DNA contained seven repeat sequences that, when transcribed into nascent RNA, could each be hybridized to a complementary DNA probe labeled with a Cy3 fluorophore. When supplied with eight purified human elongation factors and/or complexes, namely P-TEFb, DSIF, PAF1C, RTF1, SPT6, IWS1, TFIIS and ELOF1 (Extended Data Fig. 1a), more than half of the ECs exhibited active elongation, indicated by the movement of the RNA fluorescence signal and the appearance and/or movement of the probe fluorescence signal (Fig. 1c and Extended Data Fig. 3). As shown later in this paper where fluorescently labeled elongation factors were used, the RNA signal colocalized with the protein factor signal during elongation, indicating that the 5′ region of the nascent RNA remained in close proximity to the EC even as the transcript grew in length. Thus, in this study we used these RNA-bound probes as a proxy to track Pol II movement.

We found mammalian transcription elongation to be a highly dynamic and heterogeneous process, frequently transitioning between different kinetic states. Such heterogeneity likely arose in part from the dynamic association and dissociation of the elongation factors. To quantitatively characterize the kinetic behavior of the EC, we developed a Bayesian framework to divide individual elongation trajectories into discrete linear segments. We defined the segments with a positive slope of ≥1 nt s^{-1} as active elongation, and those with a slope of <1 nt s^{-1} or negative values as pausing events. The observed transient pauses may result from Pol II arrest or backtracking, which can be rescued by the elongation factor TFIIS that induces RNA cleavage[24,25]. Indeed, when TFIIS was omitted from the single-molecule assay, we saw a significant increase in the pause duration (Extended Data Fig. 4a). Moreover, we found that most EC trajectories ended with a long-lasting stall that featured extensive backtracking (Fig. 1d and Extended Data Fig. 4b). Unlike the internal pauses, these stalling events cannot be rescued by TFIIS (Extended Data Fig. 4c). We also examined the potential effect of RNase H on these pausing or stalling events and did not observe a noticeable change in their occurrence, indicating that DNA–RNA hybrid structures such as R loops were not the main cause of these events. We surmise that these events may correspond to the 'persistent backtracking' phenomenon found in mammalian cells[26]. It remains to be determined which additional factor(s) are required to resolve the long-lasting stall for Pol II to transcribe long genes in vivo.

https://doi.org/10.1038/s41594-025-01707-1

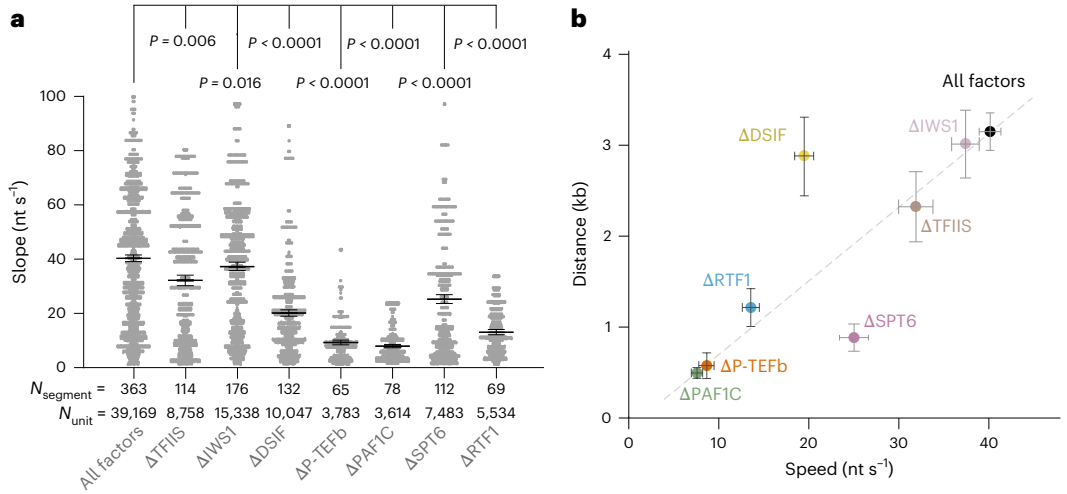


Fig. 2 | Differential effects of elongation factors on EC kinetics. a, Dot plot showing the slope of active elongation segments from EC trajectories under conditions where individual elongation factors were omitted. Segments are weighted by their transcribed length (that is, each dot corresponds to a 10-bp unit; the total number of units for each condition is listed as N_{unit}). Bars represent mean ± s.e.m. (sample size is the number of transcription segments, $N_{segment}$). *P* values were calculated from unpaired two-tailed *t*-tests with Welch's correction compared with the All-factors condition. **b**, The average distance traveled before stalling by individual mammalian ECs at a given condition (normalized by the active EC fraction) is plotted against the average pause-free elongation speed weighted by the transcribed length of each segment. Error bars represent s.e.m. (sample size for speed is $N_{segment}$; sample size for distance is the number of ECs). The dashed line represents the linear regression of the speed versus distance values for all conditions.

Elongation factors exert differential effects on EC activity

Next, we focused on the active elongation segments, whose slopes vary over two orders of magnitude (~1–100 nt s^{-1}) (Fig. 2a). This range is in general agreement with previous in vivo measured values[2,8], notwithstanding differences in the reaction conditions. When all the elongation factors were supplied at concentrations above their respective dissociation constants, the average pause-free speed on the 8-kb template, weighted by the transcribed length of each active segment, was 40 ± 1 nt s^{-1} and the average distance an EC traveled before stalling was 3.1 ± 0.2 kb (mean ± standard error of the mean (s.e.m.); Fig. 2b). Similar values were obtained when we used a longer (16-kb) DNA template with a different guanine-cytosine profile (Extended Data Fig. 3c,d), indicating these parameters are not dictated by the template but rather intrinsic properties of the EC. We then recorded single-molecule trajectories using the 8-kb template under conditions where individual elongation factors were omitted, which allowed us to determine their respective contribution to the elongation kinetics (Fig. 2a and Extended Data Fig. 5). These single-omission experiments yielded a spectrum of phenotypes. Some conditions (such as ΔP-TEFb and ΔPAF1C) exhibited a drastic reduction in speed and distance, while others (such as ΔIWS1 and ΔTFIIS) showed only a mild effect. When plotting the average speed against the average distance for each condition, we found that most of the conditions followed a linear relationship (Fig. 2b). This indicates that the total time an EC spends on productive elongation (distance over speed) remained largely constant across different conditions. There were two notable exceptions: ΔDSIF resulted in significantly reduced speed with a slight decrease in distance, whereas ΔSPT6 drastically shortened the distance but only moderately affected the pause-free speed. In the following, we will further investigate those factors whose omission caused a strong phenotype in the elongation kinetics of EC on DNA, namely P-TEFb, DSIF, PAF1C, SPT6 and RTF1.

P-TEFb activates elongation by phosphorylating Pol II and DSIF

P-TEFb is a positive elongation factor that contains a cyclin-dependent kinase CDK9, which phosphorylates the C-terminal domain (CTD) of Pol II's RPB1 subunit, as well as DSIF, PAF1C and SPT6[10,16,27,28]. Among these phosphorylated substrates, which one(s) directly affect the elongation speed remains unclear. In our single-molecule assay, omitting P-TEFb reduced the elongation speed and distance by roughly fivefold (compare ΔP-TEFb to All-factors in Fig. 2b). To separate the effects of multiple CDK9-mediated phosphorylation products on the elongation kinetics, we specifically prevented the phosphorylation of either Pol II or DSIF. To prevent Pol II phosphorylation (that is, the Pol II^{P-} condition), we omitted P-TEFb during the EC assembly step (Extended Data Fig. 2b), mixed P-TEFb with all the other elongation factors during a prephosphorylation step, and included the CDK9 inhibitor flavopiridol in the single-molecule assay to prevent in situ phosphorylation of Pol II by P-TEFb. To prevent DSIF phosphorylation (that is, the $DSIF^{P-}$ condition), we excluded DSIF from the prephosphorylation step and then introduced DSIF and flavopiridol at the same time to the reaction mixture. We found that both Pol II^{P-} and $DSIF^{P-}$ conditions exhibited a severe elongation defect (Fig. 3a–d and Extended Data Fig. 6). Next, we performed experiments in which only Pol II and DSIF—but not any other factor—were allowed to be phosphorylated. To do this, Pol II was phosphorylated in the presence of P-TEFb during the EC assembly step, DSIF was incubated alone with P-TEFb during the prephosphorylation step and flavopiridol was added to the reaction mixture together with all the other elongation factors. We found that this condition (Pol II^{P+}&$DSIF^{P+}$) phenocopied the condition where All-factors and Pol II were allowed to be phosphorylated by P-TEFb (Fig. 3a,b,e and Extended Data Fig. 6d). Together, these results indicate that the phosphorylation of Pol II and DSIF is both necessary and sufficient for optimal EC activity.

DSIF regulates both elongation and pausing kinetics of EC

DSIF exerts pleiotropic effects on transcription in vivo[13,29–31]. In vitro biochemical assays also suggest that DSIF can both negatively and positively regulate elongation[32,33]. To visualize the behavior of DSIF during elongation, we fluorescently labeled its SPT4 subunit and used two-color imaging to simultaneously monitor DSIF binding and EC translocation. DSIF and EC were frequently observed to colocalize and comigrate on the DNA template (Fig. 4a and Extended Data Fig. 7a), suggesting stable association of DSIF with Pol II during elongation. We also observed non-EC-bound DSIF on DNA, often exhibiting a diffusive behavior (Fig. 4a). We then used the CDK9 inhibitor flavopiridol to inhibit DSIF phosphorylation by P-TEFb and imaged its interaction with EC (Fig. 4b and Extended Data Fig. 7b). We observed that unphosphorylated DSIF still bound EC, but with a reduced residence time compared to the phosphorylated version (47 ± 4 s for $DSIF^{P-}$ versus 182 ± 14 s for

https://doi.org/10.1038/s41594-025-01707-1

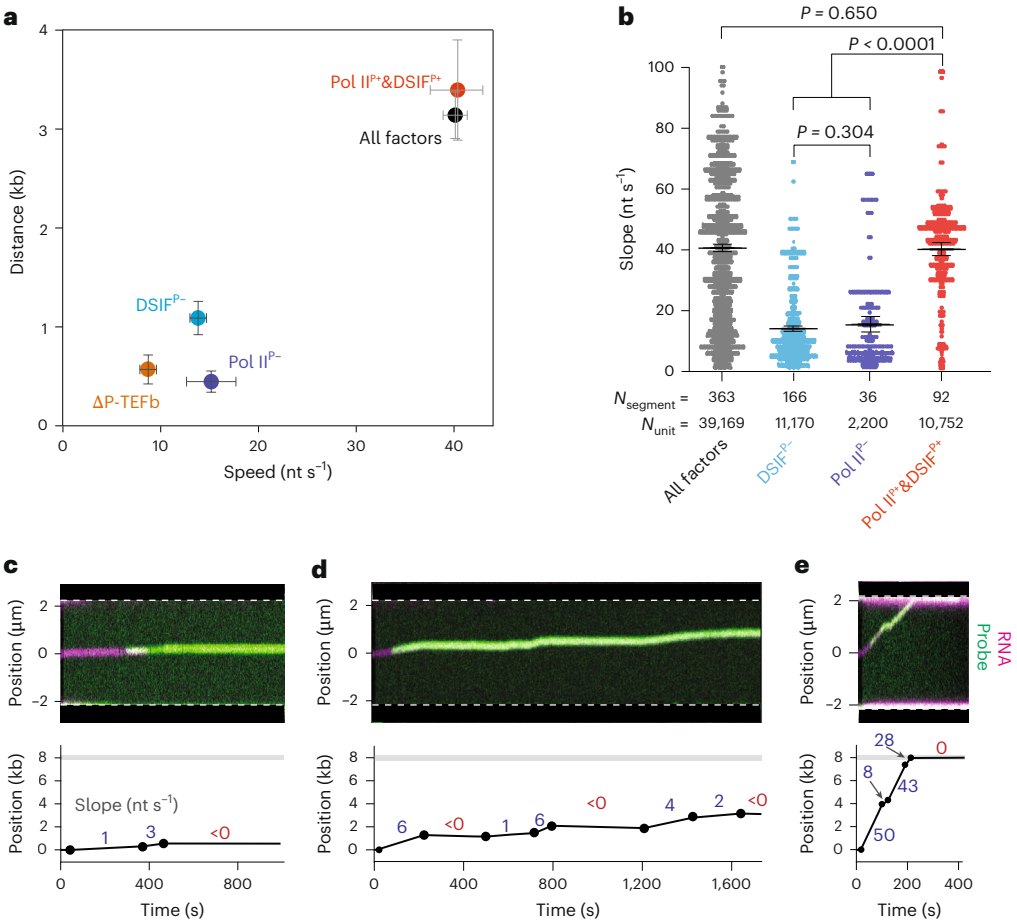


Fig. 3 | Effects of P-TEFb-mediated phosphorylation on EC activity. **a**, Two-dimensional plot showing the average elongation speed and distance for different phosphorylation conditions. All-factors: the full set of factors present, all of which were allowed to be phosphorylated by P-TEFb; ΔP-TEFb: P-TEFb omitted, hence no phosphorylation of any protein; DSIF^{P-}: All-factors present, but phosphorylation of DSIF was specifically inhibited; Pol II^{P-}: All-factors present, but Pol II was not phosphorylated; Pol II^{P+}&DSIF^{P+}: All-factors present, but only Pol II and DSIF were allowed to be phosphorylated. Error bars represent s.e.m. (sample size for speed is $N_{segment}$; sample size for distance is the number of ECs). **b**, Dot plot showing the slope of active elongation segments from EC trajectories under the All-factors, DSIF^{P-}, Pol II^{P-} and Pol II^{P+}&DSIF^{P+} conditions. Segments are weighted by their transcribed length and each dot corresponds to a 10-bp unit. Bars represent mean ± s.e.m. (sample size is $N_{segment}$). *P* values were from unpaired two-tailed *t*-tests with Welch's correction. **c**, An example of an EC trajectory for the Pol II^{P-} condition. Top: kymograph showing the overlay of ATTO647N-labeled RNA (magenta) and Cy3-labeled complementary DNA probe (green) signals that were used to track EC progression. Bottom: fitted and segmented trajectory with the slope of each segment indicated. **d**, An example of an EC trajectory for the DSIF^{P-} condition. Details are the same as **c**. **e**, An example of an EC trajectory for the Pol II^{P+}&DSIF^{P+} condition. Details are the same as **c**. To help with visualization, we flipped the kymograph vertically, when necessary, such that the EC always moves upward.

DSIF^{P+}; mean ± s.e.m.) (Fig. 4c). These results reveal that unphosphorylated DSIF can associate with the EC, but phosphorylation further stabilizes DSIF binding and is required for fast elongation. Notably, a single-molecule fluorescence study on the yeast system reported that Spt4/5 binds EC with an average residence time of 56 seconds, similar to our measurements. However, the yeast Spt4/5 was found only on ECs but not on bare DNA[34].

We found that the omission of DSIF reduced the EC speed by roughly twofold but minimally affected the distance it can travel (compare ΔDSIF to All-factors in Fig. 2b). The deviation from the linear speed–distance relationship indicates that the fraction of time an EC spends on productive elongation is affected by DSIF. We thus analyzed the pausing kinetics under different DSIF conditions (ΔDSIF, DSIF^{P+} and DSIF^{P-}) and indeed found that ECs spent significantly less time in a paused state when DSIF was omitted (Fig. 4d–f and Extended Data Fig. 7c). Specifically, both DSIF^{P+} and DSIF^{P-} increased the pause duration compared to the ΔDSIF condition, with DSIF^{P-} displaying a more pronounced effect (Fig. 4e). This is possibly related to DSIF's DNA and RNA clamping activity[35,36]. Moreover, DSIF^{P-}, but not DSIF^{P+}, exhibited an elevated pause density, a parameter that quantifies the propensity of EC entering a pause (Fig. 4f). These findings reveal that besides its positive effect on the elongation speed, DSIF^{P+} exerts a pause-stabilizing effect on the EC; whereas DSIF^{P-} also stabilizes the paused state and additionally induces pause entry without stimulating elongation, making it a negative elongation factor.

PAF1C binding directly accelerates EC

PAF1C has been shown to promote active elongation in vitro[10,15] and in vivo[14,37–39]. We found that omitting PAF1C severely reduced elongation speed and distance in our single-molecule assay (Fig. 2b and Extended Data Fig. 8a). To directly visualize the behavior of PAF1C, we attached a Cy3 fluorophore to its CTR9 subunit and monitored its interaction with EC during elongation (Fig. 5a and Extended Data Fig. 8b). We observed that in most (71%) of active trajectories, the onset of PAF1C signal coincided with a sudden acceleration of the EC (that is $\Delta t = t_{move} - t_{bind} = 0$), indicating an immediate stimulatory effect of PAF1C association on elongation. In comparison, DSIF association and EC activation co-occurred in only 29% of the trajectories. We also found that the PAF1C fluorescence signal continued to increase stepwise as elongation progressed, and this phenomenon was observed in every

