Lola-I is a promoter pioneer factor that regulates RNA polymerase II pausing during *Drosophila* embryogenesis

©2019

Vivekanandan Ramalingam

Submitted to the graduate degree program in Pathology & Laboratory Medicine and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Julia Zeitlinger, Ph.D., Co-Chair

Soumen Paul, Ph.D., Co-Chair

Committee members

Robb Krumlauf, Ph.D.

Kausik Si, Ph.D.

Patrick E. Fields, Ph.D.

Chad Slawson, Ph.D.

Date defended: December 3, 2019

The Dissertation Committee for Vivekanandan Ramalingam certifies
that this is the approved version of the following dissertation :

Lola-I is a promoter pioneer factor that regulates RNA polymerase II pausing during *Drosophila* embryogenesis

Julia Zeitlinger, Ph.D., Co-Chair

Soumen Paul, Ph.D., Co-Chair

Date approved: December 18, 2019

Abstract

Transcriptionally engaged Pol II has been observed at many inactive genes, in a paused state. Since Pol II can be found even in the inactive state, is Pol II recruited to these genes by default? If the recruitment of Pol II is regulated, how is it achieved? If the recruitment is regulated, it is presumably not by transcriptional activators, since Pol II is present even when genes are inactive. Could it be by pioneer factors, factors that regulate chromatin accessibility for the binding of other factors, but do not cause transcriptional activation?

In this thesis, I have identified Lola-I, a protein ubiquitously expressed during the late stages of *Drosophila* embryogenesis, to be necessary for the establishment of paused Pol II at many paused genes. Lola-I acts as a pioneer factor which depletes the promoter nucleosome of paused genes to allow the recruitment of Pol II. Besides, I have characterized in detail the differences between paused and non-paused genes, which are expressed during the late stages of *Drosophila* embryogenesis. The paused genes are generally broadly expressed and have a basal expression even in inactive tissues. Furthermore, utilizing the Lola-I mutants, I established the role of paused Pol II in decreasing the stochasticity in gene expression. This thesis provides first insights into how paused Pol II occupancy can selectively be established at specific promoters and how this could be utilized to fine-tune gene expression.

Acknowledgements

This dissertation has been an incredible journey made possible only because of the tremendous support and contributions of the following people.

My advisor Julia, has been tremendously supportive and understanding throughout my Ph.D., especially during times when I was feeling down. She has given me a lot of freedom to explore and try out many things. I have learned a lot of things from her, which I believe has shaped me to be a better scientist.

I wish to thank my collaborators for their contributions. I thank Jeff for guiding me through different aspects of genomics analysis and analyzing some of the INTACT data; Malini for her contributions to the INTACT experiments and general discussions (She established the INTACT method in our lab); Brian and Jay for their contributions to the smFISH experiments; Xinyang Yu and Michael Buck from the University at Buffalo helped with the *in vitro* binding assay.

I enjoyed and valued my experience working with four summer scholars during my Ph.D. I thank them for their contributions. Kaelan helped with the visual phenotypic analysis of Lola mutants. Ana helped with optimizing cell isolation for scRNAseq. Nastya helped with the optimization of the smFISH protocol. Divya helped with processing some of the smFISH data.

Cindi designed the custom antibodies and Lola-I protein expression constructs. Custom antibodies and protein production were done commercially with Genescript.

I will always be in debt to my friends for their support throughout my Ph.D. Being away from family, and the rigors of graduate school take a toll on a person. I believe I would have been lost without the support of my friends. I am grateful to have them.

I thank past and present Zeitlinger lab members for the welcoming atmosphere and making this a great place to work. I thank Cindi and Sara for keeping the lab organized. I think I could fully appreciate their help only after I leave the lab.

I am very thankful for the advice and support of my thesis committee members, especially Kausik, Robb, and Dr. Paul. Kausik was always checking in on my progress and well being.

Robb's comments were always insightful. Dr. Paul made sure I was on track.

In spite of being bothered by the pay differences between the KU grad students and the other grad students, I am still very grateful to work at the Stowers Institute. The resource availability at the institute is unbelievable, and it has made it very easy to pursue any question, only limited by the imagination. I thank Ron Yu for bringing me here as a summer scholar. The Conway and Gibson lab members were very helpful during my early days.

The Stowers Core facilities were helpful in many ways during my stay in the institute. Cytometry core (Dustin and Laura) helped to sort embryos, which made possible the numerous experiments with the mutant embryos. The microscopy core (Brian, Jay, and Jeff) were very helpful. The Sequencing core (Mike, Anoja, Kate, and Allie) has been amazing. It has always been a pleasure working with them. The quick turn around time and the responsiveness really makes the life of a genomics scientist pain-free. I thank Mark Miller for his help with some of the illustrations.

Contents

List of Figures

List of Tables

Glossary

ATAC-seq Assay for Transposase-Accessible Chromatin using sequencing.

BRE TFIIB recognition element.

ChIP-seq Chromatin Immunoprecipitation followed by sequencing.

DNA deoxyribonucleic acid.

DPE downstream promoter element.

DSIF DRB sensitivity inducing factor.

FACT facilitates chromatin transcription.

GAF GAGA factor.

GTFs General Transcription Factors.

Inr Initiator element.

MNAse-seq Micrococcal Nuclease digestion followed by sequencing.

NDR nucleosome depleted region.

NELF Negative elongation factor.

NFR nucleosome free region.

P-TEFb Positive transcription elongation factor b.

PB Pause Button.

PIC pre-initiation complex.

Pol II RNA Polymerase II.

RNAP RNA Polymerase.

scRNAseq single-cell RNA sequencing.

SEC Super elongation complex.

TAFs TATA associated factors.

TBP TATA binding protein.

Chapter 1

Introduction

How is the information embedded in the DNA read to produce precise special-temporal gene expression? This is one of the central questions in the field of transcription and developmental biology. The general picture has been that small stretches of DNA sequences in the genome can act as modules of different functionalities such as the promoters, enhancers, insulators, etc., and the integrated action of these elements produce precise transcriptional output. To understand how varied transcriptional outputs are produced in response to various developmental and disease conditions, it is imperative to know the different elements which affect transcription and the interactions between them.

Traditionally, promoters are considered as just a landing platform for the transciptional machinery and that the modulation of gene expression is achieved at the promoter proximal regulatory regions (Ma & Ptashne (1987), Hope et al. (1988)) or the distal regulatory regions (enhancers) (Spitz & Furlong (2012), Banerji et al. (1981), Small et al. (1992)) by the action of transcription factors. Recent experiments however, suggest that the promoters might play a more integral role in transcriptional regulation. For example, the promoter sequence could affect the total transcriptional output by affecting the burst size (Larsson et al. (2019), Fukaya et al. (2016), Suter et al. (2011), Bartman et al. (2016)). Promoters can also act as enhancers. In ectopic and *in vivo* enhancer activity assays and *in vivo*, promoter sequences can enhance the activity of distal promoters (Zabidi et al. (2014), Diao et al. (2017), Dao et al. (2017), Arnold et al. (2013)). These experiments challenge the distinction between promoters and enhancers. Moreover, the enhancer sequences are also transcribed (Kim et al. (2010), Arner et al. (2015), Core et al. (2014a), Kim & Shiekhattar (2015), Meers et al. (2018)) and the histone modifications at promoters are more similar to the enhancers than what has been previously characterized (Henriques et al. (2018)). Furthermore, promoter sequences might also dictate enhancer-promoter specificity (Zabidi et al. (2014), Verrijzer et al. (1995)). Thus, it is becoming increasingly clear that promoters also play a crucial role in

the regulation of gene expression.

Transcriptional regulation was thought to happen at the level of recruitment of polymerase II to the promoters (Ptashne & Gann (1997))). However, it is becoming increasingly clear that the promoter-proximal pausing of Pol II and its release is also crucial (Adelman & Lis (2012)). Paused Pol II has been found to be important for the rapid and synchronous activation of the target genes (Lagha et al. (2013), Boettiger & Levine (2009), Boettiger et al. (2011)). It has been proposed to be involved in enhancer-promoter interactions (Ghavi-Helm et al. (2014)). Moreover, Pol II has a long intrinsically disordered tail (Corden (1990), Corden et al. (1985), Allison et al. (1985)) which might be involved in phase transitions (Cho et al. (2018), Chong et al. (2018)), a mechanism by which enhancers have been proposed to activate genes in the recent models (Hnisz et al. (2017), Boehning et al. (2018)). Thus, paused RNA Pol II could act as a hub for many interactions. However, despite the many key roles paused Pol II might play in transcriptional regulation, many aspects of Pol II pausing is still not clear. In this thesis, I aim to dissect the role of paused Pol II in transcription regulation. Before going further into the aims of this thesis, I will introduce the current general understanding of the transcription process.

1.1 Different stages of transcription

The major steps in a typical transcription cycle are as follows. It begins with the recruitment of general transcription factors (GTFs) and Pol II to promoters. Following the recruitment of Pol II to promoters, Pol II initiates RNA synthesis and then pauses very close to the promoter, which usually happens 30-50 bp downstream of the TSS. The paused Pol II is transcriptionally still engaged and proceeds to further elongation, without any of the GTFs, after receiving additional activation signals. After completing the transcription of the rest of the gene, it terminates. The terminated Pol II is reused in some cases. Now we shall look in detail the different stages of transcription.

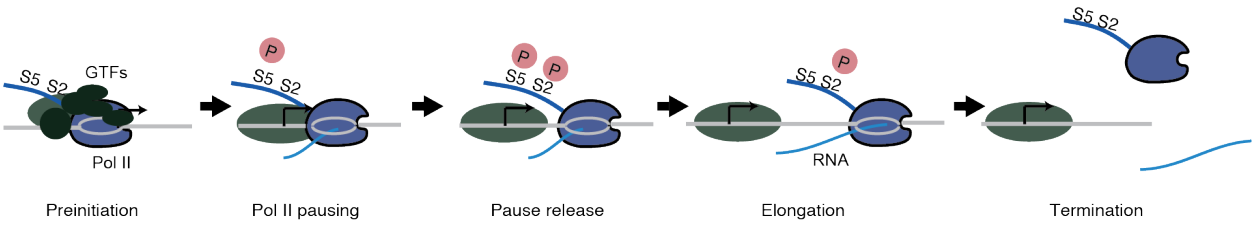

Figure 1.1: A typical transcriptional cycle of Pol II.

1.1.1 Initiation of RNA Polymerase II

Unlike the bacterial systems, RNA Polymerase in eukaryotes requires the GTFs for recognition of the promoter and early steps of the transcription cycle. Bacterial RNA Polymerase (RNAP) complex, which transcribes all the different types of RNAs in bacteria, consists of the catalytic core of two α subunits and one copy of β, β' and ω subunits and the regulatory core made up of σ subunit. σ subunit determine the promoter specificity of the holoenzyme. Following the binding of RNAP to the promoter in a closed complex, the holoenzyme melts the promoter DNA and transitions to an open complex, followed by the RNA chain initiation, promoter clearance, σ release, RNA chain elongation and termination (Browning & Busby (2004)). The above steps that take place in the prokaryotes are similar to the eukaryotes, which in addition, have promoter-proximal pausing in higher organisms. The similarity between the eukaryotic, prokaryotic and archaeal polymerase subunits and the transcription processes are remarkable (Allison et al. (1985), Chung et al. (2003), Ebright (2000), Werner & Grohmann (2011)).

In eukaryotes, transcription is achieved in most species by three different types of RNA Polymerases (RNA Pol I-III) (Roeder & Rutter (1969), Herr et al. (2005)). These polymerases differ in their subunit composition as well as the type of RNAs they transcribe. Ribosomal RNAs (rRNAs) are transcribed by RNA Polymerase I. RNA Pol II transcribes the protein-coding messenger RNAs (mRNAs) and also small regulatory RNAs, whereas RNA Pol III transcribes the transfer RNAs (tRNAs) (Lindell et al. (1970), Kedinger et al. (1970), Weinmann et al. (1974), Weinmann & Roeder (1974)).

Eukaryotic Pol II is a multi-subunit complex consisting of twelve subunits. The subunits RPB1,

RPB2, RPB3, RPB6, and RPB11 have high sequence homology with the bacterial counterparts (Werner & Grohmann (2011)). All multi-subunit RNAPs, including the eukaryotic RNA Pol II, resemble a crab claw, where the jaws feed the duplex DNA in the catalytic center. During the RNA synthesis, the duplex DNA is unwinded, creating a region of a single-stranded DNA bubble. One of the DNA strands of the bubble is utilized for the synthesis of the complementary RNA at the catalytic center. The unzipped DNA snaps back and exits the Pol II complex.

RPB1, the largest subunit of RNA Polymerase II, contains a long intrinsically disordered carboxy-terminal domain (CTD) composed of the consensus heptapeptide sequence YSPTSPS. The copy number of the heptad repeat varies across different species with 26 repeats in *S.Cerevisiae* to 52 repeats in *H.Sapiens*. The sequence composition of the heptad also varies across the tail and between different species (Chapman et al. (2008)). Major deletions or modifications of the CTD results in mortality (Hsin & Manley (2012), West & Corden (1995)). Moreover, while the CTD is essential *in vivo*, *in vitro* experiments have shown that it is not necessary for RNA synthesis (Buratowski & Sharp (1990), Zehring et al. (1988))). Therefore, the CTD might play more of a regulatory role than basic transcription (Gerber et al. (1995), Scafe et al. (1990)).

The CTD acts as a landing site for different histone-modifying enzymes such as Set1 and Set2, RNA processing enzymes involved in capping and splicing, and termination factors (Hsin & Manley (2012). Structurally, the CTD is located close to the RNA exit channel, allowing the CTD bound RNA processing enzymes access to the exiting RNA (Cramer et al. (2001)). In addition, the CTD itself is heavily post-translationally modified. The post-translational modification status could act as a code for the binding of different factors, and the modifications are also the consequence of various factors associating with the CTD at different times.

