Eukaryotic core promoters and the functional basis of transcription initiation

Vanja Haberle[1] *and Alexander Stark*[1,2]*

Abstract | RNA polymerase II (Pol II) core promoters are specialized DNA sequences at transcription start sites of protein-coding and non-coding genes that support the assembly of the transcription machinery and transcription initiation. They enable the highly regulated transcription of genes by selectively integrating regulatory cues from distal enhancers and their associated regulatory proteins. In this Review, we discuss the defining properties of gene core promoters, including their sequence features, chromatin architecture and transcription initiation patterns. We provide an overview of molecular mechanisms underlying the function and regulation of core promoters and their emerging functional diversity, which defines distinct transcription programmes. On the basis of the established properties of gene core promoters, we discuss transcription start sites within enhancers and integrate recent results obtained from dedicated functional assays to propose a functional model of transcription initiation. This model can explain the nature and function of transcription initiation at gene starts and at enhancers and can explain the different roles of core promoters, of Pol II and its associated factors and of the activating cues provided by enhancers and the transcription factors and cofactors they recruit.

Core promoter
A short sequence flanking the transcription start site that is sufficient to assemble the RNA polymerase II transcription machinery and initiate transcription.

General transcription factors
(GTFs). Proteins that together with RNA polymerase II constitute the transcription machinery at the core promoter.

[1]*Research Institute of Molecular Pathology (IMP), Vienna Biocenter (VBC), Vienna, Austria.*

[2]*Medical University of Vienna, Vienna Biocenter (VBC), Vienna, Austria.*

**e-mail: stark@starklab.org*

https://doi.org/10.1038/s41580-018-0028-8

The development of complex organisms with many morphologically and functionally diverse cell types from a single cell is largely determined by the genetic information contained within genomic DNA[1,2]. This genetic information includes both protein-coding sequences of genes and non-coding regulatory elements that govern when, where and to what level each gene will be expressed. Regulated gene expression is essential for the integrity of all eukaryotic cells and organisms[3] and has a central role in cell differentiation and metabolism, and its disruption leads to disease[4].

Gene expression starts with transcription, the copying of a DNA sequence into an RNA transcript by RNA polymerase II (Pol II), which transcribes all protein-coding and many non-coding genes. Transcription typically initiates at a defined position, the transcription start site (TSS), at the 5′ end of a gene, which we refer to as the gene start. The TSS is embedded within a core promoter, which is a short sequence encompassing ~50 bp upstream and ~50 bp downstream of the TSS (FIG. 1a). The core promoter serves as a binding platform for the transcription machinery, which comprises Pol II and its associated general transcription factors (GTFs)[5]. Core promoters are sufficient to direct transcription initiation[6] but generally have low basal activity, which can be further suppressed by chromatin or activated by often more distally located regulatory elements called enhancers[1,7,8]. Enhancers bind regulatory proteins known as transcription factors and recruit transcriptional cofactors (reviewed in REFS[1,9]) and can increase transcription from a core promoter independently of their relative distance and orientation[1,7,8]. More recently, this traditional view of gene expression and the role of enhancers and core promoters has been challenged by the observation that many genomic positions outside annotated gene starts initiate transcription, including positions within enhancers (FIG. 1b).

Genome-wide transcription initiation

Sites of transcription initiation can be identified using various methods that capture the 5′ ends of Pol II transcripts by exploiting their characteristic properties. For example, cap analysis of gene expression (CAGE)[10] and similar 5′ end-capture approaches[11,12] take advantage of the cap structure at the 5′ end of Pol II transcripts to detect the TSS and RNA abundance. Complementary methods use properties of nascent transcripts associated with Pol II to detect their TSSs and assess their transcription rates[13–16], thereby distinguishing true initiation events from sites of potential post-transcriptional cleavage and recapping[17].

Applying such large-scale approaches to map TSSs genome-wide in different cell types of various model

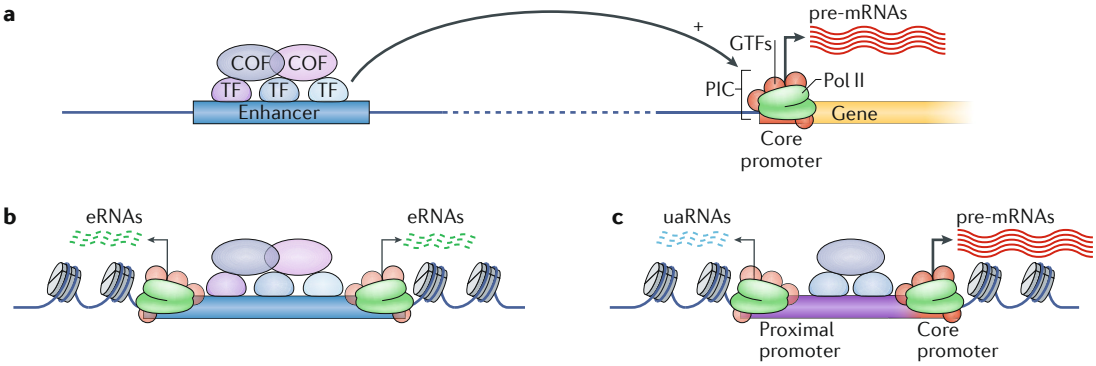


Fig. 1 | **Properties and function of core promoters and enhancers. a** | The traditional view of transcription initiation postulates that transcription initiates at gene core promoters, which recruit the transcription machinery, consisting of RNA polymerase II (Pol II) and general transcription factors (GTFs), thereby leading to the formation of the pre-initiation complex (PIC) and transcription initiation. Transcription from core promoters is activated by enhancers, which can be located distally and bind sequence-specific transcription factors (TF), which recruit cofactors (COF) that convey the activating cues to the PIC at the core promoter. **b** | Active enhancers exhibit divergent transcription of short, unstable enhancer RNAs (eRNAs) from two separate transcription start sites located at the edges of the nucleosome-depleted region (NDR) where the enhancer resides. **c** | Promoters produce long, stable mRNAs from a gene core promoter in the sense direction (orientation of the gene) and short, unstable upstream antisense RNAs (uaRNAs) from the upstream edge of an NDR that contains the transcription factor-bound proximal promoter. Separate PICs drive unidirectional transcription from each of the two TSSs.

Enhancers
DNA sequences that contain binding sites for sequence-specific transcription factors and increase the level of transcription from distal core promoters, independently of distance and orientation.

Transcription factors
Proteins that directly bind a specific DNA sequence through their DNA-binding domain and regulate the level of transcription by recruiting RNA polymerase II or transcriptional cofactors through their trans-activation domain.

Transcriptional cofactors
Proteins that do not directly bind DNA but are recruited by DNA-binding transcription factors to regulate transcription of target genes.

Enhancer RNAs
(eRNAs). Short (<2 kb) unstable non-coding RNAs, usually not spliced or polyadenylated, which are transcribed from enhancers and are rapidly degraded by the exosome.

Nucleosome-depleted region
(NDR). A genomic region depleted of canonical nucleosomes; it is usually associated with active regulatory elements such as promoters and enhancers.

Promoters
Genomic regions encompassing a gene core promoter and an upstream proximal promoter, which together autonomously drive transcription.

Proximal promoter
A transcription-activating sequence immediately upstream of the core promoter (typically up to 250 bp upstream of the transcription start site) that contains binding sites for sequence-specific transcription factors and functions like an enhancer.

organisms[12,18–22] is not only building comprehensive catalogues of gene TSSs and the regulation of transcription initiation; it has also revealed the pervasive transcription of eukaryotic genomes[23,24]. Transcription initiation at many positions distal to annotated gene starts, especially at enhancers, is challenging the traditional model of gene expression, which has implied that transcription is initiated specifically at gene core promoters and regulated by distally located enhancers[14,15,25,26] (FIG. 1a).

Transcription initiation at enhancers

Widespread transcription of mammalian enhancers has been detected in many cell types[14,25–28], and the production of enhancer RNAs (eRNAs) was suggested to be predictive of active enhancers[26,29]. Indeed, eRNA transcription correlates with target gene transcription in inducible systems[30,31] and in different cell types[26] and often, though not always, precedes target-gene activation[29,31].

Transcription from enhancers is often bidirectional[15,26] and initiates at two distinct sites, which drive divergent transcription from the edges of a nucleosome-depleted region (NDR) that is established at active enhancers (FIG. 1b). However, unlike gene core promoters, which support the production of stable transcripts, enhancers mainly produce short, unstable transcripts in both directions[15,32].

Antisense transcription at promoters

Bidirectional transcription was also detected at promoters, where the transcription of protein-coding genes is often coupled with the transcription of short non-coding RNAs in the reverse orientation[15,33–36]. These antisense transcripts, known as promoter upstream transcripts or upstream antisense RNAs (uaRNAs), are transcribed by separate Pol II complexes from divergently oriented TSSs located at the upstream edge of the nucleosome-depleted proximal promoter region, which contains transcription-factor binding sites[37,38] (FIG. 1c). Similarly to eRNAs, these antisense transcripts are typically unstable, although some promoters seem to produce long and polyadenylated divergent transcripts[39,40].

The observed divergent transcription at promoter and enhancer regions, together with other similarities, prompted the proposal of a unified architecture of transcription initiation at those elements[15,41,42]. According to this model, promoters and enhancers both initiate transcription similarly, but only at gene promoters are transcripts stabilized post-initiation by the presence of 5′ splice sites and by the absence of premature polyadenylation signals[15,43,44].

In this Review, we first summarize the insights obtained from studying core promoters of annotated genes and then discuss to what extent the properties of these bona fide core promoters can be found at TSSs within other genomic regulatory elements, including enhancers. This order of discussion reflects the notion that gene core promoters have specifically evolved to initiate stable transcripts in a highly regulated manner, whereas the cause and the role of transcription initiation outside gene starts have remained unclear. We further discuss the assembly and activation of the transcription machinery at core promoters and discuss how this machinery is regulated by distal enhancers through transcription factors and cofactors. Finally, we integrate these established promoter properties with recent results from dedicated functional assays to propose a functional model of transcription initiation that can account for transcription from promoters and from enhancers on the basis of the sequence-encoded activities of these elements.

Properties of gene core promoters

Mapping endogenous transcription initiation sites[14–16,19–22,45] has characterized different features of core promoters, including their diverse sequence and chromatin properties and the focused or dispersed distribution of transcription initiation sites, which together define three different types of core promoter[46] (BOX 1).

Pre-initiation complex
(PIC). A large complex of proteins, including RNA polymerase II and its general transcription factors, that assembles at core promoters and is required for transcription initiation.

Sequence properties

By definition, the main task of core promoters is to support the assembly of the pre-initiation complex (PIC), which consists of Pol II and GTFs, and to guide transcription initiation from precise positions at defined levels[6]. The important role of the core-promoter sequence in conferring these functions was recently corroborated by analysing single nucleotide polymorphisms and other genetic variants, which across different fruitfly strains affected both transcription levels and TSS choice within core promoters[47]. These variations were often found to disrupt crucial sequence features known as core-promoter motifs, many of which are known to recruit GTFs and mediate PIC assembly (TABLE 1).

Box 1 | Transcription initiation patterns and core-promoter types

The comprehensive mapping of gene core promoters has revealed several transcription initiation patterns and specific sequence and chromatin properties.

Dichotomy of the promoter shape

Mapping endogenous transcription initiation at single nucleotide resolution revealed striking differences between core promoters[45,58], leading to the classification of 'focused' or 'sharp' core promoters, which have a single, well-defined transcription start site (TSS; see the figure, part **a**), and 'dispersed' or 'broad' promoters[45], which have multiple closely spaced TSSs that are used with similar frequency (see the figure, parts **b,c**). These transcription initiation patterns (or 'promoter shapes') are found across species, including in fish[21] and flies[12,19,68], and are associated with distinct gene categories: focused initiation preferentially occurs in core promoters of highly cell-type-specific genes with restricted expression patterns, whereas dispersed initiation is mainly associated with housekeeping genes expressed in many cell types[19,22,45,68] and in mammals with CpG-island (CGI)-overlapping promoters of regulators of development.

Three types of core promoters

On the basis of different properties, including initiation pattern, sequence composition and motifs, chromatin configuration and gene function, three main types of core promoter in metazoa have been proposed[46]. First, core promoters with sharp initiation patterns, imprecisely positioned nucleosomes[89] and TATA-box and initiator (Inr) motifs (see the figure, part **a**). These promoters tend to have key regulatory elements near their TSSs[235] and are active in terminally differentiated cells in adult tissues, in which case they acquire histone H3 Lys 4 trimethylation (H3K4me3) and H3 Lys 27 acetylation (H3K27ac), which are associated with active transcription. Second, core promoters of broadly expressed housekeeping genes, which are associated with dispersed transcription initiation[19,45] and a well-defined nucleosome-depleted region flanked by precisely positioned nucleosomes[89] marked by H3K4me3 and H3K27ac (see the figure, part **b**). In mammals, these core promoters overlap individual CGIs[45]; in flies, they are enriched in a specific set of variably positioned motifs including Ohler 1, Ohler 6 and DNA replication-related element (DRE)[68]. Third, core promoters of key developmental transcription factors involved in patterning and morphogenesis. In mammals they resemble housekeeping gene core promoters, which in embryonic stem cells, however, are distinctly bivalently marked with both H3K4me3 and the repressive chromatin modification H3K27me3 (REF.[236]; see the figure, part **c**). This presumably primes them for activation in the correct cell lineage and for silencing in all other cells. In mammals, such 'poised' promoters are associated with long individual CGIs or multiple CGIs[75] and often produce long non-coding divergent transcripts[39,40]. In flies, promoters of this class tend to contain a downstream promoter element (DPE) and have focused initiation[62]. In both mammals and flies, they are often surrounded by arrays of highly conserved non-coding elements, which might act as distal enhancers[62,75].

