

# Real-time observation of the initiation of RNA polymerase II transcription

Furqan M. Fazal<sup>1\*</sup>, Cong A. Meng<sup>2\*</sup>, Kenji Murakami<sup>3†\*</sup>, Roger D. Kornberg<sup>3</sup> & Steven M. Block<sup>1,4</sup>

**Biochemical and structural studies have shown that the initiation of RNA polymerase II transcription proceeds in the following stages: assembly of the polymerase with general transcription factors and promoter DNA in a ‘closed’ preinitiation complex (PIC)<sup>1,2</sup>; unwinding of about 15 base pairs of the promoter DNA to form an ‘open’ complex<sup>3,4</sup>; scanning downstream to a transcription start site; synthesis of a short transcript, thought to be about 10 nucleotides long; and promoter escape. Here we have assembled a 32-protein, 1.5-megadalton PIC<sup>5</sup> derived from *Saccharomyces cerevisiae*, and observe subsequent initiation processes in real time with optical tweezers<sup>6</sup>. Contrary to expectation, scanning driven by the transcription factor IIH<sup>7–12</sup> involved the rapid opening of an extended transcription bubble, averaging 85 base pairs, accompanied by the synthesis of a transcript up to the entire length of the extended bubble, followed by promoter escape. PICs that failed to achieve promoter escape nevertheless formed open complexes and extended bubbles, which collapsed back to closed or open complexes, resulting in repeated futile scanning.**

Optical tweezers have been used in studies of transcript elongation by RNA polymerase II (Pol II)<sup>13–15</sup> with the use of a ‘dumbbell’ configuration, consisting of two beads held in separate optical traps, connected by a segment of DNA. One bead was directly attached to Pol II, and the other was attached to the opposite end of the template DNA, minimally around 3 kilobases (kb), for traversing the distance between the traps. For the study of transcription initiation, we adapted a similar approach to the PIC. Pol II, biotinylated for attachment to one bead, was assembled together with transcription factors on DNA, and end-labelled with digoxigenin for attachment to the other bead. The transcription factors consisted of six general transcription factors (GTFs) including TATA-binding protein (TBP), transcription factor IIB (TFIIB), TFIIE, TFIIF, TFIIH and TFIIA and Sub1 (yeast homologue of human PC4), which is thought to stabilize the PIC<sup>16</sup> (Extended Data Fig. 1). The DNA contained the *SNR20* (also known as *LSR1*) promoter fused to an additional 2.7 kb length of DNA, sufficient to separate the beads by roughly the wavelength of light. The *SNR20* promoter bore a mutation resulting in one, rather than many, transcription start sites (TSSs). Two versions of the promoter were used: the otherwise wild-type promoter with the single TSS located 91 base pairs (bp) downstream of the TATA box (referred to as *SNR20\** long), and a deleted version, in which the TSS was situated 31 bp downstream (*SNR20\** short), a distance characteristic of metazoan transcription. Both versions of the promoter have been characterized in bulk transcription experiments<sup>17</sup>. The PIC, assembled without the peripheral component TFIK<sup>18,19</sup>, was mixed with a 25-fold molar excess of PIC without the additional 2.7 kb DNA, to achieve an overall PIC concentration sufficient to avoid dissociation. A twofold excess of TFIK was added, and dumbbells were formed by reaction of the PIC mixture with anti-digoxigenin-coated and avidin-coated beads (Extended Data Fig. 2).

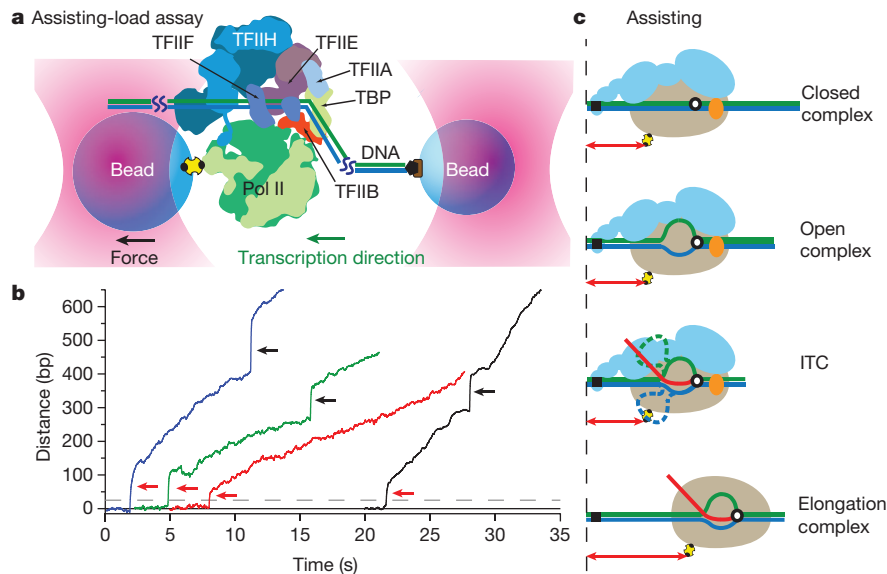
In a dumbbell carrying digoxigenin on the upstream end of the DNA, the tension exerted by the optical trap tends to pull the polymerase downstream, in the same direction as transcription, resulting in an ‘assisting-load’ assay (Fig. 1a). Transcription was initiated by the addition of saturating concentrations of all four ribonucleoside triphosphates (rNTPs). Force was maintained during measurements by the use of an optical force clamp, as the location of the polymerase on DNA was tracked with sub-nanometre-level precision. Transcription was signalled by movement of the polymerase (Fig. 1b, *SNR20\** short promoter) at  $29 \pm 3 \text{ bp s}^{-1}$  ( $n = 10$ , mean  $\pm$  s.e.m.), consistent with elongation rates observed in previous assays of transcription under similar assisting loads<sup>14,15</sup>. To confirm the identification of the moving molecules as transcription elongation complexes, we raised the force instantaneously to a value (10–15 pN) that, in our experience, can only be sustained by a stable elongation complex (Fig. 1b, black arrows). Only 2–3% of dumbbells gave rise to transcription elongation complexes, whereas in biochemical assays, about 18% of PICs gave rise to runoff transcripts (Extended Data Fig. 3a). The lower efficiency of initiation in the single-molecule system was attributable to the much lower protein concentrations used ( $<1 \text{ nM}$ , at least tenfold lower than biochemical assays; Extended Data Fig. 3b).

The onset of polymerase movement at a rate characteristic of transcript elongation was preceded by an almost instantaneous jump (Fig. 1b, red arrows), occurring around  $15 \pm 2 \text{ s}$  ( $n = 10$ ) after the addition of rNTPs. No such movement was observed in the absence of rNTPs. An interpretation consistent with all other available information is that the polymerase draws downstream DNA into the active centre region to form an extended unwound region, or transcription bubble, and then lurches forward after DNA rewinding and bubble collapse (Fig. 1c). Because one bead is attached to the upstream end of the DNA and the other bead to the polymerase, there is no change in the distance between them when DNA is drawn in from the downstream side. Only once the polymerase is released from its point of attachment at the upstream edge of the bubble (promoter escape), and DNA rewinds to collapse the bubble, does the distance between the beads change and lengthen (Fig. 1c). The size of the jump at the transition to a transcription elongation complex was  $70 \pm 13 \text{ bp}$  ( $n = 9$ , mean  $\pm$  s.e.m.), with a minimum of 32–34 bp and a maximum of about 140 bp.

The jump after promoter release and the corresponding transition to a stable elongation complex are notable in two further respects. First, the bubble does not collapse completely at the jump, because about 15 bp remain unwound in the Pol II active centre as a transcription bubble from the time of open complex formation until the end of transcript elongation<sup>20</sup>. Therefore, the entire length of the unwound region in this initial transcribing complex (ITC) is, on average, approximately 85 bp ( $70 + 15 \text{ bp}$ ). Second, because this experiment was performed with the *SNR20\** short promoter (in which the TSS is located 31 bp downstream of the TATA box), transcription was initiated within the open complex, and the nascent transcript extended

<sup>1</sup>Department of Applied Physics, Stanford University, Stanford, California 94305, USA. <sup>2</sup>Department of Chemistry, Stanford University, Stanford, California 94305, USA. <sup>3</sup>Department of Structural Biology, Stanford University, Stanford, California 94305, USA. <sup>4</sup>Department of Biology, Stanford University, Stanford, California 94305, USA. <sup>†</sup>Present address: Department of Biochemistry and Molecular Biophysics at the Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.

