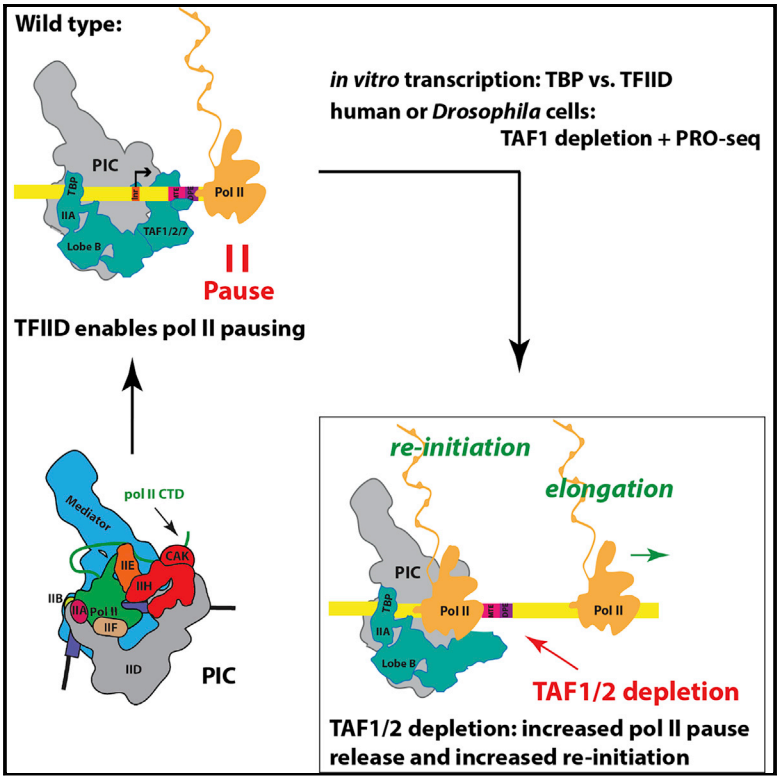
Molecular Cell

Short Article

TFIID Enables RNA Polymerase II Promoter-Proximal Pausing

Graphical Abstract



Highlights

- Reconstitution of RNAPII pausing *in vitro*, with purified factors (no extracts)
- Human PIC alone is sufficient to establish RNAPII pausing
- TFIID is required for RNAPII promoter-proximal pausing
- Rapid TFIID depletion induces RNAPII pause release genome-wide

Authors

Charli B. Fant, Cecilia B. Levandowski, Kapil Gupta, ..., Imre Berger, Robin D. Dowell, Dylan J. Taatjes

Correspondence

taatjes@colorado.edu

In Brief

Fant and co-workers use biochemical reconstitution to discover that the general transcription factor TFIID is required to establish RNAPII promoter-proximal pausing; canonical regulatory factors NELF and DSIF enhance pausing but are not required. Follow-up experiments in human and *Drosophila* cells show TFIID is a conserved regulator of RNAPII pausing.

Fant et al., 2020, Molecular Cell *78*, 785–793
May 21, 2020
https://doi.org/10.1016/j.molcel.2020.03.008

CellPress

Molecular Cell

CellPress

Short Article

TFIID Enables RNA Polymerase II Promoter-Proximal Pausing

Charli B. Fant,[1] Cecilia B. Levandowski,[1] Kapil Gupta,[2] Zachary L. Maas,[1] John Moir,[1] Jonathan D. Rubin,[1] Andrew Sawyer,[3] Meagan N. Esbin,[1] Jenna K. Rimel,[1] Olivia Luyties,[1] Michael T. Marr,[3] Imre Berger,[2] Robin D. Dowell,[4,5] and Dylan J. Taatjes[1,6,*]

[1]Department of Biochemistry, University of Colorado, Boulder, CO, USA
[2]School of Biochemistry, Bristol Research Centre for Synthetic Biology, University of Bristol, Bristol, UK
[3]Department of Biology, Brandeis University, Waltham, MA, USA
[4]Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, CO, USA
[5]BioFrontiers Institute, University of Colorado, Boulder, CO, USA
[6]Lead Contact
*Correspondence: taatjes@colorado.edu
https://doi.org/10.1016/j.molcel.2020.03.008

SUMMARY

RNA polymerase II (RNAPII) transcription is governed by the pre-initiation complex (PIC), which contains TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH, RNAPII, and Mediator. After initiation, RNAPII enzymes pause after transcribing less than 100 bases; precisely how RNAPII pausing is enforced and regulated remains unclear. To address specific mechanistic questions, we reconstituted human RNAPII promoter-proximal pausing *in vitro*, entirely with purified factors (no extracts). As expected, NELF and DSIF increased pausing, and P-TEFb promoted pause release. Unexpectedly, the PIC alone was sufficient to reconstitute pausing, suggesting RNAPII pausing is an inherent PIC function. In agreement, pausing was lost upon replacement of the TFIID complex with TATA-binding protein (TBP), and PRO-seq experiments revealed widespread disruption of RNAPII pausing upon acute depletion (t = 60 min) of TFIID subunits in human or *Drosophila* cells. These results establish a TFIID requirement for RNAPII pausing and suggest pause regulatory factors may function directly or indirectly through TFIID.

INTRODUCTION

RNA polymerase II (RNAPII) transcribes all protein-coding and many non-coding RNAs in the human genome. RNAPII transcription initiation occurs within the pre-initiation complex (PIC), which contains TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH, RNAPII, and Mediator. After initiation, RNAPII enzymes typically pause after transcribing 20–80 bases (Kwak and Lis, 2013), and paused polymerases represent a common regulatory intermediate (Core et al., 2008; Jonkers et al., 2014; Muse et al., 2007; Zeitlinger et al., 2007). Accordingly, paused RNAPII has been implicated in enhancer function (Ghavi-Helm et al., 2014; Henriques et al., 2018), development and homeostasis (Adelman et al., 2009; Lagha et al., 2013), and diseases ranging from cancer (Lin et al., 2010; Miller et al., 2017) to viral pathogenesis (Wei et al., 1998; Yamaguchi et al., 2001). Precisely how RNAPII promoter-proximal pausing is enforced and regulated remains unclear; however, protein complexes, such as NELF and DSIF, increase pausing, whereas the activity of CDK9 (P-TEFb complex) correlates with pause release (Kwak and Lis, 2013).

Although much has been learned about RNAPII promoter-proximal pausing and its regulation, the underlying molecular mechanisms remain enigmatic. One reason for this is the complexity of the human RNAPII transcription machinery, which includes the ~4.0 MDa PIC and many additional regulatory factors. Another underlying reason is that much current understanding derives from cell-based assays, which are indispensable but cannot reliably address mechanistic questions. For instance, factor knockdowns or knockouts cause unintended secondary effects and the factors and biochemicals present at each gene in a population of cells cannot possibly be defined. *In vitro* assays can overcome such limitations, but these have typically involved nuclear extracts, which contain a similarly undefined mix of proteins, nucleic acids, and biochemicals. To circumvent these issues, we sought to reconstitute RNAPII promoter-proximal pausing entirely from purified human factors (no extracts). Success with this task enabled us to address some basic mechanistic questions and opens the door for future studies to better define the contribution of specific factors in RNAPII promoter-proximal pause regulation.

RESULTS

Biochemical Reconstitution Reveals Human PIC Is Sufficient to Establish RNAPII Pausing

Past results in *Drosophila* and mammalian cells and extracts implicated the NELF, DSIF, and P-TEFb complexes as regulators of RNAPII pausing (Core et al., 2012; Li et al., 2013; Marshall and Price, 1992). We purified these factors in addition to the PIC factors TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH, Mediator, and RNAPII

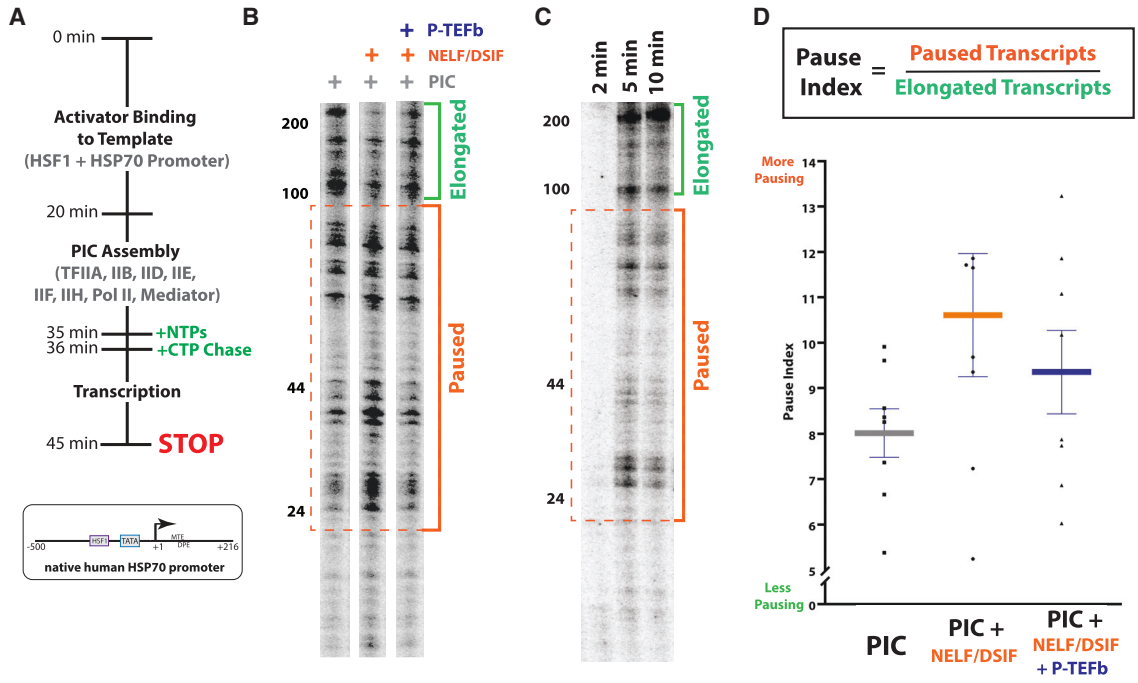


Figure 1. Biochemical Reconstitution of Promoter-Proximal RNAPII Pausing with Purified Human Factors

(A) Overview of *in vitro* transcription assay.

(B) Representative data from *in vitro* transcription reactions with the complete PIC (lane 1) or supplemented with NELF/DSIF (lane 2) and P-TEFb (lane 3). At left are approximate lengths of the RNA transcripts. Note increased transcripts in paused region and reduced transcripts in elongation region upon addition of NELF and DSIF (lane 1 versus lane 2); addition of P-TEFb reverses this trend (lane 2 versus lane 3).

(C) A time course showing maximum paused transcripts at 5 min and maximum elongated transcripts at 10 min; the elongated transcripts increased in equal proportion to the decrease in paused transcripts (5 min versus 10 min), indicating that paused complexes can ultimately generate elongated products.

(D) Calculation of an *in vitro* pause index (PI) at the HSP70 promoter, using paused and elongated regions defined in (B). A NELF/DSIF outlier (PI = 18) is not shown. Across replicate experiments (n = 8), PI almost exclusively increased with NELF/DSIF and decreased with added P-TEFb. The spread in the data is consistent with replicate-to-replicate PI variability observed in human cells (see STAR Methods). Bars represent mean ± standard error.

See also Figure S2E.

(Figure S1). Experiments were completed with the native human HSP70 promoter (*HSPA1B* gene), because others have shown that it is a quintessential model for promoter-proximal RNAPII pausing (Core et al., 2012). Because chromatin per se does not appear to be an essential regulator of RNAPII pausing in *Drosophila* or mammalian cells (Kwak et al., 2013; Lai and Pugh, 2017; Li et al., 2013), the *in vitro* transcription assays were completed on naked DNA templates (also see below).

Using purified PIC factors, primer extension assays established that transcription initiation occurred at the annotated HSPA1B start site *in vitro* (Figure S2A), as expected. An overview of the transcription assay is shown in Figure 1A, which was based in part upon *in vitro* pausing assays with nuclear extracts (Marshall and Price, 1992; Qiu and Gilmour, 2017; Renner et al., 2001). Following PIC assembly, transcription was initiated by adding ATP, guanosine triphosphate (GTP), and uridine triphosphate (UTP) at physiologically relevant concentrations, with a low concentration of cytidine triphosphate (CTP), primarily ^{32}P-CTP. After 1 min, reactions were chased with a physiologically relevant concentration of cold CTP and transcription was allowed to proceed for an additional 9 min. These "pulse-chase" assays allow better detection of short (potentially paused) transcripts, which otherwise would be drowned out by elongated transcripts that invariably possess more incorporated ^{32}P-C bases. By directly labeling all transcripts with ^{32}P-CTP, the method is highly sensitive and allowed detection of transcripts of varied lengths; furthermore, the ^{32}P-CTP pulse-chase protocol ensured that ^{32}P-labeled transcripts resulted almost exclusively from single-round transcription (see STAR Methods). Control experiments confirmed that transcripts detected were driven by the HSP70 promoter (e.g., not any contaminating nucleic acid) and that transcription was dependent on added PIC factors, as expected (Figure S2B).

A variety of methods have established that RNAPII pauses after transcribing 20–80 bases in *Drosophila* and mammalian cells (Jonkers et al., 2014; Kwak et al., 2013; Lai and Pugh, 2017; Lee et al., 2008; Li et al., 2013; Muse et al., 2007; Nechaev et al., 2010; Zeitlinger et al., 2007). The HSPA1B promoter sequence used in our assays extended 216 base pairs beyond the transcription start site (TSS); thus, elongated transcripts would migrate on a sequencing gel between 100 and 216 nt and paused transcripts would be observed between 20 and 80 nt.

Prior to testing DSIF/NELF and P-TEFb, we completed experiments with the PIC alone. As expected, elongated transcripts were prevalent; however, we observed short transcripts, between 20 and 80 nt, consistent with promoter-proximal RNAPII pausing (Figure 1B, lane 1). Potentially, these short

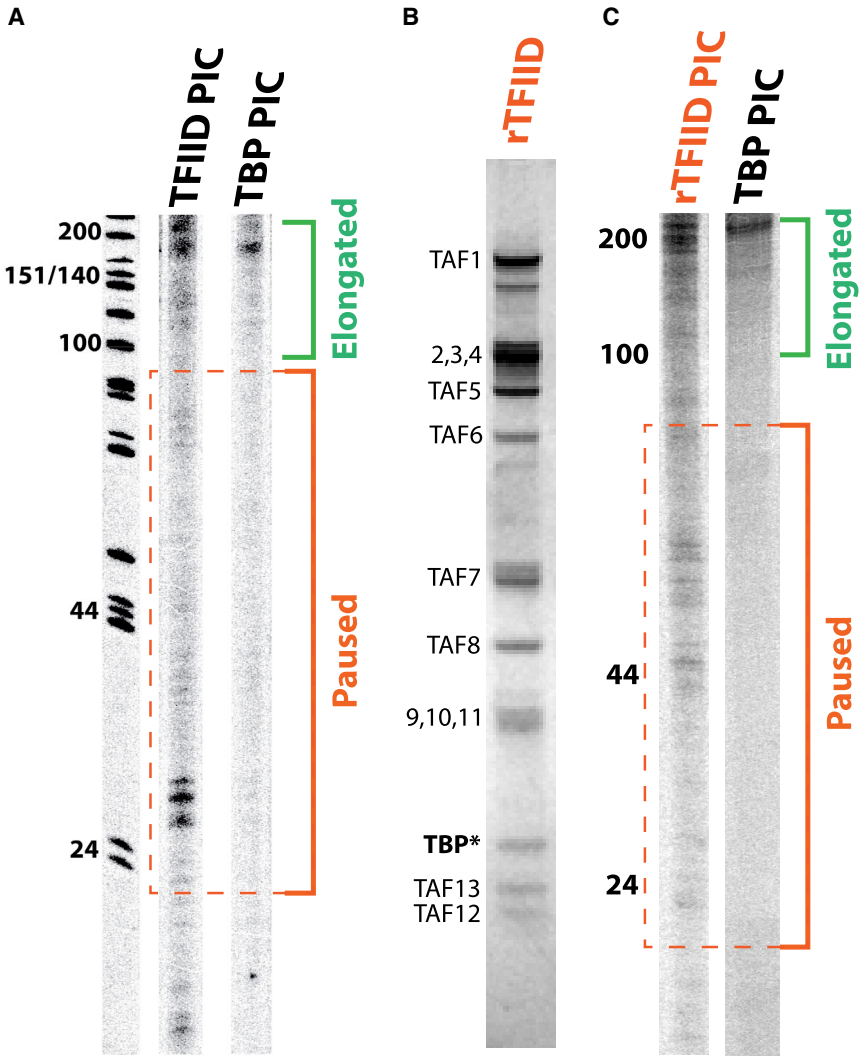


Figure 2. TFIID Is Required to Establish RNAPII Promoter-Proximal Pausing *In Vitro*

(A) Reconstituted transcription reactions with PICs containing TFIID or TBP (i.e., reactions contain TFIIA, IIB, IIE, IIF, IIH, Mediator, and RNAPII plus either TFIID or TBP). PICs with TBP still support transcription, but paused products were not observed (data from same gel). Reduced [ATP] was used in these experiments compared to Figure 1, causing an upstream shift in pausing, similar to results seen in *Drosophila* (Li et al., 2013).