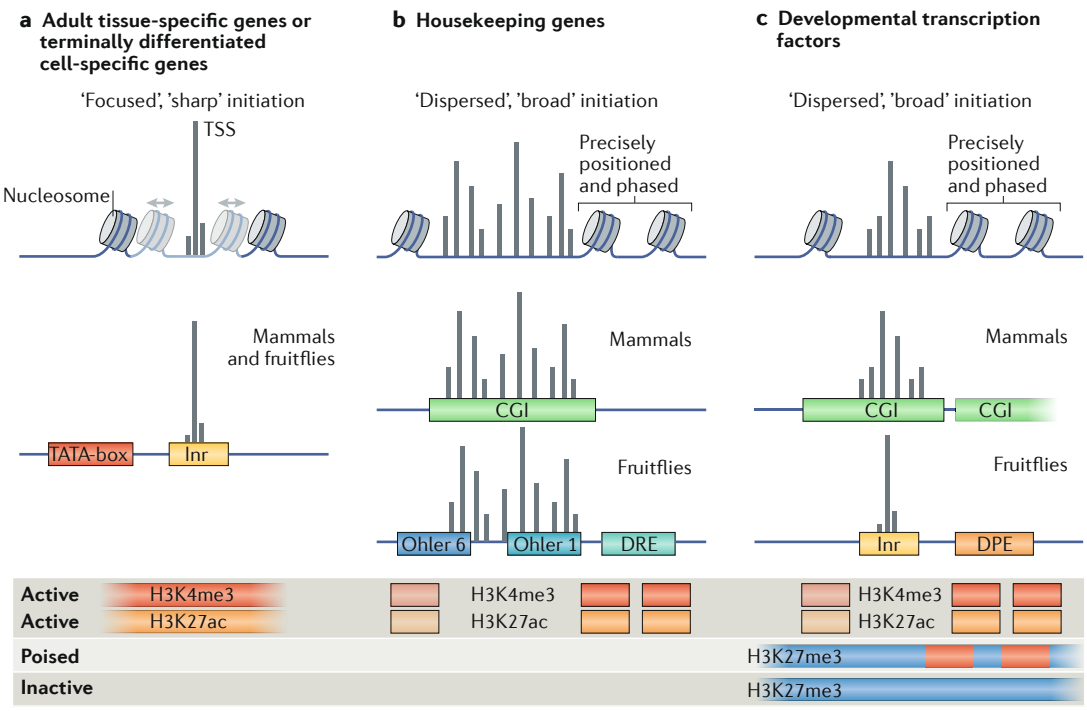


Table 1 | **Known core-promoter motifs and the (general) transcription factors that bind to them**

Core-promoter motif	Sequence logo	Consensus sequence[a]	Position relative to TSS	Bound by	Fly	Human
TATA-box		TATAWAWR[49,241]	−31 to −24	TBP[53,242]	+	+
Inr (fly)		TCAGTY[56,243]	−5 to −2	TAF1 and TAF2 (REF.[57])	+	−
Inr (human)		YR[45]	−1 to +1	NA	−	+
		BBCABW[58]	−3 to +3			
DPE		RGWCGTG[59]	+28 to +34	TAF6 and TAF9 (REF.[60]) and possibly TAF1 (REF.[55])	+	Possibly rarely
		RGWYVT[61]	+28 to +33			
		GCGWKCGGTTS[51]	+24 to +32		+	−
MTE		CSARCSSAACGS[63]	+18 to +29	Possibly TAF1 and TAF2 (REF.[55])	+	−
Ohler 1		YGGTCACACTR[51]	−60 to −1	M1BP[244]	+	−
Ohler 6		KTYRGTATWTTT[51]	−100 to −1	NA	+	−
Ohler 7		KNNCAKCNCTRNY[51]	−60 to +20	NA	+	−
DRE		WATCGATW[245]	−100 to −1	Dref[245]	+	+
TCT		YYCTTTYY[246]	−2 to +6	NA	+	+
BREu		SSRCGCC[64]	−38 to −32	TFIIB[64]	+	+
BREd		RTDKKKK[65]	−23 to −17	TFIIB[65]	+	+
DCEI–DCEIII	NA	CTTC	+6 to +11	TAF1 (REF.[67])	−	+
		CTGT	+16 to +21			
		AGC[66]	+30 to +34			
XCPE1	NA	DSGYGGRASM[247]	−8 to +2	NA	?	+
XCPE2	NA	VCYCRTTRCMY[248]	−9 to +2	NA	?	+
Pause button		KCGRWCG[155]	+25 to +35	NA	+	?

BREd, TFIIB recognition element, downstream; BREu, TFIIB recognition element, upstream; DCE, downstream core element; DPE, downstream promoter element; DRE, DNA replication-related element; Dref, DNA replication-related element factor; Inr, initiator; M1BP, motif 1-binding protein; MTE, motif ten element; NA, not available; TAF, transcription initiation factor TFIID subunit or TBP-associated factor; TBP, TATA-box-binding protein; TCT, polypyrimidine initiator; TFIIB, general transcription factor IIB; TSS, transcription start site; XCPE, X core-promoter element. [a]For the consensus sequences, B = C or G or T; D = A or G or T; K = G or T; M = A or C; R = A or G; S = C or G; V = A or C or G; W = A or T; Y = C or T. '+' indicates the presence and '−' indicates the absence of the motif in the respective species. '?' indicates that the presence or absence is not conclusive based on the literature.

Core-promoter motifs. Several core-promoter motifs have fixed positioning relative to a single, well-defined TSS. For example, the well-known TATA-box motif[48,49] is located ~30 bp upstream of a single dominant TSS[50] in 'focused' core promoters (BOX 1). Although the TATA-box is conserved from yeast to humans, it is found in only a minority of core promoters, for instance, ~5% of core promoters in flies[51,52]. The TATA-box is recognized and bound by the TATA-box-binding protein[53] (TBP; TABLE 1), one of the components of the transcription factor IID (TFIID) complex, which is a GTF that mediates Pol II recruitment and PIC assembly[54,55] and thereby might determine the choice of TSS at a fixed downstream position.

Another core-promoter motif with a fixed position relative to transcription initiation is the initiator (Inr) motif, which directly overlaps the TSS[56]. The Inr motif is more abundant than the TATA-box[52] but is not universal, and its consensus sequence differs between flies and humans. The fly Inr motif is longer and more information-rich than the human motif and encompasses several nucleotides that are adjacent to the TSS and have been shown to serve as a binding site for additional components of TFIID[57] (TABLE 1). By contrast, the human Inr motif was initially defined as a pyrimidine (C or T) followed by a purine (A or G) positioned such that the purine is the first transcribed nucleotide[45]. However, more recently, a human Inr motif with higher information content was found in focused core promoters, and several nucleotides outside the dinucleotide core motif were suggested to be important for transcription initiation in vitro[58] (TABLE 1).

In promoters that lack a TATA-box, the Inr motif is often accompanied by another motif, the downstream promoter element (DPE), which is positioned downstream of the TSS[59] (TABLE 1). The DPE motif was initially discovered in flies and, on the basis of the investigation of individual promoters, was suggested to also be present in humans[60], even though it was never found overrepresented in human promoters[45,52]. Several subunits of TFIID are suggested to bind the DPE, and a strict requirement for Inr motif–DPE spacing is thought to be essential for cooperative binding of TFIID[55,60]. Because the TATA-box and DPE rarely co-occur in flies, they were suggested to be associated with functionally distinct groups of genes[51,52,61,62] (BOX 1).

In addition to these three most abundant core-promoter motifs, other motifs with defined positions relative to the TSS include the motif ten element (MTE)[63] in flies and TFIIB recognition elements (BREs)[64,65] and downstream core elements (DCEs)[66] in humans. These motifs are bound by specific GTFs in vitro[64,67] (TABLE 1), thus potentially mediating PIC recruitment and assembly. Furthermore, analysis of large collections of core promoters allowed the computational definition of overrepresented sequences, leading to the discovery of other motifs without apparent spacing requirements relative to the TSS[51,52]. In flies, these include Ohler motifs 1, 6 and 7 and the DNA replication-related element (DRE), which were found mainly in promoters with dispersed initiation patterns associated with housekeeping genes[51,68] (BOX 1).

The described core-promoter motifs are overrepresented in gene core promoters and are more rarely associated with non-genic initiation sites. Some enhancer TSSs and promoter antisense TSSs contain weak or degenerate forms of TATA-box or Inr motifs[15,26,38], and the closer such motifs are to the consensus, the more promoter-like the enhancers are[69] (see below).

The discovery of core-promoter motifs and their importance for transcription initiation has motivated the design of synthetic core promoters that efficiently assemble the PIC and support a high level of transcription initiation for transgene expression in both fly and human systems[70–72]. Such promoters are also often used for biochemical and structural characterization of the PIC.

Characteristic (di)nucleotide composition. Apart from defined sequence motifs, gene core promoters often have distinct nucleotide compositions. For example, in vertebrates, many core promoters overlap with CpG islands (CGIs), which are regions with elevated GC content and a high density of CpG dinucleotides[73]. CGI promoters typically lack defined motifs and are mainly associated with housekeeping genes[45,74] or key developmental regulators involved in embryo patterning and morphogenesis[75] (BOX 1). The mechanisms by which CGIs confer core-promoter function are still unknown.

Characteristic patterns of dinucleotide composition have also been found downstream of the TSS, where A- or T-containing dinucleotides occur in periodic patterns[21,76]. The similarity between such patterns and the preferential sequence composition reported to underlie nucleosomal DNA[77–79] suggests a close connection between nucleosome positioning and TSS positions, especially at core promoters that lack motifs and have broad initiation patterns[21,22,76].

Chromatin configuration

Whereas most genomic DNA has limited accessibility, as it is wrapped around histone octamers to form nucleosomes, active core promoters are devoid of nucleosomes, which makes them accessible and allows PIC assembly and Pol II recruitment. Indeed, NDRs flanked by precisely positioned and phased downstream nucleosomes are hallmarks of active core promoters in all eukaryotic cells[80–82]. However, recent studies suggested that such NDRs are not depleted of nucleosomes but rather are occupied by highly dynamic nucleosomes containing the histone variants H3.3 and H2A.Z (REF.[83]) and other non-canonical or partial nucleosomal particles[84–86]. These features were proposed to ensure accessibility of the transcription machinery and associated factors to DNA, suggesting that nucleosome occupancy and accessibility to DNA at core promoters are not necessarily mutually exclusive[87,88].

Promoters with different initiation patterns differ in chromatin architecture and nucleosome positioning: dispersed promoters have more clearly defined NDRs and are associated with well-positioned nucleosomes downstream of the TSS[89] (BOX 1). Similarly, in yeast, two distinct types of promoter can be distinguished by the presence of either fragile nucleosomes or stably

CpG islands
(CGIs). GC-rich genomic sequences with a frequency of CpG dinucleotides that is higher than that found in the rest of the genome (which is generally depleted of CpG dinucleotides in mammals).

positioned nucleosomes, which correlate with distinct underlying sequences[90].

Despite the obvious correlation between open, accessible chromatin and active transcription from promoters, the causal relationship between the two is still not clear. There is evidence that some transcription factors, sometimes called pioneer factors[91], can bind to closed chromatin and recruit chromatin-remodelling factors to open the chromatin, thereby allowing Pol II binding and transcription initiation[92,93] (reviewed in REF.[9]). Similarly, the presence of H2A.Z in the first downstream (+1) nucleosome is believed to decrease the barrier this nucleosome imposes on transcribing Pol II (REF.[94]). A complementary possibility is that a low level of transcription by Pol II is required to keep the chromatin open and allow transcription factors to bind[38,95,96]. These mechanisms are not mutually exclusive and are likely combined, presumably providing different contributions at different types of core promoter[96]. H3.3, for example, appears to be both downstream and upstream of transcription: it is deposited into nucleosomes independently of DNA replication[97] and preferentially at promoters and enhancers[98], where it replaces the canonical H3 histone that is ejected during transcription. Once it accumulates at promoters, it could facilitate subsequent rounds of transcription[98].

Post-translational histone modifications. Another prominent feature of promoter-associated chromatin is the presence of specific post-translational modifications of histones[99,100]. Nucleosomes downstream of active promoters bear tri-methylation of histone H3 Lys4 (H3K4me3) and acetylation of H3 Lys27 (H3K27ac)[100] (BOX 1). Whether and how these modifications contribute to promoter function is unclear. In budding yeast, for example, H3K4 methylation occurs downstream of transcription and is mediated by the recruitment of histone-lysine *N*-methyltransferase, H3 lysine 4-specific (SET1) by the transcribing Pol II (REF.[101]). H3K4me3 was suggested to provide a memory ('bookmark') of recent transcriptional activity, thereby facilitating new rounds of transcription[101]. However, the rapid and complete loss of H3K4me3 and transcription in the absence of transcription activators suggests that H3K4me3 alone is not sufficient to maintain active transcription[102]. A bookmarking function was also proposed for H4K5ac, which can recruit the transcriptional cofactor bromodomain-containing protein 4 (BRD4) and facilitate postmitotic reactivation of a previously active genomic locus[103]. Histone acetylation might work through decreasing the affinity of DNA for nucleosomes and promoting open chromatin, similarly to acetylation of the histone core[104–106], or by directly providing binding sites for cofactors that bind acetylated Lys residues, such as BRD4 (REF.[107]).

