



## Review

## The core promoter: At the heart of gene expression

Yehuda M. Danino, Dan Even, Diana Ideses, Tamar Juven-Gershon\*



The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan 5290002, Israel

## ARTICLE INFO

## Article history:

Received 17 February 2015

Received in revised form 19 April 2015

Accepted 23 April 2015

Available online 28 April 2015

## Keywords:

Core promoter

RNA Pol II transcription

Core promoter elements/motifs

Enhancer-promoter specificity

Core promoter preferential activation

Gene expression

## ABSTRACT

The identities of different cells and tissues in multicellular organisms are determined by tightly controlled transcriptional programs that enable accurate gene expression. The mechanisms that regulate gene expression comprise diverse multiplayer molecular circuits of multiple dedicated components. The RNA polymerase II (Pol II) core promoter establishes the center of this spatiotemporally orchestrated molecular machine. Here, we discuss transcription initiation, diversity in core promoter composition, interactions of the basal transcription machinery with the core promoter, enhancer-promoter specificity, core promoter-preferential activation, enhancer RNAs, Pol II pausing, transcription termination, Pol II recycling and translation. We further discuss recent findings indicating that promoters and enhancers share similar features and may not substantially differ from each other, as previously assumed. Taken together, we review a broad spectrum of studies that highlight the importance of the core promoter and its pivotal role in the regulation of metazoan gene expression and suggest future research directions and challenges.

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

Appropriate temporal and spatial gene expression is a highly complex process underlying the fate and function of different cells and tissues. The regulation of this process is composed of multiple levels and orchestrated molecular events [1–3]. A central event in the regulation of eukaryotic gene expression is the initiation of transcription. The initiation of transcription of protein-coding genes and distinct non-coding RNAs occurs following the recruitment of RNA polymerase II (Pol II) to the core promoter region by the basal transcription machinery [4].

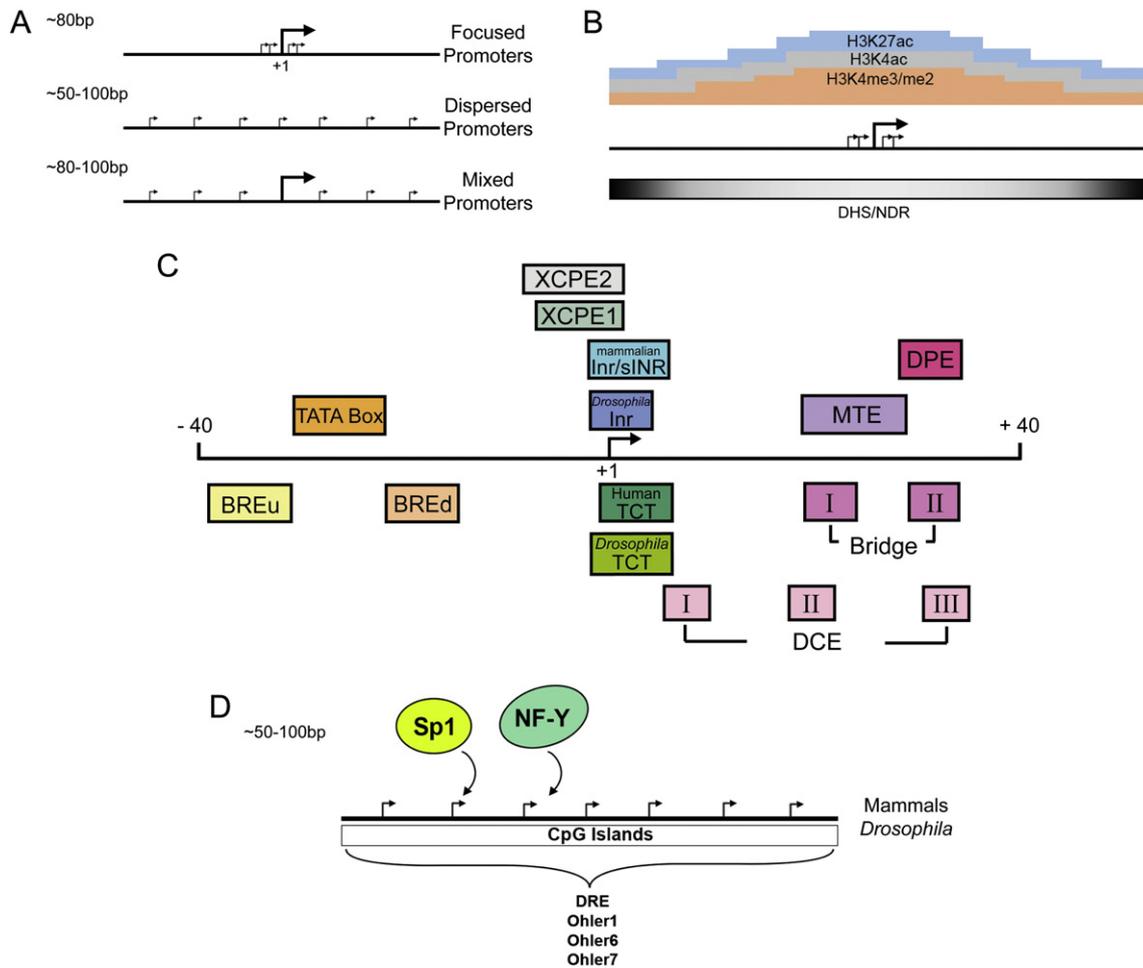
The core promoter is generally defined as the minimal DNA sequence that directs accurate initiation of transcription. The core promoter sequence encompasses the transcription start site (TSS), typically referred to as the +1 position [5,6]. Examination of the distribution of TSSs reveals that there are multiple modes of transcription initiation (Fig. 1A). Distinct molecular players can open the chromatin structure at the core promoter region and thus facilitate initiation of transcription. Interestingly, active promoters are associated with specific chromatin signatures. These include: nucleosome-depleted regions (NDR) or reduced nucleosome occupancy over the promoters, DNaseI hypersensitive sites (DHS) and the enrichment of specific histone modifications, such as di- and tri-methylation of H3K4 and acetylation of H3K4 and H3K27 (Fig. 1B) [7,8]. Notably, both human and yeast nucleosomes that are upstream and downstream of the TSSs of multiple genes have been shown to

correlate with the occupancy of the histone variant H2A.Z (termed Htz1 in yeast) (reviewed in [9]).

In the past, it was assumed that the core promoter is a generic entity that functions in a universal manner. Nowadays however, the growing convention is that the unique properties of a given promoter are a function of its architecture and core promoter motifs composition (Fig. 1C and D) [5,6,10,11].

The core promoter, which is often referred to as “the gateway to transcription”, is a central component in the initiation of transcription [12,13]. Research in the past decade has enhanced our understanding of the fundamental roles that the core promoter plays in the initiation of transcription, as well as in the regulation of additional aspects of gene expression. Insights are gained from studies of specific genes and gene networks [13–15], as well as from genome-wide studies [11,16] utilizing methodologies such as PEAT [17], 5' RACE [18], CAGE [19], FAIRE-seq [20], ChIP-seq [21], Gro-seq [22], and RNA-seq [23], and key projects and consortia (e.g. modENCODE [24], ENCODE [25] and FANTOM5 [26]), which developed following the implementation of some of the above methods. Accordingly, core promoters can be studied at different resolutions: from genomic architecture, transcription co-regulators and sequence-specific transcription factors (Fig. 2A), through basal transcription factors (Fig. 2B and C) and DNA sequence motifs (Fig. 2C). Importantly, the different experimental strategies complement each other and together, provide the elaborate view of core promoters. Here, we review the current state of knowledge relevant to the contribution of the core promoter to multiple aspects of gene expression, and discuss future directions and challenges in the field.

\* Corresponding author. Tel.: +972 3 531-8244; fax: +972 3 738 4058.  
E-mail address: [tamar.gershon@biu.ac.il](mailto:tamar.gershon@biu.ac.il) (T. Juven-Gershon).



**Fig. 1.** General features of the core promoter region. A. The three main core promoter types based on the distribution of TSSs, including focused, dispersed and mixed promoters. Small arrows represent weak TSSs, whereas a large arrow represents a single strong TSS. B. Chromatin features of active core promoters include distinct post-translational modifications and nucleosome depletion. Associated histones marks are depicted: H3K4me2/me3 (orange), H3K4ac (gray), H3K27ac (light blue). A DHS/NDR pattern ranging from nucleosome-free (light) to nucleosome-occupied regions (dark) is illustrated below. C. Schematic illustration of the most common core promoter elements found in focused promoters. The diagram is roughly to scale. D. Schematic illustration of the known factors and sequence motifs that are associated with dispersed promoters.

## 2. Diversity in the transcription initiation landscape

### 2.1. Multiple modes of transcription initiation

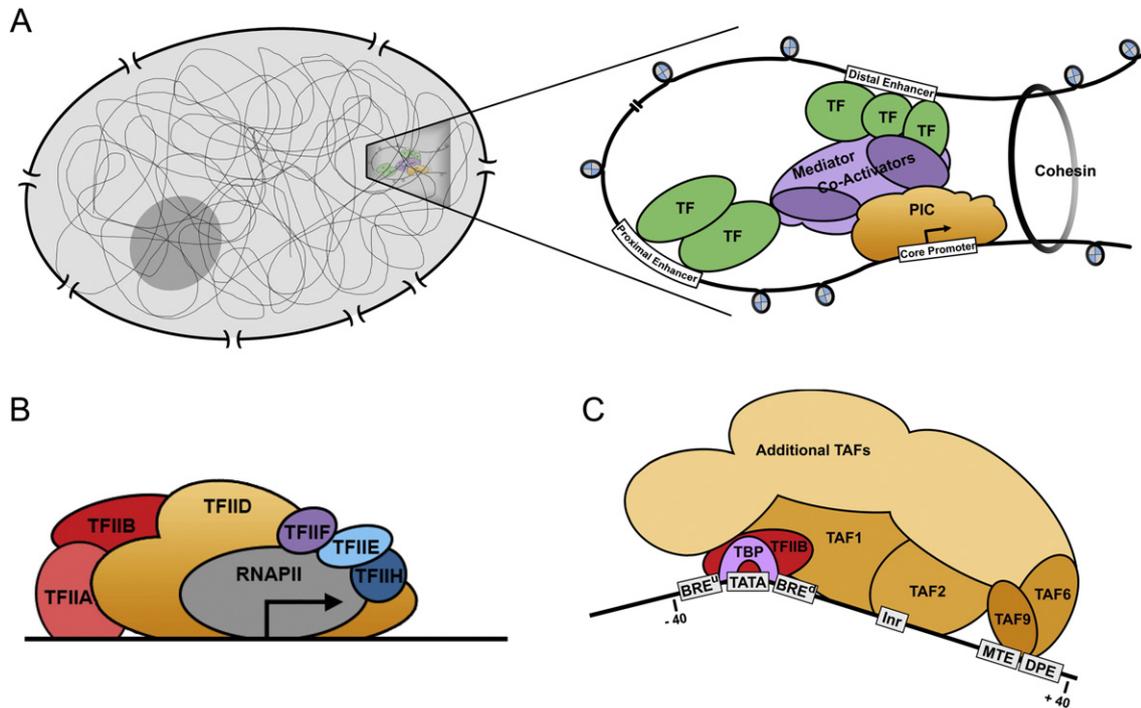
The core promoter is best known for its role in directing proper transcription initiation at the TSS. Two modes of transcription initiation, focused and dispersed, were previously noted in metazoan (Fig. 1A) (reviewed in [6,11]). Focused (also termed “sharp peak”) promoters contain a single predominant TSS or a few TSSs within a narrow region of several nucleotides [10]. Focused promoters encompass approximately between  $-40$  to  $+40$  nucleotides relative to the TSS (referred to as the  $+1$  position). Focused transcription initiation is associated with spatiotemporally regulated tissue specific genes [27] and with canonical core promoter elements that have a positional bias, such as the TATA box, Initiator, MTE and DPE [28] (Fig. 1C).

Dispersed (also termed “broad”) promoters contain multiple weak start sites that spread over 50 to 100 nucleotides at the promoter region ([10,11] and refs therein). Dispersed transcription initiation is associated with constitutive or housekeeping genes. Vertebrate dispersed promoters often contain CpG islands and Sp1 and NF-Y sites [6,10,29], whereas *Drosophila* core promoters often contain elements that have weaker positional biases (as compared to the focused promoters), but frequently co-occur in a specific order and orientation: Ohler 1, DNA replication element (DRE), Ohler 6 and Ohler 7 [28,30] (Fig. 1D). Although the focused promoter architecture exists in all the organisms

and is the predominant initiation mode in simpler organisms, the dispersed mode is more common in higher eukaryotes [10,27]. For example, over 70% of vertebrate promoters are dispersed [29,31–33]. From a teleological standpoint, the associations of sharp TSSs with regulated genes and of broad TSSs patterns with constitutively expressed genes are rather intuitive. It would be easier to achieve a more precise control of gene expression from focused TSSs, as compared with dispersed promoters of housekeeping genes, which would be constitutively transcribed with minimal variation of gene expression by usage of multiple start sites [10].

### 2.2. Focused versus Dispersed initiation patterns – recent studies, new insights

Despite the abovementioned distinction between the two modes of transcription initiation, classification of transcription initiation landscapes is not so straightforward. Functional experiments and genome-wide studies using advanced technologies imply that there are multiple ways to classify promoters. Thus, the boundaries between these two major types of promoters are sometimes unclear [6,34]. With respect to the “focused vs. dispersed” sub-classifications mentioned above, a mixed promoter (also termed “broad with peak”; [17]), an additional promoter type, was revealed. This promoter type exhibits a dispersed initiation pattern with a single strong transcription start site [6,35] (Fig. 1A). Several studies classified mammalian promoters using



**Fig. 2.** The core promoter can be studied from different angles in multiple resolutions. A. Zooming in on global genomic interactions in the nucleus, one can study long-range interactions, such as those between enhancers and promoters, by analyzing chromatin looping, cohesion function, interactions of transcription factors (TFs) with co-activators and cis-regulatory modules and interactions of the preinitiation complex (PIC) components with their target promoters. B. Zooming in on the basal transcription machinery, one can study the assembly and composition of the PIC at different Pol II-promoters and the 3D structure of different PIC components. C. Zooming in on the DNA-binding PIC components (TFIIB and TFIID), one can focus on the alternative protein components at different Pol II-promoters, on the core promoter composition of specialized transcription programs, and on the interactions of different PIC components with specific core promoter elements.

alternative criteria [27,29,33]. The Ren lab classified active promoters based on genome-wide ChIP experiments for TFIID and Pol II, as well as H3Ac and H3K4me, regardless of focused or dispersed initiation patterns [33]. Bajic et al. [29] define four promoter types, based on distribution of dinucleotides over the promoter regions, CpG Islands and TATA boxes. Moreover, Carninci et al. [27] classified promoters into four groups based on CAGE analysis: single peak, broad shape peak, bimodal/multimodal peak and broad with dominant peak. These studies also challenge the “focused vs. dispersed” classification, as some mouse and human promoters contain both CpG Islands and TATA boxes. A recent comprehensive review [11], which compared genome-wide studies in human and *Drosophila*, presented another sub-classification of three major types of promoters termed Type I, Type II and Type III. Type I promoters contain TATA boxes and focused TSSs, lack CpG islands and are associated with tissue-specific expression in adult tissues. Type II promoters contain CpG islands and dispersed TSSs. In mammals, type II promoters lack TATA boxes, and in *Drosophila* they contain DRE, Ohler 1 or Ohler 6 motifs. Genes belonging to this group are associated with broad expression throughout the organism's life. Type III promoters are associated with developmentally regulated genes, which in *Drosophila* contain combinations of Initiator and DPE motifs. In mammals, type III promoters contain large CpG islands.

Taken together, the transcriptional initiation landscape is more complex than the simple classification of two types of promoters.

### 2.3. Bidirectional and divergent transcription

Another manifestation of the complexity of transcription initiation is the phenomenon of bidirectional transcription. Bidirectional transcription, which presents two closely spaced transcription initiation events (within less than 1 kb) of head-to-head Pol II transcripts in both sense and anti-sense orientations, was originally defined for adjacent head-to-head oriented pairs of protein-coding genes [36]. The relatively short region that contains the opposite-oriented initiations

and separates between these genes, is often called a “bidirectional promoter” [37]. Experimental and computational studies have characterized many features of bidirectional promoters. In general, it is shown that 10–22% of the genes in mammals are organized in this manner [38]. Moreover, the bidirectionality was shown to be controlled in a cell-type specific manner, and these pairs of genes are coordinately regulated ([38] and refs therein). Hence, bidirectional promoters might have evolved to facilitate the regulation of transcription of different genes at the same time, and might consist of two separate, yet dependent, core promoters. Additionally, a computational analysis supports an evolutionary role for bidirectional promoters in the emergence of novel species-specific transcripts [39]. Bioinformatics analysis of the distribution of common core promoter elements (BRE<sup>u</sup>, TATA box, Inr and DPE) and CpG islands at bidirectional versus unidirectional promoters, demonstrated that while the BRE<sup>u</sup> is enriched at bidirectional promoters, the Inr and DPE elements are similarly detected at both promoter types [40]. The TATA box is rare in general, but is enriched in bidirectional promoters of histone genes. Moreover, it was shown that the CpG islands and Sp1 binding sites are common features of most of the bidirectional promoters, compared to unidirectional promoters [41]. Other studies focused on overrepresented binding-sites of different transcription factors, and in some cases - on their influence on the expression of two opposite genes regulated by a bidirectional promoter [38,42].

