REVIEW

The punctilious RNA polymerase II core promoter

Long Vo ngoc, Yuan-Liang Wang, George A. Kassavetis, and James T. Kadonaga

Section of Molecular Biology, University of California at San Diego, La Jolla, California 92093, USA

The signals that direct the initiation of transcription ultimately converge at the core promoter, which is the gateway to transcription. Here we provide an overview of the RNA polymerase II core promoter in bilateria (bilaterally symmetric animals). The core promoter is diverse in terms of its composition and function yet is also punctilious, as it acts with strict rules and precision. We additionally describe an expanded view of the core promoter that comprises the classical DNA sequence motifs, sequence-specific DNA-binding transcription factors, chromatin signals, and DNA structure. This model may eventually lead to a more unified conceptual understanding of the core promoter.

Supplemental material is available for this article.

The RNA polymerase II (Pol II) transcription system is a key component in the expression of protein-coding genes as well as many noncoding genes in eukaryotes. The initiation of Pol II transcription is mediated by a stretch of DNA known as the 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). The core promoter is sometimes referred to as the gateway to transcription, as the signals that lead to the initiation of transcription ultimately converge at the core promoter.

In the past, the core promoter was often thought to be a generic element—a stretch of DNA with a TATA box that functions universally at all genes. It then became apparent, however, that the TATA box is present in only a small fraction of metazoan core promoters and that there are no universal core promoter elements. Further studies revealed the diversity of the core promoter in terms of its composition as well as its function.

Moreover, it became apparent that the core promoter is punctilious—precise sequences at precise locations are essential for core promoter function. Some core promoter elements are involved in enhancer–core promoter specif-

icity as well as specific biological networks. In addition, there are intriguing connections between chromatin structure (including histone modifications) in the core promoter region and transcriptional activity.

Here, we discuss the initiation of transcription in bilateria (bilaterally symmetric animals) from the perspective of the Pol II core promoter. Topics include the nature of transcription start sites (TSSs), core promoter sequence motifs, enhancer–promoter specificity, TATA box-binding protein (TBP) and related factors, transcriptional directionality, and an overall view of the components that contribute to the initiation of transcription. We focus in particular on data derived from functional analyses of core promoter elements. However, we do not include discussion of CpG islands, in which mammalian promoters are frequently located, but rather direct the reader to excellent review articles on this subject (Deaton and Bird 2011; Schübeler 2015). A few underlying themes in this essay are the diversity of core promoters, the punctilious nature of core promoters, and the multifarious components that contribute to core promoter function. It is notable that many fundamental and important questions about the core promoter have yet to be answered.

Focused vs. dispersed transcription patterns

There are different transcription initiation patterns that are observed with Pol II (Fig. 1). Purified Pol II itself does not specifically recognize the core promoter. Instead, Pol II and a set of auxiliary factors (for instance, TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH at TATA box-dependent promoters) assemble into a transcription preinitiation complex (PIC) at the core promoter. (Note that TFIID, which consists of TBP and ~13–15 TBP-associated factors [TAFs], is a key factor in the recognition of sequence motifs at the core promoter.) Upon addition of the ribonucleoside 5′ triphosphates, transcription rapidly initiates from the PIC (for a recent review, see Sainsbury et al. 2015). A pattern of transcription from a single site or a narrow cluster of sites (≤5-nucleotide [nt] window) is probably derived
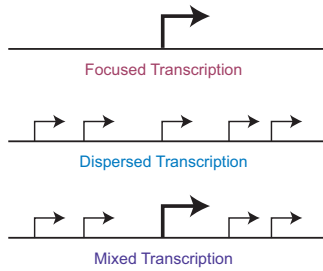
Figure 1. Focused, dispersed, and mixed transcription initiation patterns. In focused transcription, there is either a single predominant TSS or a narrow cluster of TSSs that probably derive from a single PIC. In dispersed transcription, there are multiple weak TSSs spread over an ∼50- to 100-base-pair (bp) region that likely emanate from multiple PICs. Focused and dispersed transcription patterns are two endpoints of a spectrum of possible mechanisms, and a variety of mixed TSS patterns are commonly observed. TSS patterns are also known as promoter shape.

from a specific PIC at a core promoter (for example, see Kadonaga 1990). We refer to this TSS pattern as "focused" (also known as "narrow peak," "peaked," "sharp peak," and "single peak") (see also Juven-Gershon et al. 2008a; Kadonaga 2012).

In contrast to focused transcription, there is also "dispersed" (also known as "broad" and "weak") transcription, in which there is a pattern of several weak TSSs that are distributed over a region that might span 50–100 nt. The mechanisms and factors that are involved in dispersed transcription remain to be determined. Transcription from the multiple TSSs may occur via a mechanism that involves multiple PICs. It is also notable that dispersed promoter regions are deficient in ATG codons (termed "ATG deserts") (Lee et al. 2005a). The presence of an ATG desert would enable a single protein to be encoded by a promoter region with multiple TSSs.

Focused and dispersed TSS patterns represent two endpoints of a spectrum of transcription mechanisms, and mixed promoters, such as those with multiple weak TSSs and a major predominant TSS, are often observed. The range of transcriptional patterns at promoters is sometimes referred to as "promoter shape."

Focused TSSs are frequently observed in regulated promoters, whereas dispersed TSSs are typically associated with ubiquitously expressed promoters (Hoskins et al. 2011). In addition, promoter shape is generally conserved between species (Carninci et al. 2006; Main et al. 2013). Moreover, the analysis of 81 different *Drosophila melanogaster* lines revealed that focused promoters are more evolutionarily constrained than dispersed promoters (Schor et al. 2017). From a teleological standpoint, it might be advantageous for regulated genes to be turned on and off at a single TSS at focused promoters and for constitutively active genes to maintain a steady stream of transcription via multiple TSSs at dispersed promoters.

It is also relevant to note that a key technical issue in the study of focused and dispersed promoters is the accurate determination of the TSSs. For example, processing or degradation of transcripts could lead to the inadvertent

misidentification of TSSs. To minimize this problem, it is useful to map the 5′ ends of capped nascent transcripts by using a method such as Start-seq (Nechaev et al. 2010), GRO-cap (global run on cap) (Kruesi et al. 2013), or 5′-GRO-seq (5′ end-selected GRO followed by sequencing) (Lam et al. 2013). To date, however, most studies of promoter shape have been performed with accumulated steady-state RNAs. Hence, new insights might be gained from the analysis of promoter shape with TSSs that are determined by the mapping of nascent transcripts. For instance, recent analyses of nascent transcripts suggest that most human promoters have mixed (i.e., combined focused and dispersed) transcription patterns (Lai and Pugh 2017) and that dispersed transcription occurs less frequently than previously thought from the analysis of steady-state RNAs (Core et al. 2014; Scruggs et al. 2015).

Core promoter sequence motifs

The activity of the core promoter is largely dependent on the presence or absence of specific DNA sequences known as core promoter elements or motifs. Importantly, core promoters are diverse not only in terms of the presence or absence of particular sequence motifs but also with regard to the distinct functions that are mediated by specific core promoter elements. Some of the known core promoter motifs in bilaterians are shown in Figure 2 and Table 1. These sequence elements have been studied mostly in focused promoters.

There are no universal core promoter elements. Moreover, many core promoters lack any of the known motifs. Hence, there are probably other core promoter elements that remain to be discovered. Brief summaries of some core promoter motifs are as follows.

The initiator (Inr)

The Inr motif is probably the most widely used core promoter motif in bilateria. It was originally found by Chambon and colleagues (Corden et al. 1980) and was incisively articulated as a discrete core promoter element by Smale and Baltimore (1989). The Inr encompasses the TSS and is recognized by the TAF1 and TAF2 subunits of TFIID (Chalkley and Verrijzer 1999; Louder et al. 2016).
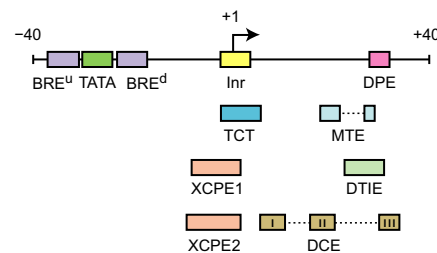

Figure 2. A plethora of core promoter sequence motifs for RNA Pol II. A typical core promoter might have zero to three of the indicated core promoter elements. The locations of the sequence motifs are roughly to scale. The consensus sequences are listed in Table 1.