https://doi.org/10.1038/s41594-025-01707-1

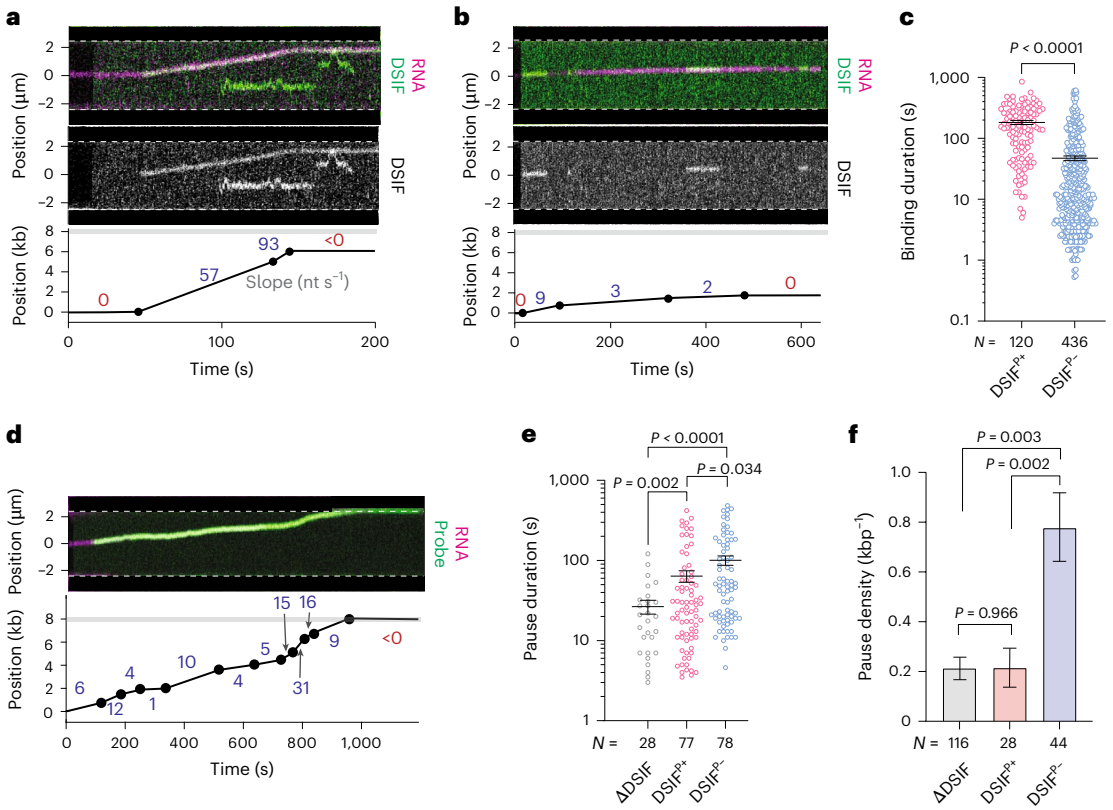


Fig. 4 | DSIF plays a dual role in EC regulation. a, An example of an EC trajectory with Cy3-labeled phosphorylated DSIF ($DSIF^{P+}$) and ATTO647N-labeled RNA. Top: kymograph showing the overlay of DSIF (green) and RNA (magenta) signals. Middle: the same kymograph but with only DSIF signals shown to more clearly visualize its binding to the EC and DNA. Bottom: fitted and segmented elongation trajectory with the slope of each segment indicated (active elongation in blue and pausing or stalling in red). **b**, An example of an EC trajectory with Cy3-labeled unphosphorylated DSIF ($DSIF^{P-}$). Details are the same as **a**. **c**, Dot plot showing the residence time of $DSIF^{P+}$ (red) and $DSIF^{P-}$ (blue) on EC. The labeled DSIF concentration used in these experiments was 12.5 nM. Bars represent mean ± s.e.m. (sample size is the number of binding events). *P* value was from unpaired two-tailed *t*-tests with Welch's correction. **d**, An example of an EC trajectory for the ΔDSIF condition. Top: kymograph showing the overlay of RNA (magenta) and DNA probe (green) signals that were used to track EC progression. Bottom: fitted and segmented trajectory with the slope of each segment indicated. **e**, Dot plot showing the duration of pauses within the elongation trajectories under the condition where DSIF was omitted (ΔDSIF) or included as either a phosphorylated ($DSIF^{P+}$) or an unphosphorylated ($DSIF^{P-}$) form. Bars represent mean ± s.e.m. (sample size is the number of pauses). *P* values were from unpaired two-tailed *t*-tests with Welch's correction. **f**, Bar plot showing the pause density, defined as the number of pauses per kb transcribed, measured for the same conditions as **e**. *N* denotes the number of EC trajectories. Error bars represent the 95% confidence interval using bootstrapping. *P* values were calculated from unpaired two-tailed *t*-tests with Welch's correction. Unlabeled $DSIF^{P+}$ and $DSIF^{P-}$ were used in the pause measurements at a concentration of 200 nM.

trajectory examined (Extended Data Fig. 8b–e). A previous study in the yeast system reported that Paf1C can directly bind RNA[40], so we speculated that additional copies of PAF1C may be recruited to the growing nascent RNA chain. Indeed, RNase A treatment led to a constant level of PAF1C signals throughout the trajectories, suggesting the association of one single copy of PAF1C with the EC without RNA (Extended Data Fig. 8d,e). This treatment allowed us to measure the lifetime of PAF1C–Pol II interaction to be 80 ± 10 seconds (mean ± s.e.m.). Our in vitro data thus show that the residence time of PAF1C on the EC is shorter than that of phosphorylated DSIF, consistent with recent live-cell imaging results[41].

Given that the ΔPAF1C and ΔP-TEFb conditions essentially phenocopied each other (Fig. 2b), we asked whether P-TEFb is responsible for PAF1C recruitment to EC. To this end, we monitored PAF1C binding and EC translocation in the absence of P-TEFb, which prevented all phosphorylation events including those on Pol II CTD. We found that the appearance of the PAF1C signal no longer coincided with EC movement (that is, $\Delta t < 0$; Fig. 5b,c and Extended Data Fig. 8f). With RNase A treatment, PAF1C binding was completely abolished (Fig. 5d), suggesting that the PAF1C binding events under the ΔP-TEFb condition resulted purely from RNA-mediated recruitment. Collectively, these results reveal that PAF1C binding to Pol II—but not its binding to RNA—causally accelerates the EC and this binding critically relies on P-TEFb, presumably via the interaction between CDK9-mediated phosphorylated CTD and the CDC73 subunit of PAF1C as suggested by previous biochemical results from the yeast system[42]. Consistent with this interpretation, no PAF1C binding events were observed when flavopiridol was used to inhibit Pol II phosphorylation by P-TEFb (that is, $Pol\ II^{P-}$ condition), ruling out a catalytic-independent function of P-TEFb in PAF1C recruitment.

SPT6 stabilizes PAF1C binding to EC

SPT6 has been shown to positively affect the elongation rate, but whether this effect is direct or not remained unclear[12,14,43–45]. In our single-molecule assay, omitting SPT6 caused a moderately reduced elongation speed and a significantly shorter distance (ΔSPT6 in Fig. 2b). Moreover, we observed that the EC trajectories under this condition frequently exhibited short elongation segments with relatively fast speeds interspersed with pauses (Fig. 5e and Extended Data Fig. 9a). This phenotype is distinct from the ΔPAF1C phenotype where the elongation segments were very slow (Fig. 5f and Extended Data Fig. 8a). Given previous in vivo depletion experiments suggesting that SPT6 helps recruit PAF1C to the EC[12], we examined the impact of SPT6 on PAF1C–Pol II interaction. We used fluorescently

https://doi.org/10.1038/s41594-025-01707-1

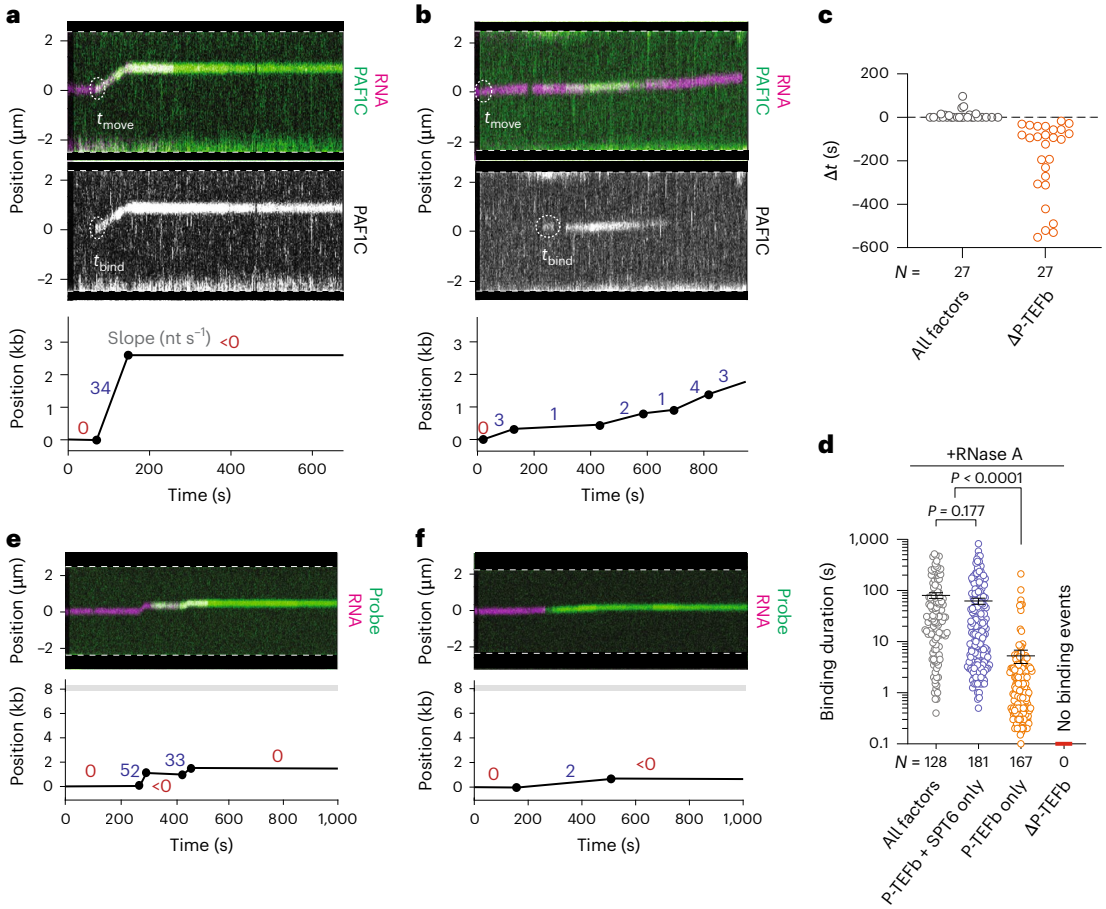


Fig. 5 | PAF1C binding directly activates EC and is stabilized by SPT6. **a**, An example of an EC trajectory with Cy3-labeled PAF1C and ATTO647N-labeled RNA when the full set of elongation factors were present (All-factors). Top: kymograph showing the overlay of PAF1C (green) and RNA (magenta) signals. t_{move} indicates the time point when the EC started translocation. Middle: the same kymograph but with only PAF1C signals shown. t_{bind} indicates the time point when PAF1C binding to EC occurred. Bottom: fitted and segmented trajectory with the slope of each segment indicated (active elongation in blue and pausing or stalling in red). **b**, An example of an EC trajectory with Cy3-PAF1C and ATTO647N-RNA when P-TEFb was omitted (ΔP-TEFb). Details are the same as **a**. **c**, Dot plot showing the Δt value (Δt = t_{move} − t_{bind}) for individual EC trajectories from All-factors (black) and ΔP-TEFb (orange) conditions. N denotes the number of EC trajectories analyzed for each condition. **d**, Dot plot showing the residence time of PAF1C on EC under different conditions. RNase A was included in these experiments to eliminate RNA-mediated PAF1C recruitment. Bars represent mean ± s.e.m. (sample size is the number of binding events). P values were from unpaired two-tailed t-tests with Welch's correction. **e**, An example of an EC trajectory under the ΔSPT6 condition. Top: kymograph showing the overlay of RNA (magenta) and DNA probe (green) signals that were used to track EC progression. Bottom: fitted and segmented trajectory with the slope of each segment indicated. **f**, An example of an EC trajectory under the ΔPAF1C condition. Details are the same as **e**.

labeled PAF1C to visualize its binding in the presence of RNase A to eliminate RNA-mediated recruitment. When only P-TEFb (which phosphorylates Pol II CTD) and SPT6 were included, the residence time of PAF1C on the EC is statistically indistinguishable from the value when All-factors were present (Fig. 5d), suggesting that P-TEFb and SPT6 by themselves are sufficient for mediating stable PAF1C recruitment. Further removing SPT6 caused a tenfold reduction in the PAF1C residence time (Fig. 5d), suggesting that SPT6 is required for stable PAF1C binding to the EC. Indeed, previous studies on the yeast EC reported a direct interaction between the SH2 domain of Spt6 and the Cdc73 subunit of Paf1C[46,47]. We used AlphaFold 3[48] to predict the interaction between human CDC73 and a CTD-mimicking peptide and then placed it in the context of the mammalian EC (Extended Data Fig. 9b). The structural model suggests that Pol II RPB1 CTD and SPT6 can simultaneously engage with PAF1C, corroborating our single-molecule data.

Next, we used fluorescently labeled SPT6 to visualize its behavior directly during elongation. We found that SPT6 stably associated with the EC during elongation (an average residence time of 412 ± 64 s; mean ± s.e.m.) in the presence of All-factors (Extended Data Fig. 9c). The residence time of SPT6 on EC was not significantly affected by DSIF (Extended Data Fig. 9d), consistent with our finding that PAF1C recruitment does not rely on DSIF (Fig. 5d). In contrast, SPT6 binding to EC was critically dependent on P-TEFb (Extended Data Fig. 9d), which could be explained by previous work showing that the RPB1 CTD linker region—a phosphorylation substrate of P-TEFb—interacts with the SH2 domain of SPT6[10]. We also found that SPT6 binding invariably preceded EC movement (Δt > 0; N = 12). Overall, these results indicate that SPT6 itself does not directly activate EC but serves to assist with PAF1C recruitment.

RTF1's stimulatory effect on EC activity relies on PAF1C

RTF1 is an essential elongation factor[11,14,17] that makes extensive contacts with PAF1C and DSIF as shown by the active EC structure[11] (Fig. 1a). Omitting RTF1 from our single-molecule experiments caused a significantly lower EC activity (Fig. 6a and Extended Data Fig. 10a), even though the phenotype was less severe than ΔPAF1C (Fig. 2b). To further understand RTF1's stimulatory function, we examined whether its effect on EC relies on the presence of PAF1C or DSIF, for which no consensus has been reached yet[11,17]. To this end, we performed

https://doi.org/10.1038/s41594-025-01707-1

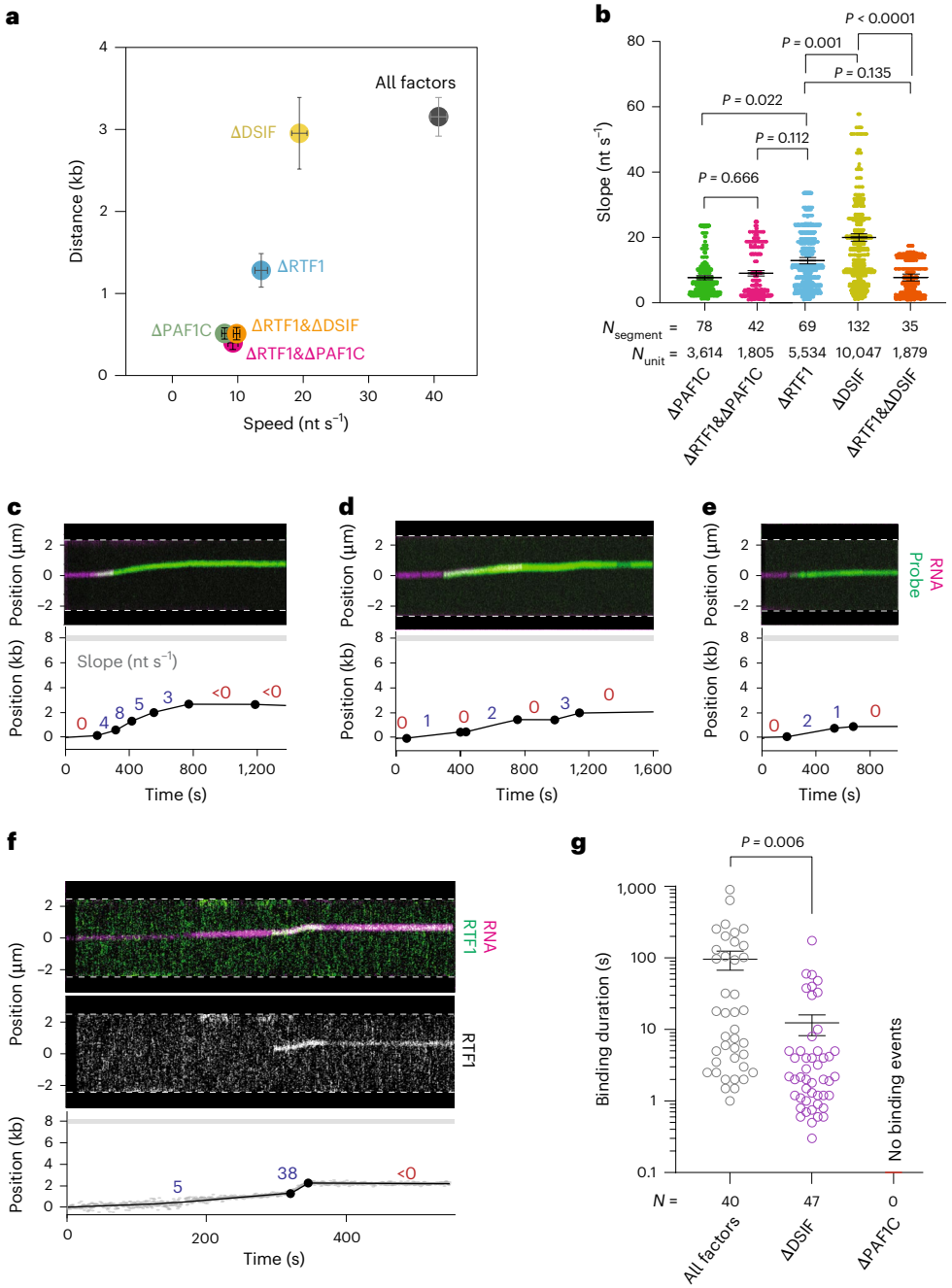


Fig. 6 | RTF1's stimulatory effect on EC differentially relies on PAF1C and DSIF. a, Two-dimensional plot showing the average speed and distance for ECs under different conditions (All-factors; single omissions: ΔRTF1, ΔPAF1C, ΔDSIF; double omissions: ΔRTF1&ΔPAF1C, ΔRTF1&ΔDSIF). Error bars represent s.e.m. (sample size for speed is $N_{segment}$; sample size for distance is the number of ECs). **b**, Dot plot showing the slope of active elongation segments from EC trajectories under ΔPAF1C, ΔRTF1&ΔPAF1C, ΔRTF1, ΔDSIF and ΔRTF1&ΔDSIF conditions. Segments are weighted by their transcribed length and each dot corresponds to a 10-bp unit. Bars represent mean ± s.e.m. (sample size is $N_{segment}$). *P* values were from unpaired two-tailed *t*-tests with Welch's correction. **c**, An example of an EC trajectory under the ΔRTF1 condition. Top: kymograph showing the overlay of RNA (magenta) and DNA probe (green) signals that were used to track EC progression. Bottom: fitted and segmented trajectory with the slope of each segment indicated. **d**, An example of an EC trajectory under the ΔRTF1&ΔPAF1C condition. Details are the same as **c**. **e**, An example of an EC trajectory under the ΔRTF1&ΔDSIF condition. Details are the same as **c**. **f**, An example of an EC trajectory with Cy3-labeled RTF1 and ATTO647N-labeled RNA when the full set of elongation factors were present (All-factors). Top: kymograph showing the overlay of RTF1 (green) and RNA (magenta) signals. Middle: the same kymograph but with only RTF1 signals shown. Bottom: fitted and segmented trajectory with the slope of each segment indicated. **g**, Dot plot showing the residence time of RTF1 on EC under the All-factors, ΔDSIF and ΔPAF1C conditions. Bars represent mean ± s.e.m. (sample size is the number of binding events). *P* values were from unpaired two-tailed *t*-tests with Welch's correction.