\*These authors contributed equally to this work.



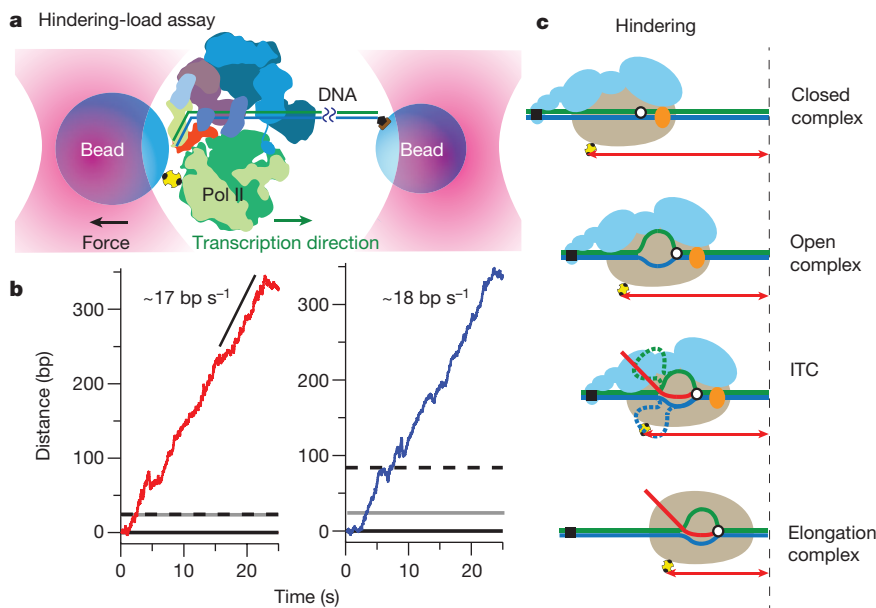
**Figure 1 | Transcription initiation in assisting-load assay.** **a**, A dumbbell tether (not to scale; subunit Sub1 not shown) is formed between beads (blue) held in separate optical tweezers (pink), with one attached to Pol II (green) in the PIC via an avidin–biotin linkage (yellow, black), and the other to upstream DNA via a digoxigenin linkage (brown, black). As transcription proceeds (green arrow indicates direction), the tether extension increases. **b**, Representative records of Pol II elongation (dashed grey line denotes the TSS location) after promoter escape (red arrows), with the applied force often stepped up after  $\sim 10$  s to confirm elongation (black arrows; the associated

vertical discontinuity is due to tether stretch). **c**, Ssl2 subunit in TFIIH (orange) unwinds the template (blue) and non-template (green) strands of DNA around the active site (white circle, black outline) of Pol II (beige), and creates a transcription bubble (open complex formation). RNA synthesis while still bound to the promoter results in DNA crumpling at the upstream edge of transcription bubble (ITC), which re-anneals after Pol II enters productive elongation (elongation complex). Distances measured by the assays are indicated (double-headed red arrows). Not to scale.

to the downstream end of the unwound region before the jump (Fig. 1c). Evidently, a transcript averaging 70 nucleotides, and as long as around 140 nucleotides, is synthesized before promoter clearance and the transition to a stable elongation complex.

To observe the process occurring before the jump, presumed to involve the drawing of downstream DNA into the Pol II active centre, we moved the point of attachment of the bead to DNA from the upstream to the downstream end of the template. In this configuration, external forces applied to polymerase tend to pull it upstream, opposite to the direction of transcription, resulting in a ‘hindering-load’ assay (Fig. 2a). Fewer dumbbells ( $<2\%$ ) yielded transcription elongation complexes in this assay, consistent with previous studies showing that

hindering loads reduce polymerase processivity. Dumbbells that elongated did so at  $17\text{--}18\text{ bp s}^{-1}$  (Fig. 2b), consistent with previous measurements of Pol II transcription under hindering loads<sup>14,15</sup>. In contrast to the assisting-load assay, there was no jump at the transition to a stable elongation complex, but rather a gradual distance change (Fig. 2b). The distance change was the same size as the jump in the assisting-load assay, and was observed for both forms of the *SNR20\** promoter. In the case of the *SNR20\** long promoter, which initiates transcription downstream (Fig. 2, dashed line), the distance change reflects open complex formation and scanning to the TSS; in the case of the *SNR20\** short promoter, which initiates transcription in the open complex (dashed line, Fig. 2b), the distance change reflects open

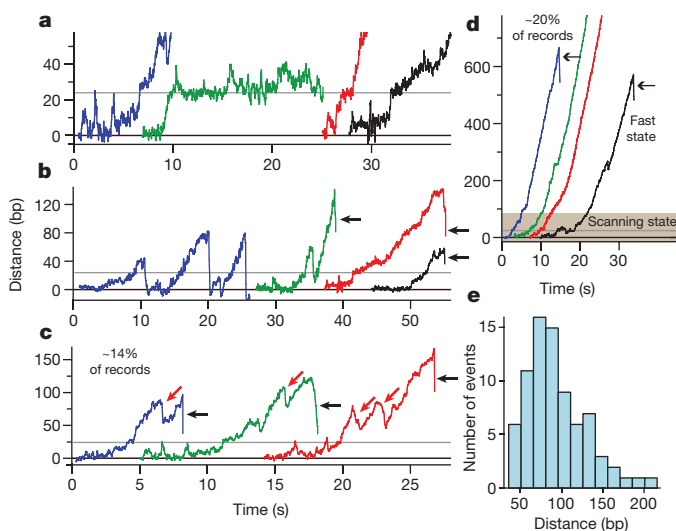


**Figure 2 | Transcription initiation in hindering-load assay.** **a**, By attaching one bead to downstream DNA, a hindering-load assay was developed (not to scale, Sub1 not shown). Scanning and subsequent transcription events (green arrow indicates transcription direction) resulted in tether-extension decrease. **b**, Records illustrating Pol II escape and elongation, with a velocity of  $\sim 17\text{--}18\text{ bp s}^{-1}$ , collected on *SNR20\** short (red, left panel) and *SNR20\** long (blue, right panel) in the presence of rNTPs. The dashed black line denotes the TSS at +1; the solid grey line marks position of the predicted  $\sim 24\text{ bp}$  distance change after open complex formation. **c**, Distances measured by the assays are indicated (colour scheme same as Fig. 1c).

complex formation, extension of the bubble, and transcription in the ITC (Fig. 2c).

In the single-molecule system, it was possible to investigate not only PICs that initiated transcription, but also those that failed to do so. Approximately 20% of dumbbells showed movement downstream without the initiation of transcription (Fig. 3 and Extended Data Fig. 4). The downstream movement began with an initial distance change of about 24 bp, which was often punctuated, at the temporal resolution of our assay ( $\sim 0.1$  s), by brief ( $< 2$  s) pauses (Fig. 3a, b). After the initial 24 bp, movement continued to a maximum of about 150–200 bp downstream (Fig. 3e), until either the bubble collapsed back to a distance of 24 bp or 0 bp, or the PIC dissociated, as evidenced by rupture of the dumbbell (Fig. 3b, black arrows). Bubble collapse was often followed by a repetition of the downstream movement.

The downstream movement was processive, and was observed under all three conditions examined: *SNR20\** short with rNTPs ( $n = 40$ ) (Fig. 3a–d); *SNR20\** long with rNTPs ( $n = 19$ ) (Extended Data Fig. 4a); and *SNR20\** long with dATP ( $n = 15$ ) (Extended Data Fig. 4b). No movement was observed in the absence of rNTPs or dATP. There were no significant differences in either processivity or velocity in the three conditions. Combining these data yielded a pause-free velocity for downstream movement of  $36 \pm 1$  bp  $s^{-1}$  ( $n = 24$ ). Because the velocity was unchanged when only dATP was present (and no rNTPs), it must have been produced by TFIIH activity, and not by polymerase. The extent of the downstream movement was  $94 \pm 36$  bp (mean  $\pm$  s.d.) (Fig. 3e). TFIIK, which contains a kinase responsible for phosphorylation of the carboxy-terminal domain of Pol II (refs 18, 19), could be omitted without effect. There was no change in the absence of TFIIK in either the distance ( $92 \pm 33$  bp,  $n = 34$ ) or the velocity ( $36 \pm 2$  bp  $s^{-1}$ ,  $n = 4$ ) of downstream movement. In about 20% ( $n = 15$  out of 74) of dumbbells that displayed TFIIH activity but failed to initiate transcription, there was a transition to a ‘fast state’, characterized by a velocity of  $61 \pm 2$  bp  $s^{-1}$  and downstream movement through hundreds of base pairs (Fig. 3d). The transition to the fast state was irreversible, and must reflect action of the TFIIH helicase subject to little or no restraint by other GTFs or Pol II.