Although H3K4me3 and H3K27ac correlate strongly with transcriptional activity, whether they are causally involved in transcription is not clear. H3K4me3 seems dispensable for transcription in flies, as cells expressing non-methylatable forms of both canonical and variant H3 histones show regulated transcription[108,109]. Similarly, cells with a Lys-to-Arg mutation at position 27 on canonical histone H3 exhibit derepression of Polycomb-silenced genes, implying that transcription does not require Lys27 acetylation at canonical H3 (REF.[110]). This suggests that Lys27 acetylation of the histone variant H3.3 is important or that histone acetylation is only a byproduct of histone acetyltransferase p300, whose relevant targets could include transcription factors[111–113] and the Pol II complex itself[114]. Such data, together with recent studies that found the pervasive enhancer mark H3K4me1 to be dispensable for enhancer activity[115,116], caution against attributing functions to histone modifications based purely on correlation and emphasize the need for functional studies to discern causation from correlation[117].

A striking example of histone modifications that causally direct transcription was recently found at Piwi-interacting RNA source loci in fly heterochromatin. Transcription of these loci is carried out by an alternative transcription machinery that is specifically recruited to the heterochromatin mark H3K9me3 through the H3K9me3 reader heterochromatin protein 1 (HP1)[118]. Although this shows that histone modifications associated with bona fide core promoters are not necessarily required for transcription, it also demonstrates that in principle modified histones are able to modulate transcription.

Transcription initiation at promoters

Transcription from gene core promoters is a step-wise process that results in defined transcriptional output. Understanding the molecular mechanisms underlying each of the individual steps is essential for understanding their activation by distal cues.

Role of the pre-initiation complex

Assembly of the PIC at core promoters and initiation of transcription involves six GTFs, which recognize and bind core-promoter elements, recruit Pol II and activate it for productive transcription[119] (FIG. 2a). A sequential model of PIC assembly, proposed on the basis of biochemical and structural studies, includes the recognition of core-promoter elements by TFIID, the binding of TFIIA and TFIIB, the recruitment of the Pol II–TFIIF complex and, finally, the binding of TFIIE followed by the binding of TFIIH (reviewed in REFS[120,121]). This model was further supported by a recent single-molecule imaging study that provided additional insight into the dynamics of GTF binding[122]. PIC assembly is followed by DNA duplex melting and the formation of an open PIC, which supports the synthesis of the first nucleotides of the nascent transcript, after which Pol II is released from the core promoter and from the GTFs that bind it ('promoter escape'; FIG. 2b). High-resolution structures of both closed and open PICs, including double-stranded DNA and melted DNA, respectively, revealed contacts between individual GTFs and core-promoter DNA and shed light on the molecular events leading to PIC assembly, promoter opening and transcription initiation at core promoters[55,123,124].

Both biochemical and structural studies agree that TFIID has a central role in recognizing and binding core-promoter elements and nucleating PIC assembly.

Piwi-interacting RNA
A type of small non-coding RNA (26–31 nucleotides) that interacts with Argonaute proteins from the Piwi family and mediates transcriptional and post-transcriptional gene silencing of transposable elements.

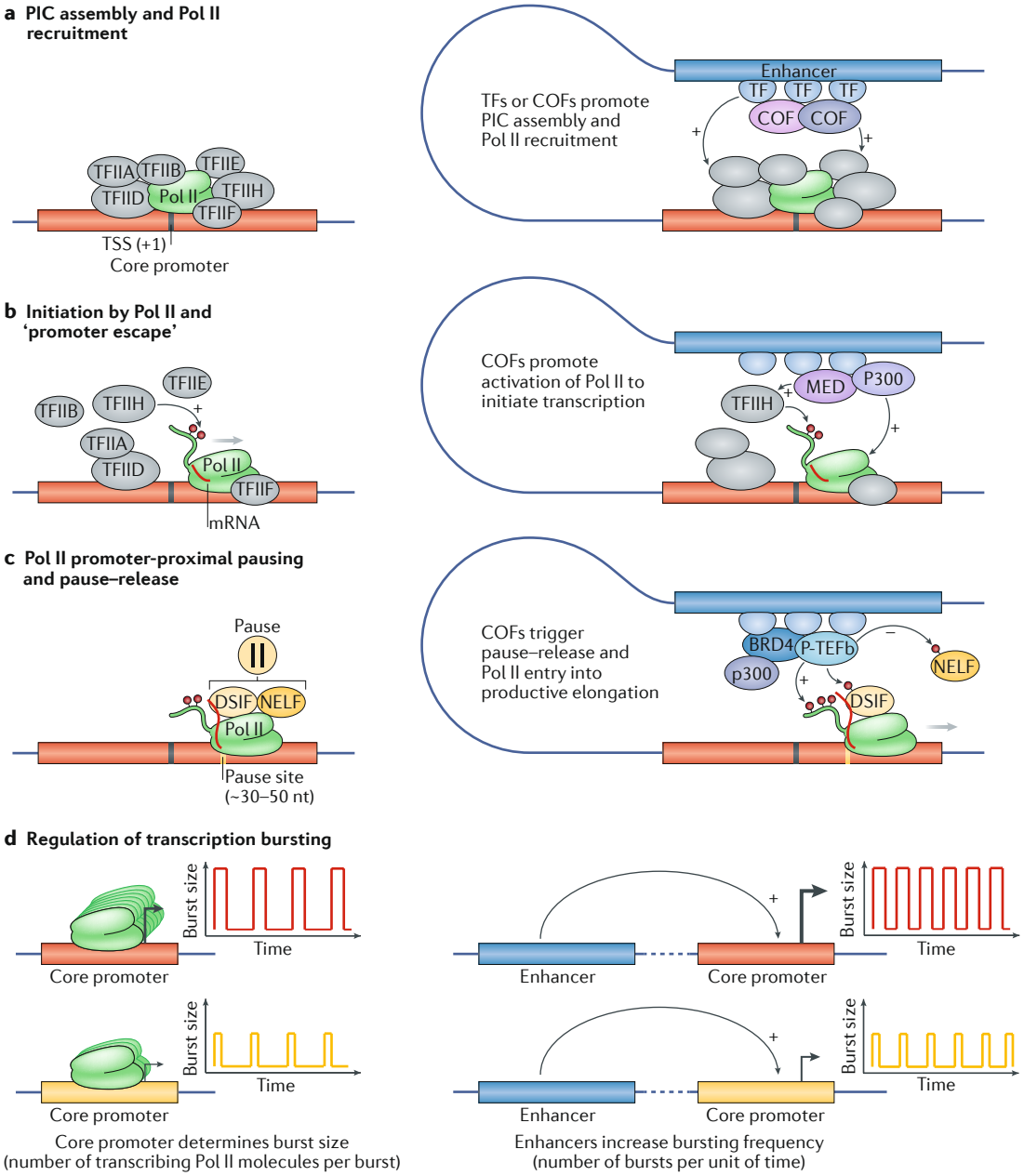


Fig. 2 | **Regulation of different steps of transcription from core promoters. a** | Pre-initiation complex (PIC) assembly and RNA polymerase II (Pol II) recruitment. The first step of transcription initiation is the assembly of the PIC, consisting of Pol II and six general transcription factors (GTFs): transcription factor IIA (TFIIA), TFIIB, TFIID, TFIIE, TFIIF and TFIIH (left image). Enhancers can promote PIC assembly by recruiting transcription factors (TFs) and cofactors (COFs) that directly interact with GTFs or Pol II (right image). **b** | Initiation by Pol II and 'promoter escape'. After PIC assembly, the DNA duplex at core promoters melts (not shown) and allows Pol II to initiate transcription at the transcription start site (TSS). To continue transcribing, Pol II has to dissociate (escape) from the TSS-binding GTFs, which is mediated by phosphorylation of Ser5 and Ser7 of the Pol II carboxy-terminal domain (CTD) by TFIIH. Enhancers can aid this process by recruiting cofactors such as the Mediator complex (MED) or the histone acetyltransferase p300 (see main text for more information). **c** | Pol II promoter-proximal pausing and pause–release. After escaping from the TSS, Pol II synthetizes a short stretch of nascent RNA (~30–50 nucleotides) and then pauses downstream of the TSS. DRB sensitivity-inducing factor (DSIF) and negative elongation factor (NELF) bind to Pol II and the nascent RNA and promote Pol II pausing. Pause–release is mediated by cyclin-dependent kinase 9 (CDK9), which is a subunit of positive transcription elongation factor b (P-TEFb) that phosphorylates DSIF, NELF and Ser2 of the Pol II CTD. This leads to dissociation of NELF and entry of Pol II into productive elongation. Enhancers promote this process by recruiting cofactors that either recruit and stimulate CDK9 or directly affect pause–release, such as bromodomain-containing protein 4 (BRD4) and p300. **d** | Regulation of transcription bursting. Transcription occurs in short 'bursts', which comprise groups of transcription initiation events and are separated by periods of inactivity. The core-promoter sequence determines burst size, that is, the number of transcribing Pol II molecules per burst (left images), whereas enhancers increase bursting frequency (right images). '+' denotes target activation and '–' denotes target inhibition.

In addition, TFIID selectively binds H3K4me3, thereby enabling crosstalk between chromatin and PIC assembly[125]. Apart from regulating accessibility to DNA (reviewed in REF.[9]), TFIID recruitment is also the first step at which transcription can be regulated, and indeed, some transcription factors can bind and potentially recruit TFIID to core promoters[126–128]. In addition, TFIID composition might also influence transcription. Canonical TFIID consists of TBP and TBP-associated factors[54], which can be replaced by different paralogues to form alternative TFIID complexes (reviewed in REFS[129–132]). For example, TBP-related factor 2 (TRF2; also known as TBPL1) substitutes for TBP at the promoters of many housekeeping genes and is essential for their activation[133–135].

As biochemical and structural studies of PIC assembly and function typically consider only a few well-defined or synthetic core promoters that contain canonical core-promoter motifs[55,70], the mechanism of GTF recruitment and regulation at other types of core promoter is unclear and might differ. Indeed, mapping the binding sites of various PIC components genome-wide in yeast revealed a distinct interplay between the PIC and nucleosomes at promoters containing strong TATA-box motifs versus those with only weak or no TATA-box motifs[136]. In yeast, the presence of a strong TATA-box has been used to distinguish between SAGA complex-dominated and TFIID-dominated promoters[137,138]. SAGA-dominated promoters more often contain strong TATA-box motifs and are associated with genes responsive to stress, whereas TFIID-dominated promoters are depleted of such strong TATA-box motifs[137,138]. However, the two complexes might not be mutually exclusively employed at distinct types of promoter; they may regulate different steps that are more or less rate-limiting at the different promoter types[138,139]. This is consistent with recent observations that the transcription of nearly all yeast genes depends to some extent on TFIID[140] and that SAGA is involved in regulating both TATA-box-containing and TATA-less promoters[139].

RNA polymerase II pausing

At many genes, once Pol II has cleared from the TSS, it transcribes only 30–50 nucleotides downstream of the TSS and then undergoes promoter-proximal pausing[141–143] (FIG. 2c). Paused Pol II was initially detected at heat-shock-responsive genes in their inactive state[144] and shown to be rapidly released into productive elongation upon heat shock[145], thereby enabling strong and rapid gene activation. Release from promoter-proximal pausing involves phosphorylation of several components of the paused transcription elongation complex, including negative elongation factor (NELF), DRB sensitivity-inducing factor (DSIF) and Pol II itself, by cyclin-dependent kinase 9 (CDK9), which is a subunit of positive transcription elongation factor b (P-TEFb)[145,146] (FIG. 2c).

The prevalence and tight regulation of Pol II promoter-proximal pausing demonstrate that PIC recruitment and transcription initiation are not necessarily the rate-limiting steps of transcription at all promoters. Rather, promoter-proximal pausing provides an additional opportunity to regulate transcription by allowing rapid release of already engaged Pol II into productive elongation[146], thereby eliminating dependencies on the slower steps of recruitment and initiation. This might be beneficial when rapid or synchronous changes in gene expression are required. For example, in early fly embryos, promoters with paused Pol II are activated synchronously across all cells[147], which is important for coordinating tissue morphogenesis[148]. Similarly, genes with paused Pol II in fly embryos were enriched for developmental regulators, and it is likely that pausing facilitates rapid changes in spatial and temporal activity of these genes during development[141]. By contrast, in mouse embryonic stem cells, paused Pol II is enriched at genes regulating cell cycle and signal transduction and is suggested to regulate development through the control of signalling pathways[149].

Different genes might, however, differ in their rate-limiting step for productive transcription. Some genes could predominantly be regulated by releasing stably paused Pol II, whereas for other genes, regulation might occur mainly at the initiation step. In addition, the stability of paused Pol II at different promoters greatly differs: half-lives of paused Pol II measured by inhibiting both pause–release and de novo initiation range from several minutes to an hour or more[150–152]. At promoters that support stable Pol II pausing with low turnover rates (half-life > 30 min), stalled Pol II seems to block new transcription initiation[151,153], presumably by steric hindrance, as previously predicted[154]. By contrast, at promoters with high turnover of paused Pol II (half-life of only minutes), there may be no interference with transcription initiation[152], potentially allowing tight regulation at the initiation step followed by non-limiting pause–release. Such an antagonistic relationship between pausing duration and transcription initiation frequency might create a pause–initiation balance[153], which could allow influencing one step by regulating another step, for example, increasing initiation frequency by stimulating CDK9-mediated release of paused Pol II (REFS[153,154]).

