

High-resolution binding data of TFIID and cofactors show promoter-specific differences

Received: 12 January 2026

Accepted: 3 September 2026

Published online: 07 October 2026

Sergio G-M Alcantara ¹, Simon Bourdareau ¹, Melanie Weilert ¹✉ & Julia Zeitlinger ^{1,2}

TFIID is instrumental in recognizing promoter sequences and initiating transcription, yet a cohesive understanding of how this complex engages and functions at different promoter types in vivo is still lacking. Here, we employed ChIP-nexus to capture high-resolution binding footprints of all *Drosophila* TFIID subunits across the genome. These footprints reveal TFIID sub-modules whose DNA contacts agree with cryo-EM structures and point to new structural details. At different promoter types, the footprints of the TAFs are very similar, suggestive of the same engaged TFIID conformation. In contrast, the binding profile of TBP is promoter-specific, enabling us to identify TATA, DPR, and TCT/housekeeping promoters de novo, along with their underlying core promoter elements. Notably, TATA promoters display specific TBP and NC2 binding footprints and reduced TAF occupancy, consistent with a fraction of transcripts being initiated in a TAF-independent manner at these promoters in vivo. Together, our results provide a comprehensive resource for linking structural and biochemical results to in vivo data and offer a potential explanation for the increased burst size observed at TATA promoters.

TFIID is a multi-subunit complex crucial for recognizing core promoter sequences, assembling the pre-initiation complex (PIC), and initiating RNA polymerase II (Pol II) transcription at transcription start sites (TSSs)¹. TFIID is composed of TATA-binding protein (TBP) and 13 TBP-associated factors (TAFs), together forming a three-lobed structure when reconstituted in vitro². As revealed by elegant cryo-EM studies, these lobes become reorganized when TFIID binds the promoter^{3–6}. One lobe of TFIID binds to the downstream region (a.k.a. lobe C), while another lobe engages the upstream core promoter region (a.k.a. lobe B) together with the third lobe (a.k.a. lobe A), which is more flexible and structurally more difficult to characterize^{5,6}. During this reorganization, TFIID places TBP near the site of transcription initiation, which is likely a critical step for PIC formation^{5–9}.

Despite remarkable advances, TFIID is best understood in vitro, while its role in vivo across various promoter types remains incompletely understood. Different promoter types produce transcripts with distinct bursting properties, but TATA promoters produce transcription bursts with more transcripts than TATA-less promoters^{10–15}. Cryo-

EM structures suggest that various promoter types are bound by engaged holo-TFIID⁵, but whether TFIID plays different functional roles across promoter types is not clear.

Since TFIID is very flexible in vitro and sometimes found in sub-stoichiometric compositions^{16–20}, it is possible that TFIID preferentially binds some promoter types, adopts different conformations, or forms distinct sub-complexes. Alternatively, TFIID could influence the transcription initiation pattern by differential TBP mobilization. TBP's contacts appear to be highly regulated since it is bound by several TAFs (TAF1, TAF11/13)⁸, as well as by NC2 and Mot1^{21–27}. These observations raise the possibility that TFIID binding differs across promoter types, an idea that remains untested in vivo.

To investigate TFIID function in vivo, we performed ChIP-nexus experiments to map the occupancy of all 14 TFIID subunits at high-resolution across the *Drosophila* genome, in which core promoters are well characterized²⁸. Our data reveal that TFIID's DNA-bound modular organization is broadly conserved across promoters, but that TBP and NC2 display striking promoter-specific binding patterns. Notably, the

¹Stowers Institute for Medical Research, Kansas City, MO, USA. ²Department of Pathology & Laboratory Medicine, The University of Kansas Medical Center, Kansas City, KS, USA. ✉e-mail: jbz@stowers.org

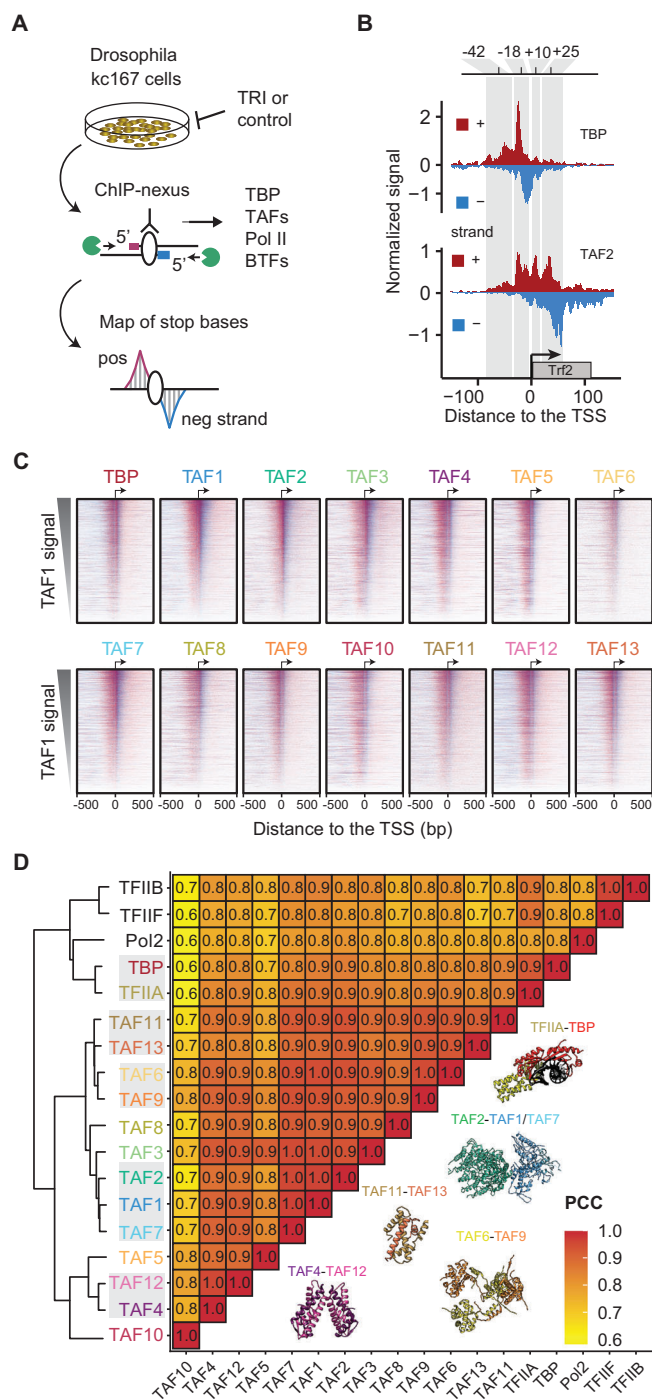


Fig. 1 | TFIID subunits show highly coordinated binding across active promoters. **A** Schematic representation of the ChIP-nexus workflow. **B** TBP and TAF2 binding profile at the *Trf2* gene, shown as normalized reads for each stop base on the positive strand (red) and the negative strand (blue). **C** Binding of each TFIID subunit across active promoters with narrow transcription initiation determined by CAGE-seq data ($n = 1769$). For each factor, the ChIP-nexus signal is shown for the positive (red) and negative (blue) strands. Promoters were sorted based on the total binding of TAF1 within 100 bp centered on the TSS. **D** Pearson correlation coefficients (PCC) and hierarchical clustering of the binding of TFIID and PIC components across narrow promoters. The total binding signal was measured at each promoter within 100 bp centered on the TSS. Proteins known to physically interact (see structures) often clustered first (marked in gray). All protein structures were obtained from PDB: TFIID-TBP (pdb_00001nvp), TAF1-7-2 (pdb_00006mzm), TAF4-12 (pdb_00001h3o), TAF6-9 (pdb_00006mzd) and TAF11-13 (pdb_00001bh8).

TBP binding profile at TATA promoters suggests an additional mode of initiation that is independent, or less reliant on TAFs. We propose that this mechanism may contribute to the large transcription burst size of TATA promoters.

Results

ChIP-nexus captures all TFIID subunits at promoters

To robustly detect the patterns by which each TFIID subunit binds to core promoters in vivo at the highest resolution, we used ChIP-nexus²⁹. ChIP-nexus is a chromatin immunoprecipitation coupled to exonuclease (ChIP-exo) technique, where a 5' exonuclease digests the DNA of formaldehyde-fixed chromatin until encountering a protein-DNA crosslink. The DNA libraries produced from pulling down each TFIID subunit then show precise footprints, consisting of strand-specific peaks on either side of where the TFIID subunit is directly or indirectly crosslinked to DNA (Fig. 1A). We performed ChIP-nexus for all TFIID subunits in *Drosophila* Kc167 cells under normal conditions (DMSO control), and after treatment with triptolide (TRI), which blocks transcription at the PIC stage and, consistent with previous studies^{30–33}, yields higher occupancy of PIC and TFIID components. After confirming that the biological replicate experiments were highly reproducible (Fig. S1), we analyzed the pooled ChIP-nexus coverage of each TFIID subunit. All experiments showed strong and distinct binding footprints of TFIID subunits at promoter regions (Fig. 1B).

We then chose active narrow promoters for our analysis since the positional information of high-resolution ChIP-nexus footprints is most interpretable when initiation occurs at a well-defined site. We used CAGE-seq data from Kc167 cell nuclear extracts (publicly available from modENCODE: ENCSR827COJ) and selected promoters where the annotated TSS shows transcription initiation within 11 bp ($n = 1769$; Fig. S2; see “Methods”). When the promoters were sorted by their TAF1 occupancy, the other TFIID subunits followed the same intensity pattern (Fig. 1C), suggesting that their binding levels are overall similar.

We next clustered the TFIID subunits based on their binding level similarity across the promoters, measured as pairwise Pearson correlation coefficients (PCCs). We included ChIP-nexus data for the PIC transcription factors TFIID, TFIIF and PolII for comparison, which showed high correlations with the TFIID subunits (median PCC = 0.85) (Fig. 1D). Notably, TBP had the highest correlation with TFIID (Fig. 1D), consistent with a direct physical interaction between TBP and TFIID^{5,6,34,35}. Importantly, the TAFs were overall all highly correlated in their binding levels (median PCC = 0.87), and the most correlated TAFs (TAF4/12, TAF1/7/2, TAF11/13 and TAF6/9) were also known to each form a physical complex⁹ (Fig. 1D). This suggests that high correlations are indicative of physical interactions, consistent with formaldehyde fixation generating abundant protein-protein crosslinks^{36,37} and causing interacting proteins to be pulled down with the same protein-DNA contacts in ChIP-nexus experiments^{33,38–41}. Taken together, these results show that TFIID subunits exhibit highly coordinated binding across active promoters and suggest an in vivo organization that aligns well with the known in vitro structure of TFIID.

High-resolution footprints reveal the promoter-bound conformation of TFIID in vivo

To determine where TFIID subunits contact DNA, we analyzed the base-resolution ChIP-nexus binding footprints. Interacting proteins are often crosslinked to each other and pulled down together, thereby producing similar footprints in ChIP-exo/nexus data^{33,38–41}. Exact protein-protein interactions cannot be inferred since formaldehyde crosslinks do not occur at all amino acids³⁷, but as a cumulative measure, we expect the ChIP-nexus binding profiles to reflect the structural organization of TFIID in vivo.