(B) Coomassie-stained gel of the complete human TFIID complex, generated by recombinant expression (Fitzgerald et al., 2006); *core TBP (residues 155–335).

(C) As with endogenously purified human TFIID, PICs with recombinant TFIID support transcription and RNAPII promoter-proximal pausing. Relative to (A), increased [ATP] caused a downstream shift in paused products.

transcripts could reflect premature termination. However, time course experiments showed that the shorter transcripts build up and then release over time (Figure 1C). In fact, the increase in elongated products between 5 and 10 min was equal to the loss of pause signal between 5 and 10 min, suggesting a transient pause followed by release into elongation (see Discussion).

Addition of NELF/DSIF to the reconstituted transcription system increased the levels of the short transcripts (20–80 nt) while decreasing the elongated products (Figure 1B, lane 2); these data were consistent with established roles for NELF/DSIF in RNAPII pausing (Kwak and Lis, 2013) and further suggested that the short transcripts represented promoter-proximal paused products. Addition of P-TEFb to reactions containing NELF/DSIF largely reversed the promoter-proximal pausing induced by NELF/DSIF (Figure 1B, lane 3); thus, P-TEFb appeared to increase RNAPII pause release *in vitro*, also consistent with current models (Kwak and Lis, 2013). A pause index (PI) was calculated and averaged across replicate experiments (n = 8; Figure 1D), which showed that NELF/DSIF increased PI, whereas P-TEFb decreased PI, as expected.

RNAPII Promoter-Proximal Pausing Requires TFIID

Because we were able to recapitulate pause enhancement with NELF/DSIF and pause release with P-TEFb at the native human HSPA1B promoter, this *in vitro* system appeared to reliably reconstitute basic mechanistic aspects of RNAPII promoter-proximal pausing. Whereas many potential questions could be addressed with this system, we focused on the unexpected result that promoter-proximal pausing was recapitulated with the PIC alone. We next tested whether RNAPII pausing would be dependent on a specific PIC factor. Although some factors could not be reliably evaluated given their requirement for transcription in this assay, removal of TFIIA, TFIIH, HSF1, or Mediator still supported transcription *in vitro*, although at reduced levels. These experiments showed little change in PI, suggesting that these factors were not required for RNAPII pausing in this assay (Figure S2C).

We also addressed a potential dependence on the large, multisubunit TFIID complex. Whereas RNAPII transcription was not supported by removal of TFIID, TBP can substitute for TFIID *in vitro*, provided that the DNA templates are not assembled into chromatin (Näär et al., 1998). Strikingly, we observed that, when PICs were assembled with TBP instead of TFIID, transcription still occurred but promoter-proximal RNAPII pausing was lost (Figure 2A). These data implicated TFIID as a key PIC factor that enabled RNAPII promoter-proximal pausing. To test further, we replaced endogenous purified human TFIID with a complete TFIID complex generated by recombinant expression (Figure 2B). As shown in Figure 2C, the recombinant human TFIID complex performed similarly to endogenous TFIID, confirming that TFIID was required for RNAPII promoter-proximal pausing *in vitro*.

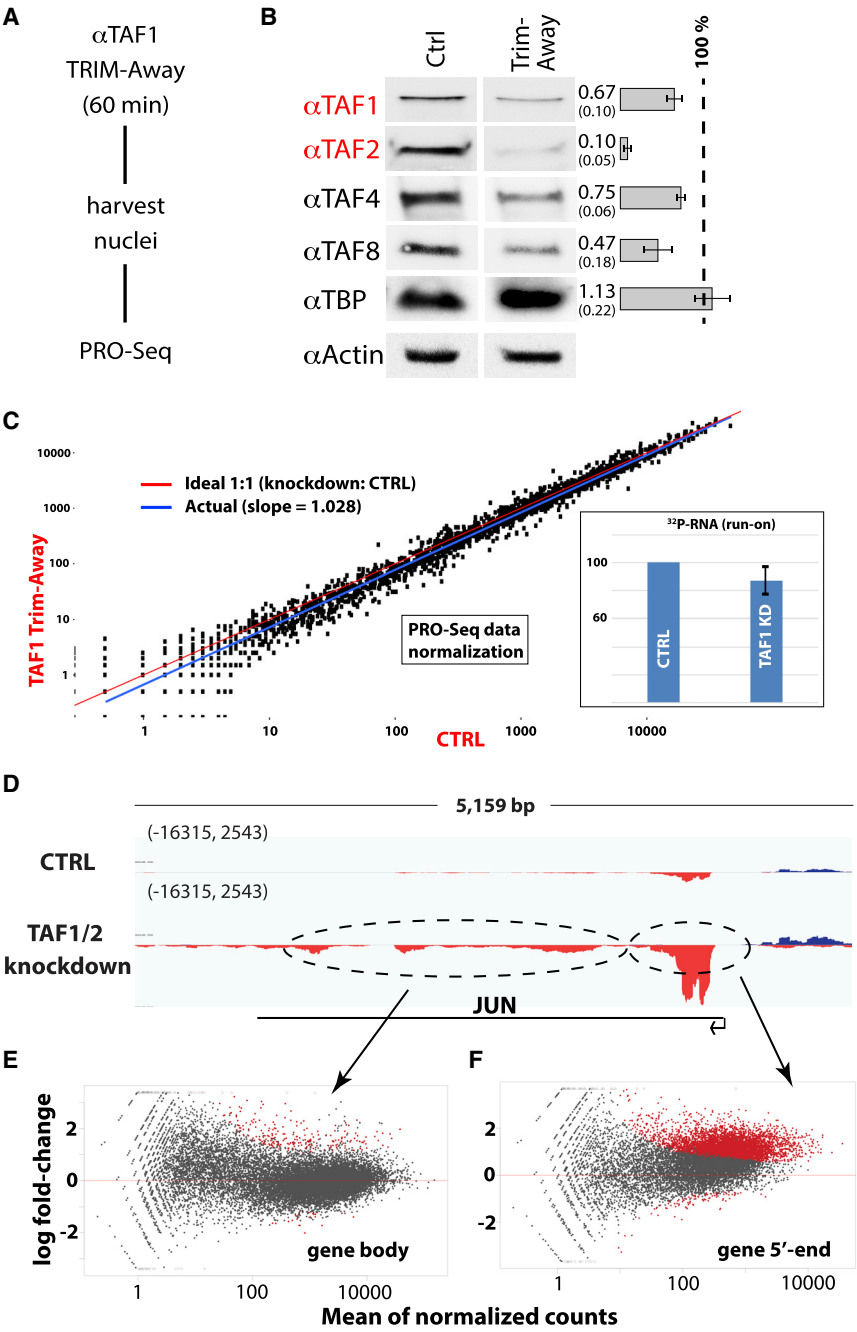


Figure 3. PRO-Seq Data Suggest Increased Pause Release upon TAF1/2 Knockdown in Human Cells

(A) Workflow for Trim-Away (Clift et al., 2017) and PRO-seq (Kwak et al., 2013).

(B) Representative western blots and quantitation (at right) for TFIID subunits. Bar plots represent mean and standard error, with actin as a loading control.

(C) Normalization of PRO-seq data based upon 3′ end reads of long genes (Mahat et al., 2016; see STAR Methods) or based upon total run-on signal comparing control versus TAF1/2-knockdown samples (inset; bar = SEM).

(D–F) Genome browser view (D) of PRO-seq reads at JUN locus (control versus TAF1/2-depleted), which reflects genome-wide trends shown in MA plots for gene bodies (+500 to −500 from transcription end site [TES]; E) and gene 5′ ends (−500 to +500; F).

Having established a TFIID dependence for RNAPII pausing, we sought to determine whether this activity could be attributed to any specific TFIID subunits. Human TFIID is approximately 1.4 MDa in size and contains TBP plus 13 different TBP-associated factors (TAFs), which are present in one or two copies each. The structures of human TFIID bound to promoter DNA reveal that lobe C—containing TAF1, TAF2, and TAF7—binds downstream DNA (Louder et al., 2016; Patel et al., 2018). In particular, TAF1/2 interact with the downstream promoter element (DPE) and the motif ten element (MTE). At the HSPA1B promoter, these elements reside at template position +18 to +33 relative to the TSS (Figure S1; Vo Ngoc et al., 2017). Because the DPE and MTE encompass part of the RNAPII pause region, we hypothesized that lobe C subunits might be important in regulation of RNAPII promoter-proximal pausing. To test this hypothesis, we expressed and purified an "S-TAF" TFIID complex that contained only a subset of TAFs (Figure S2D). The S-TAF complex contains TBP as well as lobe C subunits TAF1 and TAF7. As shown in Figure S2D, the S-TAF complex was able to support pausing, implying a role for TFIID lobe C for this function.

Rapid Depletion of TFIID Lobe C Subunits Increases RNAPII Pause Release in Human Cells

To further test the hypothesis that TFIID enables RNAPII promoter-proximal pausing, we turned to cell-based assays. To circumvent confounding issues with prolonged knockdown of essential TFIID subunits, we utilized the Trim-Away method (Clift et al., 2017), which enabled rapid (t = 60 min) TAF subunit depletion (Figures 3A, 3B, and S3). (Numerous antibodies to various TAF subunits were tested prior to identification of a TAF1 antibody that reliably immunoprecipitated TFIID from extracts and depleted TAF1 using the Trim-Away protocol.) With this approach, the effect of TFIID could be evaluated with minimal compensatory or cytotoxic consequences. Indicative of a direct TAF1-TAF2 interaction in lobe C, Trim-Away experiments targeting TAF1 also depleted TAF2 (TAF7 was not probed due to lack of reliable antibodies), and other TFIID subunits were depleted to varying degrees, except for TBP (Figures 3B and S3). Because Trim-Away works through lysine modification by an E3 ubiquitin ligase, the enhanced TAF2 depletion versus TAF1 may

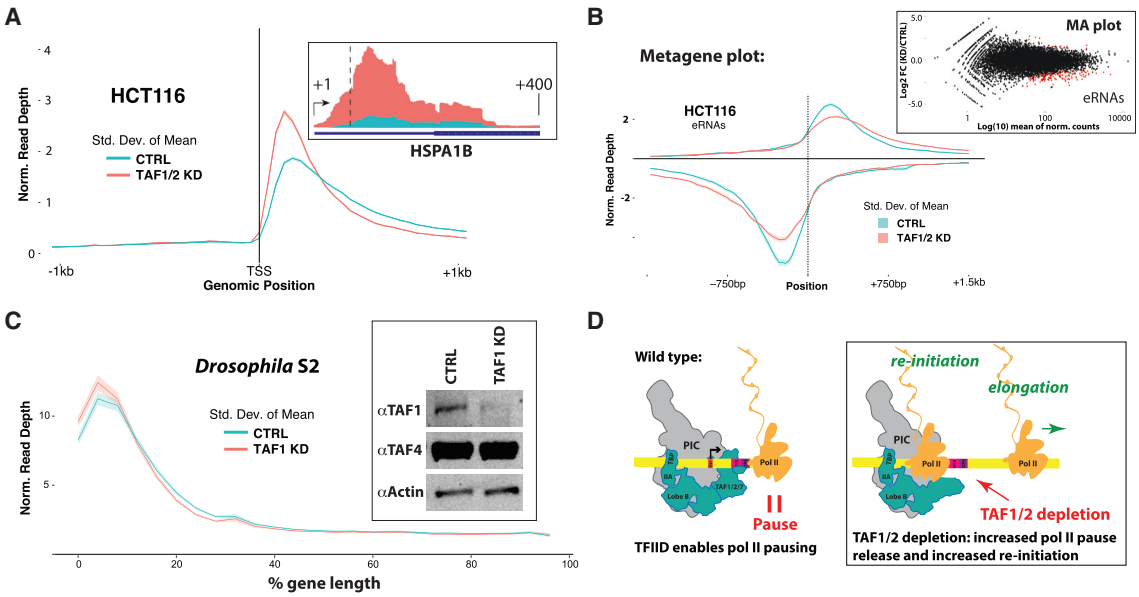


Figure 4. Depletion of TFIID Lobe C Subunits Increases Pause Release; TFIID Function Is Conserved in *Drosophila* Cells

(A) Metagene plot (all genes; n = 18,687) of promoter-proximal region that shows increased 5′ end reads in TAF1/2 knockdown cells, which extend beyond the pause site. These data are consistent with increases in pause release and re-initiation, which are coupled events in metazoan cells (Gressel et al., 2017; Shao and Zeitlinger, 2017). A sharp reduction in reads typically occurs around +300 from the TSS (e.g., HSPA1B locus, inset; vertical dashed line +60 from TSS), which suggests downstream auxiliary factors that terminate or arrest RNAPII (see Discussion).

(B) Metagene plot and MA plot (inset) of non-annotated regions, showing that eRNA transcription is reduced in TAF1/2 knockdown cells.

(C) Metagene plot of expressed genes (n = 10,995) in *Drosophila*, comparing control versus Taf1 knockdown S2 cells. As in TAF1/2 knockdown HCT116 cells, transcription increased at gene 5′ ends. Inset: representative western blots show Taf1 knockdown in S2 cells. Taf1 knockdown was determined to be 88% (±6.7%) from 3 biological replicates.

(D) Model. TFIID is required to establish RNAPII promoter-proximal pausing. Disruption of TFIID lobe C correlates with increased RNAPII pause release, which enables other RNAPII complexes to re-initiate transcription (Gressel et al., 2017; Shao and Zeitlinger, 2017).

result from its disordered C terminus, which contains 14 lysines in a 30-residue stretch.

Following acute TAF depletion using Trim-Away, we isolated nuclei and performed replicate PRO-seq experiments (TAF1/2 knockdown versus controls). The data showed good correlation between replicates (Figures S4A and S4B), and normalization tests (see STAR Methods) confirmed that rapid TAF1/2 knockdown did not dramatically shift overall transcription levels versus controls (Figure 3C). An expectation based upon our *in vitro* results (Figures 2A and 2C) and cryoelectron microscopy (cryo-EM) structural data (Louder et al., 2016; Patel et al., 2018) was that TFIID might serve as a "brake" for promoter-associated RNAPII complexes and that removal of this brake would enhance pause release. This expectation was confirmed by the PRO-seq data, which showed an overall increase in transcription genome-wide (Figures S4C–S4E), except at non-annotated enhancer RNAs (eRNAs) (see below). A representative protein-coding gene example is shown in Figure 3D, and genome-wide trends are shown as MA plots in Figures 3E and 3F.