**Figure 3 | Records of TFIIH motion for the *SNR20\** short construct with rNTPs present in hindering-load assay.** **a**, Initial transition from the closed (0 bp) to open ( $\sim 24$  bp predicted distance change, grey line) complex. **b**, Scanning behaviour, with occasional bubble collapse to the closed complex (blue record) or open complex (green record). **c**, Infrequent slips in the records ( $n = 10$  of 74) were observed (red arrows). **d**, Occasional irreversible transition from scanning (shaded region) to a highly processive fast state, occurring at a distance of  $130 \pm 21$  bp ( $n = 9$ , mean  $\pm$  s.e.m.). In all records, black arrows mark tether breakage, probably due to PIC dissociation. **e**, Histogram of TFIIH processivity, with a peak between 40 and 140 bp ( $n = 78$ ).

Biochemical evidence for movement driven by the TFIIH helicase was obtained by exonuclease III footprinting of the PIC (Extended Data Fig. 5). Extended regions of unwound DNA were previously revealed by  $KMnO_4$  reactivity of yeast and *Drosophila* promoters in studies of transcriptional activity *in vivo*<sup>21–23</sup>. The regions were similar in all cases, extending from about 20–60 bp downstream of the TATA box, with the TSSs of the *Drosophila* promoters near the upstream edge of the unwound region nearest the TATA box, and the TSSs of the yeast promoters near the downstream edge furthest from the TATA box. It had been thought that transcribing polymerases with 15-bp bubbles, at different locations on individual promoters and then revealed collectively in the  $KMnO_4$  analysis, gave the appearance of an extended bubble. Our examination of single molecules suggests instead that an extensive unwound region is a characteristic of every individual promoter, rather than some collective property. We obtained similar results for a yeast promoter in which the TSS was located near the downstream edge of the unwound region, and for the same promoter in which the TSS was moved to the upstream edge, as in *Drosophila* and other metazoans.

The formation of the unwound region is not a consequence of transcription, but instead of TFIIH action, because it occurs in the presence of dATP without rNTPs, and because it is observed even when the TSS is located at the downstream end of the unwound region. TFIIH must act continuously to maintain the unwound region, consistent with previous biochemical studies showing a requirement of TFIIH to prevent premature arrest of ITCs<sup>24,25</sup>. It is not known what determines the length of the unwound region, nor where the unwound DNA resides in the complex. In the case of bacterial polymerase transcription, approximately 10 bp of DNA drawn into the active centre before the transition to elongation are thought to be accommodated by a ‘scrunching’ mechanism<sup>26,27</sup>. The possibility of scrunching in eukaryotic transcription has previously been considered<sup>28</sup>, but no evidence obtained. Because of rotation in the direction of unwinding by TFIIH, there is unlikely to be associated torsional strain, as presumed to occur in the bacterial system. The location and conformational state of the approximately 85 bp of DNA unwound in the Pol II PIC thus remain open questions.

Although most PICs ( $\sim 80\%$  of PICs in biochemical assay conditions, and 97–98% of PICs under single-molecule assay conditions) fail to yield stable elongation complexes, they are not inert. About 20% of the dumbbells showed downstream movements of polymerase along DNA in the hindering-load assay. There was often an initial movement of  $24 \pm 2$  bp (mean  $\pm$  s.d.), which we attribute to the formation of an open complex, on the basis of previous studies of the PIC<sup>29</sup> (see Methods).

After open complex formation, downstream movement continued for a total distance of 94 bp, on average, with bubble collapse back to the open or closed complexes and repetition of downstream movement, before either final dissociation of the PIC or rupture of the dumbbell (Extended Data Fig. 6). The movement of 94 bp in the hindering-load assay is noteworthy for two reasons. First, it is in excellent agreement with the results of the assisting-load assays, in which a jump of about 70 bp was observed before the onset of transcription elongation. This jump is attributed to collapse of an extended bubble, leaving the original open complex in place, and  $70 + 24$  bp (open complex) = 94 bp. Second, the distance of downstream movement in the hindering-load assay ranged from about 30 to 150 bp, that is, about 37–157 bp from the TATA box, similar to the distribution of TSSs in yeast, which are located 40–120 bp downstream from the TATA box. Therefore, downstream movements in the hindering-load assay may be attributed to TSS scanning, which precedes the onset of transcription elongation, as observed in the assisting-load assay, and the initiation of transcription in yeast *in vivo*<sup>24,25</sup>.

It is commonly noted that TSSs for yeast promoters are spread over a wide region, rather than concentrated near the TATA box, as in metazoans. Nevertheless, as discussed above, our evidence for an extended bubble in the yeast PIC corresponds well with extended bubbles mapped

by  $\text{KMnO}_4$  reactivity in *Drosophila*<sup>21</sup>. Moreover, TATA-less promoters, which predominate in metazoans as well as in yeast, have several TSSs spread over regions of 50–100 bp in human cells<sup>30</sup>. Our findings from single-molecule studies of the yeast PIC are therefore likely to hold true for other eukaryotes as well, including metazoans.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 February; accepted 3 July 2015.

Published online 2 September 2015.

- Conaway, R. C. & Conaway, J. W. General initiation factors for RNA polymerase II. *Annu. Rev. Biochem.* **62**, 161–190 (1993).
- Kornberg, R. D. The molecular basis of eukaryotic transcription. *Proc. Natl Acad. Sci. USA* **104**, 12955–12961 (2007).
- Holstege, F. C., Fiedler, U. & Timmers, H. T. Three transitions in the RNA polymerase II transcription complex during initiation. *EMBO J.* **16**, 7468–7480 (1997).
- Pal, M., Ponticelli, A. S. & Luse, D. S. The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. *Mol. Cell* **19**, 101–110 (2005).
- Murakami, K. *et al.* Formation and fate of a complete, 31-protein, RNA polymerase II transcription initiation complex. *J. Biol. Chem.* **288**, 6325–6332 (2013).
- Fazal, F. M. & Block, S. M. Optical tweezers study life under tension. *Nature Photon.* **5**, 318–321 (2011).
- Sawadogo, M. & Roeder, R. G. Energy requirement for specific transcription initiation by the human RNA polymerase II system. *J. Biol. Chem.* **259**, 5321–5326 (1984).
- Schaeffer, L. *et al.* DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor. *Science* **260**, 58–63 (1993).
- Svejstrup, J. Q. *et al.* Different forms of TFIIF for transcription and DNA repair: holo-TFIIF and a nucleotide excision repairosome. *Cell* **80**, 21–28 (1995).
- Dvir, A. *et al.* A role for ATP and TFIIF in activation of the RNA polymerase II preinitiation complex prior to transcription initiation. *J. Biol. Chem.* **271**, 7245–7248 (1996).
- Kim, T. K., Ebright, R. H. & Reinberg, D. Mechanism of ATP-dependent promoter melting by transcription factor IIH. *Science* **288**, 1418–1421 (2000).
- Grünberg, S., Warfield, L. & Hahn, S. Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nature Struct. Mol. Biol.* **19**, 788–796 (2012).
- Galbur, E. A. *et al.* Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature* **446**, 820–823 (2007).
- Larson, M. H. *et al.* Trigger loop dynamics mediate the balance between the transcriptional fidelity and speed of RNA polymerase II. *Proc. Natl Acad. Sci. USA* **109**, 6555–6560 (2012).
- Schweikhard, V. *et al.* Transcription factors TFIIF and TFIIS promote transcript elongation by RNA polymerase II by synergistic and independent mechanisms. *Proc. Natl Acad. Sci. USA* **111**, 6642–6647 (2014).
- Sikorski, T. W. *et al.* Sub1 and RPA associate with RNA polymerase II at different stages of transcription. *Mol. Cell* **44**, 397–409 (2011).
- Murakami, K. *et al.* Uncoupling promoter opening from start-site scanning. *Mol. Cell* **59**, 133–138 (2015).
- Feaver, W. J., Svejstrup, J. Q., Henry, N. L. & Kornberg, R. D. Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIF/TFIIK. *Cell* **79**, 1103–1109 (1994).
- Roy, R. *et al.* The MO15 cell cycle kinase is associated with the TFIIF transcription-DNA repair factor. *Cell* **79**, 1093–1101 (1994).
- Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **292**, 1876–1882 (2001).
- Giardina, C., Perez-Riba, M. & Lis, J. T. Promoter melting and TFIID complexes on *Drosophila* genes *in vivo*. *Genes Dev.* **6**, 2190–2200 (1992).
- Giardina, C. & Lis, J. T. DNA melting on yeast RNA polymerase II promoters. *Science* **261**, 759–762 (1993).
- Giardina, C. & Lis, J. T. Dynamic protein–DNA architecture of a yeast heat shock promoter. *Mol. Cell Biol.* **15**, 2737–2744 (1995).
- Dvir, A., Conaway, R. C. & Conaway, J. W. A role for TFIIF in controlling the activity of early RNA polymerase II elongation complexes. *Proc. Natl Acad. Sci. USA* **94**, 9006–9010 (1997).
- Spangler, L., Wang, X., Conaway, J. W., Conaway, R. C. & Dvir, A. TFIIF action in transcription initiation and promoter escape requires distinct regions of downstream promoter DNA. *Proc. Natl Acad. Sci. USA* **98**, 5544–5549 (2001).
- Kapanidis, A. N. *et al.* Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* **314**, 1144–1147 (2006).
- Revyakin, A., Liu, C., Ebright, R. H. & Strick, T. R. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* **314**, 1139–1143 (2006).
- Miller, G. & Hahn, S. A DNA-tethered cleavage probe reveals the path for promoter DNA in the yeast preinitiation complex. *Nature Struct. Mol. Biol.* **13**, 603–610 (2006).
- Murakami, K. *et al.* Architecture of an RNA polymerase II transcription pre-initiation complex. *Science* **342**, 1238724 (2013).
- Sandelin, A. *et al.* Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nature Rev. Genet.* **8**, 424–436 (2007).