To analyze where the DNA contacts occur for each factor, we used the ChIP-nexus data of TBP and the 13 TAFs after triptolide treatment and averaged across the promoters with narrow transcription

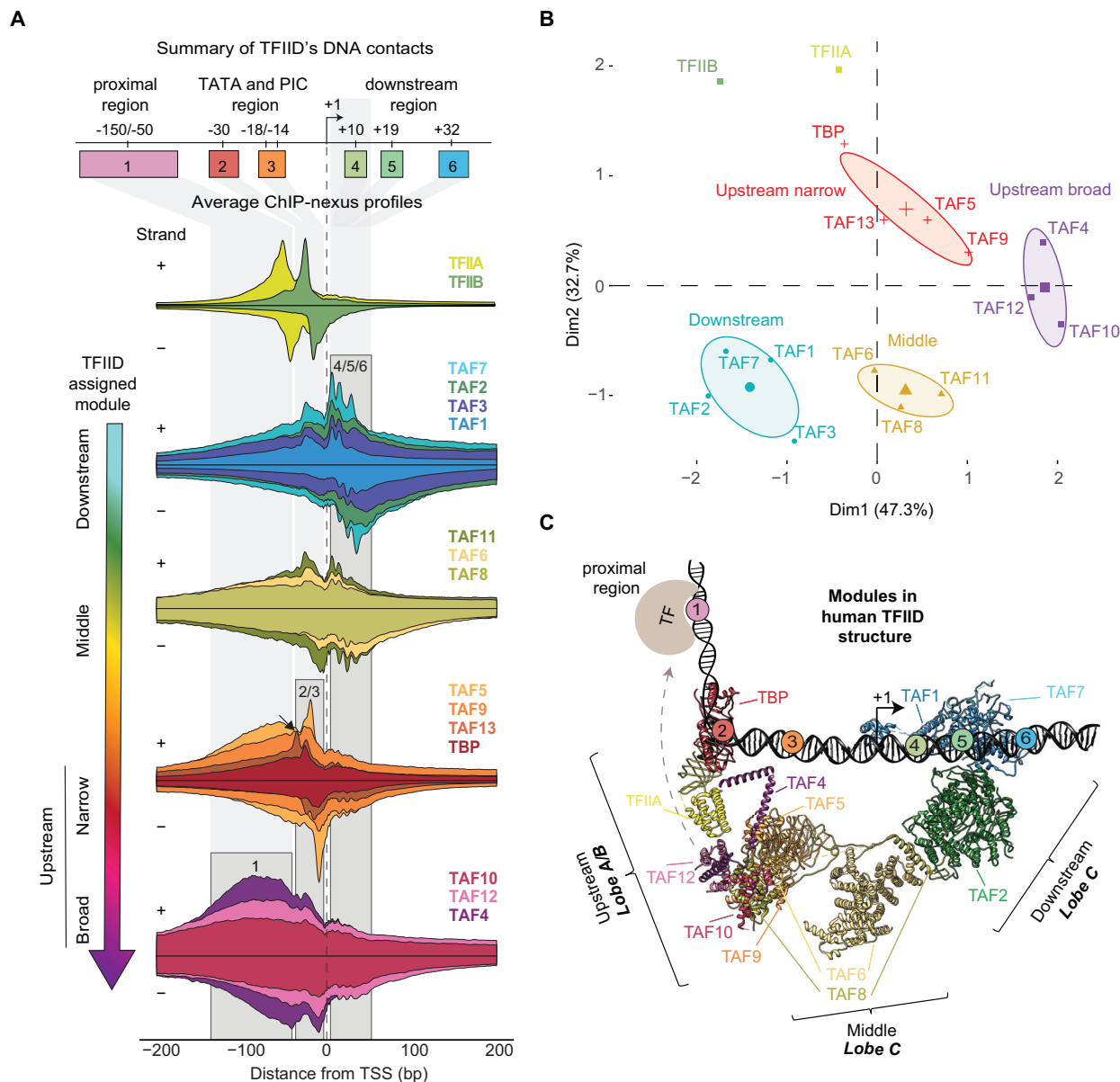


Fig. 2 | High-resolution ChIP-nexus footprints reveal a four-module organization of promoter-bound TFIID in vivo. **A** The average ChIP-nexus binding profile of the TFIID subunits centered at the TSS of narrow promoters ($n = 1769$) shows a four-module organization of the promoter-bound TFIID complex. For each subunit, the binding profile on the positive strand (top line) and negative strand (bottom line) is shown in the same color and stretched to an arbitrary scale to facilitate comparison with other subunits. The black arrow points to the footprint of TAF13 at -30. **B** Principal Component Analysis (PCA) of the averaged ChIP-nexus profiles

confirms the modular organization. Modules were defined based on PCA clustering and similarity in binding profiles. **C** Cryo-EM structure of human promoter-bound TFIID (PDB: pdb_00006mzm) is consistent with the modular organization and DNA contacts identified by the ChIP-nexus experiments. The hTFIID structure is derived from the synthetic Super Core Promoter (SCP) template³, which contains TATA, Inr, MTE, and DPE elements, all of which are enriched in the narrow promoter set selected for the analysis in (A).

initiation (Fig. 2A). As expected, promoters with broad transcription initiation exhibited substantially more diffuse average profiles (Fig. S3). Since ChIP-nexus produces strand-specific peaks on either side of the protein-DNA crosslink site, we calculated the midpoint between matching peaks on the positive strand and negative strand to identify the protein-DNA contacts in each average profile (see Fig. S3 left). We note that the DNA-bound organization of TFIID analyzed here reflects an aggregate pattern across diverse promoter classes (Fig. S1C).

The preferred DNA contacts, as well as the modular organization, align well with cryo-EM structures of promoter-bound human and yeast TFIID (Fig. 2C)³⁻⁶. The main footprint positions were found in six distinct regions, mapped at -150/-50, -30, -18/-14, +10, +19 and

+32 bp relative to the TSS (Regions 1-6 in Fig. 2A). In agreement with Principal Component Analysis (PCA) of the average profiles, we manually grouped the 14 TFIID subunits into four modules, which we call the downstream, middle, upstream narrow and upstream broad module (Fig. 2A, B). TFIIA and TFIIB were separated from the TAF modules, but closer to TBP in PCA space, consistent with their TBP association in the PIC. In the untreated (DMSO) condition, similar TFIID modules were observed (Fig. S4A, B), indicating that they are not specific for triptolide treatment. This suggests that our data can provide insights into the promoter-bound organization of TFIID in vivo.

Our downstream module is the lobe C that engages with the downstream promoter region (DPR). Here, TAF1, TAF7, and TAF2 form three clustered footprints at +10, +19, and +32 bp (Fig. 2A, regions

4–6), consistent with the human TFIID cryo-EM structures^{3,5,6} and contradicting previous biochemical experiments implicating *Drosophila* TAF6/9 in binding to DPR elements⁴². Interestingly, TAF3 is also included in our downstream module. TAF3 is part of the flexible lobe A, whose location in promoter-bound TFIID is not well established^{5,6}, but the proximity of TAF3 near TAF1/2/7 is supported by crosslinking mass spectrometry analysis in yeast⁴. The presence of TAF3 in this module should allow the PHD finger of TAF3 to interact with the H3K4me3 mark of the +1 nucleosome downstream of the TSS^{9,43,44}, though we note that the levels of TAF3 (and TFIID in general) do not correlate well with the levels of H3K4me3 (Fig. S4C).

The middle module shows footprints both upstream and downstream of the TSS (Fig. 2A, regions 2–5). As expected, it contains TAF6 and TAF8, which are known to bridge the upstream and downstream module in cryo-EM structures and thus indirectly contact the promoter both upstream and downstream (Fig. 2C). More unexpectedly, TAF11 clusters within the middle module, even though TAF11 forms a heterodimer with TAF13, which is found in the upstream module⁴⁵. TAF11/13 are known to interact with TBP and inhibit its binding to DNA in the canonical state of TFIID. Thus, in the conformation of TFIID where TBP engages the promoter, TAF11/13 must relocate or disengage from TBP^{5,6}. One possible explanation is that the TAF11/TAF13 heterodimer adopts a flexible or repositioned configuration upon TBP engagement, allowing TAF11 to shift closer to the middle module while TAF13 remains positioned upstream.

The narrow upstream module consists of TBP, TAF5, TAF9, and TAF13, which are from the two lobes that engage the upstream promoter region (lobe A and B). The density for these TAFs is poorer in cryo-EM structures^{4–6}, but their overall location agrees with our data. This module shows the strongest footprints around the site of PIC formation at –18/–14 bp (Fig. 2A, regions 2 and 3). This position seems to be a broader crosslinking hotspot in our data, since all TFIID subunits, as well as TFIIA and TFIIB, show a footprint at this location. This likely reflects the dense network of interactions within the engaged PIC at this region.

Notably, the strongest footprint of TBP is found at the –18 bp position. While this is consistent with previous *Drosophila* TBP ChIP-nexus data and human TBP ChIP-exo data^{33,38,39}, it is puzzling since the TATA box is found at –30 bp, and according to cryo-EM structures⁶, TBP binds a similar position at TATA-less promoters. This is unlikely due to a general lack of crosslinking at this position since TAF13 has a strong footprint there (Fig. 2C, arrow). Rather, the limited crosslinking efficiency may reflect a feature specific for how TBP contacts DNA. TBP interacts with DNA primarily through the minor groove, by inserting of conserved phenylalanine residues between TATA bases⁴⁶, while formaldehyde crosslinking depends on the proximity and accessibility of reactive nucleophilic groups, which may be poorly positioned for efficient crosslinking when a protein contacts DNA mainly through a hydrophobic minor-groove interface³⁷. This may explain the weaker TBP footprint at the TATA box.

Finally, we identified a broad upstream module that exhibits high levels of diffuse footprints upstream of the TSS (Fig. 2A, region 1). This module comprises TAF4, TAF10 and TAF12, each present in two copies and belonging to the two lobes that engage the upstream promoter region^{3,5,6}. These diffuse footprints may result from interactions with promoter-proximal upstream enhancers, as many TAFs are known to associate with DNA-bound transcription factors^{47–55}. Specifically, TAF4 has been identified as an interaction hub for transcription activators in biochemical studies^{56,57}, supporting the idea that broad upstream footprints may reflect activator-mediated stabilization of TFIID. Additionally, TAF10 and TAF12 could engage these regions as part of the SAGA complex⁵⁸.

Taken together, the ChIP-nexus profiles show strong overall agreement with the in vitro cryo-EM structures, provide insights for

less well-characterized subunits such as TAF3, and point to unexplained observations for TAF11/13 and TBP.

The TBP binding profile is promoter-type-specific

The high correlation of the TFIID subunits across promoter types argues against partial TFIID complexes at specific promoters but does not rule out differences by which TFIID subunits interact with distinct promoter types. To test this, we analyzed our ChIP-nexus data from untreated cells to capture all possible promoter states (see Fig. S6A for data from triptolide-treated cells). We identified four representative promoter types among our defined set of narrow promoters, selected based on the presence of core promoter elements at specific positions relative to the TSS, as done previously^{33,59,60} (Figs. 3A and S5A, B; see “Methods”).

TATA promoters ($n=191$) have a TATA box located –30 bp upstream of the TSS. DPR promoters ($n=334$) have a DPR element at +10–50 bp, thus an MTE, DPE or PB motif. DPR promoters typically show higher levels of Pol II pausing^{33,60,61}. TCT promoters ($n=78$) initiate at a TCT motif and are often found at ribosomal genes^{62,63}. Finally, housekeeping (HK) promoters ($n=282$) have at least one of the known housekeeping core promoter elements (DRE, Ohler 1, 6 and 7 motifs) and no TATA, TCT or DPR element. HK promoters tend to have broader initiation windows^{64–67}, but this dispersion was kept minimal by our selection of promoters with narrow transcription initiation.

