universität wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation /Title of the Doctoral Thesis

Identification and characterization of promoter-bound proteins

verfasst von / submitted by

Leonid Serebreni, BSc MSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Wien, 2021 / Vienna 2021

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 794 685 490
Dissertationsgebiet lt. Studienblatt / field of study as it appears on the student record sheet:	Molekulare Biologie
Betreut von / Supervisor:	Dipl.-Biochem. Dr. Alexander Stark

Aknowledgments

The achievements outlined in this thesis, the gain in scientific knowledge made during my PhD was part of a team effort, is attributed to the following
people whose contribution was invaluable form my learning process and success:

I would like to thank Alexander Stark for his supervision and support through this project, and for teaching me valuable principles in experimental design which I will continue applying through my scientific career.

I would like to thank Mamduh Zabidi for making the scientific discoveries that inspired this project, and for helping identify promoter motif matches.

I would like to thank Anna Vlasova and Vanja Haberle for help with computational analysis which enabled the identification of promoter-bound proteins and transcription initiation patterns.
I would like to thank Lisa-Marie for her persistent dedication to improve and implement NGS data analysis which was invaluable for this project.

I would like to thank Oliver Hendy and Katharina Bergauer for their help with engineering cell lines and performing PRO-seq.

I would like to thank Vincent Loubiere for help with computational analysis and for helping me improve my own computational skills.

I would like to thank Christoph Neumayr, Franziska Lorbeer, Jelle Jacobs, Ashley Woodfin, Bernardo Almeida, Loni Klaus, Lorena Hofbauer, Filip Nemcko, Martina Rath, Michaela Pagani, Felix Muerdter, Matus Vojtek and Peter Zoescher for their friendship, discussions and advice through my PhD.

If you know me, you expect me to thank my cat now, which is exactly the purpose of this section. I would like to thank my cat Dooby for keeping me sane through my time in Vienna and serving as my family in Europe.

Table of Contents

Summary

Transcription by RNA Polymerase II (Pol II) starts within promoter regions. Despite decades of *in vitro* biochemical and structural characterization of the transcription machinery, it is still unknown how the functional properties and sequence diversity of promoter regions engage with the transcription machinery for productive transcription to occur. During my PhD I set out to identify the protein binders of functionally distinct promoter regions with the aim of understanding how these proteins facilitate transcription from functionally distinct promoter types. Building on our lab's knowledge of promoters in *Drosophila melanogaster*, I have used developmental and housekeeping promoter types from the fruit fly genome in a DNA-affinity purification assay coupled to mass spectrometry. This allowed me to find shared and promoter-specific protein binders. To investigate the functional contribution to the transcription of various proteins I generated auxin-inducible degron tagged Drosophila S2 cell lines and measured nascent transcription after their acute depletion. I was able to identify proteins that were required for transcription by distinct subsets of promoters, and one, TFIIA - was required by all promoters. Surprisingly, components of the canonical Pol II transcription machinery preferentially assembled on developmental but not housekeeping promoter types. With TFIIA required by both promoter types I found that housekeeping promoters use a distinct mechanism of Pol II recruitment through housekeeping promoter-specific transcription factors and cofactors. This study opens the investigation into the interactions between transcriptional cofactors and the transcription machinery, and how these interactions are exploited in mammalian promoters.

Zusammenfassung

Die Transkription durch RNA-Polymerase II (Pol II) beginnt an Promotorregionen. Trotz jahrzehntelanger biochemischer und struktureller In-vitro-Charakterisierung der Transkriptionsmaschinerie ist immer noch unbekannt, wie die funktionellen Eigenschaften und die DNA-Sequenzvielfalt von Promotorregionen mit der Transkriptionsmaschinerie zusammenwirken, damit eine produktive Transkription stattfinden kann. Während meiner Promotion habe ich mir vorgenommen, die Proteinbinder von funktionell unterschiedlichen Promotorregionen zu identifizieren, um zu verstehen, wie diese Proteine die Transkription von funktionell unterschiedlichen Promotortypen erleichtern. Aufbauend auf den Kenntnissen unseres Labors über Promotoren in Drosophila melanogaster habe ich entwicklungs- und haushaltsübliche Promotortypen aus dem Genom der Fruchtfliege in einem DNA-Affinitätsreinigungsverfahren verwendet, das mit Massenspektrometrie gekoppelt war. Dies ermöglichte es mir, gemeinsame und promoterspezifische Proteinbinder zu finden. Um den funktionellen Beitrag verschiedener Proteine zur Transkription zu untersuchen, habe ich Auxin-induzierbare, mit Degron markierte Drosophila S2-Zelllinien erzeugt und die naszierende Transkription nach ihrer akuten Deplektion gemessen. Es gelang mir, Proteine zu identifizieren, die für die Transkription durch verschiedene Untergruppen von Promotoren erforderlich waren, und eines, TFIIA, das für alle Promotoren erforderlich war. Surprisingly, Komponenten der kanonischen Pol-II-Transkriptionsmaschinerie, wurden bevorzugt an Entwicklungspromotoren, nicht aber an Housekeeping-Promotoren montiert. Da TFIIA von beiden Promotortypen benötigt wird, konnte ich feststellen, dass Haushalte-Promotoren einen anderen Mechanismus der Pol II-Rekrutierung durch Haushalte-Promotoren-spezifische Transkriptionsfaktoren und Kofaktoren nutzen. Diese Studie eröffnet die Untersuchung der Interaktionen zwischen Transkriptions-Cofaktoren und der Transkriptionsmaschinerie und wie diese Interaktionen in Säugetierpromotoren ausgenutzt werden.

Introduction

1. Transcription by RNA Polymerase II

Control of gene expression is central for the ability of multicellular organisms to give rise to and maintain different cell types. Present in all cells of a given animal, a single genome directs both ubiquitous and cell-type specific expression of various genes. Protein coding genes are largely responsible for determining the various biochemical and cell-biological properties associated with distinct cellular identities (Orphanides and Reinberg, 2002). Out of the three nuclear RNA polymerase enzymes, transcription of protein coding genes and many non-coding RNAs is performed by RNA polymerase II (Pol II) (Schier and Taatjes, 2020). Pol II transcription originates at promoter regions, which are non-coding regulatory elements most often located upstream of their target genes (Haberle and Lenhard, 2016). Despite initiating transcription at the promoter, Pol II itself is unable to recognize the promoter DNA sequence, and requires a set of proteins to be guided to the transcription start site (TSS) (Drapkin et al., 1993). Pol II assembles into the transcription pre-initiation complex (PIC) together with highly conserved general transcription factors (GTFs) at promoter regions (Figure 1A) (Murakami et al., 2013). *In vitro* biochemical reconstitution showed that the PIC can sequentially assemble on promoter DNA (He et al., 2013). First to assemble at the promoter are TFIID and TBP, followed by TFIIA and TFIIB (Cianfrocco et al., 2013; Plaschka et al., 2015). This assembly of GTFs facilitates Pol II recruitment in complex with TFIIF. Subsequently, TFIIE and the helicase containing TFIIH complex complete assembly of the PIC. Initiation of transcription then occurs by melting/opening of the double stranded DNA helix and invasion of the template strand near the active site of RNA Pol II when nucleoside triphosphates are present (Cheung and Cramer, 2012).

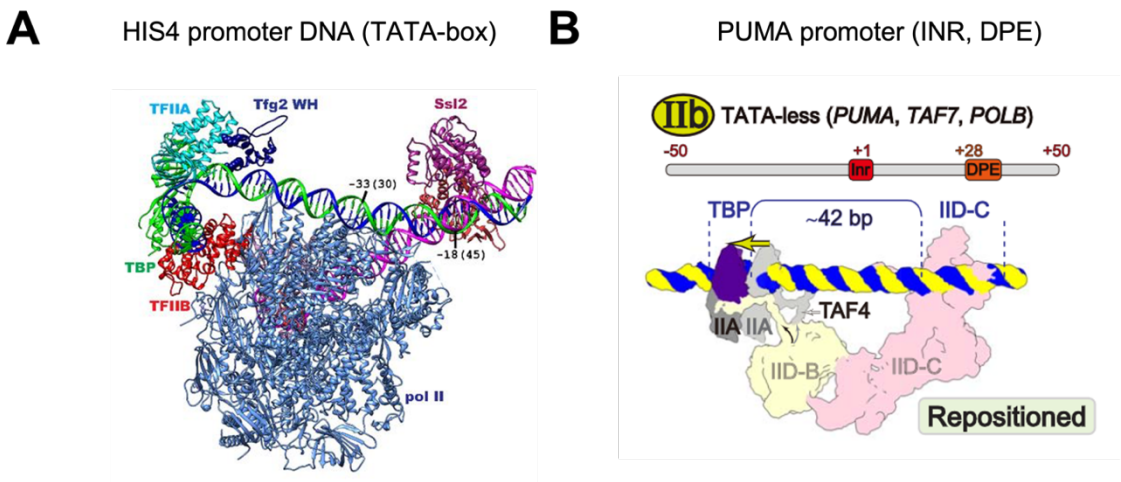


Figure 1. A, Structure of the Pol II PIC on the TATA-box containing HIS4 promoter (Murakami, K. *et al.*). B. Structure of PIC members TFIID, TBP and TFIIA on the TATA-less, INR/DPE containing PUMA promoter (Chen, X. *et al.*).

Recent structural investigations of Pol II PIC assembly on promoter DNA have shown that a canonical Pol II PIC can assemble on both TATA-box and TATA-less promoter DNA that is instead containing only an Initiator (INR) and down-stream promoter element (DPE) motifs (X. Chen et al., 2021). These promoters represent less than half of all Drosophila promoters, and less than 10% of promoters found in mammalian genomes (FANTOM Consortium and the RIKEN PMI and CLST (DGT) et al., 2014; Ohler et al., 2002). Having clear evidence of canonical PIC assembly *in vitro* for only a few tested promoter sequences, we are still lacking information on how GTF paralogs incorporate into the PIC at the majority of promoters found in the genome.

Most members of the Pol II PIC are present in the genome as highly conserved single-copy genes (Orphanides et al., 1996). However, during metazoan evolution, the rise of Bilateria and complex animal body plans were accompanied with the emergence of multiple GTF paralogs (Duttke et al., 2014). Importantly, GTF paralogs are observed predominantly in PIC components that directly interact with promoter DNA, coinciding with evolution of a greater regulatory DNA element landscape present in metazoans (Degnan et al., 2009; Peterson and Davidson, 2000). For instance, the Drosophila paralog of TFIIA, moonshiner, has been shown to recruit Pol II to transcription start sites of transposable elements in the ovaries of *Drosophila melanogaster* (Andersen et al., 2017). Additional paralogs have emerged in components of the multi-subunit TFIID complex. Specifically, TAF7L and TAF9B paralogs of TFIID components are critically important in driving transcription during adi-

pocyte and motor neuron specification (Neves and Eisenman, 2019; Zhou et al., 2014). Such evidence has shown that components of the transcription machinery can be important for cell-type specific transcription of select promoters. Regarding paralogs, perhaps none is more important in driving cell-type specific gene expression as TBP paralogs (Akhtar and Veenstra, 2011). Most interestingly, TBP paralogs such as the mammalian TRF1 and the *Drosophila melanogaster* TRF2 have been implicated in driving transcription of most Pol II target genes in germ cells or select groups of promoters in somatic tissues respectively (Baumann and Gilmour, 2017; Wang et al., 2014; Yu et al., 2020). Despite having lost the ability to directly bind DNA, TRF2 has been shown to interact with Pol II PIC components suggesting it can be incorporated into the PIC and perhaps even replace TBP (Zehavi et al., 2015). Alluding to a potential modularity of the PIC, a few studies have shown that various GTFs can remain bound as a complex on promoter DNA without Pol II. Specifically, a complex consisting of TBP, TFIIA and TFIIB; or a four GTF complex containing TBP, TFIIA, TFIIE and TFIIH which was transcription competent upon addition of TFIIB, and Pol II-TFIIF (Murakami et al., 2013). A key GTF that was excluded from the mentioned study was TFIID, a 13 subunit complex whose components have been shown to crosslink at most active promoter regions (Sun et al., 2021). TFIID has been shown to rearrange upon binding to DNA and exhibit two distinct conformations, whose interaction with DNA depends on TFIIA (Cianfrocco et al., 2013). These two different conformations have been suggested to help facilitate specific interactions with distinct factors, such as different TBP paralogs, but experimental evidence is still lacking thereof.

Upon successful recruitment of RNA pol II to the promoter region and opening of the double stranded DNA template, transcription begins but often abruptly terminates after transcribing 20 to 100 base pairs, failing to achieve productive elongation (Core and Adelman, 2019). This phenomenon has been termed ‘pausing’, and is a conserved rate-limiting feature of RNA pol II transcription among higher eukaryotes, and has been related to *E.coli* polymerase pausing at a few promoters (Gressel et al., 2019; Newell and Gray, 2010). Pol II pausing has been shown to be regulated by the factors DSIF and NELF, and correlate with a specific phosphorylated state of the Pol II carboxy-terminal domain as it transitions from initiation to elongation(Bernecky et al., 2017; Wu et al., 2003). Recent studies have shown that the paused state of Pol II can vary in duration from minutes to hours at various promoters and the pause

stability correlates with distinct promoter sequence features (K. Chen et al., 2013; Shao et al., 2019). Importantly, Pol II pausing has been shown to inhibit new rounds of transcription initiation, potentially through steric hindrance at the promoter proximal region (Shao and Zeitlinger, 2017; Gressel et al., 2019).

Recruitment, initiation and pausing are the most important stages of early Pol II transcription and are regulated by various protein factors. Despite extensive structural and biochemical studies of these early stages, the vast diversity of promoter regions at which Pol II transcription occurs cannot be reconciled with an individual model in which a distinct l PIC can recognize the DNA sequence and assemble to drive transcription. In the next section I will describe the promoter context at which transcription happens and the recent evidence for functionally distinct promoters that opens the possible existence of different modes of PIC recruitment.

2. Promoter diversity

Aptly described by James Kadonaga as the “gateway to transcription” the promoter region is the site of Pol II PIC assembly and the transcription start site (Juven-Gershon et al., 2008). Early studies in yeast determined that promoter sequences of Pol II TSSs are largely homogenous in terms of their sequence motif composition, containing a TATA-box or a TATA-like element capable replacing its function (Singer et al., 1990). Additionally, all yeast Pol II promoters were shown to require the canonical GTFs (Petrenko et al., 2019). However, in the past two decades our understanding of promoter function in higher eukaryotes has expanded. With the emergence of multicellular organisms, genomes evolved to contain greater numbers and more complex regulatory regions that facilitate the differential gene expression programs required to support specialized cell types (Degnan et al., 2009; King et al., 2003). Regulatory element diversification is observed in both enhancer and promoter regions and is associated with directional transcription driven from promoters (Rubinstein and de Souza, 2013; Tunnacliffe et al., 2018). For instance, the highly conserved TATA-box is present in a majority of yeast promoters, however in mammals it is observed in less than 10% of promoters, and in less than 2% of active promoters in most cell types (Yang et al., 2007). In fact, mammalian promoters do not exhibit unified promoter grammar, but rather contain many different TF binding motifs (Xie et al., 2005). The diversity of promoter sequences present in metazoans presents a

conflict for unifying *in-vitro* biochemical data on Pol II recruitment and assembly into a single PIC.

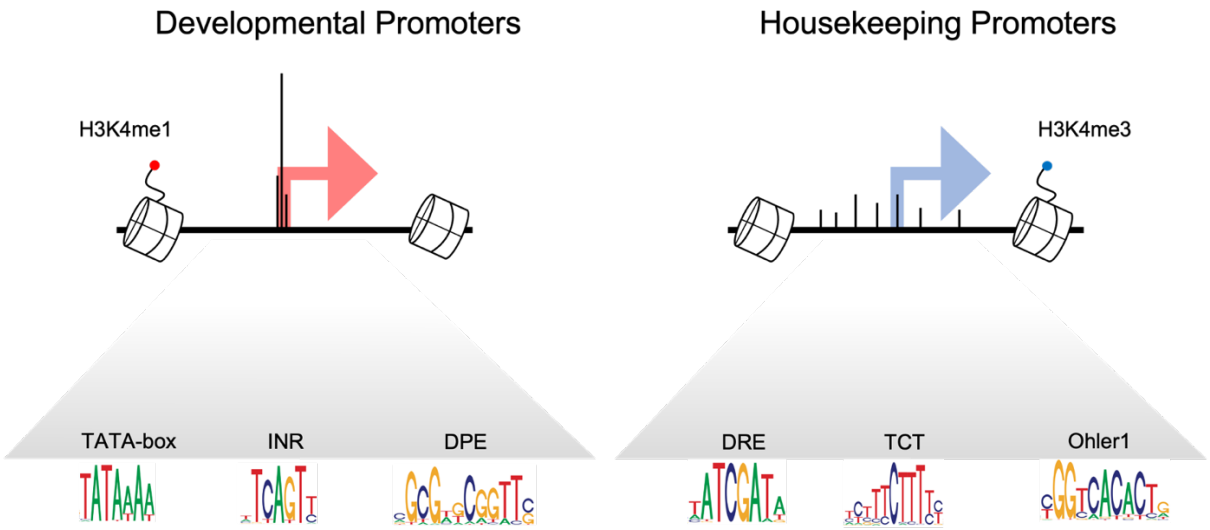


Figure 2. Developmental promoters are characterized by having focused transcription initiation patterns (black bars), and H3K4me1 upstream of the TSS, in addition to being enriched in TATA-box, INR and DPE motifs. Housekeeping promoters are characterized by focused transcription initiation patterns, H3K4me3 and enriched in motifs such as the DRE, TCT, and Ohler1.

Uncovering the relationship between promoter grammar and function is key to understanding how and if a single canonical PIC can assemble at different promoters. To aid in the dissection of promoter grammar, organisms that exhibit defined promoter sequence features are ideal to approach this question. In our lab we have used the genome of the fruit fly, *Drosophila melanogaster* as a model to dissect promoter function. In addition to the TATA-box, Drosophila promoters contain up to 10 known sequence motifs (Ohler et al., 2002). Commonly found alone, the Initiator (INR) motif is found in over 1000 promoters. Interestingly, the down-stream promoter element (DPE) commonly appears together with INR, is rarely observed together with the TATA-box, suggesting it has evolved divergent functions (Kutach and Kadonaga, 2000). Unifying the three mentioned promoter classes together is the commonality of focused transcription initiation patterns, in which Pol II preferentially initiates transcription within a short region 1-3 nucleotides in length (Figure 2). A focused initiation pattern suggests that the promoter sequence directs rigid DNA-PIC interactions that restrict transcription initiation to a narrow region. Indeed, the TATA-box, INR and DPE motifs are often found in stereotypic spatially restricted position in relation to the TSS(Haberle and Stark, 2018). With the TATA-box directing TBP binding

and subsequent PIC recruitment, and INR and DPE interacting with TFIID components further reinforce promoter-PIC positioning. Together, the TATA-box, INR and DPE are cumulatively found in approximately 40% of Drosophila promoters and are enriched in genes that show cell-type, and developmental stage specific gene expression (Haberle et al., 2019). Promoters of this class have been exclusively used as templates for *in-vitro* biochemical assays with purified PIC components which enabled us to understand how the PIC assembles on promoter DNA.

The other major class of promoters in Drosophila devoid of the motifs present in developmental promoters are primarily associated with ubiquitously expressed genes. Ubiquitously expressed genes often control 'housekeeping' processes associated with metabolic activity and cell cycle progression. Motifs identified in housekeeping promoters include the Ohler 1,6 and 7 motifs, the DNA replicated-related elements (DRE), and the TCT motif which is enriched in ribosomal and translation regulating protein promoters (Arnold et al., 2016). Unlike developmental motifs that have been shown to interact with Pol II PIC components, some housekeeping promoter motifs have been shown to bind sequence-specific transcription factors. For example, the Ohler 1 motif binds the motif-1-binding-protein (M1BP) and is sufficient to recruit M1BP and its interacting cofactor GFZF(Baumann and Gilmour, 2017). Similarly, the DRE motif binds the factor DREF, which is found at promoters of DNA replication-related genes(Hirose et al., 1993). Protein targets for additional housekeeping motifs remain unknown. Opposed to developmental promoters, the unifying features of housekeeping promoters are dispersed (or wide) transcription initiation patterns which occur in a wider region of 20 to 100 nucleotides around the dominant TSS position. The only exceptions to this observation are TCT promoters, which exhibit focused transcription initiation patterns (Parry et al., 2010). In-line with the formation of a rigid PIC, the TBP paralog TRF2 has been shown to crosslink to TCT promoters together with DREF (Hochheimer et al., 2002). However, whether and how such a PIC may form at TCT promoters is unknown.

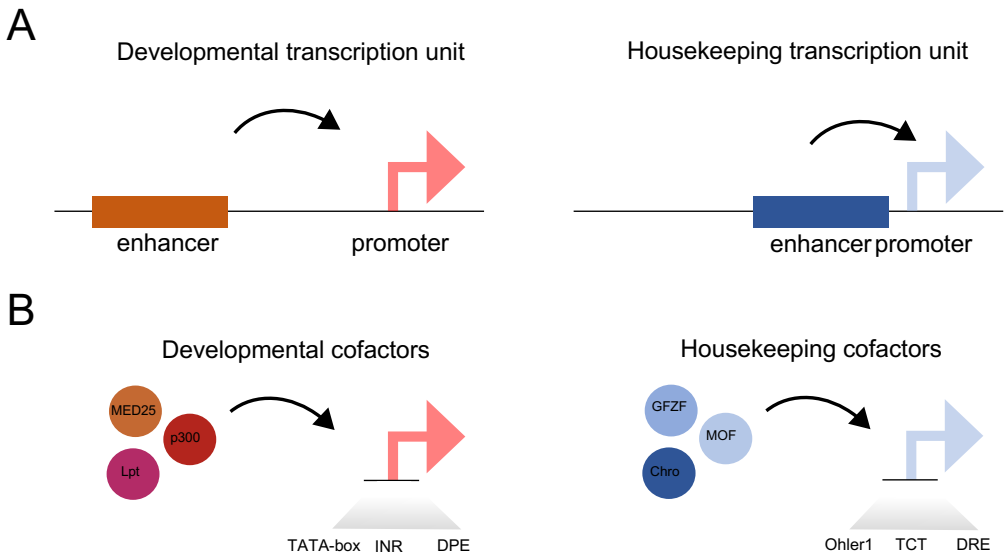
Additional differences observed between housekeeping and developmental promoters lie on the chromatin level. Nucleosomes have been shown to be a barrier for Pol II and transcription, and thus require ATP-dependent chromatin remodelers to evict or shift nucleosomes to provide accessible DNA around the TSS (Studitsky et al., 2016). Such remodeling activities are contributed by SWI/SNF and ISWI family remodelers determine the positioning or phasing of nucleosomes in relation to the

TSS (Becker and Hörz, 2002; Tolkunov et al., 2011). Examination of nucleosome positions at developmental promoters revealed lack of strong preference for specific positions, while housekeeping promoters exhibit strong nucleosome phasing with a defined +1 nucleosome position downstream of the TSS (Rach et al., 2011). It is not clear how the nucleosome structure relates to PIC occupancy as the mentioned measurements have been made in bulk cell populations. Developmental promoters may be less coordinated in their expression levels in a given population of cells than housekeeping promoters. The second major chromatin difference between housekeeping and developmental promoters lies in the post translational modifications (PTMs) of N-terminal histone tails. Specifically, histone H3 lysine 4 methylation (H3K4me) has been shown to be differentially distributed between the two promoter types. H3K4me1 is preferentially found at developmental promoters, while H3K4me3 is preferentially found at housekeeping promoters (Haberle et al., 2019). Concordantly, the methyltransferases responsible for these marks, TRR and SET1, components of COMPASS methyltransferase complex also preferentially crosslink to developmental and housekeeping promoters respectively (Haberle et al., 2019). Surprisingly, H3K4me1 was found to be dispensable for transcription in mammals and thus might only be a consequence but not a cause for distinct promoter function (Rickels et al., 2017).

3. Promoter functionality

In the Stark lab, transcription is studied from a functional perspective where important insight on promoter function has been achieved in the past decade. Promoters are the sites of Pol II PIC recruitment and assembly, but also the site where activating signals are integrated to increase rates of transcription beyond basal levels, which can be achieved by increasing the rates of Pol II recruitment or productive transcription elongation events (Haberle and Stark, 2018). Activating signals are transmitted through enhancer regions which communicate to the promoters through regulatory proteins collectively termed transcriptional cofactors (COFs) (Zabidi and Stark, 2016). Our lab and others have characterized enhancers and COFs and found that there is both a large number and diversity of regulatory DNA elements and proteins that can activate transcription from promoters (Stampfel et al., 2015). Most importantly, by coupling enhancers and COFs with different promoters, previous stud-

ies have found that not all enhancer and COFs activate the same promoters, but rather show specificity (Zabidi et al., 2015). Enhancers that preferentially activate developmental promoters are called developmental enhancers, and enhancers that preferentially activate housekeeping promoters are called housekeeping enhancers. This preference most strongly manifests itself when coupling developmental and housekeeping regulatory elements on the same piece of reporter DNA, fails to result in active transcription. Developmental enhancers tend to be more distal to their cognate promoter, while housekeeping enhancers tend to be promoter proximal and often overlap the promoter region.



Enhancer-promoter specificity indicates that this communication has evolved to utilize different proteins. Enhancers have been proposed to recruit different COFs such as chromatin remodelers, histone modifiers, and transcription elongation factors that increase rates of transcription initiation and elongation from promoter regions(Cubeñas-Potts et al., 2017). When testing COFs for their ability to activate transcription by forced tethering them to a genome-wide promoter library, a previous study found that COFs also preferentially activate some but not all promoters. For example, the Mediator subunit MED25 and the acetyltransferase p300 preferentially activated developmental promoters containing TATA-box, INR and DPE motifs, while COFs such as the acetyltransferase MOF and GFZF preferentially activate housekeeping promoters containing TCT, Ohler1 and DRE motifs (Haberle et al., 2019). The differences in COF responsiveness and the enrichment of different sequence

motifs strongly suggests that functionally distinct promoters may utilize different proteins in integrating the activating cues for transcription to occur.

4. Approaches used to identify DNA-binding proteins

Aiming to identify the protein interactors of regulatory elements, multiple studies developed various approaches to this effort. All approaches involve the biochemical isolation of naked DNA or chromatin from *in vivo* or *in vitro* material which is coupled to mass spectrometric analysis of the resulting protein binders. The first method developed for the identification of protein-DNA interactors, and the method I have used in my attached publication is the immobilized-DNA-template assay (also termed DNA-affinity purification) (Kadonaga and Tjian, 1986). Famously used to identify one of the first transcription factors, Sp1, DNA-affinity purification works by generating a biotinylated double stranded DNA probe of interest, often done by PCR amplification with a biotinylated primer (Dynan and Tjian 1983). The biotinylated DNA template is then immobilized on streptavidin beads as a solid support and incubated with nuclear extract to capture the binding partners of the region of interest. Recently, this approach has been coupled to mass spectrometry and was used to successfully identify protein binders of various non-coding genomic regions (Butter et al., 2012; Sequeria and Vermeulin, 2020; Viturawong et al., 2013). Despite its popular use, one caveat of DNA-affinity purification is the lack of chromatin/nucleosome structure on bait DNA, and lack of nuclear context such as protein concentrations and nuclear environment in an *in vitro* extract when the investigation is interested in identifying chromatin binding proteins. A limited use of chromatinized DNA template was applied but due to the challenge of *in vitro* chromatin assembly it has not been widely adopted (“A Mechanism for Coordinating Chromatin Modification and Preinitiation Complex Assembly,” 2006). Most studies employing DNA-affinity purification to promoter regions have recruited strong transcriptional co-activators such as the viral activator VP16 to the DNA template and captured the Pol II PIC (Lin and Carey, 2012). In my study, I did not recruit co-activators, but rather asked what proteins can bind promoter regions in the absence of an activator or active transcription. This strategy uncovers the sequence encoded promoter contribution to protein-DNA interactions. Applied to functionally distinct promoters, DNA-affinity purification reveals

which proteins are preferentially enriched at the differently tested promoter DNA over control DNA.

Additional methods that have been used to identify protein-DNA interactions tried to isolate endogenous loci with different affinity handles. One such method utilized locked nucleic acids (LNAs) to hybridize chemically fixed and sheared chromatin. With a higher melting temperature and duplex stability, biotinylated LNAs were successfully used to capture the repetitive telomere repeats. Despite their successful application to repetitive loci such as telomeres and pericentromeric regions, they have yet to be applied successfully to unique genomic loci (Déjardin and Kingston, 2009; Saksouk et al., 2014).

More direct methods that probe for proteins at a genomic locus of interest relied on an affinity tag placed on a protein recruited to a single site. The first one utilized dCas9 fused to an affinity tag dCas9 (Fujita and Fujii, 2016; Tsui et al., 2018). This approach is flexible in terms of the amount of sites that the dCas9 can probe, but required prohibitive amounts of input material resulting in low enrichment values. Similarly, engineering binding sites at a target locus by inserting LacO or TetR sites and recruiting their cognate binder fused to an affinity tag resulted in weak enrichments (Griesenbeck et al., 2003; Newell and Gray, 2010)Both methods have seen a limited use for identifying locus-specific proteomes, with the main technical challenge of optimizing the ratio of their recruited protein to the genomic locus as opposed to its unbound soluble concentrations.

Recently, additional approaches to study locus specific proteomes have been developed utilizing newly applied chemistries. By utilizing the biotin ligase BirA that contains a mutation, it allows the protein to promiscuously biotinylate lysine residues in a vicinity of 10nm (Roux et al., 2018). By fusing BirA to dCas9 it was then possible to recruit it to a specific genomic locus, and with the addition of biotin to examine the resulting biotinylated fraction in mass spectrometry (Schmidtmann et al., 2016). Similarly, recruiting dCas9 fused to the peroxidase APEX2 which generates biotin-phenol radicals upon the addition of hydrogen peroxide it was possible to enrich for biotinylated proteins in a short time window with hydrogen peroxide treatment (Qiu et al., 2019). Unfortunately, both methods suffer from unspecific protein enrichment and low signal to noise ratio presumably due to the lack of control over soluble vs. bound concentrations of dCas9 and have seen limited use for targeting unique genomic loci.

Aims

1. Identify promoter-DNA interacting proteins

2. Characterize the functional contribution of select proteins to transcription by rapid depletion coupled to nascent transcript sequencing

3. Consolidate models for how functionally distinct promoters recruit the Pol II PIC to initiate transcription

Results and discussions of the publications

Publication #1: “Insights into gene regulation: From regulatory genomic elements to DNA-protein and protein-protein interactions”

Authors: Leonid Serebreni & Alexander Stark
Published: Current Opinion in Cell Biology, June 2021, 70:58-66.