One of the major modifications that occurs is the phosphorylation of Ser2 and Ser5 residues. Enrichment of phosphorylated Ser5 at the promoter region and phosphorylated Ser2 towards the gene end has been observed (Li et al. (2007)). When Pol II is recruited to the promoter, it is unphosphorylated. The CTD gets phosphorylated at the Ser5 by one of the GTFs (TFIIH), which might lead Pol II to leave the promoter. The phosphorylation of Ser2 happens when it is released from

the pause site into elongation. As Pol II elongates through the gene body, CTD gets more phosphorylated by the elongation factor (P-TEFb), which travels with the Pol II, while Ser5 is gradually removed by the phosphatases. This leads to the changing ratio of Ser5/Ser2 phosphorylation of the CTD as Pol II moves over the gene body. The changing phosphorylation status serves as a signal for many of the histone-modifying enzymes, capping, splicing, and termination factors at different parts of the gene (Hsin & Manley (2012)). Moreover, the CTD has been recently shown to be also involved in phase transitions, at least in the in-vitro experiments, and the phosphorylation status of the CTD has been indicated to be involved in the phase-transition behavior (Boehning et al. (2018)).

1.1.1.1 Role of GTFs

Biochemical reconstitution systems revealed that in addition to the eukaryotic Polymerase II complex, additional factors are required for accurate transcription (Matsui et al. (1980), Weil et al. (1979)). Early *in vitro* studies with the purified GTFs suggested that transcription initiation begins with the ordered step-wise recruitment of GTFs and Pol II (TFIID, TFIIA, TFIIB, TFIIF and Pol II, TFIIE and TFIIH)(Buratowski et al. (1989), Conaway & Conaway (1993), Roeder (1996)). In this model, TFIID, which contains the TATA-binding protein (TBP) and the TBP associated factors (TAFs), first binds to the promoter. TBP is highly evolutionarily conserved (Hernandez (1993)) and is one of the well studied GTFs. Binding of TBP, a saddle-shaped protein (Nikolov et al. (1992)), to the TATA motif causes bending of the DNA (Chasman et al. (1993), Kim et al. (1993b), Kim et al. (1993a)), which could help in the formation of the initiation competent open complex. At the TATA-less promoters, TAFs have been proposed to interact with other promoter motifs to bring TFIID complex to the promoter (Burke & Kadonaga (1996), Pugh & Tjian (1991)), Purnell et al. (1994), Kaufmann & Smale (1994), Martinez et al. (1994), Chalkley & Verrijzer (1999)). For example, TAF2 is known in *in vitro* studies to bind to the initiator element (Inr) (Verrijzer et al. (1994)), which is capable of directing accurate transcription (Smale & Baltimore (1989)}). TAF9 and TAF6 can interact with the downstream promoter element (DPE) (Tora (2002), Burke

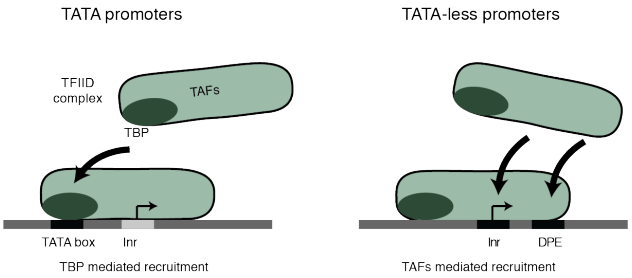

Figure 1.2: Recruitment of TFIID at TATA and TATA-less promoters: At TATA promoters the recruitment of TFIID might be driven by TBP. At TATA-less promoters it might be driven by the TAFs.

& Kadonaga (1997)). There are also proteins that are homologous to TBP and TAFs, which could assemble different TFIID complexes at different promoters. For example, TRF2 (Rabenstein et al. (1999)) is required for the transcription of the ribosomal protein genes, and cell proliferation genes (Hochheimer et al. (2002), Wang et al. (2014)). TRF3 has been shown to involved in the expression of muscle-specific genes in differentiated muscles, and genes involved in oogenesis (Deato et al. (2008), Goodrich & Tjian (2010)). Other than the promoter recognition role at the TATA-less promoters, TAFs might also play a role as coactivators at all promoters, by interacting with the activators or through their enzymatic activities (Verrijzer & Tjian (1996), Pugh & Tjian (1990), Sauer et al. (1995b), Sauer et al. (1995a), Jacq et al. (1994)). In addition to the TFIID complex, other promoter recognizing complexes containing TBP and TAFs also exist. For example, the SAGA complex, better known for its acetyltransferase activity, also can bring TBP and TAFs to promoters(Grant et al. (1997), Martinez et al. (1998), Ogryzko et al. (1998)). While it was initially thought that TFIID and SAGA complexes might act on distinct promoters (Li et al. (2000), Kuras et al. (2000)), recent studies suggest that such clear separations might not exist (Warfield et al. (2017)).

TFIIA is a multi-subunit complex, consisting of α, β, γ subunits (Cortes et al. (1992), Coulombe et al. (1992), Geiger et al. (1996), Tan et al. (1996)). Whether TFIIA is a GTF is debatable, however, it can act as an antirepressor and as a coactivator (MA et al. (1996)). It can bind to TBP dimers to induce TBP monomer formation, thereby inducing the binding of TBP to DNA (Coleman et al. (1999), Imbalzano et al. (1994)). It can also compete with other negative cofactors that destabilize

TBP binding to DNA or prevent it from interacting with other GTFs (Kokubo et al. (1998)). It can also act as a coactivator by interacting with activators (Kobayashi et al. (1995), Papai et al. (2010)), other GTFs and cofactors to effectively assemble the pre-initiation complex (PIC).

TFIIB is made of a single polypeptide (Ha et al. (1991), Hisatake et al. (1991), Malik et al. (1991)). It's binding to the TBP-TFIID complex stabilizes it (Nikolov et al. (1995), Wolner & Gralla (2001)). It also acts as an interface for the recruitment of the TFIIF and Pol II complex (Buratowski & Zhou (1993), Barberis et al. (1993)). TFIIB also interacts with the nascent RNA and the transcription bubble at the catalytic center and might stabilize the bubble (Liu et al. (2010), Pal et al. (2005)). TFIIB also plays a role in transcription start site selection (Pinto et al. (1992), Li et al. (1994), Berroteran et al. (1994), Hawkes & Roberts (1999)).

The Pol II and TFIIF complex binds next after the DNA-TFIID-TFIIA-TFIIB complex formation (Buratowski et al. (1989)). TFIIF is made up of RAP30 and RAP74 subunits (Conaway & Conaway (1989a), Price et al. (1989)). TFIIF helps in the recruitment of Pol II to the promoter (Flores et al. (1991), Killeen & Greenblatt (1992)). It might also prevent spurious initiation by preventing the binding of Pol II to non-promoter regions (Killeen et al. (1992)).

TFIIE, which is a tetramer with two copies of α and β subunits (Flores et al. (1989), Inostroza et al. (1991)), helps in the recruitment of TFIIH (Maxon et al. (1994)) and stimulates the TFIIH enzymatic activity (Ohkuma & Roeder (1994)) to form initiation competent PIC. It promotes the melting of the promoter DNA, exposing the single-stranded template for RNA synthesis (Goodrich & Tjian (1994)). Its role in promoter melting is supported by experiments which show that TFIIE and TFIIH are dispensable with premelted or supercoiled DNA templates (Pan & Greenblatt (1994), Parvin et al. (1994), Tantin & Carey (1994), Tyree et al. (1993)).

TFIIH is a multi-subunit complex (Conaway & Conaway (1989b), Feaver et al. (1991), Flores et al. (1991), Gerard et al. (1991)), which contains the subunits XPB, XPD, CDK7, and others, with ATPase, helicase, and CTD kinase activities (Tirode et al. (1999)). XPB and XPD subunits have ATPase and helicase activities. The XPB and XPD subunits unwind the DNA in the 3' to 5' and 5' to 3' direction, respectively (Fishburn et al. (2015)). The helicase activity of XPB helps

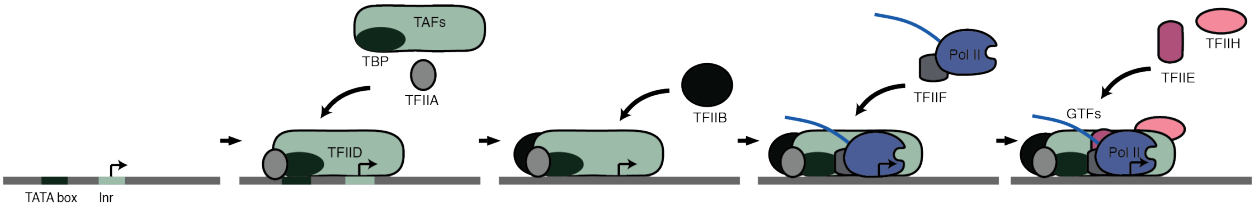

Figure 1.3: Stepwise assembly of GTFs and Pol II

in the melting of DNA by feeding the downstream DNA into fixed Pol II (Grünberg et al. (2012), Kim et al. (2000)). The ATPase and helicase activities help in the formation and the maintenance of the open complex formation (Fishburn et al. (2015), Dvir et al. (1997), Moreland et al. (1999)). During the initial synthesis of the RNA, the slippage of RNA 3'end from the active site results in the Pol II arrest. ATPase and the helicase activity could reduce Pol II arrest and abortive initiation during the early stages by stabilizing the transcription bubble (Holstege et al. (1997)). The RNA-DNA hybrid is greatly stabilized after the synthesis of about 25 nucleotides (Dvir (2002)). The phosphorylation of the CTD tail by the CDK7 subunit, which possesses the CDK kinase activity, could help further in the clearance of Pol II from the promoter and in the binding of other CTD tail associating proteins (Hsin & Manley (2012)).

While many of the *in vitro* experiments have shown that the PIC assembly happens through the step-wise recruitment of the GTFS and Pol II, isolation of unbound Pol II in a stable complex with several of the GTFs, as a holoenzyme (Koleske & Young (1994), Kim et al. (1994), Koleske & Young (1995)), suggests that the assembly of PIC could happen by recruitment of the TFIID complex and Pol II complex with GTFs and other cofactors. This will be similar to the bacterial system where the σ subunit first binds to the promoter, followed by the $\alpha\alpha\beta\beta'$ complex. Furthermore, live imaging studies have suggested that the binding stability of the GTFs on DNA is on the order of seconds (Darzacq et al. (2009), Zhang et al. (2016)). It is also important to note that the *in vitro* conditions might be different from the *in vivo* conditions leading to different stabilities. Knowing these details of PIC formation and the stability of different states would allow us to understand better the rate-limiting steps in transcription and how they might be regulated.

1.1.1.2 Promoter sequence

Transcription begins with the assembly of GTFs and Pol II on DNA. This assembly does not happen throughout the genome. Presence of nucleosomes prevents transcription initiation from most regions of the genome (Wasylyk & Chambon (1979), Knezetic & Luse (1986)). In addition to the requirement for accessible DNA, at least partially made accessible by the presence of dynamic histone variants or partial nucleosomes (Kubik et al. (2015), Chereji et al. (2017), Brahma & Henikoff (2019)), certain sequence features can lead to efficient assembly and initiation. A core promoter is a minimal sequence which can support basal levels of transcription (Weil et al. (1979)). The core promoter stretches 40 bp on either side of the TSS, consisting of several DNA motifs located generally at distinct positions with respect to the start site.

TATA sequence is one of the well-studied and well-conserved core promoter elements (Lifton et al. (1978), Goldberg (1979)). It is usually found about 25-30 bp upstream of the start site (Lenhard et al. (2012)). It recruits TFIID through its interactions with TBP (Kim et al. (1993b), Kim et al. (1993a)). The polar binding of TBP at the TATA site (Patikoglou et al. (1999)) could contribute to the orientation of the PIC. However, the orientation of the activators might have a greater contribution to the PIC orientation *in vivo*. Relatively fixed location of the TATA box suggests that it could affect the start site selection. In *S.cerevisiae*, TATA box is found 40-120 bp from the start site, where Pol II after assembly does scanning to find the start site. While in many other metazoans, including *S.pombe*, the start site selection happens where the PIC is formed without any scanning (Choi et al. (2002)). It is also important to note that TATA-box is not present at all promoters (Basehoar et al. (2004)).

Inr is another core-promoter motif. It acts as a recognition site for TFIID at TATA-less promoters (Smale & Baltimore (1989)). At TATA promoters, it can work cooperatively with the TATA motif to recruit TFIID, if they are within a certain distance (Emami et al. (1997)). If the TATA and Inr motifs are further apart, they act independently to recruit multiple PICs. Inr sequence also acts as a transcription start site at the +1A, which is preferred by Pol II (Javahery et al. (1994)). However, transcription can start at other nucleotides, if the preferred base is not found.

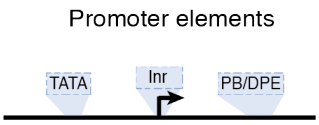

Figure 1.4: Promoter elements

Other motif sequences such as the TFIIB recognition element (BRE), Downstream promoter element (DPE, recognized by some TAFs), Pause button (PB), etc., are also enriched at promoters. At most promoters, not all the elements are present in the same promoter, suggesting that they might play a redundant role (Ohler et al. (2002), FitzGerald et al. (2006)). It has been observed that specific motifs and motif combinations are prevalent at various gene groups (Carninci et al. (2006)). TATA motif, together with a variant of Inr is found to be enriched at genes that are expressed in differentiated cells, in a very restrictive manner, the so-called effector genes (Schug et al. (2005), Lenhard et al. (2012), Carninci et al. (2006)). Meanwhile, the promoters of housekeeping genes in *Drosophila* have motifs such as Ohler1, Ohler6, and DRE (Holstege et al. (1997), Rach et al. (2009)). In addition, the initiation pattern at the housekeeping promoters is generally broad or dispersed (FANTOM Consortium and the RIKEN PMI and CLST (DGT) et al. (2014)). In mammals, the housekeeping genes are not enriched for any of the motifs found in *Drosophila* housekeeping genes. Instead, they just have elevated GC content and CpG dinucleotides, known as the CpG islands (Carninci et al. (2006)). Many developmental genes, such as the patterning and morphogenesis genes, have Inr, DPE, and PB in *Drosophila*, with focused initiation. In mammals, these genes have a broad initiation pattern and are associated with CpG dinucleotide sequences without enrichment for the motifs found in *Drosophila* (Akalin et al. (2009)). The significance of certain motif associations with specific gene groups is not clear, except in a few cases. In the developmental genes, the presence of paused Pol II has been proposed to be helpful for fast and synchronous activation, and the associated motifs might play a role in inducing pausing (Lagha et al. (2013), Boettiger & Levine (2009), Boettiger et al. (2011)).