Interestingly, another manifestation of bidirectional transcription involving non-coding RNAs (ncRNAs) was recently characterized. Multiple classes of ncRNAs were identified in different organisms (reviewed in [43]). One of these classes is promoter-associated ncRNAs. During the years, classes of promoter-associated non-coding transcripts were discovered in bacteria, yeast, *Drosophila*, mouse, human and plants ([43–45] and refs therein). Four studies, published back-to-back in 2008, described new classes of promoter-associated ncRNAs in humans and mice [22,46–49]. These ncRNAs were generally divided into two classes, termed TSS-associated RNAs (TSSa-RNAs) [48] and promoter

upstream transcripts (PROMPTs) [47] or upstream antisense RNAs (uaRNAs) [50], which share many features. They are short, present at low abundance and are associated with CpG islands and active-promoter-related histone marks (H3K4me3, H3ac), but not with elongation-related histone marks (H3K36me3, H3K79me3).

Non-coding antisense RNAs derived from bidirectional promoters have very short half-lives and are barely detectable. Two recent studies have shown that an asymmetric distribution of polyadenylation signals and U1 snRNP-binding sites surrounding TSSs control transcript stability [50–52]. Notably, bidirectional initiation is also a feature of enhancer RNAs (eRNA; see Section 7) [53,54].

The Lis lab has demonstrated that nearly 80% of active genes have bidirectional promoters, suggesting that bidirectional initiation is a general feature of mammalian genomes [22,55]. Hence, these divergent ncRNAs may be regarded as markers for active promoters of protein-coding genes [22,46–48,56]. Duttke et al. have recently analyzed transcription from human promoters in HeLa cells and have classified promoters into three types: unidirectional promoters, divergent promoters (containing an annotated gene in the forward direction and no annotated gene in the reverse direction) and bidirectional promoters (containing annotated genes in both directions) [57]. Surprisingly, they discovered that about half of human active promoters are intrinsically unidirectional. Moreover, the divergent transcripts result from their own reverse-oriented core promoters. The authors suggest that divergent transcription is not an inherent property of the transcription process, but a consequence of the presence of both forward and reverse-directed promoters. This suggestion is in line with the two occupancy peaks observed for each TBP and Pol II by the Lis lab [55]. The Lis lab observed tight spacing (estimated 110 bp) between the forward and reverse-directed promoters [55], whereas the Ohler & Kadonaga labs, observed variable, however larger, spacing between the two [57]. It remains to be determined whether the difference between these findings results from the differences between the different cell lines used or from the analysis methodology.

Despite the impressive discoveries related to bidirectional transcription in the last few years (which highlight the complexity of gene expression), the functional role of short non-coding antisense RNAs still remains elusive. From this point onwards, we only refer to the comprehensively studied focused and dispersed core promoter types.

### 3. Core promoter elements: the combinatorial code of precise transcription initiation

The Pol II core promoter is composed of short DNA sequences that are referred to as core promoter elements or motifs. The majority of core promoter motifs serve as binding sites for components of the basal transcription machinery, in particular TFIID, which is composed of TATA box-binding protein (TBP) and TBP-associated factors (TAFs), and TFIIB [4,58,59].

The basal transcription machinery recruits Pol II to the core promoter that directs the initiation of transcription [4,6,10,60–62]. Nevertheless, there are no universal core promoter elements, and diverse core promoter compositions have been reported [6,63]. In this section, we will briefly discuss the majority of core promoter elements (schematically depicted in Fig. 1C and D), which have been analyzed in *Drosophila* and mammals, with particular emphasis on their variety and the relations between them.

#### 3.1. The precisely positioned core promoter elements are common in the focused promoters

Early studies from the Chambon lab described the existence of a putative element at the TSS [64]. The function of the initiator (Inr) as a transcriptional element that encompasses the +1 TSS was articulated by Smale and Baltimore [65]. The Inr is probably the most prevalent core promoter motif in focused core promoters [66–68]. It is mainly

bound by the TAF1 and TAF2 subunits of TFIID [69–72]. The mammalian Inr consensus sequence is YYA<sub>+1</sub>NWYY (IUPAC nomenclature) [73], and the *Drosophila* consensus is TCA<sub>+1</sub>KTY [71,74]. Inr-like sequences were also identified in *Saccharomyces cerevisiae* [75]. Computational analyses of promoters argue that the Inr consensus is only YR (–1, +1 positions) in humans [11,27,76] or TCA<sub>+1</sub>GTY for *Drosophila* [66,68]. The A nucleotide (or R in the YR consensus) is generally designated as the +1 position, even when transcription does not initiate at this specific nucleotide. This critical convention is instrumental, because functional downstream elements are completely dependent on the presence of an Inr and the precise spacing from it [6,10,13].

Notably, a strict version of the mammalian initiator (sInr), which is present in 1.5% of human genes and enriched in TATA-less promoters of specific functional categories, was defined as CCA<sub>+1</sub>TYTT, with conserved sequences flanking the motif [77]. The sInr motif functions in cooperation with Sp1 and can replace the conventional Inr, but not vice versa. Similarly to the canonical Inr element, sInr is bound by TAF1 and its function depends on it [77]. The YY1 transcription factor binds sInr, but this binding is dispensable for sInr function [77].

In addition to these versions of the Inr, a few core elements that encompass the transcription start site were identified. The polypyrimidine initiator motif (TCT), which was originally identified in mouse, is conserved from *Drosophila* to humans [14,78–80]. The TCT has a consensus sequence of YYC<sub>+1</sub>TTYT in *Drosophila* and YC<sub>+1</sub>TYTYT in humans, in which C is the +1 TSS. Although the Inr consensus resembles the TCT consensus, the TCT motif cannot substitute for an Inr to initiate transcription [14]. The TCT overlaps with a motif that was previously identified in humans, termed 5'-terminal oligopyrimidine tract (5'-TOP) (reviewed in [81]), which is functionally distinct from it [14]. Both the TCT and the 5'-TOP elements are enriched and are functional in the transcription of ribosomal protein genes and proteins involved in the regulation of translation [14,78].

Two additional core promoter motifs that are located around TSSs were originally identified in the hepatitis B virus X gene promoter, which contains two TSSs. The X gene core promoter element 1 (XCPE1) drives Pol II transcription from the first TSS of the X gene promoter as well as from other human promoters, when accompanied by co-activator sites. XCPE1 is found in ~1% of the human genes (particularly TATA-less genes) and its consensus sequence DSGYGGRAS<sub>+1</sub>M spans positions –8 to +2 relative to the TSS [82]. Unlike XCPE1, the X gene core promoter element 2 (XCPE2) is sufficient to drive Pol II transcription by itself. The XCPE2 directs transcription from the second TSS of the X gene mRNA, but it also drives transcription from additional human promoters, in a TAF-free manner. Its consensus sequence VCYCRRTRCM<sub>+1</sub>Y spans positions –9 to +2 relative to the TSS [83].

There are core promoter elements that are located upstream of the TSS. The TATA box motif is the first core promoter motif to be identified [84]. Although the TATA box was previously considered to be a universal element, it is presently estimated that only 8–30% of metazoan core promoters [27,33,60,68,85] and 20–46% yeast promoters [62,86,87] are TATA-dependent. The TATA box motif is also present in plants [88, 89], however the majority of *Arabidopsis* promoters are TATA-less [90]. The TATA box is bound by the TBP subunit of TFIID ([5,6,63] and refs therein). Both the TATA box element and the TBP are conserved from archae to humans [10,91]. The consensus sequence of the TATA box is TATAWAAR, where the 5' T is usually located at –30 or –31 relative to the TSS in metazoans (or at –120 to –40 in yeast). A wide range of sequences can functionally replace the yeast TATA box for *in vivo* transcriptional activity [92]. Notably, transcription initiation in yeast, which occurs at variable and greater distances downstream of core promoter elements such as the TATA box, has been suggested to occur via Pol II scanning, where a PIC assembles at the TATA box and Pol II translocates downstream, searching the DNA for suitable start sites (reviewed in [93]). Interestingly, the TATA box has recently been

shown to contribute to high rates of transcription re-initiation of human microRNAs, resulting in reduced microRNA levels [94]. The authors suggested that the lower synthesis rates directed by the TATA box result from Pol II crowding at pause sites, thus increasing the chance for premature termination.

The TFIIB recognition elements (BRE), which are bound by the TFIIB basal transcription factor, are located immediately upstream or downstream of the TATA box, respectively [95–97]. TFIIB contacts these two elements by two independent DNA-recognition motifs within its core domain [96]. The consensus of the upstream BRE (BRE<sup>u</sup>) is SSRCGCC [97], and the consensus of the downstream BRE (BRE<sup>d</sup>) is RTDKKKK [95]. The TFIIB and the BRE elements are conserved from archae to humans [6,96]. Both BRE<sup>u</sup> and BRE<sup>d</sup> act in conjunction with the TATA box [6,10]. A bioinformatics analysis using the EPD database showed that 25% of the eukaryotic core promoters contain a potential BRE<sup>u</sup> [85]. Surprisingly, this study revealed that the BRE<sup>u</sup> is more prevalent in TATA-less promoters (28.1%) than in TATA-containing promoters (11.8%). Both elements exert positive as well as negative effects on basal transcription and on activated transcription in a manner that is context-dependent [95,97–99].

In addition to the abovementioned upstream elements, there are core promoter elements that are located downstream of the TSS. The downstream core promoter element (DPE), which was discovered as a TFIID recognition site that is downstream of the Inr, is precisely located at +28 to +33 relative to the A<sub>+1</sub> of the Inr, with a functional range set of DSWYVY [100–102]. In addition to this functional range set, the guanine at +24 was shown to contribute to DPE function [102]. The DPE is prevalent in developmental gene networks [11,15,99,103]. Importantly, a recent study provides *in vivo* evidence that expression driven by the homeotic *Antennapedia* P2 promoter during *Drosophila* embryogenesis is dependent on the DPE [103]. The motif ten element (MTE) was identified as an overrepresented core promoter sequence, which is located immediately upstream of the DPE, encompassing positions +18 to +29 relative to the A<sub>+1</sub> of the Inr [68]. As positions +28 to +29 overlap the DPE, the MTE consensus sequence was defined for positions +18 to +27 (CSARCSSAAC) [104]. Although the majority of the MTE-containing promoters contain a DPE, the MTE motif functions independently of the DPE [104,105]. Both the MTE and DPE serve as recognition sites for TFIID and appear to be in close proximity to TAF6 and TAF9 [101,105]. Interestingly, TAF1 and TAF2 may also contribute to the recognition of downstream core promoter elements [70]. Using single-nucleotide substitution analysis, the MTE and DPE together were found to consist of three functional sub-regions: positions 18–22, 27–29 and 30–33 downstream to the A<sub>+1</sub> of the Inr. The bridge configuration, which includes the first and the third functional sub-regions (Bridge I, positions 18–22 with favored nucleotides CSARC; Bridge II, positions 30–33 with favored nucleotides WYVY), was shown to be a naturally rare but functional core promoter element [105]. Both the MTE and DPE are conserved from *Drosophila* to humans [6,100,101,104–108]. The MTE, DPE and Bridge motifs are exclusively dependent on the presence of a functional Inr, and are enriched in TATA-less promoters. However, co-occurrence of putative TATA, Inr and DPE motifs was observed in a small fraction of *Drosophila* genes [15,85].

An additional downstream element was identified and characterized in the human adult  $\beta$ -globin promoter. This element, termed downstream core element (DCE), was detected by scanning mutagenesis of the +10 to +45 in the promoter region. The DCE is composed of three sub-elements, located at positions +6 to +11 (necessary motif CTTC), +16 to +21 (necessary motif CTGT), and +30 to +34 (necessary motif AGC) relative to the TSS. The DCE is distinct from the MTE, DPE and Bridge downstream elements, as the DCE is recognized and bound by TAF1 [109] and not by TAF6 or TAF9 [101,105]. Moreover, unlike the DPE, the DCE is frequently found in TATA box-containing promoters [109,110]. Table 1 summarizes the precisely positioned core promoter elements found in focused promoters and the proteins that bind them.

### 3.2. Core promoter elements with weak positional biases in dispersed promoters

Even though the vast majority of core promoter elements are precisely located in focused promoters, there are still a few variably located motifs that were also identified in dispersed promoters. These variably located elements, like some of the precisely located elements discussed above, are associated with specific gene groups.

As mentioned, there are sequence motifs such as the DNA-replicated-related element (DRE) and Ohler 1, 6 and 7 motifs, which were detected by a computational analysis as commonly expressed in dispersed promoters of *Drosophila* genes with maternally inherited transcripts [28]. The consensus sequences of the DRE, Ohler 1, 6 and 7 motifs are WATC GATW, YGGTCACACTR, KTYRGATWTTT and KNNCAKCNCTRNY, respectively [68]. The DRE is a target of the DNA replication-related-element binding factor (DREF). DREF, which was discovered in *Drosophila* and was later found to have orthologues in many other species (including humans), is involved in transcriptional regulation of proliferation-related genes [111]. A motif 1 binding protein (M1BP) has recently been identified and the enrichment of Motif 1 and M1BP was implicated in cytoskeletal organization, mitotic cell cycle and metabolism [112].

### 3.3. The interplay between core promoter elements

With the notion that there are no universal core promoter elements and that core promoter elements are a very important feature of regulation of gene expression, many studies examined the combinations between core promoter elements such as: Inr, TATA box, BRE<sup>u</sup>, BRE<sup>d</sup>, MTE and DPE, and their effects on the transcriptional output. For example, the BRE elements were originally characterized as functional elements in conjunction with the TATA box. In this context, both the BRE<sup>u</sup> and the BRE<sup>d</sup> either increase or decrease the levels of basal transcription [95,97,98,113]. Notably, the addition of a BRE<sup>u</sup> element to a core promoter of a Caudal target gene has a differential effect on transcription in a TATA box- or DPE- context [99]. The TATA box and the Inr cooperate, in certain cases, as synergistic elements [114]. An antagonistic behavior was demonstrated between TBP, which activates TATA transcription and inhibits DPE transcription, and NC2 and Mot1, which activate DPE transcription by inhibiting the function of TBP [115].

The functionality of the DPE, MTE and Bridge elements is, by definition, dependent on their precise location relative to the Inr [100,101,104,105]. Synergy was observed between the MTE and DPE, as well as between the MTE and TATA box [104]. Based on these relationships, a synthetic core promoter, termed super core promoter (SCP), containing a TATA box, Inr, MTE and DPE was designed. Remarkably, the SCP is stronger than any of the natural core promoters examined [116].

Collectively, these findings indicate that the levels of gene expression can be modulated by the core promoter composition. Such modulation is directly achieved by the impact of the combinations of core promoter elements on the architecture of the basal transcription machinery, which provides an additional level of transcriptional regulation. The core promoter may have diversified during evolution so that each element may work with the other, depending on the context and organism. Hence, simple categorization may disregard the complexity of gene expression.

## 4. Functional and structural insights regarding the role of the core promoter in the assembly of the Pol II transcription machinery

In this section, we describe the assembly of the basal transcription machinery components (primarily based on the analysis of TATA-dependent promoters) and their distinct roles in specific cellular contexts.

**Table 1**

The precisely positioned core promoter elements found in focused promoters and the proteins that bind them.

Name	Consensus	Positions (relative to the +1)	Bound by
BRE <sup>u</sup>	SSRCGCC	Immediately upstream of the TATA box	TFIIB
TATA box	TATAWAAR	–30/–31 to –23/–24	TBP
BRE <sup>d</sup>	RTDKKKK	Immediately downstream of the TATA box	TFIIB
Mammalian Inr/sINR	YYA <sub>+1</sub> NWYY /CCA <sub>+1</sub> TYTT	–2 to +5	TAF1 & TAF2 /TAF1
<i>Drosophila</i> Inr	TCA <sub>+1</sub> KTY	–2 to +4	TAF1 & TAF2
Human TCT	YC <sub>+1</sub> TYTY	–1 to +6	
<i>Drosophila</i> TCT	YYC <sub>+1</sub> TTTTY	–2 to +6	
XCPE1	DSGYGGRAS <sub>+1</sub> M	–8 to +2	
XCPE2	VCYCRTRRCM <sub>+1</sub> Y	–9 to +2	
MTE	CSARCSAACGS	+18 to +29	TAF6 & TAF9
DPE	DSWYVY (functional range set)	+28 to +33	TAF6 & TAF9
Bridge	Part I: CGANC Part II: WYGT	Part I: +18 to +22 Part II: +30 to +33	TAF6 & TAF9 (?)
DCE	Necessary motifs: Part I: CTTC Part II: CTGT Part III: AGC	Part I: +6 to +11 Part II: +16 to +21 Part III: +30 to +34	TAF1

#### 4.1. Terminology change: from “general” to “basal” transcription machinery

Classic biochemical studies performed over 30 years ago using the TATA box-containing adenovirus major late promoter identified the general transcription factors (GTFs) as accessory factors for accurate Pol II transcription initiation [117,118]. The GTFs were named TFIIA, TFIIB, TFIID, TFIIE, TFIIIF and TFIIF, based on the protein fractions they purified in (reviewed in [4]). These components, together with Pol II, were necessary and sufficient for basal transcription of the adenovirus major late promoter. They assemble into the preinitiation complex (PIC) by protein-protein interactions and by mediating core promoter recognition (Fig. 2B).