The nature of the trigger of Pol II pausing is not known, and it was suggested that the sequence downstream of the TSS plays an important role. Core promoters of the most strongly paused genes often have elevated GC content downstream of the TSS, including the GC-rich DPE or Pause button (PB) motifs[155] (TABLE 1). While these motifs might recruit specific proteins, GC-rich sequences might also simply slow down Pol II (REF.[156]). Similarly, transcription might also be hindered by the topological stress due to supercoiling of DNA downstream of the transcribing Pol II (REFS[157,158]). In addition, chromatin has been implicated in Pol II pausing, as the +1 nucleosome could represent a barrier to Pol II at essentially all genes, resulting in downstream or distal pausing[94]. However, the causal relationship between nucleosome positioning and Pol II transcription is not clear, and it was also suggested that paused Pol II is required to keep the promoter region clear of nucleosomes[96], rather than the other way around.

Interestingly, most or all genes seem to require CDK9 for productive elongation, including those without GC-rich sequences downstream of the TSSs and those for which no accumulation of paused Pol II is detected[146,150,153]. The global downregulation of transcription upon CDK9 inhibition, even at enhancers[159,160], indicates that Pol II pausing or a pausing-like checkpoint

SAGA complex
SPT–ADA–GCN5-acetyltransferase (SAGA) is a co-activator complex with different chromatin-modifying modules, including, for example, histone acetyltransferase GCN5.

Promoter-proximal pausing
Pausing of RNA polymerase II downstream of the transcription start site; this process controls the transition into productive transcription elongation.

between initiation and elongation occurs for essentially all Pol II-mediated transcription irrespective of whether paused Pol II accumulates to detectable levels. Such a checkpoint might be important to ensure RNA 5′ capping, the assembly of a functional elongation complex, including DNA topoisomerase I recruitment and activation[161], and the recruitment of other proteins required for elongation and co-transcriptional processes. This suggests that promoter-proximal pausing is an inherent property of transcription by Pol II and is triggered independently of the core-promoter sequence, potentially through the 5′ end of the nascent RNA, which after transcription of about 18 nucleotides starts protruding from Pol II. Indeed, the two pause-establishing factors DSIF and NELF require a nascent transcript longer than 18 nucleotides to stably associate with the Pol II elongation complex[162] (reviewed in REF.[163]). Furthermore, recent biochemical and structural studies of a complex containing Pol II and DSIF revealed that DSIF contacts nascent RNA exiting from Pol II, suggesting a role of this interaction in establishing Pol II pausing[164–166]. According to this model, pausing is triggered independently of the sequence and chromatin properties at the pause site, which nevertheless might influence the stability of the interactions between DNA, nascent RNA and paused Pol II. Strengthening these interactions could increase the duration of pausing and potentially explain the elevated GC content at sites that accumulate high levels of paused Pol II, that is, the formation of stable RNA–DNA hybrids owing to higher GC content at the pause site might increase the duration of pausing.

Regulation by enhancers and cofactors

Active promoters are often in spatial proximity to enhancers[167–170], and the establishment of such contacts between promoters and distal enhancers is related to the three-dimensional organization of chromatin in the nucleus[9,171–173]. Promoter activation might occur upon establishing contacts with an enhancer or by the recruitment of transcription factors to pre-formed enhancer–core-promoter interactions. The latter was found to be prevalent in fly development, where enhancer–core-promoter interactions are established before gene activation and appear stable during development[174]. In either case, promoters need to be sufficiently close to their enhancers to be activated.

Modes of core-promoter activation

The different steps required for productive transcription by Pol II provide opportunity for gene regulation: PIC assembly, Pol II activation and transcription initiation, and Pol II pausing and release into productive elongation (see above and REF.[175]). Core promoters receive regulatory input from enhancers, and this is mediated by transcription factors that directly bind short transcription-factor binding sites within enhancers and by transcriptional cofactors, which are recruited by transcription factors through protein–protein interactions. Cofactors often have enzymatic activities and can post-translationally modify components of the transcription machinery and the surrounding nucleosomes, thereby affecting the different processes taking place at target core promoters.

Promoting pre-initiation complex assembly and RNA polymerase II activation. The most straightforward way to increase transcription from a core promoter is to increase the rate of transcription initiation by promoting PIC assembly and Pol II recruitment and activation. Several transcription factors or cofactors recruited by enhancers directly interact with components of the transcription machinery, leading to stabilization of the PIC at core promoters and increased initiation (FIG. 2a). For example, the Mediator complex is recruited to enhancers, interacts with the PIC at core promoters and transduces activating cues to increase Pol II recruitment and PIC assembly[176]. In yeast, Mediator seems to directly contact TFIIH and stimulate phosphorylation of the Ser5 residues in the carboxy-terminal domain (CTD) of Pol II by the TFIIH subunit CDK7 (REF.[177]; see Supplementary Box 1). Ser5 phosphorylation is considered important for Pol II to escape from the core promoter-bound GTFs and to initiate transcription (FIG. 2b). Similarly, p300, which is a cofactor widely associated with many active enhancers[178], can acetylate GTFs or Pol II at target core promoters[112,179], and this is required for the induction of growth-factor response genes[114].

Promoting RNA polymerase II pause–release. Many core promoters support the recruitment of high levels of Pol II and are instead regulated at the level of pause–release[142,145,180]. Transition into productive elongation is coupled to phosphorylation of the Pol II CTD at Ser2 residues (Supplementary Box 1) and of DSIF and NELF by CDK9, which is the kinase subunit of P-TEFb (FIG. 2c). P-TEFb can be recruited to core promoters by the transcriptional cofactor BRD4 (REFS[181,182]), which is bound to many enhancers and is involved in regulating a specific subset of genes[183,184]. Thus, enhancers that recruit high levels of BRD4, such as those involved in oncogene activation[185,186], may preferentially function by releasing paused Pol II through CDK9. However, BRD proteins also regulate the transition to productive transcription elongation independently of CDK9 recruitment, as BRD protein degradation globally impairs transcription elongation but does not impact CDK9 recruitment to target genes[187,188]. p300 and Pol II-associated factor 1 homologue (PAF1) have also been reported to be involved in pause release[179,189]. PAF1 seems to be required for pausing at enhancers and promoters, and the loss of PAF1 leads to increased promoter activity, potentially through enhancer activation[160].

Modulating transcription bursts. Transcription occurs in short but intense 'bursts', which comprise groups of initiation events separated by periods of inactivity[190,191], as if promoters stochastically transition between inactive and active or permissive states[192,193]. This stochastic nature of transcription means that transcription activation could be achieved in one of two ways: by increasing the amplitude (size) of bursts, that is, the number of transcribing Pol II molecules per burst, or by increasing the frequency of bursts. The latter was shown to be the case in both regulation of developmental genes in fly embryos[193] and activation of the β-globin promoter by its locus-control region[194]. By contrast, burst size is a fixed property of the core-promoter sequence, which mediates

GTF binding[192,195,196] (FIG. 2d). Indeed, the presence of the TATA-box motif supports a larger burst size in yeast[195], which might enable rapid transcriptional responses to stress[196] but appears to disproportionately contribute to transcriptional noise and increased cell-to-cell transcript variability[197]. Activation of core promoters that support large burst size by an enhancer that increases the frequency of bursting will lead to high transcriptional output. This might explain the observation made in reporter assays that enhancers most highly activate TATA-box-containing core promoters[198].

Specificity and responsiveness

Although forced interaction of an enhancer with a core promoter can be sufficient to activate transcription[199], this is not the case for all promoters, suggesting that enhancers have preferences or specificities towards some promoters and, vice versa, that promoters can be activated by certain enhancers but not by others.

Sequence-encoded enhancer–core-promoter specificity. An example of sequence-encoded specificity was provided when reporter genes with TATA-box-containing promoters or with DPE-containing promoters that were integrated at identical genomic positions were differentially expressed in fly embryos[200], suggesting that they differentially responded to genomic enhancers. Similarly, core promoters derived from fly housekeeping genes or from developmental genes were differentially activated by distinct sets of enhancers in an otherwise constant plasmid environment[201]. This is indicative of a sequence-encoded enhancer–core-promoter specificity that separates developmental and housekeeping transcription programmes[201], a notion that was corroborated by a complementary approach that showed that different promoters respond specifically either to developmental enhancers or to housekeeping enhancers[198].

The specificity of core promoters towards regulatory input is not necessarily confined to different sets of genes. For example, in zebrafish, a global switch in initiation pattern from focused to dispersed occurs at many genes during embryonic development[21], suggesting that they use two different, overlapping core-promoter sequences that respond differentially to enhancers active during either maternal or zygotic transcription.

Enhancer-binding regulatory proteins mediate core-promoter specificities. Activation of core promoters by enhancers is mediated by transcription factors and cofactors that have a central role in conveying regulatory cues from enhancers to core promoters and presumably mediate enhancer specificities. Some transcription factors and cofactors can activate transcription on their own when tethered to core promoters[202–206]. Furthermore, when tested with different core promoters in a constant reporter setup, some factors displayed preferences towards certain core promoters[206,207]. An intriguing hypothesis that could explain such observations is that different types of core promoter support the assembly of structurally or compositionally distinct PIC complexes that are biochemically compatible with different types of transcription factors and cofactors. One such example is TRF2 replacing TBP in PICs assembled at housekeeping gene promoters[133–135] (FIG. 3; reviewed in REFS[9,208]).

The suggested specificity between core promoters and activating factors was further corroborated by loss-of-function studies that either specifically inhibited cofactor function[179,183,184] or depleted cofactors[139,140,209] and showed preferential downregulation of certain genes but not others. For example, in yeast, the depletion of different Mediator subunits leads to differential gene downregulation and seems to preferentially affect SAGA-regulated genes[209]. In mammals, inhibition of BRD4 leads to preferential downregulation of MYC[183,185] — a property that is exploited for therapeutic purposes. Similarly, inhibition of p300 seems to most strongly affect core promoters of highly paused genes characterized by a distinct chromatin configuration and binding of specific factors; this inhibition appears to differentially affect Pol II recruitment and initiation versus Pol II pause–release, depending on the core-promoter type[179]. These observations suggest that transcription of different genes depends on different cofactors.

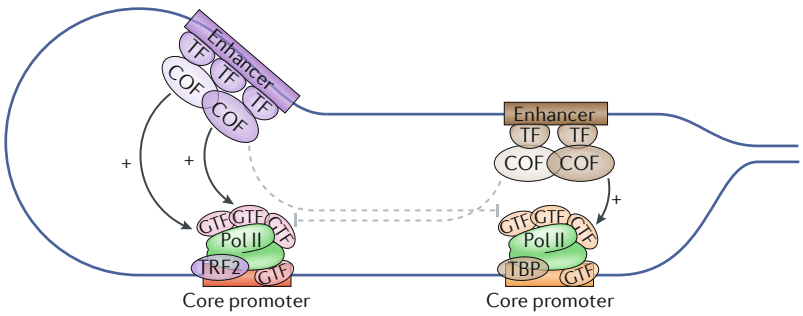


Fig. 3 | **Sequence-encoded specificity of core promoters towards enhancers and activation by specific transcription factors and cofactors.** Different types of core promoter respond differentially to distal enhancers, that is, an enhancer can activate them (solid arrows) or not activate them (dashed arrows). This selectivity or specificity is mediated by different transcription factors (TF) and cofactors (COF), which display core-promoter preferences that are likely based on biochemical compatibilities between the cofactors and core-promoter-bound general transcription factors (GTF). For example, TBP-related factor 2 (TRF2) substitutes for TATA-box-binding protein (TBP) at promoters of many housekeeping genes and is essential for their activation. Mapping and understanding preferences and compatibilities between cofactors and core promoters is an important goal for future research. Pol II, RNA polymerase II.

A functional model of transcription

The properties of core promoters establish them as specialized sequences that support transcription initiation and Pol II pause–release in response to activating cues from distal enhancers. Enhancers have been regarded as amplifiers of transcription from proximal or distal core promoters[7,8], a function mediated by transcription factors and cofactors[8]. The term 'promoters' refers to sequences at gene starts that can autonomously drive high levels of productive transcription. Promoters comprise in close proximity core promoters and supporting activating sequences, which are called proximal promoters or proximal enhancers (discussed in REFS[198,210]). Enhancers therefore share several characteristics with promoters, such as the binding of transcription factors and cofactors[26] but also — more unexpectedly — the binding of GTFs and Pol II (REFS[28,211,212]) and the ability to initiate transcription[15,26,29,42,213] (FIG. 1b).

To understand the similarities and differences between core promoters, enhancers and promoters, it is instructive to establish activity-based definitions of these elements using dedicated assays designed to specifically probe the defining function of each of these elements (BOX 2). Such assays specifically assess enhancer activity as the ability to activate transcription at a distal core promoter[214,215]; core-promoter function as the ability to initiate transcription in response to distal regulatory cues[198]; and promoter activity as the ability to autonomously drive transcription[198,216]. One recently developed assay simultaneously measures both enhancer and promoter activity[69].

Such dedicated functional assays demonstrated, for example, that promoter regions can activate transcription from distal core promoters, meaning that they can function as enhancers[198,201,214,217,218], and that enhancer regions can autonomously give rise to productive transcription and function as promoters[69,216,217]. However, these approaches also found that enhancer function and core-promoter function frequently do not co-occur[69,198,201,214,216–218], indicating that the two functions can be carried out by the same genomic region but are not strictly coupled or interdependent[219].