combinatorial omission experiments. First, we omitted both RTF1 and PAF1C and found that this condition (ΔRTF1&ΔPAF1C) yielded a phenotype more severe than ΔRTF1 but similar to ΔPAF1C (Fig. 6a–d and Extended Data Fig. 10b). The elongation speed for ΔRTF1&ΔPAF1C was significantly lower than that for ΔRTF1 but indistinguishable from that for ΔPAF1C (Fig. 6b). Next, we omitted both DSIF and RTF1 and found that this condition (ΔRTF1&ΔDSIF) caused a lower elongation speed and a shorter transcription distance compared to either the ΔRTF1 or ΔDSIF condition (Fig. 6a,e and Extended Data Fig. 10c). These results indicate that PAF1C's binding to the EC is a prerequisite for RTF1's

https://doi.org/10.1038/s41594-025-01707-1

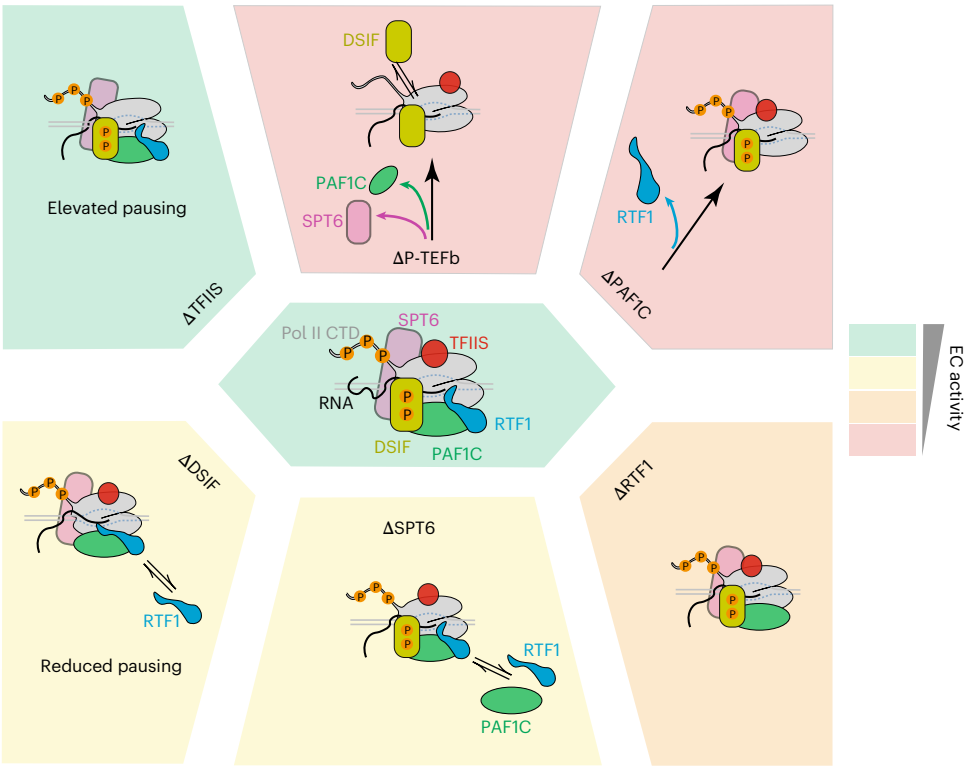


Fig. 7 | Summary of the distinct roles that each elongation factor serves in mammalian Pol II transcription on DNA. The relative EC activity for each factor omission condition is color coded as indicated on the right. Clockwise from the top left: ΔTFIIS increases pausing but has a minimal effect on the elongation speed; ΔP-TEFb abrogates the phosphorylation of Pol II CTD and DSIF, which in turn destabilizes the binding of PAF1C, SPT6 and RTF1, thereby severely reducing the elongation speed; ΔPAF1C exerts the most direct negative impact on EC speed and also completely abolishes RTF1 recruitment; ΔRTF1 shifts the EC from high speed to medium speed; ΔSPT6 reduces EC speed by destabilizing PAF1C binding even though SPT6 itself does not directly activate EC; ΔDSIF exerts a dual effect on EC kinetics by reducing its pausing and also lowering its speed partially via the destabilization of RTF1 binding.

elongation-stimulating effect, and that RTF1 and DSIF can each exert an effect on the EC without the other's presence.

Next, we attached a Cy3 fluorophore to the C terminus of RTF1 to visualize its behavior during elongation. We observed that the association of RTF1 either coincided with the onset of EC movement ($\Delta t = 0$; 6 out of 11 events) or occurred on an already translocating EC, transitioning it to a high-speed mode ($\Delta t < 0$; 5 out of 11 events) (Fig. 6f and Extended Data Fig. 10d,e). Omitting PAF1C completely abolished RTF1 binding events (Fig. 6g), confirming that PAF1C's recruitment to EC is required for RTF1 binding. To reinforce this conclusion, we simultaneously monitored PAF1C and RTF1 using differentially labeled factors and found that RTF1 and PAF1C binding occurred either concomitantly (11 out of 17 events) or sequentially with the PAF1C signal always appearing first (6 out of 17 events). On the other hand, RTF1 was still observed to bind EC in the absence of DSIF, even though the average residence time was significantly reduced (Fig. 6g). Together, these results reveal that PAF1C is required for RTF1's function in elongation but RTF1 can enhance EC activity without DSIF.

Discussion

The work presented here characterized the basic kinetic properties of mammalian transcription elongation in a fully reconstituted system. Our results show that the mammalian EC is a multi-gear supramolecular machine, displaying a remarkable level of heterogeneity and dynamic range in its speed. This feature likely reflects its need to drastically shift gears during the transcription cycle: slow during the early stage before the assembly of a complete EC, fast after the escape from promoter–proximal pausing, and then slow again when reaching the 3′ end of genes[2,6,8]. We further delineated how the kinetics of Pol II is controlled by the binding and dissociation of specific factors (Fig. 7). Although the factors studied here have all been categorically considered as positive elongation factors, our experiments revealed that they each regulate the elongation speed and distance in a unique and hierarchical manner. The fact that the absence of any single factor (P-TEFb, DSIF, PAF1C, SPT6, RTF1) results in a significantly reduced speed suggests that these factors are not functionally redundant but instead collaborate to meet the demands of the complex cellular environment in multicellular organisms.

P-TEFb is considered a master regulator of transcription elongation by phosphorylating multiple components of the EC[27]. By manipulating the phosphorylation status of specific P-TEFb substrates, we showed that its regulation of the elongation speed is mainly manifested in the phosphorylation of Pol II and DSIF. On the other hand, phosphorylation of the other CDK9 substrates, including PAF1C and SPT6, does not substantially impact the EC speed. The phosphorylation of Pol II CTD is a prerequisite for PAF1C recruitment, which we found to be one of the most critical events for EC acceleration. The phosphorylation of DSIF stabilizes its own association with EC and possibly augments RTF1's stimulatory function[11]. As such, dephosphorylation of DSIF and Pol II CTD by protein phosphatases (for example, PP1 and PP2A)[49] may induce the dissociation of speed-stimulating factors and thus the deceleration of EC, potentially facilitating mRNA processing and termination.

DSIF, specifically its SPT5 subunit, is the only universally conserved transcriptional regulator across domains of life[31]. We showed that DSIF enhances the elongation speed of the EC and also promotes its pausing, providing a mechanistic basis for its pleiotropic role in both stimulating and repressing the transcriptional activity[13,29,33,36,50]. We also found that the unphosphorylated DSIF loses the speed-enhancing function but bears a greater pause-promoting activity. As such, the exact regulatory

function of DSIF can be tuned by its phosphorylation status, making it a versatile elongation factor. Mammalian SPT5 contains multiple KOW domains and a C-terminal repeat (CTR) region that harbor multiple CDK9 phosphorylation sites[3]. Among these, phosphorylation at Ser666 within the KOWx-4/5 linker facilitates pause release[13,50], whereas the tandem threonine-rich repeats within CTR are important for the stimulation of elongation speed[51,52]. In the future, using phospho-mimic or phospho-deficient SPT5 mutants in the single-molecule assays will further clarify the roles of different phosphorylation sites in the kinetic control of Pol II.

Among all the factors examined in this work, PAF1C exerts the most direct impact on Pol II speed: its binding immediately accelerates the EC. This type of insight highlights the power of real-time single-molecule visualization in unveiling causal effects. Moreover, our data show that RTF1 provides an additional boost to the elongation speed and that it can bind Pol II without DSIF; nevertheless, DSIF significantly stabilizes RTF1's association with EC. In the EC structure that contains both RTF1 and DSIF, the latch domain of RTF1 and the bridge helix of Pol II make direct contacts that are important for robust EC activity[11]. We postulate that these contacts are allosterically enabled by RTF1–DSIF interaction, which could explain the observed synergy between these two factors. Rtf1 is a constitutive component of Paf1C in yeast but becomes a separate entity in metazoans[53]. Previous results from the *S. cerevisiae* system showed that Spt4/5 is essential for Paf1C recruitment via Rtf1[54]. In contrast, our findings from the mammalian system show that the recruitment of PAF1C to EC only relies on P-TEFb (via the phosphorylation of Pol II CTD) and SPT6 but not on RTF1 or DSIF. Further comparative studies on different eukaryotic systems using the single-molecule reconstitution approach as demonstrated here are thus warranted. Moreover, Pol II transcription is subjected to regulation by the chromatin landscape in vivo[55]. Incorporating nucleosomes in the single-molecule platform and interrogating how they dynamically interact with the EC and its regulatory factors is an important future direction.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41594-025-01707-1.

References

1. Cramer, P. Eukaryotic transcription turns 50. *Cell* **179**, 808–812 (2019).
2. Jonkers, I. & Lis, J. T. Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16**, 167–177 (2015).
3. Aoi, Y. & Shilatifard, A. Transcriptional elongation control in developmental gene expression, aging, and disease. *Mol. Cell* **83**, 3972–3999 (2023).
4. Debes, C. et al. Ageing-associated changes in transcriptional elongation influence longevity. *Nature* **616**, 814–821 (2023).
5. Core, L. & Adelman, K. Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev.* **33**, 960–982 (2019).
6. Chen, F. X., Smith, E. R. & Shilatifard, A. Born to run: control of transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **19**, 464–478 (2018).
7. Saldi, T., Cortazar, M. A., Sheridan, R. M. & Bentley, D. L. Coupling of RNA polymerase II transcription elongation with pre-mRNA splicing. *J. Mol. Biol.* **428**, 2623–2635 (2016).
8. Muniz, L., Nicolas, E. & Trouche, D. RNA polymerase II speed: a key player in controlling and adapting transcriptome composition. *Embo J.* **40**, e105740 (2021).
9. Lopez Martinez, D. & Svejstrup, J. Q. Mechanisms of RNA polymerase II termination at the 3′-end of genes. *J. Mol. Biol.* **437**, 168735 (2025).
10. Vos, S. M. et al. Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. *Nature* **560**, 607–612 (2018).
11. Vos, S. M., Farnung, L., Linden, A., Urlaub, H. & Cramer, P. Structure of complete Pol II-DSIF-PAF-SPT6 transcription complex reveals RTF1 allosteric activation. *Nat. Struct. Mol. Biol.* **27**, 668–677 (2020).
12. Aoi, Y. et al. SPT6 functions in transcriptional pause/release via PAF1C recruitment. *Mol. Cell* **82**, 3412–3423 (2022).
13. Fong, N., Sheridan, R. M., Ramachandran, S. & Bentley, D. L. The pausing zone and control of RNA polymerase II elongation by Spt5: implications for the pause-release model. *Mol. Cell* **82**, 3632–3645 (2022).
14. Zumer, K. et al. Two distinct mechanisms of RNA polymerase II elongation stimulation in vivo. *Mol. Cell* **81**, 3096–3109 (2021).
15. Kim, J., Guermah, M. & Roeder, R. G. The human PAF1 complex acts in chromatin transcription elongation both independently and cooperatively with SII/TFIIS. *Cell* **140**, 491–503 (2010).
16. Yamada, T. et al. P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Mol. Cell* **21**, 227–237 (2006).
17. Cao, Q. F. et al. Correction for Cao et al., Characterization of the human transcription elongation factor Rtf1: evidence for nonoverlapping functions of Rtf1 and the Paf1 complex. *Mol. Cell Biol.* **35**, 4093 (2015).
18. Ha, T., Kaiser, C., Myong, S., Wu, B. & Xiao, J. Next generation single-molecule techniques. Imaging, labeling, and manipulation in vitro and in cellulo. *Mol. Cell* **82**, 304–314 (2022).
19. Chua, G. N. L. & Liu, S. When force met fluorescence: single-molecule manipulation and visualization of protein–DNA interactions. *Annu. Rev. Biophys.* **53**, 169–191 (2024).
20. Dangkulwanich, M., Ishibashi, T., Bintu, L. & Bustamante, C. Molecular mechanisms of transcription through single-molecule experiments. *Chem. Rev.* **114**, 3203–3223 (2014).
21. Lee, C. Y. & Myong, S. Probing steps in DNA transcription using single-molecule methods. *J. Biol. Chem.* **297**, 101086 (2021).
22. Palangat, M. et al. Efficient reconstitution of transcription elongation complexes for single-molecule studies of eukaryotic RNA polymerase II. *Transcription* **3**, 146–153 (2012).
23. Wasserman, M. R., Schauer, G. D., O'Donnell, M. E. & Liu, S. X. Replication fork activation is enabled by a single-stranded DNA Gate in CMG helicase. *Cell* **178**, 600-+ (2019).
24. Galburt, E. A. et al. Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature* **446**, 820–823 (2007).
25. Sheridan, R. M., Fong, N., D'Alessandro, A. & Bentley, D. L. Widespread backtracking by RNA Pol II is a major effector of gene activation, 5′ pause release, termination, and transcription elongation rate. *Mol. Cell* **73**, 107–118 (2019).
26. Yang, K. B. et al. Persistence of backtracking by human RNA polymerase II. *Mol. Cell* **84**, 897–909 (2024).
27. Peterlin, B. M. & Price, D. H. Controlling the elongation phase of transcription with P-TEFb. *Mol. Cell* **23**, 297–305 (2006).
28. Buratowski, S. Progression through the RNA polymerase II CTD cycle. *Mol. Cell* **36**, 541–546 (2009).
29. Aoi, Y. et al. SPT5 stabilization of promoter-proximal RNA polymerase II. *Mol. Cell* **81**, 4413–4424 (2021).
30. Decker, T. M. Mechanisms of transcription elongation factor DSIF (Spt4-Spt5). *J. Mol. Biol.* **433**, 166657 (2021).
31. Song, A. & Chen, F. X. The pleiotropic roles of SPT5 in transcription. *Transcription* **13**, 53–69 (2022).
32. Guo, S. et al. A regulator of transcriptional elongation controls vertebrate neuronal development. *Nature* **408**, 366–369 (2000).

33. Wada, T. et al. DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev.* **12**, 343–356 (1998).
34. Rosen, G. A. et al. Dynamics of RNA polymerase II and elongation factor Spt4/5 recruitment during activator-dependent transcription. *Proc. Natl Acad. Sci. USA* **117**, 32348–32357 (2020).
35. Bernecky, C., Plitzko, J. M. & Cramer, P. Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. *Nat. Struct. Mol. Biol.* **24**, 809–815 (2017).
36. Dollinger, R. et al. Assessment of the roles of Spt5-nucleic acid contacts in promoter proximal pausing of RNA polymerase II. *J. Biol. Chem.* **299**, 105106 (2023).
37. Hou, L. et al. Paf1C regulates RNA polymerase II progression by modulating elongation rate. *Proc. Natl Acad. Sci. USA* **116**, 14583–14592 (2019).
38. Chen, F. X. et al. PAF1, a molecular regulator of promoter-proximal pausing by RNA Polymerase II. *Cell* **162**, 1003–1015 (2015).
39. Wang, Z. et al. Coordinated regulation of RNA polymerase II pausing and elongation progression by PAF1. *Sci. Adv.* **8**, eabm5504 (2022).
40. Dermody, J. L. & Buratowski, S. Leo1 subunit of the yeast paf1 complex binds RNA and contributes to complex recruitment. *J. Biol. Chem.* **285**, 33671–33679 (2010).
41. Versluis, P. et al. Live-cell imaging of RNA Pol II and elongation factors distinguishes competing mechanisms of transcription regulation. *Mol. Cell* **84**, 2856–2869 (2024).
42. Qiu, H. F., Hu, C. H., Gaur, N. A. & Hinnebusch, A. G. Pol II CTD kinases Bur1 and Kin28 promote Spt5 CTR-independent recruitment of Paf1 complex. *Embo J.* **31**, 3494–3505 (2012).
43. Ardehali, M. B. et al. Spt6 enhances the elongation rate of RNA polymerase II in vivo. *Embo J.* **28**, 1067–1077 (2009).
44. Endoh, M. et al. Human Spt6 stimulates transcription elongation by RNA polymerase II in vitro. *Mol. Cell Biol.* **24**, 3324–3336 (2004).
45. Narain, A. et al. Targeted protein degradation reveals a direct role of SPT6 in RNAPII elongation and termination. *Mol. Cell* **81**, 3110–3127 (2021).
46. Ehara, H., Kujirai, T., Shirouzu, M., Kurumizaka, H. & Sekine, S. I. Structural basis of nucleosome disassembly and reassembly by RNAPII elongation complex with FACT. *Science* **377**, eabp9466 (2022).
47. Ellison, M. A. et al. Spt6 directly interacts with Cdc73 and is required for Paf1 complex occupancy at active genes in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **51**, 4814–4830 (2023).
48. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
49. Cossa, G., Parua, P. K., Eilers, M. & Fisher, R. P. Protein phosphatases in the RNAPII transcription cycle: erasers, sculptors, gatekeepers, and potential drug targets. *Genes Dev.* **35**, 658–676 (2021).
50. Hu, S. et al. SPT5 stabilizes RNA polymerase II, orchestrates transcription cycles, and maintains the enhancer landscape. *Mol. Cell* **81**, 4425–4439 (2021).
51. Cortazar, M. A. et al. Control of RNA Pol II speed by PNUTS-PP1 and Spt5 dephosphorylation facilitates termination by a 'sitting duck torpedo' mechanism. *Mol. Cell* **76**, 896–908 (2019).
52. Sun, R. & Fisher, R. P. Tripartite phosphorylation of SPT5 by CDK9 times pause release and tunes elongation rate of RNA polymerase II. *Mol. Cell* **85**, 1743–1759 (2025).
53. Adelman, K. et al. Drosophila Paf1 modulates chromatin structure at actively transcribed genes. *Mol. Cell Biol.* **26**, 250–260 (2006).
54. Mayekar, M. K., Gardner, R. G. & Arndt, K. M. The recruitment of the *Saccharomyces cerevisiae* Paf1 complex to active genes requires a domain of Rtf1 that directly interacts with the Spt4-Spt5 complex. *Mol. Cell Biol.* **33**, 3259–3273 (2013).
55. Li, B., Carey, M. & Workman, J. L. The role of chromatin during transcription. *Cell* **128**, 707–719 (2007).

Methods

Protein purification

Pol II. Pol II was isolated from *Sus scrofa* thymus and purified following previously established protocols[56,57]. The purified Pol II was aliquoted and stored in liquid nitrogen.