In the past, it was generally accepted that the PIC composition of GTFs does not vary between promoters with different core promoter architecture, and the PIC is nucleated by the binding of the TBP subunit of TFIID, which binds the TATA box [119] (reviewed in [4,31]). Traditionally, this simple model has been considered “general”. However, due to the diversity in core promoter composition and the realization that the known GTFs are insufficient to transcribe DPE-containing promoters [120], it is suggested that the GTFs do not function in a “general” manner, and different compositions of PIC exist. Indeed, the non-ubiquitous expression pattern of certain TAFs imply that they cannot be PIC components in every cell type [58]. Moreover, many studies have presented the variability in PIC formation, specifically by the molecular flexibility in TFIID composition. Hence, GTFs should be addressed as “basal” rather than “general” transcription factors (also discussed in [58,121–123]).

#### 4.2. Compatibility between PIC components, related factors and core promoter elements

Undoubtedly, the diverse assembly of the basal transcription factors, as well as the diversity of core promoter elements, is a complex subject, both structurally and functionally. Nevertheless, due to this complexity, the PIC, which is pivotal for core promoter recognition ([58,121,124] and refs therein), can assemble at core promoters with varying compositions and regulate Pol II transcription in different cells and organisms. In agreement with that, requirements for a “match” between the PIC and the core promoter have been observed in recent years.

This compatibility has mainly been reflected in studies addressing the flexibility and modularity of TFIID subunits and the entire TFIID complex. Early footprinting assays detected differential TFIID protection patterns with respect to the presence of a TATA box and BRE in mammalian promoters [125,126], and a DPE in *Drosophila* [101]. These studies and others [127] have demonstrated the important roles of TAFs in the assembly of the PIC, and hence, in the transcription process. As mentioned earlier, sub-modules of TFIID bind specific core promoter elements, e.g. TBP binds the TATA box, TAF1/TAF2 bind the Inr, TAF1 binds the DCE and TAF6/9 bind the DPE and the MTE (Fig. 2C and Table 1) [69–71,100,101,104,107,109]. It is noteworthy that TAF4/TAF12 and TAF4b/TAF12 sub-complexes can also bind core promoters [107], and are necessary for transcription of a sub-group of genes, which are mostly associated with TATA box and Inr motifs [128]. Interestingly, TAF1 contains two distinct enzymatic activities: an acetyltransferase and a kinase activity, which are important for regulating non-overlapping, different gene sets *in vivo* [129], suggesting that different functional modules of the PIC contribute to transcription of different target genes.

While TBP and TAF1 were initially considered the nucleating subunits of holo-TFIID assembly [130], Wright et al. [131] discovered that *Drosophila* TAF4 preferentially nucleates TFIID in TATA-less, DPE-containing promoters. This study also uncovered a stable core-sub-complex, composed of TAF5 and the histone fold domain (HFD)-containing TAF4, TAF6, TAF9 and TAF12. This core sub-complex is associated with the peripheral subunits TAF1, TAF2, TAF11 and TBP. These core TAFs are incorporated into TFIID in two copies, and are organized in five heterodimer pairs with other HFD-containing TAFs (TAF3-TAF10, TAF6-TAF9, TAF4-TAF12, TAF8-TAF10 and TAF11-TAF13) ([124] and refs therein). Recent structural analysis of human TFIID demonstrated that these core TAFs exhibit two-fold symmetry [132]. Interestingly, incorporation of the TAF8-TAF10 pair breaks the symmetry and allows the entry of the single copy TAFs and TBP into the structure, resulting in an asymmetric holo-TFIID that can nucleate the PIC.

Several TBP-free complexes have been characterized [127,133,134]. One of them, the TBP-free TAF-containing complex (TFTC; also termed SAGA, STAGA and PCAF), is capable of replacing the canonical TFIID at both TATA-less and TATA-containing promoters *in vitro* [127]. The assembly of TAF-less TBP-containing complexes (such as TBP-TFIIA-containing complexes) at specific core promoters, which was somewhat surprising, has also been observed [135–137]. A TAF-free TBP-

containing PIC is important for transcription from HIV-1 LTR promoter [136]. Interestingly, a distinctive TBP-TAF complex, lacking TAF1, TAF4 and TAF10, is involved in transcription of the U2 snRNA gene [138].

These findings add to a growing body of evidence implying that distinct core promoters would be differentially recognized by PICs that contain TBP or are devoid of it. Notably, TBP activates TATA-dependent transcription and represses DPE-dependent transcription, whereas Mot1 and NC2 block TBP function and thus repress TATA-dependent transcription and activate DPE-dependent transcription [115,139]. Interestingly, Deng et al. [140] demonstrated that NC2 acts positively at promoters that lack functional BREs, while TFIIA recruitment, which is dependent on the presence of BREs, reduces transcriptional activity. The association of BRE elements with TATA boxes further supports these findings [85,97]. Interestingly, the architectural DNA-binding protein HMG1 has been shown to interact with the Mediator and activate transcription of mammalian promoters containing both a TATA box and an Inr [141].

Remarkably, the Nogales lab used electron microscopy to visualize human TFIID with promoter DNA, and discovered that TFIID exists in two structurally distinct conformations (termed canonical and rearranged) [142]. The transition between the two states is modulated by TFIIA, and the presence of TFIIA and promoter DNA facilitates the formation of the rearranged conformation [142]. Human TFIID is comprised of three main structural lobes (termed lobe A, B and C) [142, 143]. Using the super core promoter DNA [116], lobe C was shown to interact with downstream elements (DPE and MTE), while lobe A interacts with the Inr and TATA box.

Three TBP-related factors (TRF1, TRF2 and TRF3) have been discovered in the animal kingdom based on their homology to the C-terminal core domain of TBP, which is essential for interaction with the TATA box (reviewed in [121–123,144–146]). Unlike TRF1 and TRF3 (also termed TBP2 and TBPL2), TRF2 (also termed TLP, TLF, TRP and TBPL1), is unable to recognize the TATA box, as the TATA-interacting Phe residues of TBP are not conserved in TRF2 [147–149]. *Drosophila* TRF2 selectively regulates the TATA-less Histone H1 promoter, whereas TBP regulates the TATA-containing core Histones genes [137,150]. The Kadonaga lab has recently discovered that TRF2, and not TBP, regulates transcription of ribosomal protein genes that lack TATA box and contain functional TCT motifs [151]. Kedmi et al. [152] discovered that TRF2 preferentially functions as a core promoter regulator of DPE-containing promoters. These findings and others have highlighted the involvement of TRF2 in the regulation of diverse biological processes driven by distinct core promoter compositions (reviewed in [123]). Taken together, promoter recognition by multiple TAFs, TRFs, TBP-free or TBP-containing complexes, underscore a key regulatory role for core promoters in transcription initiation, and may provide an explanation for evolutionary changes affecting the PIC-promoter interface [153].

#### 4.3. Different basal transcription factors promote distinct biological processes

The diversity in the components of the PIC, especially in TFIID subunits, establishes distinct protein complexes that drive transcription of specific sets of genes (e.g. with cell type- or tissue-specific functions) (reviewed in [154]). The Wassarman lab has shown that *Drosophila* TAF1 affects multiple developmental events *in vivo* [155], and that *Drosophila* TAF6 is broadly required for cell growth and cell fate specification [156]. Moreover, *Drosophila* TAF4 and TAF6 were shown to be required for transcription of the *snail* and *twist* Dorsal-target genes *in vivo* [157]. Human TAF8 was implicated in differentiation of cultured 3 T3-L1 preadipocytes to adipocytes [158]. Interestingly, the *Drosophila* TAF10 homologues TAF10 and TAF10b, are differentially expressed during *Drosophila* embryogenesis [159]. Expression of mouse TAF10 was later shown to be required for early mouse embryogenesis of the inner cell mass, but not the trophoblast [160]. Remarkably, conditional

knock out of mouse TAF10 in embryonic and adult liver resulted in the dissociation of TFIID into individual components [161]. Based on these findings, it was suggested that TFIID is not required for the maintenance of ongoing transcription of hepatic genes. Rather, it is involved in mechanism of postnatal silencing of hepatic genes [161]. Additional studies reveal an important role for distinct TFIID complexes in regulating pluripotency of embryonic stem cells [162,163].

Multiple TAF paralogues have been implicated in different biological processes. A retroposed homologue of human TAF1 (TAF1L) and TAF7L are expressed during male germ-cells differentiation [164,165]. Similarly to humans, TAF7L in mice is required for spermatogenesis in cooperation with TRF2 [165–167]. TAF7L was recently demonstrated to be an important regulator of white- as well as brown- adipose tissue differentiation [168,169]. TAF4b was originally identified as a cell-type-specific TAF in a human B lymphocyte cell line [170]. Using knockout mice, TAF4b was shown to be important for ovarian development and spermatogenesis [171–174]. Remarkably, mouse TAF9L was recently shown to regulate neuronal gene expression *in vivo* [175]. Interestingly, tissue-specific TAF homologues of *Drosophila* TAF4 (*no hitter*), TAF5 (*cannonball*), TAF6 (*meiosis 1 arrest*), TAF8 (*spermatocyte arrest*) and TAF12 (*ryan express*) collaborate to control a testis-specific transcriptional program [176].

TBP paralogues are involved in distinct biological processes, such as embryonic development, differentiation and morphogenesis (reviewed in [121,123,145,177]). TRF2 regulates a subset of genes that differ from TBP-regulated genes. TRF2 is essential for embryonic development of *C. elegans*, *Drosophila*, zebrafish and *Xenopus* [121,123,145,177]. It is highly conserved in evolution and is present in all bilaterian organisms [147]. Since bilaterian organisms contain three germ layers (endoderm, mesoderm and ectoderm) and more ancient animals only contain two germ layers (endoderm and ectoderm), it is tempting to speculate that TRF2 may be important for mesoderm formation. This suggestion is further supported by the fact that the DPE motif is prevalent among *Drosophila* genes that are involved in embryonic development [15,99]. Mouse TRF2, unlike *C. elegans*, *Drosophila*, zebrafish and *Xenopus* TRF2, is not required for embryonic development but is essential for spermiogenesis [178,179]. A separate study demonstrated that the cleavage of TFIIA $\alpha$ - $\beta$  precursor (into the  $\alpha$  and  $\beta$  subunits of TFIIA) is necessary for activation of spermiogenic TRF2 target genes [180]. *Drosophila* *trf2* is also required for the response to the steroid hormone ecdysone during *Drosophila* metamorphosis [181]. Hence, TRF2 drives multiple transcriptional programs [123].

Zebrafish TRF3 is important for initiation of hematopoiesis during embryonic development [182,183], however, both zebrafish and *Xenopus* TRF3 are mainly expressed in oocytes and are essential for embryogenesis [184,185]. Mouse TRF3, which is exclusively expressed in oocytes, is essential for the differentiation of female germ cells but not for embryonic development [186].

These fascinating findings emphasize the motivation to investigate the regulation of gene expression at the core promoter level. It is possible that there are core promoter motifs that have not yet been discovered, and they might be bound by other PIC components. Thus, the analysis of novel core promoter elements in multiple organisms is likely to shed light on mechanistic aspects of transcriptional regulation.

## 5. Enhancer-promoter connectivity

Zooming out from the basal transcription resolution uncovers another facet of regulation of gene expression, namely, enhancer-promoter interactions that regulate the activation of specific genes in a precise spatio-temporal manner. Enhancers contain DNA binding sites for sequence-specific transcription factors that in turn, recruit co-activators and co-repressors and determine the overall activity of the enhancers (reviewed in [187–194]). Originally, scientists searched for enhancers as cis-regulatory elements that stimulate transcription levels from the nearest promoter, irrespective of

orientation. Enhancer–promoter pairs are commonly engaged by enhancer's looping, which physically brings these regulatory elements into proximity, through recruitment of multiple proteins (activators, co-activators, Mediator, cohesin and the PIC). Studies in recent years, employing advanced global methodologies such as chromatin conformation capture (3C), its derivatives (4C, 5C, Hi-C) and ChIA-PET, have led to the discovery of both intrachromosomal and interchromosomal physical contacts with promoters. While multiple enhancers can interact with multiple promoters, specificity between certain enhancers and promoters has been observed. The mechanisms that determine enhancer–promoter specificity are still poorly understood, but they are thought to include biochemical compatibility, constraints imposed by the three-dimensional architecture of chromosomes, insulator elements, and effects of local chromatin environment [194].

In the last twenty years, the compatibility of enhancer–promoter interactions has mostly been studied in *Drosophila*. One of the early studies analyzing the compatibility between enhancer–promoter pairs examined the expression of the neighboring *gooseberry* (*gsb*) and *gooseberry neuro* (*gsbn*) genes [195]. Swapping experiments revealed that although both enhancers (GsbE and GsbNE) are located between the two TSSs of the two genes (and thus cross-activation could potentially occur), the GsbE could only activate the *gsb* promoter, while the GsbNE could only activate the *gsbn* promoter. Another study showed compatibility between the *decapentaplegic* (*dpp*) promoter and its enhancer, which only activates the *dpp* gene, but not other genes that are located closer to it [196]. Erythroid-specific long-range interactions have been observed *in vivo* between the active murine  $\beta$ -globin gene and the locus control region (LCR) [197]. These long-range interactions of the  $\beta$ -globin gene were not observed in non-expressing brain cells. High-throughput imaging of thousands of transparent transgenic zebrafish embryos (which were injected with about two hundred combinations of enhancer–core promoter pairs driving the expression of the GFP reporter gene), demonstrated the specificity of individual enhancer–promoter interactions and underscored the importance of the core promoter sequence in these interactions [198]. Taken together, these results demonstrate distinct compatibilities of enhancers to their cognate promoters and the importance of the core promoters in the regulation of enhancer–promoter interactions.

While a few studies in *Drosophila* demonstrated the involvement of proximal–promoter elements in enhancer specificity [199,200], there are multiple examples of enhancer–promoter communications that are affected by specific core promoter elements. Promoter competition experiments revealed that both the AE1 enhancer from the *Drosophila* Antennapedia gene complex and the IAB5 enhancer from the Bithorax gene complex preferentially activate TATA-containing promoters when challenged with linked TATA-less promoters [201]. Nevertheless, both enhancers were able to activate transcription from a TATA-less promoter in reporters that lacked a linked TATA-containing promoter [201]. Enhancer–promoter specificity was first demonstrated in transgenic *Drosophila* sister lines that contain a DPE- or a TATA-dependent reporter gene at precisely the same genomic position relative to the enhancer [202]. Remarkably, this study identified enhancers that can discriminate between core promoters that are dependent on a TATA or a DPE motif. Furthermore, Caudal, a sequence-specific transcription factor and a key regulator of the *Drosophila* HOX gene network, activates transcription with a preference for a DPE motif relative to the TATA-box [99]. More recently, Zehavi et al. [15] analyzed the *Drosophila* dorsal–ventral developmental gene network that is regulated by the sequence-specific transcription factor Dorsal, and discovered that the majority of Dorsal target genes contain DPE sequence motifs. The DPE motif is functional in multiple Dorsal target genes, as mutation of the DPE leads to a loss of transcriptional activity. Moreover, the analysis of hybrid enhancer–promoter constructs of Dorsal targets reveals that the core promoter plays a pivotal role in the transcriptional output [103].

High-throughput analyses of enhancers in diverse biological systems have led to a wealth of information with regards to long-range

enhancer–promoter interactions and three-dimensional chromatin landscapes. We highlight several remarkable findings below. First, most of the enhancer–promoter interaction loops of regulated genes are distal, and are not localized at the nearest promoter as originally considered [203–205]. Second, enhancer looping enables cooperative regulation of genes of the same biological process by organizing them in physical proximity [203,205]. This may indicate a similar core promoter composition among these gene networks or gene clusters (as previously described for the Hox and dorsal–ventral developmental gene regulatory networks [15,99]).

A recently developed genome-wide screen termed STARR-seq (self-transcribing active regulatory region sequencing) identified thousands of enhancers that could activate transcription of a synthetic promoter containing four core promoter elements in a single promoter - the TATA-box, Inr, MTE and DPE motifs [206]. Notably, enhancers near ribosomal protein genes were under-represented among the enhancers identified in this study, which could be due to the fact that the majority of ribosomal protein gene promoters are regulated via the TCT core promoter element [14,194,206].

Remarkably, both the Furlong lab analyzing enhancer three-dimensional contacts during *Drosophila* embryogenesis, and the Ren lab analyzing long-range chromatin interactions in human cells, discovered that the majority of enhancer interactions remain unchanged during marked developmental transitions or activation following gene induction, respectively [203,207]. This “on-hold” enhancer–promoter connections, may be preparing the cell for rapid activation of transcription. The Furlong lab discovered that the pre-existing loops are associated with paused Pol II and proposed a model where through transcription factor–enhancer occupancy, an enhancer loops towards the promoter and polymerase is recruited, but paused in the majority of cases (Pol II pausing is discussed below). They suggest that the subsequent recruitment of transcription factor(s) or additional enhancers at preformed enhancer–promoter interaction hubs could trigger activation by releasing Pol II pausing [207]. Notably, enhancer–promoter interactions analyzed in these studies involve active promoters, with high enrichment for H3K27ac and H3K4me3, and active enhancers, defined by H3K27ac, Pol II and H3K79me3, indicating similarities in 3D regulatory principles from flies to humans [203,204,207].