Comparing the profiles of all TFIID subunits at the four promoter types revealed that TBP has the most promoter-specific profile (Fig. 3B), while most TAFs display broadly similar profiles across promoter classes, with only subtle differences. For example, the TAF binding profiles at TCT and HK promoters tend to be more diffuse with more pronounced upstream binding (Fig. S6A). Since these promoter types have their regulatory regions very closely nearby⁶⁴, this upstream binding could be the result of TFIID interacting with transcription activators. Furthermore, TRF2 occupancy was minimal at TATA promoters and highest at TCT promoters (Fig. S6A), in line with previous observations^{62,63,66,68–70}. Overall, TBP has the most intriguing promoter-specific profile, because it supports the hypothesis that TBP is a mobile component of TFIID^{19,71} and may indicate functional differences between promoter types.

The most notable feature is that the strong footprint of TBP at the upstream –18 bp position is specific for TATA promoters (Fig. 3B). It is the strongest after triptolide treatment, supporting its association with the PIC (Fig. S6A). TBP also displays a broad footprint +32 bp downstream of the TSS (Fig. 3B). This footprint has been previously observed³³ and is present at all promoter types to various degrees, but it is not observed after triptolide treatment (Figs. 2A and S6A) and has lower levels at promoters with stably paused Pol II (Fig. S9). This suggests that downstream TBP occurs during a yet unknown step of the transcription cycle. Since cryo-EM studies have suggested that TBP is in a promoter-downstream position before it gets loaded onto the upstream position^{3,5,6}, it is possible that downstream TBP represents the TFIID loading or re-loading state.

While TBP appears to have the most promoter-specific profile, this conclusion is so far based on average profiles across selected promoter sets. To test whether the differential TBP binding profile is robust across individual promoters, we asked whether we could re-classify promoters based on their individual TBP binding profile. Using the base-resolution TBP profiles at the previously defined narrow promoters, we performed k-means clustering allowing four possible clusters (see Methods).

The TBP profiles of the four clusters (Fig. 3C, D) and their enrichment of core promoter elements (Fig. 3E), suggest that the clusters represent TATA promoters (cluster 1), DPR promoters (cluster 2), and two sets of TCT/HK promoters, one with diffuse TBP ChIP-nexus footprints (cluster 3), and the other with low TBP ChIP-nexus footprints (cluster 4) (Fig. 3C, D). Notably, the TBP levels correlated

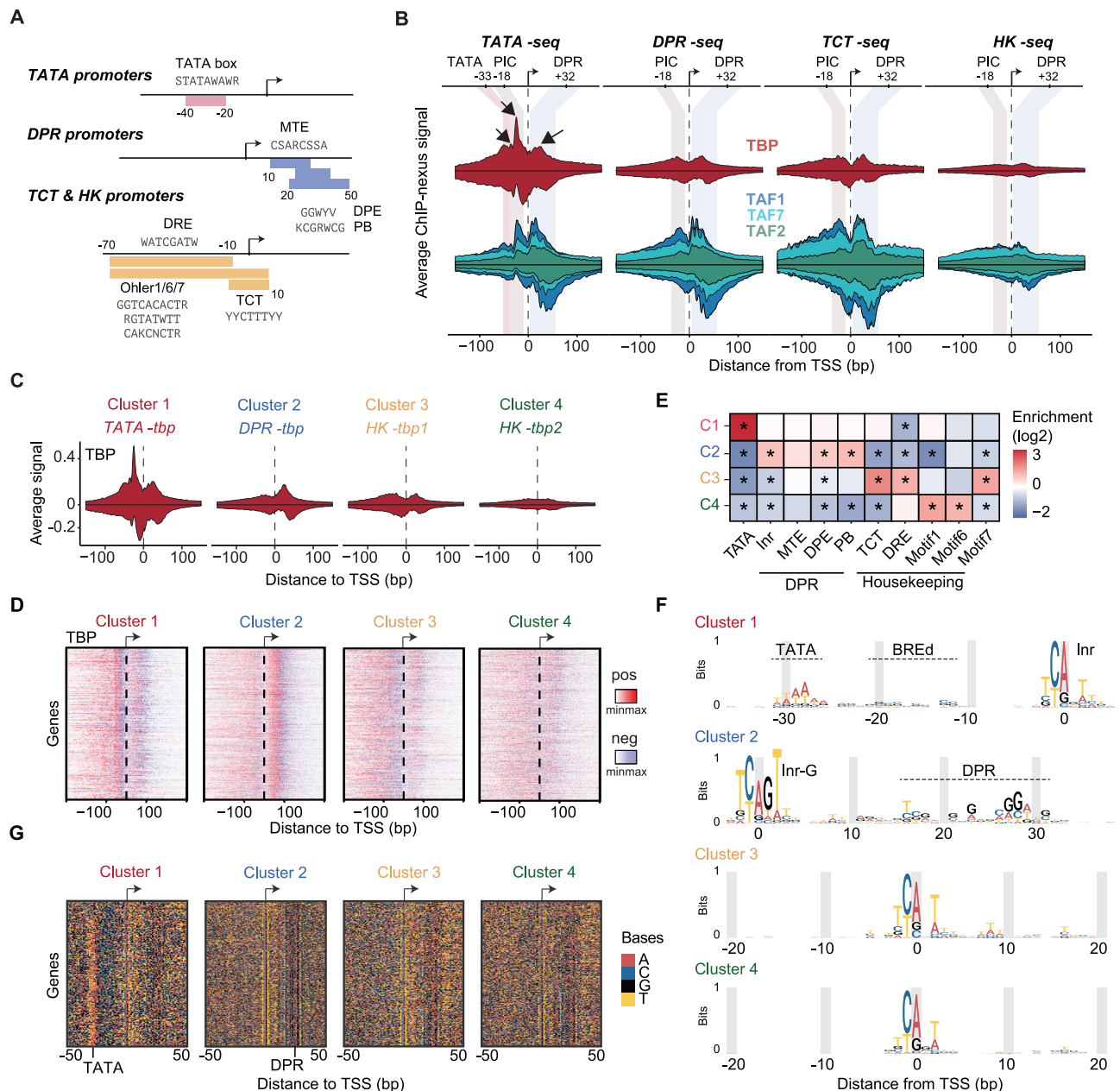


Fig. 3 | TBP is promoter-specific and allows the de novo identification of promoter types. **A** Schematic summarizing the different promoter elements used for analysis (see Methods and Supplementary Table 1 for more information). **B** The average ChIP-nexus binding profile of TBP (top; red), TAF1, TAF2 and TAF7 (second row) at TATA ($n=191$), DPR ($n=334$), TCT ($n=78$) and HK ($n=282$) promoters. Promoter classes are labeled with a “-seq” suffix to indicate sequence-based classification. **C** Average ChIP-nexus binding profiles of TBP at the four different k-means clusters (cluster 1, $n=309$; cluster 2, $n=509$; cluster 3, $n=392$; cluster 4, $n=383$) labeled with a “-tbp” suffix to indicate classification based on TBP binding patterns. **D** Distribution of TBP binding location at all promoters within each

promoter cluster. Binding to the positive and negative strand is shown in red and blue, respectively. Signal per TSS was normalized to cluster promoters based on the TBP profile shape (see “Methods”). **E** Promoter clusters are enriched for different core promoter elements. Statistical significance was calculated using a two-sided Fisher’s exact test with Benjamini–Hochberg correction, $^*p < 0.05$. Exact p values, effect sizes and 95% confidence intervals are reported in Supplementary Table 2. **F** Position weight matrix (PWM) of the identified upstream, initiator and downstream regions of the four identified promoter clusters. **G** Underlying sequence information contents across the promoter clusters colored by base.

with the mRNA levels measured by RNA-seq and CAGE-seq (Fig. SSC, D).

As further validation, we asked whether we could de novo discover the core promoter elements from each of the promoter sets clustered by TBP profile (Fig. 3F, G). We aligned the sequences at the TSS and visualized the bases coded by color as heatmap. Due to their strict positioning relative to the TSS, certain core promoter elements were discernible by eye (Fig. 3G). Clusters 3 and 4 (TCT/HK) showed clear Inr elements, while clusters 1 and 2 showed additional strong

sequence patterns indicative of TATA and DPR promoters, respectively (Fig. 3G). Consistently, the resulting motif logos closely matched the known core promoter elements^{72–74} (Fig. 3F). Notably, the TATA and DPR promoters both have Inr elements, but only those from DPR promoters were enriched for the G at position +2 (Inr-G) as previously reported⁶⁰. Furthermore, a CG-rich sequence next to the TATA motif resembles the TFIIB response element (BRE)⁷⁵, but we cannot exclude that this sequence simply enhances the crosslinking of TFIIB at this position. Taken together, we conclude that TBP displays binding

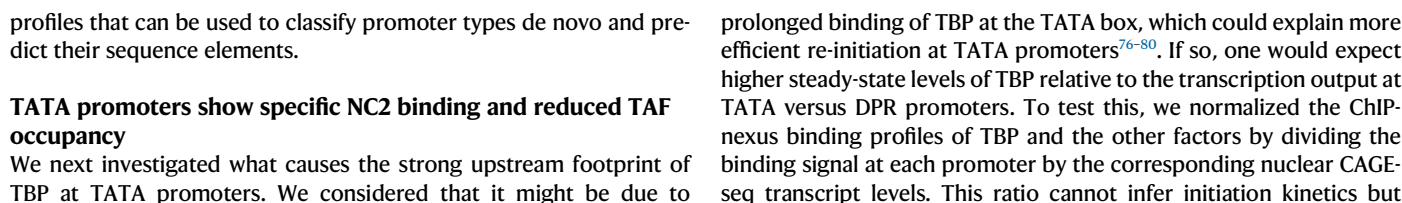


Fig. 4 | TATA promoters show reduced occupancy of TAFs. A Average transcription-normalized binding profile of TBP, TFIIA and TFIIB (top row; red, yellow and green), NC2 α , NC2 β and Mot1 (middle row; purple, blue and yellow-green), and TRF2, TAF1 and TAF7 (bottom row; pink, blue and turquoise) at TATA ($n = 191$) and DPR ($n = 334$) promoters. Transcription normalization was calculated by dividing the bp signal at each promoter by its corresponding CAGE-seq score. **B** Transcription-normalized signal distribution for the transcription factors shown in (A) at the same TATA and DPR promoters, together with TCT ($n = 78$ and HK ($n = 282$ promoters). Normalized occupancy was calculated by dividing the total ChIP-nexus signal (measured using factor-specific windows around the TSS, see Methods) at each promoter by the corresponding CAGE-seq levels. Statistical significance was performed using two-sided Wilcoxon rank-sum tests with Benjamini–Hochberg correction: ns (no significant), $p \geq 0.05$; **, $p < 0.01$; ****, $p < 0.0001$. Exact p values, effect sizes and 95% confidence intervals are reported in Supplementary Table 3. **C** Occupancy distribution of each factor relative to TBP at

TATA and DPR promoters. Relative ChIP-nexus occupancy was calculated as the log₂-transformed ratio between the total signal of each factor and the total TBP signal, measured within a 100-bp window centered on the TSS. For each factor, relative occupancy ratios were median-centered by subtracting the median log₂(*factor*/TBP) value across promoters. For panels B and C, the central line indicates the median, the box spans the 25th–75th percentiles, and whiskers extend to the most extreme values within $1.5 \times$ the interquartile range. Notches indicate the approximate 95% confidence interval around the median. Values beyond the whiskers are not shown but were included in the statistical analyses. **D** Pairwise Pearson correlation coefficients for the promoter-level occupancy of TFIID, TBP, PIC, NC2, and Mot1 at TATA and DPR promoters. **E** Difference in Pearson correlation coefficients between control (DMSO) and triptolide (TRI) when correlating upstream or downstream TBP with TAFs. Results are shown for TATA and DPR promoters, and differences were calculated as $\Delta\text{PCC} = \text{PCC}(\text{DMSO}) - \text{PCC}(\text{TRI})$.

provides an approximate measure of promoter occupancy relative to transcription initiation.