TFIIS. The full-length open reading frame (ORF) of human TFIIS was cloned into a modified pRSFDuet-1 vector and expressed in *E. coli* BL21 (DE3) cells. Cells were grown at 37 °C until reaching an optical density at 600 nm (OD_{600}) of around 0.6. The temperature was decreased to 16 °C and protein expression was induced by adding 0.1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were grown for an additional 16 h at 16 °C and were gathered and resuspended in a buffer containing 30 mM Tris-HCl 8.0, 300 mM NaCl and 20 mM imidazole. Cells were lysed by high pressure at 4 °C for 20 min and the lysate was cleared by centrifugation at 20,000*g* for 40 min at 4 °C. The supernatant was incubated with Ni-NTA resin for 1 h. Beads were collected and washed with buffer containing 30 mM Tris-HCl pH 8.0, 300 mM NaCl, 25 mM imidazole and eluted with a buffer containing 30 mM Tris-HCl pH 8.0, 300 mM NaCl and 250 mM imidazole. The elute was dialyzed against a buffer containing 30 mM Tris-HCl pH 8.0 and 300 mM NaCl and supplied with HRV 3 C protease to cleave the 6× histidine-SUMO-tag overnight. The eluate after dialysis was loaded onto a mono S 5/50 GL column (Cytiva), pre-equilibrated with buffer A (30 mM Bis-Tris pH 6.4 and 30 mM NaCl). The target protein was eluted with a gradient from 0% to 100% buffer B (30 mM Bis-Tris pH 6.4 and 1 M NaCl) over 30 column volumes. The peak fractions were pooled, concentrated and dialyzed overnight against a buffer containing 30 mM HEPES pH 7.9, 100 mM KCl, 2 mM $MgCl_2$, 2 mM dithiothreitol (DTT) and 5% (v/v) glycerol. The protein was aliquoted, flash-frozen in liquid nitrogen and stored at −80 °C.

P-TEFb. The two full-length ORFs of human P-TEFb were cloned into a modified pCAG vector, and CCNT1 was tagged with an N-terminal protein A-SUMO tag. The plasmid was transfected into Expi293F cells for overexpression. After culture at 37 °C for 72 h, cells were collected and lysed in lysis buffer containing 30 mM HEPES pH 8.0, 300 mM NaCl, 0.25% CHAPS, 5 mM ATP, 5 mM $MgCl_2$, 0.2 mM EDTA, 2 mM DTT, 10% glycerol (v/v), 1 mM PMSF, 1 μg ml^{-1} aprotinin, 1 μg ml^{-1} pepstatin and 1 μg ml^{-1} leupeptin at 4 °C for 30 min. The lysate was clarified by centrifugation at 20,000*g* for 40 min at 4 °C. The supernatant was incubated with IgG resin for 1.5 h, followed by on-column digestion with ULP1 protease at 4 °C for 1 h. The immobilized proteins were eluted and further purified by a mono S 5/50 GL column (Cytiva) pre-equilibrated in buffer A (30 mM HEPES pH 7.0, 30 mM NaCl, 0.05% CHAPS, 2 mM DTT, 2 mM $MgCl_2$ and 5% (v/v) glycerol). The target protein was eluted with a gradient from 0% to 100% buffer B (same as buffer A except with 1 M NaCl) over 30 column volumes. The peak fractions were pooled, aliquoted, flash-frozen in liquid nitrogen and stored at −80 °C.

PAF1C. The five full-length ORFs of human PAF1C were cloned into a modified pCAG vector. CDC73 was tagged with an N-terminal protein A-SUMO tag, and CTR9 was tagged with an N-terminal protein A-SUMO tag or an N-terminal protein A-SUMO-S6 tag (S6: GDSLSWLLRLLN)[58]. The plasmids were transfected into Expi293F cells (Invitrogen) for overexpression. After culture at 37 °C for 72 h, cells were collected and lysed in lysis buffer at 4 °C for 30 min. The lysate was clarified by centrifugation at 20,000*g* for 40 min at 4 °C. The supernatant was incubated with IgG resin for 1.5 h, followed by on-column digestion with ULP1 protease at 4 °C for 1 h. The immobilized proteins were eluted and further purified by a mono S 5/50 GL column (Cytiva) pre-equilibrated in buffer A (30 mM HEPES pH 8.0, 30 mM NaCl, 2 mM DTT, 2 mM $MgCl_2$ and 5% (v/v) glycerol). The target protein was eluted with a gradient from 0% to 100% buffer B (same as buffer A except with 1 M NaCl) over 30 column volumes. The peak fractions were pooled, aliquoted, flash-frozen in liquid nitrogen and stored at −80 °C.

ELOF1. The full-length ORF of human ELOF1 was cloned into a modified pRSFDuet-1 vector and expressed in *E. coli* BL21 (DE3) cells. Cells were grown at 37 °C until reaching an OD_{600} of around 0.6. The temperature was decreased to 16 °C, and protein expression was induced by adding 0.1 mM IPTG. Cells were grown for an additional 16 h at 16 °C, and gathered and resuspended in buffer C containing 50 mM Tris-HCl 8.5, 700 mM NaCl, 0.7 mM β-mercaptoethanol, 5% (v/v) glycerol and 25 mM imidazole. Cells were lysed by high pressure at 4 °C for 20 min, and the lysate was cleared by centrifugation at 20,000*g* for 40 min at 4 °C. The supernatant was incubated with Ni-NTA resin for 1 h. Beads were collected and washed with buffer C. The bound protein was digested using ULP1 protease in buffer D containing 30 mM Tris-HCl pH 8.5, 150 mM NaCl, 0.7 mM β-mercaptoethanol and 5% (v/v) glycerol overnight. The immobilized proteins were then eluted and further purified using a mono Q 5/50 GL column (Cytiva) pre-equilibrated in buffer A (30 mM Tris-HCl pH 8.5, 50 mM NaCl, 2 mM DTT and 5% (v/v) glycerol). The target protein was eluted with a gradient from 0% to 100% buffer B (same as buffer A except with 1 M NaCl) over 30 column volumes. The peak fractions were pooled, aliquoted, flash-frozen in liquid nitrogen and stored at −80 °C.

SPT6. The full-length ORF of human SPT6 was cloned into a modified pCAG vector containing an N-terminal protein A-SUMO tag. For fluorescently labeled SPT6, a Spy Tag003 (RGVPHIVMVDAYKRYK)[59] was inserted at the N terminus of the construct. The plasmid was transfected to Expi293F cells for overexpression. After culture at 37 °C for 72 h, cells were collected and lysed in a buffer containing 30 mM HEPES pH 8.0, 300 mM NaCl, 0.25% CHAPS, 5 mM ATP, 5 mM $MgCl_2$, 0.2 mM EDTA, 2 mM DTT, 10% (v/v) glycerol, 1 mM PMSF, 1 μg ml^{-1} aprotinin, 1 μg ml^{-1} pepstatin and 1 μg ml^{-1} leupeptin at 4 °C for 30 min. The lysate was clarified by centrifugation at 20,000*g* for 40 min at 4 °C, and the supernatant was incubated with IgG resin for 1.5 h, followed by on-column digestion with ULP1 protease at 4 °C for 1 h. The immobilized proteins were eluted and further purified by a mono Q 5/50 GL column (Cytiva) pre-equilibrated in buffer A (30 mM HEPES pH 8.0, 30 mM NaCl, 2 mM DTT, 2 mM $MgCl_2$ and 5% (v/v) glycerol). The target protein was eluted with a gradient from 0% to 100% buffer B (same as buffer A except with 1 M NaCl) over 30 column volumes. The peak fractions were pooled, aliquoted, flash-frozen in liquid nitrogen and stored at −80 °C.

DSIF, RTF1, IWS1. These proteins were purified in a similar way to that described for SPT6. The two full-length ORFs of human DSIF were cloned into a modified pCAG vector, and SPT4 was tagged with an N-terminal protein A-SUMO tag or an N-terminal protein A-SUMO-Spy tag. The plasmids were cotransfected into Expi293F cells for overexpression. The ORF of full-length human RTF1 with a C-terminal Spytag or truncated RTF1 (residues 126–710) was cloned into a modified pCAG containing an N-terminal protein A-SUMO tag and overexpressed in Expi293F cells. The truncated RTF1 showed a similar activity of stimulating transcription to the full-length RTF1[11]. The full-length ORF of human IWS1 was cloned into a modified pCAG vector containing an N-terminal protein A-SUMO tag and overexpressed in Expi293F cells. After cell lysis, the lysate was applied to an IgG resin followed by on-column digestion. The eluate was further purified by a mono Q 5/50 GL column (Cytiva). The peak fractions were pooled, aliquoted, snap frozen and stored at −80 °C.

Protein labeling

To site-specifically label the PAF1C complex, the S6-PAF1C, Sfp synthase and Cy3-CoA dye (SiChem) were incubated at a 1:20:100 molar ratio for 4 h at room temperature and overnight at 4 °C in the presence of 10 mM $MgCl_2$. Excess dye and Sfp synthase were removed by a 50-kDa Amicon spin filter (Millipore) with a buffer containing 30 mM HEPES pH 7.9, 250 mM NaCl, 5 mM $MgCl_2$, 5 mM DTT and 5% (v/v) glycerol.

The labeled protein was evaluated via SDS–PAGE gel electrophoresis. The gel was scanned by Typhoon FLA 7000 (GE Healthcare). The labeling efficiency was estimated to be 92%. Final protein samples were aliquoted, flash-frozen and stored at −80 °C.

DSIF, SPT6 and RTF1 were labeled through the conjugation between fluorescently labeled spycatcher3 protein and the spy tag. Spycatcher3 with a single cysteine residue (Bio-Rad) was first incubated with 10 mM tris(2-carboxyethyl)phosphine (TCEP) (Sigma-Aldrich) at room temperature for 1 h to reduce the thiol group. Cy3 maleimide mono-reactive dye (Cytiva) was then added with a 1:5 protein-to-dye molar ratio. The reaction was conducted at room temperature for 2 h and overnight at 4 °C. Excess dye and TCEP were removed by running Bio-Spin-6 column (Bio-Rad) twice. Spycatcher3 was estimated to be fully labeled by Cy3. Spy-tagged DSIF or full-length RTF1 was mixed with Cy3-spycatcher3 at room temperature for 20 min, with the optimal molar ratio predetermined through titration.

DNA substrate preparation

To generate the 8-kb DNA template for single-molecule experiments, an 8-kb plasmid ply-8 was adapted from a 13.8-kb plasmid plw83 (ref. 60) where a 6-kb fragment of plw83 containing multiple BbsI recognition sites was truncated. One new single BbsI recognition site was inserted into the truncated plw83 plasmid. A cassette harboring seven tandem repeats of a 21-bp sequence (5′-AGACACCACAGACCACACACA) was placed 76 bp downstream of the BbsI site. To generate the 16-kb template plasmid (ply-16), a 7.7-kb DNA segment was produced by PCR from λ DNA (from 14,480 bp to 22,175 bp), then inserted into ply-8 at the position 29 bp downstream of the seven tandem repeats. Ply-8 or ply-16 was digested using BbsI to generate the linear DNA template (step 1 in Extended Data Fig. 2a). The linearized DNA was ethanol precipitated at −20 °C for 1 h by 3 volumes of cold ethanol and 300 mM NaOAc pH 5.2. Precipitated DNA was recovered by centrifugation at >20,000*g* at 4 °C for 30 min. The DNA pellet was washed by 75% ethanol, air-dried and resuspended in ddH_2O. To increase the ligation efficiency and specificity, the (5′CCCA) 4-nt single-stranded DNA (ssDNA) overhang in the linearized DNA was extended to 15 nt by ligation with bridge and overhang oligonucleotides (step 2 in Extended Data Fig. 2a). Specifically, the purified linearized ply-8 was mixed with Overhang-oligo (5′Phos-TGGGTGGTGTTCCGCACATGCTGAGTCGAGCTTA) and Bridge-oligo (5′Phos-ATGTGCGGAACACCA) (Integrated DNA Technologies) with a molar ratio of 1:100:125 in a buffer containing 5 mM Tris-HCl pH 7.5 and 50 mM NaCl. The mixture was heated to 75 °C for 10 min and then cooled down to room temperature gradually for 1 h. T4 ligation buffer and T4 ligase (final concentration of 2 U μl^{-1}) were added and incubated at 16 °C overnight. The ligation product was purified twice by DNA size-selective magnetic beads (Sergi Lab Supplies) to remove the excess oligos (following the manufacturer's protocol) and eluted by ddH_2O.

The 5′ overhangs of the purified ligation product were filled in with biotinylated nucleotides by the exonuclease-deficient DNA polymerase I Klenow fragment (New England BioLabs) to create terminally biotinylated DNA for bead conjugation (step 3 in Extended Data Fig. 2a). The fill-in reaction was conducted by incubating 100–200 nM linearized plasmid DNA, 33 μM each of biotin-14-dATP and biotin-14-dCTP (Jena Bioscience) and 10 U of Klenow fragment in 1× NEB2 buffer at room temperature for 2 h. To stop the reaction, EDTA was added at a final concentration of 10 mM, and the reaction mixture was heat inactivated at 75 °C for 20 min. The DNA was then ethanol precipitated overnight at −20 °C and resuspended in ddH_2O.

Template-oligo (5′Phos-CGTTTGTGTTTTCGGGTCTCCCTCGTTTCTGGCTTGGGTTGGCTTTTCGCCGTGTCGTACATCATACTACTCCTACCAGGAAGCATAAGCTCGACTCAGC, underline indicates the RNA hybridization region) or Nontemplate-oligo (5′Phos-TGCTTCCTGGTAGGAGTAGTATGATGTACGACACGGCGAAAAGCCAACCCAAGCCAGAAACGAGGGAGACCCGAAAACACAAACGTAAGCTCGACTCAGC) (Integrated DNA Technologies) was ligated to the biotinylated 8-kb DNA from step 3 to create nontemplate-strand DNA (8k-NTS-DNA) and template-strand DNA (8k-TS-DNA) handles used for EC assembly (step 4 in Extended Data Fig. 2a). Specifically, Template-oligo or Nontemplate-oligo was mixed with biotinylated 8-kb DNA with a molar ratio of 5:1 in a buffer containing 5 mM Tris-HCl pH 7.5 and 50 mM NaCl. After annealing as described in step 2, T4 ligation buffer and T4 ligase (final concentration of 2 U μl^{-1}) were added and ligated at 16 °C overnight. The ligation product was purified once by DNA size-selective magnetic beads and eluted by ddH_2O with the final DNA concentration ~1,300 ng μl^{-1} (or ~250 nM). The final products were aliquoted and stored at −80 °C.

Single-molecule experiments

Formation of ECs. To prepare ECs as shown in Extended Data Figs. 2b, 2 μl of 8k-TS-DNA (250 nM) was mixed with 2 μl of EB40 buffer (20 mM Tris-HCl pH 8.0, 40 mM KCl and 5 mM $MgCl_2$) and 2 μl of ATTO647N-RNA primer (/5ATTO647N/UUUUUUUUUUUAUAUCGACACGGC, underline indicates the hybridization region; 10 μM, dissolved in EB40) (Integrated DNA Technologies). The mixture was subjected to an annealing procedure as follows: 55 °C for 30 s, 50 °C for 30 s, 45 °C for 30 s, 40 °C for 2 min, 37.5 °C for 2 min, 35 °C for 2 min, 30 °C for 2 min, 27 °C for 2 min, 25 °C for 10 min and 4 °C indefinitely. Following annealing, 0.5 μl of freshly prepared 0.1 M DTT, 0.5 μl of RNase Inhibitor murine (40 U μl^{-1}) and 2 μl of Pol II (3 μM) were added to the DNA–RNA assembly solution. The mixture was incubated at 30 °C for 8 min, followed by 37 °C for 2 min. Then 2 μl of 8k-NTS-DNA (250 nM), 0.52 μl of ATP (25 mM) and 2 μl of P-TEFb (4 μM) were added and incubated at 30 °C for 20 min. Finally, the nicks from 8k-TS-DNA–8k-NTS-DNA hybridization were sealed by adding 0.75 μl of T4 ligase (2,000 U μl^{-1}) and incubating at room temperature for 1 h. The assembled EC was ready to be loaded into the microfluidic chip for single-molecule assays.

Single-molecule data collection. Single-molecule experiments were performed at room temperature on a LUMICKS C-Trap instrument. Channels of the microfluidic chip were passivated by first flowing BSA (0.1% (w/v) in 1× PBS, Sigma-Aldrich) and then Pluronic F127 (0.5% (w/v) in 1× PBS) for 10 min each. Streptavidin-coated polystyrene beads (2.14-μm diameter, Spherotech, diluted in 1× PBS), 16-kb DNA-EC (100 pM in imaging buffer), imaging buffer and transcription buffer (all four ribonucleotides at 1 mM each and elongation factors in imaging buffer) were injected into channels 1–4, respectively. Imaging buffer included an oxygen scavenging system (10 nM protocatechuate-3,4-dioxygenase (Sigma-Aldrich) and 2.5 mM protocatechuic acid (Sigma-Aldrich)), 0.3 U μl^{-1} RNase inhibitor murine, 5 mM freshly made DTT and a triplet-state quenching cocktail (1 mM cyclooctatetraene, 1 mM 4-nitrobenzyl alcohol and 1 mM Trolox (Sigma-Aldrich)). The final concentrations of the elongation factors in transcription buffer for the All-factors condition were: PAF1C 50 nM, SPT6 120 nM, $RTF1_{126\text{-}710}$ 200 nM, TFIIS 0.5 μM, DSIF 200 nM, IWS1 60 nM, ELOF1 1 μM and P-TEFb 50 nM. In those experiments that involved fluorescently labeled factors, a lower concentration was used to reduce the fluorescence background (Cy3-$DSIF^{P+}$ 12.5 nM, Cy3-$DSIF^{P-}$ 12.5 nM, Cy3-PAF1C 25 nM, Cy3-$RTF1_{FL}$ 33 nM and Cy3-SPT6 10 nM). To prepare the transcription buffer in channel 4, a prephosphorylation step was conducted before mixing the elongation factors with the imaging buffer. Specifically, threefold concentrated elongation factors were incubated with 1 mM ATP at 30 °C for 30 min to allow P-TEFb-mediated phosphorylation of a specific set of factors. For the $DSIF^{P-}$ condition, DSIF was omitted from this step and subsequently added to the transcription mix together with 200 nM CDK9 inhibitor flavopiridol (Selleck Chemicals LLC). For the $Pol\ II^{P-}$ condition, P-TEFb was omitted during the steps of EC formation (Extended Data Fig. 2b) and 200 nM flavopiridol was included in the transcription buffer to prevent phosphorylation of Pol II in channel 4. For the $DSIF^{P+}$&Pol

II^{P+} condition, only DSIF was incubated with P-TEFb and ATP, and flavopiridol was included in the transcription buffer. For the experiments involving RNase A or RNase H, the concentration was used at 6.6 μg ml^{-1} (RNase A, Thermo Scientific) or 0.1 U μl^{-1} (RNase H, New England BioLabs) in the transcription buffer.