Fortuitous initiation at enhancers

Enhancer activity is mediated through the binding of transcription factors and the recruitment of cofactors, which not only mediate activation of target core promoters but also create high transcription activation potential at the enhancers themselves. Enhancers

Box 2 | **Measuring core promoter and enhancer activities**

Dedicated activity-based assays that specifically measure enhancer, core-promoter or promoter activities allow function-based definition of regulatory elements (see the figure).

Enhancers activate transcription at distal core promoters

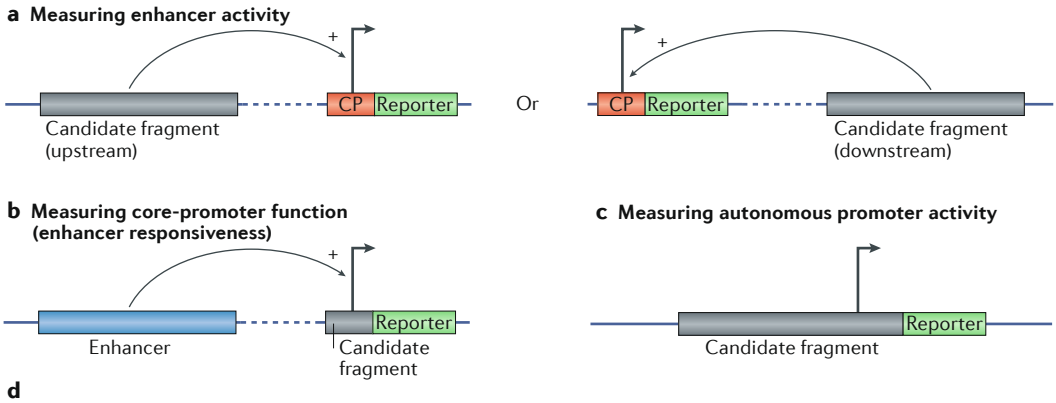
Enhancer activity is measured in reporter assays that test the ability to activate transcription at a distal core promoter (CP) and drive the expression of a reporter gene (see the figure, part **a**). Enhancer activity has been reported for intergenic and intronic sequences but also for some sequences that overlap gene promoters[201,214,215,217,218,237,238]. Such promoters support both core-promoter activity and enhancer activity through the respective sequence elements (see the figure, part **d**). The many promoter regions that do not show enhancer activity likely support only core-promoter functionality and therefore cannot activate transcription at a distal core promoter.

Core promoters initiate transcription in response to regulatory input

Analogously to directly measuring distal enhancer activity, core-promoter activity can be specifically assessed in dedicated reporter assays that measure the ability to initiate transcription in response to activating input from an enhancer, that is, assays that measure enhancer responsiveness (see the figure, part **b**). Candidates with high enhancer responsiveness mainly coincide with gene transcription start sites (TSSs) and contain core-promoter motifs such as TATA-boxes and initiator (Inr) motifs[198,239,240]. Unlike gene core promoters, TSSs within enhancers show very low or no responsiveness[198], suggesting that enhancers in general have a very weak or no sequence-based propensity to respond to distal activating cues and act as core promoters (see the figure, part **d**).

Autonomous promoter function is conferred by sequences that support both core-promoter activities and enhancer activities

Although the above methods assess core-promoter activity as the responsiveness to a defined regulatory input, promoter activity is typically defined as the ability to drive transcription autonomously[69,216] (see the figure, part **c**). Such autonomously functioning promoters typically contain both core-promoter activity and enhancer activity; an enhancer in this context is also called a proximal promoter or an upstream-activating sequence (see the figure, part **d**).



d

Candidate fragment	Enhancer activity (a)	Core-promoter function (b)	Autonomous promoter activity (c)
Enhancer	✓	✗	✗ or ✓
Core promoter	✗	✓	✗ or ✓
Proximal promoter CP Promoter	✓	✓	✓

should therefore naturally have the tendency to activate transcription close to or within the enhancer, presumably at sites that most closely resemble bona fide core promoters. Given the low sequence stringency ('information content') of many core-promoter motifs (TABLE 1), many sequences at either side of an enhancer resemble degenerate core-promoter motifs. Transcription initiation is in fact expected at any (random) sequence that is in the vicinity of strongly activating factors because achieving perfect activation specificity towards core promoters or entirely preventing background initiation at accessible DNA would be energetically costly and could evolve only under strong selective pressure.

Fortuitous transcription initiation resulting from high activator concentrations can explain several observations related to transcription initiation at enhancers, including the presence of degenerate Inr and TATA-box motifs at TSSs within enhancers[15,26,38], the bidirectional initiation pattern at enhancers[15,32] or at open chromatin in general[220] (FIG. 4a) and the observations that eRNAs are inducible[25] and cell-type specific[26] — in both cases, eRNA transcription follows the activity of the enhancer,

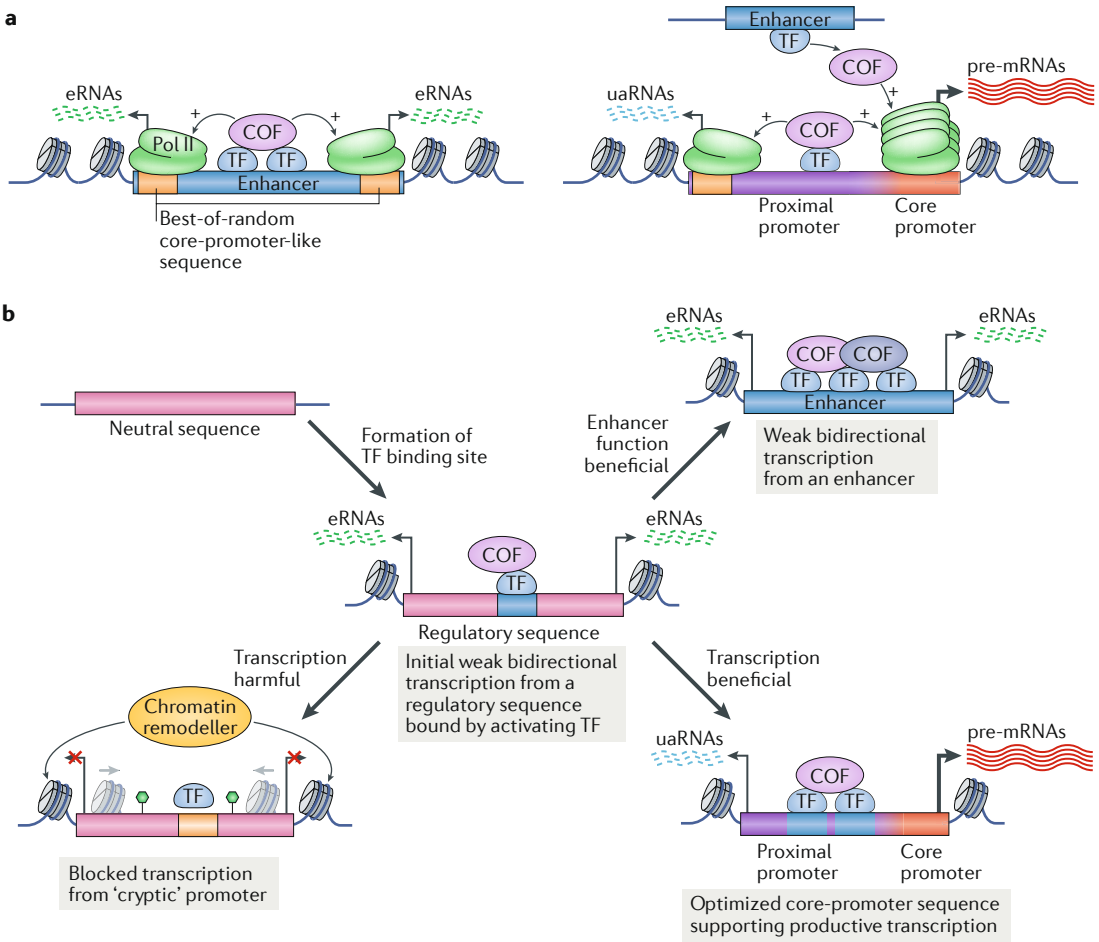


Fig. 4 | **Functional model of transcription initiation at genomic promoters and enhancers. a** | Model of transcription initiation at enhancers (left image) and promoters (right image) based on their distinct sequence-encoded activities. Enhancers bind transcription factors (TF) and recruit cofactors (COF), thereby creating a high local concentration of transcription activators. This should lead to fortuitous transcription initiation at proximal sites that resemble bona fide core promoters ('best-of-random sites'), resulting in divergent transcription of short unstable enhancer RNAs (eRNAs). Promoters transcribe stable mRNAs from a dedicated gene core promoter and — owing to a high concentration of transcription activators — also support fortuitous transcription initiation in the antisense direction. **b** | Model of evolution of functional core promoters and enhancers. Newly emerging TF binding sites (blue) support enhancer-like activity and exhibit low levels of bidirectional transcription at best-of-random sites. If such transcription is harmful, it might be actively suppressed by DNA methylation (pins)[222], repressive factors[223] or repressive chromatin, and the TF binding sites will degenerate over time (yellow). If, by contrast, the transcription in one or both directions is beneficial, the respective transcription start site will be positively selected for, evolving into a fully functional core promoter (red) with strong core-promoter motifs that is able to support high levels of regulated and productive transcription. Transcription in the non-beneficial direction will remain low and yield non-stable, upstream antisense RNAs (uaRNAs). The transcription-activator binding sites near core promoters are often referred to as 'proximal promoters'. Finally, if the transcription from a putative regulatory sequence is neutral and its enhancer activity is beneficial, the enhancer function of the regulatory sequence would be strengthened, and the resulting enhancer will bidirectionally transcribe low levels of eRNAs from best-of-random sites. Pol II, RNA polymerase II.

Enhancer responsiveness
The extent to which transcription from a core promoter is induced by a distal enhancer.

that is, the recruitment of strong transcription activators. It is also consistent with TSSs within enhancers generally showing very low enhancer responsiveness and thus having little or no capacity to support distally regulated transcription initiation as bona fide core promoters do[198] (BOX 2). Moreover, the more similar the TSSs within enhancers are to bona fide core promoters, the higher the level of productive transcription from the enhancer[69]. Therefore, although some enhancers can function as promoters, enhancers generally do not do so, and the difference stems from the presence or absence of sequence-encoded core-promoter functionality.

Evolution of enhancers and promoters

The model of fortuitous initiation at enhancers is consistent with the finding that bidirectional transcription is the ground state of evolutionarily new promoter regions and that unidirectionality is an acquired trait of gene core promoters[221]. Newly emerged transcription-factor binding sites confer enhancer-like activity, which initially leads to low levels of bidirectional transcription initiation[221] (FIG. 4b). If this transcription is harmful, it might become silenced, for example, by repressive chromatin[222,223], and the binding site might eventually decay. If, by contrast, transcription in one or both directions is beneficial, the respective TSS sequence could be positively selected and evolve into a fully functional core promoter with strong core-promoter motifs able to support regulated and productive transcription. Similarly, a core promoter that is regulated exclusively by distal enhancers could acquire proximal activator binding sites and thus promoter activity.

The functions of enhancer RNAs

In the model of fortuitous initiation at enhancers, eRNAs are unavoidable by-products of transcription activators, yet this does not exclude the possibility that eRNA transcription or eRNAs themselves are functional. It is possible that evolution took advantage of their correlation with transcriptional activity to modulate enhancer activity (reviewed in REF.[224]). For example, eRNA transcription might ensure accessibility to DNA[95,96], and eRNAs might be involved in the formation of activating microenvironments in the form of non-membrane bound compartments with high concentrations of transcription activators[225–227], which is similar to what has been reported for germline P granules[228], RNA granules[229,230] and the formation of condensed heterochromatin[231,232]. Such hypotheses that consider conceptually novel ways to understand the regulatory environment at enhancers should motivate future studies of eRNA function.

Perspective and future directions

Pol II core promoters are genomic elements that support PIC assembly and transcription initiation and function as specialized sequences that have evolved to enable highly regulated gene transcription. We propose a functional model that defines regulatory elements by their function rather than by their genomic position; we argue that core promoters and enhancers are the two principal gene regulatory elements, that they have distinct functionalities and that they have evolved for distinct purposes: initiating productive transcription locally (core promoters) versus boosting transcription locally or distally (enhancers).

We are intrigued by the widespread occurrence of Pol II pausing at most promoters and enhancers[153,159,160], which might indicate that a pausing-like checkpoint between transcription initiation and elongation is an intrinsic property of all Pol II-mediated transcription. As such, it might be triggered not by the DNA sequence at the downstream pause site but, for example, by the 5′ end of the nascent RNA as it protrudes from Pol II. It will be interesting to see whether the successful resolution of this checkpoint is necessary for productive elongation and whether this could be the main difference between transcription at promoters and enhancers.

We would like to highlight the existence of different types of core promoter with distinct properties, especially preferences towards different enhancers and cofactors that are presumably based on biochemical compatibilities (FIG. 3). Elucidating such preferences and compatibilities and determining the differences between various types of core promoter are crucial at a time when we have an increasingly complete understanding of the mechanisms that determine genome structure and spatial contacts of enhancers and their target core promoters (reviewed in REFS[233,234]) and when transcription regulation is becoming the focus of targeted intervention and novel therapeutic strategies.