Discussion and summary of the publication:
In this manuscript we summarized the most important and insightful studies published in recent years (2018-2021) in the field of gene regulation. We have chosen to highlight a breadth of new technologies and insights that encompass observations regarding the dynamic nature of transcription to emerging biophysical concepts of phase-separation at sites of transcription. Transcription is a discontinuous process often taking place in bursts of active periods. Recently live-cell imaging studies corelated the residence time and bound fraction of TFs with the kinetic parameters of transcription bursts. In additional single-cell sequencing technologies are emerging to capture transcriptional bursting dynamic at single molecule resolution. We then summarize important studies that investigated the influence of the genome’s topological structure to gene expression. A picture emerges where transcription is largely robust to perturbations of genome topology but is benefitting from pathogenic enhancer-promoter adoption at specific loci that are important for development. We continue with the proposition of encoded enhancer-cofactor compatibility as an explanation for the robustness of gene expression to most topological rearrangements. Ample emerging evidence from our lab and others is showing that various transcriptional cofactors preferentially activate distinct promoters and gene-regulatory circuits. Finally, we discuss how the specificity of various transcriptional regulators can be facilitates in the context of a liquid-liquid phase separated condensate. New studies that have tried to dissect the function and composition of various condensates have seen that they can be enriched in specific cofactors such as the Mediator complex or elongation factors. Our outlook focuses on combining single molecule live-cell imaging with novel sequencing methods to dissect the contribution of various transcriptional regulators at different transcribed loci.
Author contribution: L.S and A.S. share equal contribution for writing and choosing the topics of the manuscript.

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Cell Biology

Insights into gene regulation: From regulatory genomic elements to DNA-protein and protein-protein interactions

Leonid Serebreni[1] and Alexander Stark[1,2]

Abstract

Transcription is orchestrated by non-coding regulatory elements embedded in chromatin, which exist within the larger context of chromosome topology. Here, we review recent insights into the functions of non-coding regulatory elements and their protein interactors during transcription control. A picture emerges in which the topological environment constraints enhancer-promoter interactions and specific enhancer-bound proteins with distinct promoter-compatibilities refine target promoter choice. Such compatibilities are encoded within the sequences of enhancers and promoters and realized by diverse transcription factors and cofactors with distinct biochemical activities. An emerging property of transcription factors and cofactors is the formation of nuclear microenvironments or membraneless compartments that can have properties of phase-separated liquids. These environments are able to selectively enrich certain proteins and small molecules over others. Further investigation into the interaction of transcriptional regulators with themselves and regulatory DNA elements will help reveal the complexities of gene regulation within the context of the nucleus.

Addresses

[1] Research Institute of Molecular Pathology (IMP), Vienna BioCenter (VBC), Vienna, Austria

[2] Medical University of Vienna, Vienna BioCenter (VBC), Vienna, Austria

Corresponding author: Stark, Alexander (stark@starklab.org)

Current Opinion in Cell Biology 2021, **70**:58–66

This review comes from a themed issue on **Cell Nucleus**

Edited by **Jane Skok** and **Daniel Gerlich**

For a complete overview see the Issue and the Editorial

Available online 29 December 2020

https://doi.org/10.1016/j.ceb.2020.11.009

Keywords

Gene regulation, Transcription, Regulatory elements, Transcription factors, Cofactors, Microenvironment.

Introduction

The development and maintenance of different cell types in multicellular organisms is driven by differential gene expression programs. Control of gene expression is primarily determined at the level of transcription, the process of copying DNA to RNA. Production of RNA by RNA polymerase II (Pol II) initiates at promoter regions. Recruited to promoter regions as part of the pre-initiation complex (PIC), Pol II can only give rise to low levels of transcription and non-coding regulatory enhancer regions are required to boost the rates of productive transcriptional initiation. Enhancers bind transcription factor (TF) and cofactor proteins to increase rates of transcription from promoters. Uncovering the molecular mechanisms of how these players function together to activate transcription is a major challenge in modern biology with far-reaching implications for our understanding of how gene expression programs are regulated to enable cell-type differentiation and development.

Here, we review novel approaches to study the dynamics of the transcription process and recent insights into how regulatory specificity is mediated by genome topology and by biochemical compatibilities between regulators. We conclude by discussing emerging properties of TFs and cofactors, including their biophysical property to form liquid-liquid phase separated microenvironments that facilitate selective protein-protein interactions among themselves and the regulatory elements they bind.

Transcriptional bursting dynamics

Transcription levels differ vastly between genes, yet how such diverse transcription rates are controlled remains unclear. Transcription is not a continuous process, but occurs in discreet bursts of transcription separated by intervals of apparent inactivity, termed *refractory periods*, and can be described by two parameters: *burst size* (the number of RNA molecules produced per burst) and *burst frequency*, i.e., how many bursts occur in a given time interval [1,2]. Transcription can thus be controlled by modulating burst sizes, burst frequencies, or both.

Live-cell imaging can visualize nascent transcripts and is increasingly used to estimate bursting dynamics of individual genes. The human estrogen responsive TFF1 gene, for example, showed short bursts of transcription and very long refractory periods [3] that lasted from

hours to days, thus potentially explaining the stochasticity between individual cells in a population [4]. Imaging is also used to study which mechanisms could influence or regulate bursting dynamics. The DNA-binding dynamics of the TFs Gal4 and Glucocorticoid Receptor correlated with bursting, in particular, the TFs' dwell times and the bound fractions [5,6]. Surprisingly, changes in burst frequency and burst size seemed to be independent of Pol II recruitment as the manipulation of bursting parameters by cell differentiation or BET protein inhibition affected burst size and frequency but not Pol II recruitment rates [7]. Such observations suggest that some types of promoters may utilize a different rate-limiting step, distinct from Pol II recruitment, to license transcriptional bursts.

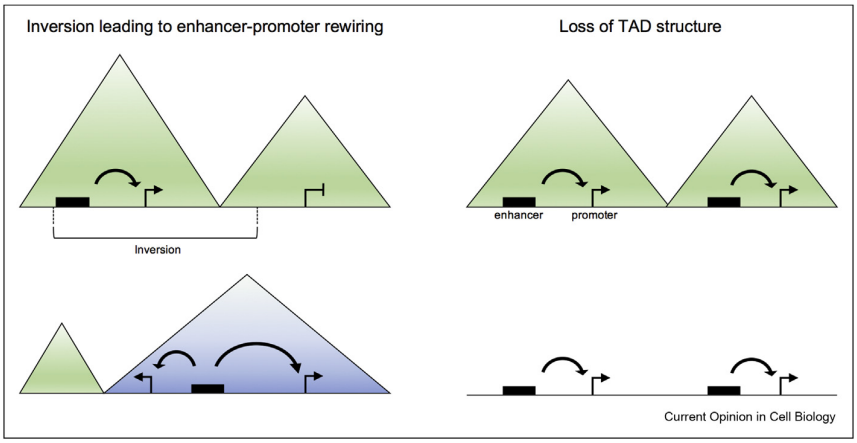
An exciting development is the application of allele-resolved single-cell RNA-sequencing to estimate transcriptional burst parameters for thousands of expressed genes in parallel [8•]. At a genome-wide scale, this confirmed that enhancers modulate burst frequency while promoters control burst size, with TATA-box promoters having the largest burst sizes. We anticipate that this method will enable exciting insights into the regulation of bursting.

Genome topology and gene expression

Genes and their regulatory elements are typically embedded within so-called topologically associating domains (TADs) defined by high(er) intra-TAD and low(er) inter-TAD chromatin contact frequencies. As TADs can therefore promote intra-TAD regulatory communication over interactions with regulatory elements in neighboring TADs, they may act as functional units to instruct or limit transcriptional control [9–12]. Examples of TAD functionality emerged from studies investigating structural variants associated with disease [13]. In a model of branchiooculofacial syndrome, the disruption of a TAD encompassing the TFAP2A gene by an inversion resulted in lower TFAP2A expression, presumably because the TFAP2A promoter was disconnected from its cognate enhancers [14]. Conversely, inversions in the Epha4 locus led to aberrant gene expression by newly formed contacts with enhancers from a neighboring gene cluster [15]. Taken together, cases of disease-associated structural variants implicate topological boundaries in maintaining gene expression robustness and restricting aberrant enhancer-promoter interactions (Figure 1).

These examples of TAD disruptions indicate a critical role of TADs in maintaining gene expression states and explain disease-phenotypes that cannot otherwise be explained. Interestingly however, other studies that perturbed TAD structures saw little effect on gene expression: deletion of CTCF sites between TADs in the Sox9-Kcnj2 locus led to TAD fusion but no major change in gene expression [16•] and changing or removing TAD boundaries in the Shh locus did not affect Shh expression [17•]. Similarly, in the Hox-D gene locus, removal of boundaries resulting in TAD fusions did not alter the levels or spatial patterns of Hox-D gene expression [18]; and gene expression from *Drosophila* balancer chromosomes, which have undergone thousands of structural rearrangements over decades of selection (which likely restricted the severity of possible outcomes), is largely intact with only mild changes near breakpoints [19•].

Figure 1



TADs constrain enhancer-promoter interactions. Structural variants such as inversions can cause misregulation of gene expression by rewiring enhancer-promoter interactions. However, the global disruption of TAD structure does not cause large changes in gene expression.

Similar results were observed when TADs were acutely disrupted such that compensatory mechanisms could be excluded: disruption of TADs by acute depletion of cohesin in mESCs abolished topological boundaries and domains [20], but found only modest changes in gene expression that were restricted to genes in proximity of highly active super-enhancers. Global abolition of TADs was also achieved by the acute depletion of the boundary-forming protein CTCF, with similarly limited impact on gene expression [21,22].

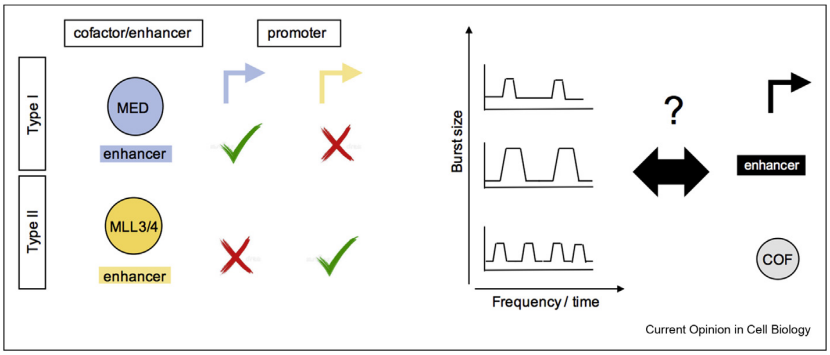
While the reasons for these distinct outcomes are still unclear and might relate to the formation of new contacts rather than the mere disruption of old ones [16], it seems that gene expression is often, but not always, robust against perturbations of the genes' topological environment. Together with the observation that not all genes within the same TADs are co-regulated, these results suggest that TADs alone cannot explain individual enhancer-promoter contacts and gene regulatory specificity and that additional mechanisms must be involved.

Transcriptional cofactors

In the disease-associated cases discussed above, aberrant enhancer-promoter interactions led to gene misexpression of some, but not all, genes, and this selectivity may correspond to an inherent property encoded within enhancers and promoters. Such compatibility has indeed been demonstrated for *Drosophila* regulatory elements by the forced pairing of different promoters with genome-wide sets of enhancers [23,24]. Moreover, such regulatory compatibilities have been recapitulated at the level of *Drosophila* and human transcriptional coactivators that mediate enhancer-promoter communication and transcriptional activation [25•]: upon forced recruitment to a complex library of promoters, coactivators such as mediator or MLL3/4 preferentially activated different promoters. Such preferences between cofactors and promoters suggest that different mechanisms are utilized by different cofactors to control specific gene expression programs, and that these mechanisms are not universally compatible with all promoters (Figure 2). Interestingly, the promoters that respond distinctly to the above-mentioned cofactors also differ in their sequences, containing, for example, TATA-box and initiator motifs versus CpG dinucleotides. These differences and the distinct regulatory compatibilities of different promoter types imply that they should also differ in the various regulatory or rate-limiting steps, the proteins they bind, and/or potentially involve distinct PICs. How regulatory compatibilities are implemented at the various promoter types and how enhancers, promoter and cofactors ultimately control transcriptional bursting and thus transcription rates are exciting open questions (Figure 2).

Regulatory compatibilities between cofactors and promoters suggest that the transcription of genes might differentially depend on different factors. Indeed, the past years saw several examples of differential cofactor requirements. For example, the cofactor Ldb1 together with the transcription factor Lhx2 is required for the transcription of olfactory receptor genes in olfactory epithelial cells [26]. Similarly, the steroid receptor coactivator-3 (SRC-3) is specifically utilized to activate the estradiol responsive gene GREB1 [27]. In both cases, the activity of the cofactors was necessary to not only maintain transcription but also spatial contacts between the promoters and their cognate enhancers. Gene-specific cofactor requirements are even seen at

Figure 2



Distinct types of enhancers and promoters communicate through different cofactors. Cofactors such as mediator and MLL3/4 preferentially activate TATA-box and CpG island promoter types, respectively, which also respond differently to different enhancer types. The relationships between promoters, enhancers, and cofactors (grey circle) with transcriptional burst dynamics remain to be explored.

the level of individual subunits of mediator [28]. Recent studies showed that depletion of the structural core mediator subunit MED14 resulted in a global perturbation of gene expression, while having no impact on the topological organization of the genome [29•,30•]. Interestingly, the depletion of other mediator subunits (specifically in the tail module) often resulted in viable cells and changes to the expression of specific gene groups only. These studies suggest subcomplexes of mediator may exist and may be differentially utilized, in agreement with earlier findings [28]. Overall, the observation of cofactor-promoter specificity indicates that not all cofactors are required to function at all enhancers or promoters, and distinct mechanisms of transcription activation may involve different cofactors.

Transcription factors

Transcription factors are the key link between regulatory DNA elements and cofactors and thus play crucial roles in gene regulation and cofactor specificities. This dual function is reflected in the prototypical structure of TFs that encompasses a DNA binding domain and a transactivation domain (tAD) that facilitates activation through cofactor binding and recruitment.

As regulatory elements are embedded in chromatin, TFs need to specifically identify their target binding sites which are often inaccessible. Profiling the ability of 220 TFs to bind naked or chromatinized DNA revealed that most TFs don't have access to chromatinized DNA [31•]. However, some TFs, for example, from the EN1 and SOX families, strongly bound nucleosomal DNA and classify together with others such as FoxA in mammals [32] or Zelda and Grainyhead in Drosophila [33,34] as *pioneer factors*. Recent structural studies now provide detailed insights into the ability of pioneer TFs to bind chromatinized DNA and enable the subsequent binding of other TFs. The binding of pioneer TFs such as Sox2 or Sox11 to nucleosomal DNA results in DNA distortions that destabilize the nucleosome, and are mediated by full or partial recognition of the TFs' DNA binding motif [35•,36•,37,38•]. Furthermore, binding of the pioneer TFs repositions the N-terminal tails of histone 4, which may be incompatible with higher order stacking of the nucleosomes, thereby opening the local chromatin structure.

The TFs' tADs are necessary and sufficient for transcription activation through their interactions with cofactors. The past years not only saw the high-throughput mapping and mutational dissection of these domains [39–41] but a potential role of tADs in determining TF binding *in vivo* seemingly independently of the TFs DNA-binding domains [42]. The latter example showed that the deletion of the Msn2 DNA binding domain did not perturb its genomic localization at a majority of binding sites. It is unknown whether the tAD achieves specific target recognition through sampling of DNA interactions or through association with other TFs that are bound at Msn2-target promoters, similar to what has been proposed for highly occupied target (HOT) regions [43].

A critical aspect of developmental gene regulation is the combinatorial function of TFs [44]. It is long known that TFs bind cooperatively and that, for example, SNPs or mutations within the binding site of one TF can influence the binding of another TF [45–47]. We have however been blind to the co-binding dynamics and the extent to which TFs are co-bound at the same time. Different strategies of DNA footprinting by DNA-modifying enzymes now enable the recording of simultaneous TF binding and of synchronous DNA-accessibility of neighboring enhancers [48•,49•,50•]. Interestingly, at both length-scales, neighboring TF binding sites and neighboring enhancers tend to be coordinated. In addition to the cooperative binding of co-expressed TFs, even TFs that are never present in the same cell can activate transcription in a combinatorial fashion by a temporal integration mechanism that involves priming and relay and enables the emergence of asymmetric neuron pairs in *C. elegans* [51•].

Liquid-liquid phase separation—mediating selective cofactor-TF interactions?

Our understanding of transcriptional regulation has been dominated by a sequential model according to which TFs bind specific sites on DNA and then, through discrete and specific protein-protein interactions, recruit cofactors within an otherwise homogeneous nuclear environment. Over the past years, this model has been complemented by an emerging biophysical property of many regulators to undergo liquid-liquid phase separation (LLPs) [52]. In this process, weak multivalent interactions, e.g., between modified histones and their reader-domains or between low-complexity and presumably disordered regions present in many transcriptional regulators [53], create local microenvironments in which some factors are concentrated while others are excluded. Indeed, HP1 and H3K9me3 drive heterochromatin formation by LLPS, which leads to chromatin compaction and the exclusion of transcription-activating factors from these domains [54–56]. Interestingly however, not all HP1 marked heterochromatin seems to adhere to an LLPS model, and heterochromatin compaction may be independent of HP1 [57]. LLPS may partition active and repressed genomic regions more generally: In *C. elegans*, H3K27-acetylated chromatin domains cluster together with the acetyl transferase p300/CBP-1 and the transcriptional activator ATF-8 [58]. These domains are spatially distinct from repressed domains containing H3K9-methylation and the chromatin anchor protein CEC-4. Removal of CEC-4 or the repressive TF Mrg1 resulted

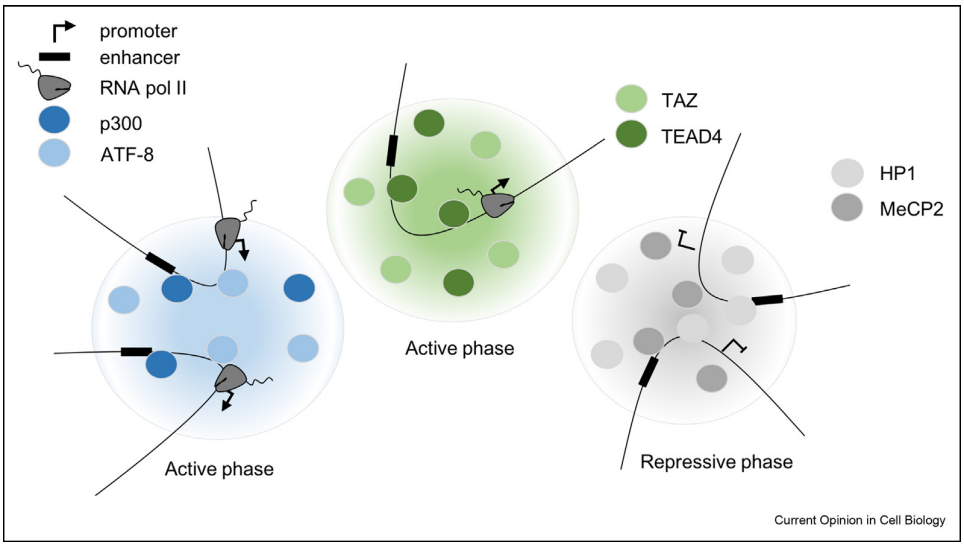
in p300/CBP-1 spreading into heterochromatin, suggesting the balance between repressors and activators is crucial for maintenance of active and repressed domains. In a more recent observation, MeCP2 has been observed to be enriched at HP1 marked LLPS heterochromatin regions [59]. MeCP2 mutations present in Rett syndrome patients correlate with the reduced ability of MeCP2 to phase separate, suggesting that LLPS could play a role in disease pathology [60]. Importantly, not all observable sites of transcription exist in LLPS condensates, and it will be important to systematically compare them with regions where transcription in LLPS occurs [61–65].

Interestingly, the presence of acetylated lysins on Histone 3 can contribute to the LLPS capacity of chromatin. Acetylation of nucleosomes on lysine residues, which reduces the net positive charge, leads to dissolution of chromatin droplets [66•]. However, in the presence of the activator Brd4, acetylated chromatin can phase separate into droplets which exclude regions not bound by Brd4 *in vitro*. It is likely that additional chemical modifications on chromatin such as ubiquitination, methylation, phosphorylation may contribute to the phase separation capacity of chromatin and the portioning of distinct phases facilitating specific protein-protein interactions [67].

LLPS has been proposed to apply to various proteins involved in transcription, including Pol II [68], mediator (in particular Med1 [69]), and TFs [70]. Low-complexity domains predicted to be intrinsically disordered (thus also called intrinsically disordered regions or *IDRs*) seem especially likely to play a prominent role in LLPS. In the case of TFs, for example, mutations that change the length of IDRs can influence the TFs' activating functions and their propensity to phase separate [71,72•]. Once formed, such condensates can sequester proteins and small molecules: mediator condensates have been shown to sequester signaling molecules such as STAT3 and SMAD3 [73], and droplets formed by the TF TAZ concentrated the DNA binding cofactor TEAD4 and other cofactors such as Brd4, p300, and Cdk9 [74]. The concentration of specific activators is not only limited to proteins, but also includes small molecules: Med1 LLPs droplets, for example, concentrated chemotherapy agents such as cisplatin [75].

It will be interesting to learn how many distinct types of nuclear environments can be formed and which molecular rules govern their formation (Figure 3). We can envision, for example, that distinct nuclear functions might each create their own type of nuclear microenvironment, including transcription repression and activation, which might potentially even create two distinct activation modes according to the developmental and housekeeping transcriptional programs, respectively [29,30,76], DNA damage, mitosis, or others. While phase-transition of heterochromatin leads to rather large domains, they seem to be of much smaller size in the case of activation, thus called microenvironments or hubs, and it remains to be seen by novel approaches in

Figure 3



TFs and cofactors participate in specific protein-protein interactions within LLPS droplets that can form transcriptionally repressive (grey) or active phases (blue and green).

electron microscopy or high-resolution optical imaging how many molecules are involved and how their intermolecular interactions are best described [77–79].

Conclusion

Gene regulation is a multifaceted process in which non-coding regulatory elements communicate with each other through various protein factors. The past years have seen tremendous progress in our understanding of how genome topology organizes regulatory domains and how enhancers and regulatory proteins further refine regulatory specificity by distinct promoter compatibilities. It will be exciting to witness further progress in understanding how distinct promoter types convert activating cues into transcriptional activity and how these cues modulate burst frequencies or sizes or both to enhance transcription.

We also expect exciting new insights into the molecular rules by which LLPS can divide the nucleus into microdomains with distinct biophysical and regulatory properties, selectively enriching some proteins and their functions while excluding others. In addition, emerging high-resolution imaging technologies will provide unprecedented insights into the molecular makeup, mechanisms, and dynamics of such LLPS at active or repressed gene loci.

Conflict of interest statement

Nothing declared.

Acknowledgments

The authors thank Clemens Plaschka (IMP) and Christa Bücker (MPL) for helpful comments on the manuscript. L.S. is supported by a DOC PhD Fellowship from the Austrian Academy of Sciences (ÖAW). 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, P29613–B28 and P33157–B). Basic research at the Institute of Molecular Pathology (IMP), Vienna, Austria, is supported by Boehringer Ingelheim GmbH and the Austrian Research Promotion Agency (FFG).

References

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest

1. Rodriguez J, Larson DR: **Transcription in living cells: molecular mechanisms of bursting**. *Annu Rev Biochem* 2020, **89**: 189–212.

2. Chubb JR, Trcek T, Shenoy SM, Singer RH: **Transcriptional pulsing of a developmental gene**. *Curr Biol* 2006, **16**: 1018–1025.

3. Rodriguez J, Ren G, Day CR, Zhao K, Chow CC, Larson DR: **Intrinsic dynamics of a human gene reveal the basis of expression heterogeneity**. *Cell* 2019, **176**:213–226.

4. Ko MS, Nakauchi H, Takahashi N: **The dose dependence of glucocorticoid-inducible**. *EMBOJ* 1990, **9**:2835–2842.

5. Donovan BT, Huynh A, Ball DA, Patel HP, Poirier MG, Larson DR, Ferguson ML, Lenstra TL: **Live-cell imaging reveals the interplay between transcription factors, nucleosomes, and bursting**. *EMBO J* 2019, **38**. e51794–18.

6. Stavreva DA, Garcia DA, Fettweis G, Gudla PR, Zaki GF, Soni V, McGowan A, Williams G, Huynh A, Palangat M, *et al.*: **Transcriptional bursting and Co-bursting regulation by steroid hormone release pattern and transcription factor mobility**. *Mol Cell* 2019, **75**:1161–1177.

7. Bartman CR, Hamagami N, Keller CA, Giardine B, Hardison RC, Blobel GA, Raj A: **Transcriptional burst initiation and polymerase pause release are key control points of transcriptional regulation**. *Mol Cell* 2019, **73**:519–532.

8. • Larsson AJM, Johnsson P, Hagemann-Jensen P, Hartmanis L, Faridani OR, Renius B, Segerstolpe A, Rivera CM, Ren B, Sandberg R: **Genomic encoding of transcriptional burst kinetics**. *Nature* 2019, **565**:251–254.
This study estimates bursting parameters genome wide using single cell RNA-seq data, confirming that enhancers control burst frequency while promoters control burst size with TATA-box promoters exhibiting the largest burst sizes.

9. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, *et al.*: **Spatial partitioning of the regulatory landscape of the X-inactivation centre**. *Nature* 2012, **485**:381–385.

10. Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G: **Three-dimensional folding and functional organization principles of the Drosophila genome**. *Cell* 2012, **148**:458–472.

11. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B: **Topological domains in mammalian genomes identified by analysis of chromatin interactions**. *Nature* 2012, **485**: 376–380.

12. de Laat W, Duboule D: **Topology of mammalian developmental enhancers and their regulatory landscapes**. *Nature* 2013, **502**: 499–506.

13. Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, *et al.*: **Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions**. *Cell* 2015, **161**: 1012–1025.

14. Laugsch M, Bartusel M, Rehimi R, Alirzayeva H, Karaolidou A, Crispatzu G, Zentis P, Nikolic M, Bleckwehl T, Kolovos P, *et al.*: **Modeling the pathological long-range regulatory effects of human structural variation with patient- specific hiPSCs**. *Stem Cell* 2019, **24**:736–752.

15. Kraft K, Magg A, Heinrich V, Riemenschneider C, Schöpflin R, Markowski J, Ibrahim DM, Acuna-Hidalgo R, Despang A, Andrey G, *et al.*: **Serial genomic inversions induce tissue-specific architectural stripes, gene misexpression and congenital malformations**. *Nat Cell Biol* 2019, **21**:305–310.

16. Despang A, Schöpflin R, Franke M, Ali S, Jerković I, Paliou C, Chan W-L, Timmermann B, Wittler L, Vingron M, *et al.*: **Functional dissection of the Sox9–Kcnj2 locus identifies nonessential and instructive roles of TAD architecture**. *Nat Genet* 2019, **51**:1263–1271.
Studies 16 & 17 manipulated the TAD structures of important developmental genes by deleting and inverting sequences at the TAD boundaries. While TAD structures were severely affected, no major changes in gene expression were observed. The authors suggest that developmentally regulated genes are robust to perturbations of their topological environment and that the acquisition of novel contacts might be required for phenotypic change.

17. • Williamson I, Kane L, Devenney PS, Flyamer IM, Anderson E, Kilanowski F, Hill RE, Bickmore WA, Lettice LA: **Developmentally regulated Shh expression is robust to TAD perturbations**. *Development* 2019, **146**:dev179523.
Studies 16 & 17 manipulated the TAD structures of important developmental genes by deleting and inverting sequences at the TAD boundaries. While TAD structures were severely affected, no major changes in gene expression were observed. The authors suggest that developmentally regulated genes are robust to perturbations of their topological environment and that the acquisition of novel contacts might be required for phenotypic change.

18. Rodríguez-Carballo E, Lopez-Delisle L, Yakushiji-Kaminatsui N, Ullate-Agote A, Duboule D: **Impact of genome architecture on the functional activation and repression of Hox regulatory landscapes**. *BMC Biol* 2019, **17**:55.

19. • Ghavi-Helm Y, Jankowski A, Meiers S, Viales RR, Korbel JO, Furlong EEM: **Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression**. *Nat Genet* 2019, **51**:1272–1282.
This study used highly rearranged Drosophila balancer chromosomes to investigate the effect of structural variants on gene expression. The authors report limited changes in gene expression despite major changes in linear DNA sequence and chromosome topologies, presumably also because the creation of balancer chromosomes selected against major changes in the expression of important genes.

20. Rao SSP, Huang S-C, Hilaire BGS, Engreitz JM, Perez EM, Kieffer-Kwon K-R, Sanborn AL, Johnstone SE, Bascom GD, Bochkov ID, *et al.*: **Cohesin loss eliminates all loop domains**. *Cell* 2017, **171**:305–309.

21. Nora EP, Goloborodko A, Valton A-L, Gibcus JH, Uebersohn A, Abdennur N, Dekker J, Mirny LA, Bruneau BG: **Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization**. *Cell* 2017, **169**:930–944.

22. Wutz G, Ladurner R, St Hilaire BG, Stocsits RR, Nagasaka K, Pignard B, Sanborn A, Tang W, Várnai C, Ivanov MP, *et al.*: **ESCO1 and CTCF enable formation of long chromatin loops by protecting cohesinSTAG1 from WAPL**. *eLife* 2020, **9**:9906.

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

24. Arnold CD, Zabidi MA, Pagani M, Rath M, Schernhuber K, Kazmar T, Stark A: **Genome-wide assessment of sequence-intrinsic enhancer responsiveness at single-base-pair resolution**. *Nat Biotechnol* 2017, **35**:136–144.

25. • Haberle V, Arnold CD, Pagani M, Rath M, Schernhuber K, Stark A: **Transcriptional cofactors display specificity for distinct types of core promoters**. *Nature* 2019, **570**:122–126.
This study shows specificity in the ability of cofactors to activate transcription from different promoters, which correlated with the DNA sequence of the promoters and their endogenous chromatin environment.

26. Monahan K, Horta A, Lomvardas S: **LHX2- and LDB1-mediated trans interactions regulate olfactory receptor choice**. *Nature* 2019, **565**:448–453.

27. Panigrahi AK, Foulds CE, Lanz RB, Hamilton RA, Yi P, Lonard DM, Tsai M-J, Tsai SY, O'Malley BW: **SRC-3 coactivator governs dynamic estrogen-induced chromatin looping interactions during transcription**. *Mol Cell* 2018, **70**:679–694.

28. Allen BL, Taatjes DJ: **The Mediator complex: a central integrator of transcription**. *Nat Rev Mol Cell Biol* 2015, **16**:155–166.

29. • Khattabi El L, Zhao H, Kalchschmidt J, Young N, Jung S, Van Blerkom P, Kieffer-Kwon P, Kieffer-Kwon K-R, Park S, Wang X, *et al.*: **A pliable mediator acts as a functional rather than an architectural bridge between promoters and enhancers**. *Cell* 2019, **178**:1145–1158.
This study provided structural and genetic dissection of mediator complex subunits, showing the essentially of some but not all subunits. Depletion of mediator does not result in perturbation of genome structure, suggesting a functional rather than a structural role for mediator in mediating enhancer-promoter communication.