As might be expected from rapid depletion of TFIID lobe C subunits, the PRO-seq data showed transcriptional changes at thousands of gene 5′ ends. In cells, increased pause release can also promote re-initiation by additional RNAPII enzymes (Gressel et al., 2017; Shao and Zeitlinger, 2017). In agreement, we observed increased 5′ end reads at thousands of genes (Figure 3F), and the reads extended beyond promoter-proximal pause regions (Figure 4A, inset), suggesting a defect in pause enforcement. Unexpectedly, however, transcription sharply declined at approximately +300 downstream of the TSS, as shown at JUN and HSPA1B (Figures 3D and 4A, inset) and in a metagene plot representing all genes (Figures 4A and S5). We note that the sharp decline in reads beyond the promoter-proximal pause site superficially resembles RNAPII pausing; however, comparisons with PRO-seq data from flavopiridol-treated cells (a positive control for RNAPII pausing) showed stark differences and confirmed that pausing was not increased in TAF1/2-depleted cells (Figure S6). The sharp decline in transcription at around +300 in TAF1/2-depleted cells explains the reduced increase in gene body reads (Figure 3E) compared with 5′ end reads (Figure 3F) and suggests the presence of a distinct factor(s) that functions at this later, post-pause release stage (see Discussion). In contrast to annotated genes, transcription of non-annotated eRNAs declined overall in TAF1/2-depleted cells (Figure 4B), suggesting alternate regulatory mechanisms at these loci. As expected, a significant decrease in the TAF1 motif was observed in the Trim-Away depleted cells (Figure S6D).

TFIID Function in RNAPII Promoter-Proximal Pausing Is Conserved in *Drosophila*

Taf1 knockdown in *Drosophila* S2 cells has minimal impact on other TFIID subunits (Pennington et al., 2013; Wright et al., 2006), and promoter-proximal pausing is widespread in *Drosophila*

(Muse et al., 2007; Nechaev et al., 2010; Zeitlinger et al., 2007). To further test the impact of TAF1 (a TFIID lobe C subunit) on RNAPII pausing, Taf1 was knocked down in S2 cells and PRO-seq experiments were completed in triplicate (Figures 4C and S7A). Consistent with TAF1/2-depleted human cells, Taf1 knockdown in *Drosophila* S2 cells showed a similar promoter-proximal increase in transcription genome-wide (Figure 4C), suggesting increased pause release and increased re-initiation with Taf1 knockdown (Figure S7B). These data suggest a conserved role for TFIID in the regulation of RNAPII promoter-proximal pausing.

DISCUSSION

Structural data indicate that TFIID lobe C subunits TAF1 and TAF2 bind promoter DNA downstream of the TSS (Louder et al., 2016; Patel et al., 2018). Past studies revealed that insertion of 10-bp DNA at the +15 site relative to the TSS disrupted RNAPII pausing at the HSP70 gene in *Drosophila* S2 cells (Kwak et al., 2013). This led to a "complex interaction" model for pausing, in which a promoter-bound factor(s) establishes an interaction (directly or indirectly) with the paused RNAPII complex. In agreement with this model, we observe a TFIID requirement for RNAPII promoter-proximal pausing *in vitro*, which is further supported by PRO-seq data in TAF-depleted human and *Drosophila* S2 cells. Additional evidence for TFIID-dependent regulation of RNAPII pausing derives from correlations among paused genes and DNA sequence elements bound by TFIID (Hendrix et al., 2008; Lee et al., 2008; Li et al., 2013; Shao et al., 2019). Defects in TFIID function are linked to numerous diseases, including cancer (Xu et al., 2018) and neurodegenerative disorders (Aneichyk et al., 2018). Its requirement for RNAPII promoter-proximal pause regulation may underlie these and other biological functions.

Biochemical reconstitution of RNAPII promoter-proximal pausing provides a level of mechanistic control that is simply not possible with cell-based assays; consequently, we were able to discover that RNAPII pausing is an inherent property of the human PIC and that TFIID is a key PIC factor that establishes pausing (Figures 2 and 4D). Our results also reveal NELF, DSIF, and P-TEFb as auxiliary factors that, although not required for pausing, enable robust regulation of this common transcriptional intermediate state. Time course experiments indicated that polymerases in the paused region remained active and generated elongated transcripts over time (Figure 1C). Experiments with P-TEFb showed enhanced release of paused intermediates, providing further evidence that polymerases in the paused region were active and competent for elongation (Figures 1B and 1D). However, some transcripts remained in the pause region after the 10-min reactions, even with added P-TEFb. This result is also consistent with current models that invoke alternative outcomes for promoter-proximal paused RNAPII, including premature termination (Erickson et al., 2018; Krebs et al., 2017), arrest (Adelman et al., 2005), or a more stable paused intermediate (Chen et al., 2015; Henriques et al., 2013). Addressing the mechanisms and factors that regulate these distinct outcomes could be explored in future studies.

Despite its advantages, the reconstituted *in vitro* transcription assay does not match the complexity of regulatory inputs that converge upon active promoters in a living cell. To test the TFIID requirement for promoter-proximal pausing in cells, we were able to rapidly deplete TFIID lobe C subunits TAF1 and TAF2 using Trim-Away (Clift et al., 2017), and genome-wide changes in nascent transcription were assessed with PRO-seq (Kwak et al., 2013). Consistent with the *in vitro* data, global transcription increased at protein-coding genes upon TAF1/2 knockdown, with evidence for enhanced pause release (Figures 3D–3F). PRO-seq reads increased at 5′ ends and downstream of promoter-proximal pause sites at thousands of genes in TAF1/2-depleted cells. These data are consistent with increased pause release and increased re-initiation (Figure 4D), two processes that are coupled in metazoan cells (Gressel et al., 2017; Shao and Zeitlinger, 2017). Unexpectedly, however, increased pause release did not yield similar genome-wide increases in gene body reads. Instead, the PRO-seq data revealed a sharp reduction in reads downstream of promoter-proximal pause sites, at around +300 from the TSS in both human and *Drosophila* cells. These results implicate additional regulatory mechanisms, downstream of the pause site, that may terminate or arrest RNAPII. Although future studies are needed to identify the factors involved, we note that the Integrator complex was recently shown to cleave nascent transcripts downstream of pause sites at hundreds of genes in *Drosophila* cells (Tatomer et al., 2019). Because promoter-proximal pausing helps ensure proper capping of transcripts at their 5′ ends (Rasmussen and Lis, 1993; Tome et al., 2018), downstream regulatory mechanisms may become important when RNAPII promoter-proximal pausing is disrupted.

A TFIID requirement for RNAPII promoter-proximal pausing implies that other pause regulatory factors may function directly or indirectly through TFIID. Although additional mechanistic aspects remain to be addressed, it is notable that pause regulatory factors, including P-TEFb and MYC, interact (directly or indirectly) with TFIID (Gegonne et al., 2008; Wei et al., 2019; Yadav et al., 2019); moreover, TFIID is conformationally flexible (Cianfrocco et al., 2013) and likely undergoes structural reorganization during RNAPII transcription initiation and pause release (Yakovchuk et al., 2010; Zhang et al., 2015). Such structural transitions may contribute to TFIID-dependent regulation of RNAPII pausing. Whereas nucleosomes likely affect promoter-proximal pausing, they are not required, based upon our results and data in *Drosophila* and mammalian systems (Benjamin and Gilmour, 1998; Kwak et al., 2013; Lai and Pugh, 2017; Li et al., 2013). TFIID possesses multiple domains that bind chromatin marks associated with transcriptionally active loci, including H3K4me3 (Jacobson et al., 2000; Vermeulen et al., 2007), which suggests TFIID function is regulated in part through epigenetic mechanisms. Future studies should help establish whether specific chromatin marks contribute to TFIID-dependent regulation of RNAPII pausing, potentially by affecting TFIID promoter occupancy or by impacting TFIID structure and function.

STAR★METHODS

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

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.molcel.2020.03.008.

ACKNOWLEDGMENTS

We thank J. Goodrich for TAF4 monoclonal antibodies, R. Tjian for ERCC3 antibodies, and T.-M. Decker for assisting with Mediator purifications. Funding support was provided from the NSF (MCB-1244175 and MCB-1818147 to D.J.T. and DBI-1759949 to R.D.D.) and the NIH (GM117370 and GM110064 to D.J.T.; GM125871 to R.D.D.; T32GM008759 to C.B.F. and J.D.R.; AG051335 to C.B.L.; and T32EB009419-10 to A.S.). I.B. is a Senior Investigator of the Wellcome Trust (106115/Z/14/Z). We acknowledge the BioFrontiers Computing Core (UC-Boulder), supported by the NIH (OD012300).

AUTHOR CONTRIBUTIONS

Study Design, C.B.F., C.B.L., R.D.D., and D.J.T.; Biochemical and Cellular Assays, C.B.F., K.G., J.M., M.N.E., J.K.R., and O.L.; Recombinant TFIID Complexes, K.G. and I.B.; S2 Cell Experiments, A.S. and M.T.M.; PRO-Seq, C.B.L., Z.L.M., and J.D.R.; Funding and Mentoring, M.T.M., I.B., R.D.D., and D.J.T.; Data Distribution, all authors; Manuscript Writing, D.J.T.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 18, 2019
Revised: November 15, 2019
Accepted: March 5, 2020
Published: March 30, 2020

REFERENCES

Adelman, K., Marr, M.T., Werner, J., Saunders, A., Ni, Z., Andrulis, E.D., and Lis, J.T. (2005). Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIS. Mol. Cell *17*, 103–112.

Adelman, K., Kennedy, M.A., Nechaev, S., Gilchrist, D.A., Muse, G.W., Chinenov, Y., and Rogatsky, I. (2009). Immediate mediators of the inflammatory response are poised for gene activation through RNA polymerase II stalling. Proc. Natl. Acad. Sci. USA *106*, 18207–18212.

Aneichyk, T., Hendriks, W.T., Yadav, R., Shin, D., Gao, D., Vaine, C.A., Collins, R.L., Domingo, A., Currall, B., Stortchevoi, A., et al. (2018). Dissecting the causal mechanism of X-linked dystonia-parkinsonism by integrating genome and transcriptome assembly. Cell *172*, 897–909.e21.

Azofeifa, J.G., and Dowell, R.D. (2017). A generative model for the behavior of RNA polymerase. Bioinformatics *33*, 227–234.

Benjamin, L.R., and Gilmour, D.S. (1998). Nucleosomes are not necessary for promoter-proximal pausing in vitro on the Drosophila hsp70 promoter. Nucleic Acids Res. *26*, 1051–1055.

Chen, F., Gao, X., and Shilatifard, A. (2015). Stably paused genes revealed through inhibition of transcription initiation by the TFIIH inhibitor triptolide. Genes Dev. *29*, 39–47.

Cianfrocco, M.A., Kassavetis, G.A., Grob, P., Fang, J., Juven-Gershon, T., Kadonaga, J.T., and Nogales, E. (2013). Human TFIID binds to core promoter DNA in a reorganized structural state. Cell *152*, 120–131.

Clemens, J.C., Worby, C.A., Simonson-Leff, N., Muda, M., Maehama, T., Hemmings, B.A., and Dixon, J.E. (2000). Use of double-stranded RNA interference in Drosophila cell lines to dissect signal transduction pathways. Proc. Natl. Acad. Sci. USA *97*, 6499–6503.

Clift, D., McEwan, W.A., Labzin, L.I., Konieczny, V., Mogessie, B., James, L.C., and Schuh, M. (2017). A method for the acute and rapid degradation of endogenous proteins. Cell *171*, 1692–1706.e18.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845–1848.

Core, L.J., Waterfall, J.J., Gilchrist, D.A., Fargo, D.C., Kwak, H., Adelman, K., and Lis, J.T. (2012). Defining the status of RNA polymerase at promoters. Cell Rep. *2*, 1025–1035.

Ebmeier, C.C., Erickson, B., Allen, B.L., Allen, M.A., Kim, H., Fong, N., Jacobsen, J.R., Liang, K., Shilatifard, A., Dowell, R.D., et al. (2017). Human TFIIH kinase CDK7 regulates transcription-associated chromatin modifications. Cell Rep. *20*, 1173–1186.

Erickson, B., Sheridan, R.M., Cortazar, M., and Bentley, D.L. (2018). Dynamic turnover of paused Pol II complexes at human promoters. Genes Dev. *32*, 1215–1225.

Fitzgerald, D.J., Berger, P., Schaffitzel, C., Yamada, K., Richmond, T.J., and Berger, I. (2006). Protein complex expression by using multigene baculoviral vectors. Nat. Methods *3*, 1021–1032.

Gegonne, A., Weissman, J.D., Lu, H., Zhou, M., Dasgupta, A., Ribble, R., Brady, J.N., and Singer, D.S. (2008). TFIID component TAF7 functionally interacts with both TFIIH and P-TEFb. Proc. Natl. Acad. Sci. USA *105*, 5367–5372.

Ghavi-Helm, Y., Klein, F.A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., and Furlong, E.E. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. Nature *512*, 96–100.

Gressel, S., Schwalb, B., Decker, T.M., Qin, W., Leonhardt, H., Eick, D., and Cramer, P. (2017). CDK9-dependent RNA polymerase II pausing controls transcription initiation. eLife *6*, e29736.

Hahne, F., and Ivanek, R. (2016). Visualizing genomic data using Gviz and Bioconductor. Methods Mol. Biol. *1418*, 335–351.

Hawley, D.K., and Roeder, R.G. (1985). Separation and partial characterization of three functional steps in transcription initiation by human RNA polymerase II. J. Biol. Chem. *260*, 8163–8172.

Hawley, D.K., and Roeder, R.G. (1987). Functional steps in transcription initiation and reinitiation from the major late promoter in a HeLa nuclear extract. J. Biol. Chem. *262*, 3452–3461.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Hendrix, D.A., Hong, J.W., Zeitlinger, J., Rokhsar, D.S., and Levine, M.S. (2008). Promoter elements associated with RNA Pol II stalling in the Drosophila embryo. Proc. Natl. Acad. Sci. USA *105*, 7762–7767.

Henriques, T., Gilchrist, D.A., Nechaev, S., Bern, M., Muse, G.W., Burkholder, A., Fargo, D.C., and Adelman, K. (2013). Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals. Mol. Cell *52*, 517–528.

Henriques, T., Scruggs, B.S., Inouye, M.O., Muse, G.W., Williams, L.H., Burkholder, A.B., Lavender, C.A., Fargo, D.C., and Adelman, K. (2018). Widespread transcriptional pausing and elongation control at enhancers. Genes Dev. *32*, 26–41.

Jacobson, R.H., Ladurner, A.G., King, D.S., and Tjian, R. (2000). Structure and function of a human TAFII250 double bromodomain module. Science *288*, 1422–1425.

Jonkers, I., Kwak, H., and Lis, J.T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. eLife *3*, e02407.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods *12*, 357–360.

Knuesel, M.T., Meyer, K.D., Bernecky, C., and Taatjes, D.J. (2009). The human CDK8 subcomplex is a molecular switch that controls Mediator co-activator function. Genes Dev. *23*, 439–451.

Krebs, A.R., Imanci, D., Hoerner, L., Gaidatzis, D., Burger, L., and Schübeler, D. (2017). Genome-wide single-molecule footprinting reveals high RNA polymerase II turnover at paused promoters. Mol. Cell *67*, 411–422.e4.

Kwak, H., and Lis, J.T. (2013). Control of transcriptional elongation. Annu. Rev. Genet. *47*, 483–508.

Kwak, H., Fuda, N.J., Core, L.J., and Lis, J.T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science *339*, 950–953.