Table 1. *Consensus sequences of some core promoter elements*

Motif	Location	Consensus
TATA box	Upstream T at −32 to −28	TATAWR
BRE[u]	Upstream of TATA box	SSRCGCC
BRE[d]	−23 to −17	RTDKKKK
Inr	−2 to +4	$TCA_{+1}GTY$ (*Drosophila*)
	−3 to +3	$BBCA_{+1}BW$ (human)
TCT	−2 to +6	$YYC_{+1}TTTYY$ (*Drosophila*)
	−1 to +6	$YC_{+1}TYTYY$ (human)
XCPE1	−8 to +2	$DSGYGGRAS_{+1}M$
XCPE2	−9 to +2	$VCYCRTTRCM_{+1}Y$
MTE	+18 to +22	CGANC
	+27 to +29	CGG
DPE	+28 to +32	RGWYV
DCE	Box I: +6 to +11	CTTC
	Box II: +16 to +21	CTGT
	Box III: +30 to +34	AGC
DTIE	+23 to +31	GSGRDNHGG

(W) A or T; (R) A or G; (S) G or C; (D) A, G, or T (not C); (K) G or T; (Y) C or T; (B) C, G, or T (not A); (M) A or C; (V) A, C, or G (not T); (N) A, C, G, or T (any base); (H) A, C, or T (not G). (BRE) TFIIB recognition element upstream (u) or downstream (d); (Inr) initiator; (XCPE1) X core promoter element 1; (MTE) motif ten element; (DPE) downstream core promoter element; (DCE) downstream core element; (DTIE) downstream transcription initiation element.

In human cells, the analysis of focused TSSs in nascent transcripts (5′-GRO-seq and GRO-cap methods) revealed the Inr consensus sequence of $BBCA_{+1}BW$ (where B is C/G/T, and W is A/T) (Vo ngoc et al. 2017; for earlier versions of the Inr consensus, see Javahery et al. 1994; Lo and Smale 1996; Carninci et al. 2006). Over half of focused human promoters contain either a perfect match to the $BBCA_{+1}$-BW Inr consensus or an Inr-like sequence with only a single mismatch outside of the CA_{+1} central core (Vo ngoc et al. 2017). To test the Inr consensus further, we analyzed focused TSSs in nascent transcripts (Start-seq method) from mouse cells (Scruggs et al. 2015) and also observed the precise placement of the same $BBCA_{+1}BW$ Inr consensus in the core promoter (Supplemental Fig. S1). This precision in the positioning of the Inr consensus sequence is an example of the punctilious nature of the core promoter.

The human and mouse $BBCA_{+1}BW$ Inr consensus is similar but not identical to the *Drosophila* Inr consensus, $TCA_{+1}GTY$ (where Y is C/T) (Purnell et al. 1994; Chalkley and Verrijzer 1999; Ohler et al. 2002; FitzGerald et al. 2006). The *Drosophila* Inr consensus appears to be a more restrictive version of the human/mouse Inr consensus. From an evolutionary perspective, it would be interesting to determine the Inr consensus sequences in diverse organisms and perhaps gain insight into whether the Inr had become more restrictive in *Drosophila* or less restrictive in mammals.

The A_{+1} in the Inr consensus sequence is usually the major site of transcription initiation and is designated as the +1 TSS position. The A_{+1} notation provides a specific reference point in the core promoter whether there is a single TSS or a cluster of TSSs. In addition, other core promoter motifs, such as the downstream core promoter element (DPE) and motif ten element (MTE) (see below), function with the Inr and are strictly positioned with respect to the A_{+1} in the Inr consensus.

The TATA box

The TATA box is the first discovered core promoter motif in eukaryotes (Goldberg 1979) and was named after the TATAAAA sequence that is present in some upstream promoter regions. It is bound by the TBP subunit of the TFIID transcription factor (Sainsbury et al. 2015). The TATA box and TBP are ancient, as both are present in Archaea and eukaryotes (for example, see Blombach et al. 2016).

The TATA box consensus has been investigated by the analysis of promoter sequences (e.g., STATAWAWR [simplified version of position-weight matrix from Bucher 1990], STATAWAAR [Ohler et al. 2002], STATAAA and TATAWRD [FitzGerald et al. 2006], and TATAAR [Vo ngoc et al. 2017], where W is A/T, R is A/G, S is C/G, and D is A/G/T) as well as the study of TBP binding to DNA (e.g., STATATAAGS [Wong and Bateman 1994] and TATATAWR [Patikoglou et al. 1999]). These TATA sequences mostly share the TATAWR motif, which is recommended as a general TATA consensus with the upstream T located at a position from −32 to −28 relative to the +1 TSS.

Promoters with a strict adherence to the TATA box consensus are somewhat rare. For instance, only ~3.5% of focused human promoters were found to have a perfect match to TATAAR (with the upstream T located from −33 to −28 relative to the +1 TSS) (Vo ngoc et al. 2017). Moreover, only ~28% of focused human promoters were observed to have WWWW (an extremely loose TATA-like sequence) in the region from −33 to −23 relative to the +1 TSS. Hence, most promoters lack TATA or TATA-like sequences, and it is important to understand the DNA sequence elements and transcription factors that mediate TATA-less transcription.

It is also useful to note that TATA-containing promoters may or may not have Inr motifs. In fact, in human focused promoters, the occurrence of the TATA box is

higher in the absence of the Inr or Inr-like sequences than in the presence of the Inr or Inr-like sequences (Vo ngoc et al. 2017). These observations suggest that some TATA boxes can drive transcription in the absence of an Inr. In other instances, the TATA and Inr can function synergistically for the recruitment of TFIID in a process that exhibits a strict spacing dependence between the two elements (Emami et al. 1997). It has also been found that transcription from TATA + Inr promoters is facilitated by high mobility group A1 (HMGA1) protein and Mediator (Xu et al. 2011).

The BRE^u and BRE^d

The basal transcription factor TFIIB binds to the TBP–TATA box complex to form a ternary complex in which TFIIB interacts with TBP as well as DNA flanking the TATA box (Sainsbury et al. 2015). The TFIIB–DNA contact sites that are upstream of and downstream from the TATA box are known as the TFIIB recognition elements BRE^u and BRE^d (Lagrange et al. 1998; Deng and Roberts 2005). Because the sequence-specific interaction of TFIIB with DNA is dependent on the binding of TBP to the TATA box, the presence of a TATA box is required for a promoter to have functional BRE motifs.

The BRE^u is immediately upstream of the TATA box, and the G in its consensus sequence (SSRCGCC) (Lagrange et al. 1998) appears to be the single most important nucleotide. The BRE^d (consensus sequence RTDKKKK) (Deng and Roberts 2005) is located immediately downstream from the 8-nt version of the TATA box (e.g., TATATAWR). It should be noted, however, that these consensus sequences have not been confirmed or revised with more recent data and methodology.

Like TBP and the TATA box, TFIIB and the BREs are present in Archaea and eukaryotes (see, e.g., Blombach et al. 2016). Hence, the BRE is an ancient promoter element. However, the functions of the BRE motifs are not yet known. Depending on the promoter context, they have been found to have a positive or negative effect on transcriptional activity (Lagrange et al. 1998; Evans et al. 2001; Deng and Roberts 2005). Intriguingly, the BRE^u was also found to suppress the ability of Caudal, a sequence-specific DNA-binding transcription factor, to activate transcription from TATA-dependent promoters (Juven-Gershon et al. 2008b). These findings indicate that further investigation of the BRE motif is likely to reveal interesting and important aspects of basal and regulated transcription.

The TCT motif

The TCT motif (also known as the polypyrimidine Inr) (Perry 2005) is present in the core promoters of nearly all of the ribosomal protein genes in *Drosophila* and humans (Parry et al. 2010). This element encompasses the TSS and has the consensus of YYC$_{+1}$TTTYY in *Drosophila* (Parry et al. 2010) and YC$_{+1}$TYTYY in humans (Parry et al. 2010; Vo ngoc et al. 2017), where transcription initiates at C$_{+1}$ rather than at A$_{+1}$ as is seen in Inr-containing pro-

moters. The term "TCT" motif refers to the TCT trinucleotide that frequently encompasses the +1 TSS.

The TCT motif is a rare motif that is found only in ribosomal protein gene core promoters and a small number of other promoters, many of which are associated with genes that encode proteins involved in translation (Parry et al. 2010). In humans, it is estimated that ∼1% of focused core promoters contain a TCT motif (Vo ngoc et al. 2017). Hence, the TCT motif is an example of a core promoter motif that is rare but biologically important. The TCT motif regulates the network of ribosomal protein genes and is thus the complement to the RNA Pol I and RNA Pol III transcription systems, which synthesize ribosomal RNAs.

The TCT motif is distinct from the Inr, but a single T-to-A substitution can convert a TCT motif into a functionally active Inr (Parry et al. 2010). These findings further reveal the punctilious nature of the core promoter. Precise sequences and precise positioning are essential features of core promoter function. In addition, TCT-dependent transcription in *Drosophila* involves the use of TBP-related factor 2 (TRF2) instead of the more commonly used TBP (discussed in more detail below; Wang et al. 2014).

