nature genetics

Paused RNA polymerase II inhibits new transcriptional initiation

Wanqing Shao[1] & Julia Zeitlinger[1,2]

RNA polymerase II (Pol II) pauses downstream of the transcription initiation site before beginning productive elongation. This pause is a key component of metazoan gene expression regulation. Some promoters have a strong disposition for Pol II pausing and often mediate faster, more synchronous changes in expression. This requires multiple rounds of transcription and thus cannot rely solely on pause release. However, it is unclear how pausing affects the initiation of new transcripts during consecutive rounds of transcription. Using our recently developed ChIP-nexus method, we find that Pol II pausing inhibits new initiation. We propose that paused Pol II helps prevent new initiation between transcription bursts, which may reduce noise.

Pol II initiation has long been known to be rate limiting in transcription and has been extensively studied using biochemical, structural and, more recently, imaging approaches[1–5]. As a result, there is a wealth of knowledge about how Pol II and the basal transcription factors TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH assemble into the pre-initiation complex (PIC) on active promoters[4,5], unwind the double-stranded DNA with the help of the XPB helicase/translocase of TFIIH, and initiate transcription[6,7].

The function of Pol II pausing is less well understood. Pol II pauses 30–50 bp downstream of the transcription start site (TSS), with the help of pausing factors such as NELF, and requires the P-TEFb kinase complex to continue the productive elongation of a full-length transcript[8,9]. Some promoters show a particularly strong tendency for Pol II pausing, and this property is tightly linked to the promoter's DNA sequence[10–12]. Strongly paused promoters can mediate faster and more synchronous gene expression in response to a developmental signal[12,13]. The high degree of pausing suggests that pause release is rate limiting at these promoters, but how this would favorably affect expression dynamics is not entirely clear.

Gene expression in response to developmental or environmental signals typically involves many rounds of transcription and thus requires new Pol II initiation in addition to pause release. This raises the question of whether initiation can take place in the presence of paused Pol II or whether paused Pol II is released first to allow new initiation to occur. This is a challenging problem to resolve because it is difficult to study Pol II pausing and initiation with the same method. Biochemical approaches for assaying the PIC do not detect Pol II pausing, while genomics techniques such as Pol II ChIP–seq[14,15], genome-wide permanganate footprinting[16] and global run-on sequencing (GRO–seq and PRO–seq)[11,17,18] do not capture the PIC. Importantly, even when Pol II was mapped by ChIP–seq in combination with lambda exonuclease digestion (ChIP–exo)[19], which has the resolution to distinguish between the initiation and pausing positions, Pol II at the PIC stage was not detected[20].

To investigate the relationship between initiation and pausing, we have used a robust ChIP–exo protocol called ChIP-nexus[21] and mapped Pol II, basal transcription factors and pausing factors in the presence or absence of transcriptional inhibitors. We show that Pol II can be detected as part of the PIC after treatment with triptolide (TRI), a drug that binds XPB and inhibits initiation[22]. Together with half-life measurements of paused Pol II, our results suggest that paused Pol II is much more stable than Pol II at the PIC stage. Further analysis shows that PIC assembly and paused Pol II do not typically occur on the same promoter and that paused Pol II inhibits new initiation. This suggests that paused Pol II is released before new initiation can occur. We discuss our findings in the context of transcriptional bursting and noise.

RESULTS

Occupancy of the transcription machinery at high resolution

We first mapped the precise positions of Pol II and a number of basal transcription factors, including TFIIA, TFIIB, TFIIF, TBP, TAF2 and XPB, by performing ChIP-nexus experiments in *Drosophila melanogaster* Kc167 cells. Exonuclease digestion is performed during the immunoprecipitation of cross-linked chromatin and stops at proteins that are directly or indirectly bound to the factor of interest[19]. After sequencing, the positions of all the stop bases are obtained across the genome. In this manner, a cross-linked protein leaves a footprint consisting of reads mapping to the positive strand upstream and to the negative strand downstream (**Fig. 1a**).

The ChIP-nexus data showed footprints for each factor at unprecedented resolution, much higher than that in ChIP–seq data (**Fig. 1a**). The data were highly reproducible without data smoothing, even at the single-gene level, identifying distinct positions for the footprints of TFIIA, TFIIB and Pol II relative to the TSS (**Fig. 1a** and **Supplementary Fig. 1**).

[1]Stowers Institute for Medical Research, Kansas City, Missouri, USA. [2]Department of Pathology and Laboratory Medicine, University of Kansas Medical Center, Kansas City, Kansas, USA. Correspondence should be addressed to J.Z. (jbz@stowers.org).

Received 15 November 2016; accepted 19 April 2017; published online 15 May 2017; doi:10.1038/ng.3867

To calculate the average footprint for each factor across the genome, we used TSS annotations determined experimentally by PRO-cap[18] and aligned and averaged the ChIP-nexus signal to the top 1,000 promoters with the highest TFIIB signal. The results showed that the average Pol II footprint was located downstream of the TSS at +29 bp, the midpoint between the upstream summit (+21 bp) and the downstream summit (+38 bp) (**Fig. 1c** and **Supplementary Fig. 2**). As observed previously[16,20], there was no distinct Pol II footprint at the expected site of PIC formation.

The basal transcription factors involved in PIC formation had a distinct binding profile, with defined footprints at various positions relative to the TSS (**Fig. 1b**,**c**), consistent with their known or predicted DNA contacts based on X-ray crystallography and cryoelectronmicroscopy (cryo-EM)[1]. The strong differences between factors suggest that the ChIP-nexus signal preferentially reflects direct DNA contacts and that indirect contacts to DNA through other components of transcription complexes do not dominate the detected signal.

The basal transcription factors made DNA contacts in roughly five regions (numbered in **Fig. 1b**). As discussed below, the regions near the initiation site closely matched those of structural models, whereas the regions furthest away (regions 1, 4 and 5) were unexpected.

TFIIA was found furthest upstream, consistent with X-ray or cryo-EM structures in which it stabilizes TBP binding to the TATA box[23–26]. Unexpectedly, the footprint of TFIIA was even stronger upstream of the TATA box (region 1, midpoint at −46 bp; **Fig. 1b**,**c**). This suggests additional contacts of TFIIA *in vivo*, beyond the length of DNA typically used in *in vitro* studies.

TFIIB, TFIIF and TBP showed footprints expected from their interactions with DNA between the TATA box and the TSS (region 2, midpoint at −12 bp; **Fig. 1c** and **Supplementary Fig. 2**). The upstream summit on the positive strand was located just upstream from where the three factors form a subcomplex that surrounds and stabilizes the promoter DNA during initiation[26,27]. The downstream summits for TFIIB and TFIIF on the negative strand were around the TSS, consistent with TFIIB and TFIIF engaging and stabilizing the melted DNA strands[26,28–30].

In addition to binding near the TATA box, TBP also made strong but unexpected contacts downstream of the TSS, with a footprint that looked similar to that of paused Pol II (region 4, midpoint at +32; **Fig. 1c** and **Supplementary Fig. 2**). This additional footprint was reproducible and was also found in TBP ChIP–exo and ChIP-nexus data in human cells[20,21] (**Supplementary Fig. 3**).