Experiments that map Pol II-associated RNA, such as the GRO-seq and start-seq, have identified the presence of pervasive transcription throughout the genome in many cell types and organ-

isms (Core et al. (2014a), Meers et al. (2018)). The transcription initiation from many sites distal to the annotated start sites (ENCODE Project Consortium et al. (2007), Kapranov et al. (2007), Kim et al. (2010)), including many known enhancers, agree with the view that the presence of bona fide core-promoter motifs might not be essential for transcription initiation. In this model, transcriptional machinery tries to enhance with any open region with sequences that distantly resemble core promoter motifs. Closer, the found motifs at these open regions are to a bona fide motif; better, the region can support transcription initiation. Presence of activators and cofactors at enhancers might further simulate this fortuitous transcription.

1.1.2 RNA Polymerase II pausing

It came as a surprise when the widespread occurrence of promoter-proximal pausing was found through genome-wide studies of the late 2000s (Muse et al. (2007); Zeitlinger et al. (2007)). The general understanding of the field based on numerous studies in yeast (*S.cerevisiae*) had been that gene regulation happens at the level of recruitment of Pol II to the promoters (Ptashne & Gann (1997)) and once recruited Pol II transcribes the genes with very little regulation. In contrast to this view, the presence of paused Pol II after the initiation suggested a secondary regulatory step, perhaps even a more important one.

The occurrence of Pol II pausing had been observed, as early as the late 1970s and 80s, at a few genes such as β-globin (Gariglio et al. (1981b)), *c-myc* (Strobl & Eick (1992), Krumm et al. (1992)), *Fos* (Plet et al. (1995)) Drosophila Heat Shock Protein *Hsp* (Rougvie & Lis (1990), Rasmussen & Lis (1993)), and HIV-LTR (Kao et al. (1987)). While some studies in the 1990s suggested widespread occurrence of Pol II pausing (Rougvie & Lis (1990), Law et al. (1998)), the general applicability of pausing as a mode of regulation was appreciated only after the genome-wide studies of the late 2000s (Zeitlinger et al. (2007), Muse et al. (2007), Guenther et al. (2007), Core et al. (2008)). ChIP-seq studies on Pol II found an accumulation of Pol II near the promoter region of many genes, similar to the heat shock genes (Zeitlinger et al. (2007), Muse et al. (2007)). Based on the ratio of Pol II signal at the promoter region to Pol II signal at the gene body, a metric

known as pausing index, it was estimated that about 30-40% of all genes might be regulated by Pol II pausing. Other methods, such as GRO-seq and Start-seq, which map Pol II occupancy by sequencing the Pol II-associated RNA, have also identified the global role played by Pol II pausing in transcriptional regulation (Core et al. (2008), Nechaev et al. (2010)). While these studies identified genes with stable pausing, other studies involving knockdown of pausing inducing factors showed that pausing might be happening at all promoters, albeit at different efficiency and stability (Gilchrist et al. (2010)). Furthermore, the promoter-proximal pausing of Pol II might be necessary at all genes, even if it happens only transiently at certain genes (Shao & Zeitlinger (2017), Chen et al. (2015b)), for regulating other co-transcriptional processes.

Promoter proximal pausing of Pol II happens 20-40 nt downstream of the TSS (Core et al. (2008), Nechaev et al. (2010), Shao & Zeitlinger (2017), Giardina et al. (1992)). Paused Pol II has been proposed to play various roles at different genes, which might depend on pausing stability. Paused Pol II could acts as a checkpoint to ensure proper phosphorylation of Pol II and the recruitment of proteins involved in co-transcriptional events. Even transient pausing of Pol II, which happens at the housekeeping and TATA genes, might be enough to achieve this function. At genes which undergo stable pausing, paused Pol II might play other roles such as rapid gene activation, reducing transcriptional noise, regulate genome organization, etc.,

Pol II pausing is induced primarily by the binding of two factors (Wu et al. (2003)), Negative elongation factor (NELF)(Yamaguchi et al. (1999), Narita et al. (2003)) and DRB sensitivity inducing factor (DSIF)(Wada et al. (1998)). DSIF and NELF are sufficient two induce Pol II pausing in in-vitro assays. However, other factors and DNA features around the promoter might also play a role in Pol II pausing.

1.1.2.1 Contributions of *trans* factors to Pol II pausing

The NELF complex consists of the four subunits NELF-A, NELF-B, NELF-C/D, and NELF-E (Yamaguchi et al. (1999), Narita et al. (2003)). NELF is generally conserved in metazoans (Wu et al. (2005), Vos et al. (2016), Narita et al. (2003)). Its absence in yeast coincides with the

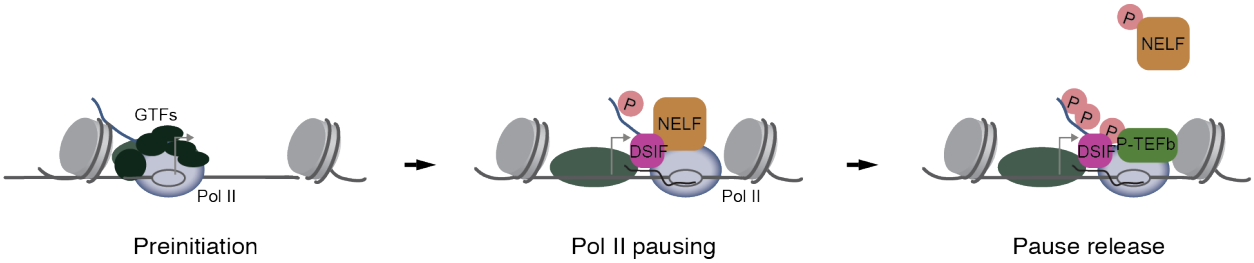

Figure 1.5: Pause inducing factors

absence of Pol II pausing, suggesting that it might be crucial for the Pol II pausing (Narita et al. (2003)). Structural studies involving the NELF-DSIF-Pol II complex provide mechanistic insights into how NELF might regulate Pol II pausing (Vos et al. (2018b)). It has been proposed that NELF could cause tilting of DNA–RNA hybrid in the paused state, and this could impair the binding of the nucleoside triphosphate substrate. NELF binding also could restrain Pol II mobility that is required for pause release. Additionally, NELF could prevent the binding of TFIIS. NELF-E has an RNA-binding domain and has been suggested to be involved in Pol II pausing (Yamaguchi et al. (2002)). Mutations in NELF subunits can cause developmental defects and cancers (McChesney et al. (2006), Midorikawa et al. (2002), Sun et al. (2008)). NELF-A is a candidate for the Wolf-Hirschhorn syndrome (Wright et al. (1999)). Similarly, the deletion of NELF-B leads to an inner cell mass deficiency and causes early embryonic legality in mice (Amleh et al. (2009)). NELF-B/E have been shown to be important for the proper endometrial function during pregnancy in mouse and humans (Hewitt et al. (2019)).

DSIF is composed of two subunits Spt4 and Spt5 (Yamaguchi et al. (1999), Wada et al. (1998)). DSIF could cause Pol II pausing by stabilizing the NELF interaction with Pol II. In contrast to NELF, DSIF also plays a positive elongation role (Wada et al. (1998)) and travels with Pol II after getting phosphorylated (Vos et al. (2018b)). DSIF is conserved from bacteria (structurally similar to the bacterial NusG) to humans, suggesting that at least its positive elongation role might be conserved. Mutations in DSIF subunits have been shown to cause developmental defects in zebrafish and *Drosophila* (Guo et al. (2000), Jennings et al. (2004)).

Paf1 complex has also been associated with both Pol II pausing, and Pol II release. Paf1

can cause Pol II pausing by restricting access the access of the super-elongation complex (SEC) (Chen et al. (2015b)), which stimulates Pol II elongation (Luo et al. (2012a), Luo et al. (2012b)). However, some studies suggest that Paf1 causes Pol II elongation may be in part by displacing the NELF complex (Vos et al. (2018a)). It is not clear how Paf1 could be involved in both stable Pol II pausing and pause release. Moreover, Paf1 can also affect transcription through histone modifications such as ubiquitination and methylation (Robzyk et al. (2000), Ng et al. (2003), Wu et al. (2014)). Other factors, such as G-down1 has also been implicated in Pol II pausing (Cheng et al. (2012)).

In addition to the above discussed DNA-binding factors, nucleosomes have also been proposed to play a role in Pol II pausing. It has been proposed that when Pol II encounters the +1 nucleosome, it might slow down. This could allow the pausing factors to bind to Pol II and cause pausing. Earlier studies came to contradicting conclusions regarding the role of +1 nucleosomes (Mavrich et al. (2008), Weber et al. (2014), Gilchrist et al. (2010)). It has been proposed that the contradiction could resolved by separating the paused genes into two groups (Li & Gilmour (2013)). The group with the GAGA motif, which is bound by GAGA factor (GAF), has strong Pol II pausing, caused by the recruitment of NELF by GAF. These genes might not be dependent on the +1 nucleosome for Pol II pausing. The other non-GAGA paused genes, have a strong +1 nucleosome but have a weaker binding of NELF and less accumulation of paused Pol II at the promoter, compared to the GAF dependent paused genes. These genes might be dependent on the +1 nucleosome for Pol II pausing. The role of +1 nucleosomes in Pol II pausing, at least in certain genes, is further supported by the studies which affect the stability of the +1 nucleosome by mutating the FACT complex (facilitates chromatin transcription) or acetylation of the +1 nucleosome (Tettey et al. (2019), Boija et al. (2017)). When the +1 nucleosome stability is reduced, the Pol II pausing is also reduced.

Pol II elongation is simulated by the recruitment of positive elongation factor (P-TEFb) (Marshall & Price (1995), Zhu et al. (1997), Peterlin & Price (2006)). P-TEFb is made up of cyclin-dependent kinase Cdk9 and CyclinT (Marshall et al. (1996), Peng et al. (1998a), Peng et al.

(1998b), Graña et al. (1994)).The human Cdk9 protein also binds cyclin K (Fu et al. (1999)). P-TEFb causes the phosphorylation of the Pol II CTD at Ser2 (Weeks et al. (1993), Egyházi et al. (1996)). It also phosphorylates NELF and DSIF (Fujinaga et al. (2004), Yamada et al. (2006)). This leads to the dissociation of NELF from Pol II and causes further elongation of Pol II with phosphorylated DSIF. P-TEFb has also been found to be part of the super-elongation complex (SEC) (Lin et al. (2010), Yokoyama et al. (2010)). P-TEFb can be recruited activators in different forms (Jang et al. (2005), Yang et al. (2005)). P-TEFb is generally present in an inactive state, in a complex with the HEXIM-1/2 and 7SK snRNA (Nguyen et al. (2001), Yang et al. (2001), Michels et al. (2004), Yik et al. (2003), Byers et al. (2005)). This inactive P-TEFb complex can be recruited to the promoter even before gene activation (Ji et al. (2013)). Pol II elongation can be achieved by the release of P-TEFb from the inactive complex (McNamara et al. (2013)).

1.1.2.2 Contributions of DNA sequences to Pol II pausing

In addition to their role in transcription initiation, the DNA sequences around promoters also strongly correlate with Pol II pausing, and mutations in some the promoter elements can affect Pol II pausing (Shao & Zeitlinger (2017), Shao et al. (2019), Hendrix et al. (2008), Chen et al. (2013)). It is not clear how these elements affect Pol II pausing. Different elements can affect Pol II pausing by various mechanisms. For example, the PB element is GC rich (Hendrix et al. (2008)). GC rich sequence can lead to strong RNA-DNA hybrids. The stability of the RNA-DNA hybrid can be an obstacle to the forward translocation of the polymerase leading to the backtracking of Pol II, with 3' end of the nascent RNA away from the active site (Palangat & Landick (2001), Nechaev et al. (2010), Tadigotla et al. (2006)). This could cause transcriptional arrest and allow the binding of pausing factors such as NELF. In addition to the GC rich motifs, promoters of many of the paused genes are GC rich such as the CpG island paused promoters in mammals. GAGA motif is also associated strongly with Pol II pausing (Hendrix et al. (2008), Lee et al. (2008), Lee et al. (1992), Shopland et al. (1995)) and can affect Pol II pausing by the recruitment of NELF by GAF (Li & Gilmour (2013)). Moreover, even sequences well upstream of the pausing site can

affect Pol II pausing, including the TATA box and the Inr sequence (Shao et al. (2019)). It has proposed that differences in the TFIID dynamics at different promoters could lead to a range of paused Pol II stabilities (Shao et al. (2019)). Certain sequences can directly recruit NELF and P-TEFb. The HIV-LTR gene produces a short hairpin loop at the 5' end of the synthesized RNA. Binding of NELF to nascent RNA can cause Pol II pausing. Binding of an activator protein TAT to the RNA loop, brings P-TEFb and dislodges NELF, leading to the elongation (He et al. (2010), Sobhian et al. (2010), Muniz et al. (2010)).

