The RNA Polymerase II Core Promoter in *Drosophila*

Long Vo Ngoc, George A. Kassavetis, and James T. Kadonaga[1]
Section of Molecular Biology, University of California, San Diego, La Jolla, California 92093-0347

ABSTRACT Transcription by RNA polymerase II initiates at the core promoter, which is sometimes referred to as the "gateway to transcription." Here, we describe the properties of the RNA polymerase II core promoter in *Drosophila*. The core promoter is at a strategic position in the expression of genes, as it is the site of convergence of the signals that lead to transcriptional activation. Importantly, core promoters are diverse in terms of their structure and function. They are composed of various combinations of sequence motifs such as the TATA box, initiator (Inr), and downstream core promoter element (DPE). Different types of core promoters are transcribed via distinct mechanisms. Moreover, some transcriptional enhancers exhibit specificity for particular types of core promoters. These findings indicate that the core promoter is a central component of the transcriptional apparatus that regulates gene expression.

KEYWORDS RNA polymerase II; core promoter; core promoter elements; sequence-specific transcription factors; TBP; TBP-related factors; FlyBook

A critically important, yet often overlooked, component of the gene regulatory apparatus is the RNA polymerase II (Pol II) core promoter (for reviews, see: Smale and Kadonaga 2003; Goodrich and Tjian 2010; Kadonaga 2012; Lenhard *et al.* 2012; Danino *et al.* 2015; Roy and Singer 2015; Vo Ngoc *et al.* 2017a; Haberle and Stark 2018). The core promoter is generally considered to be the stretch of DNA, which typically ranges from ~ -40 to $+40$ nt relative to the transcription start site (TSS), that is sufficient to direct the accurate initiation of transcription. The core promoter is thus at a strategic position in the activation of a gene, as it is the site of convergence of the regulatory signals that lead to the initiation of transcription. For this reason, the core promoter is sometimes referred to as the "gateway to transcription."

In a simple transcription unit, the core promoter acts in conjunction with an enhancer and a proximal promoter region (Figure 1). The core promoter directs the initiation of transcription by Pol II, and this signal is modulated by the sequence-specific DNA-binding transcription factors (ssTFs) that bind to the enhancer and proximal promoter. A core promoter by itself exhibits little transcription activity.

The Pol II core promoter is a diverse regulatory unit. There are different types of core promoters that function by different mechanisms and have distinct biological properties. For example, some transcriptional enhancers exhibit specificity for particular types of core promoters. The basis for this phenomenon is the diversity of the DNA sequence motifs that comprise core promoters. Some of the known core promoter sequence elements in *Drosophila* are shown in Figure 2. These motifs include the TATA box, Initiator (Inr), and downstream core promoter element (DPE). It is common, for instance, to find core promoters that contain a TATA box, an Inr, and a DPE, or an Inr and a DPE (Chen *et al.* 2014). Moreover, somewhat strikingly, the core promoter arrangements at specific genes are conserved in different *Drosophila* species (Chen *et al.* 2014). There are no universal core promoter elements. In addition, some core promoters lack any of the known sequence motifs. It is thus likely that additional core promoter elements remain to be discovered. Because of the diversity of function of the core promoter, it is important to have a fundamental understanding of its composition and function for the successful design, execution, and interpretation of experiments involving the regulation of gene expression.

In this review, we will describe our current understanding of the core promoter in *Drosophila*, an organism in which many important advances in our knowledge of the Pol II core promoter have been made. At the onset, it is useful to mention that basal transcription in *Drosophila* is somewhat simple and straightforward, at least in comparison to transcription in other animals. In *Drosophila*, a substantial fraction of the core promoters are driven by the TATA box and/or DPE motifs (Figure 2) (Kutach and Kadonaga 2000; Chen *et al.* 2014). There are extremely low levels of 5-methylcytosine in *Drosophila* (Capuano *et al.* 2014). Consistent with this property, *Drosophila* essentially lack CpG methylation and CpG islands. In addition, *Drosophila* promoters are generally amenable to biochemical analysis, as there are excellent *in vitro* transcription systems (see, for example, Soeller *et al.* 1988). Importantly, there are countless genetic tools in *Drosophila* for the study of transcription in cells as well as in the entire organism. In short, it is a truly a delight to do research on the core promoter and the basal transcription process in *Drosophila*.

Brief Overview of the Basal Transcription Initiation Factors

To study the core promoter, it is important to have a basic understanding of the factors that are involved in the initiation of Pol II transcription. Pol II by itself does not recognize the core promoter. Instead, auxiliary factors, which are known as the basal (or general) transcription factors, recognize core promoter sequence motifs, recruit Pol II to the promoter, and initiate the transcription process (for review, see Sainsbury *et al.* 2015). The best understood basal transcription factors are those that mediate transcription from TATA box-dependent core promoters. These basal factors, in the approximate order in which they assemble along with Pol II onto the promoter, are as follows: TFIID (Transcription Factor for RNA polymerase II D) and TFIIA, TFIIB, Pol II and TFIIF, TFIIE, and TFIIH. The resulting assemblage of factors is termed the transcription preinitiation complex (PIC). In the presence of ribonucleoside 5′-triphosphates, transcription initiates rapidly (in seconds) from the PIC.

For the purpose of studying the core promoter, it is important to remember that TFIID is a key core promoter recognition factor. TFIID comprises the TATA box-binding protein (TBP) along with $\sim$13–14 TBP-associated factors (TAFs), some of
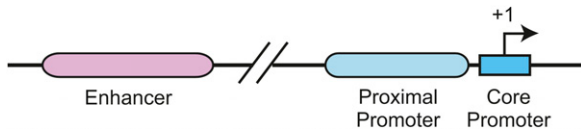
Figure 1 The core promoter is a component of the transcriptional control region of a gene. The core promoter typically comprises the region from −40 to +40 relative to the +1 transcription start site (TSS), and contains the information that directs the initiation of transcription by RNA polymerase II. The core promoter functions with the proximal promoter and enhancer regions, both of which contain sites for the binding of sequence-specific DNA-binding transcription factors. The proximal promoter region spans from ∼ −40 to −250 relative to the TSS. Many genes contain multiple enhancers that are located at variable distances both upstream and downstream of the TSS.

which have alternate forms (for reviews, see: Goodrich and Tjian 2010; Fuller 2016). TBP as well as some TAFs bind to specific core promoter motifs. In addition, TFIIA significantly affects the interaction of TFIID with the core promoter region (see, for example, Cianfrocco *et al.* 2013). The basal transcription factor TFIIB, which is a single polypeptide, also binds with some sequence specificity to a subset of promoters. Thus, particular note should be made of TFIID (including its TBP and TAF subunits), TFIIA, and TFIIB. There are also TBP-related factors (TRFs), which are discussed below.

The TSS

Identification of TSSs

The successful analysis of a promoter region is critically dependent upon the correct identification of its TSS or TSSs. The accuracy of the mapping of TSSs is continuing to improve as new techniques are being developed. At the present time, it appears that the best methods for the mapping of TSSs in cells involve the determination of the 5′ ends of capped nascent transcripts (see, for example, Nechaev *et al.* 2010; Kruesi *et al.* 2013; Kwak *et al.* 2013; Lam *et al.* 2013). The characterization of nascent transcripts ensures that minimal processing of the transcripts has occurred in cells and in the processing of the samples. In the study of any particular promoter, it is important to examine the existing TSS data carefully, and, if possible, perform additional experiments that would provide independent verification of the TSS or TSSs.

Focused vs. dispersed transcription initiation

There are different patterns of transcription initiation, which are sometimes referred to as "promoter shape." One mode of transcription initiation, termed "focused" (also known as "narrow peak," "peaked," "sharp peak," and "single peak"), occurs when transcription initiates from either a single site or a narrow (∼5 nt or less) cluster of sites. Transcription from focused promoters probably derives from a single positioned PIC. An alternate mode of transcription initiation, termed "dispersed" (also known as "broad" or "weak"), occurs when there are multiple TSSs that are spread out over a region that
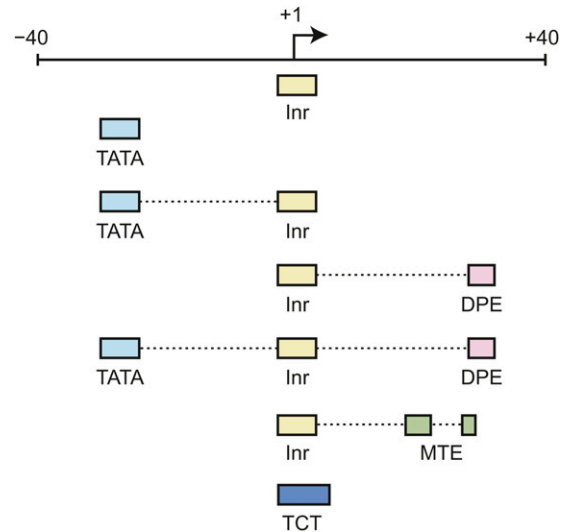
Figure 2 Schematic diagram of some RNA polymerase II core promoter motif arrangements that are observed in *Drosophila*. The specific motifs are discussed in the text. In this figure, "Inr" refers to the consensus Inr as well as to an Inr-like sequence termed DMv4, and "TATA" refers to the consensus TATA box as well as to a TATA-like sequence termed DMv5. DMv4 and DMv5 are commonly observed in *Drosophila* core promoters (FitzGerald *et al.* 2006; Chen *et al.* 2014). This drawing is roughly to scale.

might be as large as 50–100 nt. The multiple TSSs from dispersed promoters likely derive from multiple distinct PICs at different locations. In addition, many promoters have the combined features of both focused and dispersed promoters – that is, they have a distinct major TSS along with dispersed minor TSSs. Focused, dispersed, and combined transcription patterns are observed in *Drosophila* (Rach *et al.* 2009, 2011; Ni *et al.* 2010; Hoskins *et al.* 2011).

Focused promoters often contain core promoter motifs such as the TATA box, Inr, motif ten element (MTE), and DPE, and are frequently associated with regulated genes, whereas dispersed promoters are associated with Ohler core promoter motifs 1, 6, 7 (Ohler *et al.* 2002) and DNA replication-related element (DRE) sequences, and are commonly found in ubiquitously expressed genes (Rach *et al.* 2009, 2011; Ni *et al.* 2010; Hoskins *et al.* 2011). In addition, focused promoters are more evolutionarily constrained than dispersed promoters (Schor *et al.* 2017). From a teleological standpoint, it is reasonable that regulated promoters would be most facilely turned on or off at a single site and that focused transcription would be less essential at constitutive promoters at which there is sustained gene expression under different conditions.

It is also interesting to note that the comparison of promoter shape and chromatin structure showed that dispersed promoters are generally associated with periodic arrays of nucleosomes, whereas focused promoters tend to lack an organized chromatin structure (Figure 3A) (Rach *et al.* 2011). Further analysis revealed, however, that focused promoters with a TCT motif are associated with positioned arrays of nucleosomes (Figure 3B) (Rach *et al.* 2011). These

findings thus indicate that periodic arrays of nucleosomes in the promoter region are not a special characteristic of dispersed transcription, but rather appear to be associated with housekeeping genes rather than regulated genes. It is possible, for instance, that the chromatin structure in the promoter regions of regulated genes is maintained in a disrupted state that allows for more facile gene activation.

Core Promoter Sequence Motifs

In this section, we will describe core promoter sequence motifs in *Drosophila*. In this regard, it is important to mention a key study in which Ohler *et al.* (2002) identified 1941 TSS clusters in *Drosophila* and used MEME to determine the 10 most over-represented sequence motifs in the region from −60 to +40 relative to the +1 TSS. This analysis yielded the DRE (DNA replication-related element; motif 2), TATA box (motif 3), Inr (motif 4), E-box (motif 5), DPE (motif 9), and MTE (motif ten element; Lim *et al.* 2004) as well as other sequences, such as those that are currently known as Ohler motifs 1, 6, and 7. Individual motifs will be described in greater detail below.

The TATA box

The TATA box is an ancient core promoter motif that is present in Archaea as well as in eukaryotes. It was discovered in the DNA sequence analysis of the *Drosophila* histone genes (Goldberg 1979). The TATA box is recognized by the TBP subunit of the TFIID complex. In addition to conventional TATA boxes (for example, TATAAA at ∼ −30 relative to the +1 TSS), it is likely that TATA-like sequences also function as TATA box elements due at least in part to the ability of TBP to accommodate variation in the TATA sequence (see, for example, Patikoglou *et al.* 1999). In this regard, it is notable that a TATA-like sequence termed DMv5 is commonly found in *Drosophila* (FitzGerald *et al.* 2006; Chen *et al.* 2014).

The Inr

The Inr is probably the most commonly occurring core promoter motif in bilateria (bilaterally symmetric animals). It was originally described by Corden *et al.* (1980), and its functions were incisively articulated by Smale and Baltimore (1989). The Inr interacts with the TAF1 and TAF2 subunits of the TFIID complex (Chalkley and Verrijzer 1999; Louder *et al.* 2016). As a core promoter motif, the Inr does not have strong core promoter activity by itself (Smale and Baltimore 1989); rather, it functions with other core promoter elements such as the TATA box, MTE, and DPE.

The *Drosophila* Inr was first described by Hultmark *et al.* (1986) (consensus in nonheat shock genes: ATCA+1KTY), and Inr sequences that are bound by *Drosophila* TFIID (DYA+1KTG) were determined by Purnell *et al.* (1994). The current consensus for the *Drosophila* Inr is TCA+1GTY (Ohler *et al.* 2002; FitzGerald *et al.* 2006). This sequence is more distinct than the human Inr consensus of BBCA+1BW (Vo Ngoc *et al.* 2017b). In addition, an Inr-like sequence
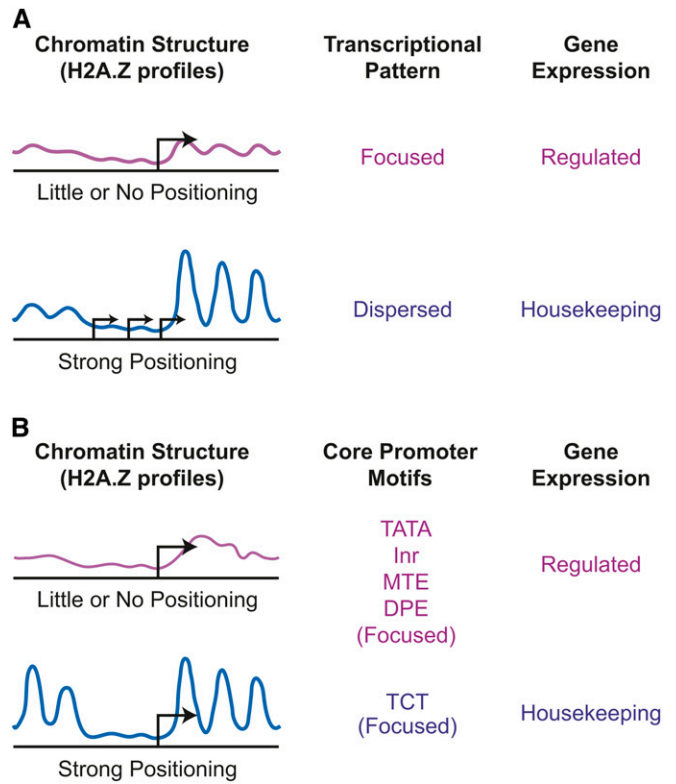
Figure 3 Analysis of chromatin structure at different types of core promoters suggests that promoters of regulated genes generally lack periodic arrays of nucleosomes. The diagrams are based on data from Rach *et al.* (2011). (A) Comparison of chromatin structure at focused *vs.* dispersed promoters. (B) Comparison of focused promoters that contain TATA, Inr, MTE, or DPE motifs *vs.* focused promoters that contain the TCT motif.

termed DMv4 is commonly found in *Drosophila* core promoters (FitzGerald *et al.* 2006; Chen *et al.* 2014).

The TCT motif (polypyrimidine initiator)

The TCT motif (polypyrimidine initiator) is a rare but biologically important core promoter motif in bilateria. The TCT motif is found in almost all ribosomal protein gene promoters as well as in the promoters of some other genes that encode factors that are involved in translation (Perry 2005; Parry *et al.* 2010). In *Drosophila*, the TCT motif is present in only ∼120 core promoters, including the ribosomal protein gene promoters (Parry *et al.* 2010); thus, it is a rare motif. The TCT consensus sequence in *Drosophila* is YYC+1TTTYY, which is similar to the TCT consensus in humans (YC+1TYTYY; Parry *et al.* 2010). Because of its scarcity, the TCT motif is frequently not identified in the sequence analysis of promoter regions. Nevertheless, it is likely to be important for the coordination of the expression of the ribosomal protein genes. Thus, in the case of the TCT motif, its rarity is a reflection of its specificity of function.

Although the TCT motif has a resemblance to the Inr, the two elements are functionally distinct. Somewhat strikingly, a single T to A nucleotide substitution (TC+1T to TCA+1) can

convert a TCT motif into an Inr (Parry *et al.* 2010). Moreover, as discussed below, transcription that is driven by the TCT motif in *Drosophila* is mediated by TBP-related factor 2 (TRF2) instead of TBP (Wang *et al.* 2014). These findings indicate that there is a specialized transcription system involving TRF2 and the TCT motif for the expression of the ribosomal protein genes in *Drosophila*. This TCT-based system complements the RNA polymerase I and RNA polymerase III transcription systems for the synthesis of the components of the ribosome.

The DPE

The DPE was discovered in the analysis of TATA-less promoters in *Drosophila* (Burke and Kadonaga 1996). The consensus for the *Drosophila* DPE is RGWYV from +28 to +32 relative to the +1 TSS (Vo Ngoc *et al.* 2017a). It is estimated that ~40% of *Drosophila* promoters contain a DPE motif (Kutach and Kadonaga 2000). A variant of the DPE is "motif 9" in Ohler *et al.* (2002). The motif 9 sequence has some of the combined features of the MTE and DPE, and is sometimes referred to as the "Ohler DPE."

The DPE and Inr elements function cooperatively for the binding of TFIID as well as for transcriptional activity (Burke and Kadonaga 1996). The mutation of either element results in a loss of TFIID binding and promoter activity. There is a precise spacing requirement between the DPE and Inr, as a single nucleotide increase or decrease in the Inr to DPE spacing can result in a several-fold decrease in transcriptional activity (Burke and Kadonaga 1997; Kutach and Kadonaga 2000). Photocrosslinking experiments with purified TFIID indicated that the DPE is in close proximity to the TAF6 and TAF9 subunits of TFIID (Burke and Kadonaga 1997); however, the cryo-EM structure of a PIC showed that the TAF1 and TAF2 subunits are in contact with the DPE (Louder *et al.* 2016). The basis for this difference is not known, but could be due to alternate conformations of promoter-bound TFIID or possibly to the presence of multiple TFIID complexes at the promoter in the photocrosslinking experiments.

The DPE is also present in humans, and is recognized by the human basal transcriptional machinery (Burke and Kadonaga 1997; Juven-Gershon *et al.* 2006), but it is rarely found in human promoters (discussed in Vo Ngoc *et al.* 2017a). The apparent scarcity of the DPE in humans may be due to our current lack of knowledge of the human DPE consensus.

One interesting feature of the DPE is that it is frequently found in the promoters of *Drosophila* long interspersed nuclear elements (LINEs; non-LTR retrotransposons), which are transcribed by Pol II. LINEs lack an LTR and must therefore have internal promoter sequences that are downstream of the +1 TSS. Thus, the DPE is well suited for the transcription of these elements, and many *Drosophila* LINEs, such as *jockey*, *Doc*, *297*, *G*, and *I*, contain DPE-dependent promoters (Burke and Kadonaga 1996; Kutach and Kadonaga 2000).

The MTE

The MTE is another core promoter element that was discovered in *Drosophila*. The MTE region was identified as an over-represented promoter sequence (motif 10) by Ohler *et al.* (2002), and the function of the MTE as a core promoter element was determined by Lim *et al.* (2004). Like the DPE, the MTE functions cooperatively with the Inr for TFIID binding and transcriptional activity (Lim *et al.* 2004). In addition, there is synergy between the MTE and the DPE (Lim *et al.* 2004; Theisen *et al.* 2010). Photocrosslinking data with purified TFIID showed that TAF6 and TAF9 are in close proximity to the MTE (Theisen *et al.* 2010). The photocrosslinking results with the MTE are similar to those seen with the DPE (Burke and Kadonaga 1997), but differ from the structural studies of TFIID in the PIC (Louder *et al.* 2016). The differences remain to be clarified.

An analysis of the downstream region of *Drosophila* core promoters suggested that the MTE through the DPE might be considered to be single functional unit with multiple contact points with TFIID (Theisen *et al.* 2010). In this study, two important DNA segments in the MTE region were found to be from +18 to +22 and from +27 to +29 relative to the +1 TSS. A comprehensive analysis might reveal a unified view of the downstream core promoter region. In this regard, it is notable that G+24 contributes to DPE-dependent transcription (Kutach and Kadonaga 2000). In addition, as mentioned above, the "Ohler DPE" has some of combined features of the MTE and DPE.

The TFIIB and TFIIA recognition elements

The TFIIB recognition elements (BREs) are sites at which the basal transcription factor TFIIB contacts the promoter DNA immediately upstream (BRE[u]) or downstream (BRE[d]) of the TATA box (Lagrange *et al.* 1998; Tsai and Sigler 2000; Deng and Roberts 2005). This interaction involves the binding of TFIIB to TATA box-bound TBP, and hence, the functions of the BREs are dependent upon the presence of a TATA box.

Like TBP, TFIIB is an ancient transcription factor that is present in Archaea and eukaryotes. In humans, the BRE[u] consensus is SSRCGCC (Lagrange *et al.* 1998), and the BRE[d] consensus is RTDKKKK (Deng and Roberts 2005). The frequency of occurrence of the BRE motifs has been estimated to be ~10–35% (Deng and Roberts 2005; Gershenzon and Ioshikhes 2005). It should be noted, however, that there have not been more recent revisions of the human BRE consensus sequences. In addition, the BRE consensus sequences in *Drosophila* have not yet been determined.

The functions of the BREs remain to be clarified, as they have been observed to increase as well as to decrease transcriptional activity in different contexts (Lagrange *et al.* 1998; Evans *et al.* 2001; Deng and Roberts 2005). The BRE[u] motif was also found to inhibit the ability of the *Drosophila* Caudal protein (an enhancer-binding factor) to activate transcription from TATA-dependent promoters (Juven-Gershon *et al.* 2008). Thus, it can be seen that further analysis of the BRE motifs is needed to determine their roles in the regulation of transcription.

In addition to the BREs, there is a TFIIA-recognition element, termed the IIARE, which was identified in the analysis of

the binding of TFIIA to promoter DNA in TFIIA-TBP-DNA complexes (Wang *et al.* 2017). The IIARE is located immediately upstream of the TATA box, and its consensus sequence is GKGVSRTKKT. In TATA-containing promoters, the IIARE enhances transcription, whereas in the absence of the TATA box, the IIARE represses transcription. The IIARE has not yet been studied in *Drosophila*.
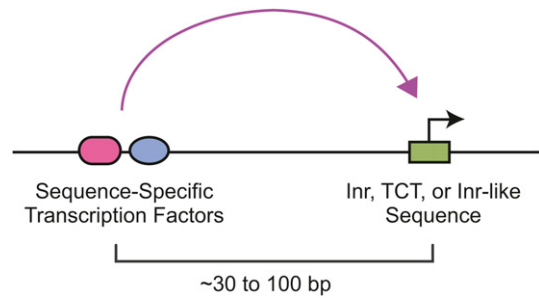
Sequence-Specific Transcription Factors and Transcription Initiation

It is important to consider the role of sequence-specific DNA-binding transcription factors (ssTFs) in the initiation of transcription, particularly at promoters that lack core promoter motifs such as the TATA box and DPE. In *Drosophila*, ssTF binding sites such as the DRE (DNA replication-related element; Ohler motif 2), E-box (Ohler motif 5), and Ohler motif 1 are frequently observed from ~ -60 to -1 relative to the +1 TSS (Ohler *et al.* 2002; FitzGerald *et al.* 2006). The DRE is bound by the DRE-binding factor (DREF; Hirose *et al.* 1993, 1996), the E-box is bound by helix-loop-helix (HLH) proteins such as Myc family transcription factors (Massari and Murre 2000), and Ohler motif 1 is bound by motif 1 binding protein (M1BP; Li and Gilmour 2013). DREF is present in *Drosophila* and humans (in which it is known as ZBED1 or hDREF) but not in yeast, HLH proteins that bind to the E-box are in yeast, *Drosophila*, and humans, and M1BP is in *Drosophila* but not in yeast or humans. Other ssTFs such as the GAGA factor (also known as GAF) and Zelda are also commonly found in the proximal promoter region (~ -250 to -40 relative to the +1 TSS) of *Drosophila* genes.

It is likely that most of the ssTFs in the proximal promoter region activate transcription, as is well established (see, for example, Kadonaga 2004). In promoters that lack core promoter motifs, or contain only an Inr or a TCT motif (which cannot direct efficient transcription by themselves; Smale and Baltimore 1989), it is possible that proximal promoter-bound ssTFs drive transcription initiation in conjunction with an Inr, a TCT motif, or an Inr-like element that might be as minimal as a CA+1 or TA+1 dinucleotide (Figure 4). It has been shown that an ssTF can activate transcription in lieu of a TATA box (Smale *et al.* 1990; Emami *et al.* 1995). It is thus reasonable to postulate that proximal promoter-bound ssTFs mediate the initiation of transcription at promoters lacking TATA box and DPE motifs. This hypothesis is relevant to the function of many promoters and should be investigated in greater depth.

TBP-Related Factors

TBP is an ancient protein that is present in Archaea and eukaryotes. At TATA-containing promoters, TBP has a critical role in the recruitment of the basal transcription machinery for assembly into the PIC. Most unicellular organisms contain a single TBP protein, but bilateria additionally have TBP-related factors (TRFs) with activities that are distinct from those of TBP.

Figure 4 Postulated model for the role of sequence-specific DNA-binding transcription factors (ssTFs) in the initiation of transcription from promoters that lack core promoter motifs such as the TATA box and DPE. This model is discussed in the text.

The first identified TRF, which was initially termed "TRF" and is now known as TRF1, was discovered in *Drosophila* (Crowley *et al.* 1993). TRF1 appears to exist only in insects. A second TRF, termed TRF2 (also known as TBPL1, TLP, TRP, and TLF), is found in bilateria (reviewed in Goodrich and Tjian 2010; Akhtar and Veenstra 2011; Duttke *et al.* 2014). A third TRF, denoted TRF3 (also known as TBPL2 and TBP2), is present in vertebrates, and is the TRF that is most closely related to TBP (Persengiev *et al.* 2003; for reviews, see: Goodrich and Tjian 2010; Akhtar and Veenstra 2011; Vo Ngoc *et al.* 2017a). In *Drosophila*, two additional TRFs, termed TRF4 and TRF5, have been identified (Kurshakova *et al.* 2018). Analysis of TRF4 revealed that it lacks a nuclear localization signal and appears to be cytoplasmic. Here, we will focus on the roles of *Drosophila* TBP, TRF1, and TRF2 in transcription (Figure 5).

TRF1

TRF1 is an insect-specific TRF that binds to the TATA box as well as to a TC-rich cluster and substitutes for TBP in the transcription of some Pol II promoters *in vitro* (Hansen *et al.* 1997; Holmes and Tjian 2000). In addition, TRF1 associates with the RNA polymerase III (Pol III) transcription factor termed BRF1 (also known as BRF), and has been found to be required for the transcription of several *Drosophila* Pol III promoters *in vitro* and to colocalize with BRF1 at Pol III genes in *Drosophila* cells and polytene chromosomes (Takada *et al.* 2000; Isogai *et al.* 2007a; Verma *et al.* 2013). Thus, TRF1 functions in transcription by Pol II and Pol III. It is interesting to note that the combined functions of TBP and TRF1 in *Drosophila* are similar to those of TBP in organisms (such as the yeast *Saccharomyces cerevisiae*) that lack TRFs (Figure 5). Hence, the emergence of TRF1 generally resulted in the subdivision or sharing of the original functions of TBP between TBP and TRF1.

TRF2

TRF2 is present in bilateria (Duttke *et al.* 2014) and is the TRF that is least closely related to TBP. Unlike TBP, TRF1, and TRF3, TRF2 does not exhibit any detectable DNA-binding activity (see, for example, Dantonel *et al.* 1999; Rabenstein

TBP	TRF1	TRF2
RNA Pol I	RNA Pol III – tRNA	RNA Pol II – TCT
RNA Pol II – TATA	RNA pol II – TC box	RNA Pol II – DPE
RNA Pol III – U6 snRNA	RNA pol II – TATA?	RNA Pol II – H1
Other Programs?	Other Programs?	Other Programs?

Functions that are mostly similar to those in organisms with only TBP (no TRFs) — *New functions with TRF2*

Figure 5 Summary of our current knowledge of transcriptional programs that are carried out by TBP, TRF1, and TRF2 in *Drosophila*. The combined functions of *Drosophila* TBP and TRF1 are approximately the same as the functions of TBP in organisms, such as *Saccharomyces cerevisiae*, that contain only TBP. In contrast, TRF2 mediates new functions that are not present in organisms that lack TRFs.

et al. 1999; Wang *et al.* 2014; Zehavi *et al.* 2015). TRF2 does, however, interact with the TFIIA and TFIIB basal transcription factors (Rabenstein *et al.* 1999; Teichmann *et al.* 1999). Thus, TRF2 is similar to TBP except that it does not bind to the TATA box.

The emergence of TRF2 may have facilitated the evolution of TATA-less transcription systems in bilateria. Consistent with this idea, *Drosophila* TRF2 is mainly associated with TATA-less promoters (Isogai *et al.* 2007b). In addition, TRF2, but not TBP, has been found to be essential for transcription from *Drosophila* TATA-less promoters that contain either a TCT or a DPE core promoter motif (Hsu *et al.* 2008; Kedmi *et al.* 2014; Wang *et al.* 2014). Moreover, TRF2, but not TBP, is required for transcription of the TATA-less (and TCT- and DPE-deficient) histone H1 genes in *Drosophila* (Isogai *et al.* 2007b). It is further notable that the TCT and DPE core promoter motifs are found only in bilateria. Importantly, the TRF2-driven TATA-less transcription systems have added new gene regulatory functions (Figure 5), and the biological functions of TRF2-regulated genes suggests that TRF2 may have had an important role in the evolution of bilateria (Duttke *et al.* 2014).

TRF2 is essential in many bilateria. For instance, the loss of TRF2 is embryonic lethal in *Drosophila* (Kopytova *et al.* 2006), *Caenorhabditis elegans* (Dantonel *et al.* 2000; Kaltenbach *et al.* 2000), zebrafish (*Danio rerio*; Müller *et al.* 2001), and *Xenopus* (Veenstra *et al.* 2000). In contrast, TRF2-deficient mice are viable but have a defect in spermiogenesis (Martianov *et al.* 2001; Zhang *et al.* 2001; Zhou *et al.* 2013). It is possible that mice have another factor that can compensate for the absence of TRF2. This question remains to be resolved.

Because TRF2 does not bind to DNA, it is probably recruited to core promoters via interactions with other factors. In *Drosophila*, the ssTFs DREF and M1BP have been found to bind to TRF2. DREF recruits TRF2 to some promoters via its interaction with DREF binding sites (*i.e.*, DREs) (Hochheimer *et al.* 2002). M1BP recruits TRF2 to a majority of the TCT motif-containing ribosomal protein (RP) gene promoters via its binding to motif 1 sequences (Baumann and Gilmour 2017). M1BP does not, however, generally recruit TRF2 to promoters, as <3% of M1BP ChIP-exo peaks correlate with TRF2 ChIP-exo peaks at non-RP gene promoters. As mentioned above, E-box sequences

are also commonly found in core promoter regions (Ohler *et al.* 2002; FitzGerald *et al.* 2006), and it is possible that some E-box-binding proteins recruit TRF2 to core promoters. It will be interesting to determine how promoter-associated TRF2 functions to mediate the initiation of transcription.

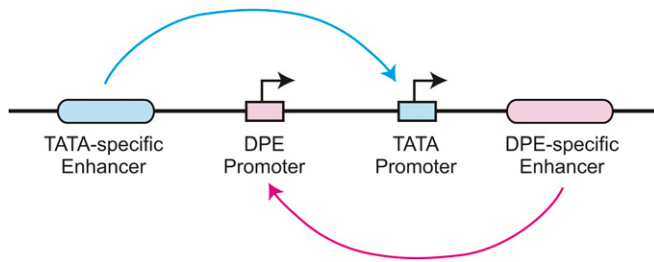## Enhancer-Core Promoter Specificity

Transcriptional enhancers are critical determinants of the spatial and temporal regulation of gene expression. In the *Drosophila* genome, it has been estimated that there are 50,000–100,000 enhancers (Kvon *et al.* 2014), and it is important to have the appropriate functional connections between each of these tens of thousands of enhancers and their cognate promoters. One means by which these connections are properly established is by enhancer-core promoter specificity (Figure 6).

The phenomenon of enhancer-core promoter specificity has been well studied in *Drosophila*. For instance, when the *Drosophila* AE-1 and IAB5 enhancers were each placed in a competitive situation in which they could activate transcription from either a TATA-dependent core promoter or a DPE-dependent core promoter, they exhibited a distinct preference for the TATA-dependent core promoter, even though they could, in the absence of competition, activate the DPE-dependent core promoter (Ohtsuki *et al.* 1998). In another set of experiments, DPE-specific as well as TATA-specific enhancers were identified on the basis of their ability to activate transcription from a TATA-dependent *vs.* a DPE-dependent core promoter (Butler and Kadonaga 2001). More recently, enhancer-core promoter specificity, such as with "developmental core promoters" (with TATA, Inr, MTE, DPE motifs) *vs.* "housekeeping core promoters" (with TCT, DRE, E-box, and Ohler motifs 1, 6, 7), was seen on a genome-wide level (Arnold *et al.* 2013; Zabidi *et al.* 2015).

It will be important to understand the molecular basis of enhancer-core promoter specificity. In this regard, it was found that the *Drosophila* Caudal protein, a sequence-specific transcription factor (ssTF) and a master developmental regulator, is a DPE-specific transcription factor that functions preferentially with DPE-dependent core promoters relative to TATA-dependent core promoters (Juven-Gershon *et al.* 2008; Shir-Shapira *et al.* 2015). In addition, the transcriptional cofactors NC2 (negative cofactor 2, also known as Dr1-Drap1) and Mot1 were observed to activate DPE-dependent transcription and to inhibit TATA-dependent transcription (Willy *et al.* 2000; Hsu *et al.* 2008). Thus, ssTFs such as Caudal and cofactors such as NC2 and Mot1 may function to mediate enhancer-core promoter specificity. We are, however, only at the early stages of understanding the factors and mechanisms that are involved in this phenomenon.

Promoter-Proximal Pausing by RNA Polymerase II

At many genes, Pol II pauses at ∼30–50 nt downstream of the TSS. This phenomenon was discovered in the analysis of

Figure 6 Schematic diagram of enhancer-core promoter specificity, in which enhancers establish functional interactions with their cognate core promoters.

the *Drosophila hsp70* promoter (Rougvie and Lis 1988). Promoter-proximal pausing involves factors such as NELF (negative elongation factor) and DSIF (DRB sensitivity-inducing factor), and is released by P-TEFb (positive transcription elongation factor b), which is a kinase that phosphorylates NELF, DSIF, and the C-terminal domain of the largest subunit of Pol II (for review, see Adelman and Lis 2012). Recent studies have shown that the paused Pol II is not a static entity, but is rather subject to rapid turnover (Krebs *et al.* 2017; Erickson *et al.* 2018). Thus, the paused Pol II is frequently regenerated via new initiation events at the core promoter.

It is interesting to consider whether the core promoter might have a role in the promoter-proximal pausing and premature termination of Pol II. Promoter elements that have been found to be associated with Pol II pausing include GAGA factor binding sites, Ohler motif 1, the Inr, and a sequence termed the pause button (PB) (Hendrix *et al.* 2008; Lee *et al.* 2008; Li and Gilmour 2013; Duarte *et al.* 2016). Notably, the PB is frequently located in the vicinity of the DPE, and its consensus sequence (KCGRWCG) has a resemblance to that of the DPE (Hendrix *et al.* 2008). It is not known, however, whether the DPE and PB are functionally related.

Practical Applications of Synthetic Core Promoters

It is possible to increase gene expression by using synthetic core promoters with optimized sequences. For instance, super core promoters (SCPs) were designed with highly active versions of the TATA box, Inr, MTE, and DPE motifs (Juven-Gershon *et al.* 2006). In addition to studies involving gene expression, SCPs have been employed in biochemical and biophysical analyses of TFIID and the PIC (see, for example: Cianfrocco *et al.* 2013; Louder *et al.* 2016). There are also CpG-deficient SCPs that are resistant to repression by CpG methylation (Theisen *et al.* 2013). For studies in *Drosophila*, the *Drosophila* synthetic core promoter (DSCP; Pfeiffer *et al.* 2008) has been used in the analysis of enhancer activity (see, for example: Pfeiffer *et al.* 2008; Arnold *et al.* 2013; Zabidi *et al.* 2015). The DSCP contains both TATA box and DPE motifs, and is thus able to function with both TATA-specific enhancers and DPE-specific enhancers.

For any particular gene, the replacement of the core promoter region ($\sim$ −40 to +40 relative to the +1 TSS, as in

Figure 1) with an SCP sequence would likely result in an increase in gene expression. The core promoter region can be identified by mapping of the +1 TSS (in *Drosophila*, see: http://labs.biology.ucsd.edu/Kadonaga/drosophila.tss. data/), and by the possible presence of position-specific core promoter motifs such as the TATA box, Inr, MTE, and DPE. For instance, if there is a distinct +1 TSS that correlates with a consensus Inr motif and perhaps a correctly positioned TATA-like sequence or DPE, then there is a good likelihood that it is a core promoter. In addition, when possible, core promoters can be confirmed and characterized by *in vitro* transcription analysis.

Perspectives

The RNA polymerase II core promoter is a diverse regulatory element. Here, we have described some of DNA sequence motifs and transcription factors that function in the core promoter region. It is our hope that this knowledge will be useful for studies of the regulation of gene activity in *Drosophila* and other bilateria.

In the future, we hope to gain new insights into the many remaining unsolved issues relating to the core promoter. From the perspective of the basal transcription process, it would be important to determine the factors and mechanisms that mediate TATA-less transcription, as the majority of promoters in *Drosophila* as well as in humans lack a TATA box or a TATA-like sequence. A related matter is the function of ssTFs that bind near the TSS of TATA-less promoters. It is possible, for instance, that ssTFs bound to their recognition sites act in a manner that is analogous to TBP bound to the TATA box (Figure 4). The activation of TATA-less transcription by ssTFs might also be related to the synthesis of species such as enhancer RNAs (eRNAs). More generally, as discussed here and elsewhere (Vo Ngoc *et al.* 2017a), it would likely be useful to expand the concept of the core promoter to incorporate the functions of ssTFs and chromatin structure and modifications.

From the perspective of gene regulation, it will be essential to elucidate the mechanism of enhancer-core promoter specificity. Such studies should reveal new insights not only into enhancer-promoter communication, but also into the factors that mediate transcriptional activation by enhancers. At the level of gene networks, it would be important to understand how core promoter motifs control groups of genes, such as in the regulation of ribosomal protein gene expression by the TCT motif, which is present in almost all ribosomal protein gene promoters in *Drosophila* and humans (Parry *et al.* 2010). With respect to gene networks, it is also notable that nearly all of the *Drosophila* Hox genes have TATA-less, DPE-driven promoters (Juven-Gershon *et al.* 2008). Thus, core promoter motifs have roles in broader biological processes.

In conclusion, the core promoter is a central component of the transcriptional apparatus. It is a downstream target of signals that lead to gene activation, and it is upstream of events that occur subsequent to transcription initiation. It can thus be

seen that our understanding of the mechanisms of gene regulation are critically dependent upon our knowledge of the multifarious functions of this "gateway to transcription."

Acknowledgments

Literature Cited

Adelman, K., and J. T. Lis, 2012   Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat. Rev. Genet. 13: 720–731. https://doi.org/10.1038/nrg3293

Akhtar, W., and G. J. Veenstra, 2011   TBP-related factors: a paradigm of diversity in transcription initiation. Cell Biosci. 1: 23. https://doi.org/10.1186/2045-3701-1-23

Arnold, C. D., D. Gerlach, C. Stelzer, Ł. M. Boryń, M. Rath *et al.*, 2013   Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science 339: 1074–1077. https://doi.org/10.1126/science.1232542

Baumann, D. G., and D. S. Gilmour, 2017   A sequence-specific core promoter-binding transcription factor recruits TRF2 to coordinately transcribe ribosomal protein genes. Nucleic Acids Res. 45: 10481–10491. https://doi.org/10.1093/nar/gkx676

Burke, T. W., and J. T. Kadonaga, 1996   *Drosophila* TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. Genes Dev. 10: 711–724. https://doi.org/10.1101/gad.10.6.711

Burke, T. W., and J. T. Kadonaga, 1997   The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAF$_{II}$60 of *Drosophila*. Genes Dev. 11: 3020–3031. https://doi.org/10.1101/gad.11.22.3020

Butler, J. E., and J. T. Kadonaga, 2001   Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. Genes Dev. 15: 2515–2519. https://doi.org/10.1101/gad.924301

Capuano, F., M. Mülleder, R. Kok, H. J. Blom, and M. Ralser, 2014   Cytosine DNA methylation is found in *Drosophila melanogaster* but absent in *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and other yeast species. Anal. Chem. 86: 3697–3702. https://doi.org/10.1021/ac500447w

Chalkley, G. E., and C. P. Verrijzer, 1999   DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250-TAF(II)150 complex recognizes the initiator. EMBO J. 18: 4835–4845. https://doi.org/10.1093/emboj/18.17.4835

Chen, Z. X., D. Sturgill, J. Qu, H. Jiang, S. Park *et al.*, 2014   Comparative validation of the *D. melanogaster* modENCODE transcriptome annotation. Genome Res. 24: 1209–1223. https://doi.org/10.1101/gr.159384.113

Cianfrocco, M. A., G. A. Kassavetis, P. Grob, J. Fang, T. Juven-Gershon *et al.*, 2013   Human TFIID binds to core promoter DNA in a reorganized structural state. Cell 152: 120–131. https://doi.org/10.1016/j.cell.2012.12.005

Corden, J., B. Wasylyk, A. Buchwalder, P. Sassone-Corsi, C. Kedinger *et al.*, 1980   Promoter sequences of eukaryotic protein-coding genes. Science 209: 1406–1414. https://doi.org/10.1126/science.6251548

Crowley, T. E., T. Hoey, J. K. Liu, Y. N. Jan, L. Y. Jan *et al.*, 1993   A new factor related to TATA-binding protein has highly restricted expression patterns in *Drosophila*. Nature 361: 557–561. https://doi.org/10.1038/361557a0

Danino, Y. M., D. Even, D. Ideses, and T. Juven-Gershon, 2015   The core promoter: at the heart of gene expression. Biochim. Biophys. Acta 1849: 1116–1131. https://doi.org/10.1016/j.bbagrm.2015.04.003

Dantonel, J. C., J. M. Wurtz, O. Poch, D. Moras, and L. Tora, 1999   The TBP-like factor: an alternative transcription factor in Metazoa? Trends Biochem. Sci. 24: 335–339. https://doi.org/10.1016/S0968-0004(99)01436-X

Dantonel, J. C., S. Quintin, L. Lakatos, M. Labouesse, and L. Tora, 2000   TBP-like factor is required for embryonic RNA polymerase II transcription in *C. elegans*. Mol. Cell 6: 715–722. https://doi.org/10.1016/S1097-2765(00)00069-1

Deng, W., and S. G. Roberts, 2005   A core promoter element downstream of the TATA box that is recognized by TFIIB. Genes Dev. 19: 2418–2423. https://doi.org/10.1101/gad.342405

Duarte, F. M., N. J. Fuda, D. B. Mahat, L. J. Core, M. J. Guertin *et al.*, 2016   Transcription factors GAF and HSF act at distinct regulatory steps to modulate stress-induced gene activation. Genes Dev. 30: 1731–1746. https://doi.org/10.1101/gad.284430.116

Duttke, S. H. C., R. F. Doolittle, Y. L. Wang, and J. T. Kadonaga, 2014   TRF2 and the evolution of the bilateria. Genes Dev. 28: 2071–2076. https://doi.org/10.1101/gad.250563.114

Emami, K. H., W. W. Navarre, and S. T. Smale, 1995   Core promoter specificities of the Sp1 and VP16 transcriptional activation domains. Mol. Cell. Biol. 15: 5906–5916. https://doi.org/10.1128/MCB.15.11.5906

Erickson, B., R. M. Sheridan, M. Cortazar, and D. L. Bentley, 2018   Dynamic turnover of paused Pol II complexes at human promoters. Genes Dev. 32: 1215–1225. https://doi.org/10.1101/gad.316810.118

Evans, R., J. A. Fairley, and S. G. Roberts, 2001   Activator-mediated disruption of sequence-specific DNA contacts by the general transcription factor TFIIB. Genes Dev. 15: 2945–2949. https://doi.org/10.1101/gad.206901

FitzGerald, P. C., D. Sturgill, A. Shyakhtenko, B. Oliver, and C. Vinson, 2006   Comparative genomics of *Drosophila* and human core promoters. Genome Biol. 7: R53. https://doi.org/10.1186/gb-2006-7-7-r53

Fuller, M. T., 2016   Differentiation in stem cell lineages and in life: explorations in the male germ line stem cell lineage. Curr. Top. Dev. Biol. 116: 375–390. https://doi.org/10.1016/bs.ctdb.2015.11.041

Gershenzon, N. I., and I. P. Ioshikhes, 2005   Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. Bioinformatics 21: 1295–1300. https://doi.org/10.1093/bioinformatics/bti172

Goldberg, M. L., 1979   Sequence analysis of *Drosophila* histone genes. Ph.D. Dissertation, Stanford University, Stanford, CA.

Goodrich, J. A., and R. Tjian, 2010   Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. Nat. Rev. Genet. 11: 549–558. https://doi.org/10.1038/nrg2847

Haberle, V., and A. Stark, 2018   Eukaryotic core promoters and the functional basis of transcription initiation. Nat. Rev. Mol. Cell Biol. 19: 621–637. https://doi.org/10.1038/s41580-018-0028-8

Hansen, S. K., S. Takada, R. H. Jacobson, J. T. Lis, and R. Tjian, 1997   Transcription properties of a cell type-specific TATA-binding protein, TRF. Cell 91: 71–83. https://doi.org/10.1016/S0092-8674(01)80010-6

Hendrix, D. A., J.-W. Hong, J. Zeitlinger, D. S. Rokhsar, and M. S. Levine, 2008   Promoter elements associated with RNA Pol II stalling in the *Drosophila* embryo. Proc. Natl. Acad. Sci. USA 105: 7762–7767. https://doi.org/10.1073/pnas.0802406105

Hirose, F., M. Yamaguchi, H. Handa, Y. Inomata, and A. Matsukage, 1993   Novel 8-base pair sequence (*Drosophila* DNA replication-related element) and specific binding factor involved in the expression of *Drosophila* genes for DNA polymerase α and proliferating cell nuclear antigen. J. Biol. Chem. 268: 2092–2099.

Hirose, F., M. Yamaguchi, K. Kuroda, A. Omori, T. Hachiya *et al.*, 1996   Isolation and characterization of cDNA for DREF, a promoter-activating factor for *Drosophila* DNA replication-related genes. J. Biol. Chem. 271: 3930–3937. https://doi.org/10.1074/jbc.271.7.3930

Hochheimer, A., S. Zhou, S. Zheng, M. C. Holmes, and R. Tjian, 2002   TRF2 associates with DREF and directs promoter-selective gene expression in *Drosophila*. Nature 420: 439–445. https://doi.org/10.1038/nature01167

Holmes, M. C., and R. Tjian, 2000   Promoter-selective properties of the TBP-related factor TRF1. Science 288: 867–870. https://doi.org/10.1126/science.288.5467.867

Hoskins, R. A., J. M. Landolin, J. B. Brown, J. E. Sandler, H. Takahashi *et al.*, 2011   Genome-wide analysis of promoter architecture in *Drosophila melanogaster*. Genome Res. 21: 182–192. https://doi.org/10.1101/gr.112466.110

Hsu, J. Y., T. Juven-Gershon, M. T. Marr II, and K. J. Wright, 2008   TBP, Mot1, and NC2 establish a regulatory circuit that controls DPE-dependent versus TATA-dependent transcription. Genes Dev. 22: 2353–2358. https://doi.org/10.1101/gad.1681808

Hultmark, D., R. Klemenz, and W. J. Gehring, 1986   Translational and transcriptional control elements in the untranslated leader of the heat-shock gene hsp22. Cell 44: 429–438. https://doi.org/10.1016/0092-8674(86)90464-2

Isogai, Y., S. Takada, R. Tjian, and S. Keleş, 2007a   Novel TRF1/BRF target genes revealed by genome-wide analysis of *Drosophila* Pol III transcription. EMBO J. 26: 79–89. https://doi.org/10.1038/sj.emboj.7601448

Isogai, Y., S. Keleş, M. Prestel, A. Hochheimer, and R. Tjian, 2007b   Transcription of histone gene cluster by differential core-promoter factors. Genes Dev. 21: 2936–2949. https://doi.org/10.1101/gad.1608807

Juven-Gershon, T., S. Cheng, and J. T. Kadonaga, 2006   Rational design of a super core promoter that enhances gene expression. Nat. Methods 3: 917–922. https://doi.org/10.1038/nmeth937

Juven-Gershon, T., J. Y. Hsu, and J. T. Kadonaga, 2008   Caudal, a key developmental regulator, is a DPE-specific transcription factor. Genes Dev. 22: 2823–2830. https://doi.org/10.1101/gad.1698108

Kadonaga, J. T., 2004   Regulation of RNA polymerase II transcription by sequence-specific DNA-binding factors. Cell 116: 247–257. https://doi.org/10.1016/S0092-8674(03)01078-X

Kadonaga, J. T., 2012   Perspectives on the RNA polymerase II core promoter. Wiley Interdiscip. Rev. Dev. Biol. 1: 40–51. https://doi.org/10.1002/wdev.21

Kaltenbach, L., M. A. Horner, J. H. Rothman, and S. E. Mango, 2000   The TBP-like factor CeTLF is required to activate RNA polymerase II transcription during *C. elegans* embryogenesis. Mol. Cell 6: 705–713. https://doi.org/10.1016/S1097-2765(00)00068-X

Kedmi, A., Y. Zehavi, Y. Glick, Y. Orenstein, D. Ideses *et al.*, 2014   *Drosophila* TRF2 is a preferential core promoter regulator. Genes Dev. 28: 2163–2174. https://doi.org/10.1101/gad.245670.114

Kopytova, D. V., A. N. Krasnov, M. R. Kopantceva, E. N. Nabirochkina, J. V. Nikolenko *et al.*, 2006   Two isoforms of *Drosophila* TRF2 are involved in embryonic development, premeiotic chromatin condensation, and proper differentiation of germ cells of both sexes. Mol. Cell. Biol. 26: 7492–7505. https://doi.org/10.1128/MCB.00349-06

Krebs, A. R., D. Imanci, L. Hoerner, D. Gaidatzis, L. Burger *et al.*, 2017   Genome-wide single-molecule footprinting reveals high RNA polymerase II turnover at paused promoters. Mol. Cell 67: 411–422.e4. https://doi.org/10.1016/j.molcel.2017.06.027

Kruesi, W. S., L. J. Core, C. T. Waters, J. T. Lis, and B. J. Meyer, 2013   Condensin controls recruitment of RNA polymerase II to achieve nematode X-chromosome dosage compensation. eLife 2: e00808. https://doi.org/10.7554/eLife.00808

Kurshakova, M. M., E. N. Nabirochkina, S. G. Georgieva, and D. V. Kopytova, 2018   TRF4, the novel TBP-related protein of *Drosophila melanogaster*, is concentrated at the endoplasmic reticulum and copurifies with proteins participating in the processes associated with endoplasmic reticulum. J. Cell. Biochem. https://doi.org/10.1002/jcb.28070

Kutach, A. K., and J. T. Kadonaga, 2000   The downstream promoter element DPE appears to be as widely used as the TATA box in *Drosophila* core promoters. Mol. Cell. Biol. 20: 4754–4764. https://doi.org/10.1128/MCB.20.13.4754-4764.2000

Kvon, E. Z., T. Kazmar, G. Stampfel, J. O. Yáñez-Cuna, M. Pagani *et al.*, 2014   Genome-scale functional characterization of *Drosophila* developmental enhancers in vivo. Nature 512: 91–95. https://doi.org/10.1038/nature13395

Kwak, H., N. J. Fuda, L. J. Core, and J. T. Lis, 2013   Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science 339: 950–953. https://doi.org/10.1126/science.1229386

Lagrange, T., A. N. Kapanidis, H. Tang, D. Reinberg, and R. H. Ebright, 1998   New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. Genes Dev. 12: 34–44. https://doi.org/10.1101/gad.12.1.34

Lam, M. T. Y., H. Cho, H. P. Lesch, D. Gosselin, S. Heinz *et al.*, 2013   Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. Nature 498: 511–515. https://doi.org/10.1038/nature12209

Lee, C., X. Li, A. Hechmer, M. Eisen, M. D. Biggin *et al.*, 2008   NELF and GAGA factor are linked to promoter-proximal pausing at many genes in *Drosophila*. Mol. Cell. Biol. 28: 3290–3300. https://doi.org/10.1128/MCB.02224-07

Lenhard, B., A. Sandelin, and P. Carninci, 2012   Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nat. Rev. Genet. 13: 233–245. https://doi.org/10.1038/nrg3163

Li, J., and D. S. Gilmour, 2013   Distinct mechanisms of transcriptional pausing orchestrated by GAGA factor and M1BP, a novel transcription factor. EMBO J. 32: 1829–1841. https://doi.org/10.1038/emboj.2013.111

Lim, C. Y., B. Santoso, T. Boulay, E. Dong, U. Ohler *et al.*, 2004   The MTE, a new core promoter element for transcription by RNA polymerase II. Genes Dev. 18: 1606–1617. https://doi.org/10.1101/gad.1193404

Louder, R. K., Y. He, J. R. López-Blanco, J. Fang, P. Chacón *et al.*, 2016   Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. Nature 531: 604–609 (erratum: Nature 536: 112). https://doi.org/10.1038/nature17394

Martianov, I., G. M. Fimia, A. Dierich, M. Parvinen, P. Sassone-Corsi *et al.*, 2001   Late arrest of spermiogenesis and germ cell apoptosis in mice lacking the TBP-like TLF/TRF2 gene. Mol. Cell 7: 509–515. https://doi.org/10.1016/S1097-2765(01)00198-8

Massari, M. E., and C. Murre, 2000   Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. Mol. Cell. Biol. 20: 429–440. https://doi.org/10.1128/MCB.20.2.429-440.2000

Müller, F., L. Lakatos, J. Dantonel, U. Strähle, and L. Tora, 2001   TBP is not universally required for zygotic RNA polymerase II transcription in zebrafish. Curr. Biol. 11: 282–287. https://doi.org/10.1016/S0960-9822(01)00076-8

Nechaev, S., D. C. Fargo, G. dos Santos, L. Liu, Y. Gao *et al.*, 2010   Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in *Drosophila*. Science 327: 335–338. https://doi.org/10.1126/science.1181421

Ni, T., D. L. Corcoran, E. A. Rach, S. Song, E. P. Spana *et al.*, 2010   A paired-end sequencing strategy to map the complex landscape of transcription initiation. Nat. Methods 7: 521–527. https://doi.org/10.1038/nmeth.1464

Ohler, U., G. C. Liao, H. Niemann, and G. M. Rubin, 2002   Computational analysis of core promoters in the *Drosophila* genome. Genome Biol. 3: RESEARCH0087.

Ohtsuki, S., M. Levine, and H. N. Cai, 1998   Different core promoters possess distinct regulatory activities in the *Drosophila* embryo. Genes Dev. 12: 547–556. https://doi.org/10.1101/gad.12.4.547

Parry, T. J., J. W. M. Theisen, J. Y. Hsu, Y.-L. Wang, D. L. Corcoran *et al.*, 2010   The TCT motif, a key component of an RNA polymerase II transcription system for the translational machinery. Genes Dev. 24: 2013–2018. https://doi.org/10.1101/gad.1951110

Patikoglou, G. A., J. L. Kim, L. Sun, S.-H. Yang, T. Kodadek *et al.*, 1999   TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. Genes Dev. 13: 3217–3230. https://doi.org/10.1101/gad.13.24.3217

Perry, R. P., 2005   The architecture of mammalian ribosomal protein promoters. BMC Evol. Biol. 5: 15. https://doi.org/10.1186/1471-2148-5-15

Persengiev, S. P., X. Zhu, B. L. Dixit, G. A. Maston, E. L. Kittler *et al.*, 2003   TRF3, a TATA-box-binding protein-related factor, is vertebrate-specific and widely expressed. Proc. Natl. Acad. Sci. USA 100: 14887–14891. https://doi.org/10.1073/pnas.2036440100

Pfeiffer, B. D., A. Jenett, A. S. Hammonds, T. B. Ngo, S. Misra *et al.*, 2008   Tools for neuroanatomy and neurogenetics in *Drosophila*. Proc. Natl. Acad. Sci. USA 105: 9715–9720. https://doi.org/10.1073/pnas.0803697105

Purnell, B. A., P. A. Emanuel, and D. S. Gilmour, 1994   TFIID sequence recognition of the initiator and sequences farther downstream in *Drosophila* class II genes. Genes Dev. 8: 830–842. https://doi.org/10.1101/gad.8.7.830

Rabenstein, M. D., S. Zhou, J. T. Lis, and R. Tjian, 1999   TATA box-binding protein (TBP)-related factor 2 (TRF2), a third member of the TBP family. Proc. Natl. Acad. Sci. USA 96: 4791–4796. https://doi.org/10.1073/pnas.96.9.4791

Rach, E. A., H.-Y. Yuan, W. H. Majoros, P. Tomancak, and U. Ohler, 2009   Motif composition, conservation and condition-specificity of single and alternative transcription start sites in the *Drosophila* genome. Genome Biol. 10: R73. https://doi.org/10.1186/gb-2009-10-7-r73

Rach, E. A., D. R. Winter, A. M. Benjamin, D. L. Corcoran, T. Ni *et al.*, 2011   Transcription initiation patterns indicate divergent strategies for gene regulation at the chromatin level. PLoS Genet. 7: e1001274. https://doi.org/10.1371/journal.pgen.1001274

Rougvie, A. E., and J. T. Lis, 1988   The RNA polymerase II molecule at the 5′ end of the uninduced hsp70 gene of *D. melanogaster* is transcriptionally engaged. Cell 54: 795–804. https://doi.org/10.1016/S0092-8674(88)91087-2

Roy, A. L., and D. S. Singer, 2015   Core promoters in transcription: old problem, new insights. Trends Biochem. Sci. 40: 165–171. https://doi.org/10.1016/j.tibs.2015.01.007

Sainsbury, S., C. Bernecky, and P. Cramer, 2015   Structural basis of transcription initiation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. 16: 129–143. https://doi.org/10.1038/nrm3952

Schor, I. E., J. F. Degner, D. Harnett, E. Cannavò, F. P. Casale *et al.*, 2017   Promoter shape varies across populations and affects promoter evolution and expression noise. Nat. Genet. 49: 550–558. https://doi.org/10.1038/ng.3791

Shir-Shapira, H., J. Sharabany, M. Filderman, D. Ideses, A. Ovadia-Shochat *et al.*, 2015   Structure-function analysis of the *Drosophila melanogaster* Caudal transcription factor provides insights into core promoter-preferential activation. J. Biol. Chem. 290: 17293–17305 (erratum: J. Biol. Chem. 290: 20747). https://doi.org/10.1074/jbc.M114.632109

Smale, S. T., and D. Baltimore, 1989   The "initiator" as a transcription control element. Cell 57: 103–113. https://doi.org/10.1016/0092-8674(89)90176-1

Smale, S. T., and J. T. Kadonaga, 2003   The RNA polymerase II core promoter. Annu. Rev. Biochem. 72: 449–479. https://doi.org/10.1146/annurev.biochem.72.121801.161520

Smale, S. T., M. C. Schmidt, A. J. Berk, and D. Baltimore, 1990   Transcriptional activation by Sp1 as directed through TATA or initiator: specific requirement for mammalian transcription factor IID. Proc. Natl. Acad. Sci. USA 87: 4509–4513. https://doi.org/10.1073/pnas.87.12.4509

Soeller, W. C., S. J. Poole, and T. Kornberg, 1988   In vitro transcription of the *Drosophila engrailed* gene. Genes Dev. 2: 68–81. https://doi.org/10.1101/gad.2.1.68

Takada, S., J. T. Lis, S. Zhou, and R. Tjian, 2000   A TRF1:BRF complex directs *Drosophila* RNA polymerase III transcription. Cell 101: 459–469. https://doi.org/10.1016/S0092-8674(00)80857-0

Teichmann, M., Z. Wang, E. Martinez, A. Tjernberg, D. Zhang *et al.*, 1999   Human TATA-binding protein-related factor-2 (hTRF2) stably associates with hTFIIA in HeLa cells. Proc. Natl. Acad. Sci. USA 96: 13720–13725. https://doi.org/10.1073/pnas.96.24.13720

Theisen, J. W. M., C. Y. Lim, and J. T. Kadonaga, 2010   Three key subregions contribute to the function of the downstream RNA polymerase II core promoter. Mol. Cell. Biol. 30: 3471–3479. https://doi.org/10.1128/MCB.00053-10

Theisen, J. W. M., J. S. Gucwa, T. Yusufzai, M. T. Khuong, and J. T. Kadonaga, 2013   Biochemical analysis of histone deacetylase-independent transcriptional repression by MeCP2. J. Biol. Chem. 288: 7096–7104. https://doi.org/10.1074/jbc.M112.438697

Tsai, F. T. F., and P. B. Sigler, 2000   Structural basis of preinitiation complex assembly on human Pol II promoters. EMBO J. 19: 25–36. https://doi.org/10.1093/emboj/19.1.25

Veenstra, G. J., D. L. Weeks, and A. P. Wolffe, 2000   Distinct roles for TBP and TBP-like factor in early embryonic gene transcription in *Xenopus*. Science 290: 2312–2315. https://doi.org/10.1126/science.290.5500.2312

Verma, N., K. H. Hung, J. J. Kang, N. H. Barakat, and W. E. Stumph, 2013   Differential utilization of TATA box-binding protein (TBP) and TBP-related factor 1 (TRF1) at different classes of RNA polymerase III promoters. J. Biol. Chem. 288: 27564–27570. https://doi.org/10.1074/jbc.C113.503094

Vo Ngoc, L., Y. L. Wang, G. A. Kassavetis, and J. T. Kadonaga, 2017a   The punctilious RNA polymerase II core promoter. Genes Dev. 31: 1289–1301. https://doi.org/10.1101/gad.303149.117

Vo Ngoc, L., C. J. Cassidy, C. Y. Huang, S. H. Duttke, and J. T. Kadonaga, 2017b   The human initiator is a distinct and abundant element that is precisely positioned in focused core promoters. Genes Dev. 31: 6–11. https://doi.org/10.1101/gad.293837.116

Wang, J., S. Zhao, W. He, Y. Wei, Y. Zhang *et al.*, 2017   A transcription factor IIA-binding site differentially regulates RNA polymerase II-mediated transcription in a promoter context-dependent manner. J. Biol. Chem. 292: 11873–11885. https://doi.org/10.1074/jbc.M116.770412

Wang, Y.-L., S. H. C. Duttke, K. Chen, J. Johnston, G. A. Kassavetis *et al.*, 2014   TRF2, but not TBP, mediates the transcription of ribosomal protein genes. Genes Dev. 28: 1550–1555. https://doi.org/10.1101/gad.245662.114

Willy, P. J., R. Kobayashi, and J. T. Kadonaga, 2000 A basal transcription factor that activates or represses transcription. Science 290: 982–984. https://doi.org/10.1126/science.290.5493.982

Zabidi, M. A., C. D. Arnold, K. Schernhuber, M. Pagani, M. Rath et al., 2015 Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. Nature 518: 556–559. https://doi.org/10.1038/nature13994

Zehavi, Y., A. Kedmi, D. Ideses, and T. Juven-Gershon, 2015 TRF2: TRansForming the view of general transcription factors. Transcription 6: 1–6. https://doi.org/10.1080/21541264.2015.1004980

Zhang, D., T. L. Penttila, P. L. Morris, M. Teichmann, and R. G. Roeder, 2001 Spermiogenesis deficiency in mice lacking the Trf2 gene. Science 292: 1153–1155. https://doi.org/10.1126/science.1059188

Zhou, H., I. Grubisic, K. Zheng, Y. He, P. J. Wang et al., 2013 Taf7l cooperates with Trf2 to regulate spermiogenesis. Proc. Natl. Acad. Sci. USA 110: 16886–16891. https://doi.org/10.1073/pnas.1317034110

Communicating editor: B. Oliver