Published online: 26 June 2018

1. Spitz, F. & Furlong, E. E. M. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613–626 (2012).
2. Levine, M. & Tjian, R. Transcription regulation and animal diversity. *Nature* **424**, 147–151 (2003).
3. Levine, M., Cattoglio, C. & Tjian, R. Looping back to leap forward: transcription enters a new era. *Cell* **157**, 13–25 (2014).
4. Herz, H.-M., Hu, D. & Shilatifard, A. Enhancer malfunction in cancer. *Mol. Cell* **53**, 859–866 (2014).
5. Hampsey, M. Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol. Mol. Biol. Rev.* **62**, 465–503 (1998).
6. Kadonaga, J. T. Perspectives on the RNA polymerase II core promoter. *Wiley Interdiscip. Rev. Dev. Biol.* **1**, 40–51 (2012).
7. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299–308 (1981).
8. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* **15**, 272–286 (2014).
9. Zabidi, M. A. & Stark, A. Regulatory enhancer–core-promoter communication via transcription factors and cofactors. *Trends Genet.* **32**, 801–814 (2016).
10. Shiraki, T. et al. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc. Natl Acad. Sci. USA* **100**, 15776–15781 (2003).
11. Gu, W. et al. CapSeq and CIP-TAP identify Pol II start sites and reveal capped small RNAs as *C. elegans* piRNA precursors. *Cell* **151**, 1488–1500 (2012).
12. Ni, T. et al. A paired-end sequencing strategy to map the complex landscape of transcription initiation. *Nat. Methods* **7**, 521–527 (2010).
13. Nechaev, S. et al. Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in *Drosophila*. *Science* **327**, 335–338 (2010).
14. Lam, M. T. Y. et al. Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature* **498**, 511–515 (2013).
15. Core, L. J. et al. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat. Genet.* **46**, 1311–1320 (2014).
This work proposes a unified model of transcription initiation at promoters and enhancers and emphasizes that post-initiation transcript stability is the main distinction between the two elements.
16. Kwak, H., Fuda, N. J., Core, L. J. & Lis, J. T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339**, 950–953 (2013).
17. Affymetrix/Cold Spring Harbor Laboratory ENCODE Transcriptome Project. Post-transcriptional processing generates a diversity of 5′-modified long and short RNAs. *Nature* **457**, 1028–1032 (2009).
18. The FANTOM Consortium and RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group). The transcriptional landscape of the mammalian genome. *Science* **309**, 1559–1563 (2005).
19. Hoskins, R. A. et al. Genome-wide analysis of promoter architecture in *Drosophila melanogaster*. *Genome Res.* **21**, 182–192 (2011).

20. Chen, R. A.-J. et al. The landscape of RNA polymerase II transcription initiation in *C. elegans* reveals promoter and enhancer architectures. *Genome Res.* **23**, 1339–1347 (2013).
21. Haberle, V. et al. Two independent transcription initiation codes overlap on vertebrate core promoters. *Nature* **507**, 381–385 (2014).
This study reveals a widespread switch in TSS usage associated with distinct sequence properties during early embryonic development of zebrafish.
22. The FANTOM Consortium, The RIKEN PMI & CLST (DGT). A promoter-level mammalian expression atlas. *Nature* **507**, 462–470 (2014).
23. The ENCODE Project Consortium et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
24. Kapranov, P. et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**, 1484–1488 (2007).
25. Kim, T.-K. et al. Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).
This study reports widespread bidirectional transcription from enhancers giving rise to eRNAs.
26. Andersson, R. et al. An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).
This study uses bidirectional transcription initiation to predict enhancers and their activity across numerous human cell types.
27. De Santa, F. et al. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol.* **8**, e1000384 (2010).
28. Koch, F. et al. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat. Struct. Mol. Biol.* **18**, 956–963 (2011).
29. Arner, E. et al. Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science* **347**, 1010–1014 (2015).
30. Li, W. et al. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* **498**, 516–520 (2013).
31. Schaukowitch, K. et al. Enhancer RNA facilitates NELF release from immediate early genes. *Mol. Cell* **56**, 29–42 (2014).
32. Andersson, R. et al. Nuclear stability and transcriptional directionality separate functionally distinct RNA species. *Nat. Commun.* **5**, 5336 (2014).
33. Seila, A. C. et al. Divergent transcription from active promoters. *Science* **322**, 1849–1851 (2008).
34. Preker, P. et al. RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322**, 1851–1854 (2008).
35. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848 (2008).
References 33–35 report widespread antisense transcription from gene promoters giving rise to short unstable upstream antisense RNAs.
36. Andersson, R. et al. Human gene promoters are intrinsically bidirectional. *Mol. Cell* **60**, 346–347 (2015).
37. Duttke, S. H. C. et al. Human promoters are intrinsically directional. *Mol. Cell* **57**, 674–684 (2015).
38. Scruggs, B. S. et al. Bidirectional transcription arises from two distinct hubs of transcription factor binding and active chromatin. *Mol. Cell* **58**, 1101–1112 (2015).
References 37 and 38 demonstrate that bidirectional transcription from promoters arises from two separate and intrinsically directional transcription complexes.
39. Sigova, A. A. et al. Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proc. Natl Acad. Sci. USA* **110**, 2876–2881 (2013).
40. Lepoivre, C. et al. Divergent transcription is associated with promoters of transcriptional regulators. *BMC Genomics* **14**, 914 (2013).
41. Andersson, R. Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model. *Bioessays* **37**, 314–323 (2015).
42. Kim, T.-K. & Shiekhattar, R. Architectural and functional commonalities between enhancers and promoters. *Cell* **162**, 948–959 (2015).
43. Almada, A. E., Wu, X., Kriz, A. J., Burge, C. B. & Sharp, P. A. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* **499**, 360–363 (2013).
44. Ntini, E. et al. Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nat. Struct. Mol. Biol.* **20**, 923–928 (2013).
45. Carninci, P. et al. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.* **38**, 626–635 (2006).
This work uses genome-wide maps of human and mouse TSSs to describe two classes of promoters that differ in initiation pattern: focused TATA-box-enriched promoters and broad CpG-rich promoters.
46. Lenhard, B., Sandelin, A. & Carninci, P. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat. Rev. Genet.* **13**, 233–245 (2012).
47. Schor, I. E. et al. Promoter shape varies across populations and affects promoter evolution and expression noise. *Nat. Genet.* **49**, 550–558 (2017).
This work identifies natural genetic variants that affect TSS distributions within core promoters in flies.
48. Lifton, R. P., Goldberg, M. L., Karp, R. W. & Hogness, D. S. The organization of the histone genes in *Drosophila melanogaster*: functional and evolutionary implications. *Cold Spring Harb. Symp. Quant. Biol.* **42**, 1047–1051 (1978).
49. Goldberg, M. L. *Sequence Analysis of Drosophila Histone Genes*. Thesis, Stanford Univ. (1979).
50. Ponjavic, J. et al. Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters. *Genome Biol.* **7**, R78 (2006).
51. Ohler, U., Liao, G.-C., Niemann, H. & Rubin, G. M. Computational analysis of core promoters in the *Drosophila* genome. *Genome Biol.* **3**, RESEARCH0087 (2002).
52. FitzGerald, P. C., Sturgill, D., Shyakhtenko, A., Oliver, B. & Vinson, C. Comparative genomics of drosophila and human core promoters. *Genome Biol.* **7**, R53 (2006).
References 51 and 52 provide comprehensive computational analyses of over-represented sequence motifs in fly and human core promoters.
53. Patikoglou, G. A. et al. TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes Dev.* **13**, 3217–3230 (1999).
54. Burley, S. K. & Roeder, R. G. Biochemistry and structural biology of transcription factor IID (TFIID). *Annu. Rev. Biochem.* **65**, 769–799 (1996).
55. Louder, R. K. et al. Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature* **531**, 604–609 (2016).
This work reports a structure of the human promoter-bound TFIID, revealing contacts between TFIID subunits and specific core-promoter motifs.
56. Smale, S. T. & Baltimore, D. The 'initiator' as a transcription control element. *Cell* **57**, 103–113 (1989).
57. Chalkley, G. E. & Verrijzer, C. P. DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250-TAF(II)150 complex recognizes the initiator. *EMBO J* **18**, 4835–4845 (1999).
58. Vo Ngoc, L., Cassidy, C. J., Huang, C. Y., Duttke, S. H. C. & Kadonaga, J. T. The human initiator is a distinct and abundant element that is precisely positioned in focused core promoters. *Genes Dev.* **31**, 6–11 (2017).
59. Burke, T. W. & Kadonaga, J. T. *Drosophila* TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes Dev.* **10**, 711–724 (1996).
60. Burke, T. W. & Kadonaga, J. T. The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAFII60 of *Drosophila*. *Genes Dev.* **11**, 3020–3031 (1997).
61. Kutach, A. K. & Kadonaga, J. T. The downstream promoter element DPE appears to be as widely used as the TATA box in *Drosophila* core promoters. *Mol. Cell. Biol.* **20**, 4754–4764 (2000).
62. Engstrom, P. G., Ho Sui, S. J., Drivenes, O., Becker, T. S. & Lenhard, B. Genomic regulatory blocks underlie extensive microsynteny conservation in insects. *Genome Res.* **17**, 1898–1908 (2007).
63. Lim, C. Y. et al. The MTE, a new core promoter element for transcription by RNA polymerase II. *Genes Dev.* **18**, 1606–1617 (2004).
64. Lagrange, T., Kapanidis, A. N., Tang, H., Reinberg, D. & Ebright, R. H. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev.* **12**, 34–44 (1998).
65. Deng, W. & Roberts, S. G. E. A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes Dev.* **19**, 2418–2423 (2005).
66. Lewis, B. A., Kim, T. K. & Orkin, S. H. A downstream element in the human beta-globin promoter: evidence of extended sequence-specific transcription factor IID contacts. *Proc. Natl Acad. Sci. USA* **97**, 7172–7177 (2000).
67. Lee, D.-H. et al. Functional characterization of core promoter elements: the downstream core element is recognized by TAF1. *Mol. Cell. Biol.* **25**, 9674–9686 (2005).
68. Rach, E. A., Yuan, H.-Y., Majoros, W. H., Tomancak, P. & Ohler, U. Motif composition, conservation and condition-specificity of single and alternative transcription start sites in the drosophila genome. *Genome Biol.* **10**, R73 (2009).
This work shows differential motif enrichment and spatiotemporal utilization of core promoters with focused and dispersed initiation patterns in flies.
69. Mikhaylichenko, O. et al. The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. *Genes Dev.* **32**, 42–57 (2018).
This work introduces a functional assay to simultaneously measure enhancer and promoter activity of candidate fragments and shows that core-promoter motifs confer promoter functionality to enhancers.
70. Juven-Gershon, T., Cheng, S. & Kadonaga, J. T. Rational design of a super core promoter that enhances gene expression. *Nat. Methods* **3**, 917–922 (2006).
71. Pfeiffer, B. D. et al. Tools for neuroanatomy and neurogenetics in *Drosophila*. *Proc. Natl Acad. Sci. USA* **105**, 9715–9720 (2008).
72. Even, D. Y. et al. Engineered promoters for potent transient overexpression. *PLoS ONE* **11**, e0148918 (2016).
73. Gardiner-Garden, M. & Frommer, M. CpG islands in vertebrate genomes. *J. Mol. Biol.* **196**, 261–282 (1987).
74. Saxonov, S., Berg, P. & Brutlag, D. L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl Acad. Sci. USA* **103**, 1412–1417 (2006).
75. Akalin, A. et al. Transcriptional features of genomic regulatory blocks. *Genome Biol.* **10**, R38 (2009).
76. Dreos, R., Ambrosini, G. & Bucher, P. Influence of rotational nucleosome positioning on transcription start site selection in animal promoters. *PLoS Comput. Biol.* **12**, e1005144 (2016).
77. Satchwell, S. C., Drew, H. R. & Travers, A. A. Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.* **191**, 659–675 (1986).
78. Widom, J. Role of DNA sequence in nucleosome stability and dynamics. *Q. Rev. Biophys.* **34**, 269–324 (2001).
79. Segal, E. et al. A genomic code for nucleosome positioning. *Nature* **442**, 772–778 (2006).
80. Yuan, G.-C. et al. Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* **309**, 626–630 (2005).
81. Mavrich, T. N. et al. Nucleosome organization in the *Drosophila* genome. *Nature* **453**, 358–362 (2008).
This work maps nucleosome positions across a metazoan genome and reveals the organization of nucleosomes around active promoters.
82. Jiang, C. & Pugh, B. F. Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.* **10**, 161–172 (2009).
83. Jin, C. et al. H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nat. Genet.* **41**, 941–945 (2009).
84. Fei, J. et al. The prenucleosome, a stable conformational isomer of the nucleosome. *Genes Dev.* **29**, 2563–2575 (2015).
85. Henikoff, J. G., Belsky, J. A., Krassovsky, K., MacAlpine, D. M. & Henikoff, S. Epigenome characterization at single base-pair resolution. *Proc. Natl Acad. Sci. USA* **108**, 18318–18323 (2011).
86. Rhee, H. S., Bataille, A. R., Zhang, L. & Pugh, B. F. Subnucleosomal structures and nucleosome asymmetry across a genome. *Cell* **159**, 1377–1388 (2014).
87. Mueller, B. et al. Widespread changes in nucleosome accessibility without changes in nucleosome occupancy during a rapid transcriptional induction. *Genes Dev.* **31**, 451–462 (2017).
88. Mieczkowski, J. et al. MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. *Nat. Commun.* **7**, 11485 (2016).
89. Rach, E. A. et al. Transcription initiation patterns indicate divergent strategies for gene regulation at the chromatin level. *PLoS Genet.* **7**, e1001274 (2011).