Once normalized, the levels of TFIIA, TFIIB and TFIIF were not statistically different between TATA and DPR promoters (Figs. 4A, B and S7), suggesting that PIC factors occupancy is roughly proportional to the amount of transcripts initiated at these promoters. An exception was TCT promoters, which had lower levels of TFIIA and higher levels of TFIIB, indicating functional differences (Fig. 4B). For TBP, we analyzed the binding levels upstream of the TSS because the downstream signal was not associated with the PIC. This revealed that TBP binding was significantly higher at TATA promoters than at DPR promoters (Fig. 4B). Thus, whereas other PIC components largely scaled with the transcription output, TBP showed elevated binding levels at TATA promoters. These observations suggest that TBP may persist longer at the TATA box, potentially facilitating PIC reassembly during transcription reinitiation.

We next investigated whether the strong upstream TBP footprint at TATA promoters could be associated with reduced TAF occupancy. This is in agreement with classic studies showing that TATA promoters do not require TAFs for transcription initiation *in vitro*^{81–87}, as well as with *in vivo* studies showing that TATA promoters have reduced TAF occupancy and a lower dependency on TAFs for transcription^{16,38,88–93}. When we analyzed the transcript-normalized levels of TFIID, we found that all TAFs showed significantly lower levels at TATA promoters, while TBP, TFIIA, TFIIB and TFIIF were nevertheless present at similar or higher levels (Figs. 4A, B and S7). To make sure that this conclusion was not dependent on the normalization, we confirmed that the expected relative ratios of TBP and TAF binding were also present in the non-normalized data (Figs. 4C and S6B). These data together show that TATA promoters have substantially lower occupancy of TAFs and support the hypothesis that to some extent, TBP can initiate PIC formation from the TATA box without TFIID *in vivo*.

We then examined whether other TBP-associated proteins might provide clues to TBP's promoter-specific distribution. We performed ChIP-nexus experiments on Mot1 and both NC2 subunits, two factors known to regulate TBP binding on DNA and mostly have negative effects on transcription^{21,23,24,26,94}. Mot1 showed elevated binding at all locations where TBP binds at TATA and DPR promoters, especially at the downstream footprint (Figs. 4A and S8A). In contrast, both NC2 subunits, NC2 α and NC2 β , were specifically bound to TATA promoters, significantly higher than at DPR promoters even after normalization ($p < 0.001$) (Fig. 4A, B). Each NC2 subunit showed two sharp ChIP-nexus footprints at TATA promoters: one at -18 bp, coinciding with that of TBP, and another at -33 bp near the TATA box (Figs. 4A and S8A). These results match remarkably well the crystal structure of TBP-NC2-DNA ternary complex where NC2 α contacts the TATA box underneath TBP (Lys18, Lys19) and NC2 β also interacts with the downstream region (Lys63, Lys64)⁹⁵ (Fig. S8B).

Finally, we performed additional analyses to support the differences in binding patterns between TATA and DPR promoters. Because TATA promoters show strong NC2 binding and reduced TAF occupancy, we reasoned that with comparable PIC levels and transcription output, promoters with high NC2 should still be associated with higher TBP and lower TAF binding than promoters with low NC2, and this was indeed the case (Fig. S8C–E). Moreover, to obtain additional evidence not based on binding differences between promoter types, we analyzed each promoter type separately and performed a pairwise correlation analysis between the binding levels of TBP, TAFs, TFIIA, TFIIB, TFIIF, NC2 and Mot1 (Fig. 4D). At TATA promoters, TBP shows stronger correlations with TFIIA, TFIIB, and TFIIF, NC2 and Mot1 than with the TAFs, implying that TBP does not always co-occur with the TAFs (Fig. 4D, left). This pattern was not observed at DPR promoters, where the binding levels of all factors were more homogenous, consistent with TFIID being an integral part mediating PIC formation at DPR promoters (Fig. 4D, right). Taken together, these results support a model in which TATA promoters do not initiate all rounds of transcription through TFIID. Instead, relatively stable TBP binding at the TATA box may enable PIC formation independently of TFIID, a mechanism that may be antagonized by NC2.

Interestingly, this conclusion also sheds light on the binding of TBP downstream. TATA promoters not only have lower levels of TAFs but also lower levels of downstream TBP compared to its upstream binding. In contrast, DPR promoters whose downstream sequences favor TFIID binding have higher levels of TBP downstream than upstream. This suggests that downstream TBP is associated with TFIID and thus may be part of the TFIID loading state. If so, we wondered whether we could identify TAFs that correlate more highly with the TFIID loading state and thus might be physically close to TBP when downstream. For this, we took advantage of the fact that TBP is found downstream under steady-state control conditions, but not after triptolide treatment. We asked which TAFs have the highest increase in correlation with downstream TBP under control *versus* triptolide conditions. We found TAF2 to have the greatest increase in correlation, followed by TAF1 and TAF7, and was overall higher at DPR promoters versus TATA promoters (Fig. 4D), consistent with the downstream module mediating the loading of TBP. We would have expected TAF1 to have the highest signal since it can contact TBP directly through its TAND domain, but it is possible that the DNA-facing surface of TAF2 is more easily cross-linked to DNA than TAF1.

These findings propose a model that explains the strong upstream TBP footprint at TATA promoters and the higher downstream footprint at DPR promoters. At both promoter types, TFIID initiates transcription by relocating TBP from the downstream to the upstream promoter position (Fig. 5). At TATA-containing promoters, however, TBP could either directly bind to the TATA box and mediate PIC formation or promote TAF-independent transcription re-initiation after being initially loaded by TFIID.

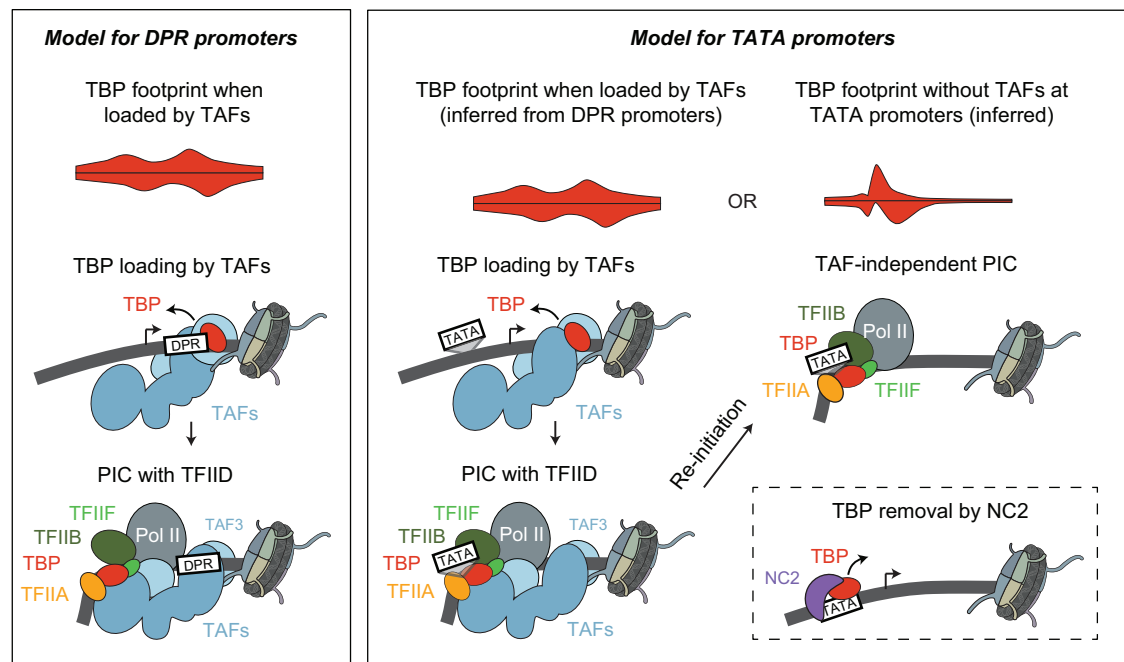


Fig. 5 | Model for transcription initiation at TATA and DPR promoters. The model summarizes the promoter-dependent modes of TBP recruitment and PIC assembly inferred from the data. At DPR promoters (left), TBP is recruited as part of TFIID through interactions of TAFs with downstream core-promoter sequences. This mode of recruitment produces the characteristic downstream TBP footprint shown at the top. Following TFIID engagement, TBP is positioned upstream, where TFIIA, TFIIB, TFIIF and Pol II assemble to form the PIC. At TATA promoters (right), two alternative modes are proposed. In the first, TBP is loaded through TFIID, analogous to the mechanism inferred at DPR promoters. Following loading by TFIID, TBP may remain bound longer at the TATA box and support successive rounds of re-initiation through a second mode that functions independently of TAFs. In addition, we cannot exclude the possibility that TBP occasionally binds directly to the TATA box and mediates initiation without TAFs. The presence of TAF-independent PICs at TATA promoters may explain the sharply localized TBP

footprints upstream of the TSS and the reduced TAF occupancy observed at these promoters. The dashed box illustrates a potential mechanism for terminating or resetting this TAF-independent promoter state. NC2 associates strongly with promoter-bound TBP at TATA promoters and may antagonize prolonged TBP binding and TAF-independent PIC formation. The TBP footprints shown above the proposed pathways represent the characteristic patterns observed experimentally and illustrate the different modes of TBP engagement inferred at DPR and TATA promoters. Together, these promoter-specific pathways provide a model in which DPR promoters predominantly use TFIID-mediated TBP loading, whereas TATA promoters can additionally exploit persistent or directly recruited TBP to support repeated PIC assembly. Such alternative modes of initiation could facilitate more efficient transcriptional re-initiation and thereby contribute to the larger transcription burst sizes characteristic of TATA promoters.

Discussion

While cryo-EM studies have unraveled the structure of TFIID bound to specific promoters in yeast and humans, it has remained speculative how TFIID's structure and function might differ between promoter types. By mapping all 14 TFIID subunits at base-pair resolution, we provide a comprehensive view of how TFIID binds across different promoter types in vivo. Our data are remarkably consistent with cryo-EM structures and suggest that the holo-TFIID complex is found at active promoters, in line with recent functional and structural investigations^{6,96–98}. However, we found that TBP displays significant promoter-specific differences, allowing us to connect structural studies with previous observations from biochemistry and genetics and construct a model of the function of TFIID at different promoter types.