Lagha, M., Bothma, J.P., Esposito, E., Ng, S., Stefanik, L., Tsui, C., Johnston, J., Chen, K., Gilmour, D.S., Zeitlinger, J., and Levine, M.S. (2013). Paused Pol II coordinates tissue morphogenesis in the Drosophila embryo. Cell *153*, 976–987.

Lai, W.K., and Pugh, B.F. (2017). Genome-wide uniformity of human 'open' pre-initiation complexes. Genome Res. *27*, 15–26.

Lee, C., Li, X., Hechmer, A., Eisen, M., Biggin, M.D., Venters, B.J., Jiang, C., Li, J., Pugh, B.F., and Gilmour, D.S. (2008). NELF and GAGA factor are linked to promoter-proximal pausing at many genes in Drosophila. Mol. Cell. Biol. *28*, 3290–3300.

Leek, J.T., Johnson, W.E., Parker, H.S., Fertig, E.J., Jaffe, A.E., Storey, J.D., Zhang, Y., and Torres, L.C. (2019). SVA: Surrogate Variable Analysis. R package version 3.32.1, https://bioconductor.org/packages/release/bioc/html/sva.html.

Li, J., Liu, Y., Rhee, H.S., Ghosh, S.K., Bai, L., Pugh, B.F., and Gilmour, D.S. (2013). Kinetic competition between elongation rate and binding of NELF controls promoter-proximal pausing. Mol. Cell *50*, 711–722.

Li, D., Zhang, B., Xing, X., and Wang, T. (2015). Combining MeDIP-seq and MRE-seq to investigate genome-wide CpG methylation. Methods *72*, 29–40.

Liao, Y., Smyth, G.K., and Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. *41*, e108.

Lin, C., Smith, E.R., Takahashi, H., Lai, K.C., Martin-Brown, S., Florens, L., Washburn, M.P., Conaway, J.W., Conaway, R.C., and Shilatifard, A. (2010). AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia. Mol. Cell *37*, 429–437.

Louder, R.K., He, Y., López-Blanco, J.R., Fang, J., Chacón, P., and Nogales, E. (2016). Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. Nature *531*, 604–609.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Mahat, D.B., Salamanca, H.H., Duarte, F.M., Danko, C.G., and Lis, J.T. (2016). Mammalian heat shock response and mechanisms underlying its genome-wide transcriptional regulation. Mol. Cell *62*, 63–78.

Marr, M.T., 2nd, Isogai, Y., Wright, K.J., and Tjian, R. (2006). Coactivator cross-talk specifies transcriptional output. Genes Dev. *20*, 1458–1469.

Marshall, N.F., and Price, D.H. (1992). Control of formation of two distinct classes of RNA polymerase II elongation complexes. Mol. Cell. Biol. *12*, 2078–2090.

Miller, T.E., Liau, B.B., Wallace, L.C., Morton, A.R., Xie, Q., Dixit, D., Factor, D.C., Kim, L.J.Y., Morrow, J.J., Wu, Q., et al. (2017). Transcription elongation factors represent in vivo cancer dependencies in glioblastoma. Nature *547*, 355–359.

Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grissom, S.F., Zeitlinger, J., and Adelman, K. (2007). RNA polymerase is poised for activation across the genome. Nat. Genet. *39*, 1507–1511.

Näär, A.M., Beaurang, P.A., Robinson, K.M., Oliner, J.D., Avizonis, D., Scheek, S., Zwicker, J., Kadonaga, J.T., and Tjian, R. (1998). Chromatin, TAFs, and a novel multiprotein coactivator are required for synergistic activation by Sp1 and SREBP-1a in vitro. Genes Dev. *12*, 3020–3031.

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

Patel, A.B., Louder, R.K., Greber, B.J., Grünberg, S., Luo, J., Fang, J., Liu, Y., Ranish, J., Hahn, S., and Nogales, E. (2018). Structure of human TFIID and mechanism of TBP loading onto promoter DNA. Science *362*, eaau8872.

Pennington, K.L., Marr, S.K., Chirn, G.W., and Marr, M.T., 2nd (2013). Holo-TFIID controls the magnitude of a transcription burst and fine-tuning of transcription. Proc. Natl. Acad. Sci. USA *110*, 7678–7683.

Qiu, Y., and Gilmour, D.S. (2017). Identification of regions in the Spt5 subunit of DRB sensitivity-inducing factor (DSIF) that are involved in promoter-proximal pausing. J. Biol. Chem. *292*, 5555–5570.

Rasmussen, E.B., and Lis, J.T. (1993). In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. Proc. Natl. Acad. Sci. USA *90*, 7923–7927.

Renner, D.B., Yamaguchi, Y., Wada, T., Handa, H., and Price, D.H. (2001). A highly purified RNA polymerase II elongation control system. J. Biol. Chem. *276*, 42601–42609.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Shao, W., and Zeitlinger, J. (2017). Paused RNA polymerase II inhibits new transcriptional initiation. Nat. Genet. *49*, 1045–1051.

Shao, W., Alcantara, S.G., and Zeitlinger, J. (2019). Reporter-ChIP-nexus reveals strong contribution of the *Drosophila* initiator sequence to RNA polymerase pausing. eLife *8*, e41461.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

Tatomer, D.C., Elrod, N.D., Liang, D., Xiao, M.-S., Jiang, J.Z., Jonathan, M., Huang, K.-L., Wagner, E.J., Cherry, S., and Wilusz, J.E. (2019). The Integrator complex cleaves nascent mRNAs to attenuate transcription. Genes Dev. *33*, 1525–1538.

Tome, J.M., Tippens, N.D., and Lis, J.T. (2018). Single-molecule nascent RNA sequencing identifies regulatory domain architecture at promoters and enhancers. Nat. Genet. *50*, 1533–1541.

Tripodi, I.J., Allen, M.A., and Dowell, R.D. (2018). Detecting differential transcription factor activity from ATAC-seq data. Molecules *23*, E1136.

Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W., van Schaik, F.M., Varier, R.A., Baltissen, M.P., Stunnenberg, H.G., Mann, M., and Timmers, H.T. (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell *131*, 58–69.

Vo Ngoc, L., Wang, Y.L., Kassavetis, G.A., and Kadonaga, J.T. (2017). The punctilious RNA polymerase II core promoter. Genes Dev. *31*, 1289–1301.

Wei, P., Garber, M.E., Fang, S.M., Fischer, W.H., and Jones, K.A. (1998). A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. Cell *92*, 451–462.

Wei, Y., Resetca, D., Li, Z., Johansson-Åkhe, I., Ahlner, A., Helander, S., Wallenhammar, A., Morad, V., Raught, B., Wallner, B., et al. (2019). Multiple direct interactions of TBP with the MYC oncoprotein. Nat. Struct. Mol. Biol. *26*, 1035–1043.

Weinzierl, R.O., Dynlacht, B.D., and Tjian, R. (1993). Largest subunit of Drosophila transcription factor IID directs assembly of a complex containing TBP and a coactivator. Nature *362*, 511–517.

Wright, K.J., Marr, M.T., 2nd, and Tjian, R. (2006). TAF4 nucleates a core subcomplex of TFIID and mediates activated transcription from a TATA-less promoter. Proc. Natl. Acad. Sci. USA *103*, 12347–12352.

Xu, Y., Milazzo, J.P., Somerville, T.D.D., Tarumoto, Y., Huang, Y.H., Ostrander, E.L., Wilkinson, J.E., Challen, G.A., and Vakoc, C.R. (2018). A TFIID-SAGA perturbation that targets MYB and suppresses acute myeloid leukemia. Cancer Cell *33*, 13–28.e8.

Yadav, D., Ghosh, K., Basu, S., Roeder, R.G., and Biswas, D. (2019). Multivalent role of human TFIID in recruiting elongation components at the promoter-proximal region for transcriptional control. Cell Rep. *26*, 1303–1317.e7.

Yakovchuk, P., Gilman, B., Goodrich, J.A., and Kugel, J.F. (2010). RNA polymerase II and TAFs undergo a slow isomerization after the polymerase is recruited to promoter-bound TFIID. J. Mol. Biol. *397*, 57–68.

Yamaguchi, Y., Filipovska, J., Yano, K., Furuya, A., Inukai, N., Narita, T., Wada, T., Sugimoto, S., Konarska, M.M., and Handa, H. (2001). Stimulation of RNA polymerase II elongation by hepatitis delta antigen. Science *293*, 124–127.

Zeitlinger, J., Stark, A., Kellis, M., Hong, J.W., Nechaev, S., Adelman, K., Levine, M., and Young, R.A. (2007). RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat. Genet. *39*, 1512–1516.

Zhang, Z., Boskovic, Z., Hussain, M.M., Hu, W., Inouye, C., Kim, H.J., Abole, A.K., Doud, M.K., Lewis, T.A., Koehler, A.N., et al. (2015). Chemical perturbation of an intrinsically disordered region of TFIID distinguishes two modes of transcription initiation. eLife *4*, e07777.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
TAF1	Santa Cruz Biotechnology	Cat#sc-735 X; RRID: AB_671202
TAF2	abcam	Cat#ab103468; RRID: AB_10716140
TAF4	Millipore Sigma	Cat#07-1803; RRID: 10846677
TAF8	abcam	Cat#ab204894
TBP	Santa Cruz Biotechnology	Cat#sc-273; RRID: AB_2200059
Actin	Santa Cruz Biotechnology	Cat#sc-47778; RRID: AB_2714189
Histone H3	A. Shilatifard (Ebmeier et al., 2017).	N/A
TAF1 C413 for Trim-Away	Recombinant Antibody Network	Anti-TAF1-RAB-C413
30H9 (*Drosophila* Taf1)	Weinzierl et al., 1993	N/A
3E12 (*Drosophila* Taf4)	Marr et al., 2006	N/A
Chemicals, Peptides, and Recombinant Proteins		
DTT	Sigma-Aldrich	cat#D0632
Benzamidine	Sigma-Aldrich	cat#B6506-100G
Sodium Metabisulfite	Sigma-Aldrich	cat#255556-100G
Phenylmethylsulfonyl Fluoride	American Bioanalytical	cat#AB01620
TRIzol Reagent	Invitrogen	cat#15596018
TRIzol LS Reagent	Invitrogen	cat#10296028
T4 Polynucleotide Kinase	NEB	cat#M0201S
RNA 5′ Pyrophosphohydrolase (RppH)	NEB	cat#M0356S
ThermoPol® Reaction Buffer Pack	NEB	cat#B9004S
T4 RNA Ligase 1 (ssRNA Ligase)	NEB	cat#M0204S
SuperScript III Reverse Transcriptase	Invitrogen	cat#18080044
Phusion® High-Fidelity DNA Polymerase	NEB	cat#M0530L
SUPERase-In	Invitrogen	cat#AM2694
Deposited Data		
Raw and analyzed data - PRO-seq experiment	This paper	GEO: GSE132764; https://www.ncbi.nlm.nih.gov/geo/
Oligonucleotides		
/5Phos/rGrArUrCrGrUrCrGrGrArCrUr GrUrArGrArArCrUrCrUrGrArArC/3InvdT/	IDT	VRA3
rCrCrUrUrGrGrCrArCrCrCrGrAr GrArArUrUrCrCrA	IDT	VRA5
AATGATACGGCGACCACCGAGATCTA CACGTTCAGAGTTCTACAGTCCGA	IDT	RP1
CAAGCAGAAGACGGCATACGA GAT NNNNNN GTGACTGGAG TTCCTTGGCACCCGAGAATTCCA	IDT	RPI-n
Experimental Models: Cell Lines		
Human HCT116 Cells	ATCC	CCL-247
Drosophila S2 Cells	Mike Marr Lab	N/A
Software and Algorithms		
NascentFlow Pipeline v1.2	Dowell Lab (DOI 10.17605/ OSF.IO/NDHJ2)	https://github.com/Dowell-Lab/Nascent-Flow
BBDuk v38.05	Bushnell, B. (Joint Genome Institute)	https://jgi.doe.gov/data-and-tools/bbtools/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bedtools v2.25.0		https://github.com/arq5x/bedtools2
FastQC v0.11.8	Andrews, S. (Babraham Bioinformatics)	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
HiSat2 v2.1.0	Kim et al., 2015	https://ccb.jhu.edu/software/hisat2/manual.shtml
IGV Tools v2.3.75		https://software.broadinstitute.org/software/igv/
Nextflow v19.04.1		https://www.nextflow.io/
Preseq v2.0.3		https://github.com/smithlabcode/preseq
Samtools v1.8		http://www.htslib.org/
FStitch v1.1		https://github.com/Dowell-Lab/FStitch
TFit v1.1	Azofeifa and Dowell, 2017	https://github.com/dowell-lab/tfit
Python v3.6.3	Python Software Foundation	https://www.python.org/
R v3.5.0	R Foundation	https://www.r-project.org/
Tidyverse v1.3.0		https://www.tidyverse.org/
Argparse v2.0.1		https://cran.r-project.org/web/packages/argparse/index.html
matrixStats v0.55.0		https://cran.r-project.org/web/packages/matrixStats/index.html
ggthemes v4.2.0		https://cran.r-project.org/web/packages/ggthemes/index.html
ggfortify v0.4.8		https://cran.r-project.org/web/packages/ggfortify/index.html
ggsci v2.9		https://cran.r-project.org/web/packages/ggsci/index.html
ggrepel v0.8.1		https://cran.r-project.org/web/packages/ggrepel/index.html
readxl v1.3.1		https://cran.r-project.org/web/packages/readxl/index.html
reshape2 v1.4.3		https://cran.r-project.org/web/packages/reshape2/index.html
nortest v1.0-4		https://cran.r-project.org/web/packages/nortest/index.html
Limma 3.42.2	Ritchie et al., 2015	https://bioconductor.org/packages/release/bioc/html/limma.html
SVA v3.32.1	Leek et al., 2019	https://bioconductor.org/packages/release/bioc/html/sva.html
DESeq2 v1.26.0	Love et al., 2014	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
TFEA v1.0.1		https://github.com/Dowell-Lab/TFEA https://www.biorxiv.org/content/10.1101/2020.01.25.919738v3.full
Other		
Micro Bio-Spin P-30 Gel Columns, Tris Buffer (RNase-free) #7326250	Bio-Rad	cat#7326250
1.7 mL pre-lubricated tube	Costar	cat#3207
AMPURE XP 60ML	Beckman Coulter	cat#NC9933872
Ultra Low Range DNA Ladder	Invitrogen	cat#10597012
Streptavidin M280 beads	Invitrogen	cat#11206D
Biotin-11-CTP	Perkin Elmer	cat#NEL542001EA
Ribonucleotide Triphosphate rATP, 100mM	Promega	cat#E6011
Ribonucleotide Triphosphate rCTP, 100mM	Promega	cat#E6041
Ribonucleotide Triphosphate rGTP, 100mM	Promega	cat#E6031
Ribonucleotide Triphosphate rUTP, 100mM	Promega	cat#E6021

LEAD CONTACT AND MATERIALS AVAILABILITY

Requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Dylan Taatjes (taatjes@colorado.edu).

All reagents generated in this study are available from the Lead Contact without restriction or with a Materials Transfer Agreement.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Drosophila cell culture

D. melanogaster Schneider line 2 (S2) cells were maintained at 25°C in Schneider's medium containing 10% (vol/vol) Fetalplex (Gemini), 100 units/mL penicillin, and 0.1 mg/mL streptomycin.

HCT116 cell culture

HCT116 cells were grown in McCoy's media (GIBCO, 16600082) with GIBCO 100x Antibiotic-Antimycotic (Fisher Sci, 15240062) penicillin-streptomycin and 10% fetal bovine serum (FBS) supplementation.

