Published in final edited form as:

Curr Opin Cell Biol. 2008 June ; 20(3): 253–259. doi:10.1016/j.ceb.2008.03.003.

The RNA Polymerase II Core Promoter – the Gateway to Transcription

Tamar Juven-Gershon, Jer-Yuan Hsu, Joshua W. M. Theisen, and James T. Kadonaga

Section of Molecular Biology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093–0347, USA

Summary

The RNA polymerase II core promoter is generally defined to be the sequence that directs the initiation of transcription. This simple definition belies a diverse and complex transcriptional module. There are two major types of core promoters – focused and dispersed. Focused promoters contain either a single transcription start site or a distinct cluster of start sites over several nucleotides, whereas dispersed promoters contain several start sites over 50 to 100 nucleotides and are typically found in CpG islands in vertebrates. Focused promoters are more ancient and widespread throughout nature than dispersed promoters; however, in vertebrates, dispersed promoters are more common than focused promoters. In addition, core promoters may contain many different sequence motifs, such as the TATA box, BRE, Inr, MTE, DPE, DCE, and XCPE1, that specify different mechanisms of transcription and responses to enhancers. Thus, the core promoter is a sophisticated gateway to transcription that determines which signals will lead to transcription initiation.

Introduction

The RNA polymerase II core promoter comprises the sequences that direct the initiation of transcription (for reviews, see [1,2*,3*,4*,5*]). Thus, in principle, the core promoter could be as simple as a single motif that serves as a universal transcription start site, or as complex as a unique set of sequence instructions for each promoter. Historically, the former model has often been presumed to be true, but emerging data indicate that there is considerable diversity in core promoter structure and function.

The objective of this review is to provide an overview of current topics that relate to the core promoter, with a particular emphasis on sequence motifs in core promoters. In addition, we have annotated core promoter-related data in papers that were published in the past two years. It should further be noted that the properties of core promoters and their cognate factors are not likely to be strictly absolute; hence, the principles and ideas described in this essay should be taken only as current working models.

Focused versus dispersed core promoters

The vast majority of research on core promoters has been devoted to the study of focused core promoters (Fig. 1). In focused core promoters (also referred to as single-peak, or SP, promoters), there is either a single transcription start site or a distinct cluster of start sites in a

Correspondence should be addressed to J.T.K. (jkadonaga@ucsd.edu).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

short region of several nucleotides. Most eukaryotic core promoters appear to be focused core promoters. In vertebrates, however, only about one-third or less of core promoters are focused core promoters; instead, the vast majority of genes appear to contain dispersed core promoters (also known as BR [broad distribution], MU [multimodal], or PB [broad with dominant peak] promoters), in which there are a number of transcription start sites distributed over a broad region that might typically range from 50 to 100 nucleotides (Fig. 1). [Note that dispersed core promoters should not be confused with alternate promoters, which are distinct and sometimes differentially-regulated promoters that are typically located hundreds or thousands of nucleotides apart.]

Core promoter elements such as the TATA box, BRE, Inr, MTE, DPE, and DCE (Fig. 2; discussed in greater detail below) are typically found in focused core promoters. These core promoter elements are not universal; rather, each is present in only a subset of core promoters. Moreover, some core promoters appear to lack all of the known core promoter elements. It is interesting to note that the TATA box and BRE are the most ancient of the core promoter motifs. The TATA box and BRE along with their cognate protein factors, TBP (TATA box-binding protein) and TFIIB (transcription factor IIB), are conserved from Archaea to humans (for review, see [6]). The TATA box is also present in plant promoters [7,8*]. The MTE and DPE appear to be conserved among metazoans. In contrast, dispersed core promoters are typically found in CpG islands in vertebrates and generally lack TATA, DPE, and MTE motifs (see, for example [1,3*,5*,9,10*]).

Thus, focused core promoters are more ancient and used in a much broader range of organisms than dispersed promoters. In addition, several of the key sequence motifs that contribute to the activity of focused core promoters have been identified. On the other hand, in vertebrates, dispersed core promoters are more common than focused promoters. Moreover, little is known about the sequences and factors that are responsible for transcription from dispersed core promoters. It is interesting to note, however, that the promoter region of dispersed TATA-less promoters are generally deficient in ATG triplets [11]. There may be fundamental differences in the basic mechanisms of transcription from dispersed versus focused core promoters.

The initiator (Inr)

The initiator (Inr) motif encompasses the transcription start site [1,12]. Based on functional assays, the Inr consensus was determined to be YYANWYY in humans and TCAKTY in *Drosophila* (degenerate nucleotides are indicated according to the IUPAC nucleotide code). The A nucleotide in the middle of the Inr consensus is often the +1 start site in focused core promoters. Inr-like sequences have also been described in *Saccharomyces cerevisiae* (for example, see [13*] and references therein). The Inr is probably the most commonly occurring sequence motif in focused core promoters (see, for example [14,15*,16*]). The Inr is a recognition site for the binding of TFIID. Although a number of proteins have been found to bind to Inr sequences, the binding of TFIID to the Inr appears to be particularly important because the sequence specificity of TFIID binding to the Inr region of the core promoter is identical to the Inr consensus sequence [17].