The DPE

The DPE functions cooperatively with the Inr for TFIID binding and transcriptional activity (Burke and Kadonaga 1996). The DPE is commonly found in *Drosophila* (∼30% of core promoters) and appears to be rare in humans (Burke and Kadonaga 1997; Kutach and Kadonaga 2000). The *Drosophila* DPE consensus is RGWYV from +28 to +32 (or RGWYVT from +28 to +33) relative to the A$_{+1}$ in the Inr (Kutach 2000; Kutach and Kadonaga 2000), and the human DPE consensus remains to be determined. There is a strict spacing requirement in the positioning of the DPE and Inr, as an increase or decrease of only a single nucleotide between the two elements results in a several-fold decrease in transcriptional activity as well as a reduction in the binding of TFIID (Kutach and Kadonaga 2000). This strict positioning requirement is another example of the punctilious nature of the core promoter.

As mentioned above, the DPE has rarely been found in human core promoters. This may be due to the scarcity of the DPE in humans and/or the lack of understanding of the human DPE consensus sequence. Functional human DPE motifs that resemble the *Drosophila* DPE have been found in the human *IRF1*, *CALM2*, and *TAF7* gene promoters (Burke and Kadonaga 1997; Zhou and Chiang 2001, 2002; Duttke 2014). Notably, human transcription factors exhibit higher activity with wild-type DPE motifs than with mutant DPE motifs (with nucleotide substitutions or alteration of the Inr to DPE spacing) in both cells and biochemical experiments (Burke and Kadonaga 1997; Zhou and Chiang 2001, 2002; Lewis et al. 2005; Juven-Gershon et al. 2006; Duttke 2014). These findings indicate that human transcription factors can recognize and function with the DPE. It was also found that DPE-specific transcription in humans involves Mediator, casein kinase II (CK2), and positive coactivator 4 (PC4) (Lewis et al. 2005). In the future, it will be important to analyze further

the sequence consensus, Inr–DPE spacing, abundance, and transcription factor requirements of the human DPE.

MTE

The MTE was identified as an overrepresented sequence in *Drosophila* core promoters (Ohler et al. 2002) and then found to be a TFIID-binding site and a core promoter element that functions cooperatively with a precisely positioned Inr (Lim et al. 2004). The original consensus of the MTE was CSARCSSAAC from +18 to +27 relative to the A+1 in the Inr. A more detailed analysis revealed that there are three key contact points for the binding of TFIID to the downstream core promoter region and that the first and second contact points constitute the MTE and the second and third contact points constitute the DPE (Theisen et al. 2010). Hence, a tentative revised MTE consensus is CGANC from +18 to +22 and CGG from +27 to +29 (Table 1). In the future, it will be important to gain a unified understanding of the TFIID–DNA interactions in the downstream core promoter region that support core promoter activity.

In this regard, the structure of human TFIID bound to a super core promoter that contains TATA, Inr, MTE, and DPE motifs (Juven-Gershon et al. 2006) revealed contacts of the TAF1 and TAF2 subunits of TFIID with the downstream core promoter region (Louder et al. 2016). Consistent with these findings, TFIID–DNA photocross-linking experiments with a reagent that extends from the DNA backbone phosphate detected the close proximity of the TAF1 subunit of TFIID with the downstream promoter (Kutach 2000). In contrast, photocross-linking studies with a reagent that extends from the DNA major groove indicated the close proximity of TAF6 and TAF9 (but not TAF1 or TAF2) to the MTE and DPE sequences (Burke and Kadonaga 1997; Theisen et al. 2010). Moreover, TAF6–TAF9 complexes were found to interact with the DPE (Shao et al. 2005). These different TFIID–DNA contacts may be due to different conformations of TFIID, as seen, for example, by Cianfrocco et al. (2013). It will also be important to determine the functions of these TAF–DNA contacts in the formation of the PIC.

Other core promoter elements

Some additional core promoter motifs include the following. The X core promoter element 1 (XCPE1) and XCPE2 motif were identified in the hepatitis B virus X gene promoter (Tokusumi et al. 2007; Anish et al. 2009). The downstream core element (DCE) was found in the human β-globin promoter (Lewis et al. 2000) and comprises three subelements in the +6 to +34 region situated in close proximity to TAF1 (Lee et al. 2005b). Three downstream elements, termed GLE, DPE-L1, and DPE-L2, were found in the +4 to +32 region of MHC class I promoters (Lee et al. 2010). Another downstream motif, the DTIE (downstream transcription initiation element), was identified in the microRNA miR-22 promoter (Marbach-Bar et al. 2016). Furthermore, because there are many promoters with no known core promoter elements, there may be as yet undis-covered motifs with interesting and important biological functions.

Enhancer–core promoter specificity

In addition to their role in the basal transcription process, core promoter motifs such as the DPE and TATA box are involved in the regulation of gene expression by transcriptional enhancers (Fig. 3). For instance, when test enhancers were placed between divergently transcribed promoters, the *Drosophila* AE1 and IAB5 enhancers were found to activate transcription preferentially from the TATA-dependent *even-skipped* promoter relative to the DPE-dependent *white* promoter (Ohtsuki et al. 1998; the *white* promoter was found to be DPE-dependent by Kutach and Kadonaga 2000). Moreover, in studies that directly compared the ability of enhancers to activate transcription from a TATA- or DPE-dependent core promoter in the same context, both DPE- and TATA-specific enhancers were observed (Butler and Kadonaga 2001). Enhancer–core promoter specificity was also seen at the genome-wide level in the comparison of a developmental core promoter (a synthetic core promoter with TATA, Inr, MTE, and DPE motifs) with a housekeeping core promoter (the TCT motif-containing ribosomal protein S12 gene promoter) (Zabidi et al. 2015). Hence, these findings reveal that transcriptional enhancers can distinguish between different core promoters and indicate that the specificity between enhancers and their cognate promoters can be achieved at least in part via core promoter motifs.

Enhancer–core promoter specificity was further examined in the context of the homeotic (Hox) gene network in *Drosophila* (Juven-Gershon et al. 2008b). Nearly all of the *Drosophila* Hox genes contain DPE-dependent core promoters, and Caudal, a sequence-specific DNA-binding transcription factor and key regulator of the Hox gene network, preferentially activates transcription from DPE-dependent promoters relative to some, but not all, TATA-dependent promoters (Juven-Gershon et al. 2008b; Shir-Shapira et al. 2015). In addition, the presence of the BRE^u motif suppresses the ability of Caudal to function in conjunction with the TATA box. These results show
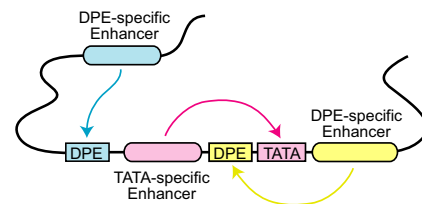

Figure 3. Enhancer–core promoter specificity. This diagram depicts transcriptional enhancers that function selectively with DPE-dependent or TATA-dependent core promoters (Butler and Kadonaga 2001; Juven-Gershon et al. 2008b). Enhancer–core promoter specificity has also been observed with a developmental core promoter (with TATA, Inr, MTE, and DPE motifs) versus a housekeeping core promoter (with the TCT motif) (Zabidi et al. 2015). (Adapted from Butler and Kadonaga 2001.)

that the DPE is an important component of the Hox gene network and that Caudal, an important regulator of this network, can function as a DPE-specific activator. Moreover, the ability of the BRE[u] to suppress Caudal activation of a TATA box-containing promoter suggests a role of TFIIB and the BRE[u] in the regulation of the activity of sequence-specific transcription factors.

The DPE motif is also overrepresented in the core promoters of *Drosophila* genes that are regulated by Dorsal, a sequence-specific transcription factor that is a member of the NF-κB family of proteins (Zehavi et al. 2014). The DPE is essential for Dorsal-mediated activation of many genes that control dorsal–ventral patterning. In addition, in some promoter contexts, Dorsal preferentially activates transcription via the DPE relative to the TATA box.

How might transcription factors activate transcription preferentially via the DPE relative to the TATA box? It is known, for instance, that NC2 (negative cofactor 2; also known as Dr1–Drap1) as well as the Mot1 ATPase repress TATA-dependent transcription and activate DPE-dependent transcription (Willy et al. 2000; Hsu et al. 2008). It is therefore possible that DPE-specific activators can recruit factors such as NC2 and/or Mot1 to the core promoter and thus promote DPE-dependent transcription relative to TATA-dependent transcription. However, the mechanisms of core promoter motif-specific activation remain to be determined.

Last, it is relevant to mention that specificity for core promoter motifs applies to not only distant transcriptional enhancers but also promoter-proximal activator binding sites. For instance, an activating region that is 60 base pairs (bp) upstream of the mouse terminal deoxynucleotidyltransferase gene promoter exhibits a preference for the Inr relative to the TATA box (Garraway et al. 1996).