Strikingly, the Stark lab has recently demonstrated that distinct sets of enhancers activate transcription with core promoter specificity using two types of *Drosophila* cultured cells [208]. They used the core promoter of a ribosomal protein gene driven by the TCT motif, as a representative of housekeeping promoters, and a synthetic promoter (derived from the *even skipped* promoter), which contains four core promoter elements in a single promoter - the TATA-box, Inr, MTE and DPE motifs, as a representative of developmental promoters. Thousands of enhancers exhibit a marked specificity to one of the two core promoters - the housekeeping promoter or the developmental promoter. Interestingly, TSSs next to housekeeping enhancers were enriched in Ohler motifs 1, 5, 6 and 7 (consistent with the ubiquitous expression and housekeeping functions of these genes), whereas TSSs next to developmental enhancers were enriched in TATA box, Inr, MTE and DPE motifs (which are associated with cell-type-specific gene expression).

Taken together, these observations strengthen the concept that the core promoter composition is not only a pivotal component in basal transcription and initiation, but also an active regulator of transcription that is instrumental for activating developmental and housekeeping gene regulatory programs via sequence-encoded enhancer–promoter specificity.

## 6. Transcription initiation, Pol II recycling and steps in between: the crosstalk between the core promoter and other modules in the transcription cycle

Apart from transcription initiation, Pol II-driven transcription cycle contains additional steps: elongation and termination. These steps

contain at least eight transition points at which transcription is regulated by multiple dedicated factors, and each can be rate limiting (reviewed in [209,210]). Moreover, maturation of mRNA precursors occurs co-transcriptionally [211]. Below, we briefly describe these highly regulated steps with a focus on the direct or indirect role of the core promoter.

### 6.1. Timing and synchrony – Pol II pausing and productive elongation

Early elongation, following proper transcription initiation and preceding productive elongation, contains two sequential steps: promoter-escape and promoter-proximal pausing of Pol II. Pol II pausing is a highly regulated step, which is characterized by accumulation of Pol II, typically at 20–60 nucleotides downstream of the TSS (reviewed in [210,212,213]). The transition from initiation to early elongation is regulated by multiple factors and phosphorylation events of the heptad repeats within the C-terminal domain (CTD) of the largest subunit of Pol II. The CTD is mostly unphosphorylated when Pol II is recruited to the promoter. Serine 5 (Ser5) of the CTD is then phosphorylated by TFIIF, which causes destabilization of the interaction between Pol II and other PIC components and thus, permits promoter escape and early elongation. Following Ser5 phosphorylation, association of DRB sensitivity-inducing factor (DSIF) and Negative elongation factor (NELF) complexes with the phosphorylated Pol II leads to pausing at the promoter-proximal region [214]. Next, positive transcription elongation factor b (P-TEFb) complex phosphorylates the Ser2 residue of the Ser5-phosphorylated CTD, and the DSIF and NELF factors. These post-translational modifications result in productive elongation (reviewed in [210,212,213]).

Pol II pausing was originally identified in *Drosophila heat-shock* and human *c-myc* genes [215–218]. Although Pol II pausing was originally considered to be restricted to a few specific genes, nowadays, the pausing of Pol II appears to be a common step in the transcription process of multiple genes, and generally prevalent in metazoans [22,219–223]. Specifically, multiple genome-wide assays and studies *in vitro* and *in vivo*, mostly in *Drosophila*, showed that the Pol II pausing has a role in facilitating metazoan developmental control genes and genes that respond to environmental stimuli ([224] and refs therein). Thus, Pol II pausing contributes to developmental dynamics, along with designated transcription initiation programs [225,226]. Notably, Pol II pausing under normal growth conditions is very rare in *C. elegans* [227] and this is consistent with the lack of a *C. elegans* NELF [214], which contributes to pausing in other organisms. It was previously argued that Pol II pausing prepares genes for a rapid and synchronous induction. Recent studies, however, suggest that paused Pol II is not absolutely required for rapid gene induction, as genes in which Pol II is not paused, can be induced just as quickly, and to even higher levels than paused genes ([213,224] and refs therein). Promoters regulated by pausing possess a distinct chromatin architecture that may facilitate the plasticity of gene expression in response to signaling events [213]. Notably, paused Pol II complexes were recently shown to be more stable than originally considered, and thus, pausing may serve as a time-window to integrate regulatory signals [228]. There are two known sequence-specific transcription factors that regulate pausing: the GAGA factor (GAF) [215, 216,221,229] and the more recently identified M1BP factor [112].

Pausing allows synchronous gene expression of developmentally regulated genes following their induction during embryogenesis [224, 230–233]. Differences in synchronicity are most likely due to the core promoter composition, as demonstrated by promoter-swapping experiments [231] and the relationship between Pol II pausing and core promoter sequence during *Drosophila* development [230,234].

The positive elongation factor P-TEFb controls NFκB target genes driven by TATA-containing promoters, whereas the negative elongation factor DSIF controls weak TATA and TATA-less genes [235]. Interestingly, *Drosophila* TATA-dependent promoters are associated with a low degree of pausing [230,234], suggesting that the TATA box prevents Pol II

pausing and promotes P-TEFb activity, leading to a more productive elongation [235].

Remarkably, the Levine lab has shown that at least one fourth of paused *Drosophila* promoters contain a shared sequence motif, the “pause button” (PB), whose consensus (KCGRWCG) [236] is similar to that of the DPE (DSWYVY) [10]. The PB motif is typically located between +25 and +35 (somewhat overlapping the DPE, although it has a wider distribution with regards to its location relative to the TSS). Over one-fifth of the paused *Drosophila* promoters are enriched for the DPE, MTE and PB core promoter motifs, all of which are located close to the pause site [236]. Notably, 75% of the genes in the dorsal-ventral network were identified as paused genes [236]. Over two thirds of Dorsal target genes contain a DPE motif [15]. These correlations, in addition to the fact that PB and DPE are GC-rich and share the ‘GGWC’ sub-consensus, and that both motifs overlap with the paused Pol II (see above), may indicate that the DPE, as opposed to the TATA box, could contribute to Pol II pausing. The Adelman lab has later found out that both the DPE and PB precisely align with the peak of Pol II pausing [222].

In addition, a current study indicates that whereas proximity of Pol II pausing to the TSSs is correlated with focused initiation, pausing at dispersed promoters is located more distally, and with a wider pattern [224,237]. Moreover, it seems that in contrast to dispersed promoters, Pol II pausing at focused promoters is not dependent on nucleosome regulation. When the core promoter elements are not located at optimal position, or do not match the consensus sequence, pausing appears to be weaker and located more downstream (+60 to +80) than its typical location. Thus, initiation modes and core promoter architecture affect the strength and location of pausing [237].

It is well known that enhancers have a major effect on activity and synchrony of gene expression in development. Remarkably, Lagha et al. [231] used a promoter swapping strategy and advanced imaging methods and discovered that promoters of key developmental genes play a pivotal role in pausing, which in turn determines the “time to synchrony”- the time it takes to achieve coordinated gene expression in over 50% of the nuclei in the developing *Drosophila* embryo. The authors demonstrate that substitutions of paused promoters (e.g. *tup*), which show rapid and synchronous activity, with non-paused promoters (such as *pnr*), result in slow and stochastic activation of gene expression. Moreover, elements associated with pausing (e.g. GAGA) influence the timing and synchrony of the gene expression. The synchronous activation is essential for proper mesoderm invagination in the developing *Drosophila* embryo. They provide evidence for a positive correlation between pausing, synchrony and gene expression levels, which are necessary for morphogenesis. Hence, it is the promoter, and not the enhancer, that determines the levels of paused Pol II and the synchrony of gene activation [231,232].

To summarize, these studies provide evidence regarding different aspects of regulation of Pol II pausing via the core promoter. However, additional biochemical studies are needed to elucidate the mechanisms underlying pausing.

### 6.2. Termination, polyadenylation and recycling of Pol II – back to square one

The promoter and terminator modules define the boundaries of the transcribed region of protein-coding genes. Transcription termination includes dephosphorylation of the Pol II CTD, its disassociation from the 3'-end and cleavage of the pre-mRNA. Furthermore, this highly regulated event is coupled with the 3'-end polyadenylation processing [238]. Numerous factors in multi-subunit protein complexes and several RNA elements mediate the termination/polyadenylation processes, including two central complexes: cleavage and polyadenylation specificity factor (CPSF) and cleavage stimulation factor (CstF) [239,240]. Although several factors are shared, the termination mechanism for metazoan replication-dependent core histone genes, which are not

polyadenylated, is different than the termination mechanism of polyadenylated genes (reviewed in [239,241,242]).

There are mutual links between transcription initiation and termination/polyadenylation. It should be noted that although many studies were done using yeast, we focus here on metazoan transcriptional termination. The CPSF complex was first immunoprecipitated and co-purified with holo-TFIID from nuclear extracts of human cell-lines almost twenty years ago [243]. The authors showed that CPSF is recruited to the core promoter by TFIID and later dissociates from TFIID and continues to be associated with the elongating Pol II and later with the polyA site. Specifically, the CPSF-160 subunit mainly interacts with TAF5, TAF7 and TAF12, but not with TAF1, TAF10 and TAF15 and minimally, if at all, with TBP. Overexpression of TBP reduced polyadenylation of transcripts initiated from a TATA-containing promoter, while both polyadenylated transcripts and non-polyadenylated transcripts that initiated from a TATA-less promoter were unaffected [243]. Furthermore, the recruitment of CstF by TFIIB to the core promoter through PIC assembly was also demonstrated ([244] and refs therein). Thus, subunits of the main termination factors CPSF and CstF are brought to the PIC and transferred to Pol II, which eventually leads to transcription termination. Moreover, components of the core histone termination machinery were also found associated with histone promoters ([239] and refs therein). Nevertheless, it was previously observed that the termination/polyadenylation machinery influences PIC assembly and the efficiency of transcription re-initiation through Pol II recycling ([245] and refs therein). These transcription initiation-termination/polyadenylation connections are mediated by two different chromatin and genomic mechanisms: gene looping from 3'-end processing sites to core promoters, which brings both modules into spatial and physical proximity, and compartmentalization of genes into "gene factories" [3,239,246]. It is noteworthy that these connections and couplings are conserved throughout eukaryotes. In this regard, it is possible that the PIC assemblies and 3'-associated machineries of the core histone genes are particularly specialized, as compared to other protein-encoding genes [137,239].

In a recent paper, Oktaba et al. [247] demonstrated that the promoters are involved in the regulation of alternative cleavage and polyadenylation. The nuclear RNA-binding protein embryonic lethal abnormal visual system (ELAV) is known to inhibit the canonical polyadenylation processing at the 3' UTRs of genes, which causes Pol II read-through and 3' UTR extension, during the development of the nervous system in *Drosophila* and vertebrates. The authors provide evidence that ELAV-mediated 3' UTR extension is dependent on the promoter and Pol II pausing in the developing *Drosophila* nervous system [247]. Using double-labeling assays and swapping promoters experiments, they show that only reporter constructs that were driven by promoters of known extended genes *in vivo*, produced extended transcripts in transgenic *Drosophila* embryos. Ectopic expression of ELAV in non-neural tissues resulted in the induction of 3' UTR extension. Moreover, sequence analysis of 252 neural-specific transcripts with 3' UTR extensions revealed the enrichment of the GAGA motif and Pol II pausing. Indeed, reduced 3' UTR extension levels were observed in GAGA-binding protein Tritorax-like (Trl)-mutant *Drosophila* embryos. ChIP-seq analysis revealed the enrichment of ELAV in promoter regions of extended genes, as well as in 3' UTRs and introns. Thus, ELAV is selectively recruited to the 3' UTRs of extended genes through paused Pol II promoters, perhaps via looping between the promoters and the termination regions. Taken together, the above studies strengthen the link between transcription initiation and termination and the pivotal role of the promoter in this linkage.

## 7. Is the dogma really composed of sequential steps? The transcription-translation linkage

Traditionally, eukaryotic translation has been defined as a separate process that is independent from transcription. However, the

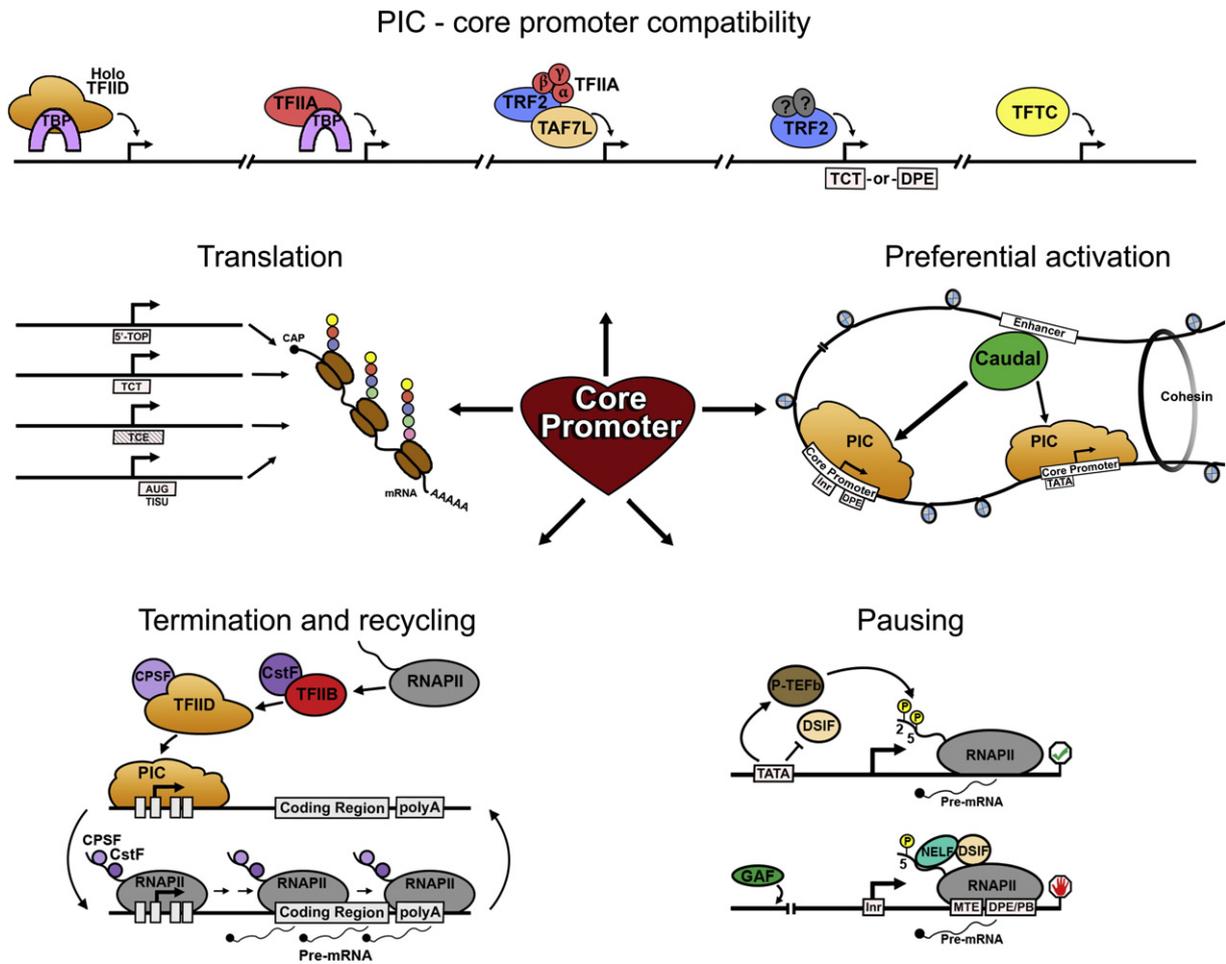
translation machinery depends on mRNA-maturation processing, such as the m7G cap structure at the 5' UTR and its associated protein complexes [248]. These complexes recruit the small ribosomal subunit that in turn reaches the first codon, AUG, via a 5' UTR scanning mechanism (reviewed in [249]). A common element for translation initiation is the Kozak element (RCCAUGG), which contains the AUG [250,251]. In addition to this well-defined translational initiator, a distinguished element, Translation Initiator of Short 5' UTR (TISU), was recently identified. Remarkably, this element is important for transcription and initiation of translation of a specific set of genes [252]. The TISU is found in 4.5% of the mammalian protein-coding genes, with consensus sequence of 'SAASATGGCGGC' with rigid core-sequence of 'ATG' located at +5 to +30, and particularly positioned around the +10 relative to the TSS [60,252,253]. This core promoter element is enriched in TATA-less promoters of genes mostly involved in cellular functions such as protein metabolism and RNA processing. As a transcriptional element, it was shown to be necessary for transcription and its function was mediated, at least in part, by YY1 [250,252]. As a translational element, it was defined as an optimized translation initiator for protein-coding genes possessing a very short 5' UTR (median of 12 nt) that mediates translation in cap-dependent but ribosomal-scanning independent manner, as opposed to the Kozak sequence [250,253]. The 5'-TOP, a mammalian pyrimidine-tract regulatory element, was previously characterized as a transcriptional and translational element [78,79,254,255]. It was identified as a core promoter motif used as a transcriptional "initiator" in many protein-biogenesis genes, and its translational activity is critical under stress conditions. The translational control element (TCE) [256], another transcription/translation element, was previously shown to regulate translation in *Drosophila* testes [257]. Katzenberger et al. [258] recently showed that the overlapping transcriptional motifs, testis element 1 (TE1) and testis element 2 (TE2), which are overrepresented in testis-specific core promoters, are together identical (TE1/2 motif) to the original TCE. Thus, this element is a transcriptional element, too. The TCE is identified as a transcriptional element in 45% of *Drosophila* testis-specific genes that are driven by focused promoters. Its consensus sequence is "CTCAAATTT", with enrichment in the -5 to +25 region, but without precise location relative to the TSS [258].