The simplest explanation is that the downstream footprint of TBP occurs through TFIID, a large complex, of which TBP is a part, that extends to the downstream promoter region[31,32]. However, we favor the idea that TBP occupies this downstream location more directly. Cross-linking TBP to this position indirectly through TFIID would require multiple protein–protein cross-links[32], which would be predicted to result in low ChIP signal. Furthermore, the ChIP-nexus profile of one of the TFIID components that mediates the downstream DNA contacts, TAF2 (refs. 32,33), was different from that of TBP. Consistent with structural studies[32], TAF2 had a strong footprint downstream of the TSS (region 3, midpoint +15 bp; similar to XPB), but in addition there was a footprint further downstream, beyond that of TBP and paused Pol II (region 5, midpoint +44 bp; **Fig. 1c** and **Supplementary Fig. 2**). This difference between the downstream footprints of TBP and TAF2 makes it unlikely that TBP contacts this position through TAF2.

Detection of the PIC after triptolide treatment

Our data captured each factor in a population of cells, where promoters are found in different stages of transcription. Under these conditions, Pol II as part of the PIC is not detected, either because the vast majority of promoter-bound Pol II is paused rather than in a PIC or because the PIC stage is hard to detect in ChIP experiments. We therefore set out to specifically capture the ChIP-nexus profile of PICs or very early transcription intermediates. TRI covalently binds XPB and inhibits ATP-dependent promoter melting *in vitro*[22]. Treatment with TRI inhibits mRNA production *in vivo*[34] and causes the ChIP signal of Pol II to shift upstream[35]. When we treated cells with TRI for 1 h, we confirmed a decrease in Pol II Ser2 phosphorylation consistent with loss of transcription and found that Pol II and basal transcription factors were not degraded (**Supplementary Fig. 4**).

In ChIP-nexus experiments with TRI-treated cells, we found that the profiles showed changes consistent with increased detection of PICs. Although some Pol II was still detected at the pause position, the strongest Pol II footprint was located upstream of the TSS (midpoint at −19 bp; **Fig. 1d** and **Supplementary Fig. 2**), exactly where Pol II contacts DNA as part of the PIC *in vitro*[26,30]. This Pol II footprint was similar to those of TFIIB, TFIIF and TBP, all of which were markedly stronger and narrower after TRI treatment as compared to control conditions (**Fig. 1c**,**d**). While the positive-strand summits of TFIIB and TFIIF remained at the same location, the negative-strand summits shifted upstream (TFIIB from +2 to −15 bp, TFIIF from +2 to −8 bp; **Fig. 1c**,**d** and **Supplementary Figs. 2** and **5**). These footprint changes are consistent with a transition from an open complex to a closed state, in which TFIIB and TFIIF do not engage with the melted DNA strands[26,30,36]. This suggests increased detection of PICs or other early intermediates.

Another notable change after TRI treatment was loss of the downstream footprint of TBP, suggesting that downstream binding of TBP is not prevalent during PIC formation. Furthermore, XPB had a much more defined footprint downstream of the TSS (midpoint at +15 bp), consistent with XPB serving as a block to initiation in the presence of TRI[37]. The footprints of TFIIA and TAF2, the outermost components of the PIC, remained mostly unchanged (**Fig. 1d** and **Supplementary Fig. 2**). Thus, some PIC components such as TBP appear to change their footprint during transcription, while other components may remain behind as scaffold.

Because ChIP-nexus was able to detect Pol II as part of the PIC after TRI treatment, the likely reason for not detecting it under normal conditions is that the Pol II ChIP signal predominantly comes from paused Pol II. This is plausible because PIC formation occurs on the order of seconds *in vitro* and *in vivo*[38,39], while paused Pol II can be very stable, with half-lives estimated from approximately 10 min at the *Hsp70* promoter as measured by imaging studies or run-on assays[40,41] to as long as 1 h by treating cells with TRI and analyzing the loss of paused Pol II (ref. 35). On the basis of these measurements, paused Pol II is possibly orders of magnitude more stable than PIC Pol II. This could explain why some paused Pol II signal is still detectable after 1 h of TRI treatment.

Genome-wide half-life measurements of paused Pol II

We performed Pol II half-life measurements across the genome to globally analyze the relationship between PIC formation and Pol II pausing times. We treated cells with TRI and performed Pol II ChIP-nexus in a time-course experiment (examples in **Fig. 2a**). As TRI treatment prevents new Pol II from reaching the pause position, previously existing paused Pol II is eventually lost, either by transitioning into elongation or due to transcript termination[34]. Unlike previous measurements, the high resolution of the ChIP-nexus data allowed us to specifically measure the Pol II signal at the pausing position with minimal influence from the site of PIC formation, where Pol II signal