TRFs

TBP and TFIIB are present in Archaea and eukaryotes. Prior to the evolution of eukaryotes, it is likely that the mechanism of transcription involved the binding of TBP to the TATA box and the subsequent assembly of TFIIB, the RNA polymerase, and other factors into the PIC (for example, see Blombach et al. 2016). In this manner, the central role of TBP in the transcription process would have been established.

In bilateria, three additional TRFs have been identified (for example, see Goodrich and Tjian 2010; Akhtar and Veenstra 2011). These factors possess many of the key features of TBP, such as sites of interaction with TFIIB and TFIIA, and therefore have much of the transcriptional potency of TBP. We refer to TBP and the TRFs as "system factors" (Duttke et al. 2014; Duttke 2015). TBP and the TRFs regulate gene expression via the basal transcription process. Some of the properties of the TRFs are discussed next.

TRF1

TRF1 (also known as TRF) was the first TRF to be identified (Crowley et al. 1993). TRF1 has been found only in insects. It can bind to the TATA box along with TFIIA and TFIIB and substitute for TBP in the transcription of some Pol II promoters in vitro (Hansen et al. 1997; Holmes and Tjian 2000). Moreover, TRF1 associates with BRF1 (an RNA Pol III transcription factor) and mediates tRNA gene transcription from Pol III promoters (Takada et al. 2000; Isogai et al. 2007a; Verma et al. 2013). Thus, TRF1 participates in both Pol II and Pol III transcription. It is also interesting to note that the emergence of TRF1 did not appear to add any new transcriptional functions but rather resulted in the subdivision and/or sharing of the pre-existing functions between TBP and TRF1.

TRF2

TRF2 (also known as TBPL1, TLP, TRP, and TLF) is present in bilateria (Duttke et al. 2014). Unlike TBP and the other TRFs, TRF2 does not bind to the TATA box and does not appear to possess any sequence-specific DNA-binding activity (Dantonel et al. 1999; Rabenstein et al. 1999; Wang et al. 2014). It does, however, interact with TFIIA and TFIIB (Rabenstein et al. 1999; Teichmann et al. 1999).

In *Drosophila*, TRF2 is involved in several different transcriptional programs. First, TRF2 associates with DREF (DNA replication-related element-binding factor) and activates transcription via the binding of DREF to DRE (DNA replication-related element) motifs in promoters (Hochheimer et al. 2002). Second and third, by DRE-independent processes, TRF2, but not TBP, is required for transcription from TCT-dependent as well as DPE-dependent core promoters (Hsu et al. 2008; Kedmi et al. 2014; Wang et al. 2014). The DPE functions with the Inr but not with the TCT element (Parry et al. 2010); hence, TRF2-driven transcription via the TCT motif probably occurs by a different mechanism than TRF2-mediated transcription via the DPE. Fourth, TRF2, but not TBP, is required for transcription of the histone H1 promoter by a process that does not appear to involve the DRE, TCT, or DPE (Isogai et al. 2007b). The partitioning of the transcriptional functions of TBP, TRF1, and TRF2 in *Drosophila* is depicted in Figure 4.

The majority of the TRF2-dependent promoters in *Drosophila* lacks a TATA box (for example, see Isogai et al. 2007b; Wang et al. 2014). Given that TRF2 does not bind to the TATA box, these findings suggest that a key early function of TRF2 may have been to mediate TATA-less transcription (for instance, see Duttke et al. 2014). It thus appears that, in contrast to the situation with TBP and TRF1 (see above), the combination of TBP and TRF2 has resulted in an expansion of the range of transcriptional mechanisms relative to those used by TBP alone. This increase in the number of transcriptional programs led to the suggestion that the emergence of TRF2 facilitated the evolution of the bilateria (Duttke et al. 2014).

The loss of TRF2 is embryonic lethal in *Drosophila* (Kopytova et al. 2006), *Caenorhabditis elegans* (Dantonel et al. 2000; Kaltenbach et al. 2000), zebrafish (Müller et al. 2001), and *Xenopus* (Veenstra et al. 2000). In mice, however, TRF2 is not essential but is required for spermiogenesis
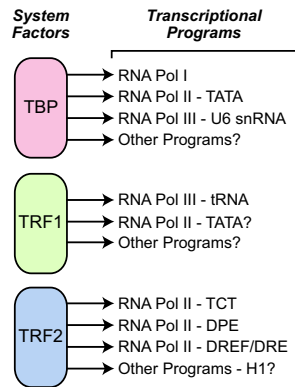
Figure 4. Transcriptional programs that are directed by TBP, TRF1, and TRF2 in *Drosophila*. This diagram shows the partitioning of transcriptional functions between TBP, TRF1, and TRF2 in *Drosophila*. It appears that each of these system factors is responsible for a set of transcriptional programs. As discussed in the text, humans lack TRF1 and contain TBP, TRF2, and TRF3. Moreover, in humans, the specific functions of factors such as TRF2 remain to be clarified. (Adapted from Duttke et al. 2014.)

(Martianov et al. 2001; Zhang et al. 2001; Zhou et al. 2013). The viability of TRF2-deficient mice could be due to the presence of a functionally analogous protein that can compensate for the absence of TRF2. It is also possible that the role of TRF2 changed substantially between frogs and mice. For instance, TBP or some other factor might have usurped the transcriptional function of TRF2 at most promoters and thus rendered TRF2 dispensable.

TRF3

TRF3 (also known as TBPL2 and TBP2) is found in vertebrates and is the TRF that is most closely related to TBP (Persengiev et al. 2003). It can bind to the TATA box, interact with TFIIA and TFIIB, and mediate Pol II transcription in vitro (Bártfai et al. 2004; Jallow et al. 2004). TRF3 is present in a variety of mouse and human cell lines and tissues (Persengiev et al. 2003) but has particularly high expression in the testes and ovaries in zebrafish (Bártfai et al. 2004) and *Xenopus* (Xiao et al. 2006) and in the ovaries in mice (Xiao et al. 2006; Gazdag et al. 2007). TRF3 is required for normal embryonic development in zebrafish (Bártfai et al. 2004; Hart et al. 2007) and *Xenopus* (Jallow et al. 2004). In mice, however, the loss of TRF3 has no apparent phenotype except for female sterility, which is due to the requirement of TRF3 for the differentiation of female germ cells (Gazdag et al. 2009).

In zebrafish, TRF3 interacts with TAF3 and is essential for the expression of the *mespa* gene, which is required for hematopoiesis (Hart et al. 2007, 2009). Moreover, as seen with TRF3, the depletion of TAF3 also results in the failure to undergo hematopoiesis. In mouse cells, a complex that contains TRF3 and TAF3 was found to be involved in muscle cell differentiation (Deato and Tjian 2007; Deato et al. 2008). However, further studies, which in-

cluded the analysis of TRF3 knockout mice, suggest that TBP, and not TRF3, remains active during muscle cell differentiation despite the rapid and dramatic loss of TBP protein in myotubes (Gazdag et al. 2009; Li et al. 2015; Malecova et al. 2017). It is possible that some of the reported differences could be due to the presence of an unknown factor that bypasses the need for TRF3 in muscle differentiation and acts in the organism but not in cells in culture.

The core promoter is unidirectional

In mammals, promoter regions frequently exhibit divergent transcription, with noncoding reverse direction transcription that initiates upstream of the forward direction TSS (Core et al. 2008; Preker et al. 2008; Seila et al. 2008; Scruggs et al. 2015). Further analysis of this phenomenon led to a simple model in which core promoters are unidirectional, and divergent promoter regions consist of forward and reverse direction core promoters (Fig. 5; Duttke et al. 2015a; see also Andersson et al. 2015; Duttke et al. 2015b). The two opposing core promoters flank a central nucleosome-free region with binding sites for sequence-specific transcription factors. More generally, however, it is important to note that the analysis of the directionality of any particular promoter region should include the positions and orientations of all of the transcriptional elements, which include not only the core promoter motifs but also the binding sites for sequence-specific factors (for example, see O'Shea-Greenfield and Smale 1992). In the future, it will be interesting to elucidate the biological functions of divergent transcription, such as a possible role in facilitating the evolution of new genes (Wu and Sharp 2013). In addition, transcription at one locus can increase transcription at a nearby locus (for instance, see Engreitz et al. 2017); hence, reverse direction transcription might enhance the level of forward direction transcription.

An expanded view of the core promoter

Traditionally, the core promoter has been thought to comprise the TATA box, Inr, and other DNA sequence motifs that direct the assembly of the basal transcription machinery (i.e., Pol II, TFIID, TFIIB, and other auxiliary factors) into the PIC. However, it is now useful to expand
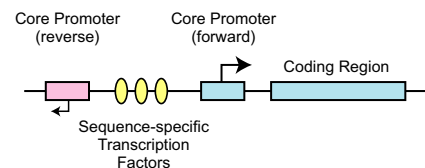

Figure 5. A model for divergent transcription. In this model, a promoter region that exhibits divergent transcription contains a unidirectional forward core promoter and a unidirectional reverse core promoter that flank binding sites for sequence-specific transcription factors.

our perspective of the core promoter. Specifically, we could view the core promoter as a multidimensional element with some of the following components.