METHOD DETAILS

HSPA1B promoter DNA template

The native human HSPA1B promoter was amplified from genomic DNA (HeLa) by PCR (forward primer: CTCCTT CCCATT AAGACG GAAAAA ACATCC GGGAGA GCCGGT CCG; reverse primer: ACCTTG CCGTGT TGGAAC ACCCCC ACGCAG GAGTAG GTGGTG CCCAGGTC) and cloned into a pCR-Blunt-TOPO plasmid. The HSPA1B promoter corresponding to −500 to +216 base pairs relative to the transcription start site was amplified off this plasmid using Phusion polymerase (Thermo-Fisher #F530S). The resulting PCR product was then purified using the E.Z.N.A. Gel Extraction kit (Omega BioTek #D2500). The DNA was then ethanol precipitated, washed, resuspended to 100 nM in milliQ water, and stored frozen in single-use aliquots.

Reconstituted *in vitro* transcription

The HSPA1B promoter template (5nM in 10 μL) was incubated with 400 nM HSF1 in template mix buffer (20 mM HEPES pH 7.6, 1 mM DTT, 8 mM $MgCl_2$) at 30°C for 20 minutes. Then, 10 μL of PIC mix was added, which contained ~5-100 nM of each GTF (TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH, RNAPII, and Mediator) with added DB(100) buffer (10% glycerol, 10 mM Tris pH 7.9, 180 mM KCl, 1 mM DTT). This mix was incubated for 15 minutes to allow PIC assembly. Transcription was initiated by addition of A/G/UTP (to 350 μM) and ^{32}P-CTP (400 nM) in DB(100) buffer. Reactions were chased one-minute post-initiation by addition of cold CTP to bring [CTP] to 350 μM. In some cases, [ATP] was increased to 2 mM. This protocol ensured predominantly single-round transcription based upon 1) the short time-frame of the assay, and 2) chase with cold CTP ensures that any potential re-initiation would have ^{32}P-signal reduced by orders of magnitude from the initial ^{32}P-CTP pulse. Single-round transcription was confirmed based upon quantitation of total ^{32}P signal (i.e., encompassing all detected transcripts ~20 to 216 nt in length) over time and further verified using experiments with added sarcosyl (0.2%), which prevents re-initiation (Hawley and Roeder, 1985; Hawley and Roeder, 1987). Transcription was stopped by addition of 150 μL of stop buffer (20 mM EDTA, 200 mM NaCl, 1% SDS). RNA was then ethanol precipitated, washed with 75% cold ethanol, resuspended in formamide loading buffer, boiled, and loaded onto an acrylamide sequencing gel (typically 18%) for analysis.

For pause index (PI) calculations, ^{32}P-signal was quantitated (with background correction; using "subtract background" feature in ImageJ, using a 1000-pixel rolling ball radius across the entire gel) from sequencing gels using ImageJ; the ratio of paused (20-80nt) versus elongated (100-216nt) transcripts was reported as the PI. For each PI plot (Figure 1D or 2F), the data were taken from a set of experiments that used the same PIC factors (i.e., TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH, RNAPII, and Mediator from the same preparation). Because the scale of some PIC factor purifications could not accommodate all experiments included in this study, some experiments used PIC factors from different preparations at different titrations (determined empirically). PI data for Figure 1D versus Figure 2F used different preparations of some PIC factors, and this likely contributes to the different PI for TFIID-containing PICs. For any given experiment, regardless of the specific PIC factor preparation or titration, NELF/DSIF increased PI (more pausing) and P-TEFb decreased PI (less pausing, more pause release).

Primer Extension

Transcription assays were carried out as described above up until the point of PIC assembly. To initiate transcription, a final concentration of 400 μM A/G/C/UTP was added, and reactions were allowed to proceed for 30 minutes then stopped with 150 μL stop buffer. RNA was then phenol:chloroform extracted and ethanol precipitated. Primer extension was carried out using AMV reverse transcriptase (RT) as described (https://www.promega.com/products/pcr/rt-pcr/amv-reverse-transcriptase/?catNum=M5101). The first ~80 nt of the HSPA1B transcript is GC-rich and predicted to adopt secondary structures, which can block RT extension. Thus, extension assays were carried out at elevated temperatures that could be tolerated by AMV RT.

Purification of human PIC factors

Factors were purified as described: TFIIA, TFIIB, TFIID, TFIIE, TFIIF, RNAPII, Mediator (Knuesel et al., 2009); TFIIH (Ebmeier et al., 2017). Briefly, TFIIA, TFIIB, TFIIE, and TFIIF were expressed in *E. coli* and purified over several ion exchange and affinity columns. For RNAPII, nuclear extract from approximately 20 L of HeLa cells was used to start. After the ammonium sulfate (AS) cut, protein was resuspended in 20 mM AS buffer (20 mM HEPES, pH 7.6; 20% glycerol, 0.2 mM EDTA, 2 mM $MgCl_2$) to a concentration of ca. 10 mg/mL and loaded onto an anti-Rpb1 column. Following incubation, the resin was washed with 50 column volumes 0.5M AS buffer B (50 mM Tris, pH 7.9; 10% glycerol, 1 mM EDTA, 10 μM $ZnCl_2$, 0.025% NP-40) and 10 column volumes 0.1M AS buffer B. After elution with CTD peptide (4 repeats; 1 mg/mL in 0.1M AS buffer B), the sample was loaded onto a UNO-Q column (BioRad) in 0.1M buffer B and eluted with a linear gradient of 0.1-0.5M AS. Pol II eluted at approximately 0.3M AS. For TFIIH, purification started with nuclei from approximately 200L of HeLa cells. The P1M/Q0.4M fraction was loaded onto an anti-ERCC3 monoclonal antibody column. After binding, the column was washed extensively with 0.5M KCl HEGN (20 mM HEPES, pH 7.9; 0.1 mM EDTA, 10% glycerol, 0.1% NP-40) followed by elution with 1 mg/mL peptide in 0.2M HEGN. TFIID purification typically started with nuclei from approximately 200 L HeLa cells. The P1M/Q0.4M fraction (ca. 2 mg/mL) or the P1M/Q1M fraction was loaded onto an anti-TAF4 resin and washed with 40 column volumes 0.7M KCl TGEM (10 mM Tris, pH 7.9; 20% glycerol, 0.2 mM EDTA, 4 mM $MgCl_2$) followed by 20 column volumes 0.2M KCl TGEM. TFIID was then eluted from the resin with 1 mg/mL peptide in 0.2M KCl TGEM. A typical Mediator purification started with nuclei from 100L HeLa cells. The P1M/Q1M fraction is enriched in core Mediator (little/no detectable CDK8). GST-immobilized activation domains of VP16 (aa 413-490) or SREBP-1a (aa 1-50) were used. After binding, the resin was washed with 50 column volumes 0.5M KCl HEGN (0.1% NP-40), followed by 10 column volumes 0.15M KCl HEGN. Elution with buffer containing 30 mM glutathione (pH 7.6, 20 mM Tris, 0.1 mM EDTA, 10% glycerol, 0.15M KCl) was followed by separation over a 15%–40% glycerol gradient in 0.15M KCl HEGN with centrifugation at 50K RPM for 6 h at 4°C.

Purification of DSIF, NELF, HSF1, and P-TEFb

The two subunit DSIF complex (SPT4 and SPT5) was expressed in Rosetta2(DE3)pLysS cell (Novagen #71403). The expression plasmid was a gift from Dr. Rob Fisher. The four subunit NELF complex (NELF A, B, C, and E) was expressed in Rosetta2(DE3)pLysS (Novagen #71403) cells. The expression plasmids were a gift from Dr. W. Lee Kraus. The two subunit P-TEFb complex (CDK9 and CCNT1) was expressed in Sf9 cells at the UC Tissue Culture Shared Resource. For each of these complexes, the cell lysate was treated with benzonase, clarified, and loaded onto a Ni-NTA agarose column (Invitrogen #R90101). The column was then extensively washed with 0.5M NaCl buffer (pH 7.5; 50mM Tris, 5mM b-mercatoethanol, 10mM imidazole) and eluted in 0.15M KCl buffer (pH 7.5; 20mM Tris, 1mM b-mercaptoethanol, 0.5M imidazole). HSF1 was expressed in BL21(DE3) cells (Novagen #69450). The lysate from bacterial expression was clarified and purified in batch over a Ni-NTA agarose resin and washed and eluted as described above.

Recombinant expression and purification of holo-TFIID and partial TAF complexes

All TAFs and TFIID complexes, including holo-TFIID, were expressed using the MultiBac baculovirus system following published protocols (Fitzgerald et al., 2006). All proteins were full-length unless indicated otherwise.

For the recombinant human TAF1,7,11,13,TBP (S-TAF) complex, TAF1 comprising an N-terminal maltose binding protein (MBP) tag with a tobacco etch virus (TEV) NIa proteolytic cleavage site, TAF7 comprising an N-terminal hexa-histidine tag with a TEV NIa site, TAF11, TAF13 and TBP core (AA155-335) were used. Insect cell pellets comprising the S-TAF complex were resuspended in IMAC Buffer A (1x PBS pH 7.5, 400mM NaCl, 50mM imidazole, 10mM $MgCl_2$, 5% glycerol (w/v), complete protease inhibitor). Cells were lysed by freeze-thawing (twice), followed by centrifugation at 45,000xg for 60 min to clear the lysate. Cleared lysate was then applied to Ni-NTA resin, pre-equilibrated with IMAC Buffer A, followed by washing (10 column volumes) with IMAC Buffer A, then IMAC Buffer HS (1x PBS pH 7.5, 1000mM NaCl, 50mM imidazole, 10mM $MgCl_2$, 5% glycerol (w/v), complete protease inhibitor (Roche Molecular Biochemicals)) and then again with IMAC Buffer A. Bound S-TAF complex was eluted using IMAC buffer B (1x PBS pH 7.5, 400mM NaCl, 500mM imidazole, 10mM $MgCl_2$, 5% glycerol (w/v), complete protease inhibitor). Fractions containing the S-TAF complex were dialyzed overnight against MonoQ Buffer A (1x PBS pH 7.5, 100mM NaCl, 10mM $MgCl_2$, 5mM β-mercaptoethanol with complete protease inhibitor). S-TAF was further purified using ion exchange chromatography with a MonoQ column pre-equilibrated with MonoQ Buffer A. After binding, column was washed (5 column volumes) with MonoQ Buffer A and S-TAF eluted using a continuous gradient to MonoQ Buffer B (1x PBS pH 7.5, 1 M NaCl, 10mM $MgCl_2$, 5mM β-mercaptoethanol, complete protease inhibitor (Roche Molecular Biochemicals)) from 0%–100% linear gradient. The complex was further purified by size exclusion chromatography (SEC) with a Sephacryl S400 16/60 column in SEC Buffer 2 (25mM HEPES pH 7.5, 300mM NaCl, 10mM $MgCl_2$, 5mM β-mercaptoethanol, complete protease inhibitor). The human TAF3/TAF10 complex was purified from insect cell pellets using the same protocol that was used for the S-TAF complex. Both TAF3 and TAF10 possess a C-terminal hexa-histidine tag.

Recombinant human holo-TFIID was reconstituted *in vitro* from purified 8TAF, S-TAF and the TAF3/TAF10 complex in SEC Buffer 2 (25mM HEPES pH 7.5, 300mM NaCl, 10mM $MgCl_2$, 5mM β-mercaptoethanol, complete protease inhibitor) and purified by SEC in SEC Buffer 2. Recombinant human TFIID eluted in a symmetric peak. Peak fractions were pooled, concentrated to 0.6 mg/ml and holo-TFIID stored frozen in SEC Buffer 2 supplemented with 10% glycerol.

Statistical analysis of in vitro *data*

Statistical comparison of in-vitro data was performed using a Welch's t test, to determine a p value while accounting for variance in sample sizes. In the case where fold-change data was compared, the data was first log transformed and an Anderson-Darling test

was used to ensure that the data was not significantly different from normal ($p < 0.05$). A one-sided t test was then used to calculate a p value for the data relative to 0, which is the log-transformed value of a 1-fold change.

Calculation of Differential Motif Displacement Scores

Differential Motif Displacement scores were calculated as described (Tripodi et al., 2018), using the implementation provided by the Transcription Factor Enrichment Analysis program (https://github.com/Dowell-Lab/TFEA). The HOCOMOCO core database version 11 was used to indicate motifs of interest.

Analysis of Pause Index variability in cells, replicate-to-replicate

Variability in PI between biological replicates was determined by comparing the calculated PI between the replicates, as a fold change. PI was calculated as described (Ebmeier et al., 2017). Comparisons at the HSPA1B gene were used to match the *in vitro* promoter. PRO-seq data from this study (HCT116 cells) yielded inter-sample variability of 2.0 in control cells, and 2.7 in TAF1/2-depleted cells (fold change). For comparison, we also performed a similar analysis on the 12.5 minute heat shock time-point a different PRO-seq study (Mahat et al., 2016). In agreement, these data gave an inter-sample PI variability of 2.0 (fold change). This level of variability is consistent with that observed, replicate-to-replicate, in our *in vitro* reconstituted transcription system (Figure 1D).

Purification of recombinant TAF1 antibody

The antibody was purchased from the Recombinant Antibody Network (RAN; https://recombinant-antibodies.org); several TAF1 antibodies were tested and we determined that anti-TAF1-RAB-C413 performed best in TFIID IP experiments from nuclear extracts. Anti-TAF1 expression plasmids were transformed into OverExpress C43(DE3) chemically competent cells and expressed according to standard protocols. Protein was then isolated via batch purification over Protein A beads according to the Recombinant Antibody Network (RAN) protocol. For the IP tests, 1mL of anti-TAF1 C413 antibody lysate was added to 100 μl of washed and equilibrated protein A beads. The tube was nutated at 4°C for 1 hour and centrifuged at 400xg to removed unbound material. The resin was then washed 3 times in 0.5M HEGN and 3 times in 0.15M HEGN buffer. The anti-TAF1 resin was then incubated with 1mL of HCT116 nuclear extract and nutated at 4°C for 2 hours. The tube was centrifuged at 400xg, and the flow through was removed. The resin was then washed three times in 0.5M HEGN and three times in 0.15M HEGN buffer. Elutions were performed with 2% sarkosyl in 0.15M HEGN solution (2 × 100 μl at 4°C for 30 minutes). A negative control consisted of HCT116 nuclear extract added to Protein A beads without antibody.

TRIM-Away

The method used was adapted from Clift et al. (2017). HCT116 cells cultured in McCoy's 5A medium were grown to approximately 70% confluency. Media was aspirated off, and the cells were washed with PBS. 2ml of trypsin per plate were used to harvest adherent cells, after which an equal volume of Opti-MEM was added to each plate to neutralize the trypsin. Cells were combined in a 50ml centrifuge tube and spun down at 2,000xg for 5 minutes, then washed in PBS and spun down again at 2,000xg for 5 minutes. Cells were counted using a hemocytometer and diluted to 25 million cells/mL. 100 μl reactions were prepared, and cells were re-suspended in Buffer R and anti-TAF1 C413 antibody. A pulse only control was prepared, which consisted of cells suspended only in Buffer R. Transfections were performed using the Neon Transfection Kit (1530V, 1ms width, 1 pulse). Transfected cells were then pipetted into 1 mL of Opti-MEM in a 35mm dish and incubated at 37°C for 1 hour. The Opti-Mem media (containing some suspended cells) was then pipetted off and saved. 500 μl of PBS was added to the cells on the plates, which were then harvested and centrifuged at 6,000xg for 5 minutes. Supernatant was aspirated off, and cell nuclei were subsequently isolated. A small sample of cells (50 μl) were saved for analysis via western blot.