A single DNA tether was caught between a pair of streptavidin beads held in optical traps, and the tether was extended at 5 pN of tension. The tether was moved to channel 3 for confocal scanning using a 638-nm laser to verify the existence of assembled EC. Empty tethers and those with protein aggregates were discarded. Kymographs were generated using the Bluelake software (LUMICKS) via confocal line scanning through the center of the two beads at 0.2 s per line (pixel time: 0.2 ms). Data acquisition started in channel 3 for 20 s for baseline calibration and continued in channel 4. When applicable, 10 nM of Cy3-labeled DNA oligo probes (/5Cy3/AGTGTGTGGTCTGTGGTGTCT, Integrated DNA Technologies) were included in channel 4 to detect nascent RNA.

Single-molecule data analysis

Data acquisition. Single-molecule force and fluorescence data from the .h5 files generated by Bluelake were processed using the lumicks.pylake Python library, supplemented by other Python modules integrated into a custom graphical user interface script, CTrapVis.py (http://harbor.lumicks.com/single-script/c5b103a4-0804-4b06-95d3-20a08d65768f). This script was used to export confocal scans and kymographs in TIFF format, as well as to extract real-time data for force, distance and photon counts. The positions of labeled RNA and DNA probes on the template over time were determined from kymographs using a custom script, kymotracker_calling_script.py (https://harbor.lumicks.com/single-script/4db9d63e-1f93-0c0-9b90-e99066469578), which implemented the greedy line-tracking algorithms provided by Python package lumicks.pylake to define line traces. Optimal tracking of EC was achieved by setting a pixel threshold of 1 photon count for ATTO647N signals and 2 photon counts for Cy3 signals, with a line width of 12 pixels, a window of 35 frames and a minimum line length of 20 frames. The extracted trajectories were aligned based on the starting position of the ATTO647N-RNA using data collected in channel 3. Distance traveled in nanometers was converted to base pairs using a scaling factor of 0.29 nm per bp, reflecting the extension of B-form DNA under 5 pN of tension.

Segmentation analysis of extracted transcription trajectories. *Structure of the data*. A condition is specified by the composition of the EC. In this study, we tested 14 different conditions that correspond to different subsets of the 8 elongation factors (not exhaustive). We refer to each condition by the variable $c = 1, 2, \ldots, 14$. For example, the EC with All-factors is $c = 1$; the EC without TFIIS is $c = 2$; the EC without IWS1 is $c = 3$ and so on. In an Excel data file, each condition has its own sheet.

For each EC c, we have $\mathbf{T}(c)$ trajectories. Different conditions c may have different numbers of trajectories $\mathbf{T}(c)$. In condition c, each trajectory $t = 1, 2, \ldots, \mathbf{T}(c)$ consists of a pair of column vectors ($\mathbf{S}(c,t)$, $\mathbf{K}(c,t)$) that are of the same length $L(c,t)$. The length $L(c,t)$ measures the number of recorded observations of trajectory t in condition c in $\mathbf{S}(c,t)$ and also in $\mathbf{K}(c,t)$. These vectors appear in adjacent vertical columns in an Excel spreadsheet. The first row of each sheet gives a unique alphanumeric identifier for each trajectory. **S** stands for 'seconds' and **K** stands for 'kilobases'. For trajectory t in condition c, the jth element of $\mathbf{S}(c,t)$, which we label $\mathbf{S}(c,t,j)$, is the time at which the jth observation is recorded, and $j = 1, 2, \ldots, L(c,t)$. Time is measured in seconds from the beginning of a trajectory. The corresponding jth element $\mathbf{K}(c,t,j)$ of $\mathbf{K}(c,t)$ is the transcription length at time $\mathbf{S}(c,t,j)$. Transcription length is measured in kilobases.

Each trajectory is segmented separately. We simplify the notation used to describe how a trajectory is segmented: assume a fixed condition c and a fixed trajectory t in condition c. We describe the segmentation analysis as applied to one pair of vectors (**S**, **K**) each of length L. For $j = 1, 2, \ldots, L$, the jth time of observation is $\mathbf{S}(j)$ and the transcription length at time $\mathbf{S}(j)$ is $\mathbf{K}(j)$. The goal of the following analysis is to approximate the plot of **K** as a function of **S** by a sequence of linear segments (Fig. 1c and Extended Data Fig. 3) that may have slope 0 (horizontal segment) or may have a positive slope (as transcription proceeds) or a negative slope (as transcription is reversed or backtracking of Pol II). Each segment is separated from its preceding segment (if any) and from its following segment (if any) by a change point, a time at which the estimated slope changes.

MATLAB software 'FitPiecewiseLinearContinuous.m'. Our code is written in MATLAB (MathWorks v.R2023b). The user chooses several parameters that govern the software. These parameters are explained below.

(1) readmat = 0 means read the data from an Excel spreadsheet; readmat = 1 means read data from a .mat file where it was stored previously;
(2) Nsheets = C is the number of conditions;
(3) globalmaxchangepoints = 15 limits maximum number of change points (NCP);
(4) mintimegap = 3 seconds sets minimal time (s) between change points; if two change points are initially estimated within 'mintimegap', they are replaced by one change point at the midpoint;
(5) $\delta = 10^{-6}$, the width of the interval of random perturbations of the recorded times, to break ties.

When the data are read from an Excel file, the user must specify which columns house data. For each condition (Excel sheet) c, our software calculates the number $\mathbf{T}(c)$ of trajectories from the number of columns that house data. When the data are read from a .mat file, the saved variables are:

(1) Nsheets = C, the number of sheets or conditions;
(2) sheetnames, a list of alphanumeric names for each condition;
(3) Ntrajectories = $T = (T(1), \ldots, T(C))$, a vector of the number $\mathbf{T}(c)$ of trajectories in each condition c, for $c = 1, 2, \ldots, C$;
(4) outnum, all the numerical values in each sheet;
(5) outtxt, all the alphanumeric column headings, which are the names of the individual trajectories;
(6) sheetnames, stringsheetnames, two formats for the names of the sheets or conditions;
(7) time = $\mathbf{S}(c,t,j)$, a cell array of the time observations in condition c for trajectory t for $j = 1, \ldots, L(c,t)$, where $L(c,t)$ is the number of observations in condition c for trajectory t;
(8) bp = $\mathbf{K}(c,t,j)$, a cell array of the observations of transcription length at each time in condition c for trajectory t for $j = 1, \ldots, L(c,t)$, where $L(c,t)$ is the number of observations in condition c for trajectory t;
(9) trajectorynames, a cell array of length c (one cell for each condition) of the names of the individual trajectories, taken from outtxt; the jth cell contains a vector of length $\mathbf{T}(c)$, each element of which is a string with the name of the corresponding trajectory.

Sometimes the identical reported time will occur more than once. To break all such ties so that the observations occur in a unique order, we add to each time an independently chosen small random quantity uniformly distributed between $-\delta$ and $+\delta$. We set $\delta = 10^{-6}$ s; the user can change the value at will. After this random perturbation, the vectors **S** and **K** are sorted in increasing order of the (perturbed) time element in **S**.

Bayesian estimation of the number of change points. We used the MATLAB version of Rbeast. Rbeast is available in MATLAB, R, Python and C. The source code and installation instructions are available at https://github.com/zhaokg/Rbeast. According to the website, 'BEAST (Bayesian Estimator of Abrupt change, Seasonality, and Trend) is a fast, generic Bayesian model averaging algorithm to decompose time series

or 1D [one-dimensional] sequential data into individual components, such as abrupt changes, trends and periodic/seasonal variations[61]. BEAST is useful for change-point detection (for example, breakpoints, structural breaks, joinpoints, regime shifts or anomalies), trend analysis, time series decomposition (for example, trend versus seasonality), time series segmentation, and interrupted time series analysis.' The specific version we used is beast_irreg with parameters 'deltat', '1', 'season' and 'none'.

Rbeast operates by estimating a probability (called $q(1)$) for the possibility that there is just one change point (without specifying the location of that change point). Then, conditional on one change point, Rbeast estimates a probability (called $q(2) < q(1)$) for the possibility that there is just one more change point. Then, conditional on two change points, Rbeast estimates a probability (called $q(3) < q(2)$) for the possibility that there is just one more change point (that is, a total of three) and so on, possibly up to our chosen globalmaxchangepoints. We find the maximal number of change points, called MaxCP, such that the cumulative sum of the first MaxCP elements of q does not exceed 0.99.

Locating the change points for linear segments. With this number MaxCP of change points, we identify the times at which those change points occur using the MATLAB function ischange with parameter 'linear'. We approximate K as a piecewise (not necessarily continuous) linear function of the perturbed times while limiting the maximum number of changes to MaxCP (using 'MaxNumChanges', MaxCP). The ischange algorithm is based on ref. 62. With the parameter 'linear', ischange fits straight lines to each segment between consecutive change points but the successive straight lines may not connect into one continuous trajectory. If the difference in seconds between two change points is less than our chosen value for the parameter mintimegap (3 s), then we replace the times at which both of those change points occur by a single time located at the arithmetic average of the original two times. This adjustment excludes any segments of duration less than mintimegap. We record the number of change points before and after this adjustment to the data as NchangepointsBefore and NchangepointsAfter, respectively.

Finding a continuous piecewise linear approximation. Given the (perturbed) times S and corresponding kilobase pairs transcribed K, and given the locations of the change points, we fit a continuous piecewise linear approximation to K using fitPiecewiseLinearFunctionGolovchenko.m written by N. Golovchenko in 2004 (https://www.golovchenko.org/home/pwl_fit). This code fits 'a piecewise continuous [linear] function $f(x)$ to the pairs of data points (x,y) such that the sum of squares of error is minimal'. It outputs $x0$, the values of x that define ends of segments of the fitted function $f(x)$, and p, the end points of the linear segments, $p = f(x0)$.

Tabulated results. A segment is defined by four numbers: its initial (perturbed) time S_0, its initial kb K_0, its final (perturbed) time S_1 and its final kb K_1. The duration of a segment is defined as $S_1 - S_0$. The length of a segment is defined as $K_1 - K_0$. The slope of a segment BpPerSecBetwCPs is defined as the ratio of length divided by duration, $(K_1 - K_0)/(S_1 - S_0)$. We define a segment to belong to active transcription (slopeClass 1) if and only if its slope satisfies BpPerSecBetwCPs greater than or equal to 1 nt s^{-1} and its length exceeds 100 nt. Such a segment has progressing transcription. We define a segment to belong to pause (slopeClass 0) if and only if its slope is less than 1 nt s^{-1}. We tabulated the principal characteristics of all the segments of all the trajectories of all the conditions in a table with these columns: sheet (condition), trajectory (or trace) in that condition, segment number in that trajectory, start time for that segment, duration (seconds) of that segment, segment length (kbp), slope and slopeClass. This table has one row for each segment. Further statistical analysis is based on this table.

Analysis of active EC fraction and transcribed length. The parameter 'active EC fraction' (Extended Data Fig. 5a) indicates the fraction of EC that displayed transcription activity (emergence of the DNA probe signals) under a given condition. For the All-factors condition, the maximal distance that an EC traveled before stalling was averaged and reported as the parameter 'transcribed length' (Extended Data Fig. 5b). For all the other conditions (for example, condition X), the transcribed length was calculated by multiplying the averaged maximal distance by a scaling factor (active EC fraction$_X$/active EC fraction$_{\text{All-factors}}$).

Bulk transcription assays

To prepare the EC for bulk assays, 5 µl of Bulk-template-TS (0.6 µM, dissolved in EB40 buffer) was mixed with 1.25 µl of Cy5-RNA primer (/5Cy5/UUUUUUUUUUUAUAUCGACACGGC; 3 µM, dissolved in EB40 buffer). The mixture was subjected to an annealing procedure as follows: 55 °C for 30 s, 50 °C for 30 s, 45 °C for 30 s, 40 °C for 2 min, 37.5 °C for 2 min, 35 °C for 2 min, 30 °C for 2 min, 27 °C for 2 min, 25 °C for 10 min and 4 °C indefinitely. Following annealing, 2.5 µl of freshly prepared 0.1 M DTT, 2.5 µl of RNase inhibitor murine (40 U μl^{-1}) and 5 µl of Pol II (1 µM) were added to the DNA–RNA hybrid solution. The mixture was incubated at 30 °C for 8 min, followed by 37 °C for 2 min. Next 5 µl of Bulk-template-NTS (0.6 µM, dissolved in EB40) was added and incubated at 30 °C for 10 min. Then, 2.4-µl aliquots of EC were mixed with 7.6 µl of elongation factor mixture (final concentration: 100 nM P-TEFb, 120 nM SPT6, 50 nM PAF1C, 200 nM RTF1$_{126\text{-}710}$, 500 nM TFIIS, 200 nM DSIF, 1 µM ELOF1 and 60 nM IWS1) or under conditions where individual factors were omitted (Extended Data Fig. 1b). 1 mM ATP was added to the mixture and incubated for 20 min at 30 °C to allow phosphorylation. All four ribonucleotides were added to 1 mM each to allow transcription for 20 min at 30 °C. Then 2 µl of DNase 10× buffer and 2 µl of DNase I (RNase-free) (2 U μl^{-1}, New England BioLabs) were added and incubated for 10 min at 37 °C to digest the DNA. Then 8 µl of stop buffer (100 mM Tris-HCl pH 7.5, 150 mM EDTA and 4 M urea) and 1 µl of protease K were added and incubated at 37 °C for 20 min. Finally, the RNA products were mixed with 2× RNA loading buffer and loaded on a 10% urea–PAGE gel after incubating at 90 °C for 5 min. The gel was scanned by Typhoon FLA 7000 (GE Healthcare) and quantified by ImageJ.

Bulk-template-TS (146 nt, Integrated DNA Technologies). 5′CTGCGCGTAATCTGCTGCTTGCTTGCAAACTATCCGGTAACTATCGTCTTGAGTCCTATGCTTGCCGCTGAGTCGAGCTTAGTCACCATCGATTGTCGCTTGGGTTGGCTTTTCGCCGTGTCGTGGCTGAGCTCTGGCTGTAAGTG (underline indicates the RNA hybridization region).

Bulk-template-NTS (146 nt, Integrated DNA Technologies). 5′CACTTACAGCCAGAGCTCAGCCACGACACGGCGAAAAGCCAACCCAAGCGACAATCGATGGTGACTAAGCTCGACTCAGCGGCAAGCATAGGACTCAAGACGATAGTTACCGGATAGTTTGCAAGCAAGCAGCAGATTACGCGCAG.

Bulk phosphorylation assays

To phosphorylate Pol II, 1 µl of Pol II (3 µM) was mixed with 1 µl, 2 µl, 3 µl or 4 µl of P-TEFb (1 µM) in the presence of 1 mM ATP and 2 mM $MgCl_2$. To phosphorylate DSIF, 2 µl of DSIF was mixed with 2.5 µl of P-TEFb (1 µM) in the presence of 1 mM ATP and 2 mM $MgCl_2$. The reaction was conducted at 30 °C for 30 min. As a negative control, 2 µl of DSIF was mixed with 1 µl of Lambda Protein Phosphatase (New England BioLabs) in the presence of 1× NEBuffer and 0.1 mM $MnCl_2$. The reaction was conducted at 30 °C for 30 min. The products were evaluated by SuperSep Phos-tag Precast 7.5% Gels (FUJIFILM Irvine Scientific).

Structural modeling

To generate a composite structural model (Fig. 1a) that includes all the elongation factors (except P-TEFb) used in our system, we first downloaded the model of human EC containing PAF1C, SPT6, DSIF

and RTF1 (Protein Data Bank (PDB) 6TED) using ChimeraX, and then added a human Pol II structure containing TFIIS (PDB 8A40), a TC-NER structure containing ELOF1 (PDB 8B3F) and a yeast EC structure containing the IWS1 homolog Spn1 (PDB 7XN7), superimposing the Pol II in each structure. Residues 550–687 of the AlphaFold 2 model of human IWS1 (AF Q96ST2) was superimposed to Spn1 in the yeast EC. We used AlphaFold 3 to model the interaction between PAF1C CDC73 and Pol II CTD. For the CDC73Ras-like domain ($CDC73^{Ras}$), residues 351–531 were chosen. For $CTD^{S2P,S5P}$, we used the sequence (YpSPTpSPS) that was shown to bind yeast Cdc73[42]. The highest ranked model was chosen (predicted template modelling score (pTM) = 0.88). The proposed model in Extended Data Fig. 9b was based on the human EC structure (PDB 6TED). Our AlphaFold prediction was loaded using ChimeraX and placed according to a yeast structure containing the $Spt6^{SH2}$–$Cdc73^{Ras}$ complex[46]. The AlphaFold webserver (https://alphafoldserver.com) was used and accessed on 1 August 2025.

Reporting summary

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

Data availability

All kymographs used for analysis are available via Zenodo at https://doi.org/10.5281/zenodo.17193150 (ref. 63). Source data are provided with this paper.

Code availability

Kymographs were processed and analyzed using a custom script that can be accessed via LUMICKS Harbor at https://harbor.lumicks.com/single-script/c5b103a4-0804-4b06-95d3-20a08d65768f. For segmentation analysis of extracted transcription trajectories, the software has been uploaded to GitHub at https://github.com/yorquant/Fitting-Piecewise-Linear-Continuous/tree/main for downloading for noncommercial purposes.

References

56. Hu, X. et al. A Mediator-responsive form of metazoan RNA polymerase II. *Proc. Natl Acad. Sci. USA* **103**, 9506–9511 (2006).
57. Chen, X. et al. Structural insights into preinitiation complex assembly on core promoters. *Science* **372**, eaba8490 (2021).
58. Zhou, Z. et al. Genetically encoded short peptide tags for orthogonal protein labeling by sfp and AcpS phosphopantetheinyl transferases. *ACS Chem. Biol.* **2**, 337–346 (2007).
59. Keeble, A. H. et al. Approaching infinite affinity through engineering of peptide-protein interaction. *Proc. Natl Acad. Sci. USA* **116**, 26523–26533 (2019).
60. Wang, L., Watters, J. W., Ju, X., Lu, G. & Liu, S. Head-on and co-directional RNA polymerase collisions orchestrate bidirectional transcription termination. *Mol. Cell* **83**, 1153–1164 (2023).
61. Zhao, K. G. et al. Detecting change-point, trend, and seasonality in satellite time series data to track abrupt changes and nonlinear dynamics: a Bayesian ensemble algorithm. *Remote Sens. Environ.* https://doi.org/10.1016/j.rse.2019.04.034 (2019).
62. Killick, R., Fearnhead, P. & Eckley, I. A. Optimal detection of changepoints with a linear computational cost. *J. Am. Stat. Assoc.* **107**, 1590–1598 (2012).
63. Wang, Y. Kymographs of mammalian transcription elongation. *Zenodo* https://doi.org/10.5281/zenodo.17193150 (2025).

Acknowledgements

We thank F. Chen (Fudan University) and G. Chua (Rockefeller University) for critical reading of the manuscript and B. Katch (Rockefeller University) for technical assistance. S.L. is supported by the Alfred P. Sloan Foundation, the Robertson Foundation, the Jensam Foundation and the Marlene Hess Center for Research on Women's Health and Biomedicine at Rockefeller University. Y.X. and X.C. are supported by grants from the National Key R&D Program of China (grant no. 2021YFA1300100), the National Natural Science Foundation of China (grant no. 32271242), the XPLORER Prize and the New Cornerstone Investigator Program.