The computational analysis of thousands of mammalian transcription start sites suggests that the mammalian Inr consensus is YR, where R corresponds to the +1 start site [10*,18*]. In contrast, the computational analysis of thousands of *Drosophila* core promoters reveals a much more strict consensus sequence of TCAGTY [14,15*]. This sharp difference in the specificity of the Inr consensus between *Drosophila* and mammals suggests that mammalian transcription factors have evolved to function with a broader range of Inr sequences than *Drosophila* transcription factors. This property may be related to the prevalence of dispersed core promoters in mammals but not in *Drosophila*.

The TATA box and BRE

The TATA box, which is the most ancient and most widely-used core promoter motif throughout nature, was aptly the first eukaryotic core promoter element to be identified [19]. The TATA box has a consensus of TATAWAAR, where the upstream T nucleotide is most commonly at -31 or -30 relative to the A+1 (or G+1) in the Inr (see, for instance [10*,20*]). The TATA box is recognized and bound by TBP, which is a subunit of the TFIID complex in eukaryotes. As noted above, the TATA box is present in a subset of focused core promoters. Consequently, the TATA box appears to be uncommon in vertebrates (see, for example [10*,21,22*]), because only about one-third or less of vertebrate core promoters is focused and only a fraction of the focused core promoters contains a TATA box.

Yet, in this context, it is useful to note parenthetically that it is probably a good practice to be somewhat cautious in the interpretation of expected frequencies of core promoter motifs. For instance, the frequency of occurrence of the TATA box is an estimate that depends on the parameters (*i.e.*, sequences and positions) that are used to define the TATA box as well as the accuracy and promoter coverage of the particular database that is used in the analysis. Thus, although it is clearly apparent that there are many TATA-less promoters, the contribution of the TATA box or functionally equivalent sequences to vertebrate transcription remains to be determined unambiguously, and may be greater than is currently believed. Ultimately, it will be necessary to determine the actual function of each potential motif in each core promoter.

The BRE (TFIIB recognition element) was originally identified as a TFIIB-binding sequence that is immediately upstream of a subset of TATA boxes [23]. It was then found that TFIIB can bind upstream or downstream of the TATA box at the BRE^u (upstream BRE, which is the same as the original BRE) or BRE^d (downstream BRE) sequences [24,25*]. The BRE^u consensus is SSRCGCC [23]. BRE^d is located immediately downstream of the TATA box and has a consensus of RTDKKKK [24]. Depending on the promoter context, the BRE^u and BRE^d can act in either a positive or negative manner [23,24,25*].

The DPE and MTE

The DPE (downstream core promoter element) was identified as a downstream TFIID recognition sequence that is important for basal transcription activity [26]. The DPE is conserved from *Drosophila* to humans, and is located from +28 to +33 relative to the A+1 in the Inr. The DPE consensus is RGWYVT in *Drosophila* [27]. The DPE consensus in humans has yet to be determined; however, mammalian core promoters containing sequences that conform to the *Drosophila* consensus have been found to possess DPE activity. The DPE functions cooperatively with the Inr, and the spacing between the Inr and DPE is critical for optimal transcription. Photocrosslinking studies revealed that the DPE is in close proximity to the TFIID subunits TAF6 (TAFII60) and TAF9 (TAFII40), which contain histone fold motifs and are related to histones H4 and H3 [28]. It is thus possible that the TAF6-TAF9 subunits of TFIID interact with the DPE in a manner that is similar to binding of histones H3-H4 to DNA in nucleosomes [29].

The MTE (motif ten element) was found by a combination of computational and biochemical studies [14,30]. The MTE has a consensus of CSARCSSAAC from +18 to +27 relative to A +1 in the Inr in *Drosophila*. Mutation of the nucleotides from +18 to +22 can abolish MTE activity in vitro and in cultured cells. Like the DPE, the MTE functions cooperatively with the Inr with a strict Inr-MTE spacing requirement. The addition of an MTE can compensate for the loss of basal transcription activity that occurs upon mutation of a TATA box or a DPE. Moreover, the MTE exhibits synergy with the TATA and DPE motifs. This synergy between the MTE and other core promoter motifs inspired the design of a Super Core Promoter, SCP, which contains optimized versions of the TATA box, Inr, MTE, and DPE [31*]. The SCP is

the strongest known core promoter in vitro and in cultured cells. In addition, the SCP exhibits unusually high affinity for the binding of TFIID.

The available evidence indicates that the MTE is present in humans. First, mutation of the MTE in a *Drosophila* core promoter causes a reduction in transcription by human factors both in vitro and in cultured cells [30,31*]. Hence, the human transcriptional machinery recognizes the MTE. Second, a human core promoter containing a functional MTE has been identified [30]. However, the MTE generally does not emerge as an overrepresented sequence in computational analyses of mammalian promoter databases (see, for example [15*,18*]). As described above for the Inr, it is possible that the MTE as well as the DPE may have broader and less restrictive consensus sequences in mammals than in *Drosophila*. In this regard, it is interesting to note that the analysis of mammalian core promoter sequences [18*] revealed a 'gcg motif' (which may be identical to 'motif8' of [32*]) at +20 as well as a 'gcg echo motif' at +30. The gcg motif overlaps with the mutationally sensitive +18 to +22 region of the MTE [30], and the gcg echo motif overlaps with the location of the DPE. Thus, the gcg motif may be the mammalian version of the MTE, whereas the gcg echo motif may correspond to the DPE.