1.1.2.3 Elongation

The release of Pol II from the promoter-proximal pause site by P-TEFb leads to the synthesis of the rest of the gene. The polymerase progression through this phase is relatively few slowdowns and backtracking. One such region is at the intron-exon boundaries presumably through its interactions with the spliceosomal complex. The intron retention or alternate spicing is dependent on the Pol II elongation rate (Bentley (2014), Saldi et al. (2016)). However, the details of such regulations are not clear. The binding of several elongation factors with Pol II enables the increased processivity of Pol II with fewer backtrackings post-pause-release. Some the major elongation factors include elongin (Bradsher et al. (1993a), Bradsher et al. (1993b)), TFIIS (Sekimizu et al. (1976)), TFIIF, DSIF, ELL (Shilatifard et al. (1996)), which is also a component of SEC, etc. These factors stimulate elongation in various ways, including by stabilizing the 3' nascent RNA near the catalytic site, stabilizing Pol II in a more active conformation, cleave backtracked RNA, etc. The phosphorylation of the CTD tail could also play a role in this by recruiting the elongation factors (Phatnani & Greenleaf (2006), Hsin & Manley (2012)). The association of chromatin remodelers allows Pol II to travel through nucleosomes. It is not clear if certain genes regulate gene output by primarily controlling the elongation rate post-pause-release. Following the complete transcription of the gene, Pol II and the nascent RNA are released from the DNA. The released Pol II is recycled for re-initiation in some cases. I will not go into the mechanisms of termination further here.

1.1.2.4 Establishment of paused Pol II

How is paused Pol II established at gene promoters? One model is that the presence of paused Pol II is the default state of promoters. Since it is the elongation of Pol II from the paused site that leads to genes expression, the establishment of paused Pol II could be inconsequential and could be the default promoter state. The other possibility is that the paused Pol II establishment is tightly regulated. The predicted nucleosome occupancy at paused promoters is very high (Gaertner et al. (2012), Xi et al. (2010), Tillo et al. (2010)), and it presumably has to be actively removed to allow access to GTFs and Pol II (Wasylyk & Chambon (1979), Knezetic & Luse (1986), Gilchrist et al. (2010)). This argues against the model that the presence of paused Pol II is the default state of promoters. Moreover, the paused Pol II occupancy has been observed to change during the *Drosophila* embryogenesis (Gaertner et al. (2012), Chen et al. (2013)). While many promoters recruit paused Pol II during the zygotic genome activation, certain genes recruit Pol II only during the late stage of embryogenesis ((Gaertner et al. (2012)). This suggests that the paused Pol II occupancy is regulated, and the presence of paused Pol II is not the default state. Similarly, in mammals, where the promoters sequences are also favorable for nucleosome assembly (Tillo et al. (2010)), it is also observed that the promoter accessibly changes during development (Lu et al. (2016)). This suggests the paused Pol II occupancy might also be regulated in mammals. However, it is not clear how is the paused Pol II establishment regulated and what might be the purpose of such a regulation. Since the promoter nucleosome undergoes a major change with the recruitment of Pol II, we propose that chromatin plays an important regulatory role in the regulation of Pol II establishment at the paused genes.

1.1.3 Chromatin and Gene expression

There is a necessity for all living organisms to compact the DNA, in order to be able to package the DNA into a cell which is many-fold smaller in diameter compared to the length of DNA. This is primarily achieved by the binding of basic histone proteins to the negatively charges phosphate backbone of DNA. Other non-histone proteins, like transcription factors, polymerases, RNA, etc.

also bind to the DNA to form a complex called as chromatin (because of its ability to bind to certain dyes, chroma means color). The compaction of DNA limits the access of the genetic information for replication and transcription and also protects it from damage. Thus, chromatin play a crucial role in transcriptional regulation (Struhl (1999)).

The fundamental unit of chromatin is nucleosome, composed of highly conserved histone subunits H2A, H2B, H3, H4 (Kornberg & Thomas (1974), Kornberg (1974), Kornberg (1977)). The nucleosome core is made of the tetramer formed by the subunits H3-H4. H2A and H2B dimers flank the nucleosome core on either side (Richmond et al. (1984), Luger et al. (1997)). This octamer forms the repeating structure of the chromatin, positioned approximately every 175bp; of this, about 147bp of DNA is wrapped around the octamer core. The DNA sequence in between the octamers is known as the linker region bound by the linker histone, H1. This arrangement of chromatin leads to 11nm fibers seen in electron micrographs. The chromatin can be further compacted at certain regions of the genome. The compaction can be influenced by the proteins binding to the regions; many of them regulate and are regulated by the post-translational motifications of histones. The post-translational modifications of histones play an important regulatory role in the regulation of gene expression. Acetylation, and methylation are some the most common histone modifications, and they can affect/correlate with the transcriptional state of the gene. Histone modifications also function as epigenetic marks that can be inherited by daughter cells.

Compaction of the genome varies with transcriptional state and location along the chromosome. Traditionally, chromatin was divided into heterochromatin and euchromatin, which corresponds to the transcriptionally inactive and active regions (Heitz (1928), Elgin (1996), Grewal & Elgin (2002)). The heterochromatin can be further divided into facultative and constitute heterochromatin, also known as the blue and green chromatin, respectively (Filion et al. (2010)). The constitutively silent heterochromatin region is present at the centromeric and telomeric regions, marked by the H3K9Me2 marks and are bound by proteins such as HP1, SU(VAR)3-9, etc. The facultative heterochromatin are regions silenced in a developmentally regulated manner, marked by H3K27Me3 marks and are bound by the Polycomb group of proteins. Other than the above

mentioned classical repressive region, another major form of repressive chromatin, known as the black chromatin, has been found (Filion et al. (2010)). How repression is achieved in the black chromatin is not clear. The active regions also further classified into the housekeeping (Yellow) and developmentally regulated (Red) regions, which have been found to bind to distinct proteins. This broad classification of the genome captures, in essence, the diverse proteins associated with the chromatin. The functions of many of the proteins associated with these regions are not known. Moreover, while there is a strong correlation between the gene expression and the chromatin-associated proteins and histone marks, the nature of causality is also clear.

The chromatin is modified constantly during different transcriptional states. The chromatin is inhibitory to both transcription initiation and elongation (Wasylyk & Chambon (1979), Knezetic & Luse (1986)). During initiation, the promoter nucleosome has to be removed or modified to allow GTFs and Pol II access to the DNA. In yeast, the majority of genes, which are constitutively expressed, have a nucleosome-free/depleted region (NFR or NDR) upstream of the TSS, covering the also proximal regulatory sequences (Bernstein et al. (2004), Lee et al. (2004), Yuan et al. (2005)). The NDR is caused by strong anti-nucleosomal sequences such as poly(dA:dT) (Field et al. (2008), Segal & Widom (2009)). However, the regulated genes have a promoter nucleosome covering the TATA box, which is frequently present at the regulated promoters, and the near-by proximal regulatory region. In higher eukaryotes, the promoter sequences intrinsically favor nucleosome occupancy. However, many of them nucleosome depleted, presumably by active chromatin remodeling processes. During elongation, nucleosomes present a challenge to the transcribing Pol II (Hodges et al. (2009), Bintu et al. (2012), Churchman & Weissman (2011), Weber et al. (2014)). The Pol II-associated factors constantly modify, evict, and reassemble the nucleosome to allow passage of Pol II, meanwhile also preventing cryptic initiations. Pol II-associated factors such as FACT (Orphanides et al. (1998), LeRoy et al. (1998), Orphanides et al. (1999)), Spt16 (Bortvin & Winston (1996)), Asf1 (Schwabish & Struhl (2006)) has been shown to be involved in transcription-associated remodeling.

Similarly, at enhancers, the presence of nucleosomes prevent the transcription factors from

accessing the DNA. Binding of many transcription factors depends on a special class of proteins known as pioneer factors that can bind to their DNA motifs even when they are masked by nucleosomes (Zaret & Carroll (2011)). Binding of pioneer factors can allow the binding of other factors by altering the chromatin structure. Pioneer factors can alter the chromatin, either simply through their binding, freeing up more DNA which could be bound by other factors or through the recruitment of chromatin remodelers. The mechanism of action by many of the pioneer factors is not clear. In addition to the pioneer factor-mediated binding, transcription factors can bind to their motif through cooperative interactions, when multiple motifs are present (Workman & Kingston (1992)). They can bind through just simple mass action when they are present in very high concentrations (Workman & Kingston (1992), Morse (1993)). They can also out-compete histones when they undergo natural turn over. Transcription factors can thus have pioneering activity along a spectrum.

The pioneering activity was initially discovered at the Alb1 locus, which is occupied by GATA1 and FoxA1 in undifferentiated gut endoderm, where Alb1 is silent (Liu et al. (1991), Gualdi et al. (1996), Bossard & Zaret (1998)). The GATA1 and FoxA1 binding sites are crucial for the proper Alb1 expression in the later stages. The FoxA1 homolog in *C.elegans*, PHA-4 also acts as a pioneer factor in endoderm development (Gaudet & Mango (2002), Fakhouri et al. (2010)). Some of the other well-known pioneer factors include Oct2, Sox4, PU.1 in mammals. In Drosophila, Zelda is a well-known pioneer factor in the early embryo (Liang et al. (2008), Harrison et al. (2011)). The role of pioneer factors is not always to open the chromatin; some of the pioneer factors act as place holders by preventing further repression. FoxD3 binds to many of the late endodermal genes in ES cells and prevent the methylation of the region. It might be acting as a place holder for the late binding of FoxA1 (Xu et al. (2009)).

Enhancer activation, in many cases, starts with the expression of a pioneer factor at a particular stage and tissue (Spitz & Furlong (2012), Zaret & Carroll (2011)). This can initiate chromatin occupancy or position changes and prime the enhancer for future activation. In the right context, the enhancer might become active, while in other contexts, it might become repressed.

Enhancers in different states are associated with various epigenetic signatures. Active enhancers usually have H3K4me1, H3K79me3, H3K27ac, and RNA polymerase II binding (Creyghton et al. (2010), Heintzman et al. (2009), Heintzman et al. (2007), Rada-Iglesias et al. (2011), Bonn et al. (2012b)). Many of the repressed enhancers have histone marks such as H3K27me3. Certain enhancers can have only some of the active marks (H3K4me1 and low H3K27Ac) (Creyghton et al. (2010), Rada-Iglesias et al. (2011), Zentner et al. (2011)). These enhancers are in a 'poised' state. This can be an intermediate state ready for rapid future activation. The majority of active enhancers can get activated without any observable poised state in the previous developmental stage (Bonn et al. (2012b), Choukrallah et al. (2015)). Therefore this type of priming might be required only for a subset of enhancers. The poised mark has been proposed to be a result of active competition between histone acetylases and deacetylases, leading to ultra-fast zero-order activation kinetics (Koenecke et al., 2017). Therefore, such priming might be important at enhancers with rapid activation kinetics. The presence of H3K4me1 in the absence of H3K27ac can also be a result of active repression(Koenecke et al. (2017)).

While the histone modification enzymes which catalyze these activities are known, how they are recruited is not always known. Moreover, mutations in the histone modifiers do not always affect the enhancer function. For example, while the loss of P300 (enzyme which catalyzes the H3K27Ac) activity leads to reduced activation (Raisner et al. (2018)), mutations in COMPASS leads to loss of H3K4Me1, but the enhancer function is not lost (Dorighi et al. (2017), Rickels et al. (2017)). Thus, the functions of many of the histone modifications are not clear.

Similar to a poised enhancer, promoters with both active and inactive histone marks, in a bivalent or poised state have been found. The promoters of many of the developmental genes exist in a bivalent state, with both H3K4Me3 and H3K27Me3, in the ES cells, presumably allowing them to differentiate upon differentiation cues (Bernstein et al. (2006), Marks et al. (2012), Young (2011), Min et al. (2011),Brookes et al. (2012), Ku et al. (2008)). It is not clear if the bivalent promoters have paused Pol II. While some earlier studies, found cooccurrence of paused Pol II with the bivalent promoters, it was later found that the overlap between the bivalent and the paused

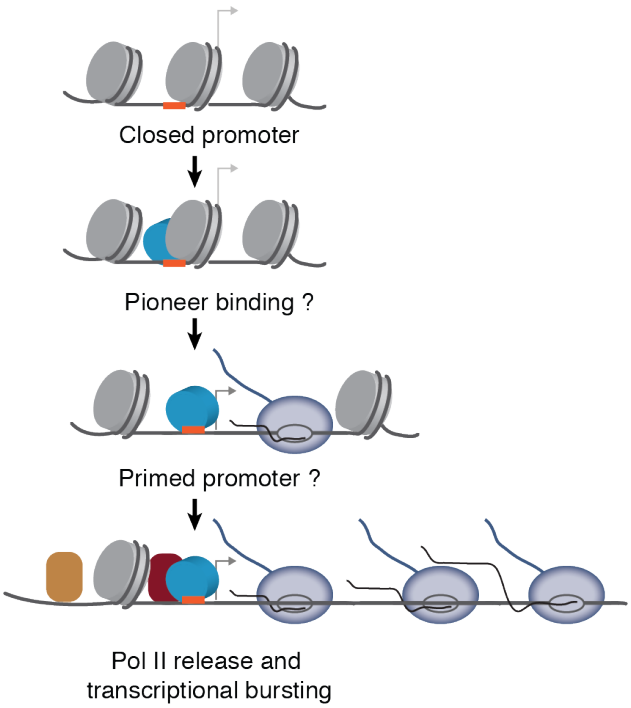

Figure 1.6: Establishment of paused Pol II

promoters are no more than the overlap between all promoters and the paused promoters (Williams et al. (2015)). The paused Pol II is enriched at the signaling genes and might regulated the pluripotency by fine-tuning the expression of the signaling molecules (Williams et al. (2015), Gilchrist et al. (2012)). Thus, the promoter state with paused Pol II might be functionally equivalent to the primed or poised state at the enhancers.

1.2 Aims of this thesis

In this thesis, I aim to address several important open questions regarding the establishment of Paused Pol II. We will analyze the late paused genes that recruit Pol II during the late stages of *Drosophila* embryogenesis but show no Pol II during the early stages (Gaertner et al. (2012)). These genes provide a clear example of PoII occupancy changes at paused promoters. The promoter nucleosome occupancy undergoes a major change during Pol II recruitment (Gaertner et al. (2012)). I hypothesize that the factors involved in the recruitment of Pol II at the paused genes might function by regulating the promoter nucleosome occupancy.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Fly stocks

Wild type flies (OregonR) were obtained from the Bloomington Drosophila Stock Center at Indiana University.