Author contributions

Y.W., X.C., Y.X. and S.L. designed the experiments. Y.W. performed the experiments and conducted data analysis. X.C. prepared the protein reagents. M.K. assisted with single-molecule data collection and performed structural modeling. J.W.W. wrote data acquisition scripts. J.E.C. developed data analysis software. S.L., Y.W. and Y.X. wrote the paper with inputs from all authors.

Competing interests

The authors declare no competing interests.

Additional information

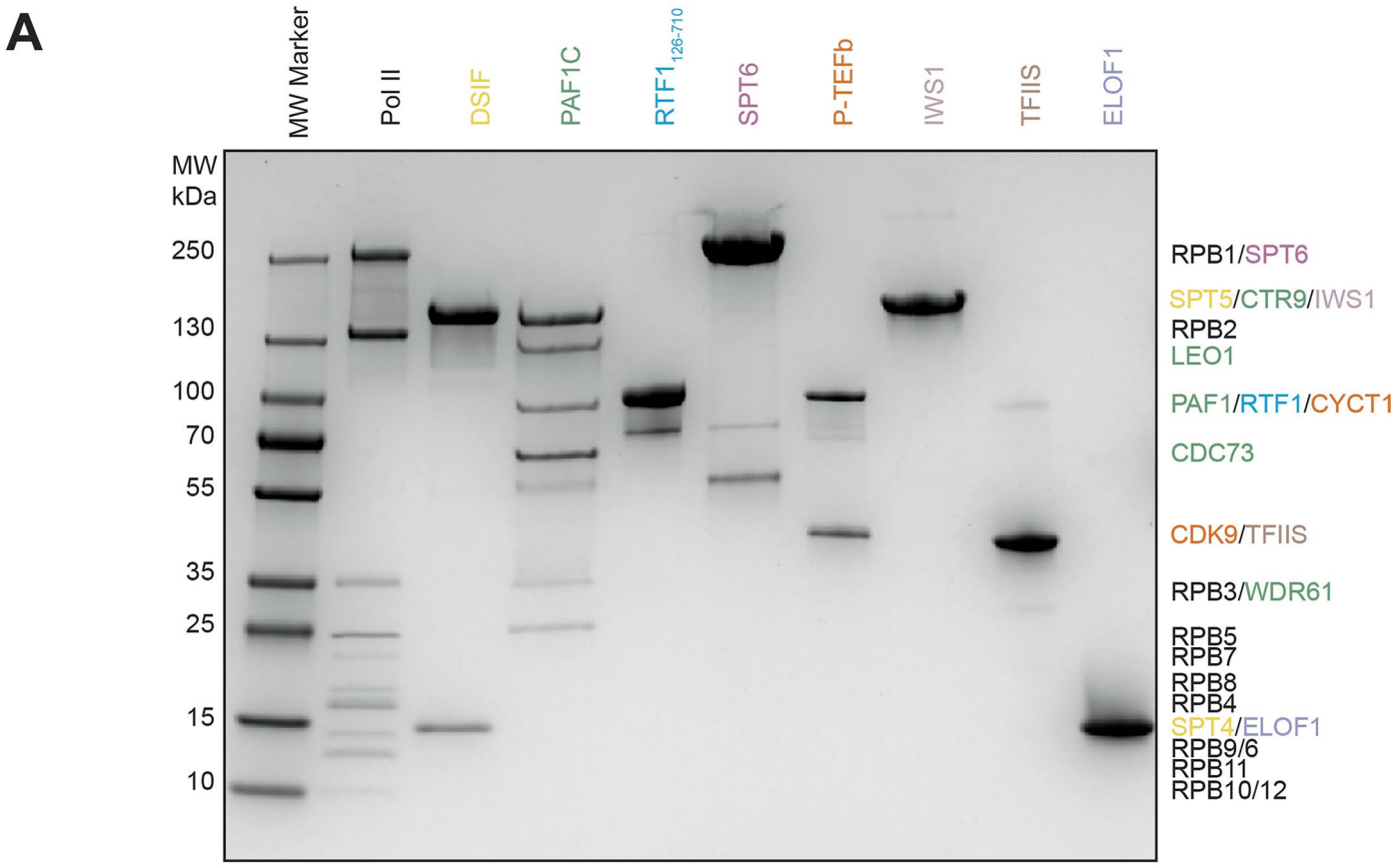
Extended data is available for this paper at https://doi.org/10.1038/s41594-025-01707-1.

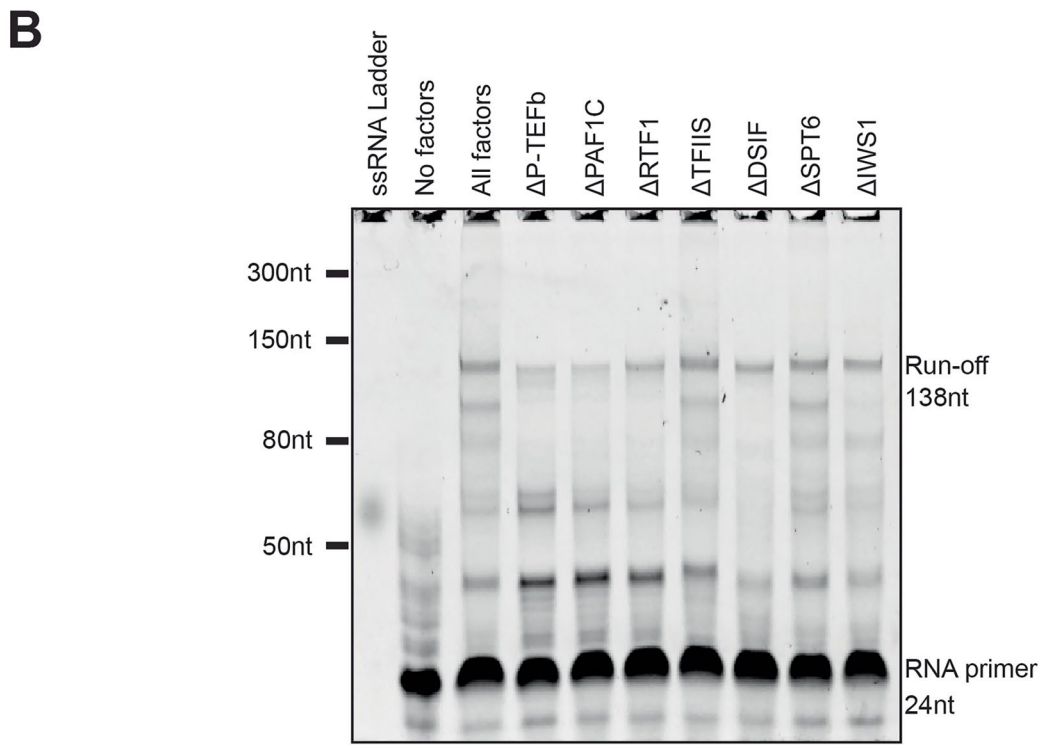
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41594-025-01707-1.

Correspondence and requests for materials should be addressed to Yanhui Xu or Shixin Liu.

Peer review information *Nature Structural & Molecular Biology* thanks Takashi Fukaya and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available. Primary Handling Editor: Dimitris Typas, in collaboration with the *Nature Structural & Molecular Biology* team.

https://doi.org/10.1038/s41594-025-01707-1



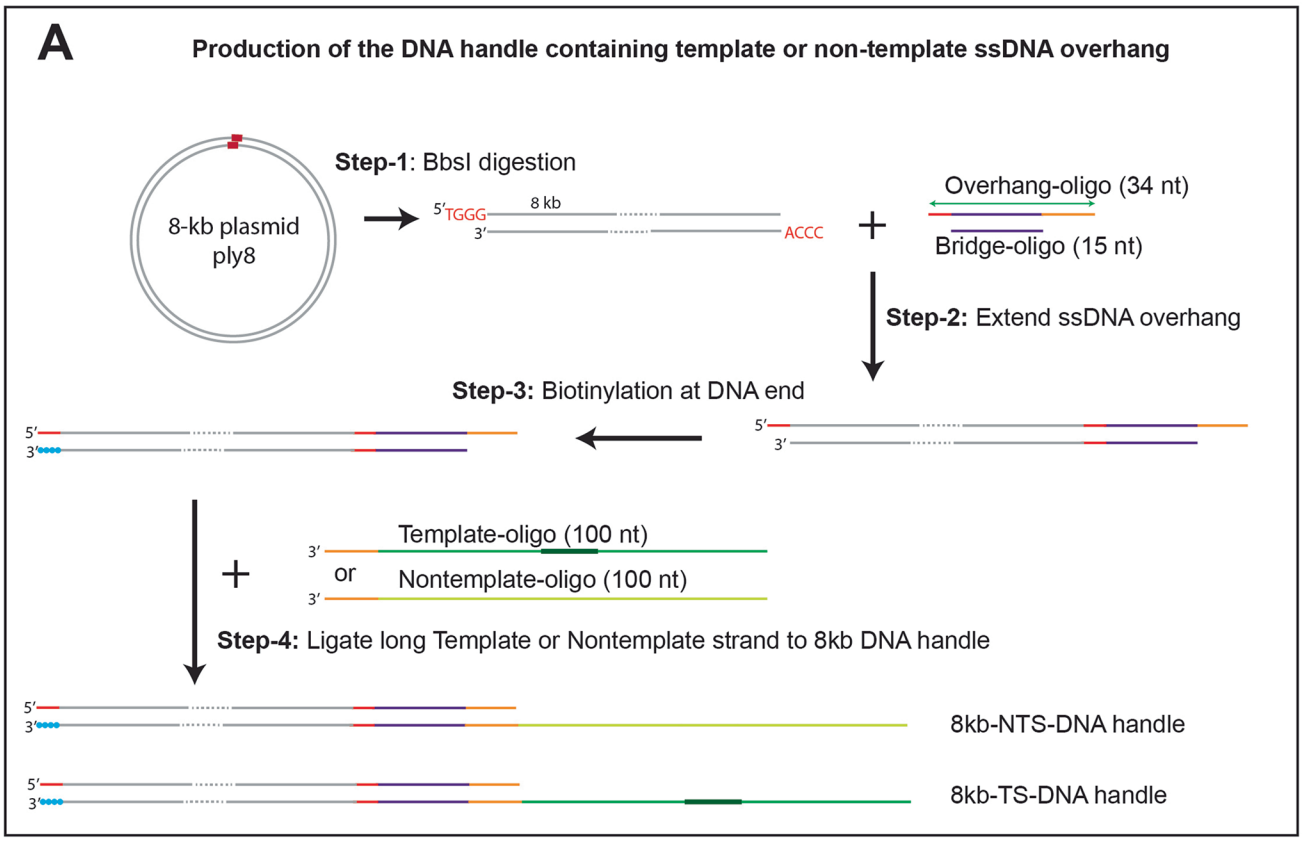


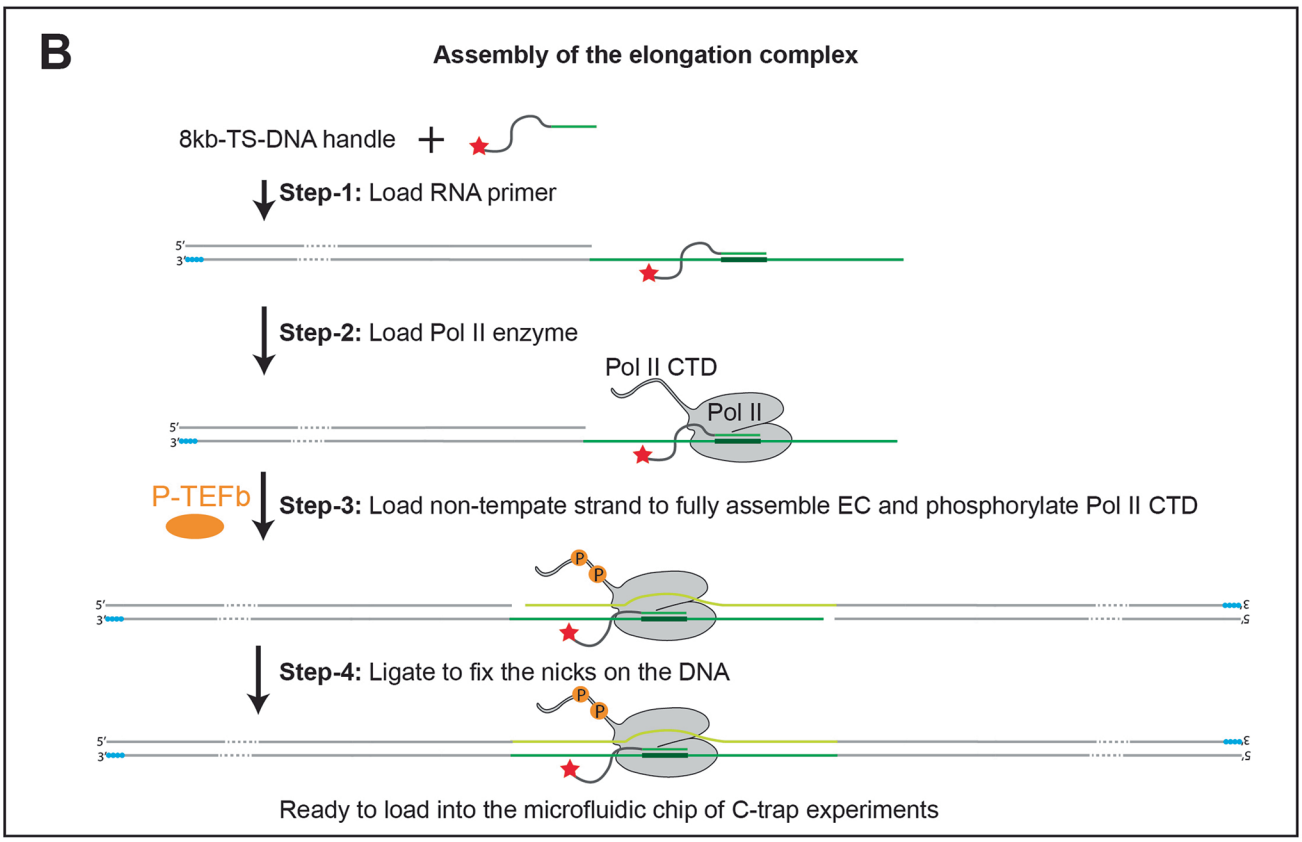
Extended Data Fig. 1 | See next page for caption.

https://doi.org/10.1038/s41594-025-01707-1

Extended Data Fig. 1 | Biochemical evaluations of the proteins used in this study. A, Coomassie-stained SDS-PAGE gel showing the purified mammalian RNA polymerase II (Pol II) and elongation factors. Protein subunit identities based on their molecular weights are annotated on the right. **B**, Bulk transcription assay using the proteins above and a synthetic DNA scaffold with a Cy5-labeled RNA primer. Reconstituted EC (50 nM) was incubated with the full set of elongation factors (All-factors) or under conditions where one factor was omitted in the presence of NTPs (1 mM each). RNA extension was monitored on a denaturing gel by Cy5 fluorescence. The positions of the RNA primer (24 nt) and the run-off product (138 nt) are indicated. The experiment was repeated three times with similar results.

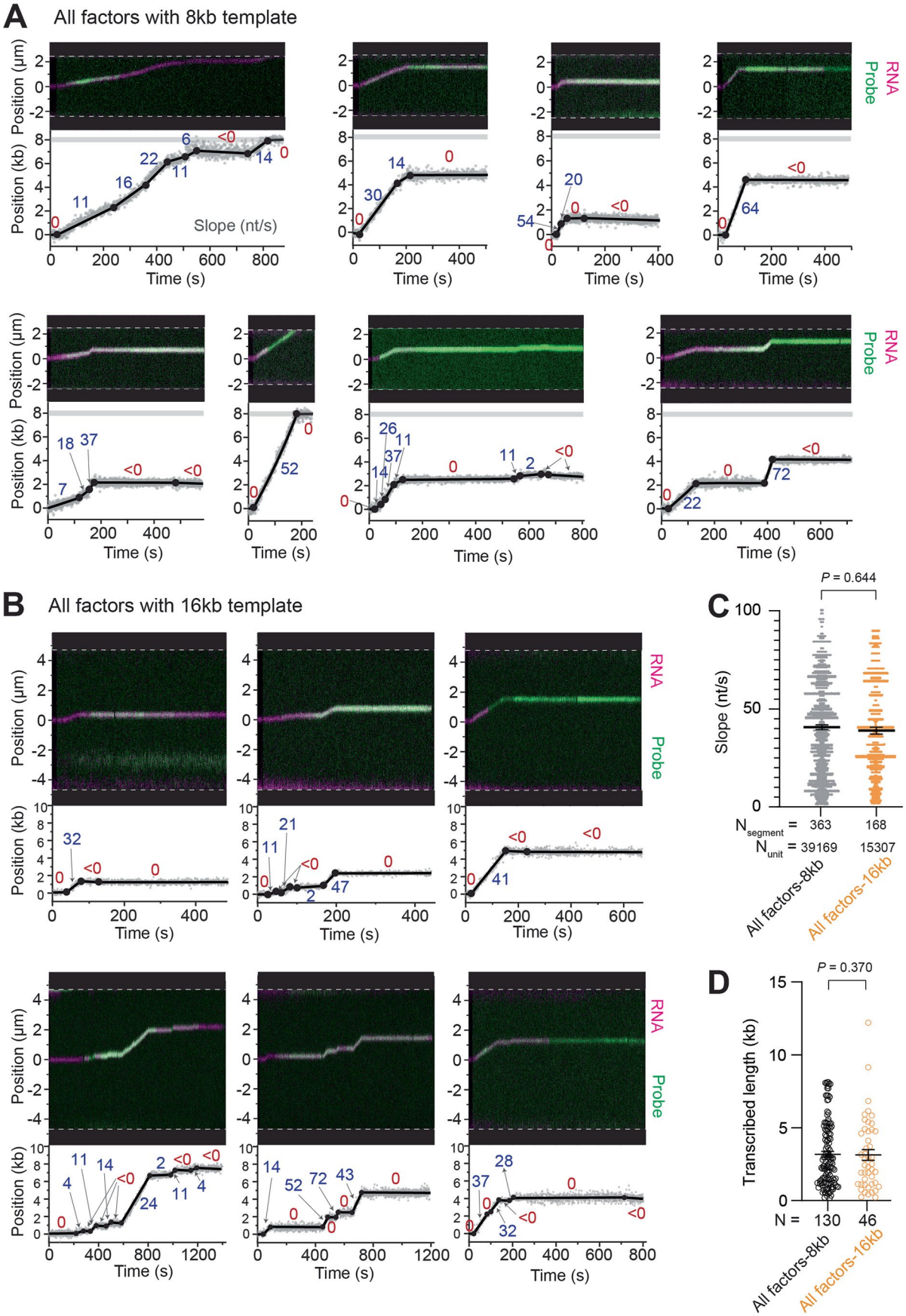
https://doi.org/10.1038/s41594-025-01707-1