Our in vivo ChIP-nexus profiles support a distinct alternative mechanism of initiation at TATA promoters. Through various analyses, we conclude that TATA promoters have elevated TBP binding but significantly lower TAF occupancy relative to PIC levels and transcription output. These findings are consistent with TATA promoters producing a substantial fraction of transcription in a TAF-independent manner⁸⁸. Whether TFIID is universally required for transcription initiation and the extent to which this requirement depends on promoter type has been debated for long^{38,82,88–93,97,99}. Early biochemical studies demonstrated that basal transcription from TATA promoters can be supported in vitro by recombinant TBP together with the general transcription factors in the absence of TAFs^{81–87}. Consistent with these observations, subsequent genetic experiments in yeast and

Drosophila showed that TAF-dependent and TAF-independent forms of transcriptionally active TBP can exist on promoters^{16,90–93}. However, later work challenged a strict division between TAF-dependent and TAF-independent promoters and suggested that TFIID contributes broadly to transcription^{25,82,97,99}. In particular, genome-wide perturbation experiments in yeast showed that rapid depletion of TAFs widely reduces transcription from both TATA and TATA-less genes, though a subset of TATA-rich promoters were less dependent on TAF^{82,97,99}. Conversely, parallel studies in yeast and mammals provided strong evidence that transcription of TATA genes persists upon depletion of certain TAF subunits^{43,88}, and that TBP and the PIC are bound at active TATA promoters even when TAFs show minimal occupancy^{38,89}. Because these studies relied on individual or limited subsets of TAFs as proxies for the entire TFIID complex, our high-resolution in vivo analysis of the complete set of TAFs extends and helps reconcile previous observations, supporting a model in which TFIID contributes to the transcription of most genes, while PIC formation at TATA promoters can occur to some extent in the absence of TAFs.

Our high-resolution profiles also provide more conclusive information on the promoter-specific binding of NC2 and TRF2. TRF2's binding was particularly low at TATA promoters and higher at DPR, HK, and TCT promoters, in agreement with previous studies^{62,68,69,100–102}. While it is unclear how TRF2 relates to TFIID, they both seem to preferentially function at TATA-less promoter types. In contrast, NC2 showed strong, sharp ChIP-nexus footprints specifically at TATA promoters, but not at TATA-less promoters. These footprints agree

with structural data of the TATA-TBP-NC ternary complex⁹⁵ and provide indirect evidence that TBP strongly binds at the TATA box, although its strongest footprint is detected slightly more downstream at -18 bp. The specificity of NC2 for TATA genes suggests that NC2 primarily inhibits transcription at TATA genes and that effects on TATA-less promoters^{24,25,103–106} may be largely indirect. Since previous evidence suggests that NC2 prevents PIC assembly and destabilizes TBP binding^{24,95}, our data support a model in which NC2 antagonizes the prolonged binding of TBP at TATA promoters.

A more puzzling observation is that under steady-state conditions, all promoters with TFIID show TBP binding downstream of the TSS³³. This downstream binding is most strongly associated with TAF2 at DPR promoters and is relatively lower at TATA promoters, where the TAF occupancy is also lower, suggesting that downstream TBP is part of TFIID. While we cannot exclude alternative explanations, several observations point to downstream TBP being part of the DNA-bound loading or reloading state of TFIID. Downstream TBP is not observed after triptolide treatment, when the DNA-binding contacts suggest that TFIID is preferentially in a PIC state and thus would sterically hinder other conformations of TFIID. Downstream TBP is also reduced at promoters with relatively stable Pol II pausing, either due to sterical occlusion as well or because these promoters are less active. Given that cryo-EM structures suggest that the downstream position corresponds to the site where TFIID initially binds DNA to place TBP upstream of the promoter⁵, we favor the hypothesis that downstream TBP reflects the DNA-bound loading state of TFIID.

In light of these findings, our results inform models for the larger transcription burst sizes observed at TATA promoters^{11,12,14,59,107–109}. By having alternative modes of transcription initiation, TATA promoters may allow more efficient reloading of TBP during a burst. TBP may either bind the TATA box directly to promote PIC formation, or TFIID could initially load TBP and subsequently disengage, while TBP remains promoter-bound. Such prolonged binding of TBP at TATA promoters is supported by the higher TBP occupancy relative to other PIC components, as well as recent live-cell single-molecule tracking experiments in yeast showing that TBP has longer residency times (~28 s) than RNA Pol II and PIC factors, which are short-lived (~2–5 s)¹¹⁰. Thus, once TBP is loaded onto the TATA box, it could remain bound through multiple rounds of transcription, as suggested by previous studies^{76–80}. Alternatively, TBP may remain chromatin-bound but must re-bind the TATA box for each round of transcription since recent structural studies in vitro suggest that Pol II elongation leads to TBP being evicted from the TATA box¹¹¹. In either case, TBP's long residence time and ability to initiate PIC formation from the TATA box likely explain the larger burst size of TATA promoters. Altogether, our high-resolution mapping of TFIID binding and other co-factors provide a resource for refining models of promoter-specific transcription initiation.

Methods

Statistics and reproducibility

ChIP-nexus experiments were performed in at least two independent biological replicates for each factor and treatment condition (except non-treated Mot1, which has only one replicate). Replicate reproducibility was assessed using Pearson correlation of promoter-level binding signals (Fig. S1). RNA-seq experiments were performed in two biological replicates in DMSO (control) condition. Publicly available CAGE-seq data from *Drosophila melanogaster* Kc167 cell nuclear extracts were derived from two biological replicates. Publicly available H3K4me3 ChIP-seq data conducted on *Drosophila melanogaster* Kc167 cells were derived from one replicate.

The number of promoters included in each analysis is reported in the corresponding figure legend. Selection of promoter subsets for each analysis are described in the “Methods”. No statistical method was used to predetermine sample size and no data were excluded from

the analyses. Extreme values were omitted only from box-plot displays to improve visualization, but were retained for all statistical tests.

Comparisons between promoter classes were performed using two-sided Wilcoxon rank-sum tests and *p* values were adjusted for multiple comparisons using the Benjamini–Hochberg procedure. Effect sizes were reported as the rank-biserial correlation (*r*), calculated from the Mann–Whitney U statistic: $r = 2U / (n_1n_2) - 1$, where *n*₁ and *n*₂ are the sample sizes of the two groups. Two-sided 95% confidence intervals were estimated by bootstrap resampling. Fisher's two-sided exact tests with Benjamini–Hochberg correction, were used for motif-enrichment analyses. Odds ratios and 95% confidence intervals were calculated as the conditional maximum-likelihood estimate and exact conditional confidence interval. For all statistical tests, exact *p* values, effect sizes, and 95% confidence intervals are reported in the corresponding Supplementary Tables. PCCs were used to assess similarity between promoter-level binding profiles.

All statistical analyses were performed in R, and the analysis code is publicly available at the GitHub repository specified in the Code Availability statement.

Cell culture and transcription inhibitor treatment

Drosophila melanogaster 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 10 μM triptolide (TOCRIS Bioscience, 3253; dissolved in DMSO) at room temperature for 1 h. Equivalent amounts of DMSO treatment (2%) were used as the control. The identity of the cell line was not independently authenticated in this study. Kc167 is not listed in the International Cell Line Authentication Committee Register of Misidentified Cell Lines (version 14, accessed 20 July 2026).

ChIP-nexus experiments

For each ChIP-nexus experiment, 10 million Kc167 cells were fixed with 1% formaldehyde in culture media at room temperature for 10 min with rotation. Fixed cells were washed with cold PBS, incubated with Orlando and Paro's Buffer (0.25% Triton X-100, 10 mM EDTA, 0.5 mM EGTA, 10 mM Tris-HCl pH 8.0, with freshly added Protease Inhibitor) for 10 min at room temperature with rotation, and then centrifuged and re-suspended in ChIP 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, with freshly added Protease Inhibitor). Sonication was performed with a Bioruptor Pico for five cycles of 30 s on and 30 s off at HIGH settings. Chromatin extracts were then centrifuged at 16,000 × *g* for 5 min at 4 °C, and supernatants were used for ChIP.

To couple Dynabeads with antibodies, 50 μl Protein A and 50 μl Protein G Dynabeads were used for each ChIP-nexus experiment and washed twice with ChIP Buffer. After removing all the liquid, Dynabeads were resuspended in 400 μl ChIP Buffer. 10 μg antibodies were added, and tubes were incubated at 4 °C for 2 h with rotation. After the incubation, antibody-bound beads were washed twice with ChIP Buffer.

For chromatin immunoprecipitation, chromatin extracts were added to the antibody-bound beads and incubated at 4 °C overnight with rotation and then washed with Nexus washing buffer A to D (wash buffer A: 10 mM Tris, 1 mM EDTA, 0.1% Triton X-100; wash buffer B: 150 mM NaCl, 20 mM Tris-HCl, pH 8.0, 5 mM EDTA, 5.2% sucrose, 1.0% Triton X-100, 0.2% SDS; wash buffer C: 250 mM NaCl, 5 mM Tris-HCl, pH 8.0, 25 mM HEPES, 0.5% Triton X-100, 0.05% sodium deoxycholate, 0.5 mM EDTA; wash buffer D: 250 mM LiCl, 0.5% IGEPAL CA-630, 10 mM Tris-HCl, pH 8.0, 0.5% sodium deoxycholate, 10 mM EDTA). End repair and dA-tailing were performed using the NEBNext End Repair Module and the NEBNext dA-Tailing Module. ChIP-nexus adaptors with mixed fixed barcodes (CTGA, TGAC, GACT, ACTG) were ligated with Quick T4 DNA ligase and converted to blunt ends with Klenow

fragment and T4 DNA polymerase. The samples were treated with lambda exonuclease and RecJf exonuclease. After each enzymatic reaction, the chromatin was washed with the Nexus washing buffers A to D and Tris buffer (10 mM Tris, pH 7.5, 8.0, or 9.5, depending on the next enzymatic step). The chromatin was then eluted and subjected to reverse crosslinking and ethanol precipitation. Purified single-stranded DNA was then circularized with CirLigase, annealed with oligonucleotides complementary to the BamHI restriction site and linearized by BamHI digestion. The linearized single-stranded DNA was purified by ethanol precipitation and subjected to PCR amplification with NEBNext High-Fidelity 2X PCR Master Mix and ChIP-nexus primers. The ChIP-nexus libraries were then gel-purified before sequencing with Illumina NextSeq 500.

ChIP-nexus data processing

ChIP-nexus data generated in this work and reprocessed from GSE85741³³ were first preprocessed by trimming and reassigning fixed and random barcode sequence information, stored as read names for use in downstream processing. Reads were next trimmed to remove adapter sequences using Cutadapt (v.4.2)¹¹² and aligned to the dm6 genome assembly using bowtie2 (v.2.3.5.1)¹¹³. Using aligned coordinates alongside the stored fixed and random barcode sequence information, reads were deduplicated. Coverage was generated by collecting only the first “stop” base of each aligned read. In order to compare reproducible, base-resolution footprinting profiles, each replicate ChIP-nexus experiment was first normalized via reads-per-million (RPM) scaling of coverage to the total aligned reads of a replicate experiment. Next, normalized replicate footprints were pooled and averaged by number of replicates per TF and condition.