The DCE and XCPE1 motifs

The DCE (downstream core element) was originally found in the human beta-globin promoter [33], and has also been characterized in the adenovirus major late promoter [34]. The DCE occurs frequently with the TATA box, and appears to be distinct from the DPE. The DCE consists of three subelements: S_I, CTTC from +6 to +11; S_{II}, CTGT from +16 to +21; and S_{III}, AGC from +30 to +34. Photocrosslinking studies revealed that the DCE is in close proximity to TAF_{II}.

The XCPE1 (X core promoter element 1) motif is located from -8 to +2 relative to the +1 start site and has the consensus sequence of DSGYGGGRASM [35*]. It is present in about 1% of human core promoters, most of which are TATA-less. XCPE1 exhibits little activity by itself. Instead, it acts in conjunction with sequence-specific activators, such as NRF1, NF-1, and Sp1. Thus, XCPE1 may be a member of a larger family of motifs that work along with sequence-specific activators in CpG islands to direct transcription initiation.

In the future, it will also be important to investigate functional interactions between different core promoter motifs. Along these lines, computational studies have revealed the co-occurrence of various combinations of core promoter motifs [15*,16*,36,37*,38*]. These studies not only confirm previously known interactions between core promoter motifs, but also identify new potential interactions.

Diversity in core promoter function

The existence of different core promoter elements results in diversity in core promoter function (reviewed in [39]). For instance, enhancers are functionally linked to core promoters (see, for example [40*]), and some transcriptional enhancers have been found to exhibit specificity for TATA versus DPE core promoter motifs [41]. In addition, different factors mediate the basic transcription process from different types of core promoters. For example, a set of purified transcription factors (TFIIA, TFIIB, TFIID, TFIIE, TFIIIF, TFIIF, TFIIF, RNA polymerase II, PC4, and Sp1) that are sufficient to transcribe a TATA-dependent core promoter were found to be unable to transcribe a DPE-dependent core promoter [42]. Moreover, in transcription reactions performed with a crude nuclear extract, it was found that NC2 (also known as Dr1-Drap1) stimulates DPE-dependent transcription and inhibits TATA-dependent transcription [43]. However, the enhancement of DPE-dependent transcription by NC2 was not seen in a purified reconstituted system [42], perhaps due to the absence of an additional auxiliary factor that was

present in the crude extract but not in the purified system. In separate work, it was also found that the Inr element contributes to resistance of repression of TATA-dependent transcription by NC2 [44*]. Furthermore, in RNAi depletion studies, TAF1 and TAF4 were observed to be important for transcription from a DPE-dependent reporter gene but not from a TATA-dependent reporter gene [45*]. At the present time, we have an intriguing yet incomplete picture of the factors involved in transcription from different types of core promoters. This subject is an important area of future investigation.

TBP-related factors

Diversity in the function of the transcription machinery can be seen with TBP and TBP-related factors (TRFs) (for recent reviews, see [46*,47*,48*]). There are three TRFs: TRF1, TRF2 (also known as TLF, TLP, TRF, and TRP), and TRF3 (also known as TBP2). TRF1 is absent in vertebrates but is present in *Drosophila*, in which it binds to a TC-rich sequence and mediates transcription by RNA polymerases II and III [49,50*]. TRF2 does not bind to TATA box sequences, and is involved in RNA polymerase II transcription in *Drosophila* and vertebrates. *Drosophila* TRF2 is in a multisubunit complex that contains ISWI and DREF proteins [51]. TRF2 in *Drosophila* has also been found to exist in a short form and a long form, both of which associate with ISWI [52*]. TRF3 is present in vertebrates but not *Drosophila*, and is the TRF that is most closely related to TBP. TRF3 binds to TATA box sequences and participates in transcription by RNA polymerase II.

It is becoming further apparent that TBP, which had been previously thought to be a universal transcription factor, is not required for transcription of many genes (see, for instance 53*, 54*]). Instead, the available data indicate that differential functions of TBP and the TRFs are involved in many different regulatory networks. For example, TRF2 is required for transcription of the TATA-less histone H1 gene but not the TATA-containing S-phase regulated core histone genes, whereas TBP exhibits the opposite effect on those genes [55*]. In addition, opposite effects of TRF2 and TBP were seen with the TATA-less neurofibromatosis type 1 (NF1) promoter and the TATA-containing *c-fos* promoter [56]. Strikingly, a novel complex containing TRF3 and TAF3 replaces the TBP-containing TFIID complex during differentiation of myoblast cells into myotubes [57*]. These findings provide an example of a TBP to TRF3 switch during cellular differentiation. In this regard, it is interesting to note that TBP and TRF3 are expressed differently during mouse oogenesis [58*]. TRF3 has also been found to be important for hematopoiesis in zebrafish [59*].