**Acknowledgements** We thank A. Chakraborty and B. Milic for careful reading of the manuscript, P.-J. Mattei for help with protein purification, and R. Landick and J. Gelles for discussions. This research was supported by NIH grants GM36659 and AI21144 to R.D.K. and GM57035 to S.M.B., and an NSF graduate fellowship to F.M.F.

**Author Contributions** F.M.F., C.A.M. and K.M. designed the experiments with input from, and supervision by, S.M.B. and R.D.K. F.M.F. and C.A.M. collected the single-molecule data. K.M. purified and reconstituted the PIC components, and performed the bulk experiments. F.M.F. analysed the single-molecule data, and C.A.M. and K.M. carried out the modelling. F.M.F., C.A.M., K.M., S.M.B. and R.D.K. wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.M.B. ([sblock@stanford.edu](mailto:sblock@stanford.edu)) or R.D.K. ([kornberg@stanford.edu](mailto:kornberg@stanford.edu)).

## METHODS

**Single-molecule optical-trapping assay.** The 29-subunit yeast PIC containing biotinylated Pol II was assembled on SNR20\* promoter DNA fused to a 2.7 kb DNA 'handle'. The DNA handle allowed us to form tethers in both the hindering-load and assisting-load assays by incorporating a digoxigenin tag via PCR at either the downstream or upstream end of the DNA, respectively. The constructs containing the handle were mixed in a 1:25 molar ratio with identical PICs assembled on the same promoter DNA, but without the handle, such that the overall concentration of the PIC was 100 nM. PIC complexes assembled on DNA lacking a handle sequence are unable to form tethers, and instead serve to increase the overall concentration of PIC by mass action. The resulting mixture was incubated with twofold excess of TFIIF<sup>18,19,31,32</sup> at room temperature for 20 min to form the complete 32-subunit PIC. This complete PIC was incubated with both anti-digoxigenin-coated 0.9- $\mu$ m diameter beads and avidin-coated 0.6- $\mu$ m diameter beads, resulting in tethers being formed with the digoxigenin-containing handle at one end and the biotinylated Pol II in the PIC at the other end. On completion of this step, the concentration of PIC was ~25 nM. In this latter step (and all subsequent steps), the buffer used (50 mM HEPES, pH 7.5, 80 mM potassium acetate, 5 mM MgSO<sub>4</sub>, 10 mM dithiothreitol (DTT), 10% glycerol) was always supplemented with 250 nM TFIIB and 250 nM TFIIE to stabilize and maintain the PIC. The assembled dumbbell tethers were flowed into a ~5  $\mu$ l flow chamber, and rinsed with ~10  $\mu$ l of additional buffer to remove excess beads. Each bead that formed the tether was held in a separate optical trap, allowing controlled load to be applied on the dumbbell by using an active force clamp as previously described<sup>14</sup>. Force uncertainties were estimated at ~15% owing to variations in bead size and systematic calibration errors. The temperature on the trap was estimated to be  $26 \pm 1$  °C (mean  $\pm$  s.d.)<sup>33</sup>. Single tethers were identified as described<sup>33</sup>, and held at ~4 pN constant force for ~15–20 s, after which transcription buffer (50 mM HEPES, pH 7.5, 80 mM potassium acetate, 10 mM magnesium acetate, 10 mM DTT, 10% glycerol, 1 U of RNaseOUT (Life Technologies), 250 nM TFIIB and 250 nM TFIIE) containing either 1.6 mM (2 $\times$ ) NTPs or 1.6 mM (2 $\times$ ) dATP, was flowed into the flow cell while holding the tether at the same force. In the absence of nucleotides, the dumbbells could be held at about 4 pN without breakage for extended periods. An oxygen-scavenging system (8.3 mg ml<sup>-1</sup> glucose (Sigma), 46 U ml<sup>-1</sup> glucose oxidase (Calbiochem), 94 U ml<sup>-1</sup> catalase (Sigma)) was used to reduce photodamage. Data were collected at 2 kHz sampling frequency, filtered at 1 kHz with an 8-pole Bessel filter (Krohn-Hite) and boxcar averaged over a 20-point window to provide positional feedback to an active force clamp at a rate of 100 Hz. The resulting data was analysed using Igor Pro (Wavemetrics).

**Protein purification.** TFIIA, TFIIB and TBP were expressed in bacteria, and TFIIE, TFIIF and TFIH were isolated from yeast<sup>5,17,29,34</sup>. Biotinylated Pol II was isolated as previously published<sup>15</sup>. For the expression of recombinant Sub1 (refs 16, 35–37), the *Escherichia coli* Rosetta2 (DE3) strain (Stratagene) was transformed with pCold II vector (Clontech) containing the *SUB1* (also known as *TSP1*) gene fused to sequence encoding a C-terminal His<sub>6</sub>-tag. The cells were grown in 2 $\times$  YT media at 30 °C, and induced with 1 mM isopropyl-1-thio- $\beta$ -D-galactopyranoside (IPTG) for 16 h at 15 °C. The cells were then lysed by sonication in a lysis buffer (20 mM Na/K-phosphate buffer, pH 7.5, 500 mM potassium acetate, 10 mM imidazole, 0.1% Triton X-100, 1 mM DTT, 1 mM benzimidazole, 100  $\mu$ M leupeptin, 10  $\mu$ M pepstatin A and 1 mM PMSF), and was eluted by a gradient of 10–500 mM imidazole in a buffer containing 20 mM Na/K-phosphate buffer, pH 7.5, 300 mM potassium acetate and 5% glycerol. The eluent was further purified using HiTrap Heparin 1 ml (GE healthcare) and CaptoSP ImpRes (GE Healthcare).

**DNA constructs for single-molecule experiments**

**SNR20\* short (–62/+96) promoter sequence for hindering-load assay.** The sequence of the non-template DNA strand was as follows: 5'-GCCGTTTCCGATGGGCCACTCGGTGAAAAACATATAAAAAAGGGCTCTACATTCATTTT TTTAAATGCCACGAATCTCTTTTCTTTTCGGGTGGATCAAGTGTAGT ATCTGTTCTTTTCAGTGAACAACTGAAATGACCTCAATGAGGCTCATT ACC-3'.