30. • Jaeger MG, Schwalb B, Mackowiak SD, Velychko T, Hanzl A, Imrichova H, Brand M, Agerer B, Chorn S, Nabet B, *et al.*: **Selective Mediator dependence of cell-type-specifying transcription**. *Nat Genet* 2020, **52**:719–727.
This study investigated mediator subunit dependence for cell-type specific transcription programs. Genes whose transcription was unaffected by mediator depletion relied on a Cdk9 mediated feedback loop that increased pol II pause-release rates captured by kinetic measurements.

31. • Zhu F, Farnung L, Kaasinen E, Sahu B, Yin Y, Wei B, Dodonova SO, Nitta KR, Morgunova E, Taipale M, *et al.*: **The interaction landscape between transcription factors and the nucleosome**. *Nature* 2018, **562**:76–81.
This study profiled the binding of 220 TFs to nucleosome bound DNA. The authors found TFs that preferentially bind nucleosome DNA over naked DNA. TFs preferred specific positions along the nucleosome, and some exhibited directionality in their binding.

32. Iwafuchi-Doi M, Donahue G, Kakumanu A, Watts JA, Mahony S, Pugh BF, Lee D, Kaestner KH, Zaret KS: **The pioneer transcription factor FoxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation**. *Mol Cell* 2016, **62**:79–91.

33. Liang H-L, Nien C-Y, Liu H-Y, Metzstein MM, Kirov N, Rushlow C: **The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila**. *Nature* 2008, **456**:400–403.

34. Jacobs J, Atkins M, Davie K, Imrichova H, Romanelli L, Christiaens V, Hulselmans G, Potier D, Wouters J, Taskiran II , *et al.*: **The transcription factor Grainy head primes epithelial enhancers for spatiotemporal activation by displacing nucleosomes**. *Nat Genet* 2018, **50**:1011–1020.

35. • Michael AK, Grand RS, Isbel L, Cavadini S, Kozicka Z, Kempf G, Bunker RD, Schenk AD, Graff-Meyer A, Pathare GR, *et al.*: **Mechanisms of OCT4-SOX2 motif readout on nucleosomes**. *Science* 2020, **368**:1460–1465.
Using a structure of SOX2-OCT4 bound to a nucleosome, this study revealed that OCT4 can unwind DNA from the nucleosome at two different positions on the nucleosome relative to its DNA binding motif, and that partial recognition of the motif is sufficient to unwind the DNA.

36. • Dodonova SO, Zhu F, Dienemann C, Taipale J, Cramer P: **Nucleosome-bound SOX2 and SOX11 structures elucidate pioneer factor function**. *Nature* 2020, **580**:669–672.
This study investigated SOX2 and SOX11 binding to DNA and found that they can locally distort DNA. Binding of SOX-factors repositioned the N-terminal tails of histone 4, which may not be compatible with higher order nucleosome stacking in which tails of histone 4 subunits contact each other.

37. Mivelaz M, Cao A-M, Kubik S, Zencir S, Hovius R, Boichenko I, Stachowicz AM, Kurat CF, Shore D, Fierz B: **Chromatin fiber invasion and nucleosome displacement by the Rap1 transcription factor**. *Mol Cell* 2020, **77**:488–499.

38. • Garcia MF, Moore CD, Schulz KN, Alberto O, Donague G, Harrison MM, Zhu H, Zaret KS: **Structural features of transcription factors associating with nucleosome binding**. *Mol Cell* 2019, **75**:921–932.
This study profiled the binding of pioneer transcription factors to nucleosomal DNA and uncovered known and new factors exhibiting binding with nanomolar affinity. The strongest binding factors contained a short alpha helix with which they recognized nucleosomal DNA.

39. Staller MV, Holehouse AS, Swain-Lenz D, Das RK, Pappu RV, Cohen BA: **A high-throughput mutational scan of an intrinsically disordered acidic transcriptional activation domain**. *Cell Systems* 2018, **6**:444–455.

40. Erijman A, Kozlowski L, Sohrabi-Jahromi S, Fishburn J, Warfield L, Schreiber J, Noble WS, Söding J, Hahn S: **A high-throughput screen for transcription activation domains reveals their sequence features and permits prediction by deep learning**. *Mol Cell* 2020, **78**:890–902.

41. Arnold CD, Nemčko F, Woodfin AR, Wienerroither S, Vlasova A, Schleiffer A, Pagani M, Rath M, Stark A: **A high-throughput method to identify trans-activation domains within transcription factor sequences**. *EMBO J* 2018, **37**:5–13.

42. Brodsky S, Jana T, Mittelman K, Chapal M, Kumar DK, Carmi M, Barkai N: **Intrinsically disordered regions direct transcription factor in vivo binding specificity**. *Mol Cell* 2020, **79**:459–471.

43. Kvon EZ, Stampfel G, Yanez-Cuna JO, Dickson BJ, Stark A: **HOT regions function as patterned developmental enhancers and have a distinct cis-regulatory signature**. *Genes & Development* 2012, **26**:908–913.

44. Reiter F, Wienerroither S, Stark A: **ScienceDirect Combinatorial function of transcription factors and cofactors**. *Curr Opin Genet Dev* 2017, **43**:73–81.

45. Bardet AF, He Q, Zeitlinger J, Stark A: **A computational pipeline for comparative ChIP-seq analyses**. *Nat Protoc* 2012, **7**:45–61.

46. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK: **Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities**. *Mol Cell* 2010, **38**:576–589.

47. Deplancke B, Alpern D, Gardeux V: **The genetics of transcription factor DNA binding variation**. *Cell* 2016, **166**:538–554.

48. • Stergachis AB, Debo BM, Haugen E, Churchman LS, Stamatoyannopoulos JA: **Single-molecule regulatory architectures captured by chromatin fiber sequencing**. *Science* 2020, **386**:1449–1454.
This study developed a footprinting method to profile the occupancy of nucleosomes and transcription factors on a single DNA molecule based on long-read DNA sequencing. With this method it is now possible to distinguish which sites are occupied and co-bound on a single regulatory element, and further investigate which neighboring regulatory elements are co-bound.

49. • Sönmezer C, Kleinendorst R, Imanci D, Villacorta L, Schübeler D, Benes V, Krebs AR: **Single molecule occupancy patterns of transcription factors reveal determinants of cooperative binding in vivo**. *bioRxiv* 2020, **6**, https://doi.org/10.1101/2020.06.29.167155.
This study developed a footprinting method to profile TF binding on single DNA molecule based on short-read sequencing. Applied to mouse regulatory elements, the authors show that the binding of TFs to neighboring sites is correlated and that co-bound TFs can be present in the same enhancer without strict spacing between their motifs.

50. • Abdulhay NJ, McNally CP, Hsieh LJ, Kasinathan S, Keith A, Estes LS, Karimzadeh M, Underwood JG, Goodarzi H, Narlikar GJ, Ramani V: **Massively multiplex single-molecule oligonucleosome footprinting**. *bioRxiv* 2020, vol. 5, https://doi.org/10.1101/2020.05.20105379.
This study developed a footprinting assay to measure nucleosome positions on individual chromatin fibers. The authors capture DNA-nucleosome interactions in heterogeneous states across different epigenomic domains suggesting a complex range of oligonucleosome patterns.

51. • Charest J, Daniele T, Wang J, Bykov A, Mandlbauer A, Asparuhova M, Roehsner J, Gutierrez-Perez P, Cochella L: **Combinatorial action of temporally segregated transcription factors**. *Dev Cell* 2020, **55**:1–17.
This study shows that the TFs TBX-37/38 and CHE-1 function cooperatively despite not being present at the same time and describes a "priming-and-relay" model. Priming by TBX-37/38 and relay by common regulators involving transcription is sufficient to subsequently enable CHE-1 to activate transcription in a cell-type specific manner giving rise to the left ASE neuron.

52. Clarke J, Pappu RV: **Editorial overview: protein folding and binding, complexity comes of age**. *Curr Opin Struct Biol* 2017, **42** [:v–vii.

53. Banai SF, Lee HO, Hyman AA, Rosen MK: **Biomolecular condensates: organizers of cellular biochemistry**. *Nat Rev Mol Cell Biol* 2020, **18**:285–298.

54. Larson AG, Elnatan D, Keenen MM, Trnka MJ, Johnston JB, Burlingame AL, Agard DA, Redding S, Narlikar GJ: **Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin**. *Nature* 2017, **547**:236–240.

55. Strom AR, Emelyanov AV, Mir M, Fyodorov DV, Darzacq X, Karpen GH: **Phase separation drives heterochromatin domain formation**. *Nature* 2017, **547**:241–245.

56. Klosin A, Hyman AA: **A liquid reservoir for silent chromatin**. *Nature* 2017, **547**:168–169.

57. • Erdel F, Rademacher A, Vlijm R, Tunnermann J, Frank L, Weinmann R, Schweigert E, Yserentant K, Hummert J, Bauer C, *et al.*: **Mouse heterochromatin adopts digital compaction states without showing hallmarks of HP1-driven liquid-liquid phase separation**. *Mol Cell* 2020, **78**:236–249.
This study tests predictions of the LLPS model versus a *collapsed polymer globules* model for mouse heterochromatin and assesses the dependency of heterochromatin compaction in cells on HP1.

58. Cabianca DS, Muñoz-Jiménez C, Kalck V, Gaidatzis D, Padeken J, Seeber A, Askjaer P, Gasser SM: **Active chromatin marks drive spatial sequestration of heterochromatin in C. elegans nuclei**. *Nature* 2019, **569**:734–739.

59. Li CH, Coffey EL, Dall'Agnese A, Hannett NM, Tang X, Henninger JE, Platt JM, Oksuz O, *et al.*: **MeCP2 links heterochromatin condensates and neurodevelopmental disease**. *Nature* 2020.

60. Liang W, Hu M, Zuo MQ, Zhao J, Wu W, Huang L, Wen Y, Li Y, Chen P, Bao X, *et al.*: **Rett syndrome-causing mutations compromise MeCP2-mediated liquid-liquid phase separation of chromatin**. *Cell Res* 2020, **30**:393–407.

61. Chong S, Dugast-Darzacq C, Liu Z, Dong P, Dailey GM, Cattoglio C, Heckert A, Banala S, Lavis L, *et al.*: **Imaging dynamic and selective low-complexity domain interactions that control gene transcription**. *Science* 2018, **361**:eaar2555.

62. McSwiggen DT, Hansen AS, Teves SS, Marie-Nelly H, Hao Y, Heckert AB, Umemoto KK, Dugast-Darzacq C, Tjian R, Darzacq X: **Evidence for DNA-mediated nuclear compartmentalization distinct from phase separation**. *eLife* 2019, **8**, e47098.

63. McSwiggen DT, Mir M, Darzacq X, Tjian R: **Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences**. *Genes Dev* 2019, **33**:1619–1634.

64. Mir M, Stadler MR, Ortiz SA, Hannon CE, Harrison MM, Darzacq X, Eisen MB: **Dynamic multifactor hubs interact transiently with sites of active transcription in *Drosophila* embryos**. *eLife* 2018, **7**, e40497.

65. Quintero-Cadena P, Lenstra TL, Sternberg PW: **RNA pol II length and disorder enable cooperative scaling of transcriptional bursting**. *Mol Cell* 2020, **79**:207–220.

66. • Gibson BA, Doolittle LK, Schneider MWG, Jensen LE, Gamarra N, Henry L, Gerlich DW, Redding S, Rosen MK: **Organization of chromatin by intrinsic and regulated phase separation**. *Cell* 2019, **179**:470–484.
This study shows that chromatin undergoes phase-separation. Parameters such as DNA length and histone acetylation modulate the ability of chromatin to phase-separate. Chromatin acetylation dissolves phase-separated droplets, however acetylated chromatin can still undergo phase-separation upon binding of factors such as Brd4.

67. Gallego LD, Schneider M, Mittal C, Romanauska A, Carrillo RMG, Schubert T, Pugh F, Koehler A: **Phase separation directs ubiquitination of gene-body nucleosomes**. *Nature* 2020, **579**:592–597.

68. Boehning M, Dugast-Darzacq C, Rankovic M, Hansen AS, Yu T, Marie-Nelly H, McSwiggen DT, Kokic G, Dailey GM, Cramer P, *et al.*: **RNA polymerase II clustering through carboxy- terminal domain phase separation**. *Nat Struct Mol Biol* 2018, **25**:833–840.

69. Sabari BR, Dall'Agnese A, Boija A, Klein IA, Coffey EL, Shrinivas K, Abraham BJ, Hannett NM, Zamudio AV, Manteiga JC, *et al.*: **Coactivator condensation at super-enhancers links phase separation and gene control**. *Science* 2018, **361**. eaar3958–13.

70. Boija A, Klein IA, Sabari BR, Dall'Agnese A, Coffey EL, Zamudio AV, Li CH, Shrinivas K, Manteiga JC, Hannett NM, *et al.*: **Transcription factors activate genes through the phase-separation capacity of their activation domains**. *Cell* 2018, **175**:1842–1855.

71. Gemayel R, Chavali S, Pougach K, Legendre M, Zhu B, Boeynaems S, van der Zande E, Gevaert K, Rousseau F, Schymkowitz J, *et al.*: **Variable glutamine-rich repeats modulate transcription factor Activity**. *Mol Cell* 2015, **59**:615–627.

72. • Basu S, Mackowiak SD, Niskanen H, Knezevic D, Asimi V, Grosswendt S, Geertsema H, Ali S, Jerković I, Ewers H, *et al.*: **Unblending of transcriptional condensates in human repeat expansion disease**. *Cell* 2020, **181**:1062–1079.
In this study the authors show that a disease associated repeat expansion in the Hoxd13 transcription factor alters its phase separation capacity. Furthermore, the authors indicate that the repeat expansion alters the ability of Hoxd13 to sequester the cofactor mediator, correlating with perturbed gene expression in several different cell types.

73. Zamudio AV, Dall'Agnese A, Henninger JE, Manteiga JC, Afeyan LK, Hannett NM, Coffey EL, Li CH, Oksuz O, Sabari BR, *et al.*: **Mediator condensates localize signaling factors to key cell identity genes**. *Mol Cell* 2019, **76**:753–766.

74. Lu Y, Wu T, Gutman O, Lu H, Zhou Q, Henis YI, Luo K: **Phase separation of TAZ compartmentalizes the transcription machinery to promote gene expression**. *Nat Cell Biol* 2020, **22**. 453-453.

75. Klein IA, Boija A, Afeyan LK, Hawken SW, Fan M, Dall'Agnese A, Oksuz O, Henninger JE, *et al.*: **Partitioning of cancer therapeutics in nuclear condensates**. *Science* 2020, **368**: 1386–1392.

76. Haberle V, Stark A: **Eukaryotic core promoters and the functional basis of transcription initiation**. *Nat Rev Mol Cell Biol* 2018, **19**:621–637.

77. Li J, Dong A, Saydaminova K, Chang H, Wang G, Ochiai H, Yamamoto T, Pertsinidis A: **Single-molecular nanoscopy elucidates RNA polymerase II transcription at single genes in live cells**. *Cell* 2019, **178**:491–506.

78. Su JH, Zheng P, Kinrot SS, Bintu B, Zhuang X: **Genome-scale imaging of the 3D organization and transcriptional activity of chromatin**. *Cell* 2020, **182**:1641–1659.

79. Li J, Hsu A, Hua Y, Wang G, Cheng L, Ochiai H, Yamamoto T, Pertsinidis A: **Single-gene imaging links genome topology, promoter-enhancer communication and transcription control**. *Nature* 2020:3136.

Publication #2: “Functionally distinct promoter classes initiate transcription via different mechanisms”

Authors: Leonid Serebreni, Lisa-Marie Pleyer, Vanja Haberle, Oliver Hendy, Anna Vlasova, Vincent Loubiere, Filip Nemcko, Katharina Bergauer, Elisabeth Roitinger, Karl Mechtler, Alexander Stark. [In submission]

Discussion and summary of the publication:

In this work I present the results from DNA-affinity purification of functionally distinct promoters and the functional characterization of various proteins’ requirement for transcription. The main observation of the DNA-affinity purifications informed us that *in-vitro* the Pol II PIC preferentially assembled on developmental promoter types but not housekeeping. While PIC components such as TBP were required at only a subset of promoters, TFIIA was required for transcription by all promoters. These observations prompted us to propose that housekeeping promoters which are TBP independent, but TFIIA dependent may recruit the PIC through a different mechanism. By identifying TFIIA and housekeeping promoter DNA interactors, we found that housekeeping cofactors such as GFZF, Chromator and Putzig can transactivate housekeeping promoters independently of TBP or TRF2. Forced recruitment of housekeeping cofactors to promoter DNA recapitulates broad transcription initiation patterns observed at these sites. These sets of experiments indicates that unlike developmental promoters that recruit the PIC through defined DNA-protein interactions, housekeeping promoters utilize transcriptional cofactors to recruit the PIC into a nucleosome-free region where transcription initiates in a broad pattern. Direct implication of these observations are attributable to mammalian promoters, which are largely exhibiting broad transcription initiation patterns and may recruit the transcription machinery similarly to Drosophila housekeeping promoters

Author contribution: AS and LS conceived and designed the experiments and wrote the manuscript. LS, KB and OH generated endogenously tagged AID cell lines and performed PRO-seq. LP and AV performed analysis of PRO-seq and ChIP-seq data. FN & LS performed TFIIA immunoprecipitation for mass spectrometry. LS performed DNA affinity purifications. KM & ER performed mass spectrometry. AV & VL analyzed mass spectrometry data. VH performed STAP-seq data analysis.

Functionally distinct promoter classes initiate transcription via different mechanisms

Leonid Serebreni[1], Lisa-Marie Pleyer[1], Vanja Haberle[1], Oliver Hendy[1], Anna Vlasova[1], Vincent Loubiere[1], Filip Nemcko[1], Katharina Bergauer[1], Elisabeth Roitinger[1,2], Karl Mechtler[1,2], Alexander Stark[1,3]

[1]Research Institute of Molecular Pathology (IMP), Vienna BioCenter (VBC), Campus-Vienna-Biocenter 1, Vienna, Austria. [2]Institute of Molecular Biotechnology (IMBA), Vienna BioCenter (VBC), Dr. Bohr-Gasse 3, Vienna, Austria. [3]Medical University of Vienna, Vienna BioCenter (VBC), Vienna, Austria.

Correspondence should be addressed to A.S. (stark@starklab.org)

Abstract

Recruitment of RNA polymerase II (Pol II) to promoter regions is essential for transcription. The Pol II Pre-Initiation Complex (PIC) is generally thought to assemble at all promoters via an identical universal mechanism. Unexpectedly, we discover that promoter classes with distinct functions and initiation patterns assemble the PIC differently: developmental promoter DNA readily associates with the canonical Pol II PIC, whereas housekeeping promoters do not. Consistently, the TATA binding protein is not required for housekeeping promoters, which relies on different DNA-binding factors such as DREF. TFIIA is required for transcription from all promoters, and we identify factors that can recruit TFIIA to housekeeping promoters and activate transcription. We show that promoter activation by these factors is sufficient to induce the dispersed transcription initiation patterns characteristic of housekeeping promoters. Thus, different promoter classes direct distinct mechanisms of PIC assembly.

Introduction

Transcription of protein-coding genes by RNA polymerase II (Pol II) is a highly regulated process orchestrated by non-coding regulatory elements, namely enhancers and promoters. Pol II recruitment at promoters leads to transcription initiation from the core-promoter region, a roughly 100 base-pair region around the transcription start site (TSS) at the 5'end of protein-coding genes (Butler and Kadonaga, 2002).

Recruitment of general transcription factors (GTFs) and assembly of the Pol II preinitiation complex (PIC) at promoters is a central regulatory step in transcription and has

been shown to occur hierarchically: the TATA-binding protein (TBP) within TFIID binds to the TATA-box in promoter DNA and recruits TFIIA, followed by the remaining GTFs TFIIB, TFIIE, TFIIF, and TFIIH (Cosma, 2002; Orphanides et al., 1996) and Pol II. TFIIA cooperates with TFIID to commit PIC assembly into an active state on promoters *in vitro* (Buratowski et al., 1989; Papai et al., 2010; Warfield et al., 2017).

Current models assume that this highly ordered and stereotypical assembly of the PIC occurs at all promoters, despite differences in sequence, transcription initiation patterns, chromatin structures, and abilities to respond to regulatory input (Haberle and Stark, 2018; He et al., 2013; Lenhard et al., 2012; Rach et al., 2011). In *Drosophila melanogaster* for example, the promoters of developmentally regulated genes contain TATA-boxes, downstream promoter elements (DPE), and/or Initiator (INR) motifs, and display focused initiation at a single, dominant TSS (Carninci et al., 2006; Ohler et al., 2002; Vo Ngoc et al., 2017; 2020). In contrast, housekeeping promoters contain TCT, DRE and Ohler1/6 motifs and show dispersed initiation at multiple TSSs (Figure 1A). Moreover, both types of promoters exhibit distinctive regulatory properties, respond differently towards activating cues (Arnold et al., 2016; Zabidi et al., 2015) and are activated by distinct sets of coactivators (Haberle et al., 2019). These distinct regulatory functions of promoters seem incompatible with a single mechanism of PIC assembly and transcription initiation, but the nature of the PIC and PIC assembly at different promoter types and whether they relate to these promoters' distinct functions is unknown.

Some evidence indeed suggests that different promoters utilize different PIC components. For example, some promoters require only a subset of GTFs *in vitro* (Parvin et al., 1994; 1992), which is in-line with the existence of different stable intermediates or alternative arrangements of the PIC on promoter DNA (Buratowski et al., 1989; Gazdag et al., 2016; Martianov et al., 2002; Murakami et al., 2013; Wieczorek et al., 1998; Yudkovsky et al., 2000).

Further, promoter-bound multi-subunit protein complexes that are part of the PIC, such as TFIID, can exhibit different arrangements. For instance, the Taf9 subunit of TFIID regulates cell-type-specific genes in Drosophila neural stem cells (Neves & Eisenman, Open Biol 2019) in specifying brown adipose cell identity (Neves and Eisenman, 2019), whereas Taf3 activates specific genes and instructs the differentiation of myoblast into myotubes (Deato et al., 2008; Stijf-Bultsma et al., 2015).

In addition, GTF paralogs may regulate transcription in distinct cell types or at specific promoters (Akhtar and Veenstra, 2011; Duttke et al., 2014; Zehavi et al., 2015). TBP-related factors, such as TBP2 (also known as TRF3) in complex with TFIIA, has been implicated in regulating transcription in mouse germ cells (Martianov et al., 2016; Yu et al., 2020), whereas the *Drosophila* Trf2 (TBP-like 1 in mouse) has been suggested to regulate the transcription of ribosomal protein genes, histone H1, and DPE motif-containing promoters (Baumann and Gilmour, 2017; Isogai et al., 2007; Kedmi et al., 2020; Wang et al., 2014). This cumulative evidence suggests that different promoter-bound GTF assemblies may exist on different promoter types and/or in different cell types.

Here, we discovered that two broad classes of promoters in *Drosophila*, developmental and housekeeping promoters, recruit fundamentally different sets of proteins and activate transcription via distinct mechanisms. Although TFIIA is required for transcription at all Pol II promoters, it assembles differently at these two types of promoters, associating more strongly with developmental promoters. Our work suggests that the more indirect recruitment of TFIIA by housekeeping promoters leads to the widely observed dispersed initiation pattern at these promoters.

Results

Different types of core promoters bind distinct sets of proteins

Roughly 37% of promoters in the Drosophila genome can be classified as developmental (TATA+INR, DPE+INR, INR only), and 38% as housekeeping (Ohler1/6, DRE, TCT), based on previous work by others and us (Figure 1A) (Haberle and Stark, 2018; Ohler et al., 2002; Vo Ngoc et al., 2019). Given the distinct regulatory functions and DNA sequence compositions of these two types of promoters and their general chromatin accessibility, we hypothesized that the core-promoter DNA recruits distinct transcription-related proteins. To test whether the core-promoter DNA sequence alone is sufficient to instruct protein recruitment, we investigated the proteins that associate with short (121 bp) core promoter DNA fragments *in vitro*. Specifically, we used DNA-affinity purification coupled to quantitative label-free mass spectrometry to profile proteins that are recruited to distinct promoter types (Figure 1B, Baek et al., 2006; Johnson et al., 2004; Kadonaga and Tjian, 1986; Köcher et al., 2012; Lin and Carey, 2012).

We first examined TATA-box-containing developmental and DRE-containing housekeeping promoter subtypes. To identify proteins that bind different promoter sequences of the same subtype, we pooled 16-32 representative promoters per subtype, and used a pool of 18

non-promoter control DNA fragments as a negative control (Figure 1B). We coupled the fragments of each pool to streptavidin-coated beads, incubated the beads with S2 cell nuclear extract, washed and cross-linked associated proteins, and quantified the enriched proteins by label-free mass spectrometry (Figure 1B). We performed three replicate experiments per pool and detected between 30-35 thousand peptides each, which allowed the label-free quantification of 3,465 proteins in total across all samples. Using the three replicates, we found 1094 proteins significantly enriched at the TATA-box promoters over the control pool; and 98 proteins significantly enriched at the DRE promoters (enrichment p-value < 0.05; limma (Ritchie et al., 2015)).

As expected from previous biochemical and structural work (Geiger et al., 1996; Nikolov et al., 1995; Plaschka et al., 2015; Tan et al., 1996), the TATA-box containing promoters were enriched for the canonical Pol II PIC, including TBP, GTFs and TFIID, and most Mediator subunits (Figure 1C, and supplementary figure 1A), suggesting that TATA-box promoter DNA is sufficient to recruit these proteins *in vitro*.

Unexpectedly, the DRE-containing promoters did not enrich for any of the Pol II PIC subunits; indeed some Tafs and GTFs were even depleted when compared to control DNA. In contrast, the DRE promoters were enriched for the core-promoter-element binding factor DREF, BEAF-32 and Ibf1/2 among other proteins (Figure 1D). Directly plotting the enrichments at DRE versus TATA promoters confirmed the strong differential recruitment of GTFs and PIC components specifically to TATA promoters but not to DRE promoters (Figure 1E).

The inability of housekeeping core promoters to recruit known PIC components was unexpected, so we tested longer, 350, 450 and 1000 bp DNA fragments derived from DRE promoters (Figure 2A). These DNA fragments bound DREF (Supplementary Figure 2A) and were highly active in luciferase reporter assays (Supplementary figure 2B). In contrast to the TATA-box containing SCP1 promoter, the longer DRE promoters were still unable to bind and enrich the PIC component TFIIA (Figure 2B). The differential ability of short core-promoter DNA fragments to enrich for known GTFs and PIC components under identical conditions *in vitro* is intriguing and suggests that the promoters' functional differences might arise at the level of GTF recruitment and PIC assembly, presumably via distinct DNA-binding factors.

To determine if the results above generalize to other promoter subtypes, we extended our analysis to developmental promoters containing DPE or Inr motifs, and to housekeeping promoters containing TCT or Ohler 1/6 motifs. We found that developmental promoter subtypes enriched for 892 to 1093 proteins, whereas housekeeping promoter subtypes enriched

between 98 and 432 proteins (enrichment p-value<0.05) (Figure 2C, Supplementary Table 1). Developmental and housekeeping promoters enriched for different sets of proteins and housekeeping promoters enriched for much fewer proteins overall compared to developmental promoters (Figure 2C). These data reveal that short fragments corresponding to functionally distinct core promoters are sufficient to recruit distinct transcription-related proteins under identical conditions *in vitro*.

GTFs and PIC components were preferentially enriched at all three types of developmental promoters but were not or only weakly enriched at housekeeping promoters (Figure 2D). Similarly, multiple components of the Mediator and TFIID complexes were also preferentially enriched at developmental promoters, with TATA-box containing promoters showing the highest levels of binding (Figure 2D). In contrast, none of the housekeeping promoter subtypes exhibited preferential binding of GTFs, TFIID or Mediator subunits; instead, they enriched for various TFs that bind core-promoter elements and chromatin regulators. For example, DRE containing promoters exhibited the highest binding levels of DREF and BEAF-32, whereas Ohler 1/6 promoters exhibited the highest binding levels of the Motif 1 Binding protein (M1BP) and the cofactor GFZF (Figure 2D).

Re-analyzing published ChIP-seq and ChIP-nexus data from Drosophila cells or embryos confirmed the differential binding preferences at the respective promoter subtypes (Figure 2E) (Baumann and Gilmour, 2017; Liang et al., 2014; Shao and Zeitlinger, 2017). GTFs were generally more enriched on developmental promoters than housekeeping promoters (except for TFIIB, TFIIF that also bind to TCT promoters), whereas housekeeping promoters preferentially bound TFs according to their motif contents: M1BP showed the highest ChIP-seq signals at Ohler 1/6 promoters, and DREF and BEAF-32 showed highest signals at DRE promoters (Figure 2E & Supplementary Figure 2C).

Thus, the DNA sequence of developmental promoters form a close association with the PIC whereas housekeeping promoters instead closely associate sequence-specific TFs through their cognate DNA binding motifs both *in vitro* and *in vivo*. Additionally, the markedly lower number of proteins enriched at housekeeping promoters suggests that the DNA-protein interface is weaker, more indirect, and/or transient nature.

Proteins bound to core promoters in vitro are required for transcription in cells

To determine if the differential recruitment of promoter-associated factors *in vitro* reflects distinct functional requirements *in vivo*, we used the auxin inducible degron (AID) system to

deplete endogenously labeled proteins from *D. melanogaster* S2 cells and measured nascent transcription by PRO-seq (Fig. 3A; (Nishimura et al., 2009)).

We examined TBP and DREF first and observed the near complete degradation of both proteins three hours after auxin addition (Fig. 3B) and their complete depletion six hours after auxin addition (Supplementary Figure 3A). To ensure complete protein degradation and avoid potential secondary effects from prolonged protein depletion, we measured changes to Pol II nascent transcription six hours after auxin treatment.

We performed two biological replicates of PRO-seq that were highly similar (PCC>0.99 Supplementary Figure 3B), and revealed 243 downregulated genes after TBP depletion and 161 downregulated genes after DREF depletion (fold-change < 0 and FDR < 0.05; Figure 3C). Notably, only a single gene was shared between the two conditions, indicating that distinct sets of promoters require TBP and DREF (Figure 2D). Motif enrichment analysis of the downregulated promoters revealed a strong enrichment of the TATA-box in the TBP-dependent promoters, and of the DRE motif in the DREF-dependent promoters (Figure 2E), as expected. The differential dependency on TBP versus DREF is apparent at the TATA promoter upstream of *Glucose dehydrogenase* (*Gld*) and the DRE promoter upstream of *Fermitin 2* (*Fit2*) (Figure 2F), and generalizes to the promoters used for the DNA affinity purification experiments, and to all active TATA- versus DRE-containing promoters genome-wide (Figure 2G and Supplementary Figure 2C). These results show that a relatively small number of active promoters require TBP, and that these are specifically TATA-box containing promoters. Similarly, only a subset of promoters require DREF, which are different from the TBP-requiring promoters and specifically contain DRE motifs. Overall, these results imply that different promoter types differentially depend on the two core-promoter element binders and utilize distinct DNA-protein interfaces and/or interactors to recruit Pol II and initiate transcription.