CAGE-seq data processing and *Drosophila* TSS annotation

Publicly available CAGE-seq data from *Drosophila* melanogaster Kc167 cells nuclear extracts (modENCODE: ENCSR827COJ) were processed using the STAR aligner (v.2.7.1a)¹¹⁴ against the *Drosophila* melanogaster UCSC dm6 reference genome. The resulting coordinate-sorted BAM files were analyzed using the CAGER (v.2.14.0) package in R to obtain TSSs (Transcription Start Sites)¹¹⁵. Quality control was assessed using Pearson correlations and biological replicates were merged and normalized using a power-law approach ($\alpha = 1.19$, $T = 10^6$) within the fit range of 3–40,000 counts.

Individual replicate TSSs were grouped into tag clusters within each sample using distance-based clustering (distclu, CAGER package) with a maximum merging distance of 30 bp. To delineate precise coordinates for each tag cluster, the cumulative TSS signal was calculated and the 10th and 90th percentiles of the cumulative distribution were determined to delineate the central coordinates which contain 80% of transcription signal.

Tag clusters with a TPM ≥ 0.5 in at least one sample were then merged across replicates (aggregateTagClusters, CAGER package) with a maximum inter-cluster distance of 100 bp to define consensus TSS clusters. The genomic coordinates of each consensus cluster were defined as the union of the constituent clusters' central activity regions. Finally, cumulative TSS distributions and quantile positions were recomputed for the consensus clusters to refine their boundaries. The dominant TSS of each consensus cluster was identified as the genomic position with the highest summed TSS signal across the two replicate samples. For promoters that exhibited a single TSS, only those with ≥ 3 TPM were kept.

TSSs from FlyBase (r6.64, FB2025_03) were used to annotate the TSS clusters. If multiple clusters mapped to the same gene or transcript, only the cluster with the highest TPM was selected. In total, we identified 6314 active promoters. Active focused promoters where transcription initiation occurred within a narrow window ($n = 1769$) used for subsequent analysis were selected based on their inter-quantile cluster width (< 11 bp).

RNA-seq library preparation

Total mRNA was extracted for *Drosophila* melanogaster Kc167 cells using Invitrogen TRIzol reagent. cDNA libraries were prepared using Illumina TruSeq Stranded mRNA Preparation Kit with polyA enrichment and sequenced on Illumina NextSeq 500.

RNA-seq data processing

All RNA-seq pair-ended sequencing data were aligned to the UCSC dm6 reference genome using STAR (v.2.7.1a)¹¹⁴ and raw read counts per gene were quantified with “--quantMode GeneCounts” option. RNA-seq levels are TPM-normalized across replicates, measured using the UCSC dm6 transcriptome annotations¹¹⁶.

H3K4me3 ChIP-seq data processing

Publicly available H3K4me3 ChIP-seq data conducted on *Drosophila* melanogaster Kc167 cells was downloaded from GEO accession GSE62904 and processed using the bowtie¹¹⁷ against the *Drosophila* melanogaster UCSC dm6 reference genome. H3K4me3 ChIP-seq reads were compared against the corresponding input control using the SES (Signal Extraction Scaling)¹¹⁸ normalization method available in deepTools¹¹⁹ (v.3.4.2) to account for differences in sequencing depth and background signal. Input-normalized signal was calculated by subtracting the scaled input coverage from the H3K4me3 coverage “--operation subtract” in 50-bp bins across the genome. The resulting normalized signal was exported as a bigWig file for downstream analysis.

Principal component analysis

Average ChIP-nexus binding profiles under DMSO control and triptolide treatment were quantified per base pair within a 151 bp window centered on each annotated TSS. The average signal for each ChIP-nexus binding signal (x) was normalized via $x - \min(x) / \max(x) - \min(x)$ and subjected to principal component analysis (PCA) using the function prcomp (center = TRUE, scale = FALSE) from the base stats package in R. Modules were defined based on PCA clustering and similarity in the average binding profiles.

Promoter type characterization

Promoter types were selected based on the presence or absence of the motifs listed in Supplementary Table 1. Each genomic sequence had to match the consensus sequence, with up to one mismatch for the TATA box and zero mismatches for the other elements. This sequence matching occurred at each promoter within the coordinate boundaries reported in Supplementary Table 1. TATA and TCT promoters were only required to contain the TATA box or the TCT, respectively, and were not excluded for having other elements. DPR promoters were allowed the presence of MTE, DPE and PB, but not TATA, TCT, Ohler1/6/7, or DRE. Housekeeping promoters were filtered to not have motif elements from any other of the promoter categories. Thus, DPR promoters were allowed the presence of MTE, DPE, and PB, but not TATA, TCT, Ohler1/6/7, or DRE, whereas housekeeping promoters were allowed to contain Ohler1/6/7 or DRE, but not TATA, MTE, DPE, PB or TCT.

TBP profile clustering

To cluster promoters based on the TBP ChIP-nexus profile we selected the top 90% narrow promoters given the total TBP ChIP-nexus signal across an 81 bp window centered at the TSS. To emphasize binding pattern shape, signals for each promoter and position relative to the TSS were ranked from the smallest to the largest values and divided by the number of columns, so the transformed values went from 0 to 1. We then performed K-means clustering (centers = 4, nstart = 100 and $x = 30$) using the base stats package in R.

Promoter element enrichment

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 1). For each k-means cluster and each promoter element, the enrichment was calculated by determining the ratio between the fraction of promoter elements in the cluster and the fraction of the same promoter element in the other four clusters. The significance for the observed frequencies was calculated using two-sided Fisher's exact test applied to 2×2 contingency tables comparing the numbers of promoters with and without each motif in the cluster of interest versus all other clusters and corrected for multiple testing with the Benjamini–Hochberg method. Odds ratios and two-sided 95% confidence intervals were calculated as the conditional maximum-likelihood estimate and exact conditional confidence interval, respectively, using `stats::fisher.test()` in R.

TAF/TBP occupancy ratios at TATA and DPR promoters

Relative ChIP-nexus occupancy was calculated as the $\log_2$ -transformed ratio between the total signal of each factor and the total TBP signal, measured within a 100-bp window centered on the TSS. For each factor, relative occupancy ratios were median-centered by subtracting the median $\log_2(\text{factor}/\text{TBP})$ value across promoters. Extreme outliers were excluded from the plot for visualization purposes.

Normalization of the ChIP-nexus binding by CAGE-seq expression

To visualize average binding effects around active TSSs, the ChIP-nexus signal was scaled relative to the measured CAGE-seq expression score across the corresponding promoter (−150 bp to +150 bp). Next, ChIP-nexus signal for each factor at each promoter was quantified using windows defined by the predominant binding footprint of each factor relative to the TSS and scaled to the promoter's CAGE-seq expression levels in order to compare signals between promoter types. TFIIA, TFIIB, TBP, TAF10, TAF4, TAF12, TAF9, TAF5 and TAF13 (upstream broad and narrow modules) were measured within 50 bp upstream of the TSS; TAF6, TAF8 and TAF11 (middle module) were measured within 100 bp centered on the TSS; and TAF1, TAF2, TAF7 and TAF3 (downstream module) were measured within 50 bp downstream of the TSS. Statistical significance was performed using two-sided Wilcoxon rank-sum tests with Benjamini–Hochberg correction. Extreme outliers were excluded from the plot for visualization purposes. Effect sizes for two-sided Wilcoxon rank-sum tests were reported as the rank-biserial correlation (r), calculated from the Mann–Whitney U statistic: $r = 2U / (n_1n_2) - 1$, where n_1 and n_2 are the sample sizes of the two groups. Two-sided 95% confidence intervals were estimated by bootstrap resampling.

Hierarchical clustering of TFIID and PIC binding

To compare similarity of ChIP-nexus binding of TFIID subunits and PIC components, binding levels were measured around annotated TSSs (−50 bp to +50 bp) and pairwise Pearson correlations ($\log_2 + 1$). Correlations were next transformed to distance values ($1 - r$) and hierarchically clustered using the Ward's D linkage method. Hierarchical clustering of binding correlations were performed across all narrow promoters (Fig. 1D), as well as for each promoter type separately (Fig. 4C).

Paused Pol II half-life measurement

Paused Pol II half-life was re-measured at CTSS promoters with narrow transcription initiation windows from Kc167 cells³³ with an updated genome (UCSC dm6). Briefly, to analyze the half-life of paused Pol II from these data, promoters were selected if (1) the total Pol II signal was high (average ChIP-nexus signal >20), (2) a typical footprint was observed for Pol II (distance between positive- and negative-strand

maximum <80 bp) and (3) the position of the Pol II footprint was less than 80 bp downstream of the TSS. Upon re-processing, 1307 promoters fulfilled these criteria. For each promoter, the Pol II ChIP-nexus signal was calculated in a 51 bp window centered on the pausing position (the midpoint between Pol II-positive and Pol II-negative maximum). To calculate the half-life of paused Pol II at each promoter, previously generated Pol II time-course measurements³³ were fitted into an exponential decay model using nonlinear regression. Promoters with a paused Pol II half-life of longer than 60 min were floored to 60 min to eliminate inflated values due to noise. The 1307 promoters were then ranked based on half-life and divided into five quintiles. Promoters within quintile 1 ($n = 262$; Pol II pausing half-lives <5 min) were selected as promoters with short Pol II pausing and promoters within quintile 5 ($n = 262$; Pol II pausing half-lives >45 min) were selected as promoters with long or stable Pol II pausing.

Antibodies

Custom rabbit polyclonal antibodies against TFIID subunits were generated by GenScript for the Zeitlinger laboratory. Additional custom rabbit polyclonal antibodies against TBP, NC2 α , NC2 β , Mot1 and TRF2 were provided by the James T. Kadonaga laboratory. Antibodies generated in this study are available from the corresponding author upon request. Antibodies were not independently validated by Western blotting. Instead, their specificity was assessed from the characteristic and reproducible ChIP-nexus binding footprints observed at promoter regions. See Supplementary Table 8 for details on epitope design.

Reporting summary

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

Data availability

The raw RNA-seq and ChIP-nexus data generated in this study are available under GEO accession number [GSE314461](https://www.ncbi.nlm.nih.gov/geo/access/10000134461). Gene expression table counts and ChIP-nexus tracks are readily available under the same GEO accession. CAGE-seq data were downloaded from mod-ENCODE, accession number [ENCSR827COJ](https://www.encodeproject.org/ENCSR827COJ). Previously published TBP, Pol II and Nelf ChIP-nexus are available under GEO accession [GSE85741](https://www.ncbi.nlm.nih.gov/geo/access/10000134461) and H3K4me3 ChIP-seq is available under GEO accession [GSE62904](https://www.ncbi.nlm.nih.gov/geo/access/10000134461). All protein structures, the DNA-bound human TFIID [<https://www.rcsb.org/structure/6M2M>], yeast TFIID [<https://www.rcsb.org/structure/6HQA>], TFIIA-TBP [<https://www.rcsb.org/structure/1NVP>], TAF1-7-2 [<https://www.rcsb.org/structure/6M2M>], TAF4-12 [<https://www.rcsb.org/structure/1H30>], TAF6-9 [<https://www.rcsb.org/structure/6M2D>], TAF11-13 [<https://www.rcsb.org/structure/1BH8>] and NC2-TBP-DNA [<https://www.rcsb.org/structure/1JFI>], were downloaded from the RCSB Protein Data Bank and colored appropriately using UCSF CHIMERA¹²⁰. The Source data used to generate all figures is available at [<https://doi.org/10.5281/zenodo.21753147>].

Code availability

The analysis code is available on GitHub at [https://github.com/zeitlingerlab/G-M_Alcantara_TFIID_2025].