**SNR20\* long (–122/+96) promoter sequence for hindering-load assay.** The SNR20\* long differed from the SNR20\* short by containing an additional 60 bp of DNA between the TATA site and the start site. The sequence of non-template strand of the longer promoter is shown, with the bases underlined not present in the shorter promoter: 5'-GCCGTTTCCGATGGGCCACTCGGTGAAAAACATA TAAAAAGGGCTCTACATTCATTTTTCATCGATGAGTACTTTTACTTTGTT ATCAGATTTAATTCATTTTGTCTTCTACTTGTTTTTTTTTTAAATGCCACG AATCTCTTTTCTTTTCGGGTGGATCAAGTGTAGTATCTGTTCTTTTCAGT GTAACAACTGAAATGACCTCAATGAGGCTCATTACC-3'.

**SNR20\* short (–62/+636) promoter sequence for assisting-load assay.** The SNR20\* short sequence for the assisting-load assay contained an additional 540 bp of DNA downstream derived from the wild-type SNR20 gene relative to the hindering-load assay, as is shown below: 5'-GCCGTTTCCGATGGGCCACT

CGGTGAAAAACATATAAAAAAGGGCTCTACATTCATTTTTTTTAAATGCC ACGAATCTCTTTTCTTTTCGGGTGGATCAAGTGTAGTATCTGTTCTTTTC AGTGTAAACAACTGAAATGACCTCAATGAGGCTCATTACCTTTTAAATTTG TTACAATACACATTTTGGCCACCCAAAATAATAAAATGGACGGGAAG AGACTTTTAAAGCAAGTTGTTTTCCGCTAATGTCAGGCTCACTACTTT TTGGTGCTATTTTCTTCGCTCATGGTTTCTTCATAAAGCCGTTTTATG ATGGTTTTTCGAAATGGTTTTTGGACGACGGAATACGAAATCTGG ATCCTTGCTCAAGGTTATTGTTTTTGTCTTCTCTGGTTGTTTTCTATTT TCTTTTTTTAGCTTTCTGTTTCTCCCTTAGTTTGGCTTTTTGCTTCATA CTCTCCCTGTCTTCCGAGCCGTTTATGTCCAACGCGGGATTTGGTTT TTCTTTATCGATGGGAAGAAATGGTGTATAGTAGGTTGGGAGATAAT ATTTATGGTATGGGGTGTCTAGTCCGGATGGGGCTCTTATTGTTGAT TTCTTCGCTCGCTTCTTTTCTGTTGGCCTGCAAGGAAGTAAATTTTC GACTTTGTTATGATTTTTGGTTTGAAGGAAAGGTGTCTTAC-3'.

**Generation of DNA templates.** A 2.7 kb DNA fragment that served as a 'handle' in our single-molecule assay was amplified by PCR from the plasmid pRL702 as previously described<sup>14</sup>, and was subcloned into the pDrive Cloning Vector (Qiagen). To obtain the three DNA constructs used in this study, three different plasmids were constructed, each containing the handle adjacent to one of the following promoter sequences: SNR20\* short (–62/+636), SNR20\* short (–62/+96), and SNR20\* long (–122/+96). For the assisting-load assay, the handle was located upstream of the SNR20\* short (–62/+636) promoter sequence. For the hindering-load assay, the handle was situated downstream of the SNR20\* short (–62/+96) or SNR20\* long (–122/+96) promoter sequence. Regions containing the promoter and the handle were amplified by PCR, using a 5'-digoxigenin labelled primer (IDT) that anneals to the end of the handle, such that PCR products carry a digoxigenin tag on the upstream end of DNA for the assisting-load assay, and on the downstream end of DNA for the hindering-load assay. The generated PCR products were loaded onto TSKgel DEAE-5PW (Tosoh), eluted by a gradient of 0.1–1 M NaCl in a buffer containing 20 mM Tris, pH 7.5, and 2 mM DTT, and concentrated up to 5–10  $\mu$ M using Vivacon 500 5K MWCO (Vivaproducts), yielding ~0.15–0.3 nmol from 4–8 ml PCR reaction.

**PIC assembly and isolation.** The PIC was isolated as previously published<sup>5</sup> with minor modifications. First, 0.15 nmol of SNR20\* promoter DNA with or without the 2.7 kb handle was separately mixed with 1.5 nmol of TFIIB, 1.5 nmol of TFIIA, 0.8 nmol of TBP, 0.65 nmol of TFIIE, 0.24 nmol of TFIH-ATFIK, and 0.8 nmol of Sub1 in 90  $\mu$ l of buffer (500) (20 mM HEPES, pH 7.6, 5 mM DTT, 2 mM MgSO<sub>4</sub> and 5% glycerol, with the mM concentration of potassium acetate in parentheses). The mixture was dialysed into buffer (300), buffer (220), buffer (150), and then combined with 0.2 nmol of biotinylated Pol II-TFIIF complex. The mixture was further dialysed into buffer (80), and loaded onto a 10–40% (v/v) glycerol gradient containing 20 mM HEPES, pH 7.6, 5 mM DTT, 2 mM magnesium acetate and 80 mM potassium acetate, and was ultra-centrifuged for 4 h at 48,000 r.p.m. (Beckman SW60 Ti rotor). The presence of the 2.7-kb handle did not significantly affect the efficiency of assembly of the PIC.

**In vitro transcription assay.** Transcription assay was performed as described before<sup>5</sup>. In brief, 1.5 pmol of DNA fragment was combined with 3.7 pmol of TFIIB, 3.7 pmol of TFIIA, 1.5 pmol of TBP, 3.7 pmol of TFIIE, 1.5 pmol of TFIH, 1.5 pmol of Pol II, 2.1 pmol of TFIIF, 2.5 pmol of Sub1 in 5  $\mu$ l of buffer (300) (50 mM HEPES, pH 7.6, 300 mM potassium acetate, 5 mM DTT and 5% glycerol), diluted with 5  $\mu$ l of buffer (10) (20 mM HEPES, pH 7.6, 10 mM potassium acetate, 5 mM MgSO<sub>4</sub> and 5 mM DTT), and incubated for more than 1 h on ice. The transcription was initiated by adding an equal volume of buffer containing 20 mM HEPES, pH 7.6, 10 mM potassium acetate, 5 mM MgSO<sub>4</sub>, 10 mM magnesium acetate, 1 U of RNaseOUT, 5 mM DTT, 1.6 mM ATP, 1.6 mM GTP, 1.6 mM CTP, 40  $\mu$ M UTP and 0.83  $\mu$ M [ $\alpha$ -<sup>32</sup>P] UTP (2.5  $\mu$ Ci). The reaction was stopped after 15 min by adding 185  $\mu$ l of stop buffer (300 mM sodium acetate (pH 5.5), 5 mM EDTA, 0.7% SDS, 0.1 mg ml<sup>-1</sup> glycogen, 0.013 mg ml<sup>-1</sup> of proteinase K (Sigma)). Transcripts were precipitated by adding 700  $\mu$ l of ethanol, dried and analysed by a denaturing 4–12% acrylamide gel.

**Exonuclease footprinting.** Exonuclease footprinting was performed as described before<sup>5</sup>. SNR20\* long (–122/+147) was amplified by PCR in 2 ml reaction using <sup>32</sup>P-labelled upstream primer (5'-GCCGTTTCCGATGGGCCACTC-3') and downstream primer (5'-CCATTTTATTATTTTGGGTGCC-3'), and was purified by electrophoresis in a 2% agarose gel. The labelled DNA (1.5 pmol) was incubated with 3.7 pmol of TFIIB, 3.7 pmol of TFIIA, 2.0 pmol of TBP, 3.7 pmol of TFIIE, 2.0 pmol of TFIH, 2.0 pmol of Pol II-TFIIF complex, and 3.8 pmol of Sub1 in 5  $\mu$ l of buffer (300) (50 mM HEPES, pH 7.6, 300 mM potassium acetate, 5 mM DTT and 5% glycerol), then combined with 5  $\mu$ l of buffer (30) (50 mM HEPES, pH 7.6, 5 mM MgSO<sub>4</sub>, 30 mM potassium acetate and 5 mM DTT), and incubated for more than 1 h at 4 °C. The reconstituted PIC was combined with an equal volume of 2 $\times$  NTP buffer (1.6 mM NTP(s) or 1.6 mM dATP,