We identified additional promoter-subtype specific binders that are also required for viability or development according to Flybase (Figure 4A and Supplementary Figure 4A). We generated five endogenously tagged AID cell lines, and three showed specific effects on transcription after protein depletion, whereas two (CG7372 and Ibf1+2) showed mild to no effects and were not pursued (Supplementary Figures 4C-E).

CG3847, an uncharacterized TF, was most strongly enriched on Ohler 1/6 and TCT promoters (Figure 4A), and rapid depletion of CG3847 from S2 cells revealed 87 genes that were downregulated and thus required this factor for transcription (Figure 4B). Motif enrichment analysis of the promoters of these genes revealed an enrichment for the Ohler 1 motif

(Figure 4B). Plotting the distribution of PRO-seq signal across the promoters used for DNA-affinity purification showed downregulation of Ohler1/6 and TCT promoters (log2 fold-change < 0 and FDR < 0.05, Figure 4E).

Chromator, a previously characterized cofactor, preferentially bound housekeeping TCT and Ohler 1/6 promoters, and to a lower extent DPE and INR promoters (Figure 4C and Supplementary Figure 4A) (Cubeñas-Potts et al., 2017; Stampfel et al., 2015). Rapid depletion of Chromator lead to the significant downregulation of 499 genes (Figure 4C), whose promoters were enriched for all housekeeping promoter motifs such as Ohler1/6 motifs, TCT, DRE and Ohler 7 (Figure 4C). Interestingly, Chromator depletion led to upregulation of DPE and INR promoters which it was weakly bound to (Supplementary Figure 4F). These observations indicate that a subset of housekeeping promoters require Chromator for transcription *in vivo*, which agrees with our previous observation that Chromator is sufficient to activate transcription from housekeeping promoters (Haberle and Stark, 2018).

CG14711 bound most strongly to DPE promoters, and rapid CG14711 depletion revealed few downregulated genes but rather 191 genes that were mildly but significantly upregulated (Figure 4D). The INR and DPE motifs were enriched in the promoters of these genes (Figure 4D), suggesting that CG14711 acts as a transcriptional repressor for a subset of DPE and INR-containing developmental promoters. Collectively, these examples illustrate that DNA affinity purification coupled to mass spectrometry identified core-promoter bound proteins that regulate transcription *in vivo.*

All promoters – including housekeeping promoters – functionally depend on TFIIA

Our observation of preferential PIC binding to developmental but not housekeeping promoters suggests that canonical GTFs are either not required for transcription initiation at housekeeping promoters, or that they are required but interact less directly and/or less stably. To discern between these possibilities, we endogenously tagged TFIIA-L, which together with TBP arrives first at promoters and nucleates PIC assembly (Ranish et al., 1999), TFIIE (α and β subunit), TFIIF (α and β subunit) and TFIIB were incompatible with tagging at either the N- or C-termini and could therefore not be assessed). Tagged TFIIA-L was efficiently depleted within one to two hours after auxin treatment, and its depletion resulted in cell death and loss of PRO-seq signal for all genes (Figure 5A-C and Supplementary 5A-B). This functional requirement of TFIIA at housekeeping promoters indicates that TFIIA is recruited to housekeeping promoters, presumably via a novel mechanism, independently of TBP.

Identification of potential TFIIA recruiters at housekeeping promoters

As housekeeping promoters depend on TFIIA for transcription in cells, but fail to enrich for TFIIA *in vitro*, we hypothesized that PIC assembly at housekeeping promoters is mediated by intermediary proteins and that the functional requirement of TIFFA might provide the means to identify these proteins (Figure 5D). We thus performed immunoprecipitation-mass spectrometry with the endogenously tagged TFIIA-L-AID-3xFLAG S2 cell line and the parental Tir1 expressing cell line as a control. We uncovered 300 TFIIA binding proteins, including all three components of the TFIIA complex and other TFIIA interactors, such as the Tbp paralog Trf2 (but not Tbp), members of the TFIID complex, and various GTFs, such as TFIIE (Supplementary Table 2).

To identify candidate intermediary proteins, we intersected the TFIIA binding proteins with the proteins enriched on housekeeping promoters *in vitro* (Figure 5D). Applying this strategy to developmental promoters as a positive control identified most known GTFs, thus validating the approach. We found 131 proteins that can associate with TFIIA and at least one housekeeping promoter subtype (Figure 5D), including DREF, Chromator, GFZF, Putzig, the nucleolar protein Nnp1, and the RNA helicase CG8611 (Supplementary Table 2).

We predicted that a protein that recruits TFIIA to housekeeping promoters would activate transcription from a housekeeping promoter. To test this prediction in S2 cells, we fused 28 candidates to the Gal4 DNA-binding domain and tethered them to a UAS upstream of a minimal housekeeping core promoter driving luciferase (Figure 5E). We found that nine proteins were able to transactivate the housekeeping promoter (fold change>4 & $p<0.05$), particularly the coactivators GFZF, Putzig and Chromator (Figure 5E), suggesting that they are able to mediate TFIIA recruitment. Indeed, when we performed DNA affinity purification with a UAS-housekeeping promoter fragment, we observed co-recruitment of TFIIA with Gal4-GFZF but not Gal4-GFP (Supplementary Figure 5C). These data suggest that GFZF can recruit TFIIA and transactivate housekeeping promoters.

Overall, these results suggest that housekeeping promoters recruit TFIIA and Pol II indirectly via intermediary proteins interacting with DNA-binding proteins, whereas developmental promoters recruit TFIIA and the PIC directly via TBP.

Housekeeping initiation sites are not determined by DNA sequence motifs

The results so far suggest that housekeeping promoters are unable to efficiently recruit a canonical PIC *in vitro* and may exhibit weaker and more indirect interactions with GTFs. We

hypothesized that a less defined DNA-TFIIA or DNA-PIC interface at housekeeping promoters might be reflected in TSSs that are only weakly aligned with any of the relevant core-promoter sequence elements such as DREF or Ohler 1/6 motifs, thereby potentially explaining the well-known dispersed initiation patterns at housekeeping promoters. Alternatively, if DRE or Ohler 1/6 motifs and the proteins they bind were to direct PIC recruitment equivalently to TATA-box elements and TBP, the initiation patterns at housekeeping promoters should exhibit a consistent positional alignment relative to these motifs, and the dispersed initiation pattern at these promoters would stem from the presence of multiple motif instances per promoter.

To discern between these alternatives, we used Cap Analysis of Gene Expression (CAGE) data to analyze the distribution of TSSs relative to the positions of various motifs across *D. melanogster* promoters. As expected (e.g. (Ohler et al., 2002; Parry et al., 2010; Rach et al., 2011)) the TSSs of developmental promoters, such as TATA-box-, INR- or DPE-containing promoters, were restricted to a narrow window at consistent and precise distances from the core-promoter sequence elements (Figure 6A). Similarly, the TCT type housekeeping promoters exhibit a focused initiation pattern precisely at the TCT motif, presumably mediated by Trf2, a paralog of TBP, as previously suggested (Wang et al., 2014). These results confirm that initiation is precisely aligned to the TATA-box, INR, DPE and TCT motifs, as expected given previous reports and the fact that these motifs direct PIC and Pol II recruitment and initiation (Rach et al., 2011; Sawadogo and Roeder, 1985; White and Jackson, 1992).

In contrast, DRE- and Ohler 1-containing housekeeping promoters showed a dispersed distribution of CAGE signal in relation to DRE and Ohler 1 motifs, even for promoters that contain only a single motif occurrence (Figure 6A and Supplementary Figure 6A and 6B). Therefore, even though these motifs directly bind the DREF and M1BP factors, which can in turn recruit TFIIA, they do not instruct TSS position.

Housekeeping initiation sites are restricted but not specified by chromatin

As transcription initiation at housekeeping promoters was not aligned to a sequence feature, we considered whether the promoter-proximal chromatin structure, especially the nucleosome-depleted region (NDR) or the +1 nucleosome might constrain initiation patterns. Although the CAGE signal is not strongly aligned with the +1 nucleosome at developmental promoters, housekeeping promoters exhibit a defined broad distribution of CAGE signal in the NDR immediately upstream of a strongly positioned +1 nucleosome (Figure 6B and Sup-

plementary Figure 6D). These data show that initiation at housekeeping promoters occurs in a rather broad NDR upstream of the +1 nucleosome and suggest that the chromatin structure might be involved in determining TSS positions (Field et al., 2008; Rach et al., 2011).

Housekeeping cofactors underlie dispersed transcription initiation patterns

If the different initiation patterns at developmental and housekeeping promoters, in particular the dispersed initiation at housekeeping promoters, result from different mechanisms of Pol II PIC recruitment, then transcriptional activation from the housekeeping-type TFIIA recruitment factors above should always lead to more dispersed TSS patterns, irrespective of the promoter sequence. To test this systematically, we recruited the developmental-type coactivator MED25 and the housekeeping-type coactivator GFZF to a library of candidate promoters and analyzed the transcription initiation patterns (data from Haberle et al., 2019); Figure 6C. Although the two coactivators preferentially activate distinct sets of promoters (Haberle et al., 2019), 1266 promoters and 1268 random control sequences were activated sufficiently strongly by both coactivators to compare the respective initiation patterns (>4 fold induction over GFP with FDR<0.05)[Supplementary Figure 6C].

MED25 and GFZF recruitment indeed led to distinct initiation patterns on identical DNA fragments, such as a DNA fragment derived from the Mcm3 promoter (Fig. 6F, top) and from a DIP-kappa intron that does not function as an endogenous promoter (Fig. 6F, bottom).

To systematically assess the initiation events across the entire dataset, we calculated the proportion of initiation events at the dominant TSS compared to the sum of all initiation events across the entire promoter fragment. For all the core promoters, initiation was at the dominant TSS for 55% of MED25-activated events but only 42% of GFZF-activated events ($p=1.629637 \times 10^{-28}$; Wilcoxon rank-sum test, Fig. 6D) and this difference persisted when housekeeping and developmental promoter sequences were analyzed separately (Supplementary Figure 6E) and even for random non-promoter sequences, for which the corresponding proportions were 59 vs. 49% ($p= 2.382643 \times 10^{-22}$; Fig. 6D). Consistently, when we examined strong but non-dominant TSSs, we found that 47% of MED25-activated random non-promoter fragments had only one non-dominant TSS and 7% had 5 or more, whereas 34% of GFZF-induced random non-promoter fragments had only one non-dominant TSS and 17% had 5 or more (Supplementary Figure 6F). Moreover, MED25-induced transcription initiated for most promoters (51%) within a narrow 20bp region, while GFZF-induced transcription

initiated in a much broader region of 30 to 75bp (only 24% promoters initiated within 20bp; Figure 6E).

Thus, cofactor recruitment under identical conditions in an identical sequence context led to initiation patterns that are characteristically different for MED25 and GFZF, suggesting coactivators impose distinct patterns due to their different mechanisms of recruiting TFIIA, PIC and Pol II.

Discussion

The Pol II PIC is generally thought to be assembled and activate transcription similarly across promoters. Unexpectedly, our evidence suggests that functionally distinct promoters recruit GTFs and assemble the Pol II PIC via distinct strategies. Developmental promoter DNA is sufficient to recruit and assemble a Pol II PIC from nuclear extract *in vitro,* by having high affinity to GTFs such as TBP and, in turn, a close connection to TFIIA. Found as part of a soluble Pol II holoenzyme in yeast, TBP and TFIIA are tightly associated with chromatin in metazoans and are important in directing Pol II PIC assembly on DNA and cofactor mediated transcription *in vitro* (Kimura et al., 1999; Koleske and Young, 1995; Lieberman et al., 1997). In contrast, housekeeping promoters do not closely interact with GTFs and are not able to recruit PIC components. Instead, housekeeping promoters bind sequence-specific transcription factors such as DREF and M1BP, which in turn interact with cofactors such as GFZF, Chromator and Putzig that – directly or indirectly – recruit with GTFs (e.g. TFIIA) and Pol II, a previously unknown interaction (Figure 6G) (Baumann et al., 2018). These differences in the assembly and stability of the DNA-protein interface and protein complexes lead to distinct transcription initiation patterns at developmental and housekeeping promoters, which generally exhibit focused and dispersed initiation patterns, respectively. Importantly, forced recruitment of housekeeping cofactors such as GFZF to arbitrary DNA sequences is sufficient to induce broad transcription initiation patterns, consistent with the initiation patters observed at housekeeping promoters *in vivo* and alternative PIC recruitment.

The alternative mechanisms converge and make TFIIA recruitment essential for transcription initiation at any promoter. A central role of TFIIA recruitment for transcription initiation is consistent with the direct interaction of the TBP paralog TBPL2 with TFIIA in oocyte transcription (Yu et al., 2020)and non-canonical Pol II transcription of transposon-rich and H3K9me3-marked piRNA source loci in Drosophila germ cells through the TFIIA paralog moonshiner (Andersen et al., 2017).

Unlike developmental promoters, the mode of recruitment utilized by housekeeping promoters may help maintain transcription and an NDR for a longer period of time and may be localized to a spatial region in the nucleus with other housekeeping promoters (Cubeñas-Potts et al., 2017). This mode of transcription initiation may be more robust to changes in the nuclear environment and preserve the fidelity of housekeeping gene transcription. Importantly, broad patterns of transcription initiation in Drosophila melanogaster housekeeping promoters are similarly observed in the majority of vertebrate CpG island promoters compromising roughly 70% of all promoters (Carninci et al., 2006; Danks et al., 2018; FANTOM Consortium and the RIKEN PMI and CLST (DGT) et al., 2014; Saxonov et al., 2006). Open questions regarding the redundancy of housekeeping cofactors in maintaining transcription, and the transcriptional kinetics imposed by developmental and housekeeping cofactors will be important to investigate, to determine the function of the two transcription units in space and time.

References

Akhtar, W., Veenstra, G.J.C., 2011. TBP-related factors: a paradigm of diversity in transcription initiation. Cell Biosci 1, 23–12. doi:10.1186/2045-3701-1-23

Andersen, P.R., Tirian, L., Vunjak, M., Brennecke, J., 2017. A heterochromatin-dependent transcription machinery drives piRNA expression. Nature 549, 54–59. doi:10.1038/nature23482

Arnold, C.D., Zabidi, M.A., Pagani, M., Rath, M., Schernhuber, K., Kazmar, T., Stark, A., 2016. Genome-wide assessment of sequence-intrinsic enhancer responsiveness at single-base-pair resolution. Nature Publishing Group 1–12. doi:10.1038/nbt.3739

Baek, H.J., Kang, Y.K., Roeder, R.G., 2006. Human Mediator enhances basal transcription by facilitating recruitment of transcription factor IIB during preinitiation complex assembly. J. Biol. Chem. 281, 15172–15181. doi:10.1074/jbc.M601983200

Baumann, D.G., Dai, M.-S., Lu, H., Gilmour, D.S., 2018. GFZF, a Glutathione S-Transferase Protein Implicated in Cell Cycle Regulation and Hybrid Inviability, Is a Transcriptional Coactivator. Molecular and Cellular Biology 38. doi:10.1128/MCB.00476-17

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

Brown, J.B., Boley, N., Eisman, R., May, G.E., Stoiber, M.H., Duff, M.O., Booth, B.W., Wen, J., Park, S., Suzuki, A.M., Wan, K.H., Yu, C., Zhang, D., Carlson, J.W., Cherbas, L., Eads, B.D., Miller, D., Mockaitis, K., Roberts, J., Davis, C.A., Frise, E., Hammonds, A.S., Olson, S., Shenker, S., Sturgill, D., Samsonova, A.A., Weiszmann, R., Robinson, G., Hernandez, J., Andrews, J., Bickel, P.J., Carninci, P., Cherbas, P., Gingeras, T.R., Hoskins, R.A., Kaufman, T.C., Lai, E.C., Oliver, B., Perrimon, N., Graveley, B.R., Celniker, S.E., 2014. Diversity and dynamics of the Drosophila transcriptome. Nature 512, 393–399. doi:10.1038/nature12962

Buratowski, S., Hahn, S., Guarente, L., Sharp, P.A., 1989. Five intermediate complexes in transcription initiation by RNA polymerase II. Cell 56, 549–561. doi:10.1016/0092-8674(89)90578-3

Butler, J.E.F., Kadonaga, J.T., 2002. The RNA polymerase II core promoter: a key component in the regulation of gene expression. Genes & Development 16, 2583–2592. doi:10.1101/gad.1026202

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A.M., Taylor, M.S., Engström, P.G., Frith, M.C., Forrest, A.R.R., Alkema, W.B., Tan, S.L., Plessy, C., Kodzius, R., Ravasi, T., Kasukawa, T., Fukuda, S., Kanamori-Katayama, M., Kitazume, Y., Kawaji, H., Kai, C., Nakamura, M., Konno, H., Nakano, K., Mottagui-Tabar, S., Arner, P., Chesi, A., Gustincich, S., Persichetti, F., Suzuki, H., Grimmond, S.M., Wells, C.A., Orlando, V., Wahlestedt, C., Liu, E.T., Harbers, M., Kawai, J., Bajic, V.B., Hume, D.A., Hayashizaki, Y., 2006. Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet 38, 626–635. doi:10.1038/ng1789

Chereji, R.V., Kan, T.-W., Grudniewska, M.K., Romashchenko, A.V., Berezikov, E., Zhimulev, I.F., Guryev, V., Morozov, A.V., Moshkin, Y.M., 2016. Genome-wide profiling of nucleosome sensitivity and chromatin accessibility in Drosophila melanogaster. Nucleic Acids Res 44, 1036–1051. doi:10.1093/nar/gkv978

Cosma, M.P., 2002. Ordered Recruitment: Gene-Specific Review Mechanism of Transcription Activation. Molecular Cell 10, 227–236.

Cubeñas-Potts, C., Rowley, M.J., Lyu, X., Li, G., Lei, E.P., Corces, V.G., 2017. Different enhancer classes in Drosophila bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture. Nucleic Acids Res 45, 1714–1730. doi:10.1093/nar/gkw1114

Danks, G.B., Navratilova, P., Lenhard, B., Thompson, E.M., 2018. Distinct core promoter codes drive transcription initiation at key developmental transitions in a marine chordate. BMC Genomics 19, 164–12. doi:10.1186/s12864-018-4504-5

Deato, M.D.E., Marr, M.T., Sottero, T., Inouye, C., Hu, P., Tjian, R., 2008. MyoD targets TAF3/TRF3 to activate myogenin transcription. Molecular Cell 32, 96–105. doi:10.1016/j.molcel.2008.09.009

Dignam, D.R., Lebovitz, R.M., Roeder, R.G., 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. NAR 11, 1–15.

Doblmann, J., Dusberger, F., Imre, R., Hudecz, O., Stanek, F., Mechtler, K., Dürnberger, G., 2018. apQuant: Accurate Label-Free Quantification by Quality Filtering. J. Proteome Res. doi:10.1021/acs.jproteome.8b00113

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

FANTOM Consortium and the RIKEN PMI and CLST (DGT), Forrest, A.R.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J.L., Haberle, V., Lassmann, T., Kulakovskiy, I.V., Lizio, M., Itoh, M., Andersson, R., Mungall, C.J., Meehan, T.F., Schmeier, S., Bertin, N., Jørgensen, M., Dimont, E., Arner, E., Schmidl, C., Schaefer, U., Medvedeva, Y.A., Plessy, C., Vitezic, M., Severin, J., Semple, C.A., Ishizu, Y., Young, R.S., Francescatto, M., Alam, I., Albanese, D., Altschuler, G.M., Arakawa, T., Archer, J.A.C., Arner, P., Babina, M., Rennie, S., Balwierz, P.J., Beckhouse, A.G., Pradhan-Bhatt, S., Blake, J.A., Blumenthal, A., Bodega, B., Bonetti, A., Briggs, J., Brombacher, F., Burroughs, A.M., Califano, A., Cannistraci, C.V., Carbajo, D., Chen, Y., Chierici, M., Ciani, Y., Clevers, H.C., Dalla, E., Davis, C.A., Detmar, M., Diehl, A.D., Dohi, T., Drabløs, F., Edge, A.S.B., Edinger, M., Ekwall, K., Endoh, M., Enomoto, H., Fagiolini, M., Fairbairn, L., Fang, H., Farach-Carson, M.C., Faulkner, G.J., Favorov, A.V., Fisher, M.E., Frith, M.C., Fujita, R., Fukuda, S., Furlanello, C., Furino, M., Furusawa, J.-I., Geijtenbeek, T.B., Gibson, A.P., Gingeras, T., Goldowitz, D., Gough, J., Guhl, S., Guler, R., Gustincich, S., Ha, T.J., Hamaguchi, M., Hara, M., Harbers, M., Harshbarger, J., Hasegawa, A., Hasegawa, Y., Hashimoto, T., Herlyn, M., Hitchens, K.J., Ho Sui, S.J., Hofmann, O.M., Hoof, I.,

Hori, F., Huminiecki, L., Iida, K., Ikawa, T., Jankovic, B.R., Jia, H., Joshi, A., Jurman, G., Kaczkowski, B., Kai, C., Kaida, K., Kaiho, A., Kajiyama, K., Kanamori-Katayama, M., Kasianov, A.S., Kasukawa, T., Katayama, S., Kato, S., Kawaguchi, S., Kawamoto, H., Kawamura, Y.I., Kawashima, T., Kempfle, J.S., Kenna, T.J., Kere, J., Khachigian, L.M., Kitamura, T., Klinken, S.P., Knox, A.J., Kojima, M., Kojima, S., Kondo, N., Koseki, H., Koyasu, S., Krampitz, S., Kubosaki, A., Kwon, A.T., Laros, J.F.J., Lee, W., Lennartsson, A., Li, K., Lilje, B., Lipovich, L., Mackay-Sim, A., Manabe, R.-I., Mar, J.C., Marchand, B., Mathelier, A., Mejhert, N., Meynert, A., Mizuno, Y., de Lima Morais, D.A., Morikawa, H., Morimoto, M., Moro, K., Motakis, E., Motohashi, H., Mummery, C.L., Murata, M., Nagao-Sato, S., Nakachi, Y., Nakahara, F., Nakamura, T., Nakamura, Y., Nakazato, K., van Nimwegen, E., Ninomiya, N., Nishiyori, H., Noma, S., Noazaki, T., Ogishima, S., Ohkura, N., Ohimiya, H., Ohno, H., Ohshima, M., Okada-Hatakeyama, M., Okazaki, Y., Orlando, V., Ovchinnikov, D.A., Pain, A., Passier, R., Patrikakis, M., Persson, H., Piazza, S., Prendergast, J.G.D., Rackham, O.J.L., Ramilowski, J.A., Rashid, M., Ravasi, T., Rizzu, P., Roncador, M., Roy, S., Rye, M.B., Saijyo, E., Sajantila, A., Saka, A., Sakaguchi, S., Sakai, M., Sato, H., Savvi, S., Saxena, A., Schneider, C., Schultes, E.A., Schulze-Tanzil, G.G., Schwegmann, A., Sengstag, T., Sheng, G., Shimoji, H., Shimoni, Y., Shin, J.W., Simon, C., Sugiyama, D., Sugiyama, T., Suzuki, M., Suzuki, N., Swoboda, R.K., 't Hoen, P.A.C., Tagami, M., Takahashi, N., Takai, J., Tanaka, H., Tatsukawa, H., Tatum, Z., Thompson, M., Toyodo, H., Toyoda, T., Valen, E., van de Wetering, M., van den Berg, L.M., Verado, R., Vijayan, D., Vorontsov, I.E., Wasserman, W.W., Watanabe, S., Wells, C.A., Winteringham, L.N., Wolvetang, E., Wood, E.J., Yamaguchi, Y., Yamamoto, M., Yoneda, M., Yonekura, Y., Yoshida, S., Zabierowski, S.E., Zhang, P.G., Zhao, X., Zucchelli, S., Summers, K.M., Suzuki, H., Daub, C.O., Kawai, J., Heutink, P., Hide, W., Freeman, T.C., Lenhard, B., Bajic, V.B., Taylor, M.S., Makeev, V.J., Sandelin, A., Hume, D.A., Carninci, P., Hayashizaki, Y., 2014. A promoter-level mammalian expression atlas. Nature 507, 462–470. doi:10.1038/nature13182

Field, Y., Kaplan, N., Fondufe-Mittendorf, Y., Moore, I.K., Sharon, E., Lubling, Y., Widom, J., Segal, E., 2008. Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. PLoS Comput Biol 4, e1000216. doi:10.1371/journal.pcbi.1000216

Gazdag, E., Jacobi, U.G., van Kruijsbergen, I., Weeks, D.L., Veenstra, G.J.C., 2016. Activation of a T-box-Otx2-Gsc gene network independent of TBP and TBP-related factors. Development 143, 1340–1350. doi:10.1242/dev.127936

Geiger, J.H., Hahn, S., Lee, S., Sigler, P.B., 1996. Crystal structure of the yeast TFIIA/TBP/DNA complex. Science 272, 830–836. doi:10.1126/science.272.5263.830

Gurudatta, B.V., Yang, J., Van Bortle, K., Donlin-Asp, P.G., Corces, V.G., 2013. Dynamic changes in the genomic localization of DNA replication-related element binding factor during the cell cycle. Cell Cycle 12, 1605–1615. doi:10.4161/cc.24742

Haberle, V., Arnold, C.D., Pagani, M., Rath, M., Schernhuber, K., Stark, A., 2019. Transcriptional cofactors display specificity for distinct types of core promoters. Nature 1–24. doi:10.1038/s41586-019-1210-7

Haberle, V., Li, N., Hadzhiev, Y., Plessy, C., Previti, C., Nepal, C., Gehrig, J., Dong, X., Akalin, A., Suzuki, A.M., van IJcken, W.F.J., Armant, O., Ferg, M., Strähle, U., Carninci, P., Müller, F., Lenhard, B., 2014. Two independent transcription initiation codes overlap on vertebrate core promoters. Nature 507, 381–385. doi:10.1038/nature12974

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

He, Y., Fang, J., Taatjes, D.J., Nogales, E., 2013. Structural visualization of key steps in human transcription initiation. Nature 495, 481–486. doi:10.1038/nature11991

Hoskins, R.A., Landolin, J.M., Brown, J.B., Sandler, J.E., Takahashi, H., Lassmann, T., Yu, C., Booth, B.W., Zhang, D., Wan, K.H., Yang, L., Boley, N., Andrews, J., Kaufman, T.C., Graveley, B.R., Bickel, P.J., Carninci, P., Carlson, J.W., Celniker, S.E., 2011. Genome-wide analysis of promoter architecture in Drosophila melanogaster. Genome Research 21, 182–192. doi:10.1101/gr.112466.110

Isogai, Y., Keles, S., Prestel, M., Hochheimer, A., Tjian, R., 2007. Transcription of histone gene cluster by differential core-promoter factors. Genes & Development 21, 2936–2949. doi:10.1101/gad.1608807

Johnson, K.M., Wang, J., Smallwood, A., Carey, M., 2004. The immobilized template assay for measuring cooperativity in eukaryotic transcription complex assembly. Methods Enzymol 380, 207–219. doi:10.1016/S0076-6879(04)80010-7

Kadonaga, J.T., Tjian, R., 1986. Affinity purification of sequence-specific DNA binding proteins. Proc Natl Acad Sci USA 83, 5889–5893.

Kedmi, A., Sloutskin, A., Epstein, N., Gasri-Plotnitsky, L., Ickowicz, D., Shoval, I., Doniger, T., Darmon, E., Ideses, D., Porat, Z., Yaron, O., Juven-Gershon, T., 2020. The transcription factor TRF2 has a unique function in regulating cell cycle and apoptosis, bioRxiv. doi:10.1101/2020.03.27.011288

Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S., Karolchik, D., 2010. BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics 26, 2204–2207. doi:10.1093/bioinformatics/btq351

Kimura, H., Tao, Y., Roeder, R.G., Cook, P.R., 1999. Quantitation of RNA polymerase II and its transcription factors in an HeLa cell: little soluble holoenzyme but significant amounts of polymerases attached to the nuclear substructure. Molecular and Cellular Biology 19, 5383–5392. doi:10.1128/MCB.19.8.5383

Koleske, A.J., Young, R.A., 1995. The RNA polymerase II holoenzyme and its implications for gene regulation. Trends in Biochemical Sciences 20, 113–116. doi:10.1016/s0968-0004(00)88977-x

Köcher, T., Pichler, P., Swart, R., Mechtler, K., 2012. Analysis of protein mixtures from whole-cell extracts by single-run nanoLC-MS/MS using ultralong gradients. Nature Publishing Group 7, 882–890. doi:10.1038/nprot.2012.036

Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, 1–10. doi:10.1186/gb-2009-10-3-r25

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., Carey, V.J., 2013. Software for computing and annotating genomic ranges. PLoS Comput Biol 9, e1003118. doi:10.1371/journal.pcbi.1003118

Lenhard, B., Sandelin, A., Carninci, P., 2012. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nature Reviews Genetics 13, 233–245. doi:10.1038/nrg3163

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. doi:10.1093/bioinformatics/btp352

Liang, J., Lacroix, L., Gamot, A., Cuddapah, S., Queille, S., Lhoumaud, P., Lepetit, P., Martin, P.G.P., Vogelmann, J., Court, F., Hennion, M., Micas, G., Urbach, S., Bouchez, O., Nöllmann, M., Zhao, K., Emberly, E., Cuvier, O., 2014. Chromatin immunoprecipitation indirect peaks highlight long-range interactions of insulator proteins and Pol II pausing. Molecular Cell 53, 672–681. doi:10.1016/j.molcel.2013.12.029

Lieberman, P.M., Ozer, J., Gürsel, D.B., 1997. Requirement for transcription factor IIA (TFIIA)-TFIID recruitment by an activator depends on promoter structure and template competition. Molecular and Cellular Biology 17, 6624–6632. doi:10.1128/MCB.17.11.6624

Lin, J.J., Carey, M., 2012. In vitro transcription and immobilized template analysis of preinitiation complexes. Curr Protoc Mol Biol Chapter 12, Unit 12.14.–12.14.19. doi:10.1002/0471142727.mb1214s97

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 1–21. doi:10.1186/s13059-014-0550-8

Mahat, D.B., Kwak, H., Booth, G.T., Jonkers, I.H., Danko, C.G., Patel, R.K., Waters, C.T., Munson, K., Core, L.J., Lis, J.T., 2016. Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). Nature Publishing Group 11, 1455–1476. doi:10.1038/nprot.2016.086

Martianov, I., Velt, A., Davidson, G., Choukrallah, M.-A., Davidson, I., 2016. TRF2 is recruited to the pre-initiation complex as a testis-specific subunit of TFIIA/ALF to promote haploid cell gene expression. Sci Rep 6, 32069–13. doi:10.1038/srep32069

Martianov, I., Viville, S., Davidson, I., 2002. RNA polymerase II transcription in murine cells lacking the TATA binding protein. Science 298, 1036–1039. doi:10.1126/science.1076327

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10–12. doi:10.14806/ej.17.1.200

Murakami, K., Elmlund, H., Kalisman, N., Bushnell, D.A., Adams, C.M., Azubel, M., Elmlund, D., Levi-Kalisman, Y., Liu, X., Gibbons, B.J., Levitt, M., Kornberg, R.D., 2013. Architecture of an RNA polymerase II transcription pre-initiation complex. Science 342, 1238724. doi:10.1126/science.1238724

Neves, A., Eisenman, R.N., 2019. Distinct gene-selective roles for a network of core promoter factors in Drosophila neural stem cell identity. Biol Open 8. doi:10.1242/bio.042168

Nikolov, D.B., Chen, H., Halay, E.D., Usheva, A.A., Hisatake, K., Lee, D.K., Roeder, R.G., Burley, S.K., 1995. Crystal structure of a TFIIB-TBP-TATA-element ternary complex. Nature 377, 119–128. doi:10.1038/377119a0

Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T., Kanemaki, M., 2009. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. Nat Meth 6, 917–922. doi:10.1038/nmeth.1401

Ohler, U., Liao, G.-C., Niemann, H., Rubin, G.M., 2002. Computational analysis of core promoters in the Drosophila genome. Genome Biol. 3, RESEARCH0087–12. doi:10.1186/gb-2002-3-12-research0087

Orphanides, G., lagrange, T., Reinberg, D., 1996. The general transcription factors of RNA polymerase II. Genes & Development 10, 2657–2683.