Role of sequence-specific DNA-binding transcription factors

Although it is well established that DNA recognition sites for the basal transcription machinery (e.g., TATA box, Inr, and DPE) are important core promoter elements, it is also likely that binding sites for sequence-specific transcription factors (SSTFs) such as Sp1 can direct transcription initiation in conjunction with a motif such as the Inr. For example, a synthetic promoter that comprises a cluster of Sp1-binding sites and an Inr exhibits transcriptional activity that is similar to that of a TATA + Inr core promoter (Smale et al. 1990; Emami et al. 1995). It is thus reasonable to postulate that a SSTF recognition site (or sites) in the immediate upstream promoter region (~50–80 nt upstream of the TSS) could function with an Inr in lieu of a TATA box (Fig. 6A). Given the presence of the Inr or Inr-like sequences in over half of human focused promoters (Vo ngoc et al. 2017), SSTF-binding site + Inr promoters may be widely used in mammals. It will therefore be important to examine this mechanism of transcription, particularly in the context of natural promoter regions.

Chromatin signals and structure

Transcription occurs in the context of chromatin, and hence variations in the structure and composition of chromatin have considerable potential to influence the events that lead to transcription initiation. Multiple new lines of evidence are revealing intriguing connections between chromatin and transcription initiation (Fig. 6B). These findings suggest that chromatin signals and structure are components of an expanded version of the core promoter. Some examples are as follows.

First, there is an interesting connection between the TAF3 subunit of the TFIID complex and trimethylated histone H3K4 (H3K4me3), which is commonly found in the region immediately downstream from active promoters (Vermeulen et al. 2007; Lauberth et al. 2013). TAF3 binds to H3K4me3, and this interaction facilitates the assembly of the PIC. Thus, the TAF3–H3K4me3 interaction provides a means of recruiting TFIID to the core promoter region. In humans, H3K4me3 is present in only ~0.1% of the total histone H3 species (Young et al. 2009). Hence, H3K4 trimethylation can potentially add considerable specificity to the recruitment of TFIID to active promoters. In addition, the interaction of TFIID to promoters could be augmented by histone acetylation, as the TAF1 subunit of TFIID contains a double bromodomain that can bind to diacetylated histone H4 (Jacobson et al. 2000).

Second, histone methylation and DNA methylation have important roles in the promoter recruitment of RNA Pol IV and Pol V, which are specialized variants of Pol II in plants (for review, see Haag and Pikaard 2011). Specifically, Pol IV is recruited to promoters containing methylated H3K9 via SHH1, a Pol IV-interacting protein that binds to unmethylated H3K4 and methylated H3K9 (Law et al. 2013; Zhang et al. 2013). Pol V is recruited to promoters containing methylated DNA via factors (DRD1 subunit of the DDR complex, SUVH2, SUVH9) that bind to the polymerase as well as to different forms of methylated DNA (Johnson et al. 2014; Liu et al. 2014). These findings show that histone methylation as well as DNA methylation can serve as chromatin-based signals for the recruitment of RNA polymerases. The positive effect of DNA methylation on Pol V transcription is in contrast to the repressive effect of CpG methylation in vertebrates. Nevertheless, it is possible that DNA modifications such as methylation or hydroxymethylation could be used to recruit Pol II to promoters in animals.

Third, the prenucleosome, a stable conformational isomer of the nucleosome that associates with ~80-bp DNA, appears to be present in the "nucleosome-depleted region" (also known as "nucleosome-free region") that is located immediately upstream of the TSSs of active promoters (Fei et al. 2015; Khuong et al. 2015). In the examination of the yeast *PHO5* promoter in vivo, prenucleosome-like particles were observed at active promoters, whereas nucleosome-like particles were seen at repressed promoters (Brown et al. 2013; Fei et al. 2015). In addition,
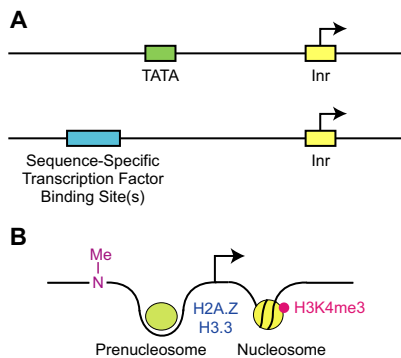

Figure 6. Potential functions of sequence-specific DNA-binding transcription factors and chromatin signals at the core promoter. (*A*) Postulated role of SSTFs in core promoter function. In this model, SSTF-binding sites in the immediate upstream promoter region (~50–80 bp upstream of the TSS) function in a manner that is analogous to a TATA box. Thus, the combination of an SSTF-binding site and an Inr could act as a core promoter. (*B*) A composite of the potential role of chromatin signals and structure in core promoter function. It may be necessary to analyze the core promoter in the broader context of chromatin. Examples discussed in the text include the following. H3K4me3 has been found to recruit TFIID via its TAF3 subunit (Vermeulen et al. 2007; Lauberth et al. 2013). Salt-labile nucleosomes containing the histone variants H2A.Z and H3.3 have been found at active chromatin (Jin et al. 2009). The prenucleosome, a conformational isomer of the nucleosome that interacts with ~80-bp DNA, appears to be present in the immediate upstream region of active promoters (Fei et al. 2015; Khuong et al. 2015). In plants, RNA Pol V is recruited to promoters via methylated DNA (Johnson et al. 2014; Liu et al. 2014). Although CpG methylation is generally repressive in vertebrates, DNA modifications such as methylation or hydroxymethylation may also function as a positive signal for the initiation of Pol II transcription.

methidiumpropyl-EDTA sequencing (MPE-seq) analysis with mouse embryonic stem cells revealed prenucleosome-like particles (i.e., histone-containing particles associated with ~61- to 100-bp DNA) in the immediate upstream region of active promoters but not inactive promoters (Ishii et al. 2015; Khuong et al. 2015). These findings suggest that prenucleosomes or prenucleosome-like particles are present in the nucleosome-depleted region of active promoters. Moreover, histone H3K56 can be acetylated by p300 in prenucleosomes but not in nucleosomes (Fei et al. 2015). It remains to be determined, however, whether prenucleosomes participate in the transcription process. Notwithstanding, the association of prenucleosomes with only ~80-bp DNA suggests that they might be more permissive to transcription than canonical nucleosomes.

Fourth, histone variants might also participate in core promoter function. For example, H2A.Z- and H3.3-containing nucleosomes have been found at sites of active chromatin, such as promoters (Jin et al. 2009). Histones H2A.Z and H3.3 may destabilize nucleosomes (for example, see Jin and Felsenfeld 2007) and thus facilitate transcription. In considering the presence of the human histone variants at promoters, it may be useful to note that H3.3 constitutes ~10% of the total histone H3 species and that H2A.Z is ~1%–3% of the total H2A species (Dang et al. 2016). Hence, the presence of H3.3 and H2A.Z at promoters could provide some specificity to core promoter function, but their roles, if any, in the initiation of transcription remain to be determined.

Thus, different aspects of the chromatin context are likely to be critical components of core promoter function. In such cases, however, it would be essential to understand the sources of the chromatin signals or structures that influence transcription.

Properties of DNA

Last, it seems likely that structural properties of DNA contribute to core promoter activity. For instance, the flexibility and curvature of DNA could facilitate interactions between transcription factors, and a decrease in the helical stability could increase the ability of Pol II to initiate transcription. However, an underlying DNA structure "code" for core promoters has yet to be determined. It is nevertheless interesting to note the general absence of core promoter DNA sequence motifs between the TATA box and Inr as well as between the Inr and MTE (Fig. 2). These regions may lack core promoter sequence elements but probably have a DNA structure that facilitates the transcription process.

Summary and perspectives

The core promoter is a rich and complex regulatory element. It is diverse in terms of its composition as well as its function. The core promoter is also punctilious: It acts unidirectionally with strict rules and precision. For instance, the change of a T nucleotide to an A can change

a TRF2-driven TCT-dependent core promoter to a TBP-driven Inr-dependent core promoter. Moreover, specific core promoter elements can be associated with biological networks. The DPE is present in nearly all of the Hox gene promoters in *Drosophila*, and the TCT motif is present in nearly all of the ribosomal protein gene promoters in *Drosophila* and humans. In addition, some transcriptional enhancers exhibit a strong preference for specific core promoter elements.