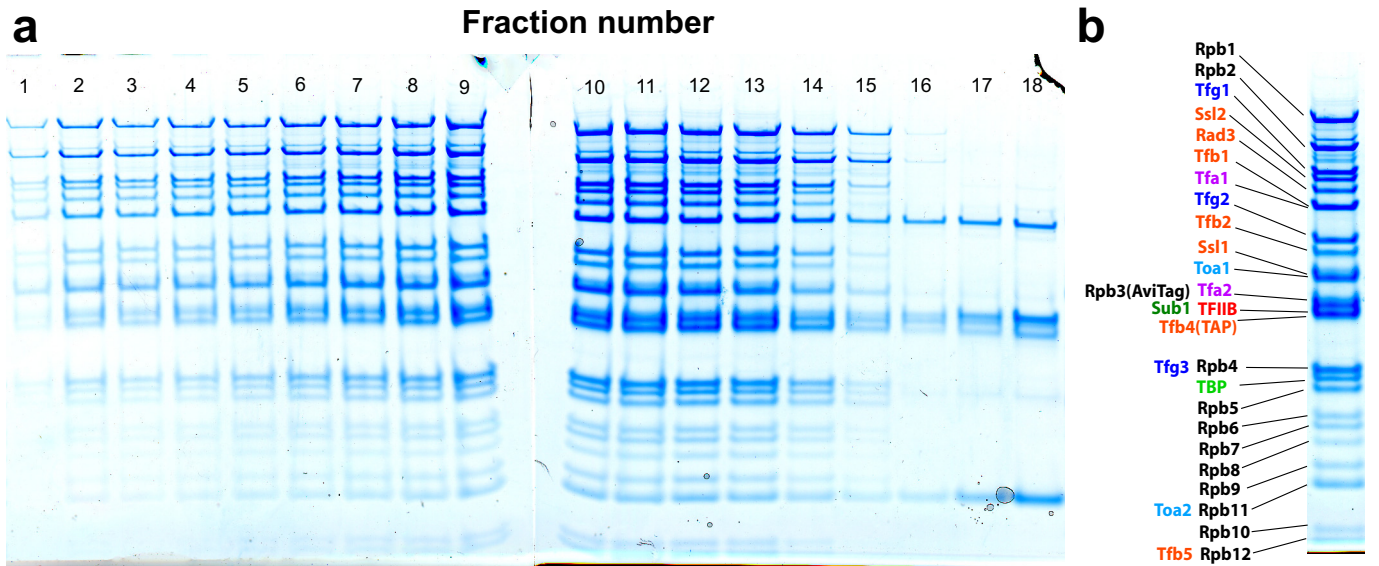
50 mM HEPES, pH 7.6, 5 mM MgSO<sub>4</sub>, 30 mM potassium acetate, 5 mM DTT, 10 mM magnesium acetate and 5 U of RNaseOUT) and incubated for 4 min at 30 °C. Exonuclease III digestion was performed with 5–10 U of the exonuclease (NEB) for 9 min at 30 °C, and was stopped by adding 185 µl of stop buffer (300 mM sodium acetate, pH 5.5, 5 mM EDTA, 0.7% SDS, 0.1 mg ml<sup>-1</sup> glycogen, 0.013 mg ml<sup>-1</sup> of proteinase K (Sigma), 0.5 mg ml<sup>-1</sup> salmon sperm DNA (Invitrogen)). DNAs were precipitated by adding 700 µl of ethanol, dried and analysed by a denaturing 6% acrylamide gel.

**Data analysis of single-molecule records.** Transcription in the expected direction resulted in a decrease in extension of the DNA tether in the hindering load assay, and an extension increase the assisting-load assay. In both geometries, the change in extension of the DNA tether, which is a function of the applied force, was converted to a distance on the template in bp ( $\sim 0.313 \text{ nm bp}^{-1}$  at  $\sim 4 \text{ pN load}^{14}$ ). The data acquired was smoothed in software by applying a low-pass filter to it (end of pass band = 0.1 Hz; start of reject band = 50 Hz, number of coefficients = 500). To align the records so that motion was defined to start at 0 bp, the mean value of the region  $\sim 1$ – $2 \text{ s}$  of positional data before the start of processive motion was set to be the starting (0 bp) distance. The velocities of Pol II and TFIIF reported were obtained by dividing the observed change in distance by the time over which the molecule moved. For both Pol II and TFIIF, these velocities were calculated over a region of at least 50 bp that did not contain any resolved pauses ( $>0.1$ – $0.2 \text{ s}$ ). As some of the pauses were short lived, especially owing to TFIIF motion, we did not have sufficient information to characterize the pause lifetimes or distributions, nor could we reliably use previous techniques<sup>14</sup> to get a pause-free velocity. We were also unable to determine pause-free velocities from the distributions of instantaneous velocities<sup>38</sup>, as there was often insufficient data to obtain reliable fits. For the TFIIF records, velocities of the scanning state and fast states, when observed, were occasionally calculated by examining different regions of the molecules that were separated in time by a relatively sharp transition (change within 0.5 s) in velocity. To estimate the processivity of TFIIF during scanning, we only included

molecules that travelled at least 30 bp, sufficient to move beyond the noise threshold and extend past the distance of the open complex ( $24 \pm 2 \text{ bp}$ ).

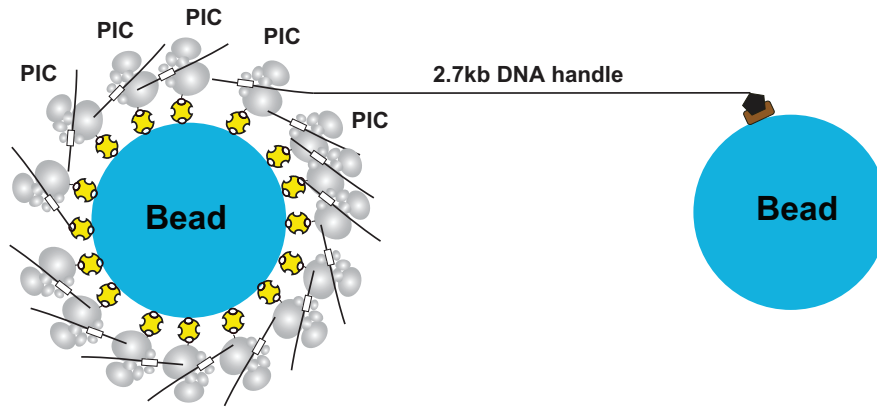
**Calculating expected open-complex distance.** Biochemical studies have shown that the minimal distance of a TSS from the TATA box is about 30 bp, and that transcription begins at this location in the initial open complex<sup>1</sup>. In the structure of the closed PIC<sup>29</sup>, this location in the promoter DNA is about 80 Å from the nucleotide addition site of Pol II. Open complex formation must bring the TSS to the nucleotide addition site, which therefore requires drawing 80 Å of downstream DNA into the Pol II cleft. In the structure of a transcribing complex<sup>20</sup>, all but 3 bp of downstream DNA are double-stranded, and 80 Å corresponds to 24 bp of dsDNA, the same as the initial movement observed in the hindering-load assay.

31. Laybourn, P. J. & Dahmus, M. E. Phosphorylation of RNA polymerase IIA occurs subsequent to interaction with the promoter and before the initiation of transcription. *J. Biol. Chem.* **265**, 13165–13173 (1990).
32. Feaver, W. J., Gileadi, O., Li, Y. & Kornberg, R. D. CTD kinase associated with yeast RNA polymerase II initiation factor b. *Cell* **67**, 1223–1230 (1991).
33. Koslover, D. J., Fazal, F. M., Mooney, R. A., Landick, R. & Block, S. M. Binding and translocation of termination factor rho studied at the single-molecule level. *J. Mol. Biol.* **423**, 664–676 (2012).
34. Murakami, K. *et al.* Tfb6, a previously unidentified subunit of the general transcription factor TFIIF, facilitates dissociation of Ssl2 helicase after transcription initiation. *Proc. Natl Acad. Sci. USA* **109**, 4816–4821 (2012).
35. Ge, H. & Roeder, R. G. Purification, cloning, and characterization of a human coactivator, PC4, that mediates transcriptional activation of class II genes. *Cell* **78**, 513–523 (1994).
36. Henry, N. L., Bushnell, D. A. & Kornberg, R. D. A yeast transcriptional stimulatory protein similar to human PC4. *J. Biol. Chem.* **271**, 21842–21847 (1996).
37. Malik, S., Guermah, M. & Roeder, R. G. A dynamic model for PC4 coactivator function in RNA polymerase II transcription. *Proc. Natl Acad. Sci. USA* **95**, 2192–2197 (1998).
38. Neuman, K. C., Abbondanzieri, E. A., Landick, R., Gelles, J. & Block, S. M. Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. *Cell* **115**, 437–447 (2003).