Papai, G., Tripathi, M.K., Ruhlmann, C., Layer, J.H., Weil, P.A., Schultz, P., 2010. TFIIA and the transactivator Rap1 cooperate to commit TFIID for transcription initiation. Nature 465, 956–960. doi:10.1038/nature09080

Parry, T.J., Theisen, J.W.M., Hsu, J.Y., Wang, Y.L., Corcoran, D.L., Eustice, M., Ohler, U., Kadonaga, J.T., 2010. The TCT motif, a key component of an RNA polymerase II transcription system for the translational machinery. Genes & Development 24, 2013–2018. doi:10.1101/gad.1951110

Parvin, J.D., Shykind, B.M., Meyers, R.E., Kim, J., Sharp, P.A., 1994. Multiple sets of basal factors initiate transcription by RNA polymerase II. J. Biol. Chem. 269, 18414–18421.

Parvin, J.D., Timmers, H.T.M., Sharp, P.A., 1992. Promoter Specificity of Basal Transcription Factors. Cell 68, 1135–1144.

Plaschka, C., Larivière, L., Wenzeck, L., Seizl, M., Hemann, M., Tegunov, D., Petrotchenko, E.V., Borchers, C.H., Baumeister, W., Herzog, F., Villa, E., Cramer, P., 2015. Architecture of the RNA polymerase II-Mediator core initiation complex. Nature 518, 376–380. doi:10.1038/nature14229

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. doi:10.1093/bioinformatics/btq033

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

Ranish, J.A., Yudkovsky, N., Hahn, S., 1999. Intermediates in formation and activity of the RNA polymerase II preinitiation complex: holoenzyme recruitment and a postrecruitment role for the TATA box and TFIIB. Genes & Development 13, 49–63. doi:10.1101/gad.13.1.49

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K., 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43, e47. doi:10.1093/nar/gkv007

Sawadogo, M., Roeder, R.G., 1985. Factors involved in specific transcription by human RNA polymerase II: analysis by a rapid and quantitative in vitro assay. Proc Natl Acad Sci USA 82, 4394–4398. doi:10.1073/pnas.82.13.4394

Saxonov, S., Berg, P., Brutlag, D.L., 2006. 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. doi:10.1073/pnas.0510310103

Shao, W., Zeitlinger, J., 2017. Paused RNA polymerase II inhibits new transcriptional initiation. Nat Genet 49, 1045–1051. doi:10.1038/ng.3867

Smyth, G.K., 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 3, Article3. doi:10.2202/1544-6115.1027

Stampfel, G., Kazmar, T., Frank, O., Wienerroither, S., Reiter, F., Stark, A., 2015. Transcriptional regulators form diverse groups with context-dependent regulatory functions. Nature 1–18. doi:10.1038/nature15545

Stijf-Bultsma, Y., Sommer, L., Tauber, M., Baalbaki, M., Giardoglou, P., Jones, D.R., Gelato, K.A., van Pelt, J., Shah, Z., Rahnamoun, H., Toma, C., Anderson, K.E., Hawkins, P., Lauberth, S.M., Haramis, A.-P.G., Hart, D., Fischle, W., Divecha, N., 2015. The basal transcription complex component TAF3 transduces changes in nuclear phosphoinositides into transcriptional output. Molecular Cell 58, 453–467. doi:10.1016/j.molcel.2015.03.009

Tan, S., Hunziker, Y., Sargent, D.F., Richmond, T.J., 1996. Crystal structure of a yeast TFIIA/TBP/DNA complex. Nature 381, 127–151. doi:10.1038/381127a0

Taus, T., Köcher, T., Pichler, P., Paschke, C., Schmidt, A., Henrich, C., Mechtler, K., 2011. Universal and confident phosphorylation site localization using phosphoRS. J. Proteome Res. 10, 5354–5362. doi:10.1021/pr200611n

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

Vo Ngoc, L., Huang, C.Y., Cassidy, C.J., Medrano, C., Kadonaga, J.T., 2020. Identification of the human DPR core promoter element using machine learning. Nature 585, 459–463. doi:10.1038/s41586-020-2689-7

Vo Ngoc, L., Kassavetis, G.A., Kadonaga, J.T., 2019. The RNA Polymerase II Core Promoter in Drosophila. Genetics 212, 13–24. doi:10.1534/genetics.119.302021

Wang, Y.L., Duttke, S.H.C., Chen, K., Johnston, J., Kassavetis, G.A., Zeitlinger, J., Kadonaga, J.T., 2014. TRF2, but not TBP, mediates the transcription of ribosomal protein genes. Genes & Development 28, 1550–1555. doi:10.1101/gad.245662.114

Warfield, L., Ramachandran, S., Baptista, T., Devys, D., Tora, L., Hahn, S., 2017. Transcription of Nearly All Yeast RNA Polymerase II-Transcribed Genes Is Dependent on Transcription Factor TFIID. Molecular Cell 68, 118–129.e5. doi:10.1016/j.molcel.2017.08.014

White, R.J., Jackson, S.P., 1992. Mechanism of TATA-Binding Protein Recruitment to a TATA-Less Class III Promoter. Cell 71, 1041–1053.

Wieczorek, E., Brand, M., Jacq, X., Tora, L., 1998. Function of TAFII-containing complex without TBP in transcription by RNA polymerase II. Nature 393, 187–191.

Yu, C., Cvetesic, N., Hisler, V., Gupta, K., Ye, T., Gazdag, E., Negroni, L., Hajkova, P., Berger, I., Lenhard, B., Müller, F., Vincent, S.D., Tora, L., 2020. TBPL2/TFIIA complex establishes the maternal transcriptome through oocyte-specific promoter usage. Nature Communications 11, 6439–13. doi:10.1038/s41467-020-20239-4

Yudkovsky, N., Ranish, J.A., Hahn, S., 2000. A transcription reinitiation intermediate that is stabilized by activator. Nature 408, 225–229. doi:10.1038/35041603

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

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

Acknowledgements. We thank all members of the Stark lab for feedback on this project. 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 IMP is supported by Boehringer Ingelheim GmbH and the Austrian Research Promotion Agency (FFG). LS is supported by a DOC PhD Fellowship from the Austrian Academy of Sciences.

Author Contributions. AS and LS conceived and designed the experiments and wrote the manuscript. LS, KB and OH generated endogenously tagged AID cell lines and performed PRO-seq. LP and AV performed analysis of PRO-seq data. FN & LS performed TFIIA immunoprecipitation for mass spectrometry. LS performed DNA affinity purifications. KM & ER performed mass spectrometry. AV & VL analyzed mass spectrometry data. VH performed STAP-seq analysis.

Figure legends

1. **DNA affinity purifications uncover differentially bound proteins at functionally distinct promoters**
   A. A pie chart representing *Drosophila melanogaster* promoters by their DNA motif content.
   B. Scheme of DNA affinity purification coupled to label free mass spectrometry.
   C. Enrichment of proteins detected by mass spectrometry on a pool of TATA-box promoters over control DNA sequences, and in D. over a pool of DRE promoters. 3 biological replicates were performed for each promtoer and control pool, significance measured with a Limma p-value<0.05.
   D. Enrichment of proteins bound to DRE promoters over TATA-box promoters. Limma p-value<0.05.

2. **Developmental and housekeeping promoters bind different sets of proteins and GTFs**
   A. An example of tested regions around the zip promoter that were tested in DNA-affinity purification and luciferase assay.
   B. Elution fractions of DNA-affinity purification assays tested with DRE promoter with various lengths with a nuclear extract expressing an endogenously tagged TFIIA-FLAG-AID protein.
   C. Total number of enriched proteins on the different tested pooled promoter types from the DNA affinity purification mass spectrometry.
   D. Enrichments from DNA affinity purification mass spectrometry of selected proteins and protein complexes on the different pooled promoter types compared to negative control DNA.
   E. Z-score normalized ChIP-seq signal of GTFs and select housekeeping binders on the different types of promoters.

3. **TBP and DREF are required by distinct sets of promoters**
   A. Strategy for generating endogenously tagged AID cell lines. An AID-3xFLAG endogenous knock-in was generated in the N-terminus of either DREF or TBP in a background cell line stably expressing the Tir1 ligase downstream of Actin5c.

B. Western blot on FLAG tagged TBP and DREF 0,1 and 3 hours after auxin addition showing protein degradation.
C. PRO-seq measurement after 6 hours of auxin addition to the TBP or DREF AID tagged cell lines, MA plots represent colored genes which are significantly down-regulated compared to no auxin control (FDR<0.05). Two biological replicates per conditions.
D. Overlap of the TBP and DREF depletion PRO-seq. Green and orange colored dots represent TBP and DREF dependent promoters, and in black, a single promoter which was dependent on both TBP and DREF.
E. Fisher's exact test for motif enrichment in TBP and DREF downregulated promoters compared to all expressed promoters. Log2 Odds ratio displayed.
F. Genome tracks of PRO-seq data indicating examples of genes that are dependent on TBP or DREF.
G. PRO-seq signal quantified across all expressed TATA-box and DREF containing promoters, or TATA-box and DRE promoters which were used for the DNA affinity purification. A two-sided Wilcoxon test was used to determine significant down-regulation with p<0.05.

4. Preferential promoter subtype bound proteins *in vitro* are required by distinct sets of promoters *in vivo*

A. Heat map representing enrichment of bound proteins on select promoter types as measured by DNA affinity purification over negative control DNA. Three biological replicates per condition with a Limma p-value<0.05.
B. PRO-seq measurement 6 hours after auxin addition to an endogenously tagged CG3847-AID cell line, and Fisher's exact test for motif enrichment on downregulated promoters, log2 odds ratio displayed. Two biological replicates per condition.
C. PRO-seq measurement 6 hours after auxin addition to an endogenously tagged Chromator-AID cell line, and Fisher's exact test of motif enrichment on upregulated promoters, log2 odds ratio displayed. Two biological replicates per condition.
D. PRO-seq measurement 6 hours after auxin addition to an endogenously tagged CG14711-AID cell line, and Fisher's exact test of motif enrichment on upregulated promoters, log2 odds ratio displayed. Two biological replicates per condition.

E. PRO-seq signal on promoters used for DNA affinity purification for CG3847- (left panel), Chromator- (center panel) and CG14711- (right panel) AID cell line depletion. Two biological replicates per condition.

5. TFIIA is required by all promoters and is recruited by housekeeping cofactors to housekeeping promoters

A. Western blot for an endogenously tagged TFIIA-L-3x-FLAG-AID cell line after addition of auxin.

B. MA-plot of PRO-seq measurement 6 hours after auxin addition to the TFIIA-L-AID cell line. Colored dots represent significant downregulation, FDR<0.05. Two biological replicates per condition.

C. PRO-seq signal at all expressed promoters, represented according to their motif content.

D. Overlap of TFIIA-L immunoprecipitation mass spectrometry data with DNA affinity purification mass spectrometry of the 3 tested housekeeping promoter types. Three biological replicates per conditions with Limma p-value<0.05 and enrichment>0.

E. Luciferase assay in which Gal4 DNA binding domain fusion proteins were recruited to 4xUAS sites upstream of a minimal housekeeping Rps12 promoter. Measurements are normalized to Renilla luciferase (transfection control) and GFP. * denotes proteins activating a housekeeping promoters with a log2FC>1.5 and p-value<0.05, two-tailed student's t-test.

6. Housekeeping cofactor recruitment is sufficient to recapitulate dispersed transcription initiation patterns

A. Distribution of CAGE signal from mixed *D.mel* embryos (0-24h) centered on the location of promoter DNA motif sequence across the 6 main promoter types investigated in this study.

B. CAGE data centered on the +1 nucleosome called from MNase-seq data on Ohler1 and DRE promoters.

C. Scheme of cofactor recruitment STAP-seq testing MED25 or GFZF Gal4 DNA-binding domain fusions recruited to a library of candidate promoter fragments.

D. Plot of the percent of STAP-seq signal originating at the dominant TSS

(normalized to GFP) at promoters and random regions that are activated by both GFZF and MED25 recruitment.

E. Histogram representing the fraction of random regions with 1 or more TSSs upon recruitment of MED25 or GFZF, and a cumulative plot of the same data. Wilcoxon rank sum test on three biological replicates.

F. Cofactor recruitment STAP-seq tracks indicating examples of promoters and random regions that are activated by both cofactors, upon recruitment of GFP, MED25 or GFZF.

G. Scheme of Pol II PIC recruitment to developmental promoters, which occurs through direct engagement between the transcription machinery and developmental promoter sequence features, resulting in narrow initiation patterns, whereas housekeeping promoters recruit Pol II through housekeeping DNA-binding proteins and intermediary cofactors that interact with TFIIA, resulting in dispersed initiation.

Resource availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Alexander Stark (alexander.stark@imp.ac.at).

Materials availability

All cell lines and plasmids generated in this study are available upon request.

Data and code availability

PRO-seq data have been deposited to the Gene Expression Omnibus (GEO), accession GSE181257, which will be publicaly available at the date of publication.
Raw mass spectrometry data have been deposited Proteome exchange.

Experimental model and subject details

Cell lines

Drosophila Schneider S2 cells were purchased from Gibco and were maintained according to the manufacturer's recommendations as described in the method details.
Engineered cell lines were created using CRISPR/Cas9 as described in the method details.

Method details

Cell culture

Drosophila melanogaster S2 cells were obtained from Thermo Fisher and maintained in Schneider's Drosophila Medium supplemented with 10% heat-inactivated fetal bovine serum.

Generation of endogenously tagged AID cell lines

A parental cell-line expressing the osTir ligase was created with a knock-in approach by introducing a vector expressing a gRNA/Cas9 targeting the carboxyl terminus of the Act5C, with a P2A before the osTir-mCherry construct, leading to constitutive expression of the osTir ligase. Wild type S2 cells were electroporated using the MaxCyte-STX system at a density of $1x10^7$ cells per 100µl and 20 µg of DNA using the pre-set protocols

Selection of promoters and controls for DNA-affinity purification

Unique CAGE corrected TSSs were scored for PWM for core-promoter motifs (TATA-box, INR, DPE, TCT, Ohler1/6, DRE), and the highest scoring TSSs that were expressed in Drosophila S2 cells (≥5tpm), and were inducible in STAP-seq (Arnold et al., 2016) were used. The following PWM score cut-offs were used for the selected motifs: TCT ≥ 95%, Ohler 1+6 ≥ 95%, DRE 100%, TATA-box and INR ≥ 95%, DPE and INR ≥ 95%, and INR only ≥ 95%. Length matched control regions were selected from the Drosophila genome to not overlap transcribed regions, and did not show any sign of transcription in any Drosophila developmental CAGE data. Selected promoters are listed in supplementary table 1.

Cloning promoter constructs

Promoter regions were PCR amplified from S2 cell genomic DNA using primers containing gibson overhangs corresponding to the BglII and HindII restriction sites on pGL3 with Q5 high-fidelity 2x master mix (NEB). PCR products were cleaned with AMPURE beads and eluted in water. Gibson reactions were performed with a Gibson assembly master mix (NEB) according to the manufacturer's recommendations.

1ul of Gibson reaction was electroporated into MegaX DH10B electrocompetent cells (Thermo). Single clones were picked and grown in 5mL bacterial cultures. Minipreps were performed using a Qiagen kit, and Sanger-sequencing was performed in-house. Correct plasmid clones were used as a template for amplification of biotinylated DNA.

Preparation and immobilization of biotinylated DNA

Biotinylated DNA was generating using a forward primer containing a Biotin TEG group on the 5' end obtained from Sigma Aldrich: Biotin TEG 5', and a reverse Reverse 3' primer (see resource table for primer sequences). At least 2mL of total PCR volume (performed in 50ul reactions) for each individual promoter sequence was amplified individually for each replicate. PCR reactions were pooled and DNA was purified using AMPURE beads and eluted in water. For each sample, 50µl of Dyna M280 Streptavidin were used and coupled to 15µg of cleaned biotinylated PCR product according to the manufacturer's recommendations. The beads were placed in an equivalent volume of DBB (150 mM NaCl, 50 mM Tris/HCl pH, 8.0, 10 mM MgCl2) and used immediately for DNA-affinity purification assay.

Preparation of nuclear extracts

Nuclear extracts from drosophila S2 cells were prepared as previously described with the following modifications (Dignam et al., 1983). Three billion drosophila S2 cells were harvested by resuspension and washed with PBS. The cell pellet was resuspended in buffer A (10mM HEPES pH7.9, 1.5mM MgCl2, 10mM KCl, 0.5mM DTT added fresh before use, and oComplete EDTA-free protease inhibitors) placed on ice for 10 minutes. Cells were spund down at 700g for 5 minutes, supernatant removed, and cells were resuspended in 5 cell-pellet volumes of buffer A supplemented with 0.5% NP-40. Cell suspension was dounced in a Beckman 15mL dounce with a 'loose' pestle for 10 strokes to isolate nuclei. Cells were spun down at 2000g for 5 minutes at 4°C, supernatant containing the cytoplasmic fraction was removed, and cell pellet containing the nuclei was resuspended in three pellet volume of buffer C (0.5M NaCl, 20mM HEPES pH7.9, 25% glycerol, 1.5mM MgCl2, 0.2mM EDTA, 0.5mM DTT added before use, oComplete EDTA-free protease inhibitors), and placed over a 10% sucrose cushion made in buffer C, and spun down at 3000g for 5 minutes at 4°C. Supernatant was removed and the pellet was resuspended in buffer

C, equivalent of 1mL per 1 billion starting cells. Nuclei were dounced in a Beckman 7mL dounce with a “tight” pestle for 20 strokes. Lysed nuclei were rotated at 4°C for 30 minutes, and then spun down at 20,000g for 10 minutes at 4°C. The supernatant was the soluble nuclear fraction was dialyzed in buffer D (20mM HEPES pH7.9, 20% glycerol, 0.1M KCl, 0.2mM EDTA, 0.5mM DTT added before use, and oComplete EDTA-free protease inhibitors) using Slide-A-Lyzer dialysis casettes with a 3.5kD molecule weight cut-off for 6 hours with two buffer exchanges. Protein concentration of the nuclear extract was determined with a Qubit protein assay kit according to the manufacturer’s instructions. Dialyzed nuclear extract was snap frozen in liquid nitrogen and stored at -80°C until use.

DNA-affinity purification and on-bead digest

50ul of DNA-immobilized beads were mixed with 400µg of nuclear extract and 1200ng sheared salmon sperm DNA in Axygen 1.5mL tubes. Reactions were incubated at room temperature for 40 minutes with rotation. Beads were then magnetically pelleted, washed once with buffer DBB (supplemented with 0.5%NP-40), and resuspended in DBB supplamented 0.75% formaldehyde for 10 minutes at room temperature with rotation. Beads were resuspended in 50 µl of 100mM ammonium bicarbonate. 600ng of Lys-C (Wako) was added to the beads and digests were incubated at 37°C for 4 hours in a thermoblock with shaking at 800rpm. Beads were magnetically pelleted, and the supernatant was transferred to a new 0.6mL Axygen tube. Samples were incubated with 6µl of a 6.25mM TCEP-HCl solution (Sigma) at 60°C for 30 minutes in a thermoblock with rotation at 400rpm. Next, 6µl of 40mM MMTS was added and incubated for 30 minutes in the dark. Finally, 600ng of trypsin gold (Promega) was added and digests were incubated at 37°C overnight. Digests were stopped with 10µl of 10% TFA solution. 30% of the reaction volume was used for Nano LC-MS/MS analysis. Results from the promoter DNA-affinity purification mass spectrometry are listed in supplementary table 1.

Nano LC-MS/MS Analysis for DNA-affinity purification

An UltiMate 3000 RSLC nano HPLC system (Thermo Fisher Scientific) coupled to a Q Exactive HF-X equipped with an Easy-Spray ion source (Thermo Fisher Scientific) or an Exploris 480 mass spectrometer equipped with a Nanospray Flex ion source

(Thermo Fisher Scientific) was used. Peptides were loaded onto a trap column (PepMap Acclaim C18, 5 mm × 300 µm ID, 5 µm particles, 100 Å pore size, Thermo Fisher Scientific) at a flow rate of 25 µl/min using 0.1% TFA as mobile phase. After 10 min, the trap column was switched in line with the analytical column (PepMap Acclaim C18, 500 mm × 75 µm ID, 2 µm, 100 Å, Thermo Fisher Scientific). Peptides were eluted using a flow rate of 230 nl/min, and a binary linear 3h gradient, respectively 225 min.

The gradient started with the mobile phases 98% A (0.1% formic acid in water) and 2% B (80% acetonitrile, 0.1% formic acid), increased to 35% B over the next 180 min, followed by a steep gradient to 90%B in 5 min, stayed there for 5 min and ramped down in 2 min to the starting conditions of 98% A and 2% B for equilibration at 30°C.

TFIIA immunoprecipitation

Drosophila S2 cells endogenously tagged with an AID-3xFLAG were used for the bait, while the parental background cells only expression the osTir ligase were used as a control immunoprecipitation. Lysates were generated from 500 million cells. Cells were washed in PBS and pelleted by centrifugation. Cell pellet was resuspended in 10mL of hypotonic swelling buffer (10mM Tris pH7.5, 2mM $MgCl_2$, 3mM $CaCl_2$, protease inhibitors) and incubated for 15 minutes at 4°C. Cells were centrifuged for 10 minutes at 700g and at 4°C. Cells were resuspended in 10mL of GRO lysis buffer (10mM Tris pH7.5, 2mM $MgCl_2$, 3mM $CaCl_2$, 0.5% NP-40, 10% glycerol, 1mM DTT, protease inhibitors) and rotated for 30 minutes at 4°C. Nuclei were centrifuged at 700g and at 4°C. Supernatant was removed and nuclei were resuspended in 1mL of IP lysis buffer (100mM NaCl, 20mM HEPES pH7.6, 2mM $MgCl_2$, 0.25% NP-40, 0.3% Tirton X-100, 10% glycerol) and rotated for 30 minutes at 4°C. Lysed nuclei were centrifuged for 5 minutes at 20000g at 4°C. The supernatant containing the soluble nucleoplasm was kept. Whie the chromatin pellet was resuspended in a 300mM NaCl IP lysis buffer (300mM NaCl, 20mM HEPES pH7.6, 2mM $MgCl_2$, 0.25% NP-40, 0.3% Tirton X-100, 10% glycerol) and sonicated Diagenode Bioruptor sonicator: 10 min (30 sec on/30 sec off) at low intensity. The sheared chromatin was centrifuged as before and the soluble supernatant was removed and mixed with the soluble nu-

cleoplasmic fraction. The resulting mixture was centrifuged again for 5 minutes at 20000g at 4°C to remove insoluble proteins. Anti-FLAG M2 beads (Sigma-Aldrich) were equilibrated by three 10 minute washes with 150mM NaCl IP lysis buffer, and resuspended back in their original volume. Immunoprecipitation reactions were set up with 50ul of Anti-FLAG beads and 1mg of the nuclear lysates overnight with rotation at 4°C. Immunoprecipitation reactions were magnetically pelleted and washed with 150mM IP lysis buffer three times, 10 minutes each with rotation at 4°C. Next, to remove detergent, the reactions were washed 4 times, 10 minutes each at 4°C with a no-detergent buffer (130mM NaCl, 20mM Tris pH7.5). Reactions were resuspended in 50µl of 100mM ammonium bicarbonate and on-bead tryptic digest was carried out as described in the DNA-affinity purification and on-bead digest section. Results of the TFIIA-L immunoprecipitation are listed in supplementary table 2.

Nano LC-MS/MS Analysis for TFIIA Immunoprecipitation

A Q Exactive HF-X mass spectrometer was operated in data-dependent mode, using a full scan (m/z range 380-1500, nominal resolution of 60,000, target value 1E6) followed by MS/MS scans of the 10 most abundant ions. MS/MS spectra were acquired using normalized collision energy of 27, isolation width of 1.4 m/z, resolution of 30.000, target value of 1E5, maximum fill time 105ms. Precursor ions selected for fragmentation (include charge states 2-6) were put on a dynamic exclusion list for 60 s. Additionally, the minimum AGC target was set to 5E3 and intensity threshold was calculated to be 4.8E4. The peptide match feature was set to preferred and the exclude isotopes feature was enabled.

LC-MS/MS analysis for TFIIA immunoprecipitation

The Orbitrap Exploris 480 mass spectrometer (Thermo Fisher Scientific), was operated in data-dependent mode, performing a full scan (m/z range 380-1200, resolution 60,000, target value 3E6) at 2 different CVs (-50, -70), followed each by MS/MS scans of the 10 most abundant ions. MS/MS spectra were acquired using a collision energy of 30, isolation width of 1.0 m/z, resolution of 45.000, the target value of 1E5 and intensity threshold of 2E4 and fixed first mass of m/z=120. Precursor ions selected for fragmentation (include charge state 2-5) were excluded for 30 s. The pep-

tide match feature was set to preferred and the exclude isotopes feature was enabled.

Mass-Spectrometry Data Processing

For peptide identification, the RAW-files were loaded into Proteome Discoverer (version 2.5.0.400, Thermo Fisher Scientific). All hereby created MS/MS spectra were searched using MSAmanda v2.0.0.16129 (Dorfer V. et al., J. Proteome Res. 2014 Aug 1;13(8):3679-84). RAW-files were searched in 2 steps: First, against the drosophila database called dmel-all-translation-r6.34.fasta (Flybase.org, 22,226 sequences; 20,310,919 residues), or against an earlier version dmel-all-translation-r6.17.fasta (21,994 sequences; 20,118,942 residues) / a small custom drosophila database (107 sequences; 61,976 residues), each case supplemented with common contaminants, using the following search parameters: The peptide mass tolerance was set to ±5 ppm and the fragment mass tolerance to ±15 ppm (HF-X) or to ±6 ppm (Exploris). The maximal number of missed cleavages was set to 2, using tryptic specificity with no proline restriction. Beta-methylthiolation on cysteine was set as a fixed modification, oxidation on methionine was set as a variable modification, the minimum peptide length was set to 7 amino acids. The result was filtered to 1 % FDR on protein level and was used to generate a smaller sub-database for further processing. As a second step, the RAW-files were searched against the created sub-database using the same settings as above plus the following search parameters: Deamidation on asparagine and glutamine were set as variable modifications. In some data sets acetylation on lysine, phosphorylation on serine, threonine and tyrosine, methylation on lysine and arginine, di-methylation on lysine and arginine, trimethylation on lysine, ubiquitinylation residue on lysine, biotinylation on lysine, formylation on lysine were set as additional variable modifications. The localization of the post-translational modification sites within the peptides was performed with the tool ptmRS, based on the tool phosphoRS (Taus et al., 2011). Peptide areas were quantified using the in-house-developed tool apQuant (Doblmann et al., 2018). Proteins were quantified by summing unique and razor peptides. Protein-abundances-normalization was done using sum normalization. Statistical significance of differentially expressed proteins was determined using limma (Smyth, 2004).

PRO-seq

PRO-seq was performed according to (Mahat et al., 2016) with the following modifications. 10 million Drosophila Schneider S2 cells were used for each replicate, spiked in with 1% human HCT116 cells. Cells were harvested by centrifugation and cells were permeabilized with cell permeabilization buffer (10mM tris Ph 7.5, 300mM sucrose, 10mM $CaCl_2$, 5mM $MgCl_2$, 1mM EGTA, 0.05% tween-20, 0.1% NP-40, 0.5mM DTT, supplemented with protease inhibitors). Permeabilization was carried by resuspending the cells in 10mM of permeabilization buffer and spinning down the cells for a total of three buffer exchanges. Nuclei were resuspended in 100µl of storage buffer (10mM tris pH 7.5, 25% glycerol, 5mM $MgCl_2$,0.1mM EDTA and 5mM DTT) and snap frozen in liquid nitrogen for later use, or immediately proceeded to the run-on reaction. Nuclear transcription run-on was carried by adding 100µl of a 2x run-on buffer (10mM tris pH8, 5mM MgC_2, 1mM DTT, 300mM KCl, 0.25mM ATP, 0.25mM GTP, 0.05mM Biotin-11-CTP, 0.05mM Biotin-11-UTP, 0.8U/µl murine RNase inhibitor, 1% sarkosyl) and incubated at 30C for 3 minutes. Reaction was terminated by adding 500ul Trizol-LS. Extraction was performed by adding 130µl of chloroform, after vortexing and centrifugation the aquesous fraction was kept and precipitated with 2.5 volumes of 100% ethanol and 1µl of glycoblue. The pellet was washed with 80% ethanol, air dried and resuspended in 50µl of water. RNA was denatured at 65C for 40 seconds before base hydrolysis with 5µl 1N NaOH for 15 minutes. Hydrolysis was quenched with 25 µl of 1M tris-HCl pH6.8. Samples were purified on a Bio-Rad P30 column. Biotinylated nascent RNA was recovered by incubating with 50µl of M280 streptavidin beads for 30 minutes at room temperature with rotation. Beads were washed twice each with high salt buffer (2M NaCl, 50mM Tris pH 7.5, 0.5% Tirton X-100) and binding buffer (300mM NaCl, 10mM Tris pH 7.5, 0.1% Tirton X-100) and once with low salt buffer (5mM Tris pH 7.5, 0.1% Tirton X-100). RNA was extracted off the beads using Trizol and cleaned on a Direct-zol column (Zymo). RNA was eluted from the column using 5 µl the 3' RNA linker. Overnight ligation at 16°C was performed with T4 RNA ligase I. The following day biotinylated RNA was recovered with 50µl of M280 streptavidin beads for 30 minutes at room temperature and washed as described. The RNA was treated with CapCLIP Pyrophosphatase (Biozyme) on the beads for 1 hour at 37°C, followed by T4 polynucleotide kinase (NEB) for 1 hour at 37°C. Beads were washed as described and an on-bead ligation was set up with T4 RNA ligase I and the 5' RNA linker at room tem-

perature with rotation at 4 hours. Next, the beads were washed as described and the RNA was extracted off the beads with 300µl Trizol and purified on a Direct-zol column, eluted in water. Eluted RNA was used for reverse transcription with Superscript III Reverse Transcriptase (Thermo) according to the manufacturer's recommendations. Half of the reverse transcription reaction was used for amplification with a KAP real-time PCR mixture (KAPA Biosystems) using the Illumina Truseq small RNA library amplification kit primers. Libraries were amplified in 8-12 cycles. Primer dimers were removed from the libraries with AMPURE beads and sent for next-generation sequencing.