Extended Data Fig. 2 | Workflow of EC assembly for single-molecule experiments. A, Procedure for producing the 8-kb non-template-strand (NTS) and template-strand (TS) DNA handles. **B**, Procedure for assembling the mammalian EC using the DNA constructs in (**A**).

https://doi.org/10.1038/s41594-025-01707-1

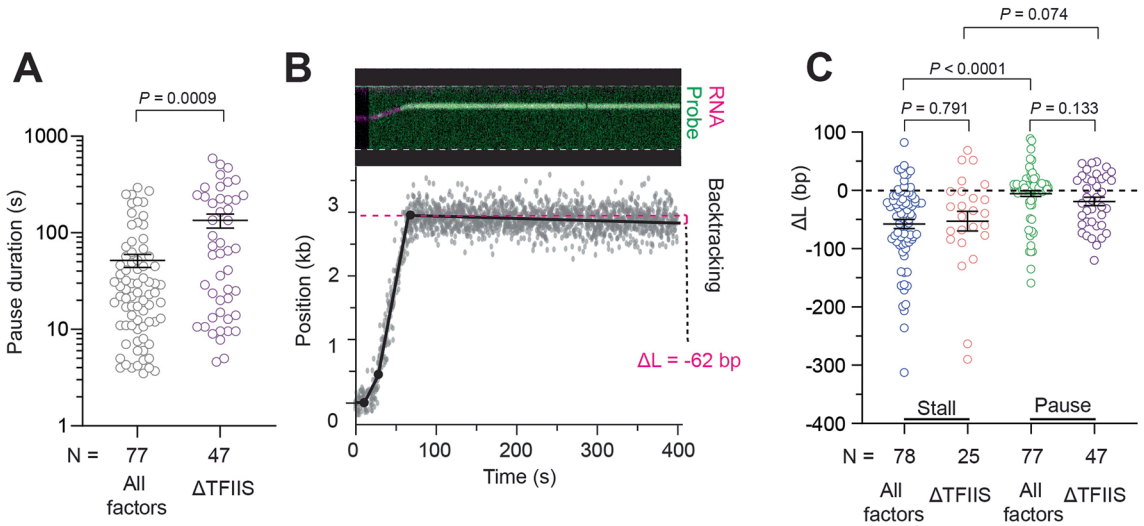


Extended Data Fig. 3 | See next page for caption.

https://doi.org/10.1038/s41594-025-01707-1

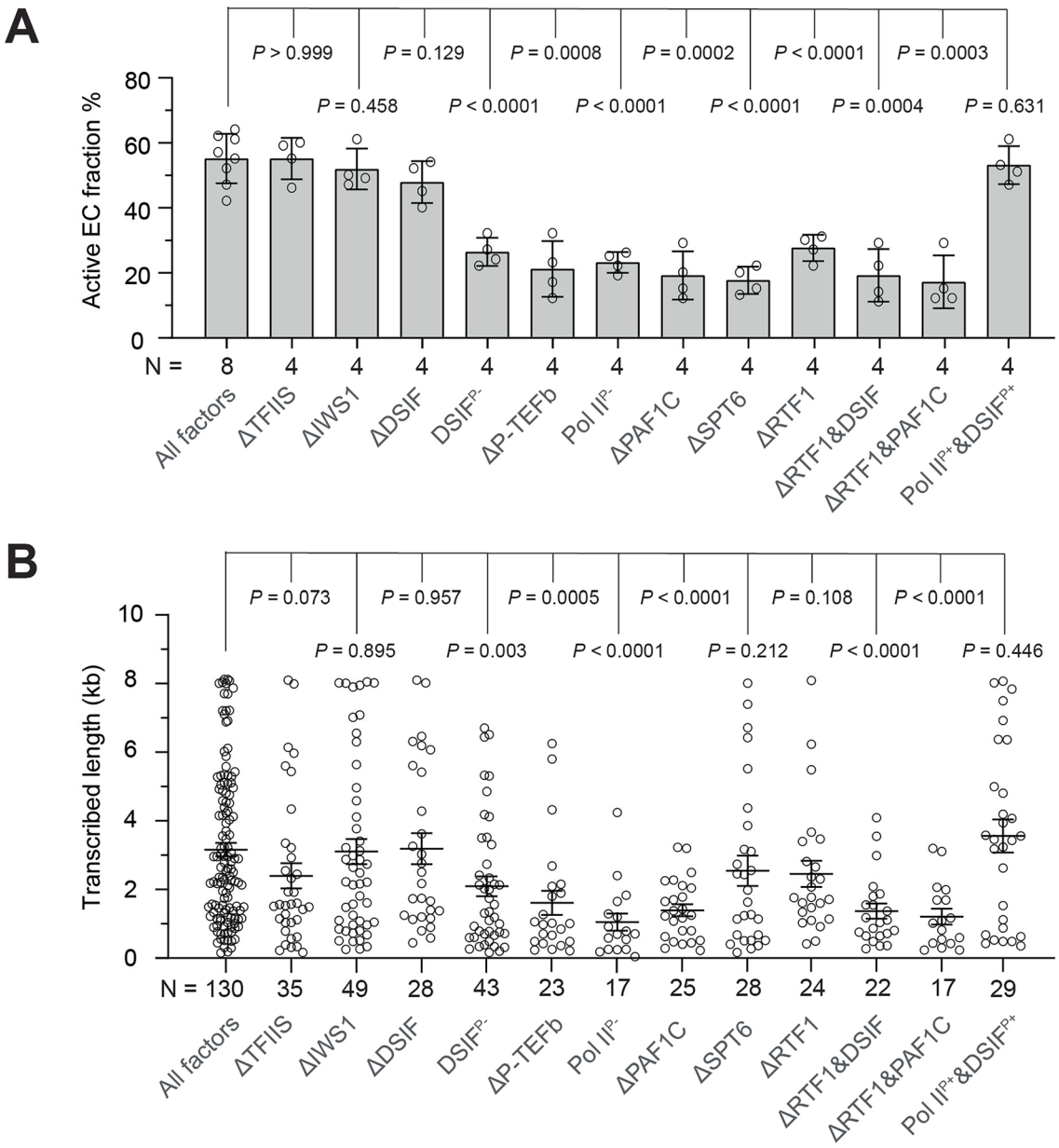
Extended Data Fig. 3 | Example trajectories of mammalian transcription elongation on two different DNA templates. A, Example trajectories on an 8-kb template. **B**, Example trajectories on a 16-kb template. In each example, the kymograph (Top) and the corresponding extracted transcription trajectory (Bottom) are shown. ATTO647N-labeled RNA (magenta) and Cy3-labeled complementary DNA probe (green) signals in the kymograph were used to track the EC progression. The gray dots represent raw data extracted from the kymograph and the black line indicates the fitted trajectory using a Bayesian framework for change point detection. Filled circles denote change points, dividing the trajectory into discrete linear segments. The slopes for each segment are indicated. The gray horizontal bar indicates the end of the template. The full set of elongation factors were present in these examples. **C**, Dot plot showing the slope of active elongation segments from EC trajectories under All-factors condition for the 8-kb and 16-kb templates. Segments are weighted by their transcribed length and each dot corresponds to a 10-bp unit. Bars represent mean ± s.e.m. (sample size is $N_{segment}$). *P* values were from unpaired two-tailed t tests with Welch's correction. **D**, Dot plot showing the distance traveled before stalling by individual ECs on the 8-kb and 16-kb templates. N represents the number of ECs analyzed for each template. Bars represent mean ± s.e.m. *P* values were from unpaired two-tailed t tests with Welch's correction.

https://doi.org/10.1038/s41594-025-01707-1



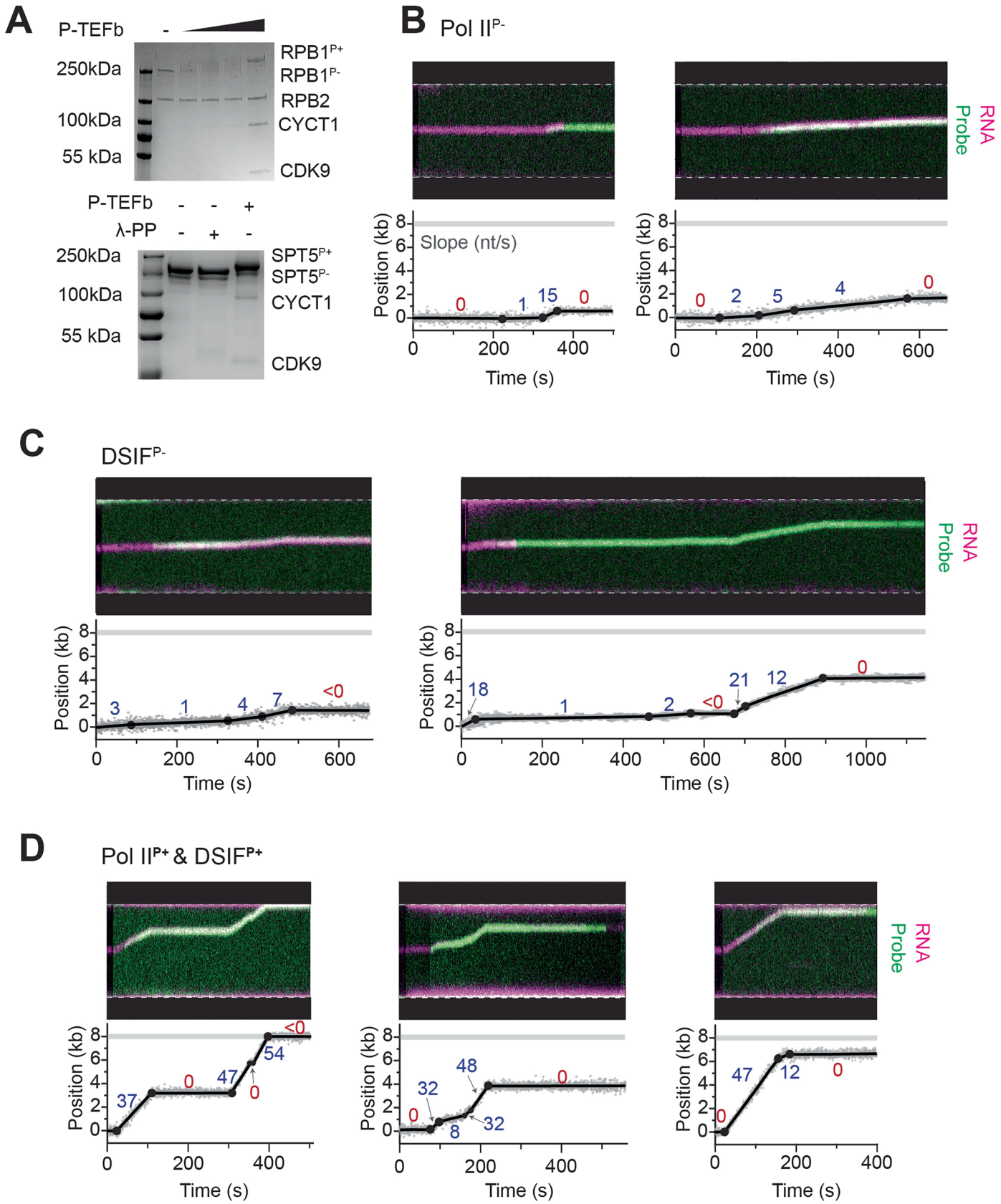
Extended Data Fig. 4 | Analysis of pausing and stalling of the mammalian EC. A, Dot plot showing the duration of pauses detected in the elongation trajectories when TFIIS was included (All-factors) or omitted (ΔTFIIS). **B**, An example kymograph (Top) and extracted transcription trajectory (Bottom) under the ΔTFIIS condition showing extensive EC backtracking during the stalling event. **C**, Dot plot showing the position change of the EC (Δ*L*) during stalling and pausing events when TFIIS was included (All-factors) or omitted (ΔTFIIS). For **A** and **C**, bars represent mean ± s.e.m. (sample size is the number of pauses or stalls analyzed for each condition). *P* values were from unpaired two-tailed t tests with Welch's correction.

https://doi.org/10.1038/s41594-025-01707-1



Extended Data Fig. 5 | Additional data on the activity of mammalian transcription elongation under different conditions. A, Bar plot showing the fraction of active ECs under each condition. An active EC is defined as a trajectory that exhibited detectable elongation activity (that is Cy3-probe signals observed). Each data point corresponds an independent dataset from one-day experiment. Bars represent mean ± s.e.m. (sample size is the number of datasets). *P* values were from unpaired two-tailed t tests with Welch's correction. **B**, Dot plot showing the total transcribed length before stalling for individual ECs. Bars represent mean ± s.e.m. (sample size is the number of active ECs for each condition). *P* values were from unpaired two-tailed t tests with Welch's correction. All comparisons were made against the All-factors condition.

https://doi.org/10.1038/s41594-025-01707-1

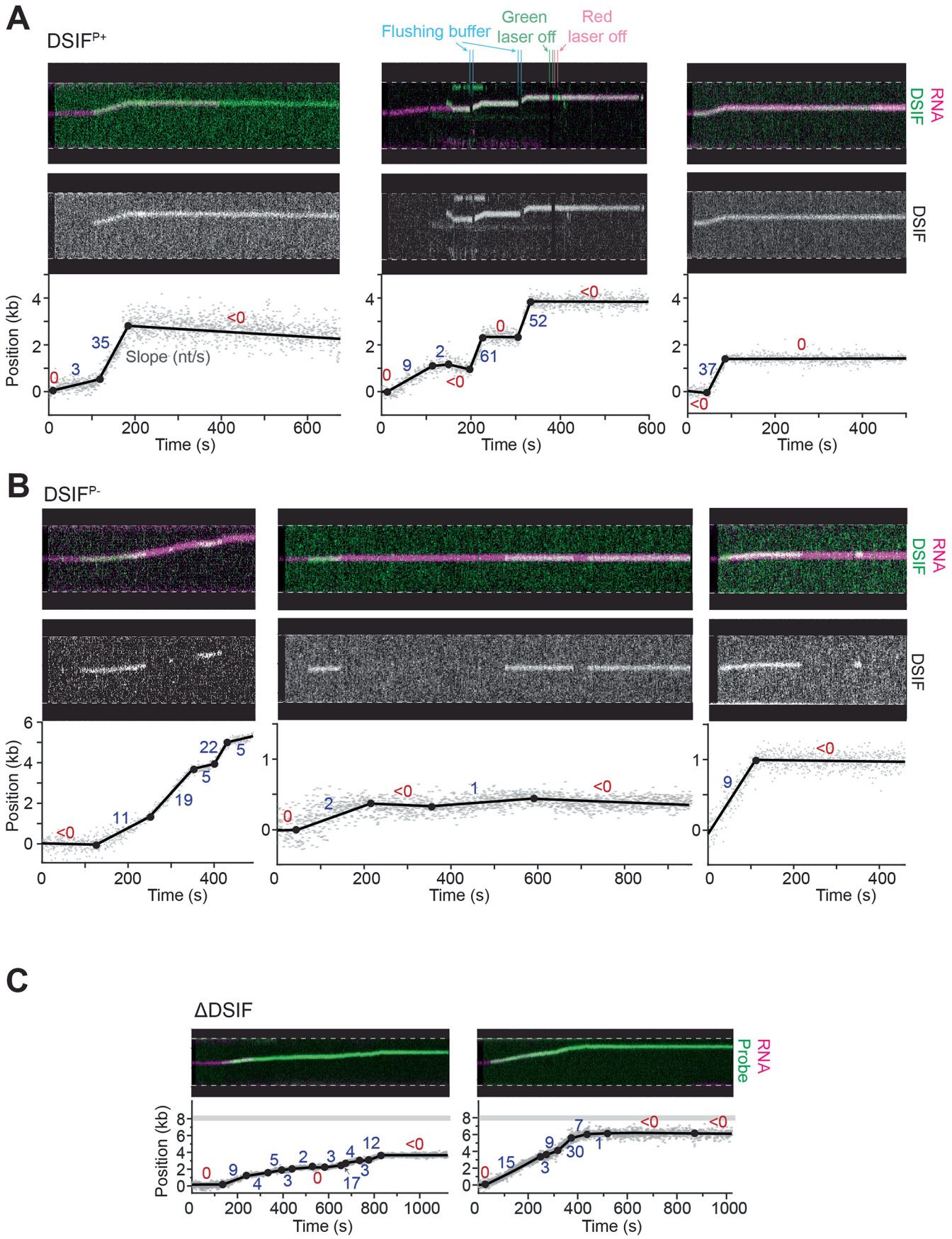


Extended Data Fig. 6 | See next page for caption.

https://doi.org/10.1038/s41594-025-01707-1

Extended Data Fig. 6 | Additional data on the effects of P-TEFb-mediated phosphorylation on EC activity. A, Coomassie-stained Phos-tag SDS-PAGE gel showing phosphorylation of the RPB1 subunit of Pol II (Top) and the SPT5 subunit of DSIF (Bottom) by P-TEFb. Phosphorylated proteins migrated more slowly compared to their unphosphorylated version. λ-phosphatase (λ-PP) treatment confirms that SPT5 was unphosphorylated prior to the addition of P-TEFb. The experiment was repeated three times with similar results. **B**, Representative EC trajectories for the Pol II^{P-} condition. (Top) Kymographs showing the overlay of ATTO647N-labeled RNA (magenta) and Cy3-labeled complementary DNA probe (green) signals which were used to track EC progression. (Bottom) Fitted and segmented trajectories with the slope of each segment indicated. **C**, Representative EC trajectories for the $DSIF^{P-}$ condition. Details are the same as (**B**). **D**, Representative EC trajectories for the Pol II^{P+}&$DSIF^{P+}$ condition. Details are the same as (**B**).