Hence, these three core promoter motifs play pivotal roles in both transcription and translation of distinct sets of genes. Moreover, correlations between the TATA box and different features of genes (e.g. gene length) have been observed [259]. This co-regulation of these processes raises questions regarding the interplay between transcription and translation, such as: Do downstream core promoter elements affect the translation of these genes? Based on the fact that the 5' UTRs of some organisms are short, are these elements evolutionarily conserved? Indeed, a recent study reveals general associations and co-occurrence between translational and transcriptional regulatory trends and features, including core promoter composition [260]. Taken together, the core promoter region is, at least in part, a central intersection for coordinating transcription and translation.

## 8. Discussion and future perspectives

In this review, we discussed diverse aspects of regulation of gene expression, particularly in metazoans, with an emphasis on the core promoter. We highlighted the complexity of the core promoter architecture. Furthermore, we presented its intricate connections and its pivotal influences on different steps of transcription: initiation, elongation, termination, polyadenylation and finally, translation (Fig. 3). Moreover, we would like to raise a few issues that are directly related to the core promoter but were not mentioned above.

First, in addition to the diversity of core promoter elements and the relationships between them, nucleotide polymorphism in the core promoter affects its activity including its binding by the PIC components. Multiple lines of evidence point towards polymorphisms in many human promoters, particularly in the TATA box sequence. These TATA



**Fig. 3.** Schematic model depicting the pivotal role of the core promoter module in diverse molecular events and stages of gene expression. The core promoter is important for (clockwise): basal transcription initiation and PIC- core promoter compatibility reflected by different compositions of basal transcription factors, which comprise diverse PIC architectures (top); enhancer-promoter compatibility (which is schematically represented by the preferential activation of DPE-dependent promoters by Caudal); promoter-proximal Pol II pausing (where different core promoter elements support Pol II pausing vs. elongating states); termination/polyadenylation and Pol II recycling (where subunits of the main termination factors CPSF and CstF are recruited to the PIC at the core promoter and continue to be associated with the elongating Pol II and later with the polyA site, which eventually leads to transcription termination. Furthermore, 3' end formation has been shown to stimulate transcription initiation); and translation (via core promoter elements that play a role in both transcription and translation). Please see the main text for detailed explanations.

box substitutions can affect TBP binding and core promoter activity, and are associated with human diseases ([261], reviewed in [262]). It is expected that like TATA box polymorphism, polymorphisms in other elements exist, and may be clinically relevant.

Second, the enhancer-promoter interactome seems to be a much more complex landscape than previously considered. In agreement with that, promoter-promoter interactions have recently been found [263]. These interactions behave as enhancer-promoter interactions, where one promoter is able to act as an enhancer of another. Hence, hypothetical, more complicated hierarchies of direct and indirect interactions between enhancers and promoters could be achieved (e.g. generating an enhancer-promoter-promoter hub).

Moreover, an additional regulatory aspect that is associated with enhancers is the discovery of enhancer-derived RNAs (eRNAs). This class of ncRNAs was only discovered a few years ago in humans [264]. eRNAs are short-lived, 5'-capped transcripts produced from enhancer regions. Their expression is correlated with histone marks of active enhancers (H3K4me1 and H3K27ac), and they are enriched for transcription factors, co-activators (such as p300/CBP), basal transcription factors and Ser5-phosphorylated Pol II. eRNAs are preferentially found in enhancers that contact their target promoters through enhancer-looping, and it is suggested that these transcripts play a role in generating or maintaining enhancer-promoter-loops and in facilitating the recruitment of sequence-specific transcription factors, chromatin remodeling

or chromatin modifying complexes to the targeted promoters [53]. Additionally, eRNAs are associated with several signaling pathways ([53,54] and refs therein). Although eRNAs are extensively investigated, also by high-scale methodologies [265], little is known about their core promoter compositions and their TSS architectures [55]. Hence, one of the future goals should be an in-depth investigation of the core promoter architectures of eRNAs and their transcriptional machineries.

Actually, in agreement with the current knowledge that many active mammalian promoters are bidirectional [22,57], a study published several months ago revealed shared architectures of bidirectional initiations at promoters and active enhancers [55]. On one hand, similar trends and profiles of transcription factor binding, nucleosome positioning, histone marks and similar frequencies of sequence motifs such as the TATA box, BRES and Inr (YR only) were present in both promoters and transcribed enhancers. On the other, these modules differ in the stability of the transcripts that they synthesize in each direction: promoters give rise to stable transcripts in the sense direction, whereas promoter upstream antisense RNA and enhancer RNAs are rapidly degraded [55]. This unifying architecture of TSSs [266] along with recent findings (e.g. promoter-promoter interactions) challenge the traditional classification of promoters and enhancers (see also [267]). It is noteworthy that Core et al. [55] indicated that although there are distinct pause modes, which include proximal focused pausing and distal dispersed pausing (see also [237]), the length between the bidirectional TSS

pairs and the peaks of TFIIB are not affected. This high-resolution analysis of nascent RNAs might also imply that the high frequency of dispersed mammalian core promoters observed previously, represents multiple independent initiation sites acting as enhancers for neighboring promoters [55]. Thus, the phenomena of dispersed mammalian promoters might be less abundant than originally perceived. Taken together, the growing body of evidence indicates that the core promoter lies at the heart of gene expression.

### Transparency document

The Transparency document associated with this article can be found, in the online version.

### Acknowledgments

We thank Ron Even for graphic design assistance. We thank Jim Kadonaga, Uwe Ohler, Sascha Duttke, Anna Sloutskin, Hila Shir-Shapira and Racheli Harshish for critical reading of the manuscript. Core promoter-related research in the Juven-Gershon lab is supported by grants from the Israel Science Foundation (no. 798/10), the European Union Seventh Framework Programme (Marie Curie International Reintegration Grant; no. 256491), the United States-Israel Binational Science Foundation (no. 2009428; joint with James T. Kadonaga) and the German-Israeli Foundation for Scientific Research and Development (no. I-1220-363.13/2012; joint with Eileen E.M. Furlong).

The authors declare that there are no potential conflicts of interest.

### References

- [1] E. Splinter, W. de Laat, The complex transcription regulatory landscape of our genome: control in three dimensions, *EMBO J.* 30 (2011) 4345–4355.
- [2] X. Dong, M.C. Greven, A. Kundaje, S. Djebali, J.B. Brown, C. Cheng, T.R. Gingeras, M. Gerstein, R. Guigo, E. Birney, Z. Weng, Modeling gene expression using chromatin features in various cellular contexts, *Genome Biol.* 13 (2012) R53.
- [3] J. Shandilya, S.G. Roberts, The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling, *Biochim. Biophys. Acta* 1819 (2012) 391–400.
- [4] M.C. Thomas, C.M. Chiang, The general transcription machinery and general cofactors, *Crit. Rev. Biochem. Mol. Biol.* 41 (2006) 105–178.
- [5] J.E. Butler, J.T. Kadonaga, The RNA polymerase II core promoter: a key component in the regulation of gene expression, *Genes Dev.* 16 (2002) 2583–2592.
- [6] J.T. Kadonaga, Perspectives on the RNA polymerase II core promoter, *Wiley Interdiscip. Rev. Dev. Biol.* 1 (2012) 40–51.
- [7] B. Li, M. Carey, J.L. Workman, The role of chromatin during transcription, *Cell* 128 (2007) 707–719.
- [8] E. Valen, A. Sandelin, Genomic and chromatin signals underlying transcription start-site selection, *Trends Genet.* 27 (2011) 475–485.
- [9] J. Zlatanova, A. Thakar, H2A.Z: view from the top, *Structure* 16 (2008) 166–179.
- [10] T. Juven-Gershon, J.T. Kadonaga, Regulation of gene expression via the core promoter and the basal transcriptional machinery, *Dev. Biol.* 339 (2010) 225–229.
- [11] B. Lenhard, A. Sandelin, P. Carninci, Metazoan promoters: emerging characteristics and insights into transcriptional regulation, *Nat. Rev. Genet.* 13 (2012) 233–245.
- [12] N.D. Heintzman, B. Ren, The gateway to transcription: identifying, characterizing and understanding promoters in the eukaryotic genome, *Cell. Mol. Life Sci.* 64 (2007) 386–400.
- [13] T. Juven-Gershon, J.Y. Hsu, J.W. Theisen, J.T. Kadonaga, The RNA polymerase II core promoter - the gateway to transcription, *Curr. Opin. Cell Biol.* 20 (2008) 253–259.
- [14] T.J. Parry, J.W. Theisen, J.Y. Hsu, Y.L. Wang, D.L. Corcoran, M. Eustice, U. Ohler, J.T. Kadonaga, The TCT motif, a key component of an RNA polymerase II transcription system for the translational machinery, *Genes Dev.* 24 (2010) 2013–2018.
- [15] Y. Zehavi, O. Kuznetsov, A. Ovadia-Shochat, T. Juven-Gershon, Core promoter functions in the regulation of gene expression of *Drosophila* dorsal target genes, *J. Biol. Chem.* 289 (2014) 11993–12004.
- [16] A. Sandelin, P. Carninci, B. Lenhard, J. Ponjavic, Y. Hayashizaki, D.A. Hume, Mammalian RNA polymerase II core promoters: insights from genome-wide studies, *Nat. Rev. Genet.* 8 (2007) 424–436.
- [17] T. Ni, D.L. Corcoran, E.A. Rach, S. Song, E.P. Spana, Y. Gao, U. Ohler, J. Zhu, A paired-end sequencing strategy to map the complex landscape of transcription initiation, *Nat. Methods* 7 (2010) 521–527.
- [18] M.A. Frohman, M.K. Dush, G.R. Martin, Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer, *Proc. Natl. Acad. Sci. U. S. A.* 85 (1988) 8998–9002.
- [19] T. Shiraki, S. Kondo, S. Katayama, K. Waki, T. Kasukawa, H. Kawaji, R. Kodzius, A. Watahiki, M. Nakamura, T. Arakawa, S. Fukuda, D. Sasaki, A. Podhajnska, M. Harbers, J. Kawai, P. Carninci, Y. Hayashizaki, Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage, *Proc. Natl. Acad. Sci. U. S. A.* 100 (2003) 15776–15781.
- [20] P.G. Giresi, J. Kim, R.M. McDaniell, V.R. Iyer, J.D. Lieb, FAIRE (formaldehyde-assisted isolation of regulatory elements) isolates active regulatory elements from human chromatin, *Genome Res.* 17 (2007) 877–885.
- [21] T.S. Furey, ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions, *Nat. Rev. Genet.* 13 (2012) 840–852.
- [22] L.J. Core, J.J. Waterfall, J.T. Lis, Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters, *Science* 322 (2008) 1845–1848.
- [23] Z. Wang, M. Gerstein, M. Snyder, RNA-Seq: a revolutionary tool for transcriptomics, *Nat. Rev. Genet.* 10 (2009) 57–63.
- [24] N.L. Washington, E.O. Stinson, M.D. Perry, P. Ruzanov, S. Conrino, R. Smith, Z. Zha, R. Lyne, A. Carr, P. Lloyd, E. Kephart, S.J. McKay, G. Micklem, L.D. Stein, S.E. Lewis, The modENCODE Data Coordination Center: lessons in harvesting comprehensive experimental details, *Database (Oxford)* 20112011. (bar023).
- [25] The ENCODE (ENCyclopedia Of DNA Elements) Projec, *Science* 306 (2004) 636–640.
- [26] A.R. Forrest, H. Kawaji, M. Rehli, J.K. Baillie, M.J. de Hoon, T. Lassmann, M. Itoh, K.M. Summers, H. Suzuki, C.O. Daub, J. Kawai, P. Heutink, W. Hide, T.C. Freeman, B. Lenhard, V.B. Bajic, M.S. Taylor, V.J. Makeev, A. Sandelin, D.A. Hume, P. Carninci, Y. Hayashizaki, A promoter-level mammalian expression atlas, *Nature* 507 (2014) 462–470.
- [27] P. Carninci, A. Sandelin, B. Lenhard, S. Katayama, K. Shimokawa, J. Ponjavic, C.A. Sempke, M.S. Taylor, P.G. Engstrom, M.C. Frith, A.R. Forrest, W.B. Alkema, S.L. Tan, C. Plessy, R. Kodzius, T. Ravasi, T. Kasukawa, S. Fukuda, M. Kanamori-Katayama, Y. Kitazume, H. Kawaji, C. Kai, M. Nakamura, H. Konno, K. Nakano, S. Mottagui-Tabar, P. Arner, A. Chesi, S. Gustincich, F. Persichetti, H. Suzuki, S.M. Grimmond, C.A. Wells, V. Orlando, C. Wahlestedt, E.T. Liu, M. Harbers, J. Kawai, V.B. Bajic, D.A. Hume, Y. Hayashizaki, Genome-wide analysis of mammalian promoter architecture and evolution, *Nat. Genet.* 38 (2006) 626–635.
- [28] E.A. Rach, H.Y. Yuan, W.H. Majoros, P. Tomancak, U. Ohler, Motif composition, conservation and condition-specificity of single and alternative transcription start sites in the *Drosophila* genome, *Genome Biol.* 10 (2009) R73.
- [29] V.B. Bajic, S.L. Tan, A. Christoffels, C. Schonbach, L. Lipovich, L. Yang, O. Hofmann, A. Kruger, W. Hide, C. Kai, J. Kawai, D.A. Hume, P. Carninci, Y. Hayashizaki, Mice and men: their promoter properties, *PLoS Genet.* 2 (2006) e54.
- [30] R.A. Hoskins, J.M. Landolin, J.B. Brown, J.E. Sandler, H. Takahashi, T. Lassmann, C. Yu, B.W. Booth, D. Zhang, K.H. Wan, L. Yang, N. Boley, J. Andrews, T.C. Kaufman, B.R. Graveley, P.J. Bickel, P. Carninci, J.W. Carlson, S.E. Celniker, Genome-wide analysis of promoter architecture in *Drosophila melanogaster*, *Genome Res.* 21 (2011) 182–192.
- [31] M. Baumann, J. Pontiller, W. Ernst, Structure and basal transcription complex of RNA polymerase II core promoters in the mammalian genome: an overview, *Mol. Biotechnol.* 45 (2010) 241–247.
- [32] S.J. Cooper, N.D. Trinklein, E.D. Anton, L. Nguyen, R.M. Myers, Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome, *Genome Res.* 16 (2006) 1–10.
- [33] T.H. Kim, L.O. Barrera, M. Zheng, C. Qu, M.A. Singer, T.A. Richmond, Y. Wu, R.D. Green, B. Ren, A high-resolution map of active promoters in the human genome, *Nature* 436 (2005) 876–880.
- [34] M.C. Frith, Explaining the correlations among properties of mammalian promoters, *Nucleic Acids Res.* 42 (2014) 4823–4832.
- [35] J.A. Stamatoyannopoulos, Illuminating eukaryotic transcription start sites, *Nat. Methods* 7 (2010) 501–503.
- [36] N. Adachi, M.R. Lieber, Bidirectional gene organization: a common architectural feature of the human genome, *Cell* 109 (2002) 807–809.
- [37] J.C. Ame, V. Schreiber, V. Fraulob, P. Dolle, G. de Murcia, C.P. Niedergang, A bidirectional promoter connects the poly(ADP-ribose) polymerase 2 (PARP-2) gene to the gene for RNase P RNA. structure and expression of the mouse PARP-2 gene, *J. Biol. Chem.* 276 (2001) 11092–11099.
- [38] A.S. Orekhova, P.M. Rubtsov, Bidirectional promoters in the transcription of mammalian genomes, *Biochem. Biokhim.* 78 (2013) 335–341.
- [39] V. Gotea, H.M. Petrykowska, L. Elnitski, Bidirectional promoters as important drivers for the emergence of species-specific transcripts, *PLoS ONE* 8 (2013) e57323.
- [40] M.Q. Yang, L.L. Elnitski, Diversity of core promoter elements comprising human bidirectional promoters, *BMC Genomics* 9 (Suppl. 2) (2008) S3.
- [41] P.G. Engstrom, H. Suzuki, N. Ninomiya, A. Akalin, L. Sessa, G. Lavorgna, A. Brozzi, L. Luzi, S.L. Tan, L. Yang, G. Kunarso, E.L. Ng, S. Batalov, C. Wahlestedt, C. Kai, J. Kawai, P. Carninci, Y. Hayashizaki, C. Wells, V.B. Bajic, V. Orlando, J.F. Reid, B. Lenhard, L. Lipovich, Complex Loci in human and mouse genomes, *PLoS Genet.* 2 (2006) e47.
- [42] G. Wang, K. Qi, Y. Zhao, Y. Li, L. Juan, M. Teng, L. Li, Y. Liu, Y. Wang, Identification of regulatory regions of bidirectional genes in cervical cancer, *BMC Med. Genet.* 6 (Suppl. 1) (2013) S5.
- [43] M.U. Kaikkonen, M.T. Lam, C.K. Glass, Non-coding RNAs as regulators of gene expression and epigenetics, *Cardiovasc. Res.* 90 (2011) 430–440.
- [44] P. Kapranov, J. Cheng, S. Dike, D.A. Nix, R. Duttagupta, A.T. Willingham, P.F. Stadler, J. Hertel, J. Hackermuller, I.L. Hofacker, I. Bell, E. Cheung, J. Drenkow, E. Dumais, S. Patel, G. Helt, M. Ganes, S. Ghosh, A. Piccolboni, V. Sementchenko, H. Tammana, T.R. Gingeras, RNA maps reveal new RNA classes and a possible function for pervasive transcription, *Science* 316 (2007) 1484–1488.
- [45] W. Wei, V. Pelechano, A.J. Jarvelin, L.M. Steinmetz, Functional consequences of bidirectional promoters, *Trends Genet.* 27 (2011) 267–276.
- [46] Y. He, B. Vogelstein, V.E. Velculescu, N. Papadopoulos, K.W. Kinzler, The antisense transcriptomes of human cells, *Science* 322 (2008) 1855–1857.