PRO-seq data mapping

PRO-seq libraries were sequenced to a depth of 3.8 - 38.9 million reads using single-end sequencing and read length of 50 bp. We used unique molecular identifiers (UMIs) to distinguish between PCR duplicated identical reads and reads stemming from distinct RNA molecules with an identical sequence. The latter will have identical sequences but different UMIs and therefore allows more accurate quantification of transcripts. RNA oligos containing UMIs of 8-10 nt in length were ligated to the 3' end of all reads before PCR amplification and then computationally removed to prevent interference during genome alignment. Cutadapt 1.18 (Martin et al., 2011) with default options was used to find and trim the sequencing adapter at the 3' end and filtered for reads ≥ 10 nts long. Only after read alignment we corrected for PCR duplicated transcripts and to more accurately quantified transcripts: reads containing the same sequence and reads aligning to the same genomic position were collapsed to unique UMIs.

To align reads, we generated an artificial genome containing sequences for tRNAs and rRNAs only, which allows for noise reduction of short reads aligning to multiple positions. Next, all unmapped reads were captured using samtools version 1.9 (Li et al., 2009) with -f 4 option, which were then aligned to the *D. melanogaster* reference genome BDGP R5/dm3. Following this, reads not aligning to the dm3 genome were aligned to the *H. sapiens* reference genome GRCh37/hg19 (used as spike-in). For genome alignment we used bowtie version 1.2.2 (Langmead et al., 2009) allowing

two mismatches (-v 2). For alignment to the artificial genome we allowed reads having up to 1000 reportable alignments, but reporting only the best alignment (-m 1000 --best --strata) to meet the highly repetitive and conserved nature of tRNAs and rRNAs. Alignment to the reference genomes was run allowing only reads aligning uniquely (-m 1).

We generated an artificial genome containing the ribosomal RNA primary transcript CR45847 (http://flybase.org/reports/FBgn0267507), all annotated tRNA genes from Dmel 5.57 and tRNAs predicted from Genomic tRNA database, published 2009, http://lowelab.ucsc.edu/GtRNAdb/ (accessed August 17th, 2020; http://lowelab.ucsc.edu/download/tRNAs/eukaryotic-tRNAs.fa.gz). We used R packages GenomicRanges 1.34.0 (Lawrence et al., 2013), Biostrings 2.50.2 (https://bioconductor.org/packages/Biostrings) and BSgenome.Dmelanogaster.UCSC.dm3 1.4.0 (Team TBD (2014). *BSgenome.Hsapiens.UCSC.hg17: Full genome sequences for Homo sapiens (UCSC version hg17)*. R package version 1.3.1000.

Since application of the usual PRO-seq protocol delivers reads corresponding to the reversed complement of the nascent RNA, the reads aligning to the minus strand originated from transcripts with the sequence on the plus strand and vice versa. Additionally, only the end of the transcript where RNA Pol II was actively transcribing was included for the downstream analysis. Reads were switched and shortened accordingly using the bigBedtoBed utility (Kent et al., 2010).

ChIP-seq and ChIP-exo data analysis

ChIP-seq and ChIP-exo data sets were taken from (Baumann and Gilmour, 2017; Gurudatta et al., 2013; Shao and Zeitlinger, 2017). Coverage was calculated over a 2-kb window centered on the TSS of each promoter type. Data was normalized across the different promoter types, and then across the different ChIP data sets.

Promoter motif annotations

We generated an R table in version 3.5.3 (R Core Team, 2019) containing transcripts of all protein-coding genes and corrected their transcription starting site with hits supported by CAGE experiments (Brown et al., 2014). First, TSSs were correct-

ed by CAGE signal from S2 cells downloaded from modENCODE 5331 that lie within a window of ±250 bps. If no hit was found, CAGE signals from mixed embryos or a developmental timecourse from modENCODE 5338-5348, 5350 and 5351 were used within the same window. If the TSS was left unsupported we repeated this using a ±500 bp window or kept the annotated TSS. We kept the longest transcript per unique TSS. We used the R packages CAGEr 1.24.0 (Haberle et al., 2014) and GenomicRanges 1.34.0(Lawrence et al., 2013). Additionally, all transcripts were annotated with known *D. melanogaster* core promoter motifs as described in a previous study (Haberle et al., 2019) with small changes regarding match thresholds for TATAbox to 90%, DPE to 98% DRE to 98% and Ohler6 to 97%. TCT motif was further limited to ribosomal proteins.

Generation of browser tracks of PRO-seq data

For visualization of PRO-seq data we converted bigBed files to bigWig files using kentUtils bigBedToBed utility (Kent et al., 2010), normalized by the number of reads aligned to dm3 (and considered number of reads aligned to hg19 for TFIIA samples) and calculated the coverage using genomeCoverageBed from bedtools 2.27.1 (Quinlan and Hall, 2010) before converting to a bigWig file using KentUtils wigToBigWig utility. BigWig files were visualized with the UCSC Genome Browser (Kent et al., 2010).

Differential expression

Differential expression was calculated using the DESeq function from the DESeq2 package v.1.30.1 (Love et al., 2014) providing the normalization factors as sizeFactors. Normalization factors were calculated based on the number of reads aligned to *D. melanogaster* reference genome and for TFIIA quantified spike-in reads were additionally considered. We used Benjamini-Hochberg adjusted p-values to determine significantly deregulated transcripts.

STAP-seq data analysis of initiation events

Cofactor recruitment STAP-seq data from (Haberle et al., 2019) was analyzed at single nucleotide resolution counting unique transcripts initiated at each position in each tested oligo. The dominants TSS was determined as the position with the high-

est count, and the relative count was calculated by dividing the count at the dominant TSS with the total count for each oligo. To determine the number of activated TSSs in each oligo, the count at each position was divided by the count at the dominant TSS, and only the positions with a ratio of more than 20% were counted as activated TSSs.

Defining Drosophila melanogaster promoter types

A set of ~17,000 promoters of protein coding genes was classified into 9 groups based on PWM scores for the different CP motifs. The data was clustered with k-means to get the representative groups defined by the occurrence of specific motifs or motif combinations.

Aligning CAGE data to promoter motif positions and +1 nucleosome centers

For the above defined promoter groups the positions of the defining CP motifs were determined relative to the dominant CAGE TSS (if they occurred within +/- 120 bp). Only promoters with a single occurrence of each motif were considered, and the position of the motif was used as a reference point to generate average plots of CAGE data. MNase-seq data from(Chereji et al., 2016), CAGE data from mixed embryos (Hoskins et al., 2011)

MNase-seq data was used to determine the position of the +1 nucleosome by taking the centers of MNase fragments between 100 and 200 bp long, calculating the coverage of such centers, and determining the position with the highest coverage in the region 150 bp downstream of the dominant CAGE TSS. These +1 nucleosome centers were used as a reference to generate average plots of CAGE data for each promoter group.

Luciferase assay

Drosophila Schneider S2 cells were plates in 96-well plates, $1x10^5$ cells per well. Cells were transfected with 100ng of luciferase plasmid contaiing a DRE promoter or negative control sequence upstream of the luciferase gene, and 100ng of a plasmid containing renilla luciferase as a transfection efficiency normalization control using Lipofectamine 2000. Cells were lysed 48 hours after transfection with 50µl passive lysis buffer for 30 minutes at room temperature with shaking. Lysates were further

diluted 10 fold in passive lysis buffer. 10µl of the diluted lysate was placed in 96-well plates compatible with luminescence read-out and measured with the Promega dual-luciferase assay kit according to the manufacturer's recommendation on a BioTek Synergy H1 plate reader.

Quantification and statistical analysis

Information regarding statistical testing for individual experiments is described in the figure legends, including statistical tests used, number of replicates, and number of observations.

Supplementary Figure Legends

Figure S1. Related to Figure 1. DNA affinity purifications uncover differentially bound proteins at functionally distinct promoters

A. Rank plot of protein binding enrichment on TATA and DRE promoters over the control DNA pool from the DNA-purification mass spectrometry assay. Highlighted proteins are the Pol II PIC components and the DRE binding factor DREF.

Figure S2. Related to Figure 2. Developmental and housekeeping promoters bind different sets of proteins and GTFs

A. Elution fractions from the DNA-purification assay with 121bp or 350bp DRE promoters and length matched negative controls were performed with a nuclear extract expressing DREF-AID-3xFLAG tag and blotted for an anti-FLAG antibody.
B. Luciferase assay performed with DRE promoter fragments that are 121bp or 450bp in length and a negative control region which is 450bp in length. Firefly luciferase values were normalized to co-transfected renilla luciferase values.
C. ChIP-seq and ChIP-exo tracks of GTFs and housekeeping promoter binders.

Figure S3. Related to Figure 3. TBP and DREF are required by distinct sets of promoters

A. Mass-spectrometric quantification of TBP protein abundance normalized to the largest RNA Pol II subunits RpII215. Samples were taken from the N-terminally tagged TBP-AID-3xFLAG treated with 6 hours with 500uM auxin, and from the parental cell line expressing the OsTir ligase.
B. Pearson correlation of PRO-seq signal along the promoter and gene body region of all protein-coding transcripts using library-normalized reads between biological replicates. Correlation coefficient displayed.
C. The number of DRE and TATA-Box expressed promoters in each of the DREF and TBP AID tagged cell lines. P-value calculated with ___ indicate down-regulation of TATA-Box or DRE promoters compared with all expressed promoters.

Figure S4. Related to Figure 4. Preferential promoter subtype bound proteins *in vitro* are required by distinct sets of promoters *in vivo*

A. Volcano plots of DNA-purification mass spectrometry showing enrichment of proteins on Ohler1/6, DPE and TCT promoters over negative control DNA. Selected proteins were highlighted.
B. Pearson correlation of PRO-seq signal along the promoter and gene body region of all protein-coding transcripts using library-normalized reads between biological replicates in the CG14711, CG3847 and Chromator AID cell lines. Pearson correlation coefficient displayed.
C. MA-plots of PRO-seq data from CG7372 tagged cell line at 6 hours of auxin treatment. Highlighted genes indicate significantly upregulated or down-regulated genes (FDR<0.05). Bottom plots: Fisher test for odds ratio for promoter motif enrichment from the highlighted down-regulated promoters in the MA-plots.
D. MA-plots of PRO-seq data from the double Ibf1 + Ibf2 tagged cell line at 6 hours of auxin treatment. Highlighted genes indicate significantly upregulated or downregulated genes (FDR<0.05). Bottom plots: Fisher test for odds ratio for promoter motif enrichment from the highlighted down-regulated promoters in the MA-plots.

E. Volcano plot of DNA-affinity purification mass spectrometry data of TATA-Box promoter protein enrichment over negative control DNA with CG7372 highlighted.
F. Fisher's exact test for odds ratio for promoter motif enrichment from the up-regulated promoters from the Chromator-AID PRO-seq experiment.

Figure S5. Related to Figure 5. TFIIA is required by all promoters and is recruited by housekeeping cofactors to housekeeping promoters

A. Growth curve of the parental and TFIIA-L AID tagged cell line treated with 500μM Auxin.
B. Ratio of reads mapping to the spike-in genome in the TFIIA-L and TBP depletion PRO-seq experiment.
C. DNA-affinity purification assay was performed with a 121bp long housekeeping DRE promoter with 4xUAS sites upstream. Initially, a nuclear extract containing a Gal4-DNA biniding domain fusion GFP or GFZF was incubated with the bead-immobilized promoter DNA (left panel). After the incubation, the extract was removed, and the beads were used for a DNA-affinity purification assay with a nuclear extract containing TFIIA-L-AID-3xFLAG as described in the materials and methods. Elution fractions were ran on an SDS-PAGE gel and blotted with a FLAG antibody (right panel).

Figure S6. Related to Figure 6. Housekeeping cofactor recruitment is sufficient to recapitulate dispersed transcription initiation patterns

A. The number of CAGE defined TSSs in each promoter type over a 120+/- bp region. TSS was defined as a position having at least 20% CAGE signal as the dominant TSS in the tested region.
B. Dinucleotides at the -1/+1 position each the dominant and secondary TSSs in each promoter type in a 120+/- bp window.
C. Fold change (log2) of STAP-seq signal upon GFZF or MED25 recruitment over GFP for oligos that are matched for their activation level by either one of both cofactors.
D. Relative CAGE signal per position on all active promoters of the indicated type aligned to the +1 nucleosome centre (point of highest coverage of MNase fragment centres in +1 to +200bp window relative to TSS).

E. Percent of STAP-seq signal at the dominant TSS for activation matched oligos (one activated oligo per gene TSS) for housekeeping and developmental promoters responding which can be activated by both MED25 and GFZF.

F. Histogram showing the number of TSS measured upon GFZF or MED25 recruitment on random regions that are responsive to both cofactors (left). Cumulative plot of the same data (right).

Figure 1

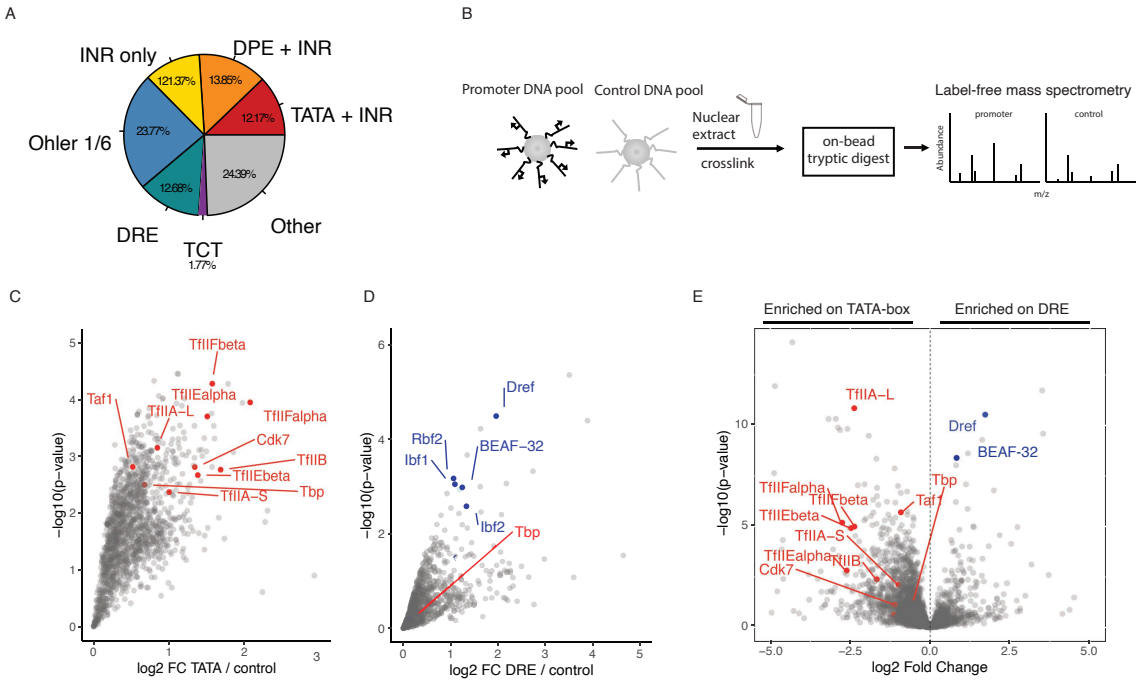

A
INR only
DPE + INR
TATA + INR
Ohler 1/6
DRE
TCT
Other
12.37%
13.85%
12.17%
23.77%
12.68%
24.39%
1.77%
B
Promoter DNA pool
Control DNA pool
Nuclear extract
crosslink
on-bead tryptic digest
Label-free mass spectrometry
promoter
control
Abundance
m/z
C
TfIIFbeta
TfIIEalpha
Taf1
TfIIA−L
TfIIFalpha
Cdk7
TfIIB
TfIIEbeta
TfIIA−S
Tbp
−log10(p−value)
log2 FC TATA / control
D
Dref
Rbf2
Ibf1
BEAF−32
Ibf2
Tbp
−log10(p−value)
log2 FC DRE / control
E
Enriched on TATA-box
Enriched on DRE
TfIIA−L
Dref
BEAF-32
Tbp
TfIIFalpha
TfIIFbeta
Taf1
TfIIEbeta
TfIIA−S
TfIIEalpha
TfIIB
Cdk7
−log10(p−value)
log2 Fold Change

Figure 2

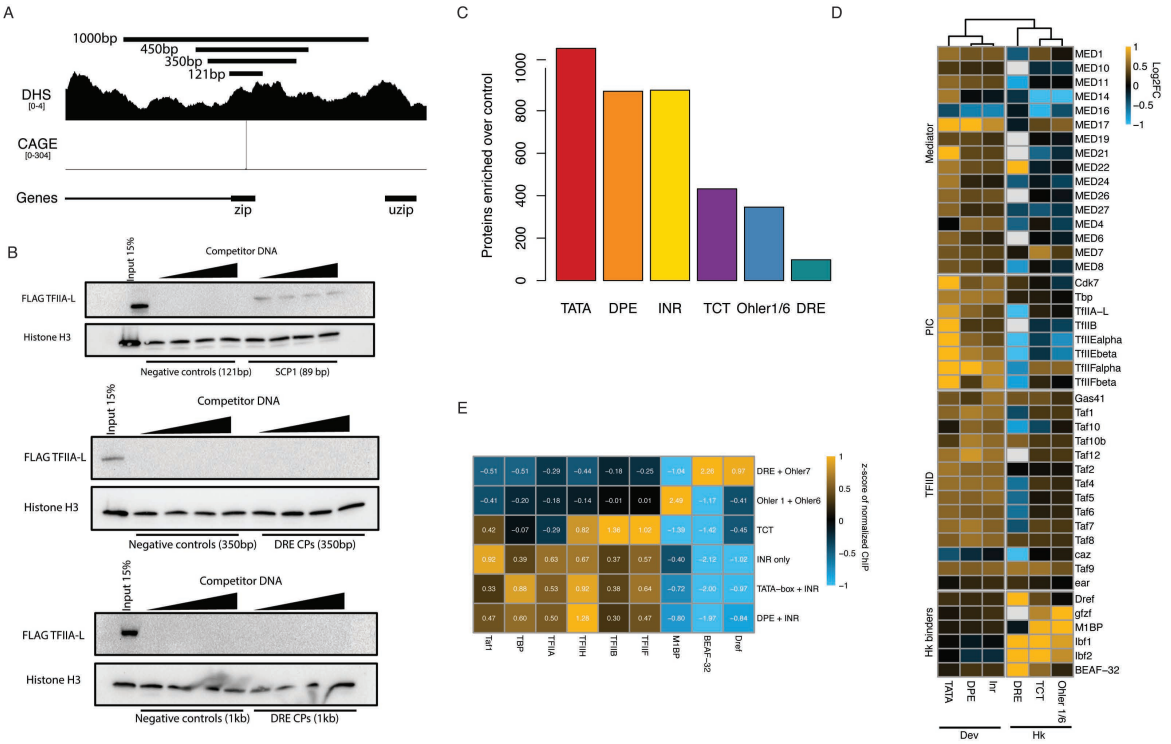

A
1000bp
450bp
350bp
121bp
DHS
[0-4]
CAGE
[0-304]
Genes
zip
uzip
B
Input 15%
Competitor DNA
FLAG TFIIA-L
Histone H3
Negative controls (121bp)
SCP1 (89 bp)
Negative controls (350bp)
DRE CPs (350bp)
Negative controls (1kb)
DRE CPs (1kb)
C
Proteins enriched over control
TATA
DPE
INR
TCT
Ohler1/6
DRE
D
Mediator
PIC
TFIID
Hk binders
MED1
MED10
MED11
MED14
MED16
MED17
MED19
MED21
MED22
MED24
MED26
MED27
MED4
MED6
MED7
MED8
Cdk7
Tbp
TfIIA–L
TfIIB
TfIIEalpha
TfIIEbeta
TfIIFalpha
TfIIFbeta
Gas41
Taf1
Taf10
Taf10b
Taf12
Taf2
Taf4
Taf5
Taf6
Taf7
Taf8
caz
Taf9
ear
Dref
gfzf
M1BP
Ibf1
Ibf2
BEAF–32
Log2FC
Dev
Hk
E
DRE + Ohler7
Ohler 1 + Ohler6
TCT
INR only
TATA–box + INR
DPE + INR
z-score of normalized ChIP

Figure 3

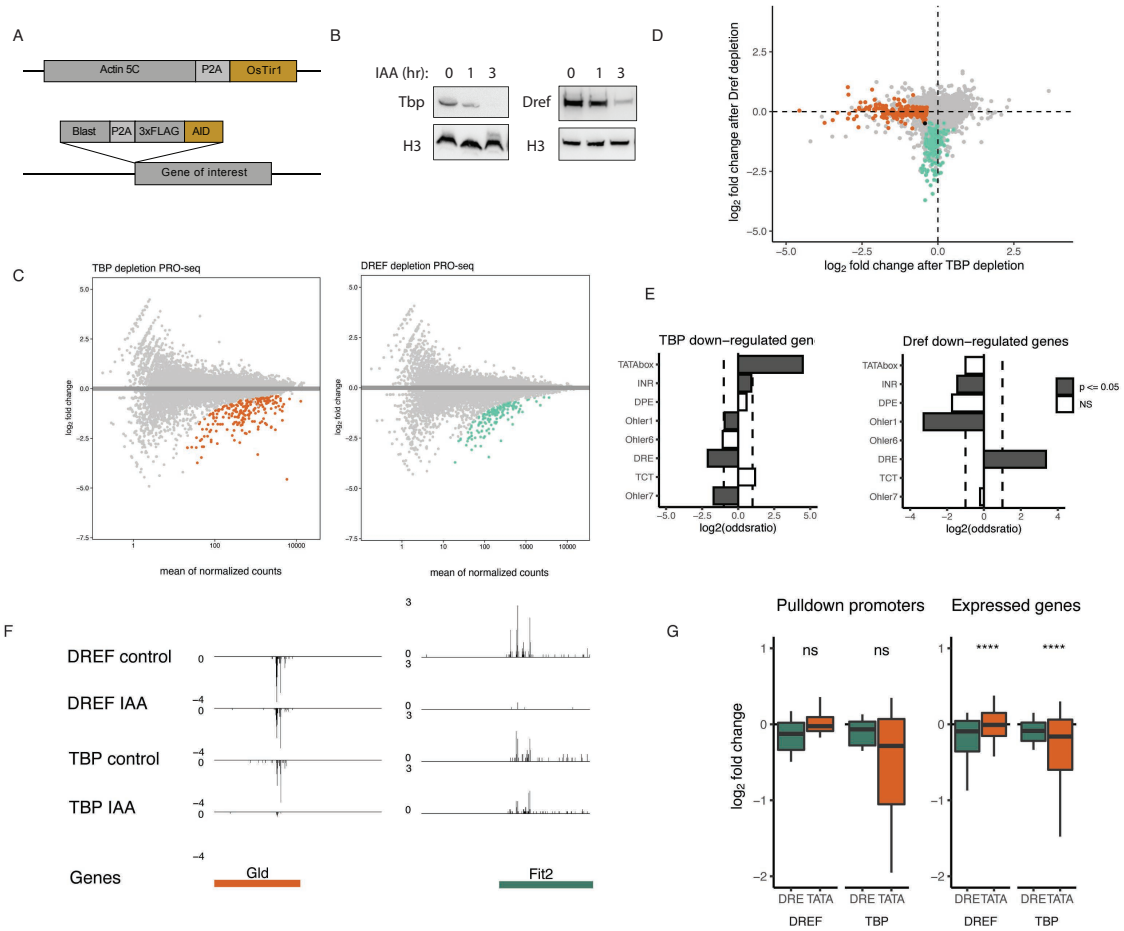

A
Actin 5C
P2A
OsTir1
Blast
P2A
3xFLAG
AID
Gene of interest
B
IAA (hr): 0 1 3
Tbp
H3
0 1 3
Dref
H3
C
TBP depletion PRO-seq
DREF depletion PRO-seq
log2 fold change
mean of normalized counts
mean of normalized counts
D
log2 fold change after Dref depletion
log2 fold change after TBP depletion
E
TBP down−regulated gen
Dref down−regulated genes
TATAbox
INR
DPE
Ohler1
Ohler6
DRE
TCT
Ohler7
log2(oddsratio)
log2(oddsratio)
p <= 0.05
NS
F
DREF control
DREF IAA
TBP control
TBP IAA
Genes
Gld
Fit2
G
Pulldown promoters
Expressed genes
log2 fold change
ns
ns
****
****
DRE TATA
DRE TATA
DRE TATA
DRE TATA
DREF
TBP
DREF
TBP

Figure 4

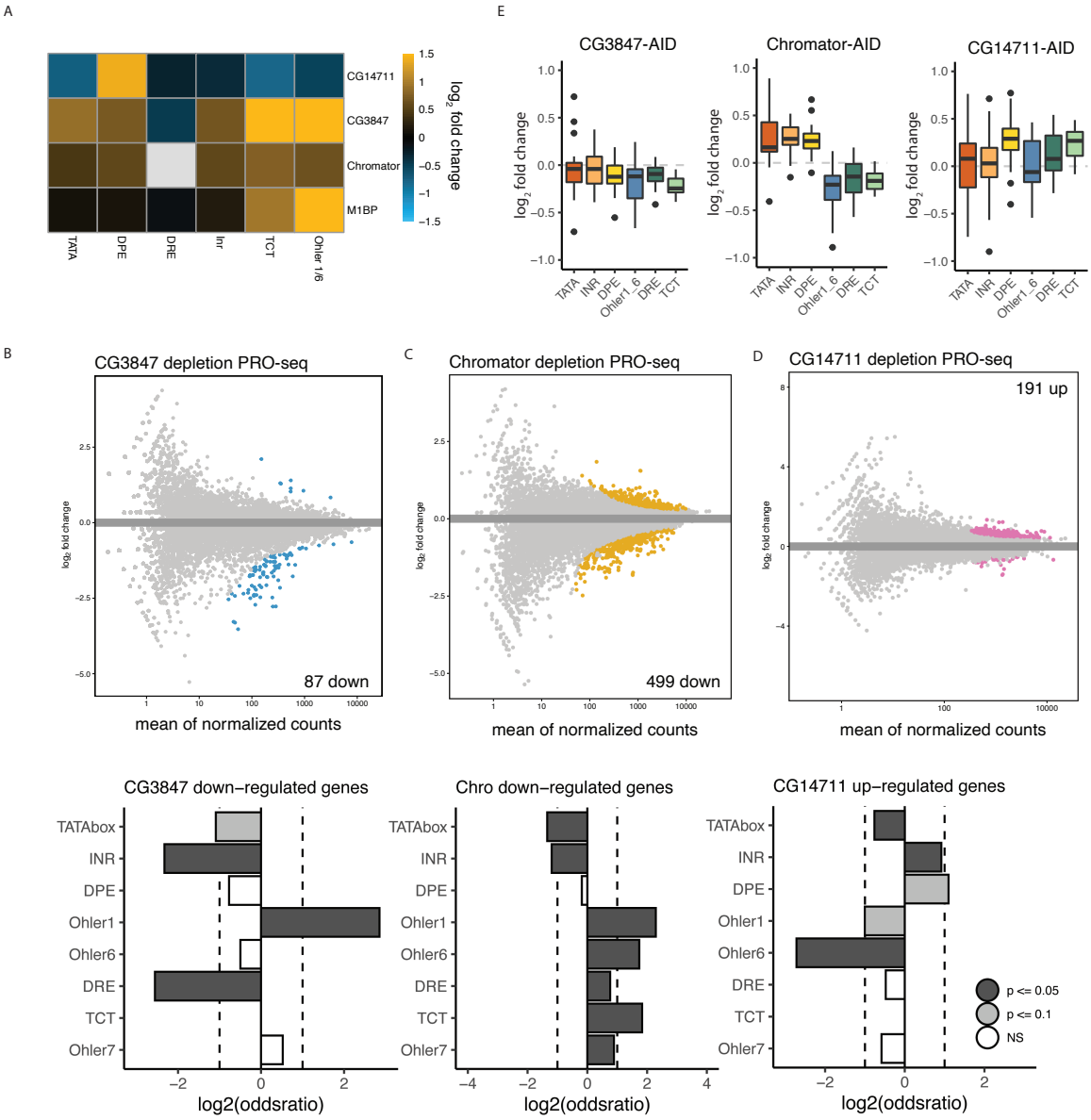

A
CG14711
CG3847
Chromator
M1BP
TATA
DPE
DRE
Inr
TCT
Ohler 1/6
log2 fold change
E
CG3847-AID
Chromator-AID
CG14711-AID
log2 fold change
TATA
INR
DPE
Ohler1_6
DRE
TCT
B
CG3847 depletion PRO-seq
87 down
mean of normalized counts
C
Chromator depletion PRO-seq
499 down
mean of normalized counts
D
CG14711 depletion PRO-seq
191 up
mean of normalized counts
CG3847 down−regulated genes
Chro down−regulated genes
CG14711 up−regulated genes
TATAbox
INR
DPE
Ohler1
Ohler6
DRE
TCT
Ohler7
log2(oddsratio)
p <= 0.05
p <= 0.1
NS