Quantitation of TFIID subunits

Whole cell extracts or nuclei + cytoplasm were isolated following either pulse-only control or TAF1/2 Trim-Away knockdown as described above. Nuclei were re-suspended in RIPA buffer with protease inhibitors, biorupted, and treated with nucleases. Protein concentrations of each fraction was determined and 10μg total protein was loaded onto 4%–20% gradient protein gels (BioRad 4%–20% Mini-PROTEAN® TGX Gel, 15 well, 15 μl cat#456-1096) and transferred onto a nitrocellulose membrane for western blotting. Westerns were scanned on an ImageQuant LAS 4000 series imager. ImageJ software was then used to measure band intensity, which was normalized to Actinβ for quantitation.

Antibodies

For western blotting, the following antibodies were used: TAF1 (1:1000, sc-735 X, Santa Cruz Biotechnology), TAF2 (1:500, ab103468, abcam), TAF4 (1:250, 07-1803, Millipore Sigma), TAF8 (1:250, ab204894, abcam), TBP (1:2000, sc-273, Santa Cruz Biotechnology), actin (1:1000, sc-47778, Santa Cruz Biotechnology), and histone H3 was a rabbit polyclonal from A. Shilatifard (Ebmeier et al., 2017). TAF1 Trim-Away was completed with the TAF1 C413 antibody from the Recombinant Antibody Network (RAN; https://recombinant-antibodies.org). Antibodies against *Drosophila* proteins were monoclonals 30H9 (Taf1) (Weinzierl et al., 1993) and 3E12 (Taf4) (Marr et al., 2006).

Drosophila RNAi and S2 nuclei isolation

RNAi was performed as described (Clemens et al., 2000) using 20-40 μg dsRNA. Cells were incubated with dsRNA for 2.5 d. Following RNAi with either TAF1 dsRNA or a LacI dsRNA control, cells were processed using the nuclei isolation steps as described (Mahat et al., 2016) before flash-freezing and storing at −80°C.

Measuring TAF1 knockdown from S2 cells

Following RNAi treatment samples were run on a SurePAGE Bis-Tris 4%–12% gel (GenScript) at 200V for 70 min. Protein was then transferred onto a nitrocellulose membrane (80V for 2hrs). For imaging and quantitation, membranes were exposed for sub-saturated times (BIO-RAD Chemidoc MP).

Sequencing data processing

The initial processing of all sequencing data was performed using the NascentFlow Pipeline (https://doi.org/10.17605/OSF.IO/NDHJ2), a data processing pipeline written in the Groovy programming language. The code for this pipeline can be found at https://github.com/Dowell-Lab/Nascent-Flow, with analysis for this experiment performed at commit 3fe1b7. Data were mapped to the hg38 reference genome for human cells, and to the dm6 reference genome for *Drosophila* S2 cells. For the remainder of the analysis, only the maximally expressed isoform of each gene was considered, which was determined by calculating the RPKM normalized expression over each isoform and selecting the one with the maximum RPKM expression. When different isoforms were determined across samples, the isoform from the first control sample was selected. In HCT116 cells, this was sample PO_1_S1_R1_001 whereas in S2 cells this was sample Control_1_S1_R1_001.

PRO-seq normalization with nuclear run-on

To test whether TAF1/2 knockdown caused a significant change in global gene expression, we directly compared run-on transcription levels in control and knockdown samples. A 50 μL aliquot (CTRL or TAF1/2-knockdown) containing 1 million nuclei was divided into 2 × 25 μL; one 25 μL aliquot was heated to 95°C for 5 minutes and the other was kept on ice. Then 25 μL of Reaction Buffer (5 mM Tris pH 8.0, 5 mM $MgCl_2$, 0.5 mM DTT, 150 mM KCl, 5 units of SUPERase-In, 0.5% Sarkosyl, 125 μM ATP, GTP, UTP, 2 μM CTP and 2.5μL a-^{32}P CTP) and incubated for 3 min at 37°C. RNA was isolated with 500μL of Trizol LS, followed by addition of 130 μL chloroform. Samples were then centrifuged for 10 min and the aqueous phase was transferred to a new tube. RNA was precipitated by adding 1 μL glyco-blue and 2.5 volumes 100% ethanol to each sample. Samples were incubated at room temperature for 10 minutes, then centrifuged for 20 minutes. Sample pellet was resuspended in 50 μL DEPC water. Buffer exchange was completed with a P-30 column according to the manufacturer's instructions (BioRad cat # 732-6250). Scintillation counts were measured by diluting 1 μL of sample in 1 mL of scintillation fluid.

Normalization using PRO-seq reads at 3′ ends of long genes

As another normalization method for PRO-seq libraries, we implemented a previously published approach (Mahat et al., 2016) that compares PRO-seq reads at the 3′ end of very long genes. The key assumption of this computational method is that genes exhibiting baseline levels of transcription will be unaffected by perturbations if they are long enough and the perturbation is short enough that its effects have not propagated through the entire gene body. Because Trim-Away was implemented for only 60 minutes, genes long enough to include in this analysis were calculated using polymerase elongation rate and treatment time (2kb/min * 60 min = 120kb) with an added 500bp to exclude the 3′ region, which often contains read pileup (final length cutoff = 120.5kb, n = 2,139). From this gene set, we determined the raw read counts within the region 120kb to −0.5kb from the TES and averaged this count between replicates. We then performed a linear regression on the count values for each of these genes (control and TAF1/2 knockdown). The results indicated that the baseline transcription between control and TAF1/2 knockdown cells was not significantly altered.

PRO-seq

Nuclei Preparation

After treatment, HCT116 cells (control or TAF1 TRIM-Away) were washed 3x with ice cold PBS, and then treated with 10 mL (per 15 cm plate) ice-cold lysis buffer (10 mM Tris–HCl pH 7.4, 2 mM $MgCl_2$, 3 mM $CaCl_2$, 0.5% NP-40, 10% glycerol, 1 mM DTT, 1x Protease Inhibitors (1mM Benzamidine (Sigma B6506-100G), 1mM Sodium Metabisulfite (Sigma 255556-100G), 0.25mM Phenylmethylsulfonyl Fluoride (American Bioanalytical AB01620), and 4U/mL SUPERase-In). Cells were centrifuged with a fixed-angle rotor at 1000 × *g* for 15 min at 4°C. Supernatant was removed and pellet was resuspended in 1.5 mL lysis buffer to a homogeneous mixture by pipetting 20-30X before adding another 8.5 mL lysis buffer. Suspension was centrifuged with a fixed-angle rotor at 1000 × *g* for 15 min at 4°C. Supernatant was removed and pellet was resuspended in 1 mL of lysis buffer and transferred to a 1.7 mL pre-lubricated tube (Costar cat. No. 3207). Suspensions were then pelleted in a microcentrifuge at 1000 × *g* for 5 min at 4°C. Next, supernatant was removed and pellets were resuspended in 500 μL of freezing buffer (50 mM Tris pH 8.3, 40% glycerol, 5 mM $MgCl_2$, 0.1 mM EDTA, 4U/ml SUPERase-In). Nuclei were centrifuged 2000 × *g* for 2 min at 4°C. Pellets were resuspended in 100 μL freezing buffer. To determine concentration, nuclei were counted from 1 μL of suspension and freezing buffer was added to generate 100 μL aliquots of 10 × 10^6 nuclei. Aliquots were flash frozen in liquid nitrogen and stored at −80°C.

Nuclear run-on and RNA preparation

Nuclear run-on experiments (HCT116 and S2 cells) were performed as described (Mahat et al., 2016) with the following modifications: the final concentration of non-biotinylated CTP was raised from 0.25 μM to 25 μM, and the final library clean-up and size selection was accomplished using 1X AMPure XP beads (Beckman).

Sequencing

Sequencing of PRO-seq libraries was performed at the BioFrontiers Sequencing Facility (UC-Boulder). Single-end fragment libraries (75 bp) were sequenced on the Illumina NextSeq 500 platform (RTA version: 2.4.11, Instrument ID: NB501447), demultiplexed and converted BCL to fastq format using bcl2fastq (bcl2fastq v2.20.0.422); sequencing data quality was assessed using FASTQC (v0.11.5) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and FastQ Screen (v0.11.0, https://www.bioinformatics.babraham.ac.uk/projects/fastq_screen/). Trimming and filtering of low-quality reads was performed using BBDUK from BBTools (v37.99) and FASTQ-MCF from EAUtils (v1.05). Alignment to the human reference genome (GRCh37/hg38) was carried out using Hisat2 (v2.1.0) (Kim et al., 2015) in unpaired, no-spliced-alignment mode with a GRCh37/hg38 index, and alignments were sorted and filtered for mapping quality (MAPQ > 10) using Samtools (v1.5) (Li et al., 2015). Gene-level count data for transcription start site (TSS, −30 to +300) and gene body (+301 to end) regions were obtained using featureCounts from the Subread package (v1.6.2) (Liao et al., 2013) with custom annotation files for single unique TSS and gene body regions per gene. Custom annotation files with single unique TSS and gene body regions per gene were generated as follows: 1) hg38 RefSeqCurated transcript-level annotation was downloaded from the UCSC genome table browser (09-07-2018), transcripts shorter than 1500bp and non-standard chromosome were removed, and only transcripts with unique start/stop coordinates per gene were retained; 2) Sense and anti-sense counts were tabulated and each candidate TSS region was ranked by sense and antisense reads to obtain a single 'most-active' TSS per gene; 3) Finally, per gene, the TSS was combined with the shortest gene body to avoid the influence of alternative transcription termination/polyadenylation sites. Differential expression analysis of gene body regions was assessed using the DESeq2 package (v1.22.1) (Love et al., 2014) with a custom R script (R v3.5.1 / RStudio v1.1.453 / Bioconductor v3.7) with cutoffs as described in text and figure legends. Analysis of RNAPII pausing was carried out using a custom R script (R v3.5.1 / RStudio v1.1.453) with the ggplot2 package (v3.1.0) used for visualizations. Gene level TSS and gene body counts were normalized by counts-per-million and by region length (cpm/bp), and Pausing Index (PI) calculated as the ratio of normalized reads in the TSS (cpm/bp) to normalized reads in the gene body (cpm/bp). Genes with < 0.5 cpm in all samples were excluded from analysis. Means of replicate values were used for plots and Wilcoxon/Mann-Whitney U tests. For genome browser snapshots, aligned reads were downsampled to the lower aligned read count per replicate using Samtools, to ensure equal contributions from each replicate, followed by merging of replicates and generation of coverage tracks in the bedgraph format using HOMER (v4.9.1) (Heinz et al., 2010). Genome browser snapshots were then generated from the bedgraph files using a custom R script (R v3.5.1 / RStudio v1.1.453 / Bioconductor v3.7) and the Gviz package (v1.26.4) (Hahne and Ivanek, 2016).

Metagene analysis

Each gene in the isoform-resolved reference sequence was either collapsed into a fixed sequence length originating from the TSS or TES (to better aggregate information, restricted to small regions) or divided into a fixed number of bins of variable length (to be used to accommodate genes of different length, but has some biases arising from length differences). The utility featurecounts (Liao et al., 2013) was then used to determine the total counts in those regions and the mean count and standard deviation of the mean were calculated. All counts over base pairs or bins were then plotted along with the standard deviation.

5′ end/3′ end gene ratio histogram

Long genes used for 3′ end normalization (n = 2139) were divided in half and reads were counted in each half (5′ end or 3′ end). For each gene, a ratio of 2nd half (3′ end) and 1st half (5′ end) reads was calculated. These results were plotted as a histogram for control cells and TAF1/2 Trim-Away cells.

Principal Component Analysis

PCA was performed using the standard prcomp function provided by the sva package for the R programming language (Leek et al., 2019). Batch effects from replicates completed on different days replicates were corrected using the removeBatchEffect function provided by the limma package (Ritchie et al., 2015) from the R programming language.

Differential Expression analysis

Differential expression analysis was performed using the DESeq2 package (Love et al., 2014) for the R programming language. Counts were generated using the utility featurecounts. Initial analysis using counts across the full annotated gene showed significant skew, indicating that the baseline assumptions of the differential expression model did not hold. To correct, counts in the region from +500 of the TSS to −500 from the TES (Transcription End Site) were used to obtain suitable model weights. Those model weights were then used when performing differential expression across the full gene, which corrected the skew effect.

Gene Set Enrichment Analysis

GSEA (Subramanian et al., 2005) was performed with the Broad Institute's GSEA software on the GenePattern Server using the pre-ranked module. Log(2) fold-change values were used as the rank metric for all genes and compared against the Hallmark gene sets database for enrichment.

QUANTIFICATION AND STATISTICAL ANALYSIS

PRO-seq experiments were completed in biological replicate. Statistical analysis of PRO-seq data is described in the Method Details. The number of replicates for each *in vitro* transcription experiment is indicated in the Figure Legends. Statistical analysis of *in vitro* transcription data is provided in Figure Legends and Method Details.

DATA AND CODE AVAILABILITY

The accession number for the PRO-Seq data (HCT116 and S2 cells) reported in this paper is GEO: GSE132764. Information about how to access the NascentFlow data processing pipeline is provided here (https://doi.org/10.17605/OSF.IO/NDHJ2). All code used in this analysis as well as additional documentation can be found in our GitHub repository at https://github.com/Dowell-Lab/Fant_2019_Analysis.

Molecular Cell, Volume *78*

Supplemental Information

TFIID Enables RNA Polymerase II Promoter-Proximal Pausing

Charli B. Fant, Cecilia B. Levandowski, Kapil Gupta, Zachary L. Maas, John Moir, Jonathan D. Rubin, Andrew Sawyer, Meagan N. Esbin, Jenna K. Rimel, Olivia Luyties, Michael T. Marr, Imre Berger, Robin D. Dowell, and Dylan J. Taatjes

Supplemental Information

TFIID enables RNA polymerase II promoter-proximal pausing

Fant, CB[1]; Levandowski, CB[1]; Gupta, K[2]; Maas, ZL[1]; Moir, J[1]; Rubin JD[1]; Sawyer, A[3]; Esbin, MN[1]; Rimel, JK[1]; Luyties, O[1]; Marr, MT[3]; Berger, I[2]; Dowell, RD[4,5]; Taatjes, DJ[1]

[1]Dept. of Biochemistry, University of Colorado, Boulder, CO, USA
[2]School of Biochemistry, Bristol Research Centre for Synthetic Biology, University of Bristol, UK
[3]Dept. of Biology, Brandeis University, Waltham, MA, USA
[4]Dept. of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, CO, USA
[5]BioFrontiers Institute, University of Colorado, Boulder, CO, USA