- [47] P. Preker, J. Nielsen, S. Kammler, S. Lykke-Andersen, M.S. Christensen, C.K. Mapendano, M.H. Schierup, T.H. Jensen, RNA exosome depletion reveals transcription upstream of active human promoters, *Science* 322 (2008) 1851–1854.
- [48] A.C. Seila, J.M. Calabrese, S.S. Levine, G.W. Yeo, P.B. Rahl, R.A. Flynn, P.A. Sharp, Divergent transcription from active promoters, *Science* 322 (2008) 1849–1851.
- [49] S. Buratowski, Transcription. Gene expression—where to start? *Science* 322 (2008) 1804–1805.
- [50] P. Richard, J.L. Manley, How bidirectional becomes unidirectional, *Nat. Struct. Mol. Biol.* 20 (2013) 1022–1024.
- [51] A.E. Almada, X. Wu, A.J. Kriz, C.B. Burge, P.A. Sharp, Promoter directionality is controlled by U1 snRNP and polyadenylation signals, *Nature* 499 (2013) 360–363.
- [52] E. Ntini, A.J. Jarvelin, J. Bornholdt, Y. Chen, M. Boyd, M. Jorgensen, R. Andersson, I. Hoof, A. Schein, P.R. Andersen, P.K. Andersen, P. Preker, E. Valen, X. Zhao, V. Pelechano, L.M. Steinmetz, A. Sandelin, T.H. Jensen, Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality, *Nat. Struct. Mol. Biol.* 20 (2013) 923–928.
- [53] F. Lai, R. Shiekhattar, Enhancer RNAs: the new molecules of transcription, *Curr. Opin. Genet. Dev.* 25 (2014) 38–42.
- [54] M.T. Lam, W. Li, M.G. Rosenfeld, C.K. Glass, Enhancer RNAs and regulated transcription programs, *Trends Biochem. Sci.* 39 (2014) 170–182.
- [55] L.J. Core, A.L. Martins, C.G. Danko, C.T. Waters, A. Siepel, J.T. Lis, Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers, *Nat. Genet.* 46 (2014) 1311–1320.
- [56] M. Uesaka, O. Nishimura, Y. Go, K. Nakashima, K. Agata, T. Imamura, Bidirectional promoters are the major source of gene activation-associated non-coding RNAs in mammals, *BMC Genomics* 15 (2014) 35.
- [57] S.H. Duttke, S.A. Lacadie, M.M. Ibrahim, C.K. Glass, D.L. Corcoran, C. Benner, S. Heinz, J.T. Kadonaga, U. Ohler, Human promoters are intrinsically directional, *Mol. Cell* 57 (2015) 674–684.
- [58] F. Muller, L. Tora, The multicoloured world of promoter recognition complexes, *EMBO J.* 23 (2004) 2–8.
- [59] L. Tora, A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription, *Genes Dev.* 16 (2002) 673–675.
- [60] R. Dikstein, The unexpected traits associated with core promoter elements, *Transcription* 2 (2011) 201–206.
- [61] J.T. Kadonaga, The DPE, a core promoter element for transcription by RNA polymerase II, *Exp. Mol. Med.* 34 (2002) 259–264.
- [62] S.T. Smale, J.T. Kadonaga, The RNA polymerase II core promoter, *Annu. Rev. Biochem.* 72 (2003) 449–479.
- [63] F. Muller, L. Tora, Chromatin and DNA sequences in defining promoters for transcription initiation, *Biochim. Biophys. Acta* 1839 (2014) 118–128.
- [64] J. Corden, B. Wasylyk, A. Buchwalder, P. Sassone-Corsi, C. Keding, P. Chambon, Promoter sequences of eukaryotic protein-coding genes, *Science* 209 (1980) 1406–1414.
- [65] S.T. Smale, D. Baltimore, The “initiator” as a transcription control element, *Cell* 57 (1989) 103–113.
- [66] P.C. FitzGerald, D. Sturgill, A. Shyakhthenko, B. Oliver, C. Vinson, Comparative genomics of *Drosophila* and human core promoters, *Genome Biol.* 7 (2006) R53.
- [67] N.I. Gershenzon, E.N. Trifonov, I.P. Ioshikhes, The features of *Drosophila* core promoters revealed by statistical analysis, *BMC Genomics* 7 (2006) 161.
- [68] U. Ohler, G.C. Liao, H. Niemann, G.M. Rubin, Computational analysis of core promoters in the *Drosophila* genome, *Genome Biol.* 3 (2002) (RESEARCH0087).
- [69] J. Kaufmann, S.T. Smale, Direct recognition of initiator elements by a component of the transcription factor IID complex, *Genes Dev.* 8 (1994) 821–829.
- [70] C.P. Verrijzer, J.L. Chen, K. Yokomori, R. Tjian, Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II, *Cell* 81 (1995) 1115–1125.
- [71] G.E. Chalkley, C.P. Verrijzer, DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250-TAF(II)150 complex recognizes the initiator, *EMBO J.* 18 (1999) 4835–4845.
- [72] C.H. Wu, L. Madabusi, H. Nishioka, P. Emanuel, M. Sypes, I. Arkhipova, D.S. Gilmour, Analysis of core promoter sequences located downstream from the TATA element in the hsp70 promoter from *Drosophila melanogaster*, *Mol. Cell. Biol.* 21 (2001) 1593–1602.
- [73] R. Javahery, A. Khachi, K. Lo, B. Zenzie-Gregory, S.T. Smale, DNA sequence requirements for transcriptional initiator activity in mammalian cells, *Mol. Cell. Biol.* 14 (1994) 116–127.
- [74] B.A. Purnell, P.A. Emanuel, D.S. Gilmour, TFIID sequence recognition of the initiator and sequences farther downstream in *Drosophila* class II genes, *Genes Dev.* 8 (1994) 830–842.
- [75] C. Yang, E. Bolotin, T. Jiang, F.M. Sladek, E. Martinez, Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters, *Gene* 389 (2007) 52–65.
- [76] M.C. Frith, E. Valen, A. Krogh, Y. Hayashizaki, P. Carninci, A. Sandelin, A code for transcription initiation in mammalian genomes, *Genome Res.* 18 (2008) 1–12.
- [77] G. Yarden, R. Elfakess, K. Gazit, R. Dikstein, Characterization of sINR, a strict version of the Initiator core promoter element, *Nucleic Acids Res.* 37 (2009) 4234–4246.
- [78] N. Hariharan, R.P. Perry, Functional dissection of a mouse ribosomal protein promoter: significance of the polypyrimidine initiator and an element in the TATA-box region, *Proc. Natl. Acad. Sci. U. S. A.* 87 (1990) 1526–1530.
- [79] A. Shibui-Nihei, Y. Ohmori, K. Yoshida, J. Imai, I. Oosuga, M. Iidaka, Y. Suzuki, J. Mizushima-Sugano, K. Yoshitomo-Nakagawa, S. Sugano, The 5' terminal oligopyrimidine tract of human elongation factor 1A-1 gene functions as a transcriptional initiator and produces a variable number of Us at the transcriptional level, *Gene* 311 (2003) 137–145.
- [80] R.P. Perry, The architecture of mammalian ribosomal protein promoters, *BMC Evol. Biol.* 5 (2005) 15.
- [81] T.L. Hamilton, M. Stoneley, K.A. Spriggs, M. Bushell, TOPs and their regulation, *Biochem. Soc. Trans.* 34 (2006) 12–16.
- [82] Y. Tokusumi, Y. Ma, X. Song, R.H. Jacobson, S. Takada, The new core promoter element XCP1 (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 (2007) 1844–1858.
- [83] R. Anish, M.B. Hossain, R.H. Jacobson, S. Takada, Characterization of transcription from TATA-less promoters: identification of a new core promoter element XCP2 and analysis of factor requirements, *PLoS ONE* 4 (2009) e5103.
- [84] M.L. Goldberg, (Ph.D. thesis) Stanford University, 1979.
- [85] N.I. Gershenzon, I.P. Ioshikhes, Synergy of human Pol II core promoter elements revealed by statistical sequence analysis, *Bioinformatics* 21 (2005) 1295–1300.
- [86] M. Mencía, Z. Moqtaderi, J.V. Geisberg, L. Kuras, K. Struhl, Activator-specific recruitment of TFIID and regulation of ribosomal protein genes in yeast, *Mol. Cell* 9 (2002) 823–833.
- [87] A.D. Basehoar, S.J. Zanton, B.F. Pugh, Identification and distinct regulation of yeast TATA box-containing genes, *Cell* 116 (2004) 699–709.
- [88] C. Molina, E. Grotewold, Genome wide analysis of *Arabidopsis* core promoters, *BMC Genomics* 6 (2005) 25.
- [89] Y.Y. Yamamoto, H. Ichida, T. Abe, Y. Suzuki, S. Sugano, J. Obokata, Differentiation of core promoter architecture between plants and mammals revealed by LDSS analysis, *Nucleic Acids Res.* 35 (2007) 6219–6226.
- [90] T. Morton, J. Petricka, D.L. Corcoran, S. Li, C.M. Winter, A. Carda, P.N. Benfey, U. Ohler, M. Megraw, Paired-end analysis of transcription start sites in *Arabidopsis* reveals plant-specific promoter signatures, *Plant Cell* 26 (2014) 2746–2760.
- [91] J.N. Reeve, Archaeal chromatin and transcription, *Mol. Microbiol.* 48 (2003) 587–598.
- [92] V.L. Singer, C.R. Wobbe, K. Struhl, A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation, *Genes Dev.* 4 (1990) 636–645.
- [93] C.D. Kaplan, Basic mechanisms of RNA polymerase II activity and alteration of gene expression in *Saccharomyces cerevisiae*, *Biochim. Biophys. Acta* 1829 (2013) 39–54.
- [94] N. Marbach-Bar, A. Ben-Noon, S. Ashkenazi, A. Tamarkin-Ben Harush, T. Avnit-Sagi, M.D. Walker, R. Dikstein, Disparity between microRNA levels and promoter strength is associated with initiation rate and Pol II pausing, *Nat. Commun.* 4 (2013) 2118.
- [95] W. Deng, S.G. Roberts, A core promoter element downstream of the TATA box that is recognized by TFIIB, *Genes Dev.* 19 (2005) 2418–2423.
- [96] W. Deng, S.G. Roberts, TFIIB and the regulation of transcription by RNA polymerase II, *Chromosoma* 116 (2007) 417–429.
- [97] T. Lagrange, A.N. Kapanidis, H. Tang, D. Reinberg, R.H. Ebricht, New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB, *Genes Dev.* 12 (1998) 34–44.
- [98] R. Evans, J.A. Fairley, S.G. Roberts, Activator-mediated disruption of sequence-specific DNA contacts by the general transcription factor TFIIB, *Genes Dev.* 15 (2001) 2945–2949.
- [99] T. Juven-Gershon, J.Y. Hsu, J.T. Kadonaga, Caudal, a key developmental regulator, is a DPE-specific transcriptional factor, *Genes Dev.* 22 (2008) 2823–2830.
- [100] T.W. Burke, J.T. Kadonaga, *Drosophila* TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters, *Genes Dev.* 10 (1996) 711–724.
- [101] T.W. Burke, J.T. Kadonaga, The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAFII60 of *Drosophila*, *Genes Dev.* 11 (1997) 3020–3031.
- [102] A.K. Kutach, J.T. Kadonaga, The downstream promoter element DPE appears to be as widely used as the TATA box in *Drosophila* core promoters, *Mol. Cell. Biol.* 20 (2000) 4754–4764.
- [103] Y. Zehavi, A. Sloutskin, O. Kuznetsov, T. Juven-Gershon, The core promoter composition establishes a new dimension in developmental gene networks, *Nucleus* 5 (2014).
- [104] C.Y. Lim, B. Santoso, T. Boulay, E. Dong, U. Ohler, J.T. Kadonaga, The MTE, a new core promoter element for transcription by RNA polymerase II, *Genes Dev.* 18 (2004) 1606–1617.
- [105] J.W. Theisen, C.Y. Lim, J.T. Kadonaga, Three key subregions contribute to the function of the downstream RNA polymerase II core promoter, *Mol. Cell. Biol.* 30 (2010) 3471–3479.
- [106] T. Zhou, C.M. Chiang, The intronless and TATA-less human TAF(II)55 gene contains a functional initiator and a downstream promoter element, *J. Biol. Chem.* 276 (2001) 25503–25511.
- [107] H. Shao, M. Revach, S. Moshonov, Y. Tzuman, K. Gazit, S. Albeck, T. Unger, R. Dikstein, Core promoter binding by histone-like TAF complexes, *Mol. Cell. Biol.* 25 (2005) 206–219.
- [108] S.H. Duttke, RNA polymerase III accurately initiates transcription from RNA polymerase II promoters in vitro, *J. Biol. Chem.* 289 (2014) 20396–20404.
- [109] D.H. Lee, N. Gershenzon, M. Gupta, I.P. Ioshikhes, D. Reinberg, B.A. Lewis, Functional characterization of core promoter elements: the downstream core element is recognized by TAF1, *Mol. Cell. Biol.* 25 (2005) 9674–9686.
- [110] B.A. Lewis, T.K. Kim, S.H. Orkin, A downstream element in the human beta-globin promoter: evidence of extended sequence-specific transcription factor IID contacts, *Proc. Natl. Acad. Sci. U. S. A.* 97 (2000) 7172–7177.
- [111] A. Matsuoka, F. Hirose, M.A. Yoo, M. Yamaguchi, The DRE/DREF transcriptional regulatory system: a master key for cell proliferation, *Biochim. Biophys. Acta* 1779 (2008) 81–89.