References

- Bhuiyan, T. & Timmers, H. T. M. Promoter recognition: putting TFIID on the spot. *Trends Cell Biol.* **29**, 752–763 (2019).
- Cianfrocco, M. A. et al. Human TFIID binds to core promoter DNA in a reorganized structural state. *Cell* **152**, 120–131 (2013).
- Louder, R. K. et al. Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature* **531**, 604–609 (2016).
- Kolesnikova, O. et al. Molecular structure of promoter-bound yeast TFIID. *Nat. Commun.* **9**, 4666 (2018).

5. Patel, A. B. et al. Structure of human TFIID and mechanism of TBP loading onto promoter DNA. *Science* **362**, (2018).
6. Chen, X. et al. Structural insights into preinitiation complex assembly on core promoters. *Science* **372**, (2021).
7. Nogales, E., Patel, A. B. & Louder, R. K. Towards a mechanistic understanding of core promoter recognition from cryo-EM studies of human TFIID. *Curr. Opin. Struct. Biol.* **47**, 60–66 (2017).
8. Gupta, K. et al. Architecture of TAF11/TAF13/TBP complex suggests novel regulatory state in General Transcription Factor TFIID function. Preprint at <https://doi.org/10.1101/168716> (2017).
9. Patel, A. B., Greber, B. J. & Nogales, E. Recent insights into the structure of TFIID, its assembly, and its binding to core promoter. *Curr. Opin. Struct. Biol.* **61**, 17–24 (2020).
10. Pimmett, V. L. et al. Quantitative imaging of transcription in living *Drosophila* embryos reveals the impact of core promoter motifs on promoter state dynamics. Preprint at <https://doi.org/10.1101/2021.01.22.427786> (2021).
11. Yokoshi, M., Kawasaki, K., Cambón, M. & Fukaya, T. Dynamic modulation of enhancer responsiveness by core promoter elements in living *Drosophila* embryos. *Nucleic Acids Res.* **50**, 92–107 (2022).
12. Larsson, A. J. M. et al. Genomic encoding of transcriptional burst kinetics. *Nature* **565**, 251–254 (2019).
13. Luo, S., Wang, Z., Zhang, Z., Zhou, T. & Zhang, J. Genome-wide inference reveals that feedback regulations constrain promoter-dependent transcriptional burst kinetics. *Nucleic Acids Res.* **51**, 68–83 (2023).
14. Hornung, G. et al. Noise-mean relationship in mutated promoters. *Genome Res.* **22**, 2409–2417 (2012).
15. Tantale, K. et al. A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting. *Nat. Commun.* **7**, 12248 (2016).
16. Wright, K. J., Marr, M. T. & Tjian, R. TAF4 nucleates a core sub-complex of TFIID and mediates activated transcription from a TATA-less promoter. *Proc. Natl. Acad. Sci. USA* **103**, 12347–12352 (2006).
17. Maston, G. A. et al. Non-canonical TAF complexes regulate active promoters in human embryonic stem cells. *eLife* **1**, e00068 (2012).
18. Bieniossek, C. et al. The architecture of human general transcription factor TFIID core complex. *Nature* **493**, 699–702 (2013).
19. Trowitzsch, S. et al. Cytoplasmic TAF2-TAF8-TAF10 complex provides evidence for nuclear holo-TFIID assembly from preformed submodules. *Nat. Commun.* **6**, 6011 (2015).
20. Antonova, S. V. et al. Chaperonin CCT checkpoint function in basal transcription factor TFIID assembly. *Nat. Struct. Mol. Biol.* **25**, 1119–1127 (2018).
21. Auble, D. T. et al. Mot1, a global repressor of RNA polymerase II transcription, inhibits TBP binding to DNA by an ATP-dependent mechanism. *Genes. Dev.* **8**, 1920–1934 (1994).
22. Klejman, M. P. et al. NC2 α interacts with BTAF1 and stimulates its ATP-dependent association with TATA-binding protein. *Mol. Cell Biol.* **24**, 10072–10082 (2004).
23. Schluesche, P., Stelzer, G., Piaia, E., Lamb, D. C. & Meisterernst, M. NC2 mobilizes TBP on core promoter TATA boxes. *Nat. Struct. Mol. Biol.* **14**, 1196–1201 (2007).
24. van Werven, F. J. et al. Cooperative action of NC2 and Mot1p to regulate TATA-binding protein function across the genome. *Genes. Dev.* **22**, 2359–2369 (2008).
25. Hsu, J.-Y. et al. TBP, Mot1, and NC2 establish a regulatory circuit that controls DPE-dependent versus TATA-dependent transcription. *Genes. Dev.* **22**, 2353–2358 (2008).
26. Zentner, G. E. & Henikoff, S. Mot1 redistributes TBP from TATA-containing to TATA-less promoters. *Mol. Cell Biol.* **33**, 4996–5004 (2013).
27. Viswanathan, R., True, J. D. & Auble, D. T. Molecular mechanism of Mot1, a TATA-binding protein (TBP)-DNA dissociating enzyme. *J. Biol. Chem.* **291**, 15714–15726 (2016).
28. Vo Ngoc, L., Kassavetis, G. A. & Kadonaga, J. T. The RNA polymerase II core promoter in *Drosophila*. *Genetics* **212**, 13–24 (2019).
29. He, Q., Johnston, J. & Zeitlinger, J. ChIP-nexus enables improved detection of in vivo transcription factor binding footprints. *Nat. Biotechnol.* **33**, 395–401 (2015).
30. Titov, D. V. et al. XPB, a subunit of TFIH, is a target of the natural product triptolide. *Nat. Chem. Biol.* **7**, 182–188 (2011).
31. Bensaude, O. Inhibiting eukaryotic transcription: Which compound to choose? How to evaluate its activity? *Transcription* **2**, 103–108 (2011).
32. He, Q.-L. et al. Covalent modification of a cysteine residue in the XPB subunit of the general transcription factor TFIH through single epoxide cleavage of the transcription inhibitor triptolide. *Angew. Chem. Int. Ed.* **54**, 1859–1863 (2015).
33. Shao, W. & Zeitlinger, J. Paused RNA polymerase II inhibits new transcriptional initiation. *Nat. Genet.* **49**, 1045–1051 (2017).
34. Murakami, K. et al. Structure of an RNA polymerase II preinitiation complex. *Proc. Natl. Acad. Sci. USA* **112**, 13543–13548 (2015).
35. Schilbach, S. et al. Structures of transcription pre-initiation complex with TFIH and mediator. *Nature* **551**, 204–209 (2017).
36. Solomon, M. J. & Varshavsky, A. Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures. *Proc. Natl. Acad. Sci. USA* **82**, 6470–6474 (1985).
37. Hoffman, E. A., Frey, B. L., Smith, L. M. & Auble, D. T. Formaldehyde crosslinking: a tool for the study of chromatin complexes. *J. Biol. Chem.* **290**, 26404–26411 (2015).
38. Rhee, H. S. & Pugh, B. F. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483**, 295–301 (2012).
39. Pugh, B. F. & Venters, B. J. Genomic organization of human transcription initiation complexes. *PLoS ONE* **11**, e0149339 (2016).
40. Rossi, M. J. et al. A high-resolution protein architecture of the budding yeast genome. *Nature* **592**, 309–314 (2021).
41. Yamada, N., Lai, W. K. M., Farrell, N., Pugh, B. F. & Mahony, S. Characterizing protein-DNA binding event subtypes in ChIP-exo data. *Bioinformatics* **35**, 903–913 (2019).
42. Burke, T. W. & Kadonaga, J. T. The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAFII60 of *Drosophila*. *Genes. Dev.* **11**, 3020–3031 (1997).
43. Vermeulen, M. et al. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131**, 58–69 (2007).
44. Lauberth, S. M. et al. H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell* **152**, 1021–1036 (2013).
45. Gupta, K. et al. Architecture of TAF11/TAF13/TBP complex suggests novel regulation properties of general transcription factor TFIID. *eLife* **6**, (2017).
46. Kim, J. L. & Burley, S. K. 1.9 Å resolution refined structure of TBP recognizing the minor groove of TATAAAAG. *Nat. Struct. Biol.* **1**, 638–653 (1994).
47. Amrolia, P. J. et al. The activation domain of the enhancer binding protein p45NF-E2 interacts with TAFII130 and mediates long-range activation of the α - and β -globin gene loci in an erythroid cell line. *Proc. Natl. Acad. Sci. USA* **94**, 10051–10056 (1997).
48. Kim, L. J., Seto, A. G., Nguyen, T. N. & Goodrich, J. A. Human Taf(II) 130 is a coactivator for NFATp. *Mol. Cell Biol.* **21**, 3503–3513 (2001).
49. Hamard, P.-J. et al. A functional interaction between ATF7 and TAF12 that is modulated by TAF4. *Oncogene* **24**, 3472–3483 (2005).