Conclusions and future prospects

The core promoter is diverse and complex. We still need to gain a better understanding of the DNA sequences that dictate core promoter function as well as the protein factors that function at the different types of core promoters. It will be particularly important to devote more effort to the study of the mechanisms of transcription at dispersed core promoters, because current evidence (see, for example [11,35*,60]) suggests that there may be fundamental differences in the strategies and mechanisms of transcription from dispersed versus focused core promoters.

It will be interesting to investigate whether the binding of RNA polymerase II to the Mediator complex (see, for example [61,62*]) influences basal transcription activity. For instance, Mediator has been found to contribute to transcription of a DPE-containing promoter [42]. Moreover, a new form of RNA polymerase II, termed Pol II(G), contains an additional protein named Gdown1 that enables RNA polymerase II to become responsive to Mediator [63*].

In addition, chromatin structure may be a component of core promoter function, as it has been found that transcription start sites are often flanked by the H2A.Z histone variant (see, for

example [64,65,66*,67*]) and are frequently located immediately upstream of trimethylated histone H3 lysine 4 (H3K4me3) (see, for example [67*,68,69*,70*]). It is possible, for instance, that transcription initiation complexes are formed with the assistance of the binding of the TAF3 subunit of TFIID to H3K4me3-containing nucleosomes [71*].

From a broader perspective, it will also be important to understand how core promoters function in biological processes, such as gene networks and development. The acquisition and integration of this information will provide us with the knowledge of the events that revolve about the core promoter – the gateway to transcription.

Acknowledgements

We are grateful to Barbara Rattner, Timur Yusufzai, and Alexandra Lusser for critical reading of this manuscript. This work was supported by a grant from the National Institutes of Health (GM041249) to J.T.K.

References

1. Smale ST, Kadonaga JT. The RNA polymerase II core promoter. *Annu Rev Biochem* 2003;72:449–479. [PubMed: 12651739]
- 2. Thomas MC, Chiang CM. The general transcription machinery and general cofactors. *Crit Rev Biochem Mol Biol* 2006;41:105–178. [PubMed: 16858867]Comprehensive review on basal transcription by RNA polymerase II
- 3. Juven-Gershon T, Hsu JY, Kadonaga JT. Perspectives on the RNA polymerase II core promoter. *Biochem Soc Trans* 2006;34:1047–1050. [PubMed: 17073747]Review on core promoter motifs
- 4. Müller F, Demény MA, Tora L. New problems in RNA polymerase II transcription initiation: matching the diversity of core promoters with a variety of promoter recognition factors. *J Biol Chem* 2007;282:14685–14689. [PubMed: 17395580]Review on core promoters
- 5. Sandelin A, Carninci P, Lenhard B, Ponjavic J, Hayashizaki Y, Hume DA. Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat Rev Genet* 2007;8:424–436. [PubMed: 17486122]Review on the generation and use of CAGE (cap analysis of gene expression) tag data for the study of mammalian RNA polymerase II core promoters
6. Reeve JN. Archaeal chromatin and transcription. *Mol Microbiol* 2003;48:587–598. [PubMed: 12694606]
7. Molina C, Grotewold E. Genome wide analysis of *Arabidopsis* core promoters. *BMC Genomics* 2005;6:25.10.1186/1471-2164-6-25 [PubMed: 15733318]
- 8. Yamamoto YY, Ichida H, Abe T, Suzuki Y, Sugano S, Obokata J. Differentiation of core promoter architecture between plants and mammals revealed by LDSS analysis. *Nucleic Acids Res* 2007;35:6219–6226. [PubMed: 17855401]The TATA box is conserved from plants to mammals
9. Bird AP. CpG-rich islands and the function of DNA methylation. *Nature* 1986;321:209–213. [PubMed: 2423876]
- 10. Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CA, Taylor MS, Engström PG, Frith MC, et al. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* 2006;38:626–635. [PubMed: 16645617]Genome-wide analysis of human and mouse transcription start sites was carried out by using cap analysis of gene expression (CAGE), which yields 20- or 21-nt sequence tags that correspond to the 5' end of capped mRNAs. TATA boxes were most commonly located from –33 to –28 (particularly, at –31 or –30) relative to the +1 start site. About 57% of the start sites had a pyrimidine-purine dinucleotide at the –1, +1 positions. This pyrimidine-purine dinucleotide appears to be the most important component of the mammalian Inr. TATA boxes were found to be associated with sharp/focussed start sites, whereas CpG islands were observed to contain more dispersed start sites
11. Lee MP, Howcroft K, Kotekar A, Yang HH, Buetow KH, Singer DS. ATG deserts define a novel core promoter subclass. *Genome Res* 2005;15:1189–1197. [PubMed: 16109972]
12. Smale ST, Baltimore D. The “initiator” as a transcription control element. *Cell* 1989;57:103–113. [PubMed: 2467742]