**Extended Data Figure 1 | The 29-component PIC assembled on *SNR20\** short promoter.** **a**, PIC excluding the kinase domain (TFIIK) was assembled on *SNR20\** short (adjacent to the 2.7-kb downstream handle sequence) and sedimented on a glycerol gradient; fractions were analysed by SDS-PAGE. **b**, The results from fraction 12, annotated in detail, indicate that all PIC

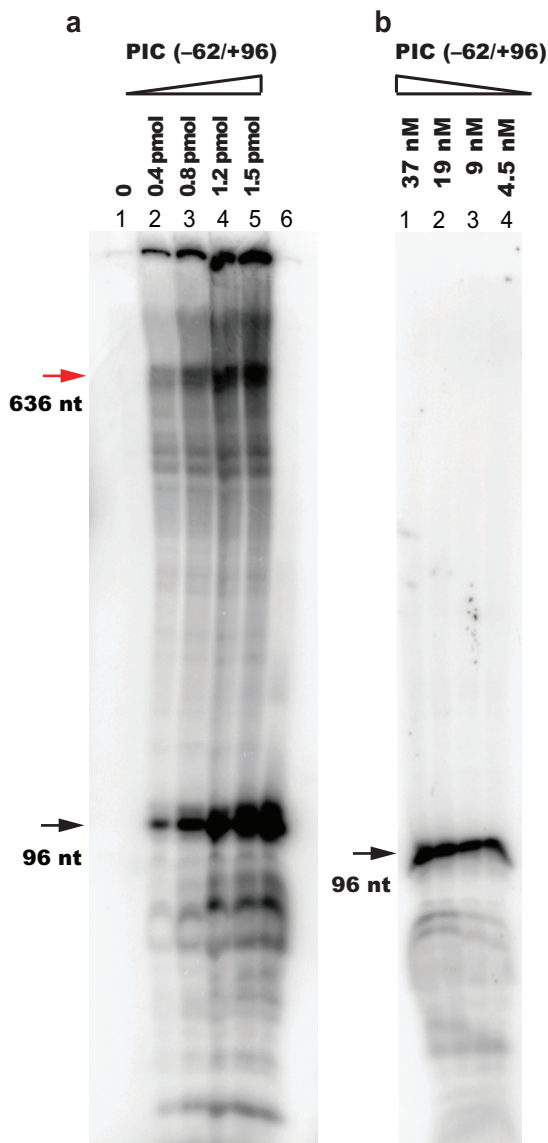
components were retained, confirming that the complex reconstituted fully from the component proteins. The subunit(s) of Pol II are labelled in black, TFIIF in blue, TFIIE in magenta, TFIIH in orange, TFIIA in cyan, TFIIB in red, TBP in light green, and Sub1 in dark green. TFIIK (3-subunits) was later added to the PIC.



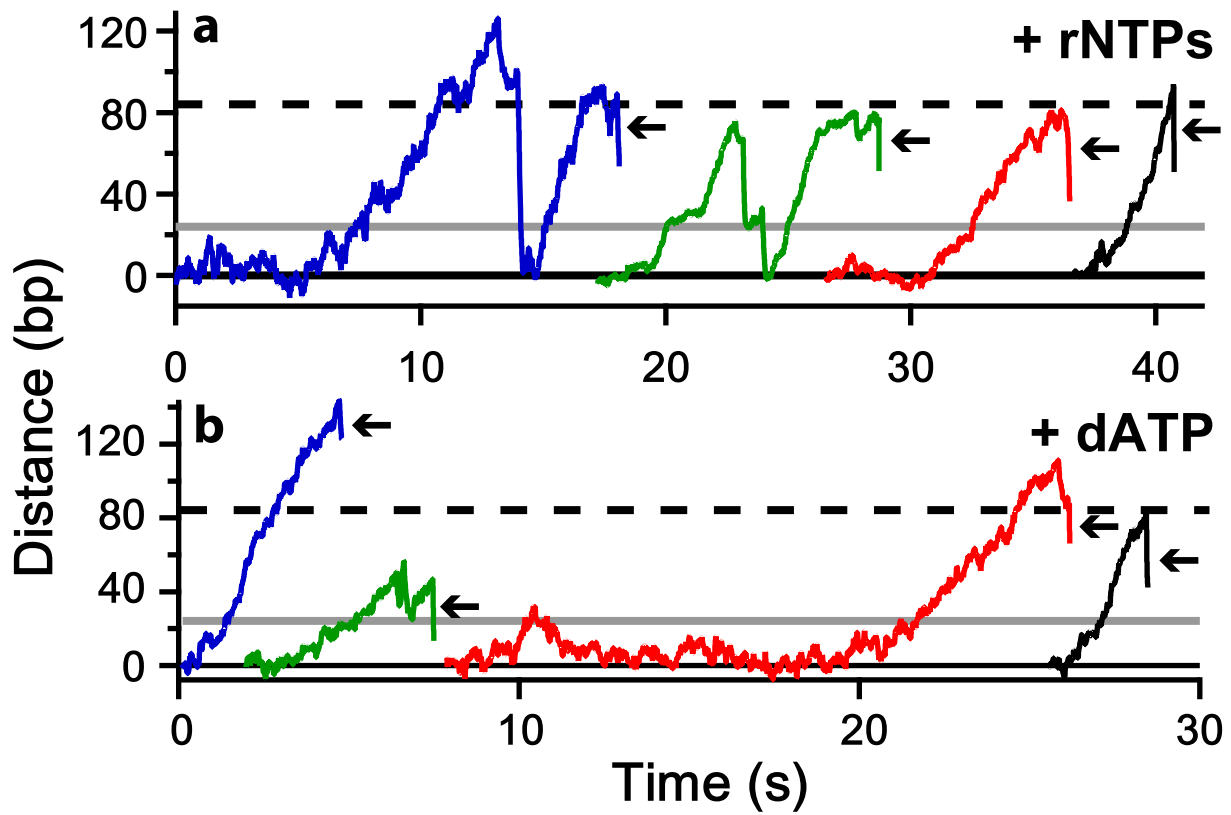
**Extended Data Figure 2 | Schematic diagram showing assembly of dumbbells, in cross-section.** PICs were attached to one bead via biotin-avidin linkages (yellow). To form dumbbell tethers, the other end of a small fraction of the PICs (4%) had digoxigenin linkages that could be tethered to anti-

digoxigenin-coated beads (black and brown) via a 2.7-kb DNA handle. PICs not involved in tether formation served to increase the local concentration of PIC components.