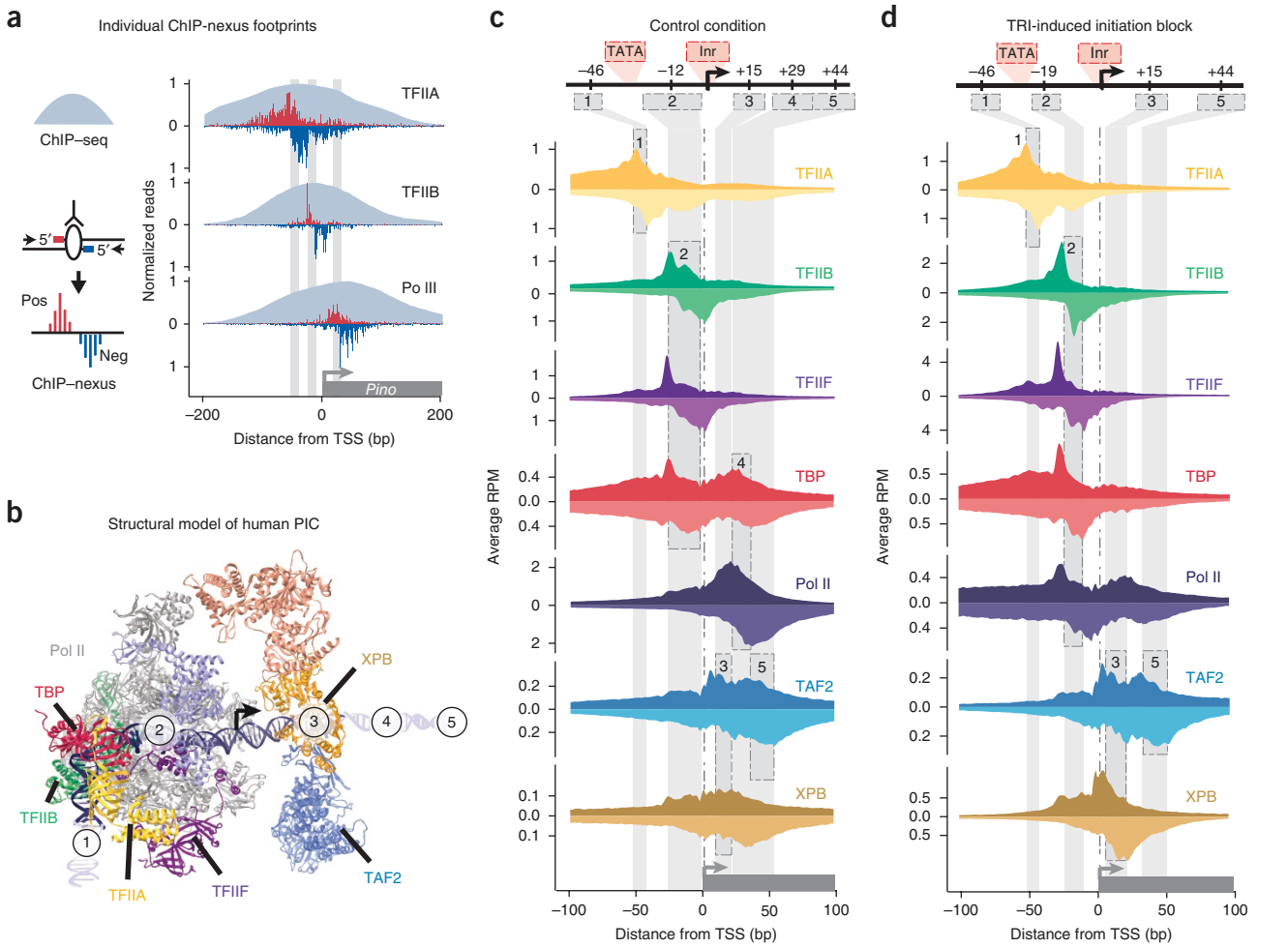

Figure 1 ChIP-nexus detects the precise locations of basal transcription factors and Pol II along the core promoter *in vivo*. (**a**) ChIP-nexus data in *Drosophila* Kc167 cells show clear footprints at distinct locations, even at individual promoters (for example, *Pino*). A footprint consists of an upstream border on the positive strand (above line, red) and a downstream boundary on the negative strand (below line, blue). Standard ChIP–seq data (gray) do not show such resolution. (**b**) A modeled crystal structure of the human PIC shows the contacts of the basal transcription factors with DNA and the five main areas of contact identified by ChIP-nexus (numbered circles). (**c**,**d**) Average footprints of basal transcription factors and Pol II in reads per million (RPM) on the positive strand (dark colored, above line) and negative strand (light colored, below line) across the top 1,000 promoters show distinct areas of DNA contact consistent with structural data (gray bars and numbered boxes). Approximate positions are shown on top in relation to the promoter elements initiator (Inr) and TATA box (TATA). Exact positions are shown in **Supplementary Figure 2**. (**c**) Structural agreement is observed under steady-state control conditions. (**d**) It becomes even more pronounced after treatment with TRI, which blocks initiation, indicating that Pol II is now detected as part of the PIC.

is retained after TRI treatment (**Fig. 2a**, left). We also included spike-in controls to account for the loss of total ChIP signal over time. For promoters with high Pol II signal and an identifiable pausing position (2,329 promoters), we fitted the Pol II time-course measurements to an exponential decay model and calculated the half-life of paused Pol II at each promoter (**Fig. 2b**,**c** and **Supplementary Table 1**). The promoters were then rank-ordered on the basis of their paused Pol II half-life and divided into five quintiles (q1 to q5; **Fig. 2d**).

We found that 1,798 promoters had a paused Pol II half-life shorter than 60 min and most half-lives fell between 5 and 20 min (example in **Fig. 2a**, left; **Fig. 2b**,**c**). However, there were some promoters where paused Pol II occupancy remained essentially unchanged after 30 min of TRI treatment ($n = 531$; example in **Fig. 2a**, right), suggesting that they represent stably paused promoters. These results are comparable to previous Pol II half-life measurements with TRI[34,40] and, on the basis of further analysis, fit our expectations.

We found that the relative differences between the Pol II half-lives among the five quintiles were consistent with mRNA–seq data and Pol II occupancy under steady-state conditions. Promoters with longer pause durations tended to have a lower steady-state level of mRNA, as well as a higher pausing index—the ratio between Pol II at the pausing position and the gene body (**Fig. 2e**).

We also performed additional experiments to test whether the Pol II pausing half-lives we measured were consistent with NELF-dependent pausing. ChIP-nexus experiments with NELF-E under steady-state conditions showed increased NELF-E occupancy with longer Pol II half-lives (**Fig. 2f**). Furthermore, stably paused promoters (q5) maintained high NELF-E occupancy after initiation was blocked with TRI, while promoters with a short Pol II half-life (q1) showed an almost complete loss of NELF-E (**Fig. 2f**). Finally, when we knocked down NELF, the observed loss of Pol II correlated with our half-life measurements; the most dramatic loss of Pol II was observed at promoters with the longer half-lives for paused Pol II (**Fig. 2g** and **Supplementary Fig. 6**).

Finally, we found that promoters with different degrees of pausing had enrichments for different core promoter elements