https://doi.org/10.1038/s41594-025-01707-1

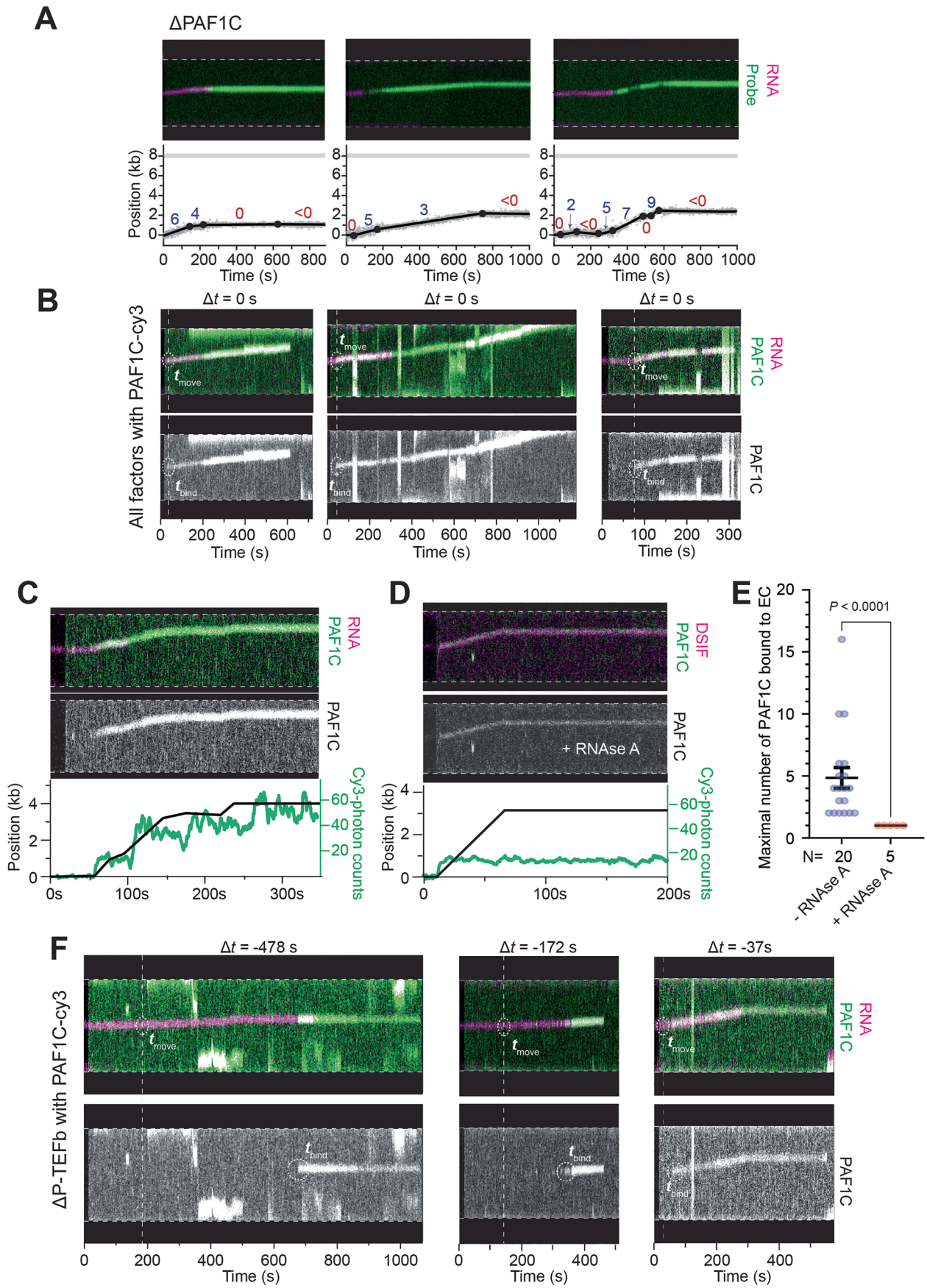


Extended Data Fig. 7 | See next page for caption.

https://doi.org/10.1038/s41594-025-01707-1

Extended Data Fig. 7 | Additional data on DSIF's role in EC activity. **A**, Representative dual-color EC trajectories with Cy3-labeled DSIF and ATTO647N-labeled RNA. (Top) Kymographs showing the overlay of the DSIF (green) and RNA (magenta) signals. (Middle) Same kymographs but with only the DSIF signals shown. (Bottom) Fitted and segmented trajectories with the slope for each segment indicated. DSIF was phosphorylated by P-TEFb in these examples ($DSIF^{P+}$). **B**, Representative EC trajectories with unphosphorylated DSIF ($DSIF^{P-}$). Details are the same as (**A**). **C**, Representative kymographs (Top) and extracted trajectories (Bottom) for ECs under the condition where DSIF was omitted (ΔDSIF). RNA (magenta) and DNA probe (green) signals were used to track EC progression.

https://doi.org/10.1038/s41594-025-01707-1

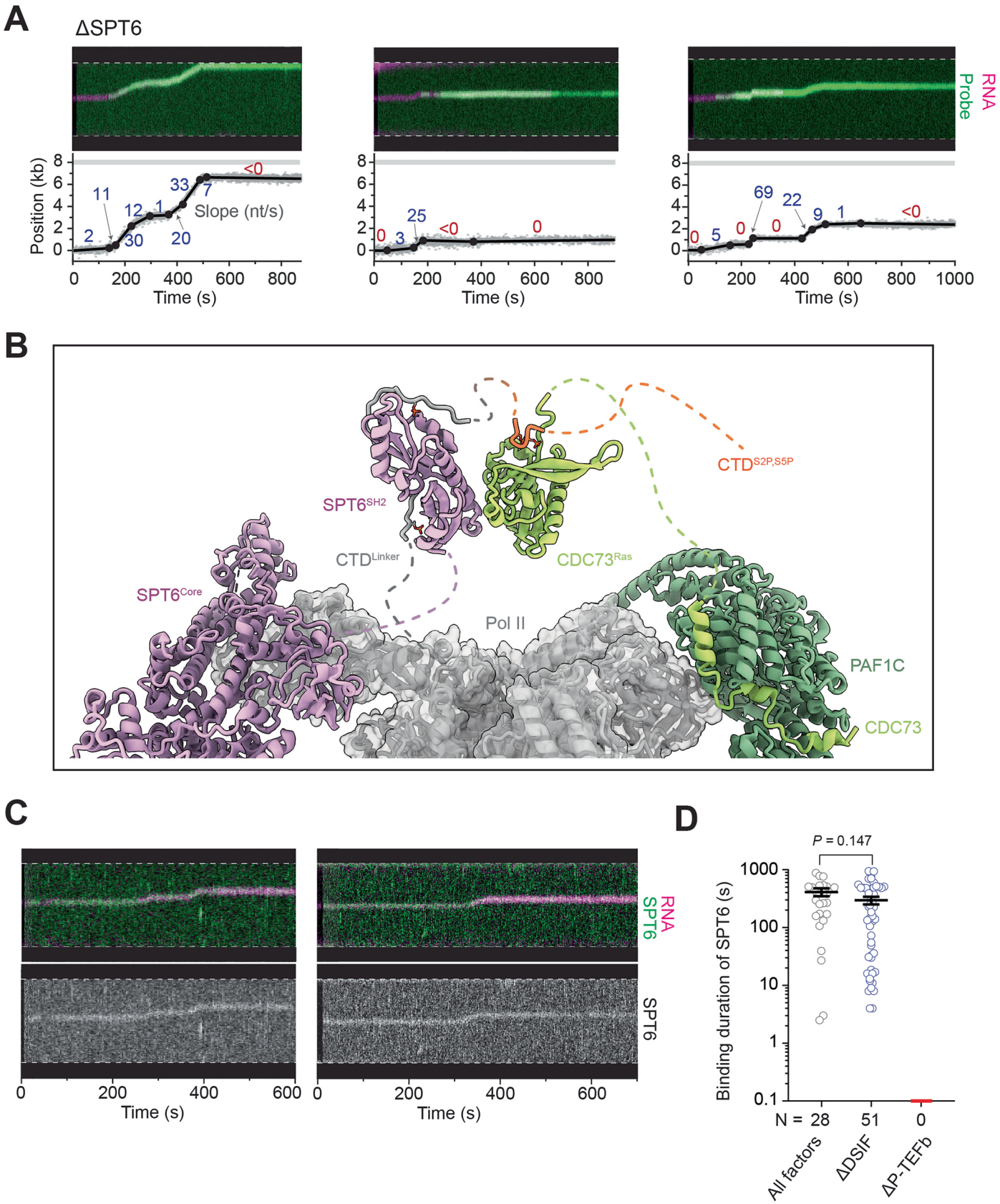


Extended Data Fig. 8 | See next page for caption.

https://doi.org/10.1038/s41594-025-01707-1

Extended Data Fig. 8 | Additional data on PAF1C's role in EC activity. **A**, Representative kymographs (Top) and extracted trajectories (Bottom) for ECs under the condition where PAF1C was omitted (ΔPAF1C). ATTO647N-labeled RNA (magenta) and Cy3-labeled complementary DNA probes (green) were used to track EC progression. In the extracted trajectories, filled circles denote change points at which the trajectory was segmented. The slope for each segment is indicated. **B**, Representative dual-color EC trajectories with Cy3-labeled PAF1C and ATTO647N-labeled RNA in the presence of All-factors. (Top) Kymographs showing the overlay of PAF1C (green) and RNA (magenta) signals. t_{move} indicates the time point when the EC started translocation. (Bottom) The same kymographs but only showing the PAF1C signals. t_{bind} indicates the time point when the initial PAF1C binding to EC occurred. Δt values are reported on top of the kymographs ($\Delta t = t_{move} - t_{bind}$). **C**, An example of an EC trajectory showing multiple PAF1C complexes being recruited during elongation. (Top) Kymograph showing the overlay of PAF1C (green) and RNA (magenta) signals. (Middle) The same kymograph with only the PAF1C signals shown. (Bottom) Extracted EC trajectory (black) and the corresponding Cy3 photon count (green) over time. **D**, An example of an EC trajectory in the presence of RNase A showing one single copy of PAF1C being recruited to the EC. (Top) Kymograph showing the overlay of labeled PAF1C (green) and DSIF (magenta; in lieu of RNA as it was digested by RNase A) signals. (Middle) The same kymograph with only the PAF1C signals shown. (Bottom) Extracted EC trajectory (black) and the corresponding Cy3 photon count (green) over time. **E**, Dot plot showing the maximal number of PAF1C recruited to EC per trajectory in the absence or presence of RNase A based on the PAF1C signal intensity. N denotes the number of EC trajectories analyzed for each condition. Bars represent mean ± s.e.m. *P* values were from unpaired two-tailed t tests with Welch's correction. **F**, Representative dual-color EC trajectories with labeled PAF1C and RNA under the ΔP-TEFb condition. Details are the same as (**B**).

https://doi.org/10.1038/s41594-025-01707-1

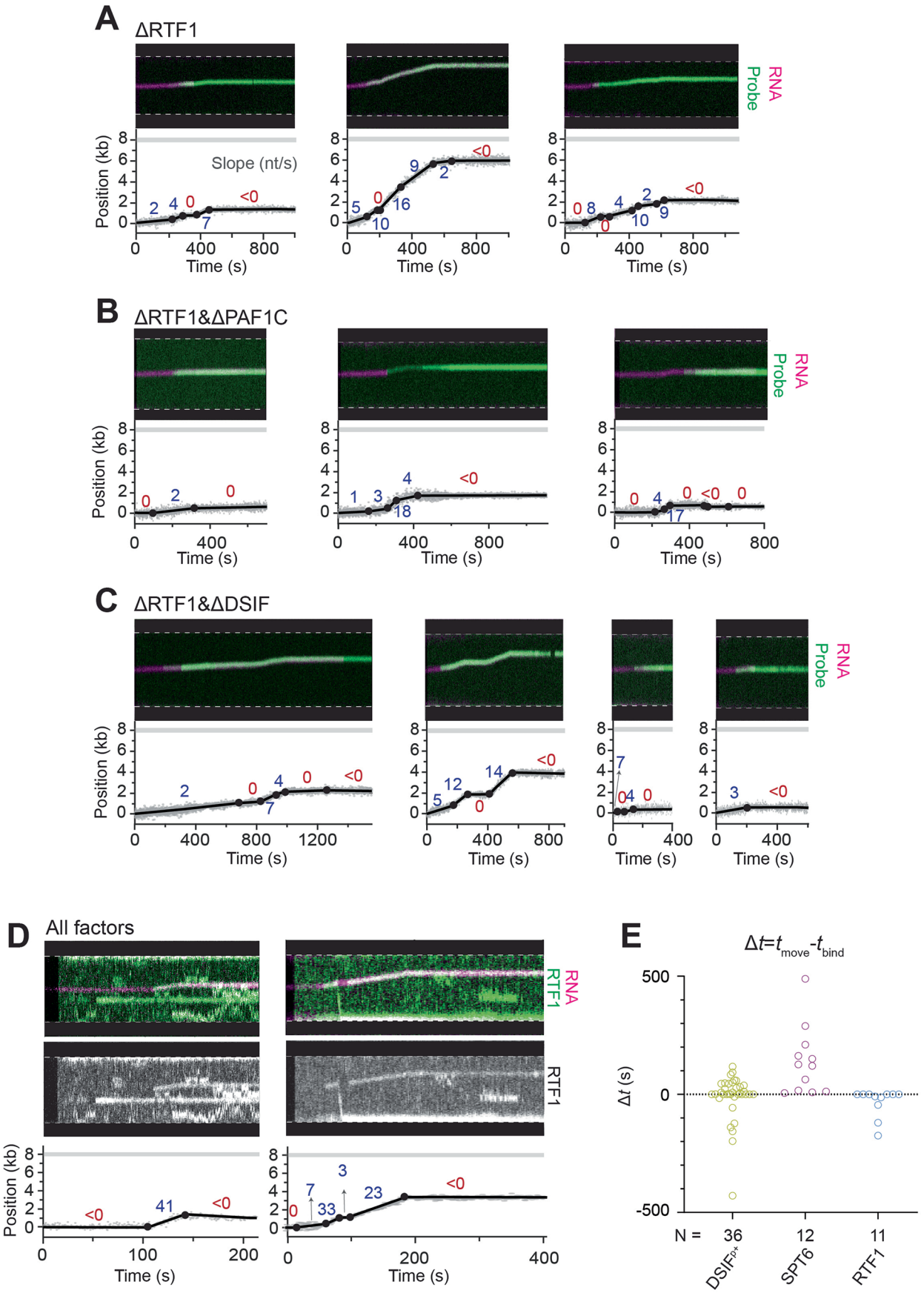


Extended Data Fig. 9 | See next page for caption.

https://doi.org/10.1038/s41594-025-01707-1

Extended Data Fig. 9 | Additional data on SPT6's role in EC activity. **A**, Representative kymographs (Top) and extracted trajectories (Bottom) for ECs under the condition where SPT6 was omitted (ΔSPT6). ATTO647N-labeled RNA (magenta) and Cy3-labeled complementary DNA probes (green) were used to track EC progression. The slope for each segment is indicated. **B**, Model for PAF1C-SPT6-Pol II interaction. Pol II, PAF1C and SPT6 chains are shown in gray, green and purple, respectively. Unresolved regions are represented by dashed lines. The Spt6:Cdc73 complex based on a yeast structure[46] is superimposed on a mammalian EC structure with the SPT6 SH2 domain interacting with the RPB1 CTD linker region[10]. Cdc73 is replaced by our AlphaFold 3 prediction of the CDC73 Ras-like domain ($CDC73^{Ras}$) bound to a CTD-mimicking heptapeptide phosphorylated at Ser2 and Ser5 (YpSPTpSPS). The predicted CTD-binding pocket in $CDC73^{Ras}$ is consistent with a previous biochemical study in the yeast system[42]. Any of the CTD repeats can in principle occupy this pocket. **C**, Representative dual-color EC trajectories with Cy3-labeled SPT6 and ATTO647N-labeled RNA in the presence of All-factors. (Top) Kymographs showing the overlay of SPT6 (green) and RNA (magenta) signals. (Bottom) The same kymographs but only showing the SPT6 signals. **D**, Dot plot showing the residence time of SPT6 on EC under All-factors, ΔDSIF, and ΔP-TEFb conditions. Bars represent mean ± s.e.m. (sample size is the number of binding events). *P* values were from unpaired two-tailed t tests with Welch's correction. Source Data.

https://doi.org/10.1038/s41594-025-01707-1



Extended Data Fig. 10 | See next page for caption.

https://doi.org/10.1038/s41594-025-01707-1

Extended Data Fig. 10 | Additional data on RTF1's role in EC activity. **A**, Representative kymographs (Top) and extracted trajectories (Bottom) for ECs under the condition where RTF1 was omitted (ΔRTF1). ATTO647N-labeled RNA (magenta) and Cy3-labeled complementary DNA probes (green) were used to track EC progression. Filled circles in the extracted trajectories denote changepoints, dividing the trajectory into discrete linear segments. The slopes for each segment are indicated. **B**, Representative kymographs (Top) and extracted trajectories (Bottom) for ECs under the condition where both RTF1 and PAF1C were omitted (ΔRTF1&ΔPAF1C). Details are the same as (**A**). **C**, Representative kymographs (Top) and extracted trajectories (Bottom) for ECs under the condition where both RTF1 and DSIF were omitted (ΔRTF1&ΔDSIF). Details are the same as (**A**). **D**, Representative dual-color EC trajectories with Cy3-labeled RTF1 and ATTO647N-labeled RNA in the presence of All-factors. (Top) Kymographs showing the overlay of RTF1 (green) and RNA (magenta) signals. (Middle) Same kymographs but only showing the RTF1 signals. (Bottom) Fitted and segmented trajectories with the slope of each segment indicated. **E**, Dot plot showing the Δt value ($\Delta t = t_{move} - t_{bind}$) distribution for a given fluorescently labeled elongation factor ($DSIF^{P+}$, SPT6, or RTF1) in the presence of All-factors. $\Delta t > 0$ indicates factor binding preceded EC activation; $\Delta t = 0$ indicates factor binding coincided with EC activation; whereas $\Delta t < 0$ indicates factor binding occurred after EC started moving. N denotes the number of EC trajectories analyzed for each factor.

nature portfolio

Corresponding author(s): Shixin Liu

Last updated by author(s): Sep 26, 2025

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	Single-molecule data was acquired using LUMICKS Bluelake software version 1.6.16.
Data analysis	Kymographs were processed and analyzed using a custom script (https://harbor.lumicks.com/single-script/c5b103a4-0804-4b06-95d3-20a08d65768f) which incorporates tools from the lumicks.pylake Python library and other Python modules (Numpy, Matplotlib, Pandas) to generate tracked lines using the kymotracker greedy algorithm. MATLAB (R2023b, MathWorks) was used for all computational analyses. Bayesian changepoint estimation was conducted using Rbeast (https://github.com/zhaokg/Rbeast), with parameters optimized for time-series segmentation. The software for Bayesian changepoint estimation and piecewise linear segments fitting was uploaded to Github (https://github.com/yorquant/Fitting-Piecewise-Linear-Continuous/tree/main).

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

Data

Policy information about availability of data

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

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Source data are provided with this paper. All kymographs used for analysis have been deposited as datasets in Zenodo (https://doi.org/10.5281/zenodo.17193150).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

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

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	For single-molecule experiments, 15-130 kymographs were collected for each condition as standard practice in the field. The number of measurements was chosen to ensure sufficient statistics for each sample.
Data exclusions	For single-molecule experiments, tethers that display abnormal force-extension behaviors (tether broke accidentally or large protein aggregate stuck to the beads) were discarded.
Replication	For single-molecule experiments, 15-130 kymographs were collected for each condition with comparable behaviors that were further quantified and described. Attempts at replication were successful.
Randomization	Randomization was not applicable as there was no grouping in this study.
Blinding	Blinding was not applicable to this study since the results are not subjective examinations by the experimenter.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	The Expi293F cell line (purchased from Invitrogen) was used for the purification of some proteins.
Authentication	Authentication of this commercial cell line is routinely performed by the manufacturer.
Mycoplasma contamination	Mycoplasma contamination for this commercial cell line is routinely tested by the manufacturer.
Commonly misidentified lines (See ICLAC register)	No misidentified cell lines were used.

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A