- 13. Yang C, Bolotin E, Jiang T, Sladek FM, Martinez E. 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* 2007;389:52–65. [PubMed: 17123746] Computational studies showed that the Inr motif is more common than the TATA box in humans, and that Inr-like sequences are present near the transcription start sites of about 40% of yeast core promoters
- 14. Ohler U, Liao GC, Niemann H, Rubin GM. Computational analysis of core promoters in the *Drosophila* genome. *Genome Biol* 2002;3:RESEARCH0087.10.1186/gb-2002-3-12-research0087 [PubMed: 12537576]
- 15. FitzGerald PC, Sturgill D, Shyakhtenko A, Oliver B, Vinson C. Comparative genomics of *Drosophila* and human core promoters. *Genome Biol* 2006;7:R53.10.1186/gb-2006-7-7-r53 [PubMed: 16827941] Computational analysis of human and *Drosophila* core promoter motifs revealed that the TATA, Inr, and DPE motifs are conserved from *Drosophila* to humans
- 16. Gershenson NI, Trifonov EN, Ioshikhes IP. The features of *Drosophila* core promoters revealed by statistical analysis. *BMC Genomics* 2006;7:161.10.1186/1471-2164-7-161 [PubMed: 16790048] Statistical analysis of *Drosophila* core promoter sequences led to predictions regarding the frequency of occurrence of the TATA (16%), Inr (66%), DPE (22%), and MTE (10%) motifs as well as specific combinations of core promoter motifs
- 17. Purnell BA, Emanuel PA, Gilmour DS. TFIID sequence recognition of the initiator and sequences farther downstream in *Drosophila* class II genes. *Genes Dev* 1994;8:830–842. [PubMed: 7926771]
- 18. Frith MC, Valen E, Krogh A, Hayashizaki Y, Carninci P, Sandelin A. A code for transcription initiation in mammalian genomes. *Genome Res* 2008;18:1–12. [PubMed: 18032727] Development of a model for the prediction of transcription start site usage in mammals. This model is based on overrepresented sequences that flank transcription start sites. Key sequence motifs include an Sp1-like recognition site, a TATA-box-like sequence, an Inr (Py, Pu at -1, +1), and two downstream sequences termed ‘gcg motif’ and ‘gcg echo motif’. The latter two motifs bear a resemblance to the downstream MTE and DPE motifs, and may function as mammalian versions of the MTE and DPE to promote the binding of TFIID. In addition, the GCG and GCG echo motifs may be related to ‘motif8’ described in [32*]
- 19. Michael, L Goldberg. PhD thesis. Stanford University; 1979.
- 20. Ponjavic J, Lenhard B, Kai C, Kawai J, Carninci P, Hayashizaki Y, Sandelin A. Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters. *Genome Biol* 2006;7:R78.10.1186/gb-2006-7-8-r78 [PubMed: 16916456] Extensive analysis of mouse CAGE tags revealed a distinct preference for the location of the TATA box at -31 or -30 relative to the +1 transcription start site
- 21. Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B. A high-resolution map of active promoters in the human genome. *Nature* 2005;436:876–880. [PubMed: 15988478]
- 22. Cooper SJ, Trinklein ND, Anton ED, Nguyen L, Myers RM. Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Res* 2006;16:1–10. [PubMed: 16344566] High-throughput transient transfection studies yielded 387 functional human promoters. About 16% of these functional promoters contained TATA box motifs. Deletion analysis of 45 promoters revealed that sequences from about -50 to -300 relative to the transcription start site generally activate transcription. Negative elements were found to be located from -350 to -1000 relative to the start site
- 23. Lagrange T, Kapanidis AN, Tang H, Reinberg D, Ebright RH. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev* 1998;12:34–44. [PubMed: 9420329]
- 24. Deng W, Roberts SG. A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes Dev* 2005;19:2418–2423. [PubMed: 16230532]
- 25. Deng W, Roberts SG. TFIIB and the regulation of transcription by RNA polymerase II. *Chromosoma* 2007;116:417–429. [PubMed: 17593382] Review on TFIIB and the BRE motifs
- 26. Burke TW, Kadonaga JT. *Drosophila* TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes Dev* 1996;10:711–724. [PubMed: 8598298]