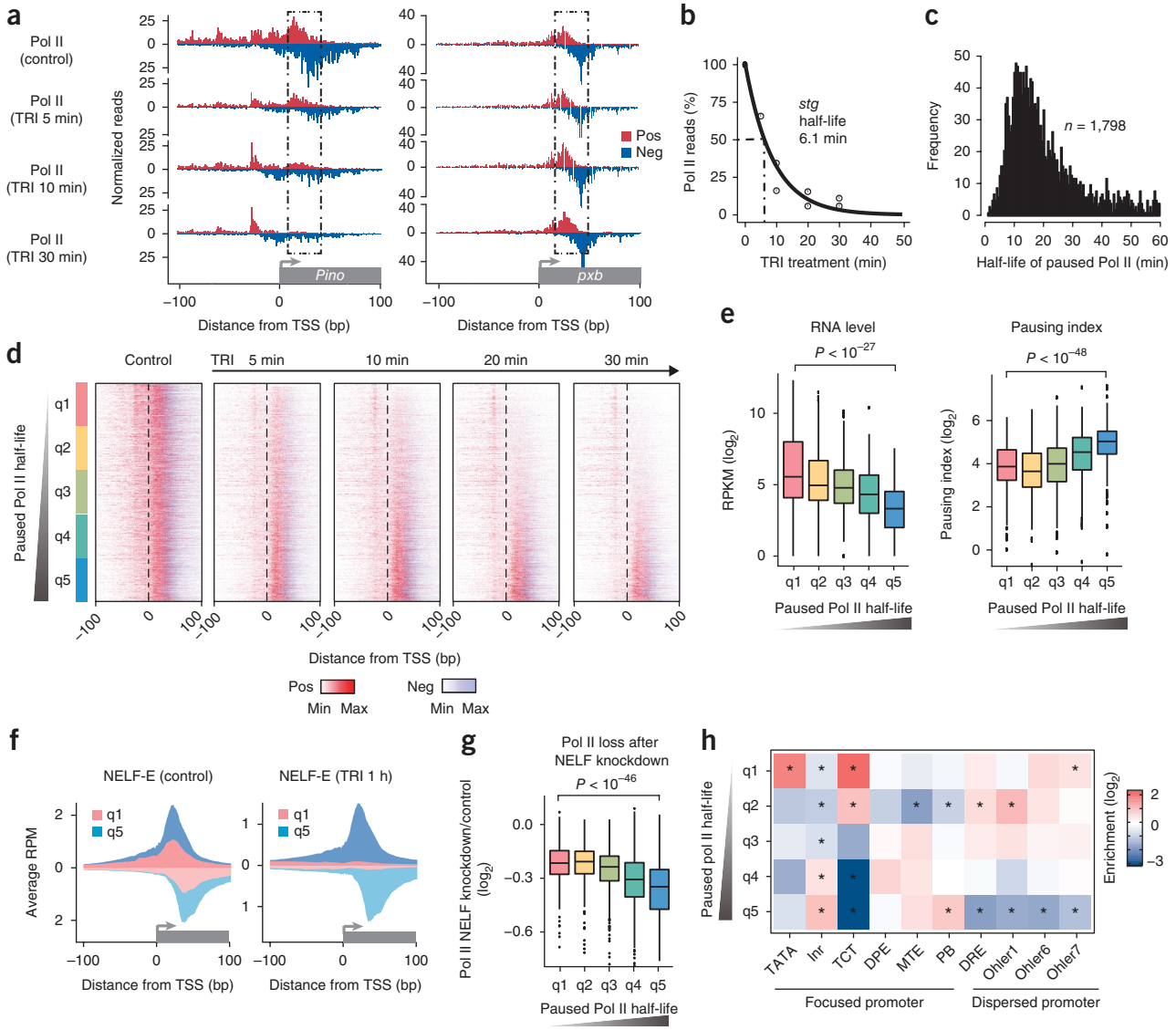

Figure 2 Measurement of the half-lives of paused Pol II across the genome using triptolide treatment. (**a**) Time-course Pol II ChIP-nexus experiments in response to TRI treatment, which prevents initiation and new occurrence of paused Pol II. Reads were normalized on the basis of a spike-in control. For each promoter with a clear Pol II pausing position, the decreasing levels of paused Pol II were calculated in the pausing window (dashed box) over time. (**b**) The half-life of paused Pol II was calculated on the basis of an exponential decay model. As an example, the paused Pol II reads for the *stg* gene are shown over time. (**c**) Histogram of the half-lives of paused Pol II for 1,798 genes showing that most half-lives are shorter than 30 min. (**d**) Heat maps of Pol II ChIP-nexus data over time sorted by Pol II half-life. The division of the promoters into five quintiles is shown on the left. (**e**) Box plots for each quintile show that, with increasing Pol II half-life, RNA levels decrease and pausing indices increase (Wilcoxon test). Boxes refer to the first quartile, median and third quartile. Whiskers refer to the lowest and highest values within 1.5 times the interquartile range. RPKM, reads per kilobase per million mapped reads. (**f**) NELF-E ChIP-nexus profile under control and TRI treatment conditions. Higher levels of NELF-E are observed at promoters with longer half-lives for paused Pol II. (**g**) After NELF knockdown, the observed loss of Pol II correlates with our half-life measurements (Wilcoxon test). (**h**) Distinct core promoter elements are differentially enriched in the quintiles (Fisher's exact test with multiple-testing correction, $*P < 0.05$).

(**Fig. 2h**). Promoters with the most stably paused Pol II were significantly enriched for pause button (PB) and initiator (Inr), whereas promoters with the shortest Pol II half-lives were significantly enriched for the TCT motif and the TATA box (**Fig. 2h**). These results are consistent with previous observations that the presence of TATA can be associated with reduced Pol II pausing under some developmental conditions[42,43].

Minimal PIC at promoters with stably paused Pol II

We took advantage of the Pol II half-life measurements at promoters to investigate the relationship between Pol II pausing and initiation. If a new PIC assembles in the presence of paused Pol II ('preloading' of the next Pol II), the percentage of upstream Pol II should be independent of the pause duration and thus similar across all quintiles with different Pol II half-lives. Conversely, if Pol II cannot form a PIC in the presence of paused Pol II, one would expect an anticorrelation between upstream Pol II and pausing duration. This latter situation was exactly what we observed.

We found that, under steady-state conditions, the shorter the half-life of paused Pol II, the higher the percentage of upstream Pol II (**Fig. 3a**). This difference was clear when comparing the average Pol II profile for promoters with short half-lives (q1) to that of stably

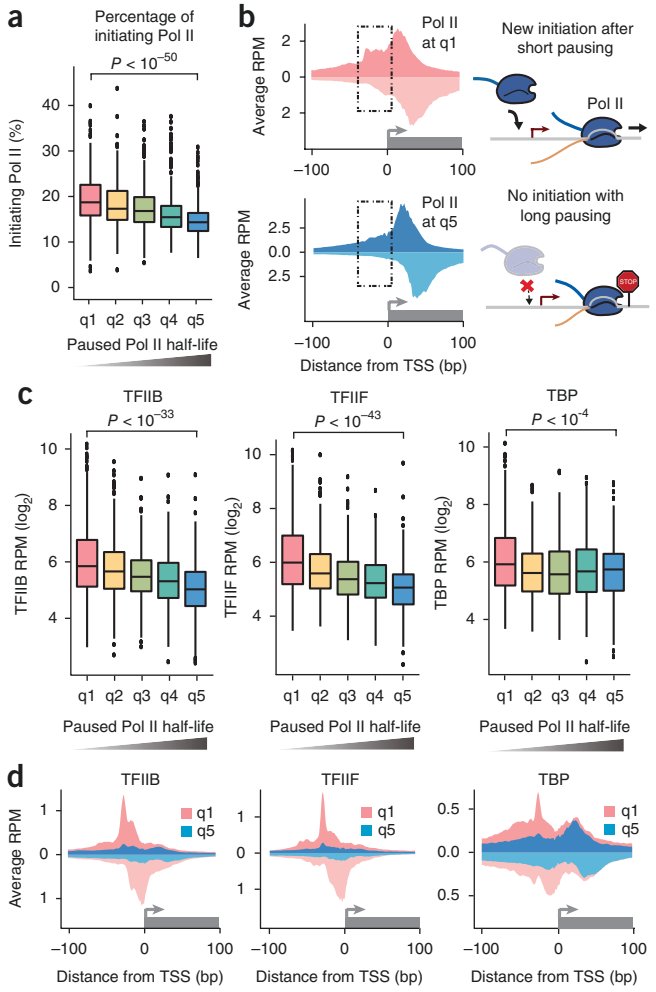

Figure 3 Promoters with stable Pol II pausing lack PICs but show downstream occupancy of TBP. (**a**) With the increasing Pol II half-life in each quintile, the percentage of initiating Pol II decreases (Wilcoxon test). (**b**) The average Pol II ChIP-nexus profile of promoters with short Pol II pausing (q1) shows initiating Pol II (dashed box), whereas the profile of those with stable Pol II pausing (q5) does not. (**c**) TFIIB and TFIIF levels decrease with increasing Pol II half-life in each quintile, while the levels of TBP remain roughly similar (Wilcoxon test). (**d**) The average ChIP-nexus profiles of TFIIB, TFIIF and TBP confirm that TFIIB and TFIIF are drastically reduced at promoters with stable Pol II pausing and show that TBP is preferentially found near the pause position at these promoters and not upstream near the TATA box where initiation occurs.