2.1.1.1 Intact flies

The UAS RAN-GAP-mcherry-FLAG-BirA lines were crossed with tissue specific gal4 driver lines to generate fly stocks expressing tissue specific RAN-GAP-mcherry-FLAG-BirA in six different tissues- Neuron, Glia, Trachea, Epidermis, Muscle, Gut (Foregut and hindgut) (Lagha et al., 2013). The genotypes of fly lines are mentioned in Table 2.1.

2.1.1.2 Lola-I mutants and overexpression flies

Lola-I mutant lines were obtained from Bloomington stock center (ORC4 - 28267) and from Ed Giniger (ORE50). Homozygous Lola-I mutant files were non-viable and were therefore maintained with a Cyo-GFP balancer to enable the sorting of the homozygous mutant embryos. Lola-I rescue lines are generated as follows. We first inserted actin-promoter driving full-length Lola-I cDNA construct into the attP40 locus and then crossing these Lola-I expressing lines with ORC4 lines to obtain the meiotic recombinants.

2.1.2 Antibodies

Lola-I, Rpb3 (Pol II), and GAF antibodies were custom made using Genescript services.

Tissue expression	Genotype	Gal4 driver bloomington stock number
Neuron	8760	w[*]; Pw[+mC]=GAL4-elav.L3, p[UAS-3xFLAG-blrp-mCherry- RanGap, UAS-BirA)6
Glia	7415	w[*]; Pw[+m*]=GAL4repo, p[UAS-3xFLAG-blrp-mCherry- RanGap, UAS-BirA)6/TM6Tb
Muscle	27390	w[*]; Pw[+mC]=GAL4-Mef2.R3, p[UAS-3xFLAG-blrp-mCherry- RanGap, UAS-BirA)6/TM3Sb
Trachea	8807	w[*]; Pw[+mC]=GAL4-btl.S2, Pw[+m*]=lacZ-un8276, p[UAS-3xFLAG-blrp-mCherry- RanGap, UAS-BirA)5/Cyo
Foregut and hindgut	110394	w[*]; PGawBNP3084 , p[UAS-3xFLAG-blrp-mCherry-RanGap, UAS-BirA)5/Cyo
Epidermis	7021	Pw[+mW.hs]=GawB112A, w[*],p[UAS-3xFLAG-blrp-mCherry- RanGap, UAS-BirA)6

Table 2.1: A list of INTACT flies used

2.2 Experimental Methods

2.2.1 Embryo collection

Adult flies were maintained in population cages for embryo collections, and embryos were collected and matured on apple juice plates at 25 °C in cages. For example, 14-17h AED embryos were collected on apple juice plates for three hours at 25 °C in cages and then matured at 25 °C for another 14 h. Embryos were dechorionated for 1 min with 67% bleach then crosslinked for 15 min with 1.8% formaldehyde (final concentration in water phase). For the smFISH experiments, the embryos were crosslinked in 1x PBS in DEPC treated water. Homozygous Lola mutant embryos were obtained by sorting for GFP negative embryos in PBT. Embryos were flash-frozen in liquid nitrogen and stored at -80°C. For ATAC-seq and scRNA-seq experiments, the embryos were processed immediately after dechorionation without crosslinking.

2.2.2 Embryo sorting

Embryos were sorted using the COPAS embryo sorter. Embryos were first gated such that the debris and the out of stage withheld larvae were removed, and the majority of embryos were still retained. The gated embryos were then sorted for the absence of the GFP signal. With the $lola^{\text{ORC4}}/^{\text{Cyo-GFP}}$ mutant lines, the GFP double-positive Cyo-GFP/Cyo-GFP embryos and the Cyo-GFP/ORC4 embryos were easily separable from the homozygous GFP negative ORC4 mutant embryos. Embryo sorting was done in PBT with 0.01% triton or 0.1% triton.

2.2.3 Isolation of tissue-specific nuclei

Nuclei isolation was performed using previously published protocols with modifications (Bonn et al. 2012; Deal and Henikoff 2011). Nuclei were isolated by douncing 0.5 g of embryos in HBS buffer (0.125 M Sucrose, 15 mM Tris (pH 7.5), 15 mM NaCl, 40 mM KCl, 2 mM EDTA, 0.5 mM EGTA, 2% BSA, protease inhibitors) in a 15 ml dounce tissue grinder followed by filtering the nuclei suspension through two layers of miracloth (Calbiochem, #475855). Nuclei were spun at

500g for 10 min at 4°C, and the supernatant was discarded. The nuclear pellet was resuspended in HBS buffer and dissociated using syringe (22.5-gauge needle) 10 times. Nuclei were spun again, and the pellet was resuspended an HBS buffer and incubated with Dynabeads M-280 Streptavidin beads (Invitrogen, # 11205D) for 30 minutes with end to end rotation at 4°C. A magnet was used to separate the bead-bound nuclei, and the beads were washed thoroughly with HBS buffer.

2.2.4 ChIP-seq experiments

ChIP-seq experiments were performed as described (Chen et al., 2013) with the following differences: 100 mg OregonR embryos were used per IP. 5 µg chromatin was used for tissue-specific ChIP-seq experiments. After incubation of magnetic beads with antibodies, tissue-specific IP samples were washed three times by rotating tubes for 3 min at 4°C to reduce background. Libraries were prepared according to the manufacturer's instructions. ChIP-seq libraries were prepared from 5-15 ng ChIP DNA or 100 ng WCE input DNA.

2.2.5 ATAC-seq experiments

ATAC-seq was performed using about 500-2000 embryos of stage 14–17 h AED. Nuclei were isolated by douncing the embryos in HBS buffer, as mentioned above, in the isolation of tissue-specific nuclei section. Whole embryo ATAC-seq was performed without the selection of nuclei from any specific tissue using the OregonR embryos. The transposition of the nuclei was performed as described (Buenrostro et al., 2013) using 2.5 μl Tn5 transposase, followed by PCR amplification (Nextera DNA Sample Preparation Kit: FC-121-1030, Illumina) and library preparation (the Nextera index kit: FC-121-1011, Illumina) was used to create libraries. Libraries were purified using Agencourt AMPure XP beads (A63881, Beckman Coulter), and paired-end sequencing was performed on the NextSeq 500 instrument (Illumina). Following sequencing, the chromatin accessibility is calculated by computationally filtering for fragments of size 0-100 bp, which represent small fragments that are supposedly from the accessible regions.

2.2.6 MNase-seq experiments

MNase digestion was performed similar to (Chen et al., 2013). Briefly, chromatin was extracted from 0.1mg of wt OregonR embryos per replicate, then digested with a concentration MNase (Worthington Biochemical Corporation #LS004798) gradient for 30 min at 37°C. All the digestion gradients were run on a gel, and the digestion concentration to be sequenced was chosen such that the digestion is complete, but the samples are not over digested as mentioned in (Chen et al. 2013). Libraries were prepared using the NEBNext DNA Library Prep kit following the manufacturer's instructions and then paired-end sequenced on an Illumina HiSeq 2500 sequencing system. The nucleosome sized fragments (100-200 bp) were selected computationally following sequencing to analyze the nucleosome occupancy.

2.2.7 mRNA-seq experiments

Total mRNA was extracted from non-cross-linked embryos using the Maxwell Total mRNA purification kit (Promega, #AS1225) according to the manufacturer's instructions. PolyA-mRNA was isolated using DynaI oligo(dT) beads (Life Technologies, #61002). Libraries were prepared following the instructions of the TruSeq DNA Sample Preparation Kit (Illumina, #FC-121-2001) and sequenced on the HiSeq 2500 and the Nextseq 500 (Illumina).

2.2.8 scRNA-seq experiments

scRNA-seq experiments were performed on 14-14.5h AED wildtype OregonR embryos. The isolation of single cells was performed similar to the previously published protocol (Karaiskos et al., 2017), with the following modifications. The embryos were dounced in the SFX medium with 0.1% PF68 + 0.1% BSA, which was found to improve the viability. The total number of dounces was increased to 120 to improve the isolation of cells from the late-stage embryos. The isolated cells were filtered and washed and then resuspended in Schneider's medium to avoid any interference with droplet formations in the subsequent steps. The resuspended cells were immediately processed in the 10x Genomics instrument with optimal loading at a targeted capture rate of about

6000 cells per run to minimize doublets. RNA isolation and cDNA synthesis and amplification were done according to the manufacturer's instructions. Libraries were then prepared following the instructions of the TruSeq DNA Sample Preparation Kit (Illumina, #FC-121-2001) and sequenced on the HiSeq 2000. scRNA-seq experiments were performed on two biological replicates on separate days from different cages.

2.2.9 Western blot

Whole-cell extracts were obtained from embryos through douncing followed by sonication in A2 lysis buffer. Samples are denatured by adding Laemmli buffer and heating at 95c for 5-10 min. Denatured proteins were separated through sodium dodecyl sulfate-polyacrylamide gel electrophoresis and transfer on to a nitrocellulose or PVDF membrane. Proteins were detected and quantified by incubating the membrane with primary antibodies against the proteins of interest, followed by incubation with secondary fluorescent antibodies. LICOR scanner was used to detect the fluorescent signal from the blot.

2.2.10 Immunostaining

Whole *Drosophila* embryos were prepared for immunostaining by removing the chorian and vitelline membranes using bleach and methanol/heptane mix, respectively. Permeabilization and preservation is achieved by washing in methanol. Methanol permeabilized dehydrated embryos were rehydrated; incubated with primary antibodies against the proteins of interest; followed by incubation with fluorescent secondary antibodies. Embryos were mounted in 2.5% DAPCO and imaged using confocal or epifluorescent microscopes.

2.2.11 In-vitro Nucleosome binding assay

2.2.11.1 Purification of Lola-I

Purified Lola-I protein was produced commercially by Genescript. Briefly, Lola-I full-length sequence was synthesized and sub-cloned, with Flag-TAG to facilitate the purification, into expres-

sion vector F1 for insect cell expression. DH10Bac strain was used for the recombinant bacmid (rbacmid) generation. Sf9 cells were grown in Sf-900 II SFM Expression Medium. The cells were maintained in Erlenmeyer Flasks at 27 °C in an orbital shaker. One day before transfection, the cells were seeded at an appropriate density in 6 wells. On the day of transfection, DNA and Cellfectin II were mixed at an optimal ratio and then added into the plate with cells ready for transfection. Cells were incubated in Sf-900 II SFM for 5-7 days at 27 °C before harvest. The supernatant was collected after centrifugation and designated as P1 viral stock. P2 was amplified for later infection.

The viral stock was later used to infect the cells. The Sf9 cells (1 L) containing 5% FBS were infected with the P2 virus at MOI=3 and harvested at 48 h post-infection. Cells were sonicated in 50 mM Tris, pH 8.0, 150 mM NaCl, 5% Glycerol containing protease inhibitor. The supernatant after centrifugation was purified by the Flag column. Cell pellets were harvested and lysed by proper cell lysis buffer. The cell lysate supernatant was incubated with Flag Columns to capture the target protein. Fractions were pooled and dialyzed with 50 mM Tris-HCl, 150 mM NaCl, 5% Glycerol, pH 8.0 followed by 0.22 μm filter sterilization. Proteins were analyzed by SDS-PAGE and Western blot by using standard protocols for molecular weight and purity measurements.

2.2.11.2 Design of the nucleosome positioning templates

Nucleosomal templates were derived from the Widom 601 strong nucleosome-positioning sequence. Lola-PI TFBS AAAGCTC were added at differing positions to ensure that the TFBS are either in an exposed or concealed orientation as determined by the crystal structure of a nucleosome with the Widom 601 sequence (Makde et al., 2010). Therefore, ten templates were designed starting from the 217-bp Widom 601 sequence and compared to non-specific binding to one control sequences (Table. 1). Starting with the Widom 601 nucleosome positioning sequence, ten templates were designed and compared to non-specific binding to the control sequence (Table. 1). The Widom 601 nucleosome positioning sequence produces highly stable and optimally positioned nucleosomes. Since the NCP structure is in principle, dyad symmetric, we primarily focused on

the right half of the core particle. With increasing distance to the dyad axis, four translational settings were tested – dyad (R0, R0.5), intermediate (R4, R4.5), edge (R6, R6.5, R7), and linker which is outside the NCP. At each translational position within the NCP, two rotational settings were tested by shifting the Lola-PI TFBS 5-bp to the right. TFBS accessibility was determined by modeling the TFBS position onto the nucleosome crystal structure formed from the Widom 601 sequence (Makde et al., 2010). In addition, to explore whether two neighboring TFBS will enhance Lola-PI binding to the nucleosomes, a template with two neighboring TFBSs was designed (R6+R7), another template with 2 TFBSs at a further-away distance was used as control (L6+R7).