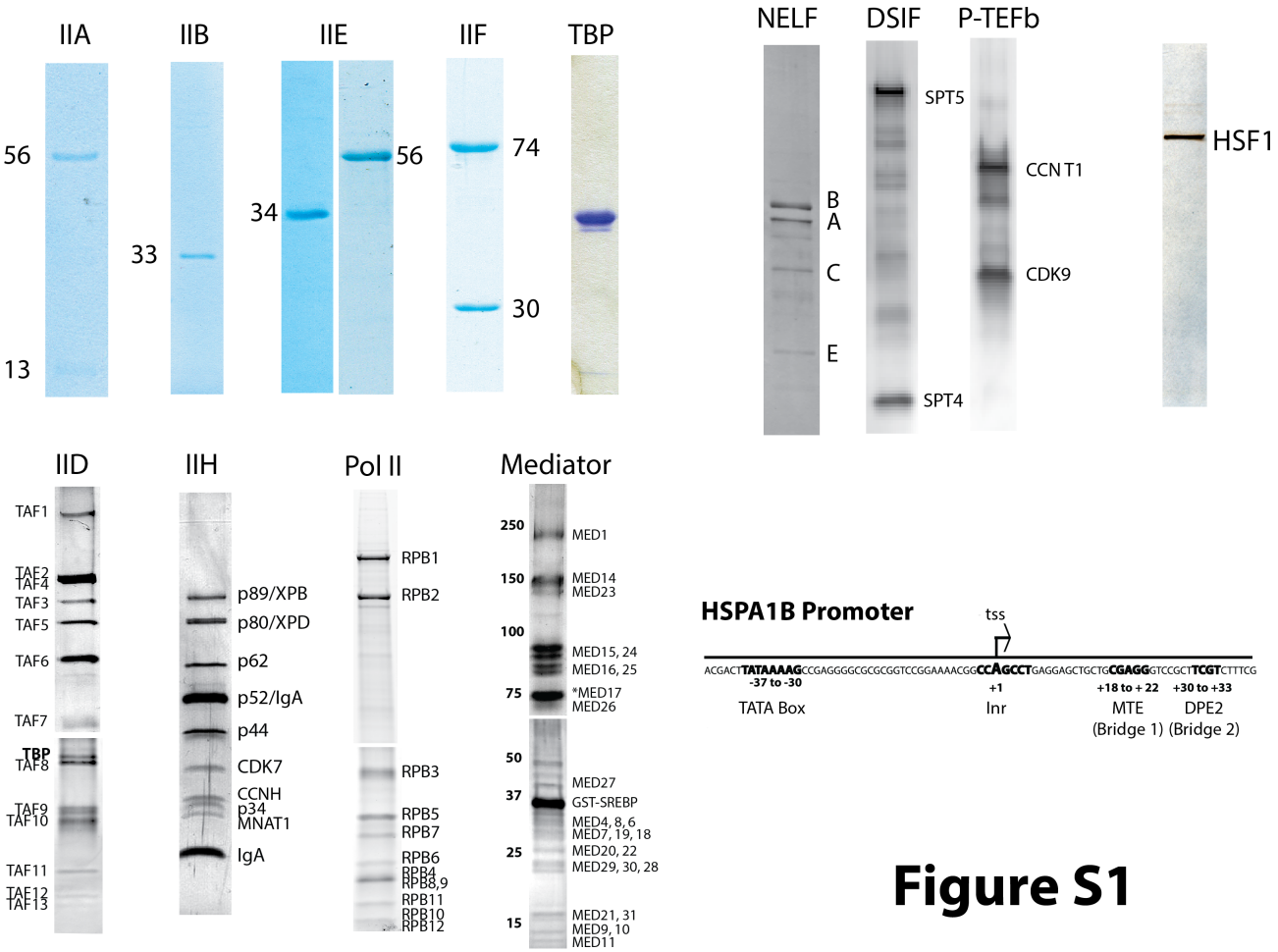


Figure S1

Figure S1. Purified human PIC factors used for biochemical reconstitution of RNAPII pausing (Related to Figure 1, 2). Also shown is a schematic of the native human HSPA1B promoter, including sequences around the transcription start site (TSS). Note that because of their many subunits with varying sizes, sections from two different silver-stained gels are shown for TFIID, RNAPII, and Mediator.

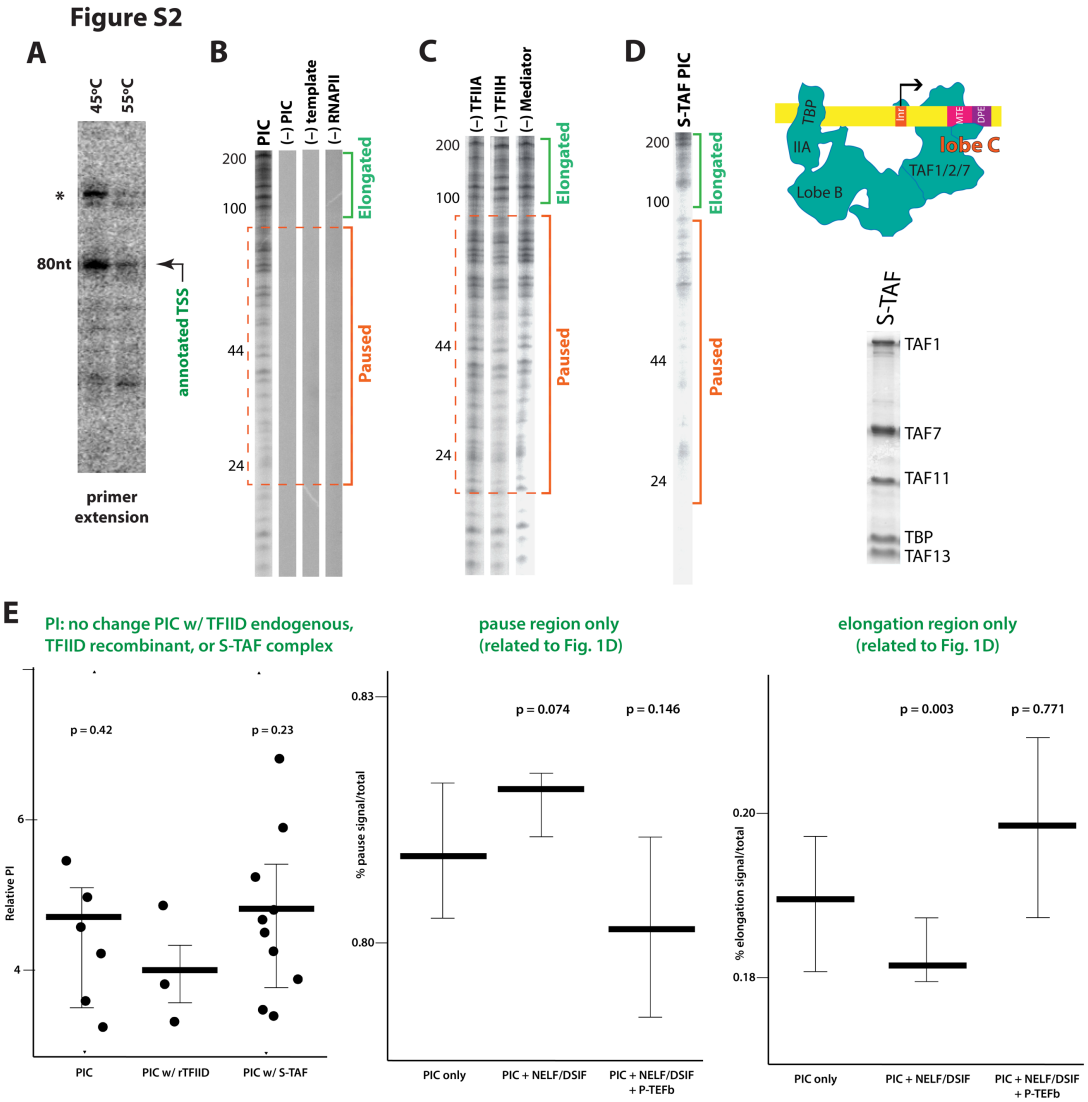


Figure S2. Additional controls for biochemical reconstitution of transcription with purified human factors (related to Figure 1). (**A**) Representative primer extension data, confirming that *in vitro* transcription from the native HSPA1B promoter initiates at the annotated transcription start site (TSS). The reverse transcriptase (RT) primer used anneals at +80 bases downstream of the TSS. Because AMV RT enzymes retain activity at higher temperatures, we could confirm that a higher-migrating band is likely an off-target annealing event (asterisk). The proportion of the annotated TSS product increases relative to the off-target (58% at 45ºC and 72% at 55ºC); furthermore, the annotated product is more robust to the increased temperature, retaining 47% of its intensity vs. only 26% for the off-target. The overall levels of PE products decrease at higher temperature because AMV RT enzymes have optimal activity at lower temperatures. (**B**) Additional *in vitro* transcription controls, in which each reaction was completed as described in **Figure 1**, except that specific factors were left out. (**C**) Examples of dropout experiments in which specific PIC factors were removed from the assay, to assess a potential dependence on RNAPII promoter-proximal pausing. Because these experiments were completed with incomplete PICs, overall transcriptional output was compromised; however, the data suggest that RNAPII pausing is not dependent on the factors excluded in these experiments. Reactions lacking HSF1 showed overall reduction of transcription but no major change in paused vs. elongation regions. (**D**) Partial TFIID S-TAF complex, which contains TBP and lobe C subunits TAF1 and TAF7, enables pol II pausing; at right is TFIID structural scheme and coomassie-stained gel of the purified S-TAF complex. (**E**) Summary of *in vitro* transcription data, related to **Figure 1** and **2**. (Left) Related to Figure 2; relative Pause Index upon comparison of PIC with endogenous purified TFIID, recombinant TFIID (rTFIID) and S-TAF. Note that PI is similar in each case; dots represent independent experiments. (Center) Related to **Figure 1D**, including p-values; analysis of ^{32}P-signal in pause region only vs. total ^{32}P-signal/lane. Comparisons show increased pausing with addition of NELF/DSIF and reduced pausing upon addition of P-TEFb, consistent with PI data shown in **Figure 1D**. (Right) Related to **Figure 1D**, including p-values; analysis of ^{32}P-signal in elongation region only vs. total ^{32}P-signal/lane. Comparisons show decreased elopngation with addition of NELF/DSIF and increased elongation upon addition of P-TEFb, consistent with PI data shown in **Figure 1D**.

Figure S3

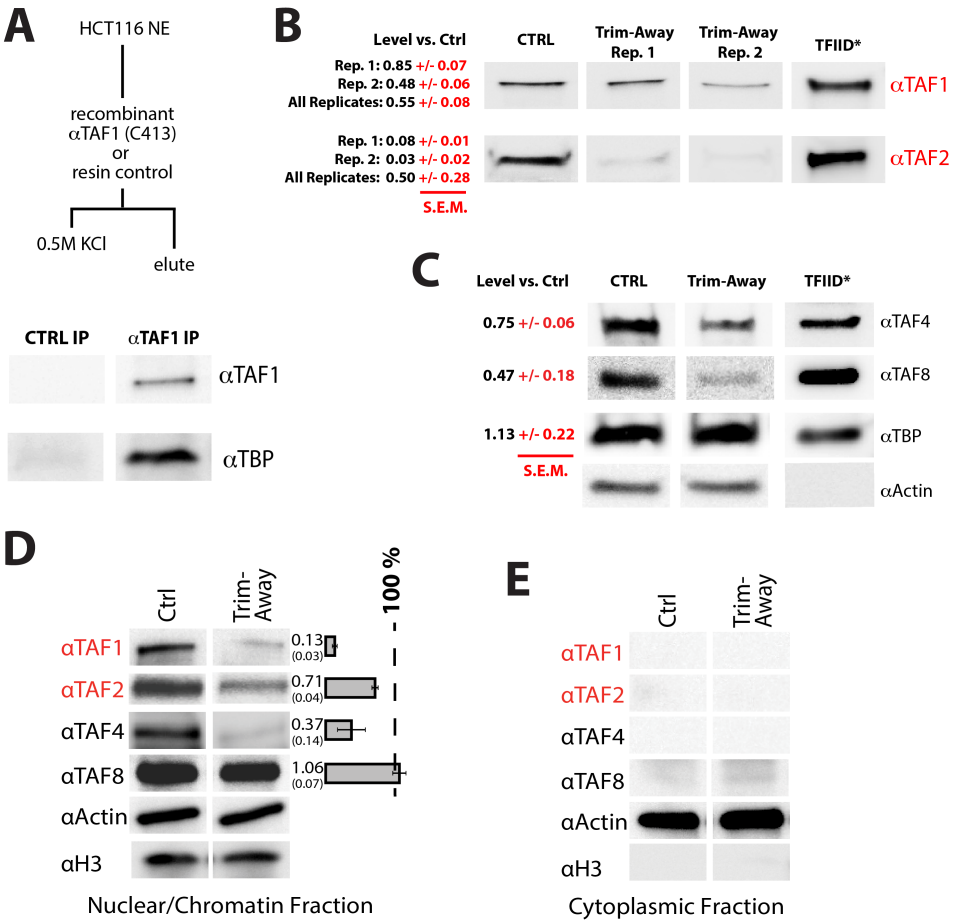
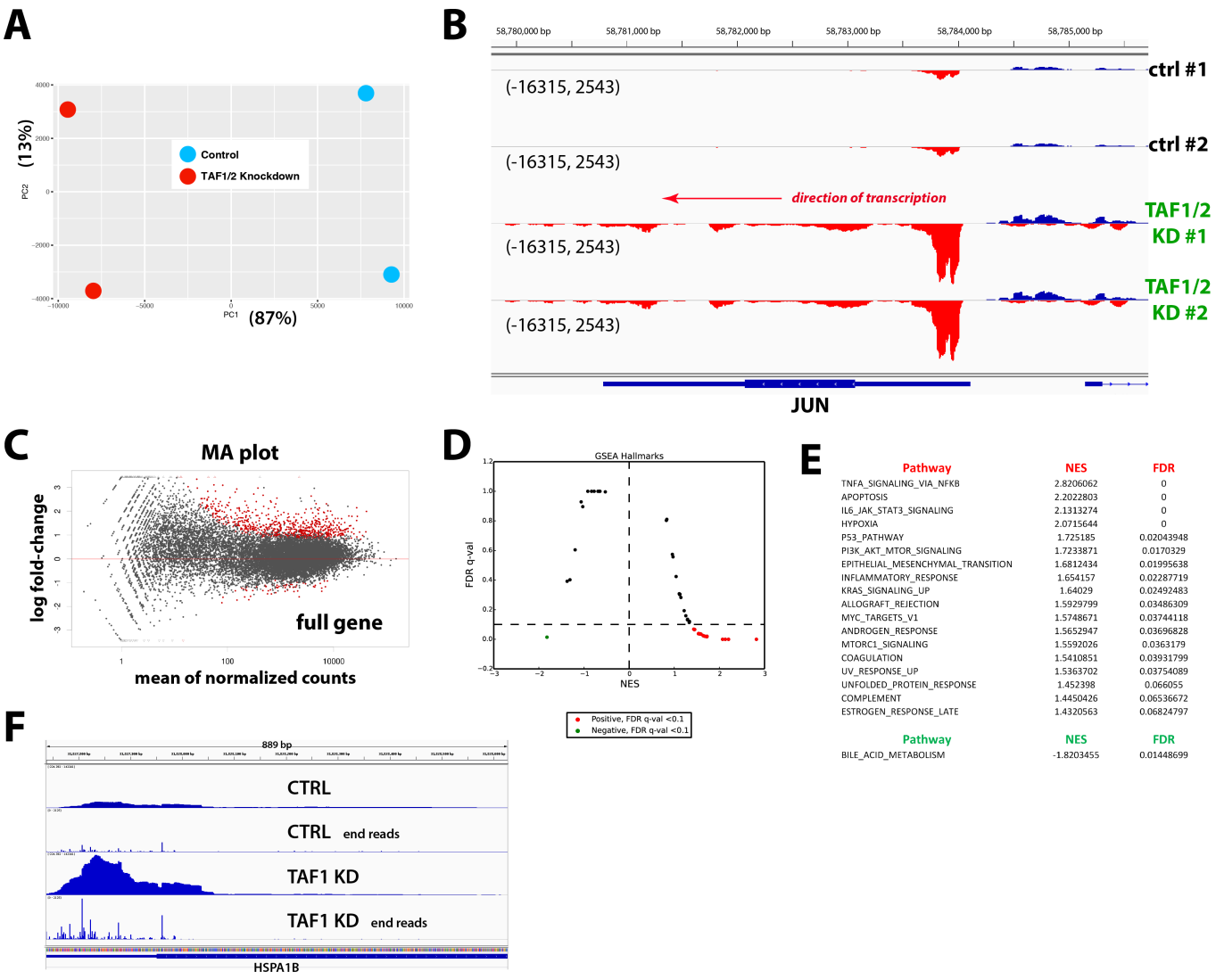


Figure S3. Additional control experiments for TAF1 TRIM-Away in HCT116 cells (related to Figure 3 & 4). (**A**) Human TAF1 antibody validation. TAF1 antibodies were immobilized onto protein A resin, incubated with HCT116 nuclear extract, washed extensively with 0.5M KCl buffer, and eluted. Eluted material was probed for TFIID subunits by western. (**B**) Quantitative westerns to compare control vs. TAF1 TRIM-Away experiments. Data for biological replicate 1 & 2 represent samples used for PRO-seq experiments (2 technical replicates for each biological replicate). Additional TRIM-Away experiments were completed to estimate reproducibility of knockdown and to assess effects on other TFIID subunits (see panel **C**). For TAF1 "all replicates" includes 4 biological and 13 technical replicates; TAF2 = 4 biological and 8 technical replicates. Numbers is red represent the standard error of the mean. (**C**) Quantitative westerns to probe TAF1 TRIM-Away effects on other TFIID subunits. Data for TAF4 represent 2 biological and 4 technical replicates; TAF8 = 1 biological and 3 technical; TBP: 2 biological and 5 technical. Numbers is red represent the standard error of the mean. *Purified human TFIID was included as a positive control for antibody specificity. (**D**) Additional experiments confirmed that TAF1 Trim-Away depleted TFIID subunits on chromatin (see Methods). Quantitation of each protein represented by 2 technical replicates. (**E**) Cytoplasmic fractions were also evaluated for comparison.