Figure 5

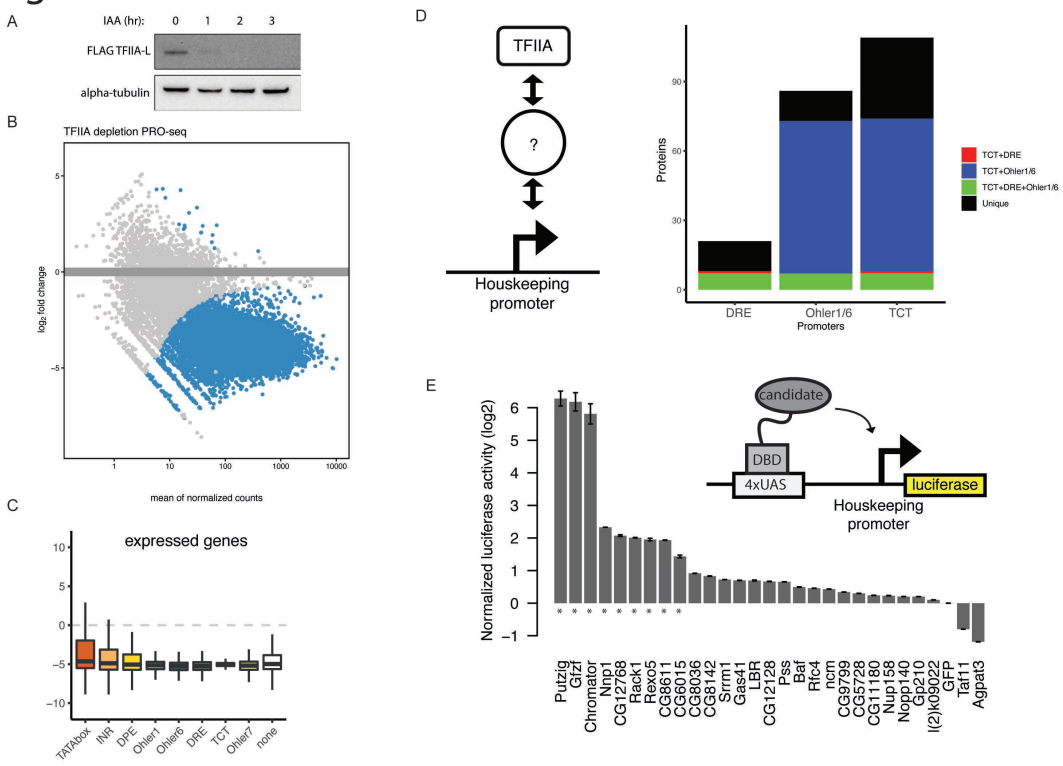
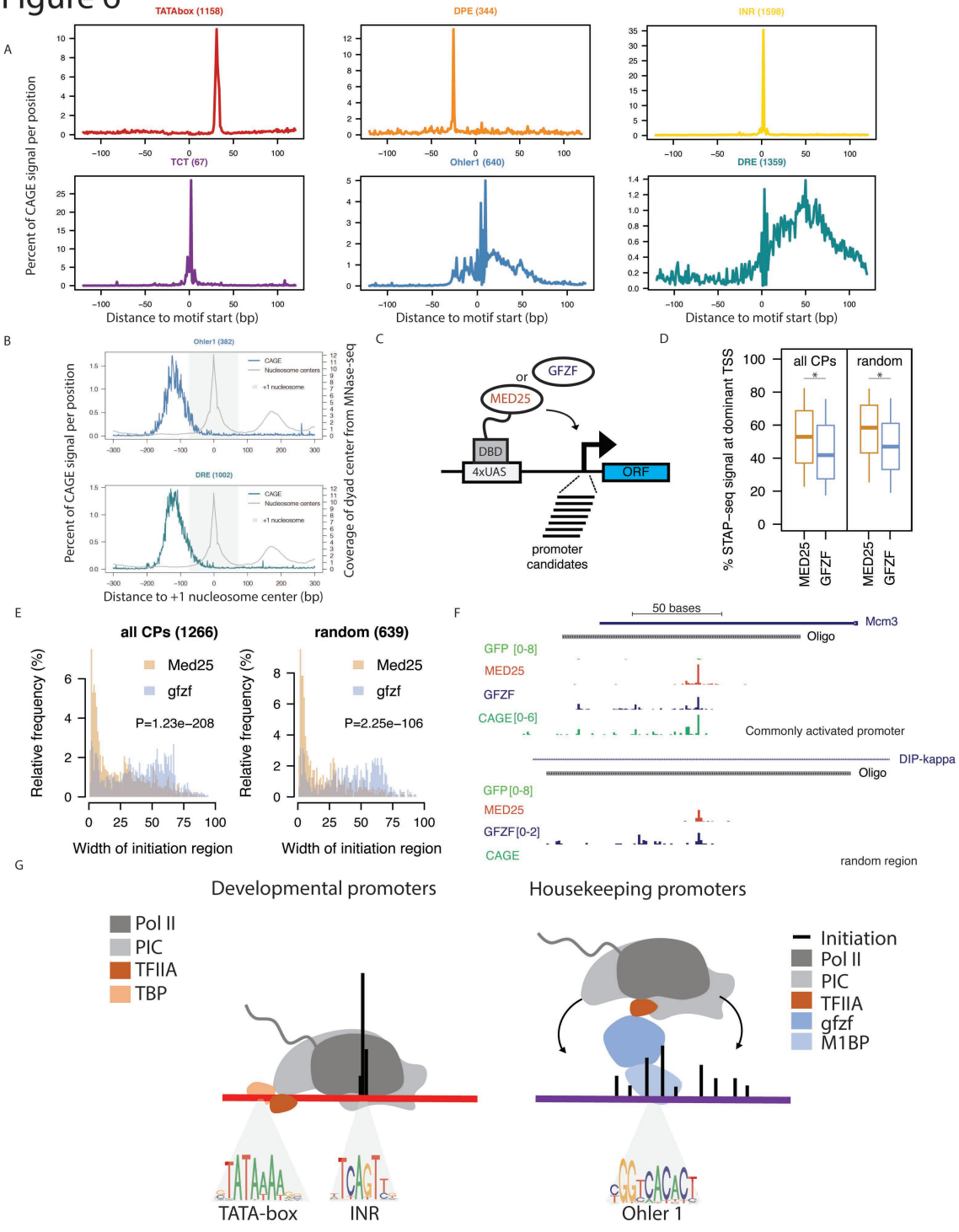
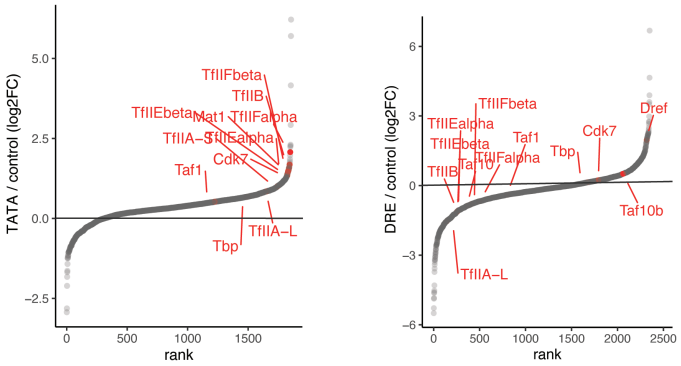

A
IAA (hr): 0 1 2 3
FLAG TFIIA-L
alpha-tubulin
B
TFIIA depletion PRO-seq
log2 fold change
mean of normalized counts
C
expressed genes
TATAbox
INR
DPE
Ohler1
Ohler6
DRE
TCT
Ohler7
none
D
TFIIA
?
Housekeeping promoter
Proteins
DRE
Ohler1/6
TCT
Promoters
TCT+DRE
TCT+Ohler1/6
TCT+DRE+Ohler1/6
Unique
E
candidate
DBD
4xUAS
luciferase
Housekeeping promoter
Normalized luciferase activity (log2)
Putzig
Gfzf
Chromator
Nnp1
CG12768
Rack1
Rexo5
CG8611
CG6015
CG8036
CG8142
Srrm1
Gas41
LBR
CG12128
Pss
Baf
Rfc4
ncm
CG9799
CG5728
CG11180
Nup158
Nopp140
Gp210
l(2)k09022
GFP
Taf11
Agpat3

Figure 6

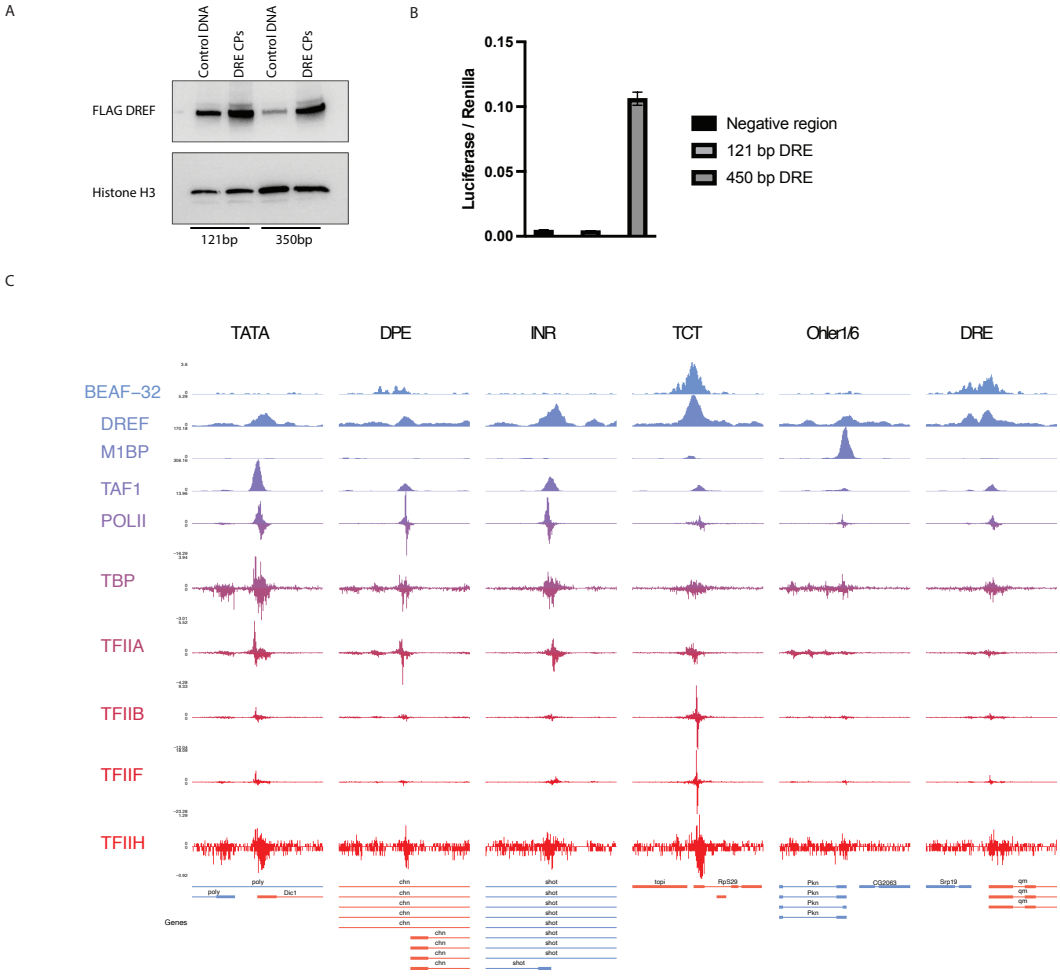

A
TATAbox (1158)
DPE (344)
INR (1598)
TCT (67)
Ohler1 (640)
DRE (1359)
Percent of CAGE signal per position
Distance to motif start (bp)
B
Ohler1 (382)
DRE (1002)
CAGE
Nucleosome centers
+1 nucleosome
Percent of CAGE signal per position
Coverage of dyad center from MNase-seq
Distance to +1 nucleosome center (bp)
C
or
GFZF
MED25
DBD
4xUAS
ORF
promoter candidates
D
% STAP–seq signal at dominant TSS
all CPs
random
MED25
GFZF
E
all CPs (1266)
random (639)
Med25
gfzf
P=1.23e–208
P=2.25e–106
Relative frequency (%)
Width of initiation region
F
50 bases
Mcm3
Oligo
GFP [0-8]
MED25
GFZF
CAGE[0-6]
Commonly activated promoter
DIP-kappa
GFZF[0-2]
CAGE
random region
G
Developmental promoters
Housekeeping promoters
Pol II
PIC
TFIIA
TBP
Initiation
gfzf
M1BP
TATA-box
INR
Ohler 1

Supplementary Figure 1

A

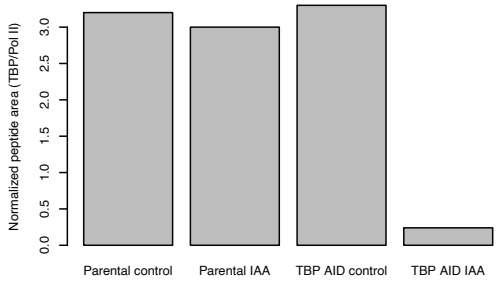


Supplementary Figure 2

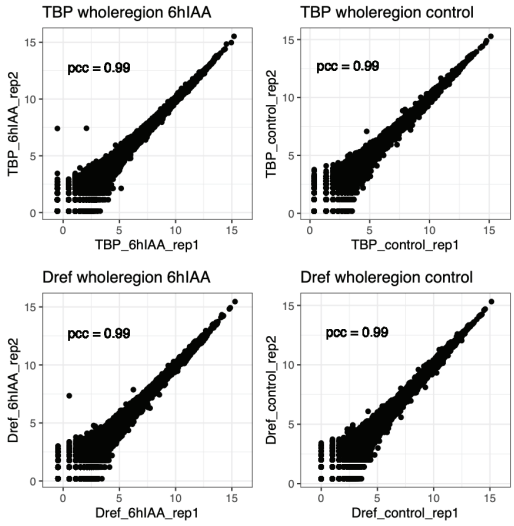


Supplementary Figure 3

A



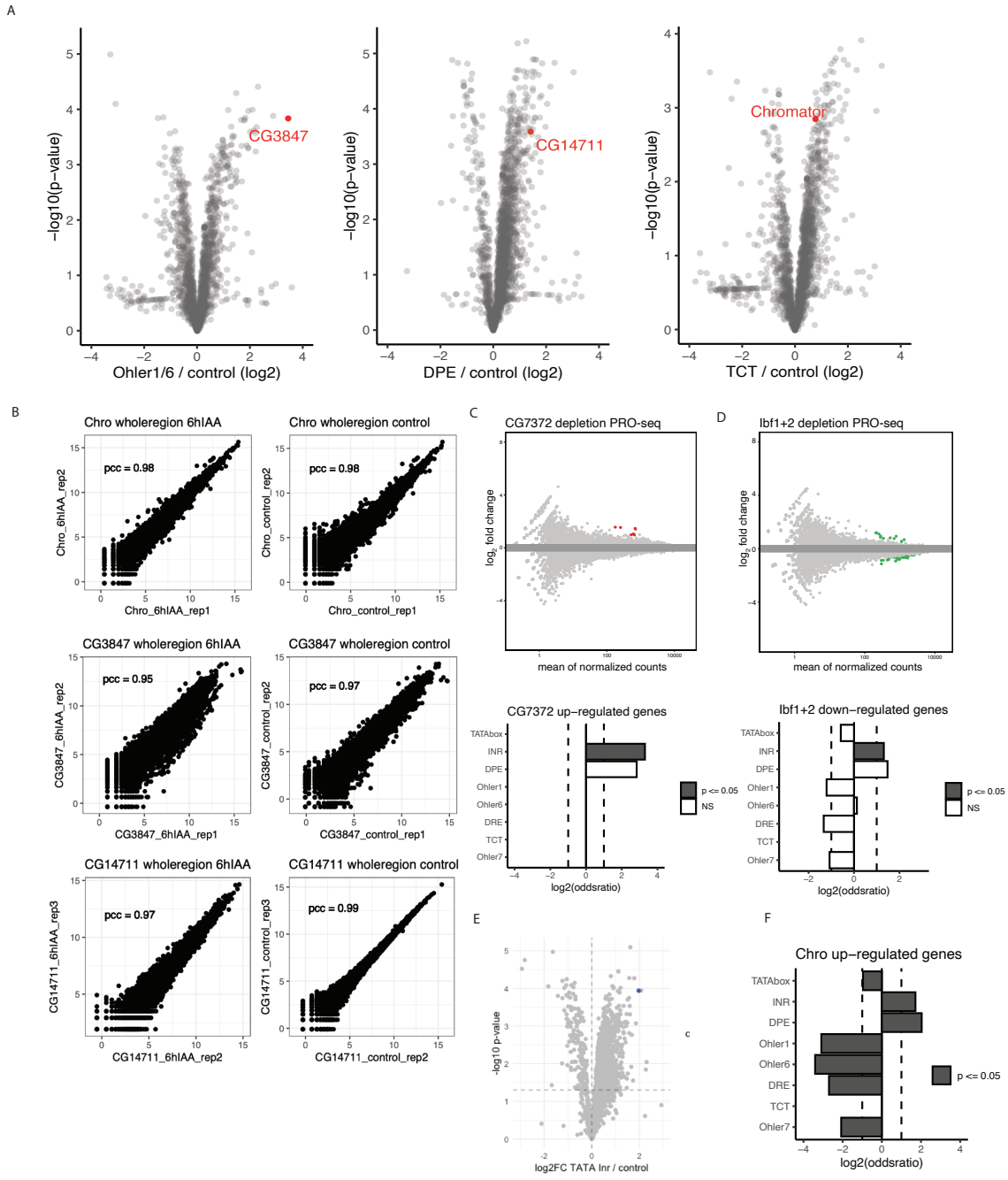
B



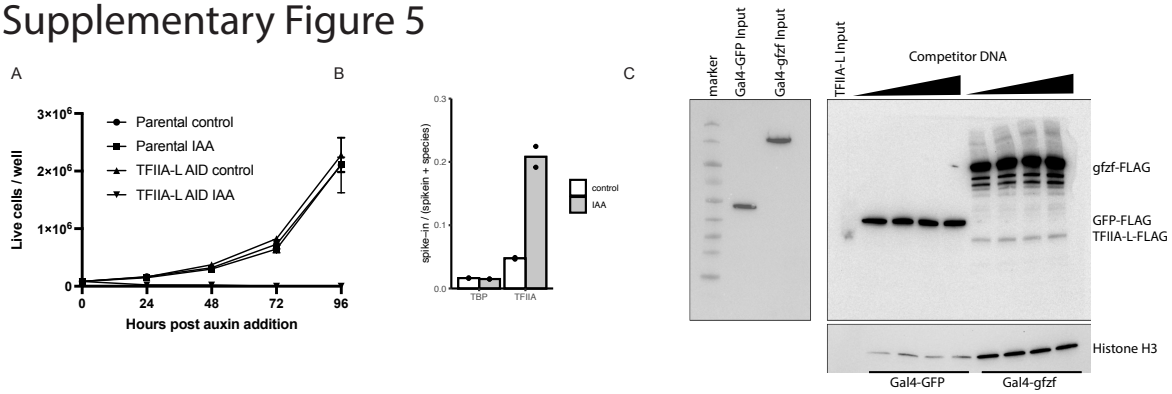
C

	motif	factor	N_expressed	pval_expressed	N_pulldown	pval_pulldown
1	DRE	DREF	1632	1.476829e−28	20	0.05412509
2	DRE	TBP	1640	2.986946e−14	20	0.20959174
3	TATAbox	DREF	741	1.476829e−28	24	0.05412509
4	TATAbox	TBP	747	2.986946e−14	24	0.20959174

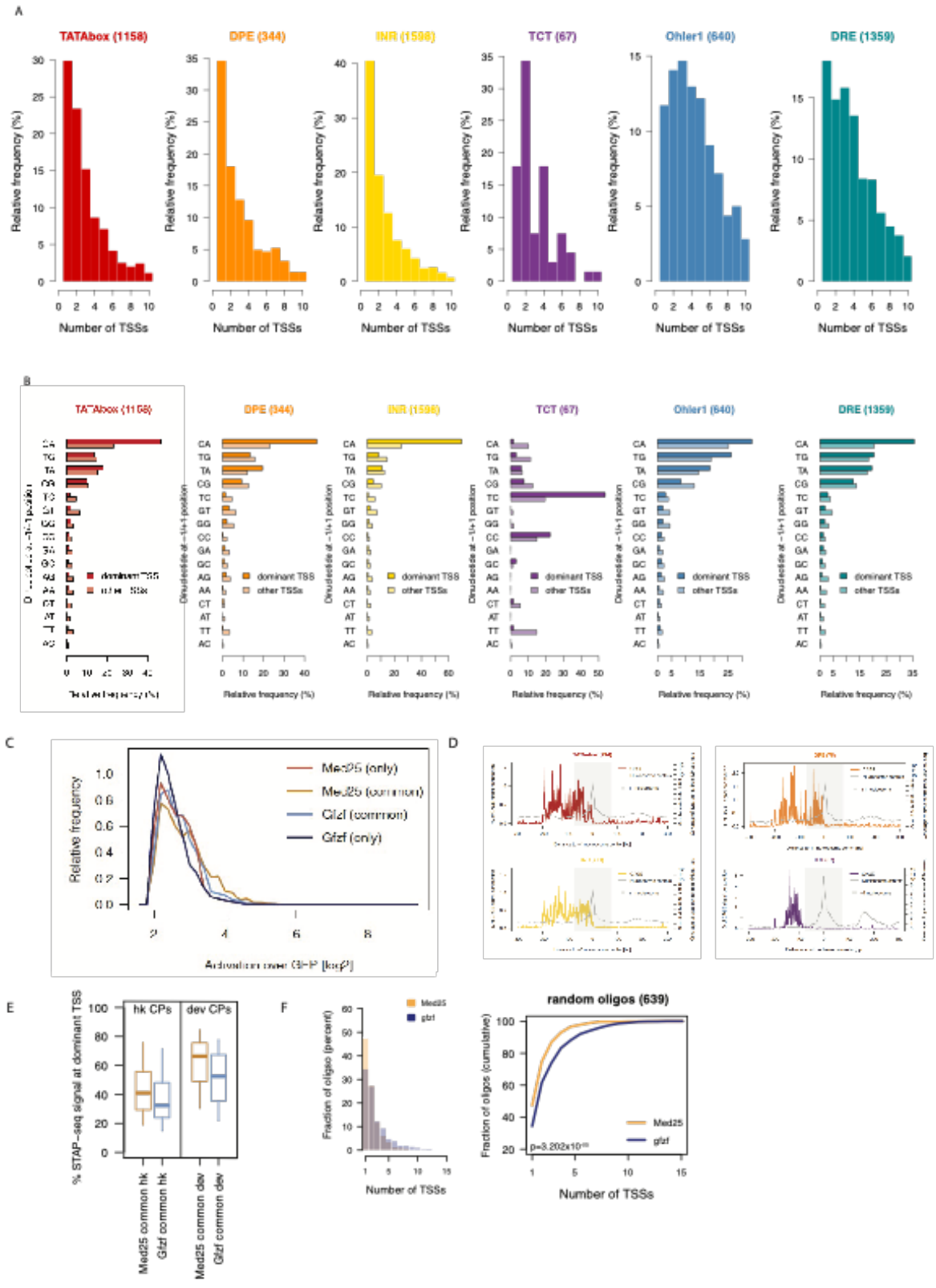
Figure 4 Supplementary



Supplementary Figure 5



Supplementary Figure 6



KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti FLAG	Sigma-Aldrich	Cat#F3165
Secondary anti mouse HRP	Sigma-Aldrich	Cat#12-349
Histone H3	Abcam	Cat#ab1791
Alpha-tubulin	Abcam	Cat#Ab18251
Secondary anti rabbit HRP	Sigma-Aldrich	Cat#12-348
Bacterial and virus strains		
MegaX DH10B T1^{R} Electrocomp^{TM} Cells	Thermo Fisher	Cat#C640003
Biological samples		
Chemicals, peptides, and recombinant proteins		
FastDigest MluI	ThermoFisher	Cat#FD0564
BspQI	NEB	Cat#R0712S
Blasticidin S HCl	ThermoFisher	Cat#R21001
3-Indoleacetic acid	Merck	Cat#I3750
QuickExtract™ DNA Extraction Solution	Lucigen	Cat#QE9059
2x Laemmli Sample Buffer	BioRad	Cat#1610737
EGTA	Merck	Cat#E4378
Biotin-11-CTP	PerkinElmer	Cat#NEL542001EA
Biotin-11-UTP	PerkinElmer	Cat#NEL543001EA
Q5 polymerase high fidelity 2x master mix	NEB	Cat#M0492S
Trizol	ThermoFisher	Cat#15596026
Trizol-LS	ThermoFisher	Cat#10296010
GlycoBlue™ Coprecipitant	ThermoFisher	Cat#AM9515
NTP Set, 100 mM Solution	ThermoFisher	Cat#R0481
N-Lauroylsarcosine sodium salt	Merck	Cat#L5125
Dynabeads™ M-280 Streptavidin	ThermoFisher	Cat#11205D
Cap-CLIP	BioZym	Cat#C-CC15011H
T4 Polynucleotide Kinase	NEB	Cat#M0201S
Murine RNAse Inhibitor	NEB	Cat#M0314L
T4 RNA Ligase	NEB	Cat#M0204L
SuperScript™ III Reverse Transcriptase	ThermoFisher	Cat#18080093
KAPA HiFi HotStart Real-Time Library Amp Kit	Roche	Cat#7959028001
AMPure XP beads	Beckman Coulter	Cat#A63882
Anti-FLAG® M2 Magnetic Beads	Merck	Cat#M8823
Lysyl endopeptidase	Wako Chemicals	Cat#7041
Ammoniumbicarbonate	Sigma-Aldrich	Cat#09830
Tris-(2-carboxyethyl)-phosphin-hydrochloride (TCEP)	Sigma-Aldrich	Cat#646547
S-Methyl-thiomethanesulfonate (MMTS)	Sigma-Aldrich	Cat#64306
Trifluoroacetic acid	Sigma-Aldrich	Cat#T6508
oComplete mini protease inhibitors	Sigma-Aldrich	Cat# 11836170001
Axygen 1.5mL MaxyClear tube	Corning	Cat#MCT-150-A
Axygen 0.6mL MaxyClear tube	Corning	Cat#MCT-060-C-S
Critical commercial assays		
Direct-zol RNA Microprep	Zymo	Cat#R2061
Micro Bio-spin P-30 gel columns	Bio-rad	7326251
Power Blotter Station	ThermoFisher	Cat#PB0010
MaxCyte STX Scalable Transfection System	Maxcyte	NA
4–20% Mini-PROTEAN® TGX™ Precast Protein Gels, 15-well, 15 µl	Bio-Rad	Cat#34561096

Mini-PROTEAN Tetra Vertical Electrophoresis Cell	Bio-Rad	Cat#1658004
Monarch Gel Extraction	NEB	Cat#T1020L
Illumina Truseq small RNA library prep kit	Illumina	Cat#RS-200-0012
Deposited data		
CAGE data	Hoskins et al., 2011	SRX015329
All NGS data from this study	This study	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE181257 accession GSE181257
DNA-affinity purification of promoter DNA	This study	http://proteomecentral.proteomexchange.org/cgi/GetDataset identifier: PXD028090
TFIIA immunoprecipitation	This study	http://proteomecentral.proteomexchange.org/cgi/GetDataset identifier: PXD028094
M1BP ChIP-seq	Baumann and Gilmour, 2017	GSE97841
DREF ChIP-seq	Gurudatta et al., 2013	GSE39664
GTF ChIP-nexus	Shao & Zeitlinger, 2017	GSE85741
MNase-seq data	Gilchrist et al. 2010	GSE22119
Oligonucleotides		
Primers for AID tagging, table 1	This Paper	N/A
5'- /5Phos/rNrNrN rNrNrN rNrNrG rArUrC rGrUrC rGrGrA rCrUrG rUrArG rArArC rUrCrU rGrArA rC/3InvdT/ -3' (3' RNA linker)	IDT	N/A
5- rCrCrU rUrGrG rCrArC rCrCrG rArGrA rArUrU rCrCrA rNrNrN rN -3 (5' RNA linker)	IDT	N/A
Biotin TEG 5' [BtnTg]GCAGGTGCCAGAACATTTCTCTATCGATAGG	Sigma-Aldrich	N/A
Reverse 3' CTTTACCAACAGTACCGGAATGC	Sigma-Aldrich	N/A
Act5C gRNA forward TTCGGACCGCAAGTGCTTCTAAGA	Sigma-Aldrich	N/A
Act5C gRNA reverse AACTCTTAGAAGCACTTGCGGTC	Sigma-Aldrich	N/A
TBP N-terminus gRNA forward TTCGACAATAAACCATCTGTAAGA	Sigma-Aldrich	N/A
TBP N-terminus gRNA reverse AACTCTTACAGATGGTTTATTGTC	Sigma-Aldrich	N/A
DREF N-terminus gRNA forward ttcGGAAGACAAGATGAGCGAAG	Sigma-Aldrich	N/A
DREF N-terminus gRNA reverse aacCTTCGCTCATCTTGTCTTCC	Sigma-Aldrich	N/A
Chromator N-terminus gRNA forward TTCGCTGGAGTCGTGAATAATGT	Sigma-Aldrich	N/A
Chromator N-terminus gRNA reverse AACACATTATTCACGACTCCAGC	Sigma-Aldrich	N/A

TFIIA-L C-terminus gRNA forward TTCGCGACGCCGAGTGGTAATGGA	Sigma-Aldrich	N/A
TFIIA-L C-terminus gRNA reverse AACTCCATTACCACTCGGCGTCGC	Sigma-Aldrich	N/A
CG3847 N-terminus gRNA forward TTCGGCTTGGCATTCATATCGAGT	Sigma-Aldrich	N/A
CG3847 N-terminus gRNA reverse AACACTCGATATGAATGCCAAGCC	Sigma-Aldrich	N/A
CG7372 N-terminus gRNA forward TTCGCGTGTCTGACATGCTGAAAA	Sigma-Aldrich	N/A
CG7372 N-terminus gRNA reverse AACTTTTCAGCATGTCAGACACGC	Sigma-Aldrich	N/A
CG14711 N-terminus gRNA forward TTCGATTCGGCACAACATGTACTC	Sigma-Aldrich	N/A
CG14711 N-terminus gRNA reverse AACGAGTACATGTTGTGCCGAATC	Sigma-Aldrich	N/A
Ibf1 N-terminus gRNA forward TTCGAATGCCCCGAAAGAAGTCCG	Sigma-Aldrich	N/A
Ibf1 N-terminus gRNA reverse AACCGGACTTCTTTCGGGGCATTC	Sigma-Aldrich	N/A
Ibf2 N-terminus gRNA forward TTCGGCTTGCGACATTTTTACATA	Sigma-Aldrich	N/A
Ibf2 N-terminus gRNA reverse AACTATGTAAAAATGTCGCAAGCC	Sigma-Aldrich	N/A
TBP AID N-terminal repair cassette forward CCGCGTTACATAGCATCGTACGCGTACGTGTTT-GGTCCACAATAAACCATCTGTAATGGCCAA-GCCTTTGTCTCAAG	Sigma-Aldrich	N/A
TBP AID N-terminal repair cassette reverse CATCAGCATTCTAGAGCATCGTACGCGTAC-GTGTTTGGCTTAGCATTT-GGTCCATCTGCGAGCCACCGCCCGATC	Sigma-Aldrich	N/A
DREF AID N-terminal repair cassette forward ccgcgttacatagcatcgtacgcgtacgtgtttggCACAGAAGA-CAAGATGAGCGATGGCCAAGCCTTTGTCTCAAG	Sigma-Aldrich	N/A
DREF AID N-terminal repair cassette reverse catcagcattctagagcatcgtacgcgtacgtgtttggGGGCGAC-GCTGGTACCCCTTCCGAGCCACCGCCCGATC	Sigma-Aldrich	N/A
TFIIA-L AID C-terminal repair cassette forward CCGCGTTACATAGCATCGTACGCGTACGTGTTT-GGCGAATGGCGACGCCGAG-TGGGGCGGTGGCTCGGGAG	Sigma-Aldrich	N/A
TFIIA-L AID C-terminal repair cassette reverse CATCAGCATTCTAGAGCATCGTACGCGTAC-GTGTTT-GGTGTTCGCTCAACTGCCATCCTTAGCCCTCCCACA CATAACCAG	Sigma-Aldrich	N/A
Chromator AID N-terminal repair cassette forward gttccgcgttacatagcatcgtacgcgtacgtgtttggGGCGCTG-GAGTCGTGAATAAATGGCCAAGCCTTTGTCTCA	Sigma-Aldrich	N/A
Chromator AID N-terminal repair cassette reverse catcagcattctagagcatcgtacgcgtacgtgttt-ggTGAAATCTCCTGTGCCAACATCGAGCCAC-CGCCCGATC	Sigma-Aldrich	N/A
CG3847 AID N-terminal repair cassette forward CCGCGTTACATAGCATCGTACGCGTACGTGTTT-GGAAACTTATTCAAAGCCAACTATGGCCAA-GCCTTTGTCTCAAG	Sigma-Aldrich	N/A