Label	Sequence
R0	GATGGACCCTATACGCGGCCGCCCTGGAGAATCCCGGTGCCGAGGC CGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTAC **AAAGCTC**CCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCT CCAGGCACGTGTCAGATATATACATCCTGTGCATGTATTGAACAGCGA CCTTGCCGGTGCCAGTCGGATAGTGTTCC
R0.5	GATGGACCCTATACGCGGCCGCCCTGGAGAATCCCGGTGCCGAGGC CGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTAC GCGCT**AAAGCTC**GCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCT CCAGGCACGTGTCAGATATATACATCCTGTGCATGTATTGAACAGCGA CCTTGCCGGTGCCAGTCGGATAGTGTTCC
R4	GATGGACCCTATACGCGGCCGCCCTGGAGAATCCCGGTGCCGAGGC CGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTAC GCGCTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCC**AAAGCT C**CAGGCACGTGTCAGATATATACATCCTGTGCATGTATTGAACAGCGA CCTTGCCGGTGCCAGTCGGATAGTGTTCC

Label	Sequence
R4.5	GATGGACCCTATACGCGGCCGCCCTGGAGAATCCCGGTGCCGAGGC CGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTAC GCGCTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTC **AAAGCTC**ACGTGTCAGATATATACATCCTGTGCATGTATTGAACAGC GACCTTGCCGGTGCCAGTCGGATAGTGTTCC
R6	GATGGACCCTATACGCGGCCGCCCTGGAGAATCCCGGTGCCGAGGC CGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTAC GCGCTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCT CCAGGCACGTGTCA**AAAGCTC**ACATCCTGTGCATGTATTGAACAGCG ACCTTGCCGGTGCCAGTCGGATAGTGTTCC
R6.5	GATGGACCCTATACGCGGCCGCCCTGGAGAATCCCGGTGCCGAGGC CGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTAC GCGCTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCT CCAGGCACGTGTCAGATAT**AAAGCTC**CTGTGCATGTATTGAACAGCG ACCTTGCCGGTGCCAGTCGGATAGTGTTCC
R7	GATGGACCCTATACGCGGCCGCCCTGGAGAATCCCGGTGCCGAGGC CGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTAC GCGCTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCT CCAGGCACGTGTCAGATATATACA**AAAGCTC**CATGTATTGAACAGCG ACCTTGCCGGTGCCAGTCGGATAGTGTTCC
R8	GATGGACCCTATACGCGGCCGCCCTGGAGAATCCCGGTGCCGAGGC CGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTAC GCGCTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCT CCAGGCACGTGTCAGATATATACATCCTGTGCAT**AAAGCTC**ACAGC GACCTTGCCGGTGCCAGTCGGATAGTGTTCC

Label	Sequence
R6+R7	GATGGACCCTATACGCGGCCGCCCTGGAGAATCCCGGTGCCGAGGC CGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTAC GCGCTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCT CCAGGCACGTGTCA**AAAGCTC**ACA**AAAGCTC**CATGTATTGAACAGCG ACCTTGCCGGTGCCAGTCGGATAGTGTTCC
L6+R7	GATGGACCCTATACGCGGCCGCCCTGGAGAAT**AAAGCTC**CCGAGGC CGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTAC GCGCTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCT CCAGGCACGTGTCAGATATATACA**AAAGCTC**CATGTATTGAACAGCG ACCTTGCCGGTGCCAGTCGGATAGTGTTCC
601 control	GATGGACCCTATACGCGGCCGCCCTGGAGAATCCCGGTGCCGAGGC CGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTAC GCGCTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCT CCAGGCACGTGTCAGATATATACATCCTGTGCATGTATTGAACAGCG ACCTTGCCGGTGCCAGTCGGATAGTGTTCC

Table 2.2: Position of Lola-I motifs in different 601 templates

2.2.11.3 In vitro nucleosome reconstitution & purification

All 11 synthesized DNA templates were amplified via PCR with the following primers: 5'-GATGGACCCTATACGCGGC-3',

5'-GGAACACTATCCGACTGGCA-3'

After PCR, all DNA fragments were column-purified (QIAGEN) and quantified. In vitro nucleosomes were generated from H2A/H2B dimer and H3.1/H4 tetramer (NEB). All 11 nucleosome sequences were mixed at equal molar amounts. Mixed DNAs were then added to histones at octamer/DNA molar ratios of 1.25:1 in 2M NaCl solution. Nucleosomes were reconstituted through

salt gradient dialysis as described (Hayes & Lee, 1997), which were further purified by 7%-20% sucrose gradient centrifuge (Fang et al., 2016) and concentrated by 50K centrifugal filter units (Millipore, AmiconR Ultra).

2.2.11.4 DNA binding assay followed by EMSA

The protein-nucleosome binding assays were carried out with purified nucleosomes mentioned above and human recombinant Lola-PI protein in 7 µl DNA binding buffer (10mM Tris-Cl, pH7.5; 50mM NaCl; 1mM DTT; 0.25mg/ml BSA; 2mM MgCl2; 0.025% P-40; and 5% glycerol), and then incubated for 10 minutes on ice and then for 30 minutes at room temperature. Increasing concentrations of Lola-PI protein were added to 0.25 pmol purified nucleosomes, from Lola/nucleosome molar ratios of 0:1, 1:1, 2:1, 4:1 to 10:1. Protein binding was detected by mobility shift assay on 4% (w/v) native polyacrylamide gels (acrylamide/bisacrylamide, 29:1, w/w, 7 × 10 cm) in 0.5 × Tris Borate-EDTA buffers at 100V at 4 ºC. After electrophoresis, DNA was imaged by staining with SYBR Green (LONZA).

2.2.11.5 DNA isolation and purification

All visual bands were excised from the gel, as well as the bands at the same locations in the other lanes. Each gel slice was processed separately for a total of 20 samples from 2 replicate experiments. In order to extract DNA from polyacrylamide gel, the chopped gel slices were soaked in diffusion buffer (0.5 M ammonium acetate; 10mM magnesium acetate; 1mM EDTA, pH8.0; 0.1% SDS), and incubated at 50 ºC overnight. The supernatant was collected, residual polyacrylamide removed with glass wool, and DNA purified with QIAquick Spin column (QIAGEN). The DNA concentration for each sample was determined by qPCR by comparing to a standard curve generated from the control 601 sequence.

2.2.11.6 Library construction & sequencing

After we obtained the concentrations of the purified DNA samples, we performed two rounds of PCR to construct sequencing libraries for Illumina sequencing. The first-round PCR used forward-amplicon-primer and reverse–amplicon primer (see below). The number of cycles for PCR was determined by the sample concentration determined by qPCR and ranged from 8 to 12 cycles. Each sample was then indexed using Nextera dual indices (Nextera XT Index Primer 1 (N7xx) and Nextera XT Index Primer 2 (S5xx)). After each PCR, reactions were cleaned up with AMPure XP beads (Beckman Coulter). All samples were multiplexed and sequenced in a single lane on the Next-Seq using 2×150-bp paired-end sequencing. Sequencing and quality control were performed at the University at Buffalo Genomics and Bioinformatics Core.

2.2.12 smFISH

smFISH technique was optimized and adapted from the Stellaris website and based on Shawn Little's protocol (personal communication). smFISH probes were designed by and ordered from Stellaris (Biosearch Technologies). The probes were designed with a C-terminal TEG-Amino tag, and the 3' ends were fluorescently labeled with a succinimidyl ester and HPLC purified as previously described (De Kumar et al., 2017)). The following probe sets were generated: GIP-Af555, PPO1-Af647, PPO2-Af647, Lola-I-Af647.

2.2.12.1 Sample preparation

1. Embryo collection and crosslinking:

   Properly staged embryos were collected and dechorionated with bleach. Embryos were then crosslinked in 4% formaldehyde in a 1:3 mixture of PBS and heptane. Following crosslinking, embryos were sorted in PBT (0.1% Triton). Following sorting embryos were devitellinated with a 1:1 mixture of methanol and heptane. Embryos were then permeabilized in methanol. Embryos can then be stored at $-20\,^{\circ}\mathrm{C}$.

2. Protease treatment: Embryos were rehydrated gradually. Following rehydration, embryos were post-fixed and treated with proteinase k (0.5ug/ml) for one hour in ice and one hour at 37 °C. The proteinase k digested samples were crosslinked once more.

3. Hybridization: The permeabilized embryos were washed with a series of PBT and Stellaris WashA buffers. Prehybridization in the Stellaris Hyb buffer was done overnight at 37 °C. Hybridization was at 25C overnight with 1µ molar concentration of probes.

4. Whole embryo mounting: Embryos were mounted in Prolong gold and curated at room temperature (22 °C) for three days to one week.

2.2.12.2 Image acquisition

3-color samples were generated in whole-mount fixed embryos, with Dapi, GIP-Af55, and a combined far-red probe set with PPO1-Af647 and PPO2-Af647. (The contribution of PPO1 and PPO2 is to find positive cells, and not quantification of transcript intensity, so we merged the PPO1 and PPO2 probe sets). For analyzing the Lola-I expression, Lola-I-Af647 and dapi stained embryos were used.

All data were acquired on a Nikon base outfitted with a CSU-W1 Yokogawa disc. Emission was collected onto a Hamamatsu ORCA-flash 4.0 scMOS. Emission filters were as follows – dapi – 455/50 nm, green – 525/36 nm, red – 605/52 nm, and far-red – 705/72 nm. The main dichroic was a 405/488/561/640 nm tetra-band dichroic, and those correspond to the laser lines used for each channel. For simple determination of GIP and PPO1 positive cells, whole embryo images were acquired with a 40x Apo LWD 1.15 NA water objective, with a z-spacing of 1 micron. Images were blurred in FIJI with a radius of 1 pixel, and a rolling background subtraction with a radius of 250 pixels was applied to all channels with the exception of the dapi channel. Positive cells were easily distinguished (see Figure 3.22A) and were manually counted.

For analyzing the Lola-I expression, RNA fish data was acquired on a PerkinElmer Ultra-view spinning disk confocal micro-scope equipped with an EM-CCD camera (model C9100-13;

Hamamatsu Photonics), using Volocity software (PerkinElmer). An Apochromat 63×, 1.46 NA oil immersion objective was used with a 405/488/561/640 nm multiband dichroic. LoLa-PI probes were fluorescently labeled with Af647 and purified, as discussed above. Dual color images of Dapi and Lola-PI-Af647 were acquired with 405 nm and 640 nm laser lines, respectively, with 415-475 nm emission filter for Dapi and a 660-750 nm emission filter for Af647.

2.3 Data Analysis

2.3.1 Sequence alignment

All sequencing reads were aligned to the Drosophila melanogaster genome (dm6) using Bowtie (v 1.1.2)(Langmead et al. (2009)) , allowing a maximum of two mismatches and including only uniquely aligned reads. The sequenced reads were trimmed to 50 bp before alignment. Aligned reads were then extended to the estimated insert size or the actual size for the paired-end libraries. For the bulk mRNAseq samples, the gene expression values were calculated by performing pseudo-alignment using the Kallisto package (Bray et al. (2016)). For the scRNA-seq samples, alignment and separations of reads from different cells were done using the Cell Ranger pipeline (v 2.1.1) from 10x Genomics.

2.3.2 smFISH analysis

2.3.2.1 Quantification of smFISH data

For modeling purposes, it was necessary to calibrate transcript intensity to convert from integrated raw intensity per cell to number of transcripts per cell. For this, data was acquired on the same Nikon spinning disc, but with a z-spacing of 0.3 microns, and a 100x Plan Apo 1.4 NA oil objective. Data was blurred with a radius of 1 pixel, and a rolling ball background subtraction was applied with a radius of 50 pixels. Data was then z-binned (summing pixels) in the z dimension by 7. In order to distinguish single transcripts and minimize the chance of fitting spots that contain multiple transcripts, cells with low overall expression were found. Within the low expressing cells, single

cytosolic spots were fit to a 2-dimensional Gaussian, and the integrated intensity of the Gaussian fit was used as an approximation for contribution of every single transcript to the integrated intensity of each cell. Cells were outlined manually in FIJI, and the integrated intensity of each cell was summed over a total z-dimension of 6 microns. Autofluorescence in the gut is considerable. For illustration purposes, we determined the ratio of autofluorescence in the red and green channels and subtracted this value from the red images per pixel based on the green pixel intensity. Note this gut area is outside the area of interest – the PPO1 and GIP positive cells, and therefore was for visual purposes only and did not affect quantification. A cell was considered to be in the 'bursting' state if it had a visible nascent transcript, which we set as each cell that possessed a secluded, nuclear spot at least three times brighter than the determined single transcript intensity (see Figure 3.22D). This will slightly underestimate the bursting percentage, as new nascent transcript sites that are not producing high numbers of transcripts are not counted.

2.3.2.2 Application of model

Single-molecule RNA FISH intensities were analyzed according to Raj et al. (2006). Intensity distributions were calibrated according to the intensity of a single transcript, as described above, in order to convert from raw integrated intensity per cell to transcripts per cell. The resulting intensity distributions were fit to simple two state gene expression model involving stochastic activation, inactivation, production, and degradation. As in Raj et al. (2006), we expressed all rates as ratios to the transcript degradation rate, which we assume is constant across conditions measured here. As mentioned in previous work, the ratio of activation to inactivation rates is poorly determined for the model. Nevertheless, our measurements provide clear evidence for the number of cells with active expression because of the visual identification of nascent transcripts in the nucleus. We combined these measurements with a count of total potentially expressing cells but co-labeling with a PPO1 + PPO2 FISH probe. In this way, we can estimate the fraction of bursting cells and, therefore, the ratio of activation to inactivation rates. This ratio was fixed in the analysis. Fitting was performed via Levenberg marquardt non-linear least squares using custom software available

at research.stowers.org/imagejplugins. Errors were determined via monte carlo analysis in which we generated 100 curves from the best fit with errors corresponding to the chi-squared parameter and then fit these curves to generate errors for the fit parameters.

2.3.3 Invitro binding assay

Quality sequence reads were mapped to each specific starting sequence using VSearch (Rognes et al., 2016). After obtaining the amount of the reads from each band, we analyzed the data relative the control 601 sequences in each sample (Eq. 2.1). Where N is one of the 11 nucleosome sequences, 601 indicates the Widom 601 control sequence.

$$Relative\,Supershift = log_2\left(\frac{reads\,supershift_N / reads\,supershift_{601}}{reads\,nucleosome\,band_N / reads\,nucleosome\,band_{601}}\right) \tag{2.1}$$

In this method each specific nucleosome sequence is measured relative to non-specific binding. This approach internally controls technical variability introduced by gel-excision, PCR, NGS-library construction, or NGS sequencing.

2.3.4 Mapping scRNA-seq data to known tissue types

We obtained the gene expression profiles of about 3500 cells in total from two independent biological replicates. The cells from both the replicates were pooled together for all the analyses. The Seurat package (Satija et al. (2015)) was used for normalization, clustering, and visualization of the scRNA-seq data. The cell-gene expression matrix is normalized by the total expression and scaled to by a scale factor of 10000 and log-transformed. The principal component analysis was performed on highly variable genes. The first 20 principal components were used as input for clustering by the Shared Nearest Neighbor method. Seurat was also used to identify the marker genes for each of the clusters. The tissue of origin for the clusters was identified by comparing the scRNA-seq expression patterns with the in situ hybridization profiles from the Berkeley Drosophila

Genome Project (BDGP) similar to the previously published method (Karaiskos et al. (2017)). Briefly, the annotated gene expression profiles for the embryonic stage 13-16 were obtained from BDGP, excluding the ubiquitously expressed genes. The scRNA-seq data was then binarized into ON/OFF, for expression values above/below a threshold. We used the expression value at the 0.9 quantile for each gene, as the threshold above which it is considered ON. The results did not vary significantly for a wide range of cutoffs. The Matthews Correlation Coefficient was calculated based on this binarized version of our data versus the binarized BDGP data. The annotation with the maximum correlation with a cell is used as the annotation for the cell. The annotation of the maximum number of cells in a cluster is considered to the tissue type of the cluster. For ambiguous clusters, we analyzed the occurrence of known tissue markers and manually merged or separated clusters such that they better matched anatomical structures. For example, there were cases when more than one cluster was annotated with the same tissue type, and we could not find meaningful differences between them. In these cases, we merged these clusters. While in other cases, when small subgroups with distinct tissue types were found within a cluster, the clusters were separated into multiple sub-clusters. The *lola*$^{\text{wt}}$ scRNAseq data was aligned to the *lola*$^{\text{ORC4}}$ embryos using the canonical correlation analysis (CCA).