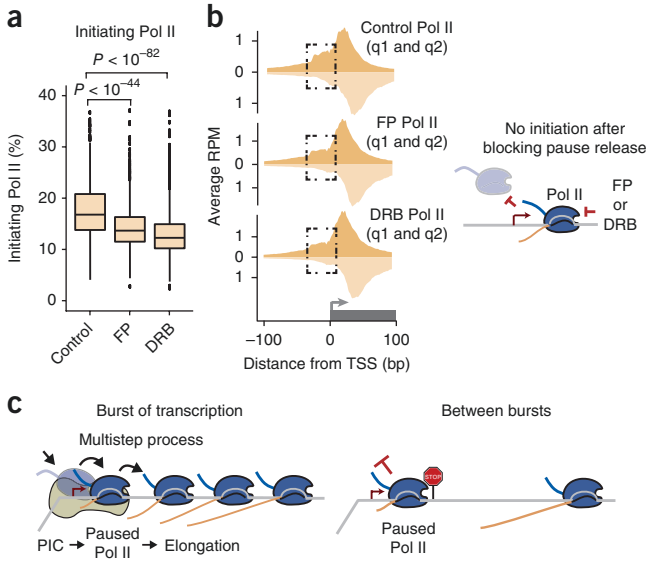

Figure 4 Paused Pol II inhibits new initiation. (**a**) After paused Pol II is stabilized with FP or DRB treatment, initiating Pol II is significantly reduced at all promoters (Wilcoxon test). (**b**) Promoters with short Pol II pausing (q1 and q2) show decreased initiating Pol II after FP or DRB treatment. (**c**) Model of the inhibitory role of paused Pol II. Left, during gene activation, the inhibitory role of paused Pol II ensures that initiation does not occur by default and thus makes sure that activation is a multistep process. Right, between transcription bursts, paused Pol II inhibits new transcriptional initiation.

paused promoters (q5). The profile for promoters in q1 showed a discernable footprint where Pol II forms a PIC and initiates, whereas that of q5 showed little evidence of Pol II initiation (**Fig. 3b**). Thus, the half-life of Pol II pausing and the occupancy of Pol II at the initiation site display an inverse correlation.

To investigate whether the binding of basal transcription factors was also lower in the presence of paused Pol II, we plotted the levels of each factor across the five quintiles (**Fig. 3c**). Similar to initiating Pol II itself, the levels of TFIIB and TFIIF decreased with longer half-lives of paused Pol II. Promoters with stably paused Pol II had much lower levels of TFIIB and TFIIF in their average profiles than promoters with short Pol II half-lives (**Fig. 3d**). This is consistent with biochemical experiments suggesting that TFIIB dissociates from the transcription machinery after initiation[44–46] and suggests that TFIIF is also released from stably paused Pol II.

In contrast to TFIIB and TFIIF, the levels of TBP did not decrease to the same extent across the five quintiles (**Fig. 3d**). However, stably paused promoters showed substantial loss of TBP from the upstream position, where initiation takes place. TBP was mostly observed at the downstream position, which is likely associated with later stages of transcription. Thus, the distribution of TBP also argues that stably paused promoters are not undergoing initiation.

Paused Pol II inhibits new initiation

The lack of Pol II initiation at stably paused promoters raises the possibility that paused Pol II blocks new Pol II from initiation. Alternatively, it could be the other way around, that Pol II pausing is stable because no new initiation takes place. To distinguish between these two possibilities, we treated cells with flavopiridol (FP) or 5,6-dichloro-1-β-D-ribofuranosyl benzimidazole (DRB), two kinase inhibitors that block pause release by P-TEFb with different mechanisms[47,48]. We confirmed that Pol II Ser2 phosphorylation was reduced without affecting total Pol II levels or Ser5 phosphorylation (**Supplementary Fig. 7**). This allowed us to analyze the effect of increasing the pause duration at promoters where new Pol II initiation normally occurs.

When we calculated the percentage of upstream Pol II after FP treatment, there was a strong decrease in comparison to the control (**Fig. 4a**). Furthermore, analysis of the Pol II profile at promoters with short half-lives showed that the upstream Pol II footprint present under normal conditions was markedly reduced after FP or DRB treatment (**Fig. 4b**). This suggests that stably paused Pol II blocks new transcriptional initiation.

DISCUSSION

Paused Pol II is thought to facilitate rapid and synchronous gene expression responses[12,13], but how this occurs is not clear. Our results rule out the model in which paused Pol II does so by allowing new

initiation to occur before pause release. Instead, we find that paused Pol II inhibits new initiation and that basal transcription factors such as TFIIB and TFIIF no longer occupy the site of initiation upon prolonged Pol II pausing. This suggests that initiation can only take place after paused Pol II is released.

The mechanism by which paused Pol II inhibits new initiation is not known. The minimum length of DNA that Pol II occupies has been estimated to be ~33 bp[49,50], and the site of PIC formation is ~47 bp upstream of paused Pol II in our data. Thus, we cannot conclude that the inhibition is simply because the two Pol II complexes would spatially constrain each other. However, our data also show that basal transcription factors involved in initiation (for example, TAF2) occupy a broader area of the promoter than previously shown by structural studies and that some of them, including TBP, appear to change their contacts during transcription. This raises the possibility that dynamic occupancy of basal transcription factors is the reason that paused Pol II interferes with initiation.

The reason we could detect the change in basal transcription factors between the initiation and pausing stages is that promoters were found to have a wide range of paused Pol II half-lives, with hundreds of them stable for over 30 min. Although we cannot rule out the possibility that TRI-based measurements lead to overestimations, the results are consistent with previous studies on individual promoters and explain why paused Pol II is the most prominent form detected by ChIP across the genome, especially in comparison to the much more transient PICs. Thus, paused Pol II appears to be generally very stable, with half-lives on the order of minutes to hours.

How does Pol II pausing fit into a framework where transcription occurs in bursts? It is becoming increasingly clear from imaging experiments that regulated metazoan genes, including those with paused Pol II, are transcribed in bursts[51,52]. Bursts are characterized by a rapid succession of numerous transcribing Pol II proteins, interspersed by periods of inactivity lasting from minutes to hours, during which the gene may be refractory to activation[53,54]. Paused Pol II cannot yet be detected in these assays, but the emerging properties of paused Pol II strongly suggest that it is present during the periods of inactivity between bursts of transcription.

The stability of paused Pol II appears to be on the order of minutes to hours, which is much longer than Pol II could pause during transcription bursts but in the same range as measured refractory periods[53,54]. Paused Pol II has been shown to prevent nucleosomes from forming over the promoter, especially at strongly paused promoters[14,55]. Nucleosomes reform in a period on the order of minutes to hours[56], presumably between prolonged periods of inactivity. This suggests that paused Pol II excludes nucleosome formation by remaining on the promoter between transcription bursts.

If paused Pol II is present during long periods of inactivity, its inhibitory effect on initiation could be beneficial. By preventing initiation, paused Pol II prevents basal transcription factors and further Pol II from being sequestered (**Fig. 4c**). Furthermore, paused Pol II itself might contribute to the refractory period. In this more speculative model, the presence of paused Pol II after a transcription burst is one of the mechanisms that prevents another transcription burst from following immediately afterward.

Theoretical considerations suggest that a multistep process of regulating transcriptional activation reduces noise by creating narrowly distributed refractory periods[54,57,58]. While a single rate-limiting step would produce periods of inactivity of exponential length, the presence of multiple steps that involve resetting before reactivation reduces the variability of the refractory periods between transcription bursts. We note that such a multistep activation process is consistent with our results (**Fig. 4c**). Namely, the inhibitory property of paused Pol II could ensure that initiation does not occur by default and that initiation and pause release are both regulated. After a transcription burst, paused Pol II could be one of the mechanisms that prevent immediate reactivation.