90. Kubik, S. et al. Nucleosome stability distinguishes two different promoter types at all protein-coding genes in yeast. *Mol. Cell* **60**, 422–434 (2015).
91. Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* **25**, 2227–2241 (2011).
92. Shimojima, T. et al. *Drosophila* FACT contributes to Hox gene expression through physical and functional interactions with GAGA factor. *Genes Dev.* **17**, 1605–1616 (2003).
93. Fuda, N. J. et al. GAGA factor maintains nucleosome-free regions and has a role in RNA polymerase II recruitment to promoters. *PLoS Genet.* **11**, e1005108 (2015).
94. Weber, C. M., Ramachandran, S. & Henikoff, S. Nucleosomes are context-specific, H2A. Z-modulated barriers to RNA polymerase. *Mol. Cell* **53**, 819–830 (2014).
95. Mousavi, K. et al. eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. *Mol. Cell* **51**, 606–617 (2013).
96. Gilchrist, D. A. et al. Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. *Cell* **143**, 540–551 (2010).
97. Ahmad, K. & Henikoff, S. The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol. Cell* **9**, 1191–1200 (2002).
98. Pradhan, S. K. et al. EP400 deposits H3.3 into promoters and enhancers during gene activation. *Mol. Cell* **61**, 27–38 (2016).
99. Raisner, R. M. et al. Histone variant H2A. Z marks the 5′ends of both active and inactive genes in euchromatin. *Cell* **123**, 233–248 (2005).
100. Barski, A. et al. High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
This study maps different histone modifications genome-wide and identified those associated with active or repressed promoters.
101. Ng, H.-H., Robert, F., Young, R. A. & Struhl, K. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol. Cell* **11**, 709–719 (2003).
102. Hathaway, N. A. et al. Dynamics and memory of heterochromatin in living cells. *Cell* **149**, 1447–1460 (2012).
103. Zhao, R., Nakamura, T., Fu, Y., Lazar, Z. & Spector, D. L. Gene bookmarking accelerates the kinetics of post-mitotic transcriptional re-activation. *Nat. Cell Biol.* **13**, 1295–1304 (2011).
104. Tropberger, P. et al. Regulation of transcription through acetylation of H3K122 on the lateral surface of the histone octamer. *Cell* **152**, 859–872 (2013).
105. Neumann, H. et al. A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. *Mol. Cell* **36**, 153–163 (2009).
106. Tessarz, P. & Kouzarides, T. Histone core modifications regulating nucleosome structure and dynamics. *Nat. Rev. Mol. Cell Biol.* **15**, 703–708 (2014).
107. Dey, A., Chitsaz, F., Abbasi, A., Misteli, T. & Ozato, K. The double bromodomain protein Brd4 binds to acetylated chromatin during interphase and mitosis. *Proc. Natl Acad. Sci. USA* **100**, 8758–8763 (2003).
108. Hödl, M. & Basler, K. Transcription in the absence of histone H3.2 and H3K4 methylation. *Curr. Biol.* **22**, 2253–2257 (2012).
109. Hödl, M. & Basler, K. Transcription in the absence of histone H3.3. *Curr. Biol.* **19**, 1221–1226 (2009).
110. Pengelly, A. R., Copur, Ö., Jäckle, H., Herzig, A. & Müller, J. A histone mutant reproduces the phenotype caused by loss of histone-modifying factor Polycomb. *Science* **339**, 698–699 (2013).
111. Sterner, D. E. & Berger, S. L. Acetylation of histones and transcription-related factors. *Microbiol. Mol. Biol. Rev.* **64**, 435–459 (2000).
112. Imhof, A. et al. Acetylation of general transcription factors by histone acetyltransferases. *Curr. Biol.* **7**, 689–692 (1997).
113. Roe, J.-S., Mercan, F., Rivera, K., Pappin, D. J. & Vakoc, C. R. BET bromodomain inhibition suppresses the function of hematopoietic transcription factors in acute myeloid leukemia. *Mol. Cell* **58**, 1028–1039 (2015).
114. Schröder, S. et al. Acetylation of RNA polymerase II regulates growth-factor-induced gene transcription in mammalian cells. *Mol. Cell* **52**, 314–324 (2013).
115. Rickels, R. et al. Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. *Nat. Genet.* **156**, 645–1653 (2017).
116. Dorighi, K. M. et al. Mll3 and Mll4 facilitate enhancer RNA synthesis and transcription from promoters independently of H3K4 monomethylation. *Mol. Cell* **66**, 568–576.e4 (2017).
117. Pollex, T. & Furlong, E. E. M. Correlation does not imply causation: histone methyltransferases, but not histone methylation, SET the stage for enhancer activation. *Mol. Cell* **66**, 439–441 (2017).
118. Andersen, P. R., Tirian, L., Vunjak, M. & Brennecke, J. A heterochromatin-dependent transcription machinery drives piRNA expression. *Nature* **549**, 54–59 (2017).
This study shows that histone modifications recruit the transcription machinery to transcribe heterochromatic loci that are a source of small RNAs.
119. Thomas, M. C. & Chiang, C.-M. The general transcription machinery and general cofactors. *Crit. Rev. Biochem. Mol. Biol.* **41**, 105–178 (2006).
120. Orphanides, G., Lagrange, T. & Reinberg, D. The general transcription factors of RNA polymerase II. *Genes Dev.* **10**, 2657–2683 (1996).
121. Sainsbury, S., Bernecky, C. & Cramer, P. Structural basis of transcription initiation by RNA polymerase II. *Nat. Rev. Mol. Cell. Biol.* **16**, 129–143 (2015).
122. Zhang, Z. et al. Rapid dynamics of general transcription factor TFIIB binding during preinitiation complex assembly revealed by single-molecule analysis. *Genes Dev.* **30**, 2106–2118 (2016).
123. He, Y. et al. Near-atomic resolution visualization of human transcription promoter opening. *Nature* **533**, 359–365 (2016).
124. Plaschka, C. et al. Transcription initiation complex structures elucidate DNA opening. *Nature* **533**, 353–358 (2016).
This study reports structures of open and closed yeast PIC complexes and proposes a mechanism of DNA duplex opening.
125. Vermeulen, M. et al. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131**, 58–69 (2007).
126. Papai, G. et al. TFIIA and the transactivator Rap1 cooperate to commit TFIID for transcription initiation. *Nature* **465**, 956–960 (2010).
127. Liu, W.-L. et al. Structures of three distinct activator-TFIID complexes. *Genes Dev.* **23**, 1510–1521 (2009).
128. Chopra, V. S. et al. Transcriptional activation by GAGA factor is through its direct interaction with dmTAF3. *Dev. Biol.* **317**, 660–670 (2008).
129. Hochheimer, A. & Tjian, R. Diversified transcription initiation complexes expand promoter selectivity and tissue-specific gene expression. *Genes Dev.* **17**, 1309–1320 (2003).
130. Taatjes, D. J., Marr, M. T. & Tjian, R. Regulatory diversity among metazoan co-activator complexes. *Nat. Rev. Mol. Cell. Biol.* **5**, 403–410 (2004).
131. Goodrich, J. A. & Tjian, R. Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. *Nat. Rev. Genet.* **11**, 549–558 (2010).
132. Jones, K. A. Changing the core of transcription. *eLife* **3**, e03575 (2014).
133. Hochheimer, A., Zhou, S., Zheng, S., Holmes, M. C. & Tjian, R. TRF2 associates with DREF and directs promoter-selective gene expression in *Drosophila*. *Nature* **420**, 439–445 (2002).
134. Wang, Y.-L. et al. TRF2, but not TBP, mediates the transcription of ribosomal protein genes. *Genes Dev.* **28**, 1550–1555 (2014).
References 133 and 134 reveal that TRF2 replaces TBP within the PIC to drive transcription of a specific subset of genes.
135. Isogai, Y., Keles, S., Prestel, M., Hochheimer, A. & Tjian, R. Transcription of histone gene cluster by differential core-promoter factors. *Genes Dev.* **21**, 2936–2949 (2007).
136. Rhee, H. S. & Pugh, B. F. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483**, 295–301 (2012).
This study maps positions of PIC components at high resolution across the entire yeast genome.
137. Basehoar, A. D., Zanton, S. J. & Pugh, B. F. Identification and distinct regulation of yeast TATA box-containing genes. *Cell* **116**, 699–709 (2004).
138. Struhl, K. Constitutive and inducible *Saccharomyces cerevisiae* promoters: evidence for two distinct molecular mechanisms. *Mol. Cell. Biol.* **6**, 3847–3853 (1986).
139. Baptista, T. et al. SAGA is a general cofactor for RNA polymerase II transcription. *Mol. Cell* **68**, 1–20 (2017).
140. Warfield, L. et al. Transcription of nearly all yeast RNA polymerase II- transcribed genes is dependent on transcription factor TFIID. *Mol. Cell* **68**, 1–18 (2017).
141. Zeitlinger, J. et al. RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nat. Genet.* **39**, 1512–1516 (2007).
References 141 and 142 report widespread Pol II pausing at developmentally regulated genes in flies.
142. Muse, G. W. et al. RNA polymerase is poised for activation across the genome. *Nat. Genet.* **39**, 1507–1511 (2007).
References 141 and 142 report widespread Pol II pausing at developmentally regulated genes in flies.
143. Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77–88 (2007).
144. Rougvie, A. E. & Lis, J. T. The RNA polymerase II molecule at the 5′ end of the uninduced hsp70 gene of *D. melanogaster* is transcriptionally engaged. *Cell* **54**, 795–804 (1988).
This study shows that Pol II pauses downstream of the TSS.
145. Lis, J. T., Mason, P., Peng, J., Price, D. H. & Werner, J. P-TEFb kinase recruitment and function at heat shock loci. *Genes Dev.* **14**, 792–803 (2000).
146. Henriques, T. et al. Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals. *Mol. Cell* **52**, 517–528 (2013).
147. Boettiger, A. N. & Levine, M. Synchronous and stochastic patterns of gene activation in the *Drosophila* embryo. *Science* **325**, 471–473 (2009).
148. Lagha, M. et al. Paused Pol II coordinates tissue morphogenesis in the *Drosophila* embryo. *Cell* **153**, 976–987 (2013).
149. Williams, L. H. et al. Pausing of RNA polymerase II regulates mammalian developmental potential through control of signaling networks. *Mol. Cell* **58**, 311–322 (2015).
150. Jonkers, I., Kwak, H. & Lis, J. T. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *eLife* **3**, e02407 (2014).
151. Shao, W. & Zeitlinger, J. Paused RNA polymerase II inhibits new transcriptional initiation. *Nat. Genet.* **16**, 129–1051 (2017).
152. Krebs, A. R. et al. Genome-wide single-molecule footprinting reveals high RNA polymerase II turnover at paused promoters. *Mol. Cell* **67**, 411–422.e4 (2017).
153. Gressel, S. et al. CDK9-dependent RNA polymerase II pausing controls transcription initiation. *eLife* **6**, R106 (2017).
References 151–153 report a wide-range of paused Pol II half-lives at promoters genome-wide.
154. Ehrensberger, A. H., Kelly, G. P. & Svejstrup, J. Q. Mechanistic interpretation of promoter-proximal peaks and RNAPII density maps. *Cell* **154**, 713–715 (2013).
155. Hendrix, D. A., Hong, J.-W., Zeitlinger, J., Rokhsar, D. S. & Levine, M. S. Promoter elements associated with RNA Pol II stalling in the *Drosophila* embryo. *Proc. Natl Acad. Sci. USA* **105**, 7762–7767 (2008).
156. Veloso, A. et al. Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Res.* **24**, 896–905 (2014).
157. Gartenberg, M. R. & Wang, J. C. Positive supercoiling of DNA greatly diminishes mRNA synthesis in yeast. *Proc. Natl Acad. Sci. USA* **89**, 11461–11465 (1992).
158. Joshi, R. S., Piña, B. & Roca, J. Positional dependence of transcriptional inhibition by DNA torsional stress in yeast chromosomes. *EMBO J.* **29**, 740–748 (2010).
159. Henriques, T. et al. Widespread transcriptional pausing and elongation control at enhancers. *Genes Dev.* **32**, 26–41 (2018).
160. Chen, F. X. et al. PAF1 regulation of promoter-proximal pause release via enhancer activation. *Science* **357**, 1294–1298 (2017).
References 159 and 160 report widespread pausing of Pol II at enhancers.
161. Baranello, L. et al. RNA polymerase II regulates topoisomerase 1 activity to favor efficient transcription. *Cell* **165**, 357–371 (2016).
162. Missra, A. & Gilmour, D. S. Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the *Drosophila* RNA polymerase II transcription elongation complex. *Proc. Natl Acad. Sci. USA* **107**, 11301–11306 (2010).
163. Yamaguchi, Y., Shibata, H. & Handa, H. Transcription elongation factors DSIF and NELF: promoter-proximal pausing and beyond. *Biochim. Biophys. Acta* **1829**, 98–104 (2013).