We also described an expanded view of the core promoter that comprises the classical DNA sequence motifs (such as the TATA box, Inr, and DPE) along with promoter-proximal SSTF-binding sites, chromatin signals, and DNA structure. Each of these components might be important to varying degrees at any particular core promoter. Even though it appears to add complexity to our definition of the core promoter, the expanded model may result in a more unified and coherent conceptual understanding of the core promoter.

A few decades ago, with the discoveries of the TATA box and Inr, it seemed like we had a good understanding of the core promoter. We have since found, however, that the core promoter is a complex multidimensional regulatory element. We hope that, in the future, we might once again at least have the impression that we understand the punctilious RNA Pol II core promoter.

References

Akhtar W, Veenstra GJ. 2011. TBP-related factors: a paradigm of diversity in transcription initiation. *Cell Biosci* **1:** 23.

Andersson R, Chen Y, Core L, Lis JT, Sandelin A, Jensen TH. 2015. Human gene promoters are intrinsically bidirectional. *Mol Cell* **60:** 346–347.

Anish R, Hossain MB, Jacobson RH, Takada S. 2009. Characterization of transcription from TATA-less promoters: identification of a new core promoter element XCPE2 and analysis of factor requirements. *PLoS One* **4:** e5103.

Bártfai R, Balduf C, Hilton T, Rathmann Y, Hadzhiev Y, Tora L, Orbán L, Müller F. 2004. TBP2, a vertebrate-specific member of the TBP family, is required in embryonic development of zebrafish. *Curr Biol* **14:** 593–598.

Blombach F, Smollett KL, Grohmann D, Werner F. 2016. Molecular mechanisms of transcription initiation—structure, function, and evolution of TFE/TFIIE-like factors and open complex formation. *J Mol Biol* **428:** 2592–2606.

Brown CR, Mao C, Falkovskaia E, Jurica MS, Boeger H. 2013. Linking stochastic fluctuations in chromatin structure and gene expression. *PLoS Biol* **11:** e1001621.

Bucher P. 1990. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J Mol Biol* **212:** 563–578.

Burke TW, Kadonaga JT. 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.

Burke TW, Kadonaga JT. 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.

Butler JE, Kadonaga JT. 2001. Enhancer–promoter specificity mediated by DPE or TATA core promoter motifs. *Genes Dev* **15:** 2515–2519.

Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CAM, Taylor MS, Engström PG, Frith MC, et al. 2006. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* **38:** 626–635.

Chalkley GE, Verrijzer CP. 1999. DNA binding site selection by RNA polymerase II TAFs: a TAF$_{II}$250–TAF$_{II}$150 complex recognizes the initiator. *EMBO J* **18:** 4835–4485.

Cianfrocco MA, Kassavetis GA, Grob P, Fang J, Juven-Gershon T, Kadonaga JT, Nogales E. 2013. Human TFIID binds to core promoter DNA in a reorganized structural state. *Cell* **152:** 120–131.

Corden J, Wasylyk B, Buchwalder A, Sassone-Corsi P, Kedinger C, Chambon P. 1980. Promoter sequences of eukaryotic protein-coding genes. *Science* **209:** 1406–1414.

Core LJ, Waterfall JJ, Lis JT. 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322:** 1845–1848.

Core LJ, Martins AL, Danko CG, Waters CT, Siepel A, Lis JT. 2014. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet* **46:** 1311–1320.

Crowley TE, Hoey T, Liu JK, Jan YN, Jan LY, Tjian R. 1993. A new factor related to TATA-binding protein has highly restricted expression patterns in *Drosophila*. *Nature* **361:** 557–561.

Dang X, Singh A, Spetman BD, Nolan KD, Isaacs JS, Dennis JH, Dalton S, Marshall AG, Young NL. 2016. Label-free relative quantitation of isobaric and isomeric human histone H2A and H2B variants by Fourier transform ion cyclotron resonance top-down MS/MS. *J Proteome Res* **15:** 3196–3203.

Danino YM, Even D, Ideses D, Juven-Gershon T. 2015. The core promoter: at the heart of gene expression. *Biochim Biophys Acta* **1849:** 1116–1131.

Dantonel JC, Wurtz JM, Poch O, Moras D, Tora L. 1999. The TBP-like factor: an alternative transcription factor in Metazoa? *Trends Biochem Sci* **24:** 335–339.

Dantonel JC, Quintin S, Lakatos L, Labouesse M, Tora L. 2000. TBP-like factor is required for embryonic RNA polymerase II transcription in *C. elegans*. *Mol Cell* **6:** 715–722.

Deato MDE, Tjian R. 2007. Switching of the core transcription machinery during myogenesis. *Genes Dev* **21:** 2137–2149.

Deato MDE, Marr MT II, Sottero T, Inouye C, Hu P, Tjian R. 2008. MyoD targets TAF3/TRF3 to activate myogenin transcription. *Mol Cell* **32:** 96–105.

Deaton AM, Bird A. 2011. CpG islands and the regulation of transcription. *Genes Dev* **25:** 1010–1022.

Deng W, Roberts SG. 2005. A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes Dev* **19:** 2418–2423.

Duttke SHC. 2014. RNA polymerase III accurately initiates transcription from RNA polymerase II promoters in vitro. *J Biol Chem* **289:** 20396–20404.

Duttke SHC. 2015. Evolution and diversification of the basal transcription machinery. *Trends Biochem Sci* **40:** 127–129.

Duttke SHC, Doolittle RF, Wang YL, Kadonaga JT. 2014. TRF2 and the evolution of the bilateria. *Genes Dev* **28:** 2071–2076.

Duttke SHC, Lacadie SA, Ibrahim MM, Glass CK, Corcoran DL, Benner C, Heinz S, Kadonaga JT, Ohler U. 2015a. Human promoters are intrinsically directional. *Mol Cell* **57:** 674–684.

Duttke SHC, Lacadie SA, Ibrahim MM, Glass CK, Corcoran DL, Benner C, Heinz S, Kadonaga JT, Ohler U. 2015b. Perspectives on unidirectional versus divergent transcription. *Mol Cell* **60:** 348–349.

Emami KH, Navarre WW, Smale ST. 1995. Core promoter specificities of the Sp1 and VP16 transcriptional activation domains. *Mol Cell Biol* **15:** 5906–5916.

Emami KH, Jain A, Smale ST. 1997. Mechanism of synergy between TATA and initiator: synergistic binding of TFIID following a putative TFIIA-induced isomerization. *Genes Dev* **11:** 3007–3019.

Engreitz JM, Haines JE, Perez EM, Munson G, Chen J, Kane M, McDonel PE, Guttman M, Lander ES. 2017. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539:** 452–455.

Evans R, Fairley JA, Roberts SG. 2001. Activator-mediated disruption of sequence-specific DNA contacts by the general transcription factor TFIIB. *Genes Dev* **15:** 2945–2949.

Fei J, Torigoe SE, Brown CR, Khuong MT, Kassavetis GA, Boeger H, Kadonaga JT. 2015. The prenucleosome, a stable conformational isomer of the nucleosome. *Genes Dev* **29:** 2563–2575.

FitzGerald PC, Sturgill D, Shyakhtenko A, Oliver B, Vinson C. 2006. Comparative genomics of *Drosophila* and human core promoters. *Genome Biol* **7:** R53.

Garraway IP, Semple K, Smale ST. 1996. Transcription of the lymphocyte-specific terminal deoxynucleotidyltransferase gene requires a specific core promoter structure. *Proc Natl Acad Sci* **93:** 4336–4341.

Gazdag E, Rajkovic A, Torres-Padilla ME, Tora L. 2007. Analysis of TATA-binding protein 2 (TBP2) and TBP expression suggests different roles for the two proteins in regulation of gene expression during oogenesis and early mouse development. *Reproduction* **134:** 51–62.

Gazdag E, Santenard A, Ziegler-Birling C, Altobelli G, Poch O, Tora L, Torres-Padilla ME. 2009. TBP2 is essential for germ cell development by regulating transcription and chromatin condensation in the oocyte. *Genes Dev* **23:** 2210–2223.

Goldberg ML. 1979. "Sequence analysis of Drosophila histone genes." Ph.D. thesis, Stanford University, Stanford, CA.

Goodrich JA, Tjian R. 2010. Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. *Nat Rev Genet* **11:** 549–558.

Haag JR, Pikaard CS. 2011. Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nat Rev Mol Cell Biol* **12:** 483–492.

Hansen SK, Takada S, Jacobson RH, Lis JT, Tjian R. 1997. Transcription properties of a cell type-specific TATA-binding protein, TRF. *Cell* **91:** 71–83.

Hart DO, Raha T, Lawson ND, Green MR. 2007. Initiation of zebrafish haematopoiesis by the TATA-box-binding protein-related factor Trf3. *Nature* **450:** 1082–1085.

Hart DO, Santra MK, Raha T, Green MR. 2009. Selective interaction between Trf3 and Taf3 required for early development and hematopoiesis. *Dev Dyn* **238:** 2540–2549.