- [112] J. Li, D.S. Gilmour, Distinct mechanisms of transcriptional pausing orchestrated by GAGA factor and M1BP, a novel transcription factor, *EMBO J.* 32 (2013) 1829–1841.
- [113] Z. Chen, J.L. Manley, Core promoter elements and TAFs contribute to the diversity of transcriptional activation in vertebrates, *Mol. Cell. Biol.* 23 (2003) 7350–7362.
- [114] E. Martinez, H. Ge, Y. Tao, C.X. Yuan, V. Palhan, R.G. Roeder, Novel cofactors and TFIIA mediate functional core promoter selectivity by the human TAFII150-containing TFIIID complex, *Mol. Cell. Biol.* 18 (1998) 6571–6583.
- [115] J.Y. Hsu, T. Juven-Gershon, M.T. Marr 2nd, K.J. Wright, R. Tjian, J.T. Kadonaga, TBP, Mot1, and NC2 establish a regulatory circuit that controls DPE-dependent versus TATA-dependent transcription, *Genes Dev.* 22 (2008) 2353–2358.
- [116] T. Juven-Gershon, S. Cheng, J.T. Kadonaga, Rational design of a super core promoter that enhances gene expression, *Nat. Methods* 3 (2006) 917–922.
- [117] T. Matsui, J. Segall, P.A. Weil, R.G. Roeder, Multiple factors required for accurate initiation of transcription by purified RNA polymerase II, *J. Biol. Chem.* 255 (1980) 11992–11996.
- [118] M. Samuels, A. Fire, P.A. Sharp, Separation and characterization of factors mediating accurate transcription by RNA polymerase II, *J. Biol. Chem.* 257 (1982) 14419–14427.
- [119] Y. He, J. Fang, D.J. Taatjes, E. Nogales, Structural visualization of key steps in human transcription initiation, *Nature* 495 (2013) 481–486.
- [120] B.A. Lewis, R.J. Sims 3rd, W.S. Lane, D. Reinberg, Functional characterization of core promoter elements: DPE-specific transcription requires the protein kinase CK2 and the PC4 coactivator, *Mol. Cell* 18 (2005) 471–481.
- [121] F. Muller, M.A. Demeny, L. Tora, New problems in RNA polymerase II transcription initiation: matching the diversity of core promoters with a variety of promoter recognition factors, *J. Biol. Chem.* 282 (2007) 14685–14689.
- [122] T.W. Sikorski, S. Buratowski, The basal initiation machinery: beyond the general transcription factors, *Curr. Opin. Cell Biol.* 21 (2009) 344–351.
- [123] Y. Zehavi, A. Kedmi, D. Ideses, T. Juven-Gershon, TRF2: TRansForming the view of general transcription factors, *Transcription* 6 (2015) 1–6.
- [124] G. Papai, P.A. Weil, P. Schultz, New insights into the function of transcription factor TFIIID from recent structural studies, *Curr. Opin. Genet. Dev.* 21 (2011) 219–224.
- [125] N. Nakajima, M. Horikoshi, R.G. Roeder, Factors involved in specific transcription by mammalian RNA polymerase II: purification, genetic specificity, and TATA box-promoter interactions of TFIIID, *Mol. Cell. Biol.* 8 (1988) 4028–4040.
- [126] C.M. Chiang, H. Ge, Z. Wang, A. Hoffmann, R.G. Roeder, Unique TATA-binding protein-containing complexes and cofactors involved in transcription by RNA polymerases II and III, *EMBO J.* 12 (1993) 2749–2762.
- [127] E. Wiczcerek, M. Brand, X. Jacq, L. Tora, Function of TAF(II)-containing complex without TBP in transcription by RNA polymerase II, *Nature* 393 (1998) 187–191.
- [128] K. Gazit, S. Moshonov, R. Elfakess, M. Sharon, G. Mengus, I. Davidson, R. Dikstein, TAF4/4b x TAF12 displays a unique mode of DNA binding and is required for core promoter function of a subset of genes, *J. Biol. Chem.* 284 (2009) 26286–26296.
- [129] T. O'Brien, R. Tjian, Different functional domains of TAFII250 modulate expression of transcriptional subsets of mammalian genes, *Proc. Natl. Acad. Sci. U. S. A.* 97 (2000) 2456–2461.
- [130] R.O. Weinzierl, B.D. Dynlacht, R. Tjian, Largest subunit of *Drosophila* transcription factor IID directs assembly of a complex containing TBP and a coactivator, *Nature* 362 (1993) 511–517.
- [131] K.J. Wright, M.T. Marr 2nd, R. Tjian, TAF4 nucleates a core subcomplex of TFIIID and mediates activated transcription from a TATA-less promoter, *Proc. Natl. Acad. Sci. U. S. A.* 103 (2006) 12347–12352.
- [132] C. Bieniossek, G. Papai, C. Schaffitzel, F. Garzoni, M. Chaillet, E. Scheer, P. Papadopoulos, L. Tora, P. Schultz, I. Berger, The architecture of human general transcription factor TFIIID core complex, *Nature* 493 (2013) 699–702.
- [133] M.A. Demeny, E. Soutoglou, Z. Nagy, E. Scheer, A. Janoshazi, M. Richardot, M. Argentini, P. Kessler, L. Tora, Identification of a small TAF complex and its role in the assembly of TAF-containing complexes, *PLoS ONE* 2 (2007) e316.
- [134] J. Bonnet, C.Y. Wang, T. Baptista, S.D. Vincent, W.C. Hsiao, M. Stierle, C.F. Kao, L. Tora, D. Devys, The SAGA coactivator complex acts on the whole transcribed genome and is required for RNA polymerase II transcription, *Genes Dev.* 28 (2014) 1999–2012.
- [135] D.J. Mitsiou, H.G. Stunnenberg, TAC, a TBP-sans-TAFs complex containing the unprocessed TFIIA $\alpha$  precursor and the TFIIA $\gamma$  subunit, *Mol. Cell* 6 (2000) 527–537.
- [136] T. Raha, S.W. Cheng, M.R. Green, HIV-1 Tat stimulates transcription complex assembly through recruitment of TBP in the absence of TAFs, *PLoS Biol.* 3 (2005) e44.
- [137] B. Guglielmi, N. La Rochelle, R. Tjian, Gene-specific transcriptional mechanisms at the histone gene cluster revealed by single-cell imaging, *Mol. Cell* 51 (2013) 480–492.
- [138] J. Zaborowska, A. Taylor, S. Murphy, A novel TBP-TAF complex on RNA polymerase II-transcribed snRNA genes, *Transcription* 3 (2012) 92–104.
- [139] F.J. van Werven, H. van Bakel, H.A. van Teeffelen, A.F. Altelaar, M.G. Koerkamp, A.J. Heck, F.C. Holstege, H.T. Timmers, Cooperative action of NC2 and Mot1p to regulate TATA-binding protein function across the genome, *Genes Dev.* 22 (2008) 2359–2369.
- [140] W. Deng, B. Malecova, T. Oelgeschlager, S.G. Roberts, TFIIIB recognition elements control the TFIIA-NC2 axis in transcriptional regulation, *Mol. Cell. Biol.* 29 (2009) 1389–1400.
- [141] M. Xu, P. Sharma, S. Pan, S. Malik, R.G. Roeder, E. Martinez, Core promoter-selective function of HMGA1 and Mediator in Initiator-dependent transcription, *Genes Dev.* 25 (2011) 2513–2524.
- [142] M.A. Cianfrocco, G.A. Kassavetis, P. Grob, J. Fang, T. Juven-Gershon, J.T. Kadonaga, E. Nogales, Human TFIIID binds to core promoter DNA in a reorganized structural state, *Cell* 152 (2013) 120–131.
- [143] M.A. Cianfrocco, E. Nogales, Regulatory interplay between TFIIID's conformational transitions and its modular interaction with core promoter DNA, *Transcription* 4 (2013) 120–126.
- [144] W. Akhtar, G.J. Veenstra, TBP-related factors: a paradigm of diversity in transcription initiation, *Cell Biosci.* 1 (2011) 23.
- [145] F. Muller, A. Zaucker, L. Tora, Developmental regulation of transcription initiation: more than just changing the actors, *Curr. Opin. Gen. Dev.* 20 (2010) 533–540.
- [146] J.H. Reina, N. Hernandez, On a roll for new TRF targets, *Genes Dev.* 21 (2007) 2855–2860.
- [147] S.H. Duttke, R.F. Doolittle, Y.L. Wang, J.T. Kadonaga, TRF2 and the evolution of the bilateria, *Genes Dev.* 28 (2014) 2071–2076.
- [148] P.A. Moore, J. Ozer, M. Salunek, G. Jan, D. Zerby, S. Campbell, P.M. Lieberman, A human TATA binding protein-related protein with altered DNA binding specificity inhibits transcription from multiple promoters and activators, *Mol. Cell. Biol.* 19 (1999) 7610–7620.
- [149] M.D. Rabenstein, S. Zhou, J.T. Lis, R. Tjian, TATA box-binding protein (TBP)-related factor 2 (TRF2), a third member of the TBP family, *Proc. Natl. Acad. Sci. U. S. A.* 96 (1999) 4791–4796.
- [150] Y. Isogai, S. Keles, M. Prestel, A. Hochheimer, R. Tjian, Transcription of histone gene cluster by differential core-promoter factors, *Genes Dev.* 21 (2007) 2936–2949.
- [151] Y.L. Wang, S.H. Duttke, K. Chen, J. Johnston, G.A. Kassavetis, J. Zeitlinger, J.T. Kadonaga, TRF2, but not TBP, mediates the transcription of ribosomal protein genes, *Genes Dev.* 28 (2014) 1550–1555.
- [152] A. Kedmi, Y. Zehavi, Y. Glick, Y. Orenstein, D. Ideses, C. Wachtel, T. Doniger, H. Waldman Ben-Asher, N. Muster, J. Thompson, S. Anderson, D. Avrahami, J.R. Yates 3rd, R. Shamir, D. Gerber, T. Juven-Gershon, *Drosophila* TRF2 is a preferential core promoter regulator, *Genes Dev.* 28 (2014) 2163–2174.
- [153] S.H. Duttke, Evolution and diversification of the basal transcription machinery, *Trends Biochem. Sci.* 40 (2015) 127–129.
- [154] J.A. Goodrich, R. Tjian, Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation, *Nat. Rev. Genet.* 11 (2010) 549–558.
- [155] D.A. Wassarman, N. Aoyagi, L.A. Pile, E.M. Schlag, TAF250 is required for multiple developmental events in *Drosophila*, *Proc. Natl. Acad. Sci. U. S. A.* 97 (2000) 1154–1159.
- [156] N. Aoyagi, D.A. Wassarman, Developmental and transcriptional consequences of mutations in *Drosophila* TAF(II)60, *Mol. Cell. Biol.* 21 (2001) 6808–6819.
- [157] J. Zhou, J. Zwicker, P. Szymanski, M. Levine, R. Tjian, TAFII mutations disrupt Dorsal activation in the *Drosophila* embryo, *Proc. Natl. Acad. Sci. U. S. A.* 95 (1998) 13483–13488.
- [158] M. Guermah, K. Ge, C.M. Chiang, R.G. Roeder, The TBN protein, which is essential for early embryonic mouse development, is an inducible TAFII implicated in adipogenesis, *Mol. Cell* 12 (2003) 991–1001.
- [159] S. Georgieva, D.B. Kirschner, T. Jagla, E. Nabirochkina, S. Hanke, H. Schenkel, C. de Lorenzo, P. Sinha, K. Jagla, B. Mechler, L. Tora, Two novel *Drosophila* TAF(II)s have homology with human TAF(II)30 and are differentially regulated during development, *Mol. Cell. Biol.* 20 (2000) 1639–1648.
- [160] W.S. Mohan Jr., E. Scheer, O. Wendling, D. Metzger, L. Tora, TAF10 (TAF(II)30) is necessary for TFIIID stability and early embryogenesis in mice, *Mol. Cell. Biol.* 23 (2003) 4307–4318.
- [161] A. Tatarakis, T. Margaritis, C.P. Martinez-Jimenez, A. Kouskouti, W.S. Mohan 2nd, A. Haroniti, D. Kafetzopoulos, L. Tora, I. Talianidis, Dominant and redundant functions of TFIIID involved in the regulation of hepatic genes, *Mol. Cell* 31 (2008) 531–543.
- [162] W.W. Pijnappel, D. Esch, M.P. Baltissen, G. Wu, N. Mischerikow, A.J. Bergsma, E. van der Wal, D.W. Han, H. Bruch, S. Moritz, P. Lijnzaad, A.F. Altelaar, K. Sameith, H. Zaehres, A.J. Heck, F.C. Holstege, H.R. Scholer, H.T. Timmers, A central role for TFIIID in the pluripotent transcription circuitry, *Nature* 495 (2013) 516–519.
- [163] G.A. Maston, L.J. Zhu, L. Chamberlain, L. Lin, M. Fang, M.R. Green, Non-canonical TAF complexes regulate active promoters in human embryonic stem cells, *eLife* 1 (2012) e00068.
- [164] P.J. Wang, D.C. Page, Functional substitution for TAF(II)250 by a retroposed homolog that is expressed in human spermatogenesis, *Hum. Mol. Genet.* 11 (2002) 2341–2346.
- [165] J.C. Pointud, G. Mengus, S. Brancorsini, L. Monaco, M. Parvinen, P. Sassone-Corsi, I. Davidson, The intracellular localisation of TAF7L, a paralogue of transcription factor TFIIID subunit TAF7, is developmentally regulated during male germ-cell differentiation, *J. Cell Sci.* 116 (2003) 1847–1858.
- [166] Y. Cheng, M.G. Buffone, M. Kouadio, M. Goodheart, D.C. Page, G.L. Gerton, I. Davidson, P.J. Wang, Abnormal sperm in mice lacking the Taf7l gene, *Mol. Cell. Biol.* 27 (2007) 2582–2589.
- [167] H. Zhou, I. Grubisic, K. Zheng, Y. He, P.J. Wang, T. Kaplan, R. Tjian, Taf7l cooperates with Trf2 to regulate spermiogenesis, *Proc. Natl. Acad. Sci. U. S. A.* 110 (2013) 16886–16891.
- [168] H. Zhou, T. Kaplan, Y. Li, I. Grubisic, Z. Zhang, P.J. Wang, M.B. Eisen, R. Tjian, Dual functions of TAF7L in adipocyte differentiation, *eLife* 2 (2013) e00170.
- [169] H. Zhou, B. Wan, I. Grubisic, T. Kaplan, R. Tjian, TAF7L modulates brown adipose tissue formation, *eLife*, 3, 2014.
- [170] R. Dikstein, S. Zhou, R. Tjian, Human TAFII 105 is a cell type-specific TFIIID subunit related to hTAFII130, *Cell* 87 (1996) 137–146.
- [171] A.E. Falender, R.N. Freiman, K.G. Geles, K.C. Lo, K. Hwang, D.J. Lamb, P.L. Morris, R. Tjian, J.S. Richards, Maintenance of spermatogenesis requires TAF4b, a gonad-specific subunit of TFIIID, *Genes Dev.* 19 (2005) 794–803.
- [172] A.E. Falender, M. Shimada, Y.K. Lo, J.S. Richards, TAF4b, a TBP associated factor, is required for oocyte development and function, *Dev. Biol.* 288 (2005) 405–419.