164. Bernecky, C., Plitzko, J. M. & Cramer, P. Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. *Nat. Struct. Mol. Biol.* **24**, 809–815 (2017).
165. Qiu, Y. & Gilmour, D. S. Identification of regions in the Spt5 subunit of DRB sensitivity-inducing factor (DSIF) that are involved in promoter-proximal pausing. *J. Biol. Chem.* **292**, 5555–5570 (2017).
References 164 and 165 provide structural and biochemical evidence that DSIF contacts nascent RNA protruding from Pol II, ascribing it a role in triggering Pol II pausing.
166. Ehara, H. et al. Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science* **357**, 921–924 (2017).
167. Li, G. et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84–98 (2012).
168. Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
169. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489**, 109–113 (2012).
170. Beagrie, R. A. et al. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519–524 (2017).
171. Symmons, O. et al. Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* **24**, 390–400 (2014).
172. Merkenschlager, M. & Nora, E. P. CTCF and cohesin in genome folding and transcriptional gene regulation. *Annu. Rev. Genomics Hum. Genet.* **17**, 17–43 (2016).
173. Spitz, F. Gene regulation at a distance: from remote enhancers to 3D regulatory ensembles. *Semin. Cell Dev. Biol.* **57**, 57–67 (2016).
174. Ghavi-Helm, Y. et al. Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**, 96–100 (2014).
175. Michel, M. & Cramer, P. Transitions for regulating early transcription. *Cell* **153**, 943–944 (2013).
176. Eychenne, T. et al. Functional interplay between mediator and TFIIB in preinitiation complex assembly in relation to promoter architecture. *Genes Dev.* **30**, 2119–2132 (2016).
177. Esnault, C. et al. Mediator-dependent recruitment of TFIIH modules in preinitiation complex. *Mol. Cell* **31**, 337–346 (2008).
178. Visel, A. et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009).
179. Boija, A. et al. CBP regulates recruitment and release of promoter-proximal RNA polymerase II. *Mol. Cell* **68**, 491–503.e5 (2017).
180. Sawado, T., Halow, J., Bender, M. A. & Groudine, M. The beta -globin locus control region (LCR) functions primarily by enhancing the transition from transcription initiation to elongation. *Genes Dev.* **17**, 1009–1018 (2003).
181. Yang, Z. et al. Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Mol. Cell* **19**, 535–545 (2005).
182. Jang, M. K. et al. The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Mol. Cell* **19**, 523–534 (2005).
183. Zuber, J. et al. RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature* **478**, 524–528 (2011).
184. Delmore, J. E. et al. BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell* **146**, 904–917 (2011).
References 183 and 184 identify BRD4 as a gene-specific regulator whose depletion affects only a subset of genes.
185. Lovén, J. et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013).
186. Rathert, P. et al. Transcriptional plasticity promotes primary and acquired resistance to BET inhibition. *Nature* **525**, 543–547 (2015).
187. Winter, G. E. et al. BET bromodomain proteins function as master transcription elongation factors independent of CDK9 recruitment. *Mol. Cell* **67**, 5–18 (2017).
188. Muhar, M. et al. SLAM-seq defines direct gene-regulatory functions of the BRD4-MYC axis. *Science* https://doi.org/10.1126/science.aao2793 (2018).
References 187 and 188 demonstrate that BRD proteins, in particular BRD4, are globally required for transition into productive elongation in a manner that is independent of CDK9 recruitment.
189. Chen, F. X. et al. PAF1, a molecular regulator of promoter-proximal pausing by RNA polymerase II. *Cell* **162**, 1003–1015 (2015).
190. Chubb, J. R., Trcek, T., Shenoy, S. M. & Singer, R. H. Transcriptional pulsing of a developmental gene. *Curr. Biol.* **16**, 1018–1025 (2006).
191. Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y. & Tyagi, S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* **4**, e309 (2006).
192. Tantale, K. et al. A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting. *Nat. Commun.* **7**, 12248 (2016).
193. Fukaya, T., Lim, B. & Levine, M. Enhancer control of transcriptional bursting. *Cell* **166**, 358–368 (2016).
This study provides evidence that enhancers regulate the frequency of transcription bursts synchronously from multiple promoters in their vicinity.
194. Bartman, C. R., Hsu, S. C., Hsiung, C. C.-S., Raj, A. & Blobel, G. A. Enhancer regulation of transcriptional bursting parameters revealed by forced chromatin looping. *Mol. Cell* **62**, 237–247 (2016).
195. Hornung, G. et al. Noise-mean relationship in mutated promoters. *Genome Res.* **22**, 2409–2417 (2012).
196. Blake, W. J. et al. Phenotypic consequences of promoter-mediated transcriptional noise. *Mol. Cell* **24**, 853–865 (2006).
197. Tirosh, I., Weinberger, A., Carmi, M. & Barkai, N. A genetic signature of interspecies variations in gene expression. *Nat. Genet.* **38**, 830–834 (2006).
198. Arnold, C. D. et al. Genome-wide assessment of sequence-intrinsic enhancer responsiveness at single-base-pair resolution. *Nat. Biotechnol.* **35**, 136–144 (2016).
This study measures enhancer responsiveness for all core promoters across the fly genome and demonstrates that core promoters show differential responses to different enhancers.
199. Deng, W. et al. Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* **149**, 1233–1244 (2012).
200. Butler, J. E. & Kadonaga, J. T. Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. *Genes Dev.* **15**, 2515–2519 (2001).
This work demonstrates that TATA-box-containing and DPE-containing core promoters can be differentially activated when integrated into the same genomic locus.
201. Zabidi, M. A. et al. Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* **518**, 556–559 (2015).
This work provides evidence for sequence-encoded enhancer–core-promoter specificity that distinguishes between housekeeping and developmental transcription programmes in the fly genome.
202. Ptashne, M. & Gann, A. Transcriptional activation by recruitment. *Nature* **386**, 569–577 (1997).
203. Brent, R. & Ptashne, M. A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. *Cell* **43**, 729–736 (1985).
204. Hope, I. A. & Struhl, K. Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. *Cell* **46**, 885–894 (1986).
205. Keung, A. J., Bashor, C. J., Kiriakov, S., Collins, J. J. & Khalil, A. S. Using targeted chromatin regulators to engineer combinatorial and spatial transcriptional regulation. *Cell* **158**, 110–120 (2014).
206. Stampfel, G. et al. Transcriptional regulators form diverse groups with context-dependent regulatory functions. *Nature* **528**, 147–151 (2015).
References 202–206 show that artificial recruitment of transcription factors and cofactors can be sufficient to drive transcription and that their activity is often context-dependent.
207. Juven-Gershon, T., Hsu, J.-Y. & Kadonaga, J. T. Caudal, a key developmental regulator, is a DPE-specific transcriptional factor. *Genes Dev.* **22**, 2823–2830 (2008).
208. van Arensbergen, J., van Steensel, B. & Bussemaker, H. J. In search of the determinants of enhancer-promoter interaction specificity. *Trends Cell Biol.* **24**, 695–702 (2014).
209. Petrenko, N., Jin, Y., Wong, K. H. & Struhl, K. Evidence that mediator is essential for Pol II transcription, but is not a required component of the preinitiation complex in vivo. *eLife* **6**, 155 (2017).
This work demonstrates that the depletion of different Mediator subunits affects transcription of a specific subset of genes more strongly than others.
210. Huminiecki, Ł. & Horbańczuk, J. Can we predict gene expression by understanding proximal promoter architecture? *Trends Biotechnol.* **35**, 530–546 (2017).
211. Bonn, S. et al. Cell type-specific chromatin immunoprecipitation from multicellular complex samples using BiTS-ChIP. *Nat. Protoc.* **7**, 978–994 (2012).
212. Lai, W. K. M. & Pugh, B. F. Genome-wide uniformity of human 'open' pre-initiation complexes. *Genome Res.* **27**, 15–26 (2017).
213. Andersson, R., Sandelin, A. & Danko, C. G. A unified architecture of transcriptional regulatory elements. *Trends Genet.* **31**, 426–433 (2015).
214. Arnold, C. D. et al. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**, 1074–1077 (2013).
This work functionally maps enhancer activity across an entire genome.
215. Muerdter, F. et al. Resolving systematic errors in widely used enhancer activity assays in human cells. *Nat. Methods* **15**, 141–149 (2018).
216. van Arensbergen, J. et al. Genome-wide mapping of autonomous promoter activity in human cells. *Nat. Biotechnol.* **35**, 145–153 (2016).
This study reports autonomous promoter activity across the human genome.
217. Nguyen, T. A. et al. High-throughput functional comparison of promoter and enhancer activities. *Genome Res.* **26**, 1023–1033 (2016).
218. Dao, L. T. M. et al. Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nat. Genet.* **49**, 1073–1081 (2017).
219. Catarino, R. R., Neumayr, C. & Stark, A. Promoting transcription over long distances. *Nat. Genet.* **49**, 972–973 (2017).
220. Young, R. S., Kumar, Y., Bickmore, W. A. & Taylor, M. S. Bidirectional transcription initiation marks accessible chromatin and is not specific to enhancers. *Genome Biol.* **18**, 242 (2017).
221. Jin, Y., Eser, U., Struhl, K. & Churchman, L. S. The ground state and evolution of promoter region directionality. *Cell* **170**, 1–21 (2017).
This work shows that transcription from newly emerged promoter regions in yeast is bidirectional and that transcription directionality is an evolutionarily selected trait.
222. Neri, F. et al. Intragenic DNA methylation prevents spurious transcription initiation. *Nature* **543**, 72–77 (2017).
223. Kim, J. et al. Blocking promiscuous activation at cryptic promoters directs cell type-specific gene expression. *Science* **356**, 717–721 (2017).
224. Lam, M. T. Y., Li, W., Rosenfeld, M. G. & Glass, C. K. Enhancer RNAs and regulated transcriptional programs. *Trends Biochem. Sci.* **39**, 170–182 (2014).
225. Tsai, A. et al. Nuclear microenvironments modulate transcription from low-affinity enhancers. *eLife* **6**, e1006441 (2017).
226. Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A. Phase separation model for transcriptional control. *Cell* **169**, 13–23 (2017).
227. Muerdter, F. & Stark, A. Gene regulation: activation through space. *Curr. Biol.* **26**, R895–R898 (2016).
228. Brangwynne, C. P. et al. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324**, 1729–1732 (2009).
229. Kato, M. et al. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767 (2012).
230. Han, T. W. et al. Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. *Cell* **149**, 768–779 (2012).
231. Larson, A. G. et al. Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240 (2017).
232. Strom, A. R. et al. Phase separation drives heterochromatin domain formation. *Nature* **547**, 241–245 (2017).
233. Dekker, J. & Mirny, L. The 3D Genome as Moderator of Chromosomal Communication. *Cell* **164**, 1110–1121 (2016).
234. Dixon, J. R., Gorkin, D. U. & Ren, B. Chromatin Domains: The Unit of Chromosome Organization. *Mol. Cell* **62**, 668–680 (2016).
235. Roider, H. G., Lenhard, B., Kanhere, A., Haas, S. A. & Vingron, M. CpG-depleted promoters harbor tissue-specific transcription factor binding signals — implications for motif overrepresentation analyses. *Nucleic Acids Res.* **37**, 6305–6315 (2009).
236. Bernstein, B. E. et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326 (2006).
237. Diao, Y. et al. A tiling-deletion-based genetic screen for *cis*-regulatory element identification in mammalian cells. *Nat. Methods* **503**, 290–635 (2017).

238. Rajagopal, N. et al. High-throughput mapping of regulatory DNA. *Nat. Biotechnol.* **34**, 167–174 (2016).
239. Lubliner, S. et al. Core promoter sequence in yeast is a major determinant of expression level. *Genome Res.* **25**, 1008–1017 (2015).
240. Patwardhan, R. P. et al. High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat. Biotechnol.* **27**, 1173–1175 (2009).
241. Bucher, P. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol.* **212**, 563–578 (1990).
242. Hahn, S., Buratowski, S., Sharp, P. A. & Guarente, L. Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and nonconsensus DNA sequences. *Proc. Natl Acad. Sci. USA* **86**, 5718–5722 (1989).
243. Arkhipova, I. R. et al. The steps of reverse transcription of *Drosophila* mobile dispersed genetic elements and U3-R-U5 structure of their LTRs. *Cell* **44**, 555–563 (1986).
244. Li, J. & Gilmour, D. S. Distinct mechanisms of transcriptional pausing orchestrated by GAGA factor and M1BP, a novel transcription factor. *EMBO J.* **32**, 1829–1841 (2013).
245. Hirose, F., Yamaguchi, M., Handa, H., Inomata, Y. & Matsukage, A. Novel 8-base pair sequence (*Drosophila* DNA replication-related element) and specific binding factor involved in the expression of *Drosophila* genes for DNA polymerase alpha and proliferating cell nuclear antigen. *J. Biol. Chem.* **268**, 2092–2099 (1993).
246. Parry, T. J. et al. The TCT motif, a key component of an RNA polymerase II transcription system for the translational machinery. *Genes Dev.* **24**, 2013–2018 (2010).
247. Tokusumi, Y., Ma, Y., Song, X., Jacobson, R. H. & 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.* **27**, 1844–1858 (2007).
248. Anish, R., Hossain, M. B., Jacobson, R. H. & Takada, S. Characterization of transcription from TATA-less promoters: identification of a new core promoter element XCPE2 and analysis of factor requirements. *PLoS ONE* **4**, e5103 (2009).

Acknowledgements
The authors thank F. Mürdter, M. A. Zabidi, P. R. Andersen, C. Plaschka and C. Bernecky for helpful comments on the manuscript. V.H. is supported by a long-term postdoctoral fellowship from the Human Frontier Science Program (HFSP, grant number LT000324/2016-L). Research in the Stark group is supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 647320) and by the Austrian Science Fund (FWF, F4303-B09). Basic research at the Institute of Molecular Pathology (IMP), Vienna, Austria, is supported by Boehringer Ingelheim GmbH and the Austrian Research Promotion Agency (FFG).

Author contributions
All authors contributed equally to all aspects of the article.

Competing interests
The authors declare no competing interests.

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

Reviewer information
Nature Reviews Molecular Cell Biology thanks S. Spicuglia and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Supplementary information
Supplementary information is available for this paper at https://doi.org/10.1038/s41580-018-0028-8.