Hochheimer A, Zhou S, Zheng S, Holmes MC, Tjian R. 2002. TRF2 associates with DREF and directs promoter-selective gene expression in *Drosophila*. *Nature* **420:** 439–445.

Holmes MC, Tjian R. 2000. Promoter-selective properties of the TBP-related factor TRF1. *Science* **288:** 867–870.

Hoskins RA, Landolin JM, Brown JB, Sandler JE, Takahashi H, Lassmann T, Yu C, Booth BW, Zhang D, Wan KH, et al. 2011. Genome-wide analysis of promoter architecture in *Drosophila melanogaster*. *Genome Res* **21:** 182–192.

Hsu JY, Juven-Gershon T, Marr MT II, Wright KJ, Tjian R, Kadonaga JT. 2008. TBP, Mot1, and NC2 establish a regulatory circuit that controls DPE-dependent versus TATA-dependent transcription. *Genes Dev* **22:** 2353–2358.

Ishii H, Kadonaga JT, Ren B. 2015. MPE-seq, a new method for the genome-wide analysis of chromatin structure. *Proc Natl Acad Sci* **112:** E3457–E3465.

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

Isogai Y, Keleş S, Prestel M, Hochheimer A, Tjian R. 2007b. Transcription of histone gene cluster by differential core-promoter factors. *Genes Dev* **21:** 2936–2949.

Jacobson RH, Ladurner AG, King DS, Tjian R. 2000. Structure and function of a human TAF$_{II}$250 double bromodomain module. *Science* **288:** 1422–1425.

Jallow Z, Jacobi UG, Weeks DL, Dawid IB, Veenstra GJ. 2004. Specialized and redundant roles of TBP and a vertebrate-specific TBP paralog in embryonic gene regulation in *Xenopus*. *Proc Natl Acad Sci* **101:** 13525–13530.

Javahery R, Khachi A, Lo K, Zenzie-Gregory B, Smale ST. 1994. DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Mol Cell Biol* **14:** 116–127.

Jin C, Felsenfeld G. 2007. Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes Dev* **21:** 1519–1529.

Jin C, Zang C, Wei G, Cui K, Peng W, Zhao K, Felsenfeld G. 2009. H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nat Genet* **41:** 941–945.

Johnson LM, Du J, Hale CJ, Bischof S, Feng S, Chodavarapu RK, Zhong X, Marson G, Pellegrini M, Segal DJ, et al. 2014. SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. *Nature* **507:** 124–128.

Juven-Gershon T, Cheng S, Kadonaga JT. 2006. Rational design of a super core promoter that enhances gene expression. *Nat Methods* **3:** 917–922.

Juven-Gershon T, Hsu JY, Theisen JW, Kadonaga JT. 2008a. The RNA polymerase II core promoter—the gateway to transcription. *Curr Opin Cell Biol* **20:** 253–259.

Juven-Gershon T, Hsu JY, Kadonaga JT. 2008b. Caudal, a key developmental regulator, is a DPE-specific transcription factor. *Genes Dev* **22:** 2823–2830.

Kadonaga JT. 1990. Assembly and disassembly of the *Drosophila* RNA polymerase II complex during transcription. *J Biol Chem* **265:** 2624–2631.

Kadonaga JT. 2012. Perspectives on the RNA polymerase II core promoter. *Wiley Interdiscip Rev Dev Biol* **1:** 40–51.

Kaltenbach L, Horner MA, Rothman JH, Mango SE. 2000. The TBP-like factor CeTLF is required to activate RNA polymerase II transcription during *C. elegans* embryogenesis. *Mol Cell* **6:** 705–713.

Kedmi A, Zehavi Y, Glick Y, Orenstein Y, Ideses D, Wachtel C, Doniger T, Ben-Asher HW, Muster N, Thompson J, et al. 2014. *Drosophila* TRF2 is a preferential core promoter regulator. *Genes Dev* **28:** 2163–2174.

Khuong MT, Fei J, Ishii H, Kadonaga JT. 2015. Prenucleosomes and active chromatin. *Cold Spring Harbor Symp Quant Biol* **80:** 65–72.

Kopytova DV, Krasnov AN, Kopantceva MR, Nabirochkina EN, Nikolenko JV, Maksimenko O, Kurshakova MM, Lebedeva LA, Yerokhin MM, Simonova OB, 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.

Kruesi WS, Core LJ, Waters CT, Lis JT, Meyer BJ. 2013. Condensin controls recruitment of RNA polymerase II to achieve nematode X-chromosome dosage compensation. *Elife* **2:** e00808.

Kutach AK. 2000. "The DPE is a core promoter element that is widely used in Drosophila promoters." Ph.D. thesis, University of California at San Diego, La Jolla, CA.

Kutach AK, Kadonaga JT. 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.

Lagrange T, Kapanidis AN, Tang H, Reinberg D, Ebright RH. 1998. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev* **12:** 34–44.

Lai WKM, Pugh BF. 2017. Genome-wide uniformity of human 'open' pre-initiation complexes. *Genome Res* **27:** 15–26.

Lam MTY, Cho H, Lesch HP, Gosselin D, Heinz S, Tanaka-Oishi Y, Benner C, Kaikkonen MU, Kim AS, Kosaka M, et al. 2013. Rev–Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature* **498:** 511–515.

Lauberth SM, Nakayama T, Wu X, Ferris AL, Tang Z, Hughes SH, Roeder RG. 2013. H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell* **152:** 1021–1036.

Law JA, Du J, Hale CJ, Feng S, Krajewski K, Palanca AM, Strahl BD, Patel DJ, Jacobsen SE. 2013. Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature* **498:** 385–389.

Lee MP, Howcroft K, Kotekar A, Yang HH, Buetow KH, Singer DS. 2005a. ATG deserts define a novel promoter subclass. *Genome Res* **15:** 1189–1197.

Lee DH, Gershenzon N, Gupta M, Ioshikhes IP, Reinberg D, Lewis BA. 2005b. Functional characterization of core promoter elements: the downstream core element is recognized by TAF1. *Mol Cell Biol* **25:** 9674–9686.

Lee N, Iyer SS, Mu J, Weissman JD, Ohali A, Howcroft TK, Lewis BA, Singer DS. 2010. Three novel downstream promoter elements regulate MHC class I promoter activity in mammalian cells. *PLoS One* **5:** e15278.

Lenhard B, Sandelin A, Carninci P. 2012. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat Rev Genet* **13:** 233–245.

Lewis BA, Kim TK, Orkin SH. 2000. A downstream element in the human β-globin promoter: evidence of extended sequence-specific transcription factor IID contacts. *Proc Natl Acad Sci* **97:** 7172–7177.

Lewis BA, Sims RJ III, Lane WS, Reinberg D. 2005. Functional characterization of core promoter elements: DPE-specific transcription requires the protein kinase CK2 and the PC4 coactivator. *Mol Cell* **18:** 471–481.

Li L, Martinez SS, Hu W, Liu Z, Tjian R. 2015. A specific E3 ligase/deubiquitinase pair modulates TBP protein levels during muscle differentiation. *Elife* **4:** e08536.

Lim CY, Santoso B, Boulay T, Dong E, Ohler U, Kadonaga JT. 2004. The MTE, a new core promoter element for transcription by RNA polymerase II. *Genes Dev* **18:** 1606–1617.

Liu ZW, Shao CR, Zhang CJ, Zhou JX, Zhang SW, Li L, Chen S, Huang HW, Cai T, He XJ. 2014. The SET domain proteins SUVH2 and SUVH9 are required for Pol V occupancy at RNA-directed DNA methylation loci. *PLoS Genet* **10:** e1003948.

Lo K, Smale ST. 1996. Generality of a functional initiator consensus sequence. *Gene* **182:** 13–22.

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

Main BJ, Smith AD, Jang H, Nuzhdin SV. 2013. Transcription start site evolution in *Drosophila*. *Mol Biol Evol* **30:** 1966–1974.

Malecova B, Dall'Agnese A, Madaro L, Gatto S, Coutinho Toto P, Albini S, Ryan T, Tora L, Puri PL. 2017. TBP/TFIID-dependent activation of MyoD target genes in skeletal muscle cells. *Elife* **5:** e12534.

Marbach-Bar N, Bahat A, Ashkenazi S, Golan-Mashiach M, Haimov O, Wu SY, Chiang CM, Puzio-Kuter A, Hirshfield KM, Levine AJ, et al. 2016. DTIE, a novel core promoter element that directs start site selection in TATA-less genes. *Nucleic Acids Res* **44:** 1080–1094.

Martianov I, Fimia GM, Dierich A, Parvinen M, Sassone-Corsi P, Davidson I. 2001. Late arrest of spermiogenesis and germ cell apoptosis in mice lacking the TBP-like TLF/TRF2 gene. *Mol Cell* **7:** 509–515.