CG3847 AID N-terminal repair cassette reverse CATCAGCATTCTAGAGCATCGTACGCGTAC-GTGTTTGGCGCGCTTGGCATTCATA-TCCGAGCCACCGCCCGATC	Sigma-Aldrich	N/A
CG7372 AID N-terminal repair cassette forward CCGCGTTACATAGCATCGTACGCGTACGTGTTT-GGTTTCGTGTCTGACCTGCTGAATGGCCAA-GCCTTTGTCTCAAG	Sigma-Aldrich	N/A
CG7372 AID N-terminal repair cassette reverse CATCAGCATTCTAGAGCATCGTACGCGTAC-GTGTTTGGACCATTTGCCATT-GCCATTTTCGAGCCACCGCCCGATC	Sigma-Aldrich	N/A
CG14711 AID N-terminal repair cassette forward CCGCGTTACATAGCATCGTACGCGTACGTGTTT-GGCGTGCAAGATAATGCCCGAGATGGCCAA-GCCTTTGTCTCAAG	Sigma-Aldrich	N/A
CG14711 AID N-terminal repair cassette reverse CATCAGCATTCTAGAGCATCGTACGCGTAC-GTGTTTGGCAGATTCGGCACAACATGTAC-GAGCCACCGCCCGATC	Sigma-Aldrich	N/A
Ibf1 AID N-terminal repair cassette forward CCGCGTTACATAGCATCGTACGCGTACGTGTTTGG-TAAAATGCCCCGAAAGAAGTATGGCCAAGCCTTT-GTCTCAAG	Sigma-Aldrich	N/A
Ibf1 AID N-terminal repair cassette reverse CATCAGCATTCTAGAGCATCGTACGCGTAC-GTGTTTGGGGTTCTGTAAAAATCCTCGGAC-GAGCCACCGCCCGATC	Sigma-Aldrich	N/A
Ibf2 AID N-terminal repair cassette forward CCGCGTTACATAGCATCGTACGCGTACGTGTTTGG-TAATTTAACACAAACCGTATATGGCCAAGCCTTT-GTCTCAAG	Sigma-Aldrich	N/A
Ibf2 AID N-terminal repair cassette reverse CATCAGCATTCTAGAGCATCGTACGCGTAC-GTGTTTGGTTTGCTTGCGACATTTTTACAC-GAGCCACCGCCCGATC	Sigma-Aldrich	N/A
OsTir ligase donor cassette forward TGGATCTCCAAGCAGGAGTACGACGAG-TCCGGCCCCTCCATTGTGCACCG-CAAGTGCTTCGGCAGCGGCGCCAC	Sigma-Aldrich	N/A
OsTir ligase donor cassette reverse CCTCCAGCAGAATCAAGAC-CATCCCGATCCTGATCCTCTTGCCCAGACAA-GCGATCCTTCCTAGCCCTCCCACACATAACCAG	Sigma-Aldrich	N/A
Genotyping Act5C OsTir forward GGCTTCGCTGTCCACCTTCCAG	Sigma-Aldrich	N/A
Genotyping Act5C OsTir reverse GAAGTCGAGGAAGCAGCAGCGA	Sigma-Aldrich	N/A
Recombinant DNA		
pBabe Puro osTIR1-9Myc	Addgene	plasmid #80074
pAc-sgRNA-Cas9	Addgene	plasmid #49330
pCRIS-PITChv2-FBL	Addgene	plasmid #63672
pGL13_tGFP	This study	N/A
Software and algorithms		
MSAmanda	N/A	https://ms.imp.ac.at/?goto=msamanda
Benchling	N/A	https://benchling.com

R version 3.5.3	R Development Core Team, 2019	https://www.r-project.org
Cutadapt	Martin et al. 2011	https://bioweb.pasteur.fr/packages/pack@cutadapt@1.18
Samtools version 1.9	Li et al. 2009	http://www.htslib.org/
bowtie version 1.2.2	Langmead et al., 2009	https://sourceforge.net/projects/bowtie-bio/files/bowtie/1.2.2/
GenomicRanges 1.34.0	Lawrence et al. 2013	https://bioconductor.org/packages/release/bioc/html/GenomicRanges.html
Biostrings 2.50.2	N/A	https://bioconductor.org/packages/Biostrings
bigBedtoBed	Kent et al. 2010	https://github.com/ENCODE-DCC/kentUtils/blob/master/src/utils/bigBedToBed/bigBedToBed.c
bedtools 2.27.1	Quinlan & Hall 2010	https://github.com/arq5x/bedtools2/releases/tag/v2.30.0
DESeq2 package v.1.30.1	Love et al. 2014	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
Experimental models: Cell lines		
D. melanogaster Schneider S2 cells	Thermo Fisher	Cat#R69007
HCT116	ATCC	Cat#CCL-247
Parental OsTir expressing S2 cell line	This study	N/A
TBP N-terminally tagged AID S2 cell line	This study	N/A
DREF N-terminally tagged AID S2 cell line	This study	N/A
TFIIA C-terminally tagged AID S2 cell line	This study	N/A
Chromator N-terminally tagged AID S2 cell line	This study	N/A
CG3847 N-terminally tagged AID S2 cell line	This study	N/A
CG7372 N-terminally tagged AID S2 cell line	This study	N/A
CG14711 N-terminally tagged AID S2 cell line	This study	N/A
Ibf1+Ibf2 double N-terminally tagged AID S2 cell line	This study	N/A

Discussion

Setting out to identify the protein binders of functionally distinct promoters in *D. melanogaster* allowed me to identify differences in the recruitment strategies of the Pol II PIC. Specifically, developmental promoters were found to use sequence motif features such as the TATA-box, INR and DPE to directly recruit the PIC. Consistent with my model, housekeeping promoters on the other hand, first recruit sequence-specific TFs that in turn recruit cofactors that recruit the PIC. Both classes of promoters utilize their encoded DNA sequence motifs to recruit the Pol II PIC, while developmental promoters directly engage the PIC through the DNA motifs, housekeeping promoter motifs may indirectly recruit the PIC. This was also in-line with the observation that developmental promoter motifs such as the TATA-box, INR and DPE are found in spatially restricted positions in relation to a focus transcription initiation site, while housekeeping promoter motifs such as Ohler1, Ohler7 and the DRE do not prefer a specific position and can be found equally up-and-down stream of the dominant TSS. My findings further confirm structural studies showing defined promoter DNA-PIC interactions in developmental type promoters which have been exclusively utilized for such purpose. In addition, my study expands our understanding of housekeeping promoter function by proposing a new mechanism of Pol II PIC recruitment which does not require defined promoter DNA-PIC interactions, but rather relies on cofactor-PIC interactions through TFIIA to recruit the PIC. Importantly, this does not exclude the participation of additional GTFs in housekeeping promoter transcription as I was not able to investigate their requirement for transcription. Having two functionally distinct promoter types recruit the transcription machinery in different ways, suggests that transcriptional cofactors may utilize this behavior to confer specificity.

Our lab has found that transcriptional cofactor proteins display a preference for activating developmental or housekeeping promoters. I thus envision that the functional specificity of cofactors can be manifested by contributing distinct biochemical activities at either preferred class of promoters. As developmental promoters can nucleate and recruit the PIC *in-vitro*, it is likely that they may be activated by signals acting post polymerase recruitment. For example, the acetyltransferase p300 and the Mediator subunit MED25 may preferentially activate developmental promoters by increasing the rates of transcription elongation through the recruitment of Cdk9.

While housekeeping cofactors may facilitate transcriptional activation by increasing the local concentration of Pol II to initiate transcription within the nucleosome free regions encompassing the TSS. Further implication for the mechanism of PIC recruitment by Drosophila housekeeping promoters may relate more generally to the way promoters with broad transcription initation patterns function. In mammalian genomes roughly 80% of all promoters exhibition broad transcription initiation patterns. This may require us to revisit PIC recruitment and assembly on a wider range of promoter DNA. Additional biochemical experiments that identify cofactor-PIC interactions would be required for us to understand how cofactors engage with the transcription machinery to recruit the PIC to a wide range of promoter DNA regions.

The data I have so far discussed indicate that within a single cell, housekeeping and developmental transcription programs coexist and utilize different transcriptional cofactors that recruit the transcription machinery. In the crowded and dynamic nuclear environment, it is hard to imagine how this specificity is reinforced. I propose that it may be possible to integrate these data in the context of liquid-liquid phase separated nuclear condensates for each transcriptional program (Boehning et al., 2018; Cramer, 2019). Developmental cofactors would preferentially be found inside transcriptional condensates with developmental enhancers and promoters, while housekeeping cofactors may be found in condensates with housekeeping enhancers and promoters. In mammalian genomes, it might be likely that many more functional types of condensates would exist with different types of cofactors. Experiments that dissect the ability of transcriptional regulators to form condensates and their ability to transactivate or integrate into different protein complexes would be imperative for our ability to understand their ability to specifically activate distinct sets of promoters.

References

A Mechanism for Coordinating Chromatin Modification and Preinitiation Complex Assembly, 2006. A Mechanism for Coordinating Chromatin Modification and Preinitiation Complex Assembly. Molecular Cell 23, 809–818. doi:10.1016/j.molcel.2006.07.018

Akhtar, W., Veenstra, G.J.C., 2011. TBP-related factors: a paradigm of diversity in transcription initiation. Cell Biosci 1, 23–12. doi:10.1186/2045-3701-1-23

Andersen, P.R., Tirian, L., Vunjak, M., Brennecke, J., 2017. A heterochromatin-dependent transcription machinery drives piRNA expression. Nature 549, 54–59. doi:10.1038/nature23482

Arnold, C.D., Zabidi, M.A., Pagani, M., Rath, M., Schernhuber, K., Kazmar, T., Stark, A., 2016. Genome-wide assessment of sequence-intrinsic enhancer responsiveness at single-base-pair resolution. Nat Biotechnol 35, 136–144. doi:10.1038/nbt.3739

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

Becker, P.B., Hörz, W., 2002. ATP-dependent nucleosome remodeling. Annu. Rev. Biochem. 71, 247–273. doi:10.1146/annurev.biochem.71.110601.135400

Bernecky, C., Plitzko, J.M., Cramer, P., 2017. Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. Nat Struct Mol Biol 24, 809–815. doi:10.1038/nsmb.3465

Boehning, M., Dugast-Darzacq, C., Rankovic, M., Hansen, A.S., Yu, T., Marie-Nelly, H., McSwiggen, D.T., Kokic, G., Dailey, G.M., Cramer, P., Darzacq, X., Zweckstetter, M., 2018. RNA polymerase II clustering through carboxy- terminal domain phase separation. Nat Struct Mol Biol 1–16. doi:10.1038/s41594-018-0112-y

Butter, F., Davison, L., Viturawong, T., Scheibe, M., Vermeulen, M., Todd, J.A., Mann, M., 2012. Proteome-Wide Analysis of Disease-Associated SNPs That Show Allele-Specific Transcription Factor Binding. PLoS Genet 8, e1002982–8. doi:10.1371/journal.pgen.1002982

Chen, K., Johnston, J., Shao, W., Meier, S., Staber, C., Zeitlinger, J., 2013. A global change in RNA polymerase II pausing during the Drosophila midblastula transition. eLife 2, e00861. doi:10.7554/eLife.00861

Chen, X., Qi, Y., Wu, Z., Wang, X., Li, J., Zhao, D., Hou, H., Li, Y., Yu, Z., Liu, W., Wang, M., Ren, Y., Li, Z., Yang, H., Xu, Y., 2021. Structural insights into preinitiation complex assembly on core promoters. Science 372. doi:10.1126/science.aba8490

Cheung, A.C.M., Cramer, P., 2012. A movie of RNA polymerase II transcription. Cell 149, 1431–1437. doi:10.1016/j.cell.2012.06.006

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

Core, L., Adelman, K., 2019. Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. Genes & Development 33, 960–982. doi:10.1101/gad.325142.119

Cramer, P., 2019. Organization and regulation of gene transcription. Nature 573, 45–54. doi:10.1038/s41586-019-1517-4

Cubeñas-Potts, C., Rowley, M.J., Lyu, X., Li, G., Lei, E.P., Corces, V.G., 2017. Different enhancer classes in Drosophila bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture. Nucleic Acids Res 45, 1714–1730. doi:10.1093/nar/gkw1114
Degnan, B.M., Vervoort, M., Larroux, C., Richards, G.S., 2009. Early evolution of metazoan transcription factors 1–9. doi:10.1016/j.gde.2009.09.008
Déjardin, J., Kingston, R.E., 2009. Purification of Proteins Associated with Specific Genomic Loci. Cell 136, 175–186. doi:10.1016/j.cell.2008.11.045
Drapkin, R., Merino, A., Reinberg, D., 1993. Regulation of RNA polymerase II transcription. Curr Opin Cell Biol 5, 469–476. doi:10.1016/0955-0674(93)90013-g
Duttke, S.H.C., Doolittle, R.F., Wang, Y.-L., Kadonaga, J.T., 2014. TRF2 and the evolution of the bilateria. Genes & Development 28, 2071–2076. doi:10.1101/gad.250563.114
FANTOM Consortium and the RIKEN PMI and CLST (DGT), Forrest, A.R.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J.L., Haberle, V., Lassmann, T., Kulakovskiy, I.V., Lizio, M., Itoh, M., Andersson, R., Mungall, C.J., Meehan, T.F., Schmeier, S., Bertin, N., Jørgensen, M., Dimont, E., Arner, E., Schmidl, C., Schaefer, U., Medvedeva, Y.A., Plessy, C., Vitezic, M., Severin, J., Semple, C.A., Ishizu, Y., Young, R.S., Francescatto, M., Alam, I., Albanese, D., Altschuler, G.M., Arakawa, T., Archer, J.A.C., Arner, P., Babina, M., Rennie, S., Balwierz, P.J., Beckhouse, A.G., Pradhan-Bhatt, S., Blake, J.A., Blumenthal, A., Bodega, B., Bonetti, A., Briggs, J., Brombacher, F., Burroughs, A.M., Califano, A., Cannistraci, C.V., Carbajo, D., Chen, Y., Chierici, M., Ciani, Y., Clevers, H.C., Dalla, E., Davis, C.A., Detmar, M., Diehl, A.D., Dohi, T., Drabløs, F., Edge, A.S.B., Edinger, M., Ekwall, K., Endoh, M., Enomoto, H., Fagiolini, M., Fairbairn, L., Fang, H., Farach-Carson, M.C., Faulkner, G.J., Favorov, A.V., Fisher, M.E., Frith, M.C., Fujita, R., Fukuda, S., Furlanello, C., Furino, M., Furusawa, J.-I., Geijtenbeek, T.B., Gibson, A.P., Gingeras, T., Goldowitz, D., Gough, J., Guhl, S., Guler, R., Gustincich, S., Ha, T.J., Hamaguchi, M., Hara, M., Harbers, M., Harshbarger, J., Hasegawa, A., Hasegawa, Y., Hashimoto, T., Herlyn, M., Hitchens, K.J., Ho Sui, S.J., Hofmann, O.M., Hoof, I., Hori, F., Huminiecki, L., Iida, K., Ikawa, T., Jankovic, B.R., Jia, H., Joshi, A., Jurman, G., Kaczkowski, B., Kai, C., Kaida, K., Kaiho, A., Kajiyama, K., Kanamori-Katayama, M., Kasianov, A.S., Kasukawa, T., Katayama, S., Kato, S., Kawaguchi, S., Kawamoto, H., Kawamura, Y.I., Kawashima, T., Kempfle, J.S., Kenna, T.J., Kere, J., Khachigian, L.M., Kitamura, T., Klinken, S.P., Knox, A.J., Kojima, M., Kojima, S., Kondo, N., Koseki, H., Koyasu, S., Krampitz, S., Kubosaki, A., Kwon, A.T., Laros, J.F.J., Lee, W., Lennartsson, A., Li, K., Lilje, B., Lipovich, L., Mackay-Sim, A., Manabe, R.-I., Mar, J.C., Marchand, B., Mathelier, A., Mejhert, N., Meynert, A., Mizuno, Y., de Lima Morais, D.A., Morikawa, H., Morimoto, M., Moro, K., Motakis, E., Motohashi, H., Mummery, C.L., Murata, M., Nagao-Sato, S., Nakachi, Y., Nakahara, F., Nakamura, T., Nakamura, Y., Nakazato, K., van Nimwegen, E., Ninomiya, N., Nishiyori, H., Noma, S., Noazaki, T., Ogishima, S., Ohkura, N., Ohimiya, H., Ohno, H., Ohshima, M., Okada-Hatakeyama, M., Okazaki, Y., Orlando, V., Ovchinnikov, D.A., Pain, A., Passier, R., Patrikakis, M., Persson, H., Piazza, S., Prendergast, J.G.D., Rackham, O.J.L., Ramilowski, J.A., Rashid, M., Ravasi, T., Rizzu, P., Roncador, M., Roy, S., Rye, M.B., Saijyo, E., Sajantila, A., Saka, A., Sakaguchi, S., Sakai, M., Sato, H., Savvi, S., Saxena, A., Schneider, C., Schultes, E.A., Schulze-Tanzil, G.G., Schwegmann, A., Sengstag, T., Sheng, G., Shimoji, H., Shimoni, Y., Shin, J.W., Simon, C., Sugiyama, D., Sugiyama, T., Su-

zuki, M., Suzuki, N., Swoboda, R.K., 't Hoen, P.A.C., Tagami, M., Takahashi, N., Takai, J., Tanaka, H., Tatsukawa, H., Tatum, Z., Thompson, M., Toyodo, H., Toyoda, T., Valen, E., van de Wetering, M., van den Berg, L.M., Verado, R., Vijayan, D., Vorontsov, I.E., Wasserman, W.W., Watanabe, S., Wells, C.A., Winteringham, L.N., Wolvetang, E., Wood, E.J., Yamaguchi, Y., Yamamoto, M., Yoneda, M., Yonekura, Y., Yoshida, S., Zabierowski, S.E., Zhang, P.G., Zhao, X., Zucchelli, S., Summers, K.M., Suzuki, H., Daub, C.O., Kawai, J., Heutink, P., Hide, W., Freeman, T.C., Lenhard, B., Bajic, V.B., Taylor, M.S., Makeev, V.J., Sandelin, A., Hume, D.A., Carninci, P., Hayashizaki, Y., 2014. A promoter-level mammalian expression atlas. Nature 507, 462–470. doi:10.1038/nature13182

Fujita, T., Fujii, H., 2016. Biochemical Analysis of Genome Functions Using Locus-Specific Chromatin Immunoprecipitation Technologies. Gene Regul Syst Bio 10, 1–9. doi:10.4137/GRSB.S32520

Gressel, S., Schwalb, B., Cramer, P., 2019. The pause-initiation limit restricts transcription activation in human cells. Nature Communications 10, 3603–12. doi:10.1038/s41467-019-11536-8

Griesenbeck, J., Boeger, H., Strattan, J.S., Kornberg, R.D., 2003. Affinity purification of specific chromatin segments from chromosomal loci in yeast. Molecular and Cellular Biology 23, 9275–9282. doi:10.1128/MCB.23.24.9275-9282.2003

Haberle, V., Arnold, C.D., Pagani, M., Rath, M., Schernhuber, K., Stark, A., 2019. Transcriptional cofactors display specificity for distinct types of core promoters. Nature 1–24. doi:10.1038/s41586-019-1210-7

Haberle, V., Lenhard, B., 2016. Promoter architectures and developmental gene regulation. Semin Cell Dev Biol 57, 11–23. doi:10.1016/j.semcdb.2016.01.014

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

He, Y., Fang, J., Taatjes, D.J., Nogales, E., 2013. Structural visualization of key steps in human transcription initiation. Nature 495, 481–486. doi:10.1038/nature11991

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

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

Sequeira, V.M., Vermeulin, M. 2020. Identifying Readers for (hydroxy)methylated DNA Using Quantitative Interaction Proteomics: Advances and Challenges Ahead. Journal of Molecular Biology 432, 1792–1800. doi:10.1016/j.jmb.2019.12.014

Juven-Gershon, T., Hsu, J.-Y., Theisen, J.W., Kadonaga, J.T., 2008. The RNA polymerase II core promoter - the gateway to transcription. Curr Opin Cell Biol 20, 253–259. doi:10.1016/j.ceb.2008.03.003

Kadonaga, J.T., Tjian, R., 1986. Affinity purification of sequence-specific DNA binding proteins. Proc Natl Acad Sci USA 1–5.

King, N., Hittinger, C.T., Carroll, S.B., 2003. Evolution of key cell signaling and adhesion protein families predates animal origins. Science 301, 361–363. doi:10.1126/science.1083853

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

Lin, J.J., Carey, M., 2012. In vitro transcription and immobilized template analysis of preinitiation complexes. Curr Protoc Mol Biol Chapter 12, Unit 12.14.–12.14.19. doi:10.1002/0471142727.mb1214s97

Murakami, K., Elmlund, H., Kalisman, N., Bushnell, D.A., Adams, C.M., Azubel, M., Elmlund, D., Levi-Kalisman, Y., Liu, X., Gibbons, B.J., Levitt, M., Kornberg, R.D., 2013. Architecture of an RNA polymerase II transcription pre-initiation complex. Science 342, 1238724. doi:10.1126/science.1238724

Neves, A., Eisenman, R.N., 2019. Distinct gene-selective roles for a network of core promoter factors in Drosophila neural stem cell identity. Biol Open 8. doi:10.1242/bio.042168

Newell, C.A., Gray, J.C., 2010. Binding of lac repressor-GFP fusion protein to lac operator sites inserted in the tobacco chloroplast genome examined by chromatin immunoprecipitation. Nucleic Acids Res 38, e145. doi:10.1093/nar/gkq413

Ohler, U., Liao, G.-C., Niemann, H., Rubin, G.M., 2002. Computational analysis of core promoters in the Drosophila genome. Genome Biol. 3, RESEARCH0087–12. doi:10.1186/gb-2002-3-12-research0087

Orphanides, G., lagrange, T., Reinberg, D., 1996. The general transcription factors of RNA polymerase II. Genes & Development 10, 2657–2683.

Orphanides, G., Reinberg, D., 2002. A unified theory of gene expression. Cell 108, 439–451. doi:10.1016/s0092-8674(02)00655-4

Parry, T.J., Theisen, J.W.M., Hsu, J.Y., Wang, Y.L., Corcoran, D.L., Eustice, M., Ohler, U., Kadonaga, J.T., 2010. The TCT motif, a key component of an RNA polymerase II transcription system for the translational machinery. Genes & Development 24, 2013–2018. doi:10.1101/gad.1951110

Peterson, K.J., Davidson, E.H., 2000. Regulatory evolution and the origin of the bilaterians. Proc Natl Acad Sci USA 97, 4430–4433. doi:10.1073/pnas.97.9.4430

Petrenko, N., Jin, Y., Dong, L., Wong, K.H., Struhl, K., 2019. Requirements for RNA polymerase II preinitiation complex formation in vivo. eLife 8. doi:10.7554/eLife.43654

Plaschka, C., Larivière, L., Wenzeck, L., Seizl, M., Hemann, M., Tegunov, D., Petrotchenko, E.V., Borchers, C.H., Baumeister, W., Herzog, F., Villa, E., Cramer, P., 2015. Architecture of the RNA polymerase II-Mediator core initiation complex. Nature 518, 376–380. doi:10.1038/nature14229

Qiu, W., Xu, Z., Zhang, M., Zhang, D., Fan, H., Li, T., Wang, Q., Liu, P., Zhu, Z., Du, D., Tan, M., Wen, B., Liu, Y., 2019. Determination of local chromatin interactions using a combined CRISPR and peroxidase APEX2 system. Nucleic Acids Res 47, e52. doi:10.1093/nar/gkz134

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

Rickels, R., Herz, H.-M., Sze, C.C., Cao, K., Morgan, M.A., Collings, C.K., Gause, M., Takahashi, Y.-H., Wang, L., Rendleman, E.J., Marshall, S.A., Krueger, A., Bartom, E.T., Piunti, A., Smith, E.R., Abshiru, N.A., Kelleher, N.L., Dorsett, D., Shilatifard, A., 2017. Histone H3K4 monomethylation catalyzed by Trr and

mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. Nat Genet 49, 1647–1653. doi:10.1038/ng.3965
Roux, K.J., Kim, D.I., Burke, B., May, D.G., 2018. BioID: A Screen for Protein-Protein Interactions. Curr Protoc Protein Sci 91, 19.23.1–19.23.15. doi:10.1002/cpps.51
Rubinstein, M., de Souza, F.S.J., 2013. Evolution of transcriptional enhancers and animal diversity. Philos Trans R Soc Lond B Biol Sci 368, 20130017. doi:10.1098/rstb.2013.0017
Saksouk, N., Barth, T.K., Ziegler-Birling, C., Olova, N., Nowak, A., Rey, E., Mateos-Langerak, J., Urbach, S., Reik, W., Torres-Padilla, M.-E., Imhof, A., Déjardin, J., Simboeck, E., 2014. Redundant mechanisms to form silent chromatin at pericentromeric regions rely on BEND3 and DNA methylation. Molecular Cell 56, 580–594. doi:10.1016/j.molcel.2014.10.001
Schier, A.C., Taatjes, D.J., 2020. Structure and mechanism of the RNA polymerase II transcription machinery. Genes & Development 34, 465–488. doi:10.1101/gad.335679.119
Schmidtmann, E., Anton, T., Rombaut, P., Herzog, F., Leonhardt, H., 2016. Determination of local chromatin composition by CasID. Nucleus 7, 476–484. doi:10.1080/19491034.2016.1239000
Shao, W., Alcantara, S.G.-M., Zeitlinger, J., 2019. Reporter-ChIP-nexus reveals strong contribution of the Drosophila initiator sequence to RNA polymerase pausing. eLife 8. doi:10.7554/eLife.41461
Shao, W., Zeitlinger, J., 2017. Paused RNA polymerase II inhibits new transcriptional initiation. Nat Genet 49, 1045–1051. doi:10.1038/ng.3867
Singer, V.L., Wobbe, C.R., Struhl, K., 1990. A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation. Genes & Development 4, 636–645. doi:10.1101/gad.4.4.636
Stampfel, G., Kazmar, T., Frank, O., Wienerroither, S., Reiter, F., Stark, A., 2015. Transcriptional regulators form diverse groups with context-dependent regulatory functions. Nature 1–18. doi:10.1038/nature15545
Studitsky, V.M., Nizovtseva, E.V., Shaytan, A.K., Luse, D.S., 2016. Nucleosomal Barrier to Transcription: Structural Determinants and Changes in Chromatin Structure. Biochem Mol Biol J 2. doi:10.21767/2471-8084.100017
Sun, F., Sun, T., Kronenberg, M., Tan, X., Huang, C., Carey, M.F., 2021. The Pol II preinitiation complex (PIC) influences Mediator binding but not promoter-enhancer looping. Genes & Development 35, 1175–1189. doi:10.1101/gad.348471.121
Dynan, W.S., Tjian, R. 1983. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell 35, 79–87. doi:10.1016/0092-8674(83)90210-6
Tolkunov, D., Zawadzki, K.A., Singer, C., Elfving, N., Morozov, A.V., Broach, J.R., 2011. Chromatin remodelers clear nucleosomes from intrinsically unfavorable sites to establish nucleosome-depleted regions at promoters. Mol Biol Cell 22, 2106–2118. doi:10.1091/mbc.E10-10-0826
Tsui, C., Inouye, C., Levy, M., Lu, A., Florens, L., Washburn, M.P., Tjian, R., 2018. dCas9-targeted locus-specific protein isolation method identifies histone gene regulators. Proc Natl Acad Sci U S A 115, E2734–E2741. doi:10.1073/pnas.1718844115
Tunnacliffe, E., Corrigan, A.M., Chubb, J.R., 2018. Promoter-mediated diversification of transcriptional bursting dynamics following gene duplication. Proc Natl Acad Sci U S A 115, 8364–8369. doi:10.1073/pnas.1800943115

Viturawong, T., Meissner, F., Butter, F., Mann, M., 2013. A DNA-Centric Protein Interaction Map of Ultraconserved Elements Reveals Contribution of Transcription Factor Binding Hubs to Conservation. CellReports 5, 531–545. doi:10.1016/j.celrep.2013.09.022

Wang, Y.L., Duttke, S.H.C., Chen, K., Johnston, J., Kassavetis, G.A., Zeitlinger, J., Kadonaga, J.T., 2014. TRF2, but not TBP, mediates the transcription of ribosomal protein genes. Genes & Development 28, 1550–1555. doi:10.1101/gad.245662.114

Wu, C.-H., Yamaguchi, Y., Benjamin, L.R., Horvat-Gordon, M., Washinsky, J., Enerly, E., Larsson, J., Lambertsson, A., Handa, H., Gilmour, D., 2003. NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila. Genes & Development 17, 1402–1414. doi:10.1101/gad.1091403

Xie, X., Lu, J., Kulbokas, E.J., Golub, T.R., Mootha, V., Lindblad-Toh, K., Lander, E.S., Kellis, M., 2005. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. Nature 434, 338–345. doi:10.1038/nature03441

Yang, C., Bolotin, E., Jiang, T., Sladek, F.M., Martinez, E., 2007. Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. Gene 389, 52–65. doi:10.1016/j.gene.2006.09.029

Yu, C., Cvetesic, N., Hisler, V., Gupta, K., Ye, T., Gazdag, E., Negroni, L., Hajkova, P., Berger, I., Lenhard, B., Müller, F., Vincent, S.D., Tora, L., 2020. TBPL2/TFIIA complex establishes the maternal transcriptome through oocyte-specific promoter usage. Nature Communications 11, 6439–13. doi:10.1038/s41467-020-20239-4

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

Zabidi, M.A., Stark, A., 2016. Regulatory Enhancer–Core- Promoter Communication via Transcription Factors and Cofactors. Trends in Genetics 1–14. doi:10.1016/j.tig.2016.10.003

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

Zhou, H., Wan, B., Grubisic, I., Kaplan, T., Tjian, R., 2014. TAF7L modulates brown adipose tissue formation. eLife 3. doi:10.7554/eLife.02811