2.3.5 % of cells with expression and coefficient of variation calculations

The percentage of cells with any detectable transcripts for each gene in each tissue was calculated. The tissue with maximum expression was considered as the expressing tissue, and the five least expressing tissues were considered as other tissues. The coefficient of variation was calculated as the ratio of standard deviation of expression divided by the mean expression in the expressing tissue. Only the cells with detectable transcripts were considered for this calculation.

2.3.6 Promoter element enrichment

The presence of known Drosophila promoter elements in each promoter is identified with zero mismatches in a specified window relative to the TSS. For each gene group and each promoter

element, the enrichment was calculated as the fraction of genes in a group with a promoter element over the fraction of all genes with the same promoter element. The statistical significance was calculated with Fisher's exact test after correcting for multiple testing by the Benjamini–Hochberg method.

2.3.7 Pausing index calculations

The pausing index was calculated as the amount of Pol II ChIP-seq signal in the 200bp window downstream of the TSS divided by the Pol II signal in the 200bp-400bp region downstream from the TSS in the gene body.

2.3.8 Data and software availability

All data analysis performed in this paper, including raw data, processed data, software tools, and analysis scripts are available through publicly accessible Amazon Linux virtual machine image (ami-id will be provided with peer-reviewed publication of this work or on request). The analysis code is also available on GitHub @zeitlingerlab. Raw and processed data associated with this manuscript have been deposited in GEO and will be available upon peer-reviewed publication of this work or on request.

Chapter 3

Characterization of Lola-I as a promoter pioneer factor that regulates RNA polymerase II pausing

A large fraction of metazoan promoters have a disposition for RNA polymerase II (Pol II) pausing. Since these promoters show paused Pol II even in the inactive state, the loading of paused Pol II is thought to occur by default in a constitutive fashion. In contrast to this view, we show here that the loading of paused Pol II is regulated over the course of *Drosophila* embryogenesis by the zinc finger transcription factor Lola-I. During the late stages of embryogenesis, ubiquitous expression of Lola-I regulates the loading of paused Pol II to a specific set of promoters. At the Lola-I target promoters, Pol II is recruited ubiquitously throughout the embryo. However, this does not lead to ubiquitous expression but rather primes them for tissue-specific activation. The mechanism by which Lola-I regulates the loading of paused Pol II occurs through the depletion of nucleosomes, similar to the role of pioneer factors at enhancers. Moreover, the priming of target promoters by Lola-I shifts the gene expression equilibrium towards the active state, and might thereby reduce the transcriptional noise and bring synchrony in gene expression between cells. Our results therefore suggest a broader role for pioneer factors in regulating chromatin accessibility in the genome and in preparing genes for activation.

3.1 Introduction

For transcription to occur within chromatin, the DNA at enhancers and promoters has to be accessible to transcription factors and RNA polymerase II. During embryonic development, enhancers are made accessible first by lineage-specific pioneer factors, which recognize and bind to specific DNA sequences even when the binding sites are part of a nucleosome. In the presence of the right combination of cell-type-specific transcription factors, these enhancers then become devoid of nucleosomes during enhancer activation. In contrast, promoters are usually nucleosome-depleted in

a more constitutive fashion (Bernstein et al. (2004), Lee et al. (2004), Yuan et al. (2005)). They either contain AT-rich sequences that make nucleosome occupancy unfavorable (Field et al. (2008), Segal & Widom (2009)) or they contain binding sequences for a constitutively expressed pioneer factors (such as GAGA in *Drosophila*), which keep the promoter nucleosome-free (Tsukiyama et al. (1994), Okada & Hirose (1998), Fuda et al. (2015)). Therefore, the accessibility of promoters is generally not considered a target for regulation during development.

Consistent with promoters being constitutively accessible, a large fraction of promoters show evidence for transcriptional initiation, as for example evident by the high occupancy of paused RNA polymerase II (Pol II) in the genome (Zeitlinger et al. (2007), Muse et al. (2007), Guenther et al. (2007), Core et al. (2008)). If a promoter is accessible and Pol II can initiate transcription, Pol II typically accumulates at the promoter-proximal pause site at around 20-60 bp downstream of the transcriptional start site (TSS). During gene activation, Pol II is increasingly released from pausing into productive elongation, in which transcripts are being produced. Since a large fraction of promoters show high levels of paused Pol II even in the absence of gene activation, it is assumed that Pol II initiation occurs at basal levels by default and that the release from Pol II pausing is the rate-limiting step for productive transcription. Consistent with a permissive state for initiation, the promoters of highly paused genes are usually strongly depleted of nucleosomes (Gaertner et al. (2012), Gilchrist et al. (2010)). However, it is possible that promoters with paused Pol II might be regulated at the level of promoter accessibility. Highly paused promoters in *Drosophila* tend to have GC-rich sequences and are predicted to intrinsically favor nucleosome formation (Tillo et al. (2010), Xi et al. (2010), Gaertner et al. (2012)), suggesting that these promoters could be protected by nucleosomes under some conditions. Indeed, we have previously observed that a certain set of genes in *Drosophila* acquire paused Pol II only at later stages of embryogenesis (Gaertner et al., 2012), and this change in the occupancy of paused Pol II is accompanied by a depletion in promoter nucleosomes. However, how this global change in paused Pol II and nucleosome occupancy might occur is not known.

Here, we show that the sequence-specific zinc-finger transcription factor Lola-I regulates the

loading of paused Pol II at this set of target promoters. Moreover, Lola-I regulates the loading of paused Pol II by functioning as a promoter pioneer factor. Lola-I motifs tend to be found at the edges of nucleosomes *in vivo* and Lola-I preferentially binds this location in *in vitro* nucleosome binding assays. Thus, the expression of Lola-I during the late stages of *Drosphila* embryogenesis leads to pre-loading of paused Pol II at the Lola-I target genes. Moreover, the pre-loading of paused Pol II leads to synchrony in gene expression between cells and reduction in transcriptional noise, but the pre-loading also leads to basal expression throughout the embryo.

3.2 Results

3.2.1 Lola-I is required for the de novo recruitment of paused Pol II in the late embryo

We previously observed that a set of promoters have no Pol II at the early stages (2-4h AED) and then acquire high levels of paused Pol II at the late stages (14-17h AED) of *Drosophila* embryogenesis (Gaertner et al. (2012)). To test whether this set of promoters (referred to as opening set) might be regulated by sequence-specific transcription factors, we analyzed the DNA sequences around these promoters (Figure 3.1A). The most highly enriched motif was AAAGCT.This motif is specifically enriched at the opening set of promoters and not at genes that have constantly high levels of paused Pol II throughout embryogenesis (constant set). Notably, the motif has previously been shown to be bound by Lola-I (Enuameh et al. (2013)) (Figure 3.1B), a splice isoform from the *lola* locus that becomes upregulated during the late stages of *Drosophila* embryogenesis (Casas-Vila et al. (2017)). This makes Lola-I a good candidate for regulating the loading of paused Pol II at the opening set.

We therefore tested whether Lola-I binds to the opening set *in vivo*. There are about 25 different splice isoforms made from the *lola* locus, with the same N-terminal BTB domain but have different C termini, encoding different zinc finger domains that bind to diverse sets of DNA motifs (Goeke et al. (2003), Enuameh et al. (2013)). We generated isoform-specific antibodies against the C-terminus of Lola-I and confirmed Lola-I's expression in late embryos on Westerns blots (Fig-

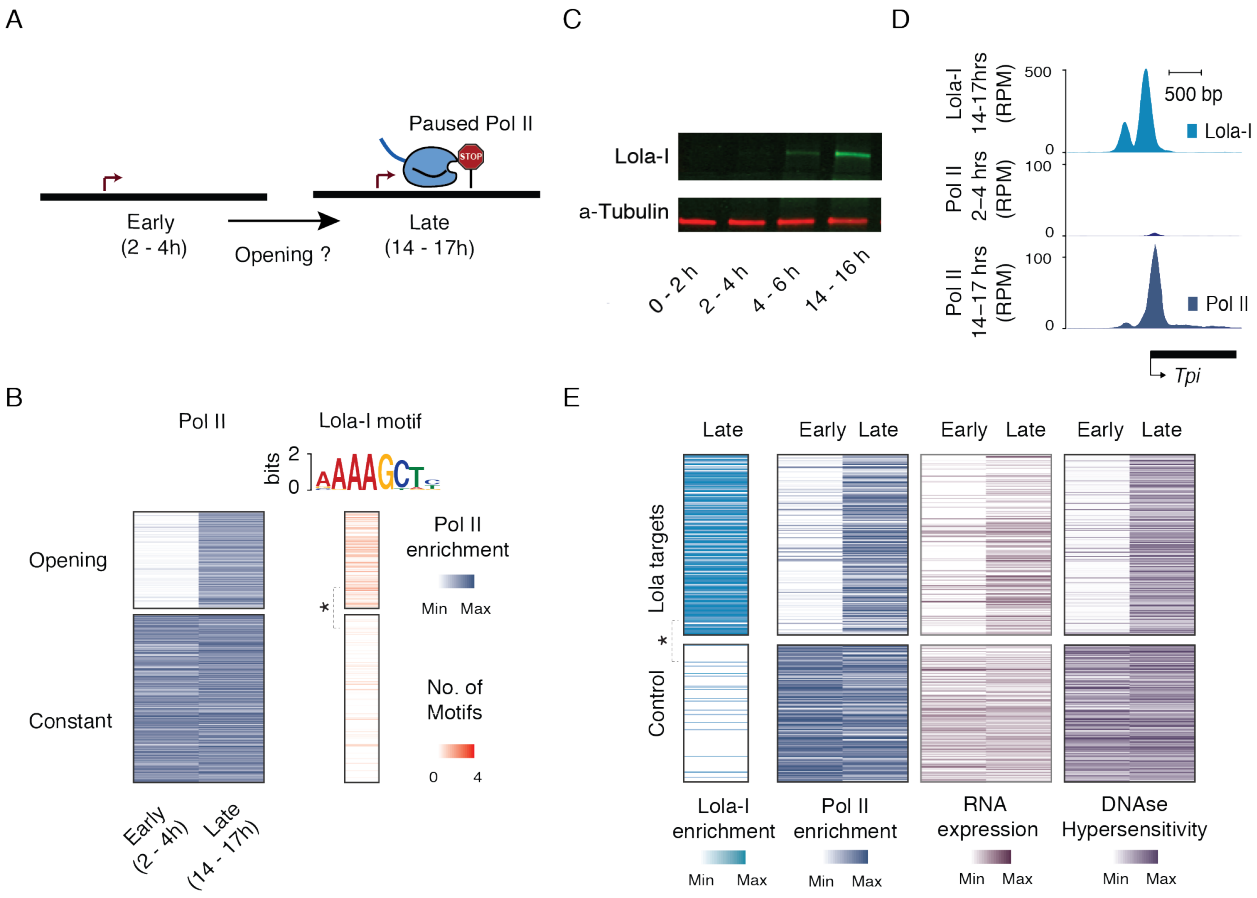

Figure 3.1: Lola-I correlates with the recruitment of paused Pol II in the late embryo: A) Schematic illustration of establishment of paused Pol II at late genes. B) Enrichment of Lola-I motif at the late paused genes. Lola-I motif was identified de-novo at the late paused genes. Differential enrichment analysis shows specific enrichment of the Lola-I motif at the late genes (Wilcoxson two-sided test, *P $<10^{-15}$). C) Lola-I shows increasing expression during *Drosophila* embryonic development, shown by western blot specifically against the Lola-I isoform. D) Single gene example shows Lola-I binding to the promoter of the late paused gene Tpi. E) Heatmap shows that Pol II occupancy, RNA levels, accessibility (measured by DNAse hypersensitivity) correlate with the binding of Lola-I at its target promoters. (Wilcoxson two-sided test, *P $<10^{-15}$)