Figure S4



Pathway	NES	FDR
TNFA_SIGNALING_VIA_NFKB	2.8206062	0
APOPTOSIS	2.2022803	0
IL6_JAK_STAT3_SIGNALING	2.1313274	0
HYPOXIA	2.0715644	0
P53_PATHWAY	1.725185	0.02043948
PI3K_AKT_MTOR_SIGNALING	1.7233871	0.0170329
EPITHELIAL_MESENCHYMAL_TRANSITION	1.6812434	0.01995638
INFLAMMATORY_RESPONSE	1.654157	0.02287719
KRAS_SIGNALING_UP	1.64029	0.02492483
ALLOGRAFT_REJECTION	1.5929799	0.03486309
MYC_TARGETS_V1	1.5748671	0.03744118
ANDROGEN_RESPONSE	1.5652947	0.03696828
MTORC1_SIGNALING	1.5592026	0.0363179
COAGULATION	1.5410851	0.03931799
UV_RESPONSE_UP	1.5363702	0.03754089
UNFOLDED_PROTEIN_RESPONSE	1.452398	0.066055
COMPLEMENT	1.4450426	0.06536672
ESTROGEN_RESPONSE_LATE	1.4320563	0.06824797

Pathway	NES	FDR
BILE_ACID_METABOLISM	-1.8203455	0.01448699

Figure S4. Additional details from PRO-seq experiments in HCT116 cells (related to Figure 3 & 4). (**A**) PCA from each replicate (batch corrected; control TRIM-Away vs. TAF1 TRIM-Away knockdown) shows controls cluster separately from TAF1 knockdown samples. (**B**) Genome browser views showing the consistency of results across biological replicates. PRO-seq data at the JUN locus is shown. (**C**) Full gene MA plot (TSS to polyA site) showing general up-regulation of transcription in TAF1 TRIM-Away cells. (**D**) Moustache plot of false discovery rate (FDR-q) vs. normalized enrichment score (NES) of hallmark gene sets analyzed by GSEA. Differences (increase = red; decrease = green) are highlighted in TAF1-knockdown cells vs. controls. (**E**) Listing of Hallmark gene sets highlighted in **D**; up-regulated hallmarks suggest onset of stress responses in TAF1/2-knockdown cells. (**F**) Mapping the ends of PRO-seq reads shows that the mode (maximum point) of 5'-end reads at HSPA1B aligns at the +70 position. This is downstream from the canonical pausing position due to the use of biotinylated CTP only in the PRO-seq protocol. Our analysis uses TFit software, which analyzes the whole read instead of only the read-ends and provides more accurate classification vs. dREG software (Azofeifa and Dowell, 2017).

Figure S5

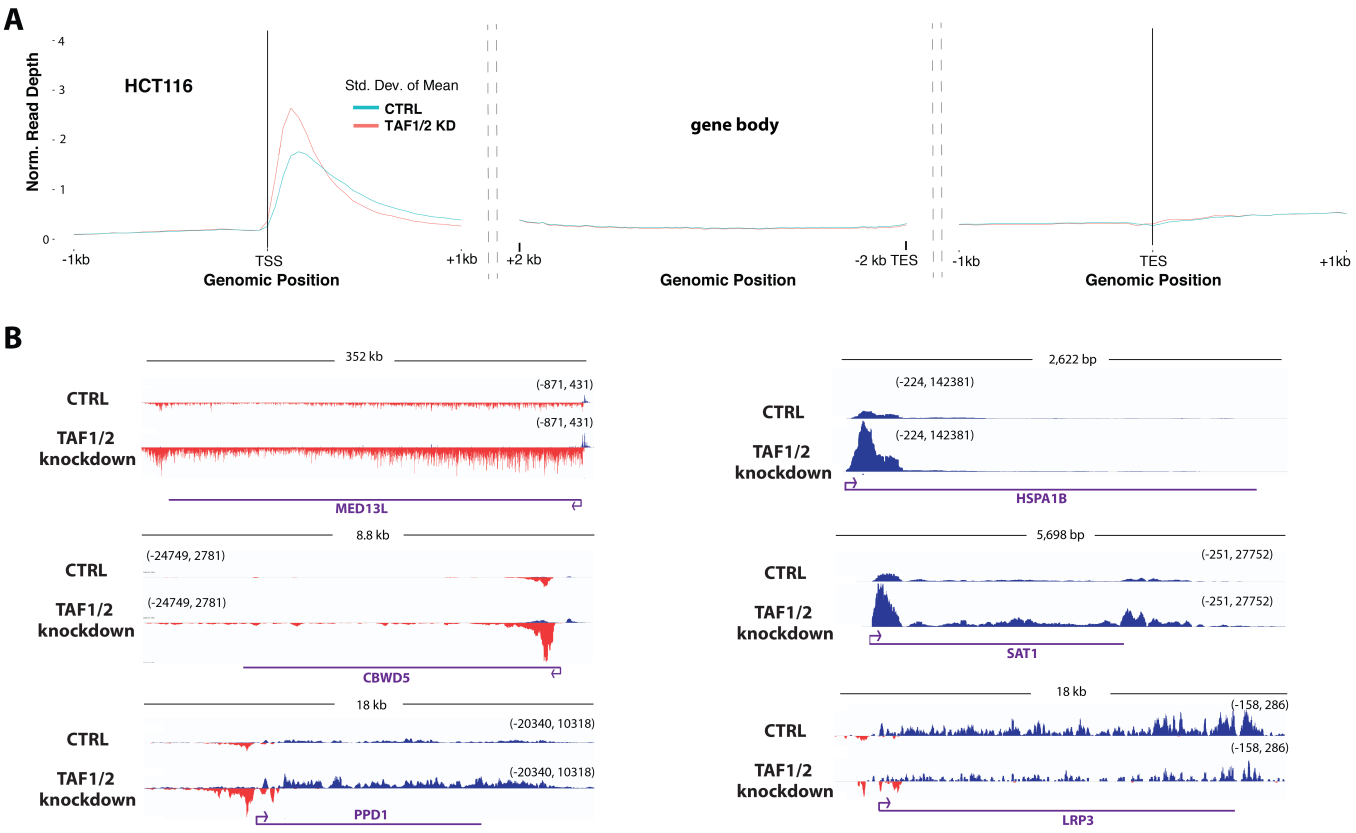


Figure S5. PRO-seq data suggest increased pause release in TAF1/2-knockdown cells (related to Figure 3 & 4). (**A**) Metagene plots for all gene regions (5'-end/TSS, gene body, 3'-end/TES) in HCT116 cells, comparing TAF1/2 knockdown with control cells (n = 18687). The increased reads at gene 5'-ends are consistent with coupled increases in pause release and re-initiation (Gressel et al., 2017; Shao and Zeitlinger, 2017). (**B**) Additional genome browser views of PRO-seq data from HCT116 cells. Note that the HSPA1B locus is shown at upper right and not all annotated genes show evidence of increased transcription upon TAF1/2 knockdown (LRP3, lower right).

Figure S6

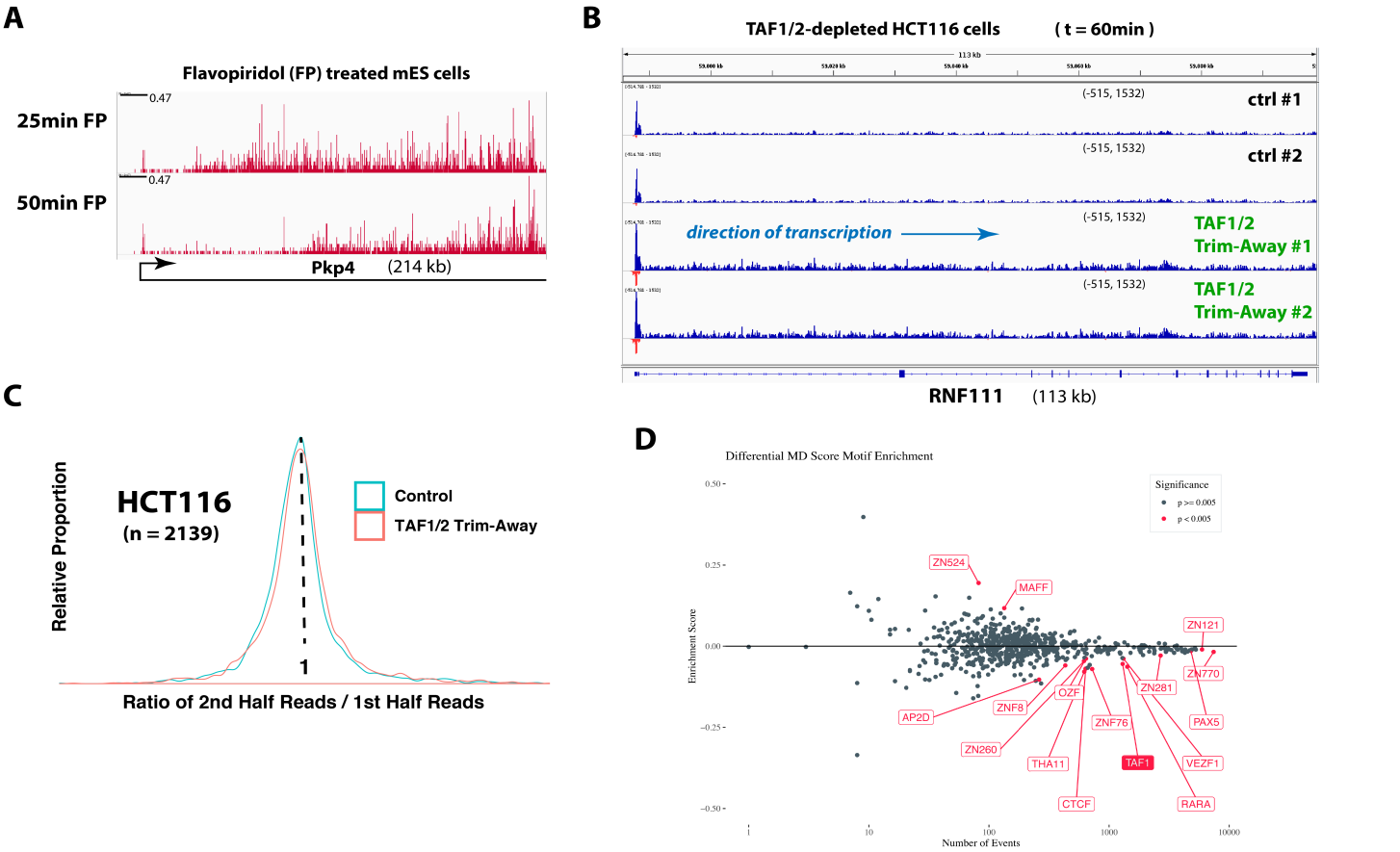


Figure S6. TAF1/2 depletion does not increase pausing (related to Figure 3 & 4). (**A**) A positive control for increased pol II pausing is PRO-seq data from flavopiridol-treated mES cells (Jonkers et al., 2014). Note that at long genes in particular, a characteristic loss of reads is observed toward the TSS as pausing increases and pause release is blocked. (**B**) Such features are absent in TAF1/2-depleted cells (t=60 min), with biological replicates from the RNF111 locus shown as an example. (**C**) Compiling data from long genes in TAF1/2 depleted cells (n=2139), no significant difference in PRO-seq reads was detected in the first half (toward 5'-end) vs. the second half (toward 3'-end). The ratio is 1:1; this result is in stark contrast to what is observed in FP-treated cells (panel A), confirming that pausing does not increase in TAF1/2-depleted cells. (**D**) Differential Motif Displacement (MDD) scores were calculated as described (Tripodi et al., 2018). The MDD score quantifies changes in transcription factor (TF) activity across samples. Points colored in red are significantly different after TAF1/2-depletion at $p < 0.005$. Note TAF1 TF activity is reduced after depletion, as expected.

Figure S7

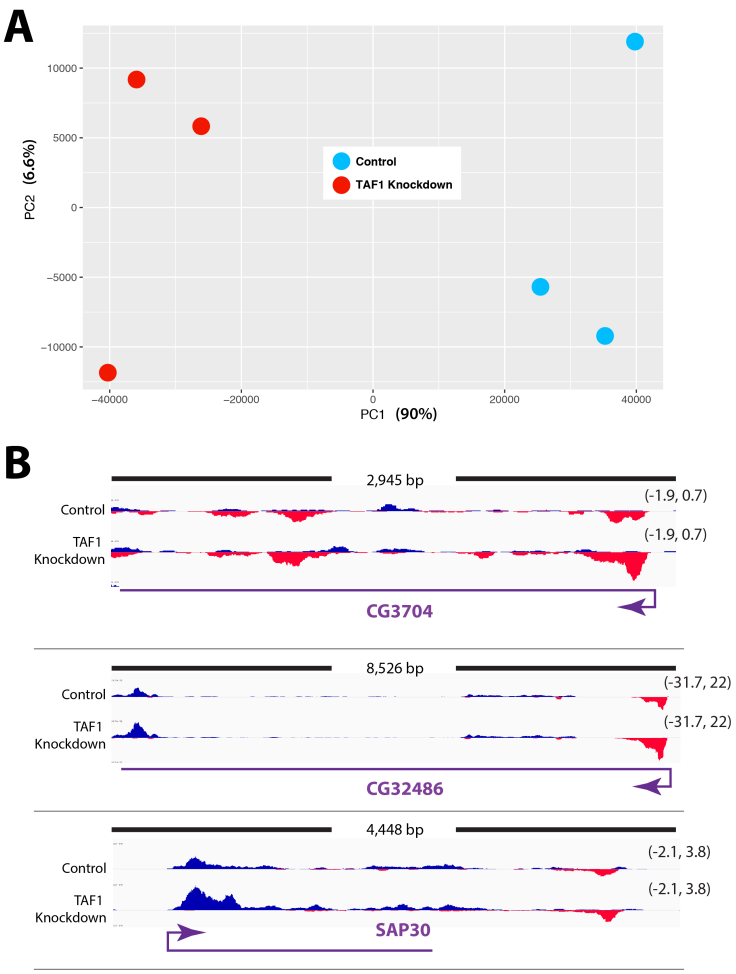


Figure S7. PRO-seq data suggest TFIID lobe C function in RNAPII pausing is conserved in *Drosophila* S2 cells (related to Figure 4). (**A**) Representative western blots showing TAF1 knockdown in S2 cells. TAF1 knockdown was determined to be 88% (± 6.7) from 3 biological replicates. (**A**) PCA from each replicate (batch corrected; control RNAi vs. TAF1 knockdown) shows that controls cluster separately from TAF1 knockdown samples. (**B**) Example genome browser views of PRO-seq data from S2 cells showing results similar to TAF1/2-knockdown HCT116 cells. Increased transcription is observed generally at gene 5'-ends, which extend beyond the promoter-proximal pause site.