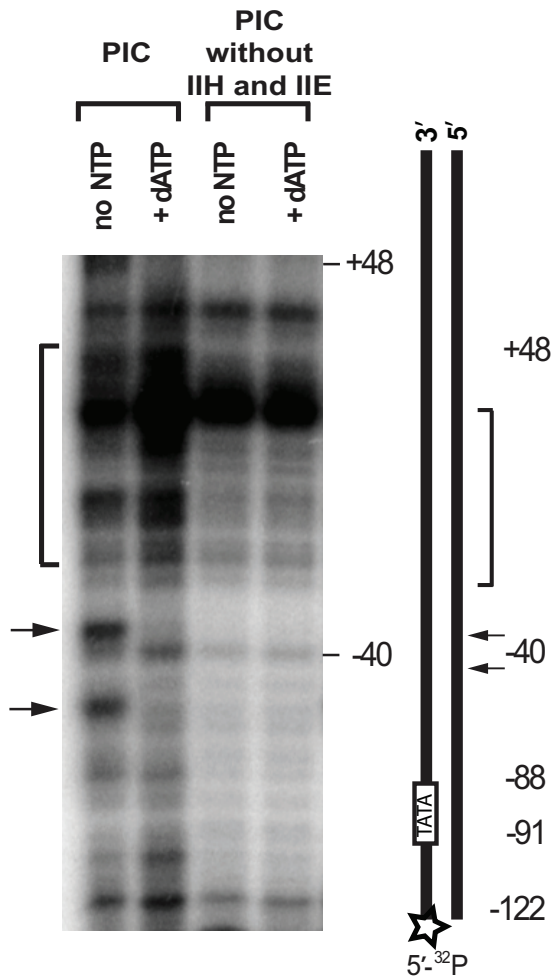


**Extended Data Figure 3 | Run-off transcription under single-molecule assay conditions.** **a**, Isolated PICs (0.1 pmol), formed on the *SNR20\** short promoter fragment fused to the transcription template (covering the region -62/+636), and attached to a 2.7 kb DNA handle, was combined with increasing amounts of PICs assembled on the *SNR20\** short promoter, but without the handle, hereafter referred to as PIC (-62/+96). These constituents were incubated with an equal volume of a 2X NTP solution (10 ml) containing 1.6 mM ATP, 1.6 mM GTP, 1.6 mM CTP, 40 mM UTP, and 0.83 mM [ $\alpha$ - $^{32}$ P] UTP (2.5  $\mu$ Ci). The resulting transcripts were analyzed by gel electrophoresis. PICs fused to the DNA handle failed to support transcription alone (lane 1), but transcription activity was restored (red arrow) when a 4-fold (lane 2), 8-fold (lane 3), 12-fold (lane 4), or 15-fold (lane 5) excess of PIC (-62/+96) was added to the reactions. In lane 6, the reaction contains 1.5 pmol PIC (-62/+96). The 96-nt run-off transcription from PIC (-62/+96) is indicated (black arrow). A 25-fold excess of PIC (-62/+96) was used for single-molecule assays (Extended Figure. 2). **b**, 1.5 pmol aliquots of PIC (-62/+96) were introduced into different volumes of transcription buffer, such that assayed concentration of PIC varied from 37 nM to 4.5 nM. Transcription efficiency (run-off band, black) decreased with PIC concentration from ~18% to just 2–3%. The low concentrations used in single-molecule assays (<1 nM) could not be assayed directly using gels, but we expect that the transcription efficiency is correspondingly low.

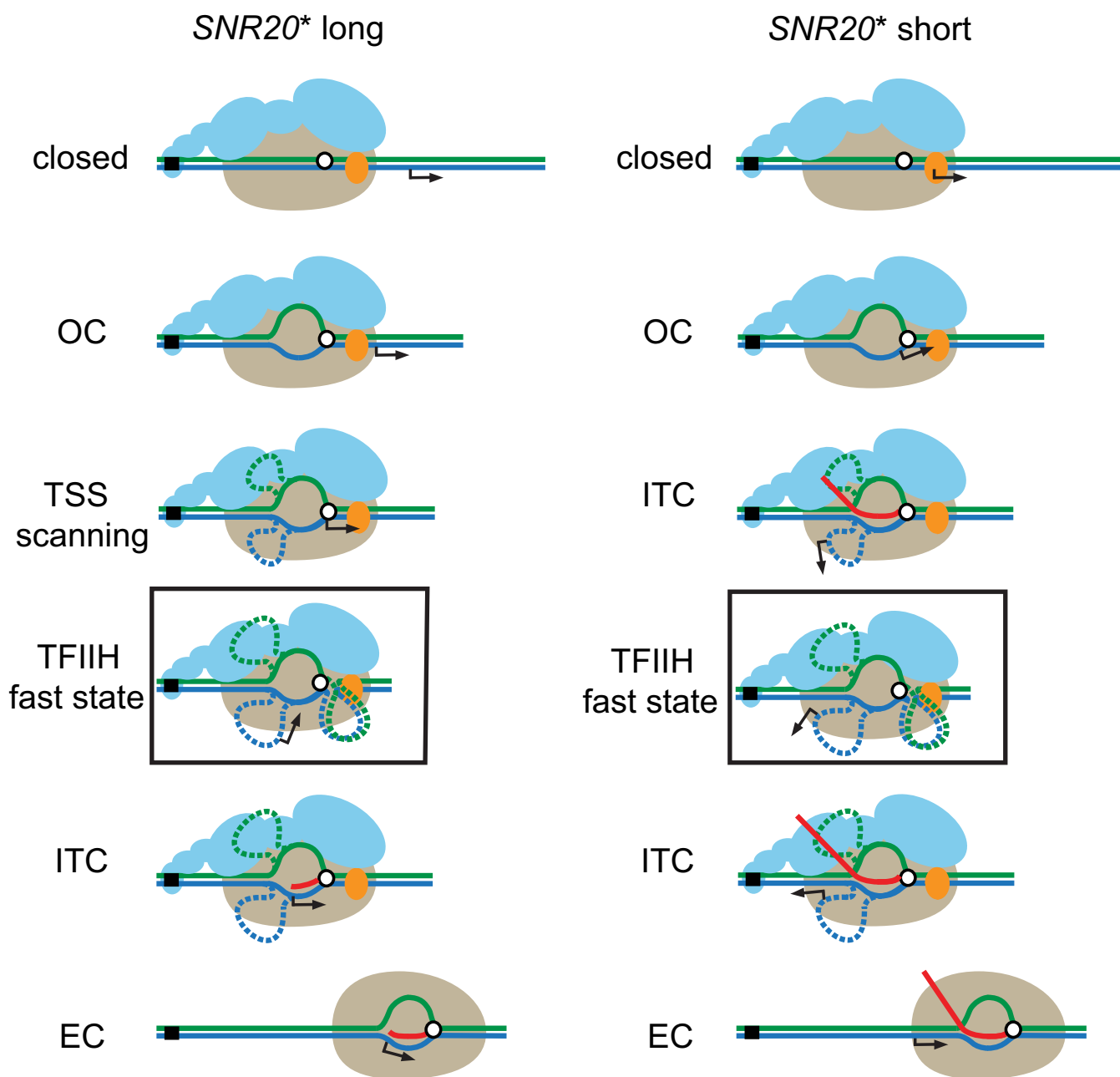


**Extended Data Figure 4 | Records of TFIIH scanning on SNR20\* long with rNTPs or dATP. a, b,** Just as for SNR20\* short (Fig. 3), the longer promoter shows TFIIH scanning with both rNTPs (a) or dATP only (b), after which

either the PIC dissociates (black arrows), or the bubble collapses to the closed (blue and green records) or open (grey line) complex and TFIIH moves again. The dashed line indicates the position of the TSS (+1).



**Extended Data Figure 5 | Exonuclease III footprinting assay of the PIC on *SNR20\*long*.** In the absence of nucleotides *in vitro*, PIC complexes bound to the *SNR20\*long* promoter produced barriers to exonuclease III digestion located ~50 bp downstream of the TATA box (about -40 nucleotides from the TSS, black arrows). These barriers depended on the presence of TFIID and also TFIIE, which interacts with TFIID. After the addition of dATP, the barriers disappeared, and the bands at pause positions were intensified between positions -30 and +30 (~60–120 bp downstream of the TATA box, bracket).



**Extended Data Figure 6 | The transcription initiation pathway for *SNR20\** long (left) and *SNR20\** short (right) promoters.** Left, a model for the initiation pathway on the *SNR20\** long promoter. States starting from the top: Pol II (beige) with attached GTFs (blue) and Ssl2 (orange) binds in its 'closed' form to the promoter element upstream of the TSS (arrow) on the DNA template (green and blue lines). Positions of the enzyme active site (open white circle) and TATA box (closed black square) are indicated. Unwinding by TFIIH produces an open complex (OC) that leads to bubble formation. Arrival of the open complex at the TSS owing to scanning, driven by TFIIH, leads to the formation of an extended bubble (dashed lines indicate the speculative position of single-stranded DNA). If the complex fails to recognize

the TSS, it can be driven beyond it by TFIIH, resulting in a 'fast state' that produces no RNA but advances at roughly twice the normal rate (black box; see text). When Pol II recognizes the TSS, it begins transcription of RNA (red line), corresponding to the ITC. Formation of the ITC leads to bubble collapse, followed by the loss of GTFs and transition to the elongation complex (EC). Right, corresponding model for the initiation pathway on the *SNR20\** short promoter. Similar states as for *SNR20\** long. In this case, the open complex does not need to scan for the TSS, which is found within its DNA footprint. As a consequence, the ITC can form and begin RNA synthesis once the active site has recognized the TSS. A longer segment of RNA can thereby be produced before the transition to the elongation complex (EC).