While our model is consistent with synchronous, less noisy expression, it does not fully explain the rapid inducibility of paused genes. As we ruled out permissive reinitiation in the presence of paused Pol II, we speculate that the inhibitory property of paused Pol II may allow the promoter to be in an otherwise permissive transcription state. For example, paused Pol II may not only keep the promoter accessible[14,43] but also allow engagement with enhancers[59]. In this manner, paused Pol II may block new initiation while at the same time keeping the promoter responsive to enhancer activation.

In summary, our results establish a clear relationship between Pol II pausing and initiation, thereby revising existing models for the role of Pol II in transcriptional regulation. They also illustrate how novel high-resolution genomics approaches such as ChIP-nexus can complement biochemical, structural and imaging approaches, thereby providing new avenues for the characterization of transcription stages such as paused Pol II that were previously challenging to study.

URLs. ChIP–seq and ChIP-nexus data, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE85741; analysis code, https://github.com/zeitlingerlab/Shao_NG_2017; instructions for accessing Linux virtual machine containing raw data, processed data, software tools and analysis scripts, http://research.stowers.org/zeitlingerlab/data.html; FlyBase RNA–seq data, ftp://ftp.flybase.net/releases/FB2014_03/precomputed_files/genes/gene_rpkm_report_fb_2014_03.tsv.gz.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank P. Verrijzer (Erasmus University Medical Center) for TAF2 antibodies, J. Kadonaga (University of California, San Diego) for TFIIA, TFIIB, TBP and TFIIF antibodies, J. Johnston for help with data analysis, and R. Krumlauf, R. Conaway, J. Conaway, C. Kaplan and R. Fropf for comments on the manuscript. The work was funded by the Stowers Institute for Medical Research.

AUTHOR CONTRIBUTIONS

W.S. and J.Z. conceived the project, W.S. performed all experiments and computational analyses, and W.S. and J.Z. interpreted the data and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Sainsbury, S., Bernecky, C. & Cramer, P. Structural basis of transcription initiation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16**, 129–143 (2015).
2. Darzacq, X. *et al.* Imaging transcription in living cells. *Annu. Rev. Biophys.* **38**, 173–196 (2009).
3. Nogales, E., Louder, R.K. & He, Y. Cryo-EM in the study of challenging systems: the human transcription pre-initiation complex. *Curr. Opin. Struct. Biol.* **40**, 120–127 (2016).
4. Conaway, R.C. & Conaway, J.W. General initiation factors for RNA polymerase II. *Annu. Rev. Biochem.* **62**, 161–190 (1993).
5. Roeder, R.G. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.* **21**, 327–335 (1996).

6. Tirode, F., Busso, D., Coin, F. & Egly, J.M. Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. *Mol. Cell* **3**, 87–95 (1999).
7. Holstege, F.C., van der Vliet, P.C. & Timmers, H.T. Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. *EMBO J.* **15**, 1666–1677 (1996).
8. Adelman, K. & Lis, J.T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13**, 720–731 (2012).
9. 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).
10. Hendrix, D.A., Hong, J.W., Zeitlinger, J., Rokhsar, D.S. & Levine, M.S. Promoter elements associated with RNA Pol II stalling in the *Drosophila* embryo. *Proc. Natl. Acad. Sci. USA* **105**, 7762–7767 (2008).
11. Nechaev, S. *et al.* Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in *Drosophila*. *Science* **327**, 335–338 (2010).
12. Lagha, M. *et al.* Paused Pol II coordinates tissue morphogenesis in the *Drosophila* embryo. *Cell* **153**, 976–987 (2013).
13. Boettiger, A.N. & Levine, M. Synchronous and stochastic patterns of gene activation in the *Drosophila* embryo. *Science* **325**, 471–473 (2009).
14. Gilchrist, D.A. *et al.* Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. *Cell* **143**, 540–551 (2010).
15. Zeitlinger, J. *et al.* RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nat. Genet.* **39**, 1512–1516 (2007).
16. Li, J. *et al.* Kinetic competition between elongation rate and binding of NELF controls promoter-proximal pausing. *Mol. Cell* **50**, 711–722 (2013).
17. Core, L.J., Waterfall, J.J. & Lis, J.T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848 (2008).
18. Kwak, H., Fuda, N.J., Core, L.J. & Lis, J.T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339**, 950–953 (2013).
19. Rhee, H.S. & Pugh, B.F. Comprehensive genome-wide protein–DNA interactions detected at single-nucleotide resolution. *Cell* **147**, 1408–1419 (2011).
20. Pugh, B.F. & Venters, B.J. Genomic organization of human transcription initiation complexes. *PLoS One* **11**, e0149339 (2016).
21. He, Q., Johnston, J. & Zeitlinger, J. ChIP-nexus enables improved detection of *in vivo* transcription factor binding footprints. *Nat. Biotechnol.* **33**, 395–401 (2015).
22. Titov, D.V. *et al.* XPB, a subunit of TFIIH, is a target of the natural product triptolide. *Nat. Chem. Biol.* **7**, 182–188 (2011).
23. Geiger, J.H., Hahn, S., Lee, S. & Sigler, P.B. Crystal structure of the yeast TFIIA/TBP/DNA complex. *Science* **272**, 830–836 (1996).
24. Tan, S., Hunziker, Y., Sargent, D.F. & Richmond, T.J. Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature* **381**, 127–151 (1996).
25. Imbalzano, A.N., Zaret, K.S. & Kingston, R.E. Transcription factor (TF) IIB and TFIIA can independently increase the affinity of the TATA-binding protein for DNA. *J. Biol. Chem.* **269**, 8280–8286 (1994).
26. He, Y. *et al.* Near-atomic resolution visualization of human transcription promoter opening. *Nature* **533**, 359–365 (2016).
27. Horn, A.E., Kugel, J.F. & Goodrich, J.A. Single molecule microscopy reveals mechanistic insight into RNA polymerase II preinitiation complex assembly and transcriptional activity. *Nucleic Acids Res.* **44**, 7132–7143 (2016).
28. Kostrewa, D. *et al.* RNA polymerase II–TFIIB structure and mechanism of transcription initiation. *Nature* **462**, 323–330 (2009).
29. Rhee, H.S. & Pugh, B.F. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483**, 295–301 (2012).
30. Plaschka, C. *et al.* Transcription initiation complex structures elucidate DNA opening. *Nature* **533**, 353–358 (2016).
31. Burley, S.K. & Roeder, R.G. Biochemistry and structural biology of transcription factor IID (TFIID). *Annu. Rev. Biochem.* **65**, 769–799 (1996).
32. Louder, R.K. *et al.* Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature* **531**, 604–609 (2016).
33. Verrijzer, C.P., Chen, J.L., Yokomori, K. & Tjian, R. Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell* **81**, 1115–1125 (1995).
34. Jonkers, I., Kwak, H. & Lis, J.T. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *eLife* **3**, e02407 (2014).
35. Chen, F., Gao, X. & Shilatifard, A. Stably paused genes revealed through inhibition of transcription initiation by the TFIIH inhibitor triptolide. *Genes Dev.* **29**, 39–47 (2015).
36. Grünberg, S., Warfield, L. & Hahn, S. Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nat. Struct. Mol. Biol.* **19**, 788–796 (2012).
37. Alekseev, S. *et al.* Transcription without XPB establishes a unified helicase-independent mechanism of promoter opening in eukaryotic gene expression. *Mol. Cell* **65**, 504–514 (2017).
38. Zhang, Z. *et al.* Rapid dynamics of general transcription factor TFIIB binding during preinitiation complex assembly revealed by single-molecule analysis. *Genes Dev.* **30**, 2106–2118 (2016).
39. Darzacq, X. *et al. In vivo* dynamics of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **14**, 796–806 (2007).
40. Henriques, T. *et al.* Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals. *Mol. Cell* **52**, 517–528 (2013).
41. Buckley, M.S., Kwak, H., Zipfel, W.R. & Lis, J.T. Kinetics of promoter Pol II on *Hsp70* reveal stable pausing and key insights into its regulation. *Genes Dev.* **28**, 14–19 (2014).
42. Chen, K. *et al.* A global change in RNA polymerase II pausing during the *Drosophila* midblastula transition. *eLife* **2**, e00861 (2013).
43. Gaertner, B. *et al.* Poised RNA polymerase II changes over developmental time and prepares genes for future expression. *Cell Reports* **2**, 1670–1683 (2012).
44. Zawel, L., Kumar, K.P. & Reinberg, D. Recycling of the general transcription factors during RNA polymerase II transcription. *Genes Dev.* **9**, 1479–1490 (1995).
45. Yudkovsky, N., Ranish, J.A. & Hahn, S. A transcription reinitiation intermediate that is stabilized by activator. *Nature* **408**, 225–229 (2000).
46. Van Dyke, M.W., Sawadogo, M. & Roeder, R.G. Stability of transcription complexes on class II genes. *Mol. Cell. Biol.* **9**, 342–344 (1989).
47. Chao, S.H. & Price, D.H. Flavopiridol inactivates P-TEFb and blocks most RNA polymerase II transcription *in vivo*. *J. Biol. Chem.* **276**, 31793–31799 (2001).
48. Marshall, N.F. & Price, D.H. Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *J. Biol. Chem.* **270**, 12335–12338 (1995).
49. Saeki, H. & Svejstrup, J.Q. Stability, flexibility, and dynamic interactions of colliding RNA polymerase II elongation complexes. *Mol. Cell* **35**, 191–205 (2009).
50. Ehrensberger, A.H., Kelly, G.P. & Svejstrup, J.Q. Mechanistic interpretation of promoter-proximal peaks and RNAPII density maps. *Cell* **154**, 713–715 (2013).
51. Bothma, J.P. *et al.* Dynamic regulation of *eve* stripe 2 expression reveals transcriptional bursts in living *Drosophila* embryos. *Proc. Natl. Acad. Sci. USA* **111**, 10598–10603 (2014).
52. Fukaya, T., Lim, B. & Levine, M. Enhancer control of transcriptional bursting. *Cell* **166**, 358–368 (2016).
53. Suter, D.M. *et al.* Mammalian genes are transcribed with widely different bursting kinetics. *Science* **332**, 472–474 (2011).
54. Zoller, B., Nicolas, D., Molina, N. & Naef, F. Structure of silent transcription intervals and noise characteristics of mammalian genes. *Mol. Syst. Biol.* **11**, 823 (2015).
55. Gilchrist, D.A. *et al.* NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. *Genes Dev.* **22**, 1921–1933 (2008).
56. Deal, R.B., Henikoff, J.G. & Henikoff, S. Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science* **328**, 1161–1164 (2010).
57. Pedraza, J.M. & Paulsson, J. Effects of molecular memory and bursting on fluctuations in gene expression. *Science* **319**, 339–343 (2008).
58. Coulon, A., Chow, C.C., Singer, R.H. & Larson, D.R. Eukaryotic transcriptional dynamics: from single molecules to cell populations. *Nat. Rev. Genet.* **14**, 572–584 (2013).
59. Ghavi-Helm, Y. *et al.* Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**, 96–100 (2014).