27. Kutach AK, Kadonaga JT. The downstream promoter element DPE appears to be as widely used as the TATA box in *Drosophila* core promoters. *Mol Cell Biol* 2000;20:4754–4764. [PubMed: 10848601]
28. Burke TW, Kadonaga JT. The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAFII60 of *Drosophila*. *Genes Dev* 1997;11:3020–3031. [PubMed: 9367984]
29. Shao H, Revach M, Moshonov S, Tzuman Y, Gazit K, Albeck S, Unger T, Dikstein R. Core promoter binding by histone-like TAF complexes. *Mol Cell Biol* 2005;25:206–219. [PubMed: 15601843]
30. Lim CY, Santoso B, Boulay T, Dong E, Ohler U, Kadonaga JT. The MTE, a new core promoter element for transcription by RNA polymerase II. *Genes Dev* 2004;18:1606–1617. [PubMed: 15231738]
- 31. Juven-Gershon T, Cheng S, Kadonaga JT. Rational design of a super core promoter that enhances gene expression. *Nat Methods* 2006;3:917–922. [PubMed: 17124735] Creation of super core promoters (SCPs) for RNA polymerase II transcription by combining the TATA, Inr, MTE, and DPE motifs in a single promoter. The SCPs exhibit strong transcription in vitro and in vivo and have unusually high affinity for TFIID. This work also shows that the MTE and DPE motifs each contribute to core promoter activity in vivo in human cells
- 32. Xi H, Yu Y, Fu Y, Foley J, Halees A, Weng Z. Analysis of overrepresented motifs in human core promoters reveals dual regulatory roles of YY1. *Genome Res* 2007;17:798–806. [PubMed: 17567998] Identified the recognition motif of YY1, a sequence-specific DNA-binding protein, as a commonly occurring sequence in the downstream region of promoters that contain short 5' untranslated regions. In most instances, the YY1 recognition motif, which contains an ATG sequence, correlated with the expected translation start site. It was suggested that the YY1 recognition motif is involved in both transcription and translation. The promoter sequence analysis also revealed a GC-rich downstream promoter motif that was termed 'motif8'. Motif8 may be related to the 'gcg motif' of [18*]
33. Lewis BA, Kim TK, Orkin SH. A downstream element in the human beta-globin promoter: evidence of extended sequence-specific transcription factor IID contacts. *Proc Natl Acad Sci USA* 2000;97:7172–7177. [PubMed: 10840054]
34. Lee DH, Gershenzon N, Gupta M, Ioshikhes IP, Reinberg D, Lewis BA. Functional characterization of core promoter elements: the downstream core element is recognized by TAF1. *Mol Cell Biol* 2005;25:9674–9686. [PubMed: 16227614]
- 35. Tokusumi Y, Ma Y, Song X, Jacobson RH, Takada S. The new core promoter element XCPE1 (X Core Promoter Element 1) directs activator-, mediator-, and TATA-binding protein-dependent but TFIID-independent RNA polymerase II transcription from TATA-less promoters. *Mol Cell Biol* 2007;27:1844–1858. [PubMed: 17210644] Identification of a new core promoter element termed XCPE1, which is located from –8 to +2 relative to the +1 transcription start site. XCPE1 exhibits little promoter activity by itself; instead, it works in conjunction with sequence-specific activators such as NRF1, NF-1, and Sp1. It was estimated that about 1% of human core promoters contain an XCPE1 motif
36. Gershenzon NI, Ioshikhes IP. Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. *Bioinformatics* 2005;21:1295–1300. [PubMed: 15572469]
- 37. Jin VX, Singer GA, Agosto-Pérez FJ, Liyanarachchi S, Davuluri RV. Genome-wide analysis of core promoter elements from conserved human and mouse orthologous pairs. *BMC Bioinformatics* 2006;7:114.10.1186/1471-2105-7-114 [PubMed: 16522199] Analysis of core promoter sequences that are conserved between humans and mice suggests the use of TATA-Inr-MTE and BRE^U-Inr-MTE combinations
- 38. Ohler U. Identification of core promoter modules in *Drosophila* and their application in accurate transcription start site prediction. *Nucleic Acids Res* 2006;34:5943–5950. [PubMed: 17068082] Computational analysis of the co-occurrence of *Drosophila* core promoter motifs revealed distinct core promoter modules
39. Butler JE, Kadonaga JT. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev* 2002;16:2583–2592. [PubMed: 12381658]
- 40. Lee AM, Wu C-t. Enhancer-promoter communication at the yellow gene of *Drosophila melanogaster*: diverse promoters participate in and regulate trans interactions. *Genetics* 2006;174:1867–1880.

[PubMed: 17057235] Analyzed the effect of the core promoter on the ability of enhancers to act in trans (via transvection) in *Drosophila*. When some core promoters were weakened by mutation of either the TATA box or the DPE, their cognate enhancers were released from the core promoters and acquired the ability to activate transcription in trans. These studies demonstrate a functional link between enhancers and core promoters