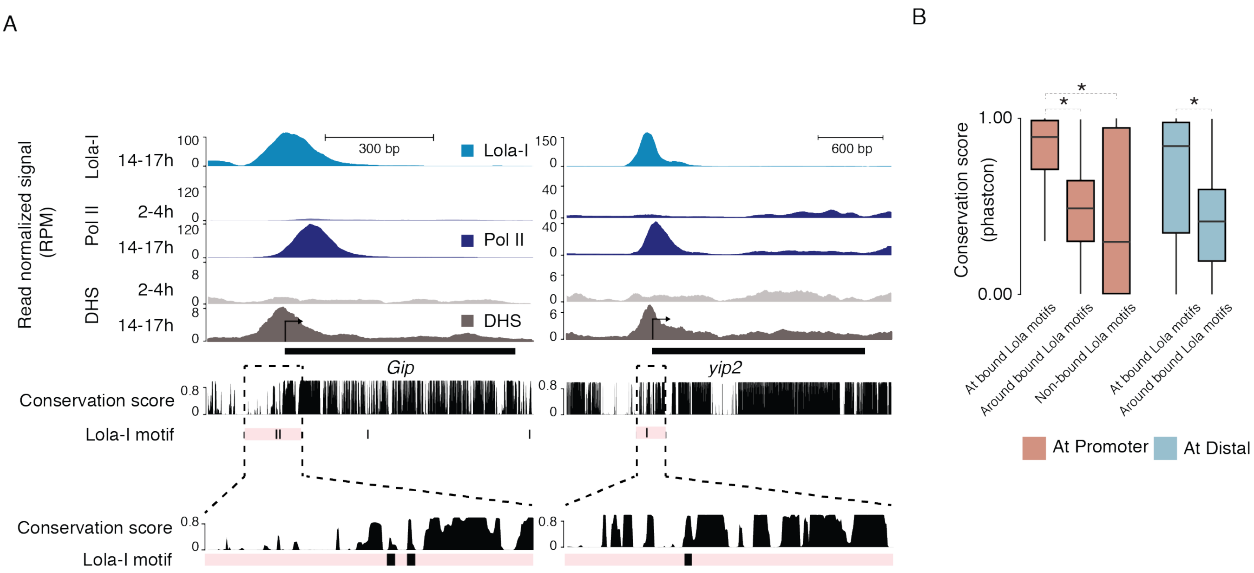

Figure 3.2: Conservation of Lola-I binding sites: A) Single gene examples showing the conservation of the Lola-I binding sites at Lola-I target promoters across different *Drosophila* species using the PhastCons score. Lola-I binding to its target sites correlate with Pol II binding and increased accessibility. B) Boxplot of the PhastCons score shows the conservation of the Lola-I binding sites. Lola-I binding sites at promoters bound by Lola-I are highly conserved compared to promoter region (200 bp around the binding site) in general or the not-bound Lola-I binding sites (Wilcoxson two-sided test, *P $<10^{-15}$). Similarly Lola-I binding sites at non-promoter regions bound by Lola-I are also highly conserved (Wilcoxson two-sided test, *P $<10^{-15}$).

ure 3.1C). We then performed chromatin immunoprecipitation (ChIP-seq) experiments with these antibodies and identified regions bound by Lola-I. These Lola-I-bound regions significantly overlap with the opening set of genes, and the most highly enriched 6-mer they contain is the Lola-I motif (AAAGCT). Furthermore, the Lola-I motifs bound by Lola-I are highly conserved among the different *Drosophila* species, consistent with these binding motifs to be functionally important (Figure 3.2). We therefore conclude that the opening set as previously identified is likely characterized by the presence of Lola-I binding.

We next analyzed how Lola-I target promoters change from the early to late stages of embryogenesis. We specifically focused on Lola-I binding targets and compared the results to a random subset of the constant set. As expected, we found that Lola-I target promoters had dramatically increased levels of paused Pol II compared to the early stages (Figure 3.1D, as an example; Figure 3.1E for all targets). At the same time, these target promoters showed a significant increase

in chromatin accessibility as measured by DNAse-seq (Thomas et al., 2011) and were associated with higher levels of transcript expression (RNA-seq) in the late embryo (Figure 3.1E). Thus, the binding of Lola-I to target promoters is associated with increased promoter chromatin accessibility, Pol II recruitment and gene expression.

To test whether Lola-I is responsible for these promoter changes, we analyzed the previously reported Lola-I mutant (lola^{ORC4}) ((Goeke et al., 2003)). In this mutant, Lola-I peptides, but not other Lola isoforms, are truncated due to a premature stop codon that leaves the protein without the zinc finger DNA-binding domain. Consistent with this, we found that homozygous lola^{ORC4} mutant embryos contained a shorter form of Lola-I that was less abundant than the full-length product in wild-type embryos in Western blots (Figure 3.3A). These lola^{ORC4} mutant embryos had significantly reduced Pol II occupancy and chromatin accessibility measured by ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing) (Buenrostro et al. (2013)) at Lola-I targets, but not at control promoters (Figure 3.3B,C). These defects were not due to a developmental delay or due to secondary mutations in the lola^{ORC4} stock (Figure 3.4). Notably, the defects in Pol II occupancy and accessibility in lola^{ORC4} mutants were rescued by the transgenic expression of Lola-I (Figure 3.3B,C). This shows that Lola-I specifically affects Pol II and chromatin accessibility at the promoters to which it directly binds.

3.2.2 Single-cell RNA-seq reveals tissue-specific genes in the late embryo

While, we have now clearly established the role of Lola-I in loading Pol II at certain late genes, it is not clear if Lola-I establishes paused Pol II or is a general activator. We have previously observed that certain paused genes have Pol II loaded at their promoters, even in tissues where they are not expressed (Gaertner et al. (2012)). We reasoned that if Lola-I is a general activator, binding of Lola-I to its target promoter will correlate with the expression of its target genes both temporally and spatially. However, if Lola-I and Pol II are bound in tissues where the Lola-I targets are not expressed, it would argue that Lola-I is involved in the establishment of paused Pol II and is not a general activator. To this effect we performed scRNAseq (single-cell RNA sequencing)

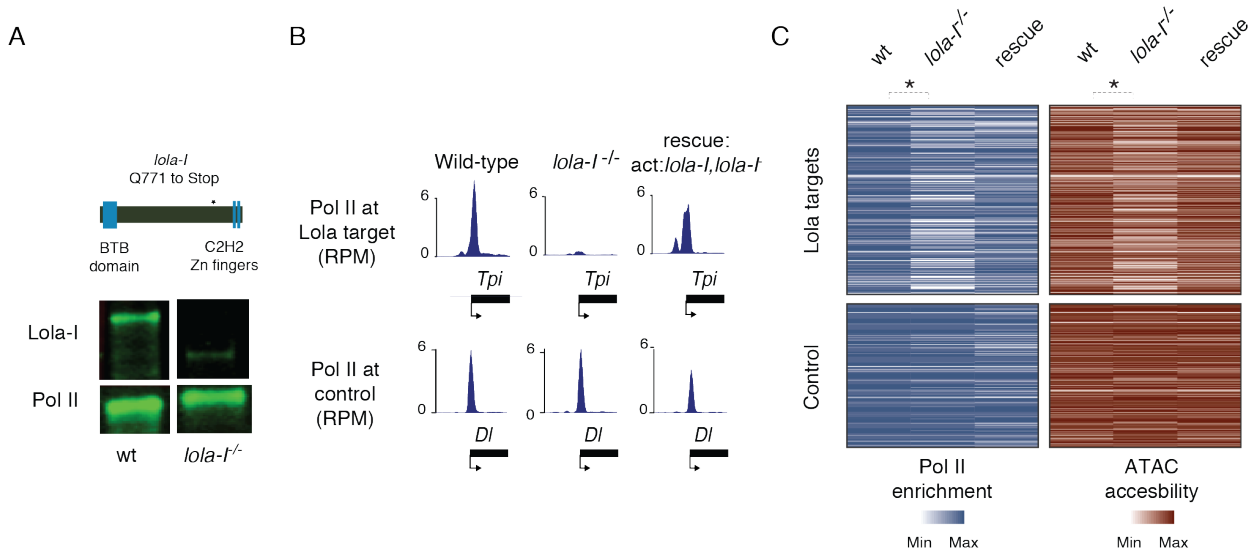

Figure 3.3: Lola-I is required for the recruitment of paused Pol II in the late embryo: A) Lola-I mutant line *lola*textsuperscriptORC4 has a premature stop codon before the DNA binding domain. Western blot analysis shows mostly degraded Lola-I protein in the $lola^{ORC4}$ mutant. B) Pol II occupancy at the Lola-I target Tpi goes down in the $lola^{ORC4}$ mutant embryos. While in the rescue line which expresses *lola-I* cDNA in the $lola^{ORC4}$ background, Pol II occupancy is rescued to the wildtype levels. In the control non-Lola-I target gene, Dl, Pol II occupancy remains the same between all the samples. C) Heatmap shows that globally Pol II occupancy (Wilcoxson two-sided test, $*P < 10^{-9}$), accessibility (measured by ATAC-seq) (Wilcoxson two-sided test, $*P < 10^{-14}$) goes down specifically at the Lola-I target promoters in the $lola^{ORC4}$ mutant embryos and is rescued by the expression of *lola-I*.

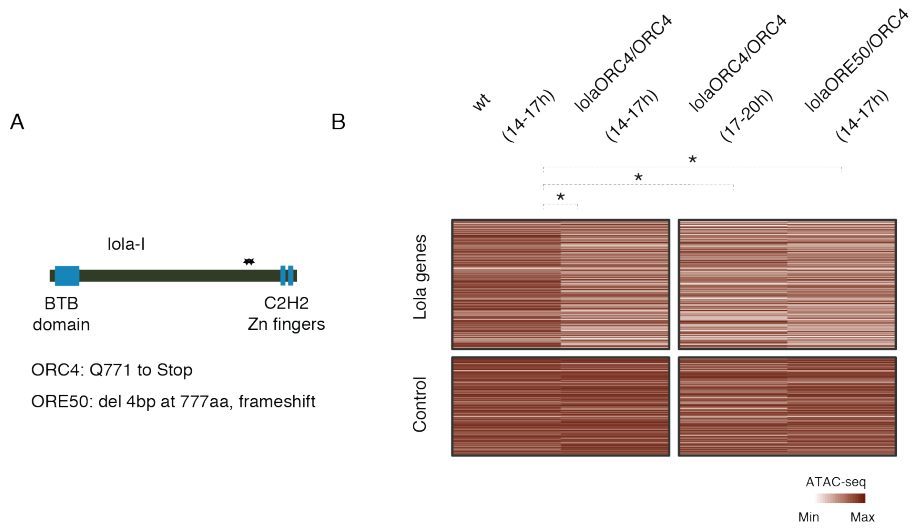

Figure 3.4: Effect of loss of *lola-I* on accessibility. A) Illustration of different *lola-I* mutants. *lola*^{ORC4} mutant has a premature stop codon. *lola*^{ORE50} mutant has a frameshift mutation. B) Heatmap showing accessibility changes at the Lola-I targets at different conditions. Lola-I targets show reduced accessibility in the *lola*^{ORC4} mutant embryos at 14-17h (Wilcoxson two-sided test, *P $<10^{-9}$) andat 17-20h (Wilcoxson two-sided test, *P $<10^{-10}$)and in the transhetreozygous *lola*$^{ORE50/ORC4}$ mutant embryos (Wilcoxson two-sided test, *P $<10^{-13}$)

and tissue-specific Pol II and Lola-I ChIP-seq experiments.

To obtain an unbiased global view of the gene expression programs in differentiated tissues, we performed single-cell RNA-seq on late *Drosophila* embryos (Figure 3.5A). We chose stage 16 embryos (14h-14.5h) since the tissues are fully formed, while the cuticle is not yet developed enough to hamper the dissociation of the cells. We obtained the expression profiles of approximately 3,500 cells after processing the cells through a 10x Genomics Chromium instrument. Cells were prepared and sequenced in two separate batches, the results of which were overall highly similar with regard to data quality and tissue clusters (Figure 3.6). We clustered the single-cell RNA-seq data and then labelled and modified the clusters based on large-scale in situ hybridizations from the Berkeley *Drosophila* Genome Project (BDGP) and existing knowledge of the *Drosophila* tissues at this stage (Figure 3.5A). For ambiguous clusters, we analyzed the occurrence of known tissue markers and manually merged or separated clusters such that they better matched anatomical structures. In this manner, we obtained scRNA-seq data for 16 tissues of the late *Drosophila* embryo: central nervous system (CNS), peripheral nervous system (PNS), glia, germ cells, epidermis, tra-

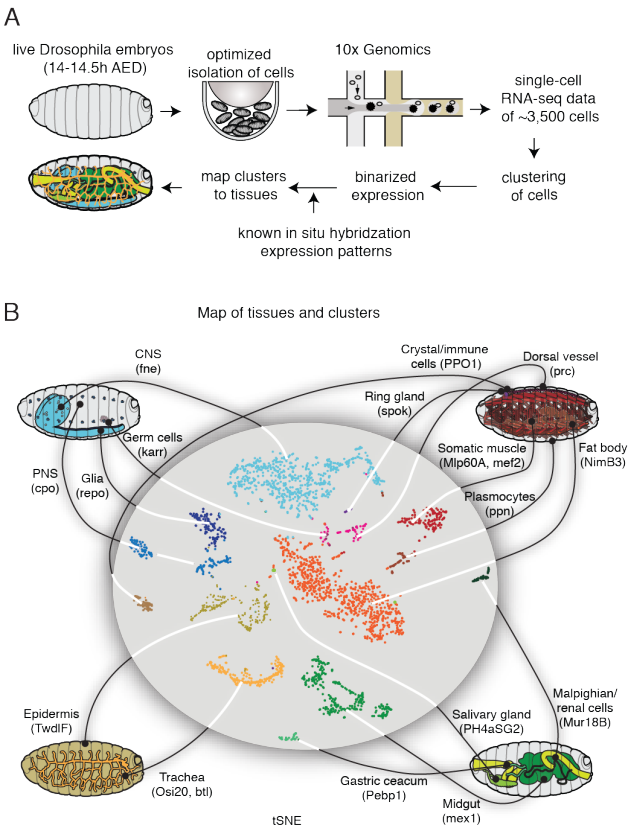

Figure 3.5: scRNA-seq captures the expression profiles of effectors genes in the late stages of *Drosophila* embryogenesis. A) Single cells were isolated from 14-14.5h *Drosophila* embryos (After Egg Deposition - AED). Isolated cells were processed through 10x Genomics instrument and the gene expression profiles were obtained after sequencing. The obtained gene expression profiles were used to map the isolated cells to known cell types by comparing against the available in situ hybridization profiles from the Berkeley *Drosophila* Genome Project. B) tSNE projection of the gene expression data from the isolated cells shows various tissue types captured in the scRNA-seq experiment. Marker genes for each tissue type is also shown.

chea, muscle, dorsal vessel, fatbody, plasmocytes, crystal cells, ring gland, salivary gland, gastric caecum, midgut and malpighian tubules. (Figure 3.5B). For each tissue, we then identified marker genes that reliably represented the cluster, some of which were previously known (Figure 3.5B and 3.7, 3.8). This gave us the tissue specific expression profile of the Lola-I targets. To test if Lola-I acts as a general activator or is involved in the establishment of paused Pol II, we also needed the tissue-specific binding profiles of Lola-I and Pol II.