ONLINE METHODS

Cell culture and transcription inhibitor treatment. *Drosophila* Kc167 cells (purchased from DGRC; negative for mycoplasma contamination) were grown at 25 °C in HyClone SFX-Insect Cell Culture Media. Transcription inhibitors were added directly into SFX media. Cells were treated with 500 μM TRI (TOCRIS Bioscience, 3253; dissolved in DMSO), 500 nM FP (Santa Cruz Biotechnology, sc-202157; in DMSO) or 50 μM DRB (Sigma, D1916; in DMSO) at room temperature for 1 h. Equivalent amounts of DMSO treatment (2%) were used as the control.

ChIP-nexus. For each ChIP-nexus experiment, 1×10^7 Kc167 cells were fixed with 1% formaldehyde in SFX media at room temperature for 10 min. Fixed cells were washed with cold PBS, incubated with Orlando and Paro's Buffer A (0.25% Triton X-100, 10 mM EDTA, 0.5 mM EGTA, 10 mM Tris-HCl, pH 8.0) for 10 min at room temperature with rotation, and then centrifuged and resuspended in S2-RIPA buffer (10 mM Tris-HCl, pH 8.0, 140 mM NaCl, 0.1% SDS, 0.1% sodium deoxycholate, 0.5% sarkosyl, 1% Triton X-100). Sonication was performed with a Bioruptor Pico for four rounds of 30 s on and 30 s off. The ChIP-nexus procedure and data processing were performed as previously published[21], except that ChIP-nexus adaptors with a single fixed barcode (CTGA) or mixed fixed barcodes (CTGA, TGAC, GACT, ACTG) were used during adaptor ligation. Two biological replicates were performed for each factor and condition. ChIP-nexus samples were used for further analysis if more than 5 million non-duplicated reads could be aligned to the genome and ChIP-nexus signal was enriched at promoter regions (**Supplementary Table 2**).

Antibodies. For each ChIP-nexus experiment, 10 μg of antibody was coupled to 100 μl of Dynabeads coupled to protein A or protein G (Thermo Fisher, 10008D). The following antibodies were used: rabbit polyclonal antibodies against full-length *Drosophila* TFIIA, TFIIB, TFIIF and TBP (a gift from J. Kadonaga, University of California, San Diego), guinea pig polyclonal antibodies against *Drosophila* TAF2 (a gift from P. Verrijzer, Erasmus University Medical Center), rabbit polyclonal antibodies against human XPB (Santa Cruz Biotechnology, sc-293X), rabbit polyclonal antibodies against full-length *Drosophila* NELF-E (custom made from GeneScript, Zeitlinger lab 126740-7) and rabbit polyclonal antibodies against the full-length *Drosophila* Pol II subunit RPB3 (custom made from GeneScript, Zeitlinger lab 163185-50). For the spike-in control of human chromatin, rabbit polyclonal antibodies against human Pol II were used (N20, Santa Cruz Biotechnology, sc-899X). For immunoblots, mouse monoclonal antibodies against α-tubulin (Sigma, T9026) were used as loading control.