41. Butler JE, Kadonaga JT. Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. *Genes Dev* 2001;15:2515–2519. [PubMed: 11581157]
42. Lewis BA, Sims RJ 3rd, Lane WS, Reinberg D. Functional characterization of core promoter elements: DPE-specific transcription requires the protein kinase CK2 and the PC4 coactivator. *Mol Cell* 2005;18:471–481. [PubMed: 15893730]
43. Willy PJ, Kobayashi R, Kadonaga JT. A basal transcription factor that activates or represses transcription. *Science* 2000;290:982–985. [PubMed: 11062130]
- 44. Malecová B, Gross P, Boyer-Guittaut M, Yavuz S, Oelgeschläger T. The initiator core promoter element antagonizes repression of TATA-directed transcription by negative cofactor NC2. *J Biol Chem* 2007;282:24767–24776. [PubMed: 17584739] The presence of the Inr confers resistance to repression of TATA-dependent transcription by NC2 (Dr1-Drap1)
- 45. Wright KJ, Marr MT 2nd, Tjian R. TAF4 nucleates a core subcomplex of TFIID and mediates activated transcription from a TATA-less promoter. *Proc Natl Acad Sci USA* 2006;103:12347–12352. [PubMed: 16895980] RNAi depletion analysis of TFIID subunits revealed a critical role of TAF4 in maintaining the integrity of the TFIID complex. TAF1 and TAF4, but not TAF5, were found to be important for transcription of a reporter gene that contains a DPE motif. In contrast, transcription from a related TATA-dependent reporter gene was not affected by depletion of TAF1 or TAF4
- 46. Jones KA. Transcription strategies in terminally differentiated cells: shaken to the core. *Genes Dev* 2007;21:2113–2117. [PubMed: 17785521] Review and analysis of TFIID and TRF3 during muscle differentiation
- 47. Reina JH, Hernandez N. On a roll for new TRF targets. *Genes Dev* 2007;21:2855–2860. [PubMed: 18006682] Review on TBP-related factors
- 48. Torres-Padilla ME, Tora L. TBP homologues in embryo transcription: who does what? *EMBO Rep* 2007;8:1016–1018. [PubMed: 17972900] Review on TBP and TBP-related factors
49. Takada S, Lis JT, Zhou S, Tjian R. A TRF1:BRF complex directs *Drosophila* RNA polymerase III transcription. *Cell* 2000;101:459–469. [PubMed: 10850489]
- 50. Isogai Y, Takada S, Tjian R, Keles S. Novel TRF1/BRF target genes revealed by genome-wide analysis of *Drosophila* Pol III transcription. *EMBO J* 2007;26:79–89. [PubMed: 17170711] ChIP-on-chip and functional analyses suggested that TRF1 and BRF in *Drosophila* are responsible for all transcription by RNA polymerase III
51. Hochheimer A, Zhou S, Zheng S, Holmes MC, Tjian R. TRF2 associates with DREF and directs promoter-selective gene expression in *Drosophila*. *Nature* 2002;420:439–445. [PubMed: 12459787]
- 52. Kopytova DV, Krasnov AN, Kopantseva MR, Nabirochkina EN, Nikolenko JV, Maksimenko O, Kurshakova MM, Lebedeva LA, Yerokhin MM, Simonova OB, et al. Two isoforms of *Drosophila* TRF2 are involved in embryonic development, premeiotic chromatin condensation, and proper differentiation of germ cells of both sexes. *Mol Cell Biol* 2006;26:7492–7505. [PubMed: 17015475] Identified a large 175 kDa form of *Drosophila* TRF2. The C-terminal region of the 175 kDa form of TRF2 is identical to the previously described 75 kDa TRF2 polypeptide. The two forms of TRF2 exist in *Drosophila*, are encoded by the same gene, and appear to have related functions
- 53. Jacobi UG, Akkers RC, Pierson ES, Weeks DL, Dagle JM, Veenstra GJ. TBP paralogs accommodate metazoan- and vertebrate-specific developmental gene regulation. *EMBO J* 2007;26:3900–3909. [PubMed: 17703192] Antisense knockdown experiments in *Xenopus* embryos (subsequent to the midblastula transition) revealed that the mRNA levels of about 69% of the expressed genes are not affected by depletion of TBP
- 54. Ferg M, Sanges R, Gehrig J, Kiss J, Bauer M, Lovas A, Szabo M, Yang L, Straehle U, Pankratz MJ, Olsz F, Stupka E, Müller F. The TATA-binding protein regulates maternal mRNA degradation and differential zygotic transcription in zebrafish. *EMBO J* 2007;26:3945–3956. [PubMed: 17703193] The steady-state mRNA levels of 65.3% of genes in dome-stage zebrafish embryos are not affected by depletion of TBP, as assessed by TBP knockdown and microarray analysis. TBP is