50. Garbett, K. A., Tripathi, M. K., Cencki, B., Layer, J. H. & Weil, P. A. Yeast TFIID serves as a coactivator for Rap1p by direct protein-protein interaction. *Mol. Cell Biol.* **27**, 297–311 (2007).
51. Layer, J. H., Miller, S. G. & Weil, P. A. Direct transactivator-transcription factor IID (TFIID) contacts drive yeast ribosomal protein gene transcription. *J. Biol. Chem.* **285**, 15489–15499 (2010).
52. Papadopoulos, P. et al. TAF10 interacts with the GATA1 transcription factor and controls mouse erythropoiesis. *Mol. Cell Biol.* **35**, 2103–2118 (2015).
53. Hibino, E. et al. Interaction between intrinsically disordered regions in transcription factors Sp1 and TAF4. *Protein Sci.* **25**, 2006–2017 (2016).
54. Hintze, S., Engelhardt, M., van Diepen, L., Witt, E. & Schüller, H.-J. Multiple Taf subunits of TFIID interact with Ino2 activation domains and contribute to expression of genes required for yeast phospholipid biosynthesis. *Mol. Microbiol.* **106**, 876–890 (2017).
55. Hibino, E. et al. Identification of heteromolecular binding sites in transcription factors Sp1 and TAF4 using high-resolution nuclear magnetic resonance spectroscopy. *Protein Sci.* **26**, 2280–2290 (2017).
56. Chen, W.-Y. et al. A TAF4 coactivator function for E proteins that involves enhanced TFIID binding. *Genes Dev.* **27**, 1596–1609 (2013).
57. Wang, X. et al. Conserved region I of human coactivator TAF4 binds to a short hydrophobic motif present in transcriptional regulators. *Proc. Natl. Acad. Sci. USA* **104**, 7839–7844 (2007).
58. Koutelou, E., Hirsch, C. L. & Dent, S. Y. R. Multiple faces of the SAGA complex. *Curr. Opin. Cell Biol.* **22**, 374–382 (2010).
59. Ramalingam, V., Natarajan, M., Johnston, J. & Zeitlinger, J. TATA and paused promoters active in differentiated tissues have distinct expression characteristics. *Mol. Syst. Biol.* **17**, e9866 (2021).
60. Shao, W., Alcantara, S. G.-M. & Zeitlinger, J. Reporter-ChIP-nexus reveals strong contribution of the Drosophila initiator sequence to RNA polymerase pausing. *eLife* **8**, (2019).
61. 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).
62. Wang, Y.-L. et al. TRF2, but not TBP, mediates the transcription of ribosomal protein genes. *Genes Dev.* **28**, 1550–1555 (2014).
63. Baumann, D. G. & Gilmour, D. S. A sequence-specific core promoter-binding transcription factor recruits TRF2 to coordinately transcribe ribosomal protein genes. *Nucleic Acids Res.* **45**, 10481–10491 (2017).
64. Zabidi, M. A. et al. Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* **518**, 556–559 (2015).
65. Haberle, V. & Stark, A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat. Rev. Mol. Cell Biol.* **19**, 621–637 (2018).
66. Serebreni, L. et al. Functionally distinct promoter classes initiate transcription via different mechanisms reflected in focused versus dispersed initiation patterns. *EMBO J.* **42**, e113519 (2023).
67. Haberle, V. et al. Transcriptional cofactors display specificity for distinct types of core promoters. *Nature* **570**, 122–126 (2019).
68. Soffers, J. H. et al. The SAGA core module is critical during Drosophila oogenesis and is broadly recruited to promoters. Preprint at <https://doi.org/10.1101/2021.06.18.448946> (2021).
69. Kedmi, A. et al. Drosophila TRF2 is a preferential core promoter regulator. *Genes Dev.* **28**, 2163–2174 (2014).
70. Hochheimer, A., Zhou, S., Zheng, S., Holmes, M. C. & Tjian, R. TRF2 associates with DREF and directs promoter-selective gene expression in Drosophila. *Nature* **420**, 439–445 (2002).
71. Greber, B. J. & Nogales, E. The structures of eukaryotic transcription pre-initiation complexes and their functional implications. *Subcell Biochem.* **93**, 143–192 (2019).
72. Vo Ngoc, L., Huang, C. Y., Cassidy, C. J., Medrano, C. & Kadonaga, J. T. Identification of the human DPR core promoter element using machine learning. *Nature* **585**, 459–463 (2020).
73. Vo Ngoc, L., Rhyne, T. E. & Kadonaga, J. T. Analysis of the Drosophila and human DPR elements reveals a distinct human variant whose specificity can be enhanced by machine learning. *Genes Dev.* **37**, 377–382 (2023).
74. Vo Ngoc, L., Wang, Y.-L., Kassavetis, G. A. & Kadonaga, J. T. The punctilious RNA polymerase II core promoter. *Genes Dev.* **31**, 1289–1301 (2017).
75. Deng, W. & Roberts, S. G. E. A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes Dev.* **19**, 2418–2423 (2005).
76. Brouwer, I., Kerklingh, E., van Leeuwen, F. & Lenstra, T. L. Dynamic epistasis analysis reveals how chromatin remodeling regulates transcriptional bursting. *Nat. Struct. Mol. Biol.* **30**, 692–702 (2023).
77. Hasegawa, Y. & Struhl, K. Promoter-specific dynamics of TATA-binding protein association with the human genome. *Genome Res.* **29**, 1939–1950 (2019).
78. Yudkovsky, N., Ranish, J. A. & Hahn, S. A transcription reinitiation intermediate that is stabilized by activator. *Nature* **408**, 225–229 (2000).
79. Krebs, A. R. et al. Genome-wide single-molecule footprinting reveals high RNA polymerase II turnover at paused promoters. *Mol. Cell* **67**, 411–422.e4 (2017).
80. Yean, D. & Gralla, J. Transcription reinitiation rate: a special role for the TATA box. *Mol. Cell Biol.* **17**, 3809–3816 (1997).
81. Tjian, R. & Maniatis, T. Transcriptional activation: a complex puzzle with few easy pieces. *Cell* **77**, 5–8 (1994).
82. Donczew, R. & Hahn, S. Mechanistic differences in transcription initiation at TATA-less and TATA-containing Promoters. *Mol. Cell Biol.* **38**, (2018).
83. Hoey, T., Dynlacht, B. D., Peterson, M. G., Pugh, B. F. & Tjian, R. Isolation and characterization of the Drosophila gene encoding the TATA box binding protein, TFIID. *Cell* **61**, 1179–1186 (1990).
84. Hoffman, A. et al. Highly conserved core domain and unique N terminus with presumptive regulatory motifs in a human TATA factor (TFIID). *Nature* **346**, 387–390 (1990).
85. Dynlacht, B. D., Hoey, T. & Tjian, R. Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. *Cell* **66**, 563–576 (1991).
86. Wu, S. Y. & Chiang, C. M. TATA-binding protein-associated factors enhance the recruitment of RNA polymerase II by transcriptional activators. *J. Biol. Chem.* **276**, 34235–34243 (2001).
87. Zhou, Q., Boyer, T. G. & Berk, A. J. Factors (TAFs) required for activated transcription interact with TATA box-binding protein conserved core domain. *Genes Dev.* **7**, 180–187 (1993).
88. Petrenko, N., Jin, Y., Dong, L., Wong, K. H. & Struhl, K. Requirements for RNA polymerase II preinitiation complex formation in vivo. *eLife* **8**, (2019).
89. Petrenko, N. & Struhl, K. Comparison of transcriptional initiation by RNA polymerase II across eukaryotic species. *eLife* **10**, e67964 (2021).
90. Moqtaderi, Z., Bai, Y., Poon, D., Weil, P. A. & Struhl, K. TBP-associated factors are not generally required for transcriptional activation in yeast. *Nature* **383**, 188–191 (1996).
91. Kuras, L., Kosa, P., Mencia, M. & Struhl, K. TAF-Containing and TAF-independent forms of transcriptionally active TBP in vivo. *Science* **288**, 1244–1248 (2000).
92. Li, X. Y., Bhaumik, S. R. & Green, M. R. Distinct classes of yeast promoters revealed by differential TAF recruitment. *Science* **288**, 1242–1244 (2000).

93. Basehoar, A. D., Zanton, S. J. & Pugh, B. F. Identification and distinct regulation of yeast TATA box-containing genes. *Cell* **116**, 699–709 (2004).
94. Dasgupta, A., Darst, R. P., Martin, K. J., Afshari, C. A. & Auble, D. T. Mot1 activates and represses transcription by direct, ATPase-dependent mechanisms. *Proc. Natl. Acad. Sci. USA* **99**, 2666–2671 (2002).
95. Kamada, K. et al. Crystal structure of negative cofactor 2 recognizing the TBP-DNA transcription complex. *Cell* **106**, 71–81 (2001).
96. Baptista, T. et al. SAGA is a general cofactor for RNA polymerase II transcription. *Mol. Cell* **68**, 130–143.e5 (2017).
97. Warfield, L. et al. Transcription of nearly all yeast RNA polymerase II-transcribed genes is dependent on transcription factor TFIID. *Mol. Cell* **68**, 118–129.e5 (2017).
98. Malik, S. & Roeder, R. G. Regulation of the RNA polymerase II preinitiation complex by its associated coactivators. *Nat. Rev. Genet.* **24**, 767–782 (2023).
99. Donczew, R., Warfield, L., Pacheco, D., Erijman, A. & Hahn, S. Two roles for the yeast transcription coactivator SAGA and a set of genes redundantly regulated by TFIID and SAGA. *eLife* **9**, (2020).
100. Bashirullah, A., Lam, G., Yin, V. P. & Thummel, C. S. dTrf2 is required for transcriptional and developmental responses to ecdysone during *Drosophila* metamorphosis. *Dev. Dyn.* **236**, 3173–3179 (2007).
101. Isogai, Y., Keles, S., Prestel, M., Hochheimer, A. & Tjian, R. Transcription of histone gene cluster by differential core-promoter factors. *Genes Dev.* **21**, 2936–2949 (2007).
102. Oyama, T. et al. Cleavage of TFIIA by Taspase1 activates TRF2-specified mammalian male germ cell programs. *Dev. Cell* **27**, 188–200 (2013).
103. Geisberg, J. V., Holstege, F. C., Young, R. A. & Struhl, K. Yeast NC2 associates with the RNA polymerase II preinitiation complex and selectively affects transcription in vivo. *Mol. Cell Biol.* **21**, 2736–2742 (2001).
104. Zhou, Z., Lin, I.-J., Darst, R. P. & Bungert, J. Maneuver at the transcription start site: Mot1p and NC2 navigate TFIID/TBP to specific core promoter elements. *Epigenetics* **4**, 1–4 (2009).
105. Willy, P. J., Kobayashi, R. & Kadonaga, J. T. A basal transcription factor that activates or represses transcription. *Science* **290**, 982–985 (2000).
106. Albert, T. K. et al. Global distribution of negative cofactor 2 subunit-alpha on human promoters. *Proc. Natl. Acad. Sci. USA* **104**, 10000–10005 (2007).
107. Hendy, O., Campbell, J., Weissman, J. D., Larson, D. R. & Singer, D. S. Differential context-specific impact of individual core promoter elements on transcriptional dynamics. *Mol. Biol. Cell* **28**, 3360–3370 (2017).
108. Ramsköld, D. et al. Single-cell new RNA sequencing reveals principles of transcription at the resolution of individual bursts. *Nat. Cell Biol.* **26**, 1725–1733 (2024).
109. Pimmett, V. L. et al. Quantitative imaging of transcription in living *Drosophila* embryos reveals the impact of core promoter motifs on promoter state dynamics. *Nat. Commun.* **12**, 4504 (2021).
110. Ling, Y. H., Liang, C., Wang, S. & Wu, C. Live-cell single-molecule dynamics of eukaryotic RNA polymerase machineries. *Science* **391**, eads0960 (2026).
111. Zhan, Y., Grabbe, F., Oberbeckmann, E., Dienemann, C. & Cramer, P. Three-step mechanism of promoter escape by RNA polymerase II. *Mol. Cell* **84**, 1699–1710.e6 (2024).
112. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j.* **17**, 10 (2011).
113. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
114. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
115. Haberle, V., Forrest, A. R. R., Hayashizaki, Y., Carninci, P. & Lenhard, B. CAGER: precise TSS data retrieval and high-resolution promoterome mining for integrative analyses. *Nucleic Acids Res.* **43**, e51 (2015).
116. Perez, G. et al. The UCSC Genome Browser database: 2025 update. *Nucleic Acids Res.* **53**, D1243–D1249 (2025).
117. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
118. Diaz, A., Park, K., Lim, D. A. & Song, J. S. Normalization, bias correction, and peak calling for ChIP-seq. *Stat. Appl. Genet. Mol. Biol.* **11**, Article 9 (2012).
119. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).
120. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

Acknowledgements

We thank Wanqing Shao for initial assistance with data generation and analysis, and James Kadonaga for providing the TBP, NC2, and Mot1 antibodies. We are also grateful to Marc Timmers, Joan and Ron Conaway, Ralf Jansen, and Ferenc Mueller for their valuable feedback on the manuscript.

Author contributions

S.G.-M.A. and J.Z. conceived the project. S.G.-M.A. performed all the experiments and data analysis. S.B. and M.W. implemented the analysis pipeline and reprocessed the data. J.Z. supervised the project. S.G.-M.A., S.B., M.W., and J.Z. prepared the manuscript.

Funding

This work was funded by the Stowers Institute for Medical Research (SIMR) and supported by NIH grant no. R01HG010211 to J.Z.

Competing interests

J.Z. owns a patent on ChIP-nexus (US-10287628). All other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-026-77758-9>.

Correspondence and requests for materials should be addressed to Melanie Weiler.

Peer review information *Nature Communications* thanks Eva Nogales, who co-reviewed with Nicholas Lue, and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.

© The Author(s) 2026