NELF knockdown. *Drosophila* Kc167 cells were treated with double-stranded RNA (dsRNA) against eGFP or the NELF-E and NELF-B subunits. dsRNA was generated from PCR templates carrying two T7 RNA polymerase promoter sequences (5′-TAATACGACTCACTATAGG-3′) on each end. dsRNA was prepared using the MEGAscript RNAi kit (Fisher, AM 1626). For each knockdown experiment, 1×10^6 cells were seeded per well in a six-well plate and incubated at 25 °C for 30 min to allow cells to attach to the bottom of the plate. Cells were treated with 10 μg of dsRNA and incubated for 24 h. The cells were again treated with 10 μg of dsRNA and collected 48 h later to study the phenotypes of NELF depletion. The primers used to amplify PCR templates from *Drosophila* genomic DNA or plasmid containing eGFP are listed in **Supplementary Table 3**.

RT–qPCR. 5×10^5 *Drosophila* Kc167 cells after NELF or eGFP knockdown were transferred to a 1.5-ml tube and washed with cold PBS. Total RNA was extracted and purified with the Direct-zol RNA MiniPrep kit (Genesee, 11-330). cDNA was generated with the High-Capacity RNA-to-cDNA kit (Thermo Fisher, 4387406). qPCR was then performed with Fast SYBR Green Master Mix (Thermo Fisher, 4385610) and primers against NELF-E and NELF-B.

Immunoblotting. 5×10^5 transcription-inhibitor- or dsRNA-treated Kc167 cells were transferred to a 1.5-ml tube and washed with cold PBS. Cells were then lysed with S2-RIPA buffer (10 mM Tris-HCl, pH 8.0, 140 mM NaCl, 0.1% SDS, 0.1% sodium deoxycholate, 0.5% sarkosyl, 1% Triton X-100). Primary antibodies were used at a 1:2,000 dilution.

Filtering and reannotating *Drosophila* TSSs. TSSs from FlyBase protein-coding genes (fb-r5.47) were reannotated to match a nearby PRO-cap (GSM1032759) peak summit if within 150 bp of it. Original TSS annotations from FlyBase were preserved if no PRO-cap peak was detected. If a TSS was found within 300 bp of another TSS, both TSSs were removed from the promoter set. This set of reannotated TSSs (n = 14,229) was used for subsequent analyses.

Mapping basal transcription factor footprints. Basal transcription factors and Pol II ChIP-nexus data were processed and aligned to the dm3 genome. Only the nucleotides where the exonuclease stopped digestion (a single stop base for each read) were kept, and no data smoothing was performed. The top 1,000 promoters with the highest TFIIB signal were selected by calculating the sum of the TFIIB ChIP-nexus signal (under normal conditions) within a 201-bp window centered on the TSS. For each factor, two ChIP-nexus replicates were merged after read count normalization. The average ChIP-nexus signal for the top 1,000 promoters was calculated in RPM after aligning to the TSS. The position of a footprint was defined as the middle point between the positive- and negative-strand summits (rounded down if between bases).

Paused Pol II half-life measurement. Kc167 cells were treated with TRI at room temperature for 5, 10, 15 and 30 min and 2% DMSO as a control. Cells were fixed and prepared for ChIP-nexus as referenced above, except that a fixed amount of spike-in was added to each ChIP to control for the loss of Pol II ChIP signal after TRI treatment (there was about 8% loss of Pol II signal after 30 min of TRI treatment). For the spike-in control, human chromatin extracts from the GM12878 cell line were incubated with Dynabeads coupled to antibodies against human Pol II (N20, Santa Cruz Biotechnology, sc-899X), washed four times (Buffer A-D). The mixture of Dynabeads and Pol II ChIP was then added equally to each Kc167 ChIP-nexus experiment. After sequencing, spiked-in samples were aligned to a dm3–hg19 combined genome. Only reads that uniquely aligned to each genome were used for analysis, and PCR duplicates with the same ChIP-nexus barcode were removed. Each *Drosophila* sample was then read count normalized on the basis of the ratio between human and *Drosophila* reads. Two biological replicates were performed for each condition.

To analyze the half-lives of paused Pol II from these data, promoters were selected if (i) the total Pol II signal in both the control and FP-treated conditions was high (top 25%), (ii) a typical footprint was observed for Pol II (distance between positive- and negative-strand peak < 80 bp) and (iii) the position of the Pol II footprint was less than 80 bp downstream of the TSS; 2,329 promoters fulfilled these criteria. For each promoter, the Pol II signal was calculated in a 41-bp window centered on the pausing position (the midpoint between Pol II–positive and Pol II–negative summits). To calculate the half-life of paused Pol II at each promoter, the Pol II time-course measurements were fitted into an exponential decay model using nonlinear regression. 1,798 promoters had a paused Pol II half-life shorter than 60 min, and promoters with a paused Pol II half-life of longer than 60 min (n = 531) were floored to 60 min to eliminate inflated values due to noise. The 2,329 promoters were then ranked and divided into five quintiles. For display purposes (heat map in **Fig. 2b**), the Pol II measurements at each promoter were normalized to the maximum Pol II signal under the control condition.

Promoter element enrichments. In **Figure 2h**, known *Drosophila* promoter elements in each promoter were identified by the presence of the known consensus sequence with zero mismatches in a specified window relative to the TSS (**Supplementary Table 4**). For each promoter quintile and each promoter element, the enrichment was calculated by determining the ratio between the fraction of promoter element in the quintile and the fraction of the same promoter element in the other four quintiles. The significance for the observed frequencies was calculated with Fisher's exact test and corrected for multiple testing with the Benjamini–Hochberg method.

Other calculations. Total Pol II signal (**Fig. 2f**) was calculated in a 201-bp window around the TSS. The pausing index (**Fig. 2f**) was calculated as the average amount of Pol II ChIP-nexus signal per base pair in a 41-bp window centered on the pausing position divided by the average amount of Pol II signal

doi:10.1038/ng.3867

per base pair in a 101-bp window located 200 bp downstream of the pausing position. Initiating Pol II (**Figs. 3a** and **4a**) was defined as the amount of Pol II signal in a 41-bp window upstream of the TSS divided by the total Pol II signal in a 201-bp window around the TSS. The basal transcription factor levels at each promoter were calculated as the number of ChIP-nexus reads in a 201-bp window centered on the TSS. *P* values in **Figures 2e**,**g**, **3a**,**c** and **4a** were calculated with a two-sample Wilcoxon test.

Public data sets used for analysis. The human PIC structure (5IY6)[26] and TFIID structure (5FUR)[32] were downloaded from the RCSB Protein Data Bank and superimposed using UCSF CHIMERA[60]. Relevant factors are shown and colored appropriately. Kc167 RNA–seq data were downloaded from FlyBase (see URLs). PRO-cap (GSM1032759)[18], human TBP ChIP-nexus (GSE55306)[21] and human TBP ChIP–exo (SRA067908)[20] data were downloaded from GEO and the Sequence Read Archive.

Data availability. ChIP–seq and ChIP-nexus data sets have been deposited in GEO under accession GSE85741. All data analysis performed in this paper, including raw data, processed data, software tools and analysis scripts, is reproduced in a publically accessible Linux virtual machine. Instructions for accessing the virtual machine can be found at http://research.stowers.org/zeitlingerlab/data.html. The analysis code is available on GitHub at https://github.com/zeitlingerlab/Shao_NG_2017.

60. Pettersen, E.F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).