- involved in both activation (17.5%) and repression (17.1%) of zebrafish promoters. In addition, TBP appears to be involved in the degradation of maternal mRNA by the miR-430 microRNA
- 55. Isogai Y, Keles S, Prestel M, Hochheimer A, Tjian R. Transcription of histone gene cluster by differential core-promoter factors. *Genes Dev* 2007;21:2936–2949. [PubMed: 17978101] In *Drosophila*, TRF2 is required for transcription of the TATA-less histone H1 genes but not the TATA-containing S-phase regulated core histone genes. Conversely, TBP is required for transcription of the S-phase regulated core histone genes but not the histone H1 genes. Genome-wide analyses revealed that TRF2 is mostly localized to TATA-less core promoters that appear to contain sequences such as Motif 1, Motif 7, and the DRE
  - 56. Chong JA, Moran MM, Teichmann M, Kaczmarek JS, Roeder R, Clapham DE. TATA-binding protein (TBP)-like factor (TLF) is a functional regulator of transcription: reciprocal regulation of the neurofibromatosis type 1 and c-fos genes by TLF/TRF2 and TBP. *Mol Cell Biol* 2005;25:2632–2643. [PubMed: 15767669]
  - 57. Deato MD, Tjian R. Switching of the core transcription machinery during myogenesis. *Genes Dev* 2007;21:2137–2149. [PubMed: 17704303] Upon differentiation of myoblast cells into myotubes, the TBP-containing TFIID complex disappears and is replaced by a novel complex containing TRF3 and TAF3. These findings suggest a role of core promoter factors in cellular differentiation
  - 58. Gazdag E, Rajkovic A, Torres-Padilla ME, Tora L. Analysis of TATA-binding protein 2 (TBP2) and TBP expression suggests different roles for the two proteins in regulation of gene expression during oogenesis and early mouse development. *Reproduction* 2007;134:51–62. [PubMed: 17641088] TBP and TRF3 (TBP2) are expressed differently during oogenesis in the mouse
  - 59. Hart DO, Raha T, Lawson ND, Green MR. Initiation of zebrafish haematopoiesis by the TATA-box-binding protein-related factor Trf3. *Nature* 2007;450:1082–1085. [PubMed: 18046332] Depletion of TRF3 in zebrafish embryos results in a defect in hematopoiesis. A key target of TRF3 was found to be the *mespa* gene
  - 60. Smale ST, Schmidt MC, Berk AJ, Baltimore D. Transcriptional activation by Sp1 as directed through TATA or initiator: specific requirement for mammalian transcription factor IID. *Proc Natl Acad Sci USA* 1990;87:4509–4513. [PubMed: 2141169]
  - 61. Malik S, Roeder RG. Dynamic regulation of pol II transcription by the mammalian Mediator complex. *Trends Biochem Sci* 2005;30:256–263. [PubMed: 15896744]
  - 62. Kornberg RD. The molecular basis of eukaryotic transcription. *Proc Natl Acad Sci U S A* 2007;104:12955–12961. [PubMed: 17670940]
  - 63. Hu X, Malik S, Negroiu CC, Hubbard K, Velalar CN, Hampton B, Grosu D, Catalano J, Roeder RG, Gnatt A. A Mediator-responsive form of metazoan RNA polymerase II. *Proc Natl Acad Sci USA* 2006;103:9506–9511. [PubMed: 16769904] Found a novel form of RNA polymerase II, termed Pol II(G), that contains an additional subunit, Gdown1, which enables RNA polymerase II to respond to the Mediator
  - 64. Raisner RM, Hartley PD, Meneghini MD, Bao MZ, Liu CL, Schreiber SL, Rando OJ, Madhani HD. Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. *Cell* 2005;123:233–248. [PubMed: 16239142]
  - 65. Guillemette B, Bataille AR, Gévy N, Adam M, Blanchette M, Robert F, Gaudreau L. Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning. *PLoS Biol* 2005;3:e384.10.1371/journal.pbio.0030384 [PubMed: 16248679]
  - 66. Albert I, Mavrich TN, Tomsho LP, Qi J, Zanton SJ, Schuster SC, Pugh BF. Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome. *Nature* 2007;446:572–576. [PubMed: 17392789] Genome-wide analysis of H2A.Z localization in *S. cerevisiae*
  - 67. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. High-resolution profiling of histone methylations in the human genome. *Cell* 2007;129:823–837. [PubMed: 17512414] Comprehensive ChIP-Seq (chromatin immunoprecipitation followed by genome-wide sequencing of the immunoprecipitated DNA) analysis of various histone methylations and chromatin-related proteins revealed the frequent occurrence of a peak of trimethylated H3K4 immediately downstream of the transcription start sites of active as well as inactive promoters
  - 68. Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, McMahon S, Karlsson EK, Kulbokas EJ 3rd, Gingeras TR, Schreiber SL, Lander ES. Genomic maps and

comparative analysis of histone modifications in human and mouse. *Cell* 2005;120:169–181. [PubMed: 15680324]

- 69. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 2007;39:311–318. [PubMed: 17277777]ChIP-on-chip analysis revealed that promoters exhibit trimethylation of H3K4, whereas enhancers contain monomethylation but not trimethylation of H3K4
- 70. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* 2007;130:77–88. [PubMed: 17632057]ChIP-on-chip analysis revealed that most active and inactive promoters in human embryonic stem cells contain trimethylated H3K4 immediately downstream of their transcription start sites. H3K9,K14 acetylation and hypophosphorylated RNA polymerase II were also observed at promoters
- 71. Vermeulen M, Mulder KW, Denissov S, Pijnappel WW, van Schaik FM, Varier RA, Baltissen MP, Stunnenberg HG, Mann M, Timmers HTM. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* 2007;131:58–69. [PubMed: 17884155]The PHD finger of the TAF3 subunit of human TFIID binds to nucleosomes containing trimethylated H3K4. This finding suggests that there is a key functional interaction between TFIID and a downstream nucleosome containing trimethylated H3K4

Focused versus Dispersed Transcription Initiation

Figure 1.

Focused versus dispersed core promoters. In focused core promoters, transcription initiates at a single site or in a cluster of sites in a narrow region of several nucleotides. Dispersed core promoters are typically found in CpG islands in vertebrates and usually yield multiple weak start sites over a region of 50 to 100 nucleotides. Focused core promoters are more ancient and widespread throughout nature than dispersed core promoters. In vertebrates, however, dispersed promoters are more common than focused promoters. There may be fundamental differences in the basic mechanisms of transcription from focused versus dispersed core promoters.

Core Promoter Elements

Figure 2.

Core promoter motifs. This diagram, which is drawn roughly to scale, shows some of the known core promoter elements for transcription by RNA polymerase II. There are no universal core promoter elements. Each of these elements is found in only a fraction (typically estimated to be from 1% to 30%, depending on the motif) of all core promoters. The Inr is probably the most commonly occurring core promoter motif. There are additional core promoter elements that remain to be discovered. The TATA box, Inr, MTE, DPE, and DCE are recognition sites for the binding of transcription factor TFIID. It should be noted, however, that there are multiple forms of TFIID and TFIID-related protein complexes that could potentially interact with the core promoter. BRE^u and BRE^d interact with TFIIB.