- [173] R.N. Freiman, S.R. Albright, S. Zheng, W.C. Sha, R.E. Hammer, R. Tjian, Requirement of tissue-selective TBP-associated factor TAFII105 in ovarian development, *Science* 293 (2001) 2084–2087.
- [174] K.J. Grive, K.A. Seymour, R. Mehta, R.N. Freiman, TAF4b promotes mouse primordial follicle assembly and oocyte survival, *Dev. Biol.* 392 (2014) 42–51.
- [175] F.J. Herrera, T. Yamaguchi, H. Roelink, R. Tjian, Core promoter factor TAF9B regulates neuronal gene expression, *eLife* 3 (2014) e02559.
- [176] M. Hiller, X. Chen, M.J. Pringle, M. Suchorolski, Y. Sancak, S. Viswanathan, B. Bolival, T.Y. Lin, S. Marino, M.T. Fuller, Testis-specific TAF homologs collaborate to control a tissue-specific transcription program, *Development* 131 (2004) 5297–5308.
- [177] U. Ohler, D.A. Wassarman, Promoting developmental transcription, *Development* 137 (2010) 15–26.
- [178] I. Martianov, G.M. Fimia, A. Dierich, M. Parvinen, P. Sassone-Corsi, I. Davidson, Late arrest of spermiogenesis and germ cell apoptosis in mice lacking the TBP-like TLF/TRF2 gene, *Mol. Cell* 7 (2001) 509–515.
- [179] D. Zhang, T.L. Penttila, P.L. Morris, M. Teichmann, R.G. Roeder, Spermiogenesis deficiency in mice lacking the Trf2 gene, *Science* 292 (2001) 1153–1155.
- [180] T. Oyama, S. Sasagawa, S. Takeda, R.A. Hess, P.M. Lieberman, E.H. Cheng, J.J. Hsieh, Cleavage of TFIIA by Taspase1 activates TRF2-specified mammalian male germ cell programs, *Dev. Cell* 27 (2013) 188–200.
- [181] A. Bashirullah, G. Lam, V.P. Yin, C.S. Thummel, dTrf2 is required for transcriptional and developmental responses to ecdysone during *Drosophila* metamorphosis, *Developmental dynamics: an official publication of the American Association of Anatomists* 2362007. 3173–3179.
- [182] D.O. Hart, T. Raha, N.D. Lawson, M.R. Green, Initiation of zebrafish haematopoiesis by the TATA-box-binding protein-related factor Trf3, *Nature* 450 (2007) 1082–1085.
- [183] D.O. Hart, M.K. Santra, T. Raha, M.R. Green, Selective interaction between Trf3 and Taf3 required for early development and hematopoiesis, *Developmental dynamics: an official publication of the American Association of Anatomists* 2382009. 2540–2549.
- [184] R. Bartfai, C. Balduf, T. Hilton, Y. Rathmann, Y. Hadzhiev, L. Tora, L. Orban, F. Muller, TBP2, a vertebrate-specific member of the TBP family, is required in embryonic development of zebrafish, *Curr. Biol.* 14 (2004) 593–598.
- [185] Z. Jallow, U.G. Jacobi, D.L. Weeks, I.B. Dawid, G.J. Veenstra, Specialized and redundant roles of TBP and a vertebrate-specific TBP paralog in embryonic gene regulation in *Xenopus*, *Proc. Natl. Acad. Sci. U. S. A.* 101 (2004) 13525–13530.
- [186] E. Gazdag, A. Santenard, C. Ziegler-Birling, G. Altobelli, O. Poch, L. Tora, M.E. Torres-Padilla, TBP2 is essential for germ cell development by regulating transcription and chromatin condensation in the oocyte, *Genes Dev.* 23 (2009) 2210–2223.
- [187] M. Bulger, M. Groudine, Functional and mechanistic diversity of distal transcription enhancers, *Cell* 144 (2011) 327–339.
- [188] M. Levine, Transcriptional enhancers in animal development and evolution, *Curr. Biology*. 20 (2010) R754–R763.
- [189] M. Levine, C. Cattoglio, R. Tjian, Looping back to leap forward: transcription enters a new era, *Cell* 157 (2014) 13–25.
- [190] J. Marsman, J.A. Horsfield, Long distance relationships: enhancer-promoter communication and dynamic gene transcription, *Biochim. Biophys. Acta* 1819 (2012) 1217–1227.
- [191] C.T. Ong, V.G. Corces, Enhancer function: new insights into the regulation of tissue-specific gene expression, *Nat. Rev. Genet.* 12 (2011) 283–293.
- [192] D. Shlyueva, G. Stampfel, A. Stark, Transcriptional enhancers: from properties to genome-wide predictions, *Nat. Rev. Genet.* 15 (2014) 272–286.
- [193] F. Spitz, E.E. Furlong, Transcription factors: from enhancer binding to developmental control, *Nat. Rev. Genet.* 13 (2012) 613–626.
- [194] B. van Arensbergen, H.J. van Steensel, Bussemaker, In search of the determinants of enhancer-promoter interaction specificity, *Trends Cell Biol.* 24 (2014) 695–702.
- [195] X. Li, M. Noll, Compatibility between enhancers and promoters determines the transcriptional specificity of gooseberry and gooseberry neuro in the *Drosophila* embryo, *EMBO J.* 13 (1994) 400–406.
- [196] C. Merli, D.E. Bergstrom, J.A. Cygan, R.K. Blackman, Promoter specificity mediates the independent regulation of neighboring genes, *Genes Dev.* 10 (1996) 1260–1270.
- [197] B. Tolhuis, R.J. Palstra, E. Splinter, F. Grosveld, W. de Laat, Looping and interaction between hypersensitive sites in the active beta-globin locus, *Mol. Cell* 10 (2002) 1453–1465.
- [198] J. Gehrig, M. Reischl, E. Kalmar, M. Ferg, Y. Hadzhiev, A. Zaucker, C. Song, S. Schindler, U. Liebel, F. Muller, Automated high-throughput mapping of promoter-enhancer interactions in zebrafish embryos, *Nat. Methods* 6 (2009) 911–916.
- [199] V.C. Calhoun, A. Stathopoulos, M. Levine, Promoter-proximal tethering elements regulate enhancer-promoter specificity in the *Drosophila* Antennapedia complex, *Proc. Natl. Acad. Sci. U. S. A.* 99 (2002) 9243–9247.
- [200] O.S. Akbari, E. Bae, H. Johnsen, A. Villaluz, D. Wong, R.A. Drewell, A novel promoter-tethering element regulates enhancer-driven gene expression at the bithorax complex in the *Drosophila* embryo, *Development* 135 (2008) 123–131.
- [201] S. Ohtsuki, M. Levine, H.N. Cai, Different core promoters possess distinct regulatory activities in the *Drosophila* embryo, *Genes Dev.* 12 (1998) 547–556.
- [202] J.E. Butler, J.T. Kadonaga, Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs, *Genes Dev.* 15 (2001) 2515–2519.
- [203] F. Jin, Y. Li, J.R. Dixon, S. Selvaraj, Z. Ye, A.Y. Lee, C.A. Yen, A.D. Schmitt, C.A. Espinoza, B. Ren, A high-resolution map of the three-dimensional chromatin interactome in human cells, *Nature* 503 (2013) 290–294.
- [204] A. Sanyal, B.R. Lajoie, G. Jain, J. Dekker, The long-range interaction landscape of gene promoters, *Nature* 489 (2012) 109–113.
- [205] Y. Zhang, C.H. Wong, R.Y. Birnbaum, G. Li, R. Favaro, C.Y. Ngan, J. Lim, E. Tai, H.M. Poh, E. Wong, F.H. Mulawadi, W.K. Sung, S. Nicolis, N. Ahituv, Y. Ruan, C.L. Wei, Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations, *Nature* 504 (2013) 306–310.
- [206] C.D. Arnold, D. Gerlach, C. Stelzer, L.M. Boryn, M. Rath, A. Stark, Genome-wide quantitative enhancer activity maps identified by STARR-seq, *Science* 339 (2013) 1074–1077.
- [207] Y. Ghavi-Helm, F.A. Klein, T. Pakozdi, L. Ciglar, D. Noordermeer, W. Huber, E.E. Furlong, Enhancer loops appear stable during development and are associated with paused polymerase, *Nature* 512 (2014) 96–100.
- [208] M.A. Zabidi, C.D. Arnold, K. Schernhuber, M. Pagani, M. Rath, O. Frank, A. Stark, Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation, *Nature* 518 (2014) 556–559.
- [209] N.J. Fuda, M.B. Ardehali, J.T. Lis, Defining mechanisms that regulate RNA polymerase II transcription in vivo, *Nature* 461 (2009) 186–192.
- [210] S. Nechaev, K. Adelman, Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation, *Biochim. Biophys. Acta* 1809 (2011) 34–45.
- [211] D.L. Bentley, Coupling mRNA processing with transcription in time and space, *Nat. Rev. Genet.* 15 (2014) 163–175.
- [212] K. Adelman, J.T. Lis, Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans, *Nat. Rev. Genet.* 13 (2012) 720–731.
- [213] D.A. Gilchrist, K. Adelman, Coupling polymerase pausing and chromatin landscapes for precise regulation of transcription, *Biochim. Biophys. Acta* 1819 (2012) 700–706.
- [214] Y. Yamaguchi, H. Shibata, H. Handa, Transcription elongation factors DSIF and NELF: promoter-proximal pausing and beyond, *Biochim. Biophys. Acta* 1829 (2013) 98–104.
- [215] D.S. Gilmour, J.T. Lis, RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in *Drosophila melanogaster* cells, *Mol. Cell. Biol.* 6 (1986) 3984–3989.
- [216] E.B. Rasmussen, J.T. Lis, In vivo transcriptional pausing and cap formation on three *Drosophila* heat shock genes, *Proc. Natl. Acad. Sci. U. S. A.* 90 (1993) 7923–7927.
- [217] D.L. Bentley, M. Groudine, A block to elongation is largely responsible for decreased transcription of c-myc in differentiated HL60 cells, *Nature* 321 (1986) 702–706.
- [218] A. Krumm, T. Meulia, M. Brunvand, M. Groudine, The block to transcriptional elongation within the human c-myc gene is determined in the promoter-proximal region, *Genes Dev.* 6 (1992) 2201–2213.
- [219] G.W. Muse, D.A. Gilchrist, S. Nechaev, R. Shah, J.S. Parker, S.F. Grissom, J. Zeitlinger, K. Adelman, RNA polymerase is poised for activation across the genome, *Nat. Genet.* 39 (2007) 1507–1511.
- [220] J. Zeitlinger, A. Stark, M. Kellis, J.W. Hong, S. Nechaev, K. Adelman, M. Levine, R.A. Young, RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo, *Nat. Genet.* 39 (2007) 1512–1516.
- [221] C. Lee, X. Li, A. Hechmer, M. Eisen, M.D. Biggin, B.J. Venters, C. Jiang, J. Li, B.F. Pugh, D.S. Gilmour, NELF and GAGA factor are linked to promoter-proximal pausing at many genes in *Drosophila*, *Mol. Cell. Biol.* 28 (2008) 3290–3300.
- [222] S. Nechaev, D.C. Fargo, G. dos Santos, L. Liu, Y. Gao, K. Adelman, Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in *Drosophila*, *Science* 327 (2010) 335–338.
- [223] M. Quinodoz, C. Gobet, F. Naef, K.B. Gustafson, Characteristic bimodal profiles of RNA polymerase II at thousands of active mammalian promoters, *Genome Biol.* 15 (2014) R85.
- [224] B. Gaertner, J. Zeitlinger, RNA polymerase II pausing during development, *Development* 141 (2014) 1179–1183.
- [225] C. Nepal, Y. Hadzhiev, C. Previti, V. Haberle, N. Li, H. Takahashi, A.M. Suzuki, Y. Sheng, R.F. Abdelhamid, S. Anand, J. Gehrig, A. Akalin, C.E. Kockx, A.A. van der Sloot, W.F. van Ijcken, O. Armant, S. Rastegar, C. Watson, U. Strahle, F. Carninci, B. Lenhard, F. Muller, Dynamic regulation of the transcription initiation landscape at single nucleotide resolution during vertebrate embryogenesis, *Genome Res.* 23 (2013) 1938–1950.
- [226] V. Haberle, N. Li, Y. Hadzhiev, C. Plessy, C. Previti, C. Nepal, J. Gehrig, X. Dong, A. Akalin, A.M. Suzuki, W.F. van Ijcken, O. Armant, M. Ferg, U. Strahle, F. Carninci, F. Muller, B. Lenhard, Two independent transcription initiation codes overlap on vertebrate core promoters, *Nature* 507 (2014) 381–385.
- [227] W.S. Kruesi, L.J. Core, C.T. Waters, J.T. Lis, B.J. Meyer, Condensin controls recruitment of RNA polymerase II to achieve nematode X-chromosome dosage compensation, *eLife* 2 (2013) e00808.
- [228] T. Henriques, D.A. Gilchrist, S. Nechaev, M. Bern, G.W. Muse, A. Burkholder, D.C. Fargo, K. Adelman, Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals, *Mol. Cell* 52 (2013) 517–528.
- [229] J. Li, Y. Liu, H.S. Rhee, S.K. Ghosh, L. Bai, B.F. Pugh, D.S. Gilmour, Kinetic competition between elongation rate and binding of NELF controls promoter-proximal pausing, *Mol. Cell* 50 (2013) 711–722.
- [230] B. Gaertner, J. Johnston, K. Chen, N. Wallaschek, A. Paulson, A.S. Garruss, K. Gaudenz, B. De Kumar, R. Krumlauf, J. Zeitlinger, Poised RNA polymerase II changes over developmental time and prepares genes for future expression, *Cell Rep.* 2 (2012) 1670–1683.
- [231] M. Lagna, J.P. Bothma, E. Esposito, S. Ng, L. Stefanik, C. Tsui, J. Johnston, K. Chen, D.S. Gilmour, J. Zeitlinger, M.S. Levine, Paused Pol II coordinates tissue morphogenesis in the *Drosophila* embryo, *Cell* 153 (2013) 976–987.
- [232] A. Saunders, H.L. Ashe, Taking a pause to reflect on morphogenesis, *Cell* 153 (2013) 941–943.
- [233] A. Saunders, L.J. Core, C. Sutcliffe, J.T. Lis, H.L. Ashe, Extensive polymerase pausing during *Drosophila* axis patterning enables high-level and pliable transcription, *Genes Dev.* 27 (2013) 1146–1158.

- [234] K. Chen, J. Johnston, W. Shao, S. Meier, C. Staber, J. Zeitlinger, A global change in RNA polymerase II pausing during the *Drosophila* midblastula transition, *eLife* 2 (2013) e00861.
- [235] L. Amir-Zilberstein, E. Ainbinder, L. Toubé, Y. Yamaguchi, H. Handa, R. Dikstein, Differential regulation of NF-kappaB by elongation factors is determined by core promoter type, *Mol. Cell. Biol.* 27 (2007) 5246–5259.
- [236] D.A. Hendrix, J.W. Hong, J. Zeitlinger, D.S. Rokhsar, M.S. Levine, Promoter elements associated with RNA Pol II stalling in the *Drosophila* embryo, *Proc. Natl. Acad. Sci. U. S. A.* 105 (2008) 7762–7767.
- [237] H. Kwak, N.J. Fuda, L.J. Core, J.T. Lis, Precise maps of RNA polymerase reveal how promoters direct initiation and pausing, *Science* 339 (2013) 950–953.
- [238] N.J. Proudfoot, Ending the message: poly(A) signals then and now, *Genes Dev.* 25 (2011) 1770–1782.
- [239] P.K. Andersen, T.H. Jensen, S. Lykke-Andersen, Making ends meet: coordination between RNA 3'-end processing and transcription initiation, *Wiley Interdiscip. Rev.* 4 (2013) 233–246.
- [240] D.C. Di Giammartino, J.L. Manley, New links between mRNA polyadenylation and diverse nuclear pathways, *Mol. Cells* 37 (2014) 644–649.
- [241] O. Calvo, J.L. Manley, Strange bedfellows: polyadenylation factors at the promoter, *Genes Dev.* 17 (2003) 1321–1327.
- [242] K. Xiang, L. Tong, J.L. Manley, Delineating the structural blueprint of the pre-mRNA 3'-end processing machinery, *Mol. Cell. Biol.* 34 (2014) 1894–1910.
- [243] J.C. Dantonel, K.G. Murthy, J.L. Manley, L. Tora, Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA, *Nature* 389 (1997) 399–402.
- [244] Y. Wang, J.A. Fairley, S.G. Roberts, Phosphorylation of TFIIB links transcription initiation and termination, *Curr. Biol.* 20 (2010) 548–553.
- [245] C.K. Mapendano, S. Lykke-Andersen, J. Kjems, E. Bertrand, T.H. Jensen, Crosstalk between mRNA 3' end processing and transcription initiation, *Mol. Cell* 40 (2010) 410–422.
- [246] S. Lykke-Andersen, C.K. Mapendano, T.H. Jensen, An ending is a new beginning: transcription termination supports re-initiation, *Cell Cycle* 10 (2011) 863–865.
- [247] K. Oktaba, W. Zhang, T.S. Lotz, D.J. Jun, S.B. Lemke, S.P. Ng, E. Esposito, M. Levine, V. Hilgers, ELAV Links Paused Pol II to Alternative Polyadenylation in the *Drosophila* Nervous System, *Mol. Cell* 57 (2015) 341–348.
- [248] T. Gonatopoulos-Poumatzis, V.H. Cowling, Cap-binding complex (CBC), *Biochem. J.* 457 (2014) 231–242.
- [249] R.J. Jackson, C.U. Hellen, T.V. Pestova, The mechanism of eukaryotic translation initiation and principles of its regulation, *Nat. Rev. Mol. Cell Biol.* 11 (2010) 113–127.
- [250] R. Dikstein, Transcription and translation in a package deal: the TISU paradigm, *Gene* 491 (2012) 1–4.
- [251] M. Kozak, Initiation of translation in prokaryotes and eukaryotes, *Gene* 234 (1999) 187–208.
- [252] R. Elfakess, R. Dikstein, A translation initiation element specific to mRNAs with very short 5'UTR that also regulates transcription, *PLoS ONE* 3 (2008) e3094.
- [253] R. Elfakess, H. Sinvani, O. Haimov, Y. Svitkin, N. Sonenberg, R. Dikstein, Unique translation initiation of mRNAs-containing TISU element, *Nucleic Acids Res.* 39 (2011) 7598–7609.
- [254] D. Avni, S. Shama, F. Loreni, O. Meyuhas, Vertebrate mRNAs with a 5'-terminal pyrimidine tract are candidates for translational repression in quiescent cells: characterization of the translational cis-regulatory element, *Mol. Cell. Biol.* 14 (1994) 3822–3833.
- [255] O. Meyuhas, Synthesis of the translational apparatus is regulated at the translational level, *Eur. J. Biochem.* 267 (2000) 6321–6330.
- [256] M. Schafer, R. Kuhn, F. Bosse, U. Schafer, A conserved element in the leader mediates post-meiotic translation as well as cytoplasmic polyadenylation of a *Drosophila* spermatocyte mRNA, *EMBO J.* 9 (1990) 4519–4525.
- [257] E. Kempe, B. Muhs, M. Schafer, Gene regulation in *Drosophila* spermatogenesis: analysis of protein binding at the translational control element TCE, *Dev. Genet.* 14 (1993) 449–459.
- [258] R.J. Katzenberger, E.A. Rach, A.K. Anderson, U. Ohler, D.A. Wassarman, the *Drosophila* Translational Control Element (TCE) is required for high-level transcription of many genes that are specifically expressed in testes, *PLoS ONE* 7 (2012) e45009.
- [259] S. Moshonov, R. Elfakess, M. Golan-Mashiach, H. Sinvani, R. Dikstein, Links between core promoter and basic gene features influence gene expression, *BMC Genomics* 9 (2008) 92.
- [260] A. Tamarkin-Ben-Harush, E. Schechtman, R. Dikstein, Co-occurrence of transcription and translation gene regulatory features underlies coordinated mRNA and protein synthesis, *BMC Genomics* 15 (2014) 688.
- [261] L. Savinkova, I. Drachkova, T. Arshinova, P. Ponomarenko, M. Ponomarenko, N. Kolchanov, An experimental verification of the predicted effects of promoter TATA-box polymorphisms associated with human diseases on interactions between the TATA boxes and TATA-binding protein, *PLoS ONE* 8 (2013) e54626.
- [262] L.K. Savinkova, M.P. Ponomarenko, P.M. Ponomarenko, I.A. Drachkova, M.V. Lysova, T.V. Arshinova, N.A. Kolchanov, TATA box polymorphisms in human gene promoters and associated hereditary pathologies, *Biochem. Biokhim.* 74 (2009) 117–129.
- [263] G. Li, X. Ruan, R.K. Auerbach, K.S. Sandhu, M. Zheng, P. Wang, H.M. Poh, Y. Goh, J. Lim, J. Zhang, H.S. Sim, S.Q. Peh, F.H. Mulawadi, C.T. Ong, Y.L. Orlov, S. Hong, Z. Zhang, S. Landt, D. Raha, G. Euskirchen, C.L. Wei, W. Ge, H. Wang, C. Davis, K.I. Fisher-Aylor, A. Mortazavi, M. Gerstein, T. Gingeras, B. Wold, Y. Sun, M.J. Fullwood, E. Cheung, E. Liu, W.K. Sung, M. Snyder, Y. Ruan, Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation, *Cell* 148 (2012) 84–98.
- [264] T.K. Kim, M. Hemberg, J.M. Gray, A.M. Costa, D.M. Bear, J. Wu, D.A. Harmin, M. Laptewicz, K. Barbara-Haley, S. Kuersten, E. Markenscoff-Papadimitriou, D. Kuhl, H. Bito, P.F. Worley, G. Kreiman, M.E. Greenberg, Widespread transcription at neuronal activity-regulated enhancers, *Nature* 465 (2010) 182–187.
- [265] R. Andersson, C. Gebhard, I. Miguel-Escalada, I. Hoof, J. Bornholdt, M. Boyd, Y. Chen, X. Zhao, C. Schmidt, T. Suzuki, E. Ntini, E. Arner, E. Valen, K. Li, L. Schwarzfischer, D. Glatz, J. Raithe, B. Lilje, N. Rapin, F.O. Bagger, M. Jorgensen, P.R. Andersen, N. Bertin, O. Rackham, A.M. Burroughs, J.K. Baillie, Y. Ishizu, Y. Shimizu, E. Furuhashi, S. Maeda, Y. Negishi, C.J. Mungall, T.F. Meehan, T. Lassmann, M. Itoh, H. Kawaji, N. Kondo, J. Kawai, A. Lennartsson, C.O. Daub, P. Heutink, D.A. Hume, T.H. Jensen, H. Suzuki, Y. Hayashizaki, F. Muller, F. Consortium, A.R. Forrest, P. Carninci, M. Rehli, A. Sandelin, An atlas of active enhancers across human cell types and tissues, *Nature* 507 (2014) 455–461.
- [266] S. Weingarten-Gabbay, E. Segal, A shared architecture for promoters and enhancers, *Nat. Genet.* 46 (2014) 1253–1254.
- [267] R. Andersson, Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model, *BioEssays: news and reviews in molecular, cellular and developmental biology*, 2014.