Müller F, Lakatos L, Dantonel J, Strähle U, Tora L. 2001. TBP is not universally required for zygotic RNA polymerase II transcription in zebrafish. *Curr Biol* **11:** 282–287.

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

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

Ohtsuki S, Levine M, Cai HN. 1998. Different core promoters possess distinct regulatory activities in the *Drosophila* embryo. *Genes Dev* **12:** 547–556.

O'Shea-Greenfield A, Smale ST. 1992. Roles of TATA and initiator elements in determining the start site location and direction of RNA polymerase II transcription. *J Biol Chem* **267:** 1391–1402.

Parry TJ, Theisen JWM, Hsu JY, Wang YL, Corcoran DL, Eustice M, Ohler U, Kadonaga JT. 2010. The TCT motif, a key component of an RNA polymerase II transcription system for the translational machinery. *Genes Dev* **24:** 2013–2018.

Patikoglou GA, Kim JL, Sun L, Yang SH, Kodadek T, Burley SK. 1999. TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes Dev* **13:** 3217–3230.

Perry RP. 2005. The architecture of mammalian ribosomal protein promoters. *BMC Evol Biol* **5:** 15.

Persengiev SP, Zhu X, Dixit BL, Maston GA, Kittler EL, Green MR. 2003. TRF3, a TATA-box-binding protein-related factor, is vertebrate-specific and widely expressed. *Proc Natl Acad Sci* **100:** 14887–14891.

Preker P, Nielsen J, Kammler S, Lykke-Andersen S, Christensen MS, Mapendano CK, Schierup MH, Jensen TH. 2008. RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322:** 1851–1854.

Purnell BA, Emanuel PA, Gilmour DS. 1994. TFIID sequence recognition of the initiator and sequences farther downstream in *Drosophila* class II genes. *Genes Dev* **8:** 830–842.

Rabenstein MD, Zhou S, Lis JT, Tjian R. 1999. TATA box-binding protein (TBP)-related factor 2 (TRF2), a third member of the TBP family. *Proc Natl Acad Sci* **96:** 4791–4796.

Roy AL, Singer DS. 2015. Core promoters in transcription: old problem, new insights. *Trends Biochem Sci* **40:** 165–171.

Sainsbury S, Bernecky C, Cramer P. 2015. Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Biol* **16:** 129–143.

Schor IE, Degner JF, Harnett D, Cannavò E, Casale FP, Shim H, Garfield DA, Birney E, Stephens M, Stegle O, et al. 2017. Promoter shape varies across populations and affects promoter evolution and expression noise. *Nat Genet* **49:** 550–558.

Schübeler D. 2015. Function and information content of DNA methylation. *Nature* **517:** 321–326.

Scruggs BS, Gilchrist DA, Nechaev S, Muse GW, Burkholder A, Fargo DC, Adelman K. 2015. Bidirectional transcription arises from two distinct hubs of transcription factor binding and active chromatin. *Mol Cell* **58:** 1101–1102.

Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, Young RA, Sharp PA. 2008. Divergent transcription from active promoters. *Science* **322:** 1849–1851.

Shao H, Revach M, Moshonov S, Tzuman Y, Gazit K, Albeck S, Unger T, Dikstein R. 2005. Core promoter binding by histone-like TAF complexes. *Mol Cell Biol* **25:** 206–219.

Shir-Shapira H, Sharabany J, Filderman M, Ideses D, Ovadia-Shochat A, Mannervik M, Juven-Gershon T. 2015. Structure-function analysis of the *Drosophila melanogaster* Caudal transcription factor provides insights into core promoter-preferential activation. *J Biol Chem* **290:** 17293–17305.

Smale ST, Baltimore D. 1989. The 'initiator' as a transcription control element. *Cell* **57:** 103–113.

Smale ST, Kadonaga JT. 2003. The RNA polymerase II core promoter. *Annu Rev Biochem* **72:** 449–479.

Smale ST, Schmidt MC, Berk AJ, Baltimore D. 1990. Transcriptional activation by Sp1 as directed through TATA or initiator: specific requirement for mammalian transcription factor IID. *Proc Natl Acad Sci* **87:** 4509–4513.

Takada S, Lis JT, Zhou S, Tjian R. 2000. A TRF1:BRF complex directs *Drosophila* RNA polymerase III transcription. *Cell* **101:** 459–469.

Teichmann M, Wang Z, Martinez E, Tjernberg A, Zhang D, Vollmer F, Chait BT, Roeder RG. 1999. Human TATA-binding protein-related factor-2 (hTRF2) stably associates with hTFIIA in HeLa cells. *Proc Natl Acad Sci* **96:** 13720–13725.

Theisen JW, Lim CY, Kadonaga JT. 2010. Three key subregions contribute to the function of the downstream RNA polymerase II core promoter. *Mol Cell Biol* **30:** 3471–3479.

Tokusumi Y, Ma Y, Song X, Jacobson RH, Takada S. 2007. The new core promoter element XCPE1 (X Core Promoter Element 1) directs activator-, mediator-, and TATA-binding protein-dependent but TFIID-independent RNA polymerase II transcription from TATA-less promoters. *Mol Cell Biol* **27:** 1844–1858.

Veenstra GJ, Weeks DL, Wolffe AP. 2000. Distinct roles for TBP and TBP-like factor in early embryonic gene transcription in *Xenopus*. *Science* **290:** 2312–2315.

Verma N, Hung KH, Kang JJ, Barakat NH, Stumph WE. 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.

Vermeulen M, Mulder KW, Denissov S, Pijnappel WWMP, van Schaik FMA, Varier RA, Baltissen MPA, Stunnenberg HG, Mann M, Timmers HTM. 2007. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131:** 58–69.

Vo ngoc L, Cassidy CJ, Huang CY, Duttke SHC, Kadonaga JT. 2017. The human initiator is a distinct and abundant element that is precisely positioned in focused core promoters. *Genes Dev* **31:** 6–11.

Wang YL, Duttke SHC, Chen K, Johnston J, Kassavetis GA, Zeitlinger J, Kadonaga JT. 2014. TRF2, but not TBP, mediates the transcription of ribosomal protein genes. *Genes Dev* **28:** 1550–1555.

Willy PJ, Kobayashi R, Kadonaga JT. 2000. A basal transcription factor that activates or represses transcription. *Science* **290:** 982–984.

Wong JM, Bateman E. 1994. TBP-DNA interactions in the minor groove discriminate between A:T and T:A base pairs. *Nucleic Acids Res* **22:** 1890–1896.

Wu X, Sharp PA. 2013. Divergent transcription: a driving force for new gene origination? *Cell* **155:** 990–996.

Xiao L, Kim M, DeJong J. 2006. Developmental and cell type-specific regulation of core promoter transcription factors in germ cells of frogs and mice. *Gene Expr Patterns* **6:** 409–419.

Xu M, Sharma P, Pan S, Malik S, Roeder RG, Martinez E. 2011. Core promoter-selective function of HMGA1 and Mediator in Initiator-dependent transcription. *Genes Dev* **25:** 2513–2524.

Young NL, DiMaggio PA, Plazas-Mayorca MD, Baliban RC, Floudas CA, Garcia BA. 2009. High throughput characterization of combinatorial histone codes. *Mol Cell Proteomics* **8:** 2266–2284.

Zabidi MA, Arnold CD, Schernhuber K, Pagani M, Rath M, Frank O, Stark A. 2015. Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* **518:** 556–559.

Zehavi Y, Kuznetsov O, Ovadia-Shochat A, Juven-Gershon T. 2014. Core promoter functions in the regulation of gene expression of *Drosophila* Dorsal target genes. *J Biol Chem* **289:** 11993–12004.

Zhang D, Penttila TL, Morris PL, Teichmann M, Roeder RG. 2001. Spermiogenesis deficiency in mice lacking the Trf2 gene. *Science* **292:** 1153–1155.

Zhang H, Ma ZY, Zeng L, Tanaka K, Zhang CJ, Ma J, Bai G, Wang P, Zhang SW, Liu ZW, et al. 2013. DTF1 is a core component of RNA-directed DNA methylation and may assist in the recruitment of Pol IV. *Proc Natl Acad Sci* **110:** 8290–8295.

Zhou T, Chiang CM. 2001. The intronless and TATA-less human TAF_{II}55 gene contains a functional initiator and a downstream promoter element. *J Biol Chem* **276:** 25503–25511.

Zhou T, Chiang CM. 2002. Sp1 and AP2 regulate but do not constitute TATA-less human TAFII55 core promoter activity. *Nucleic Acids Res* **30:** 4145–4157.

Zhou H, Grubisic I, Zheng K, He Y, Wang PJ, Kaplan T, Tjian R. 2013. Taf7l cooperates with Trf2 to regulate spermiogenesis. *Proc Natl Acad Sci* **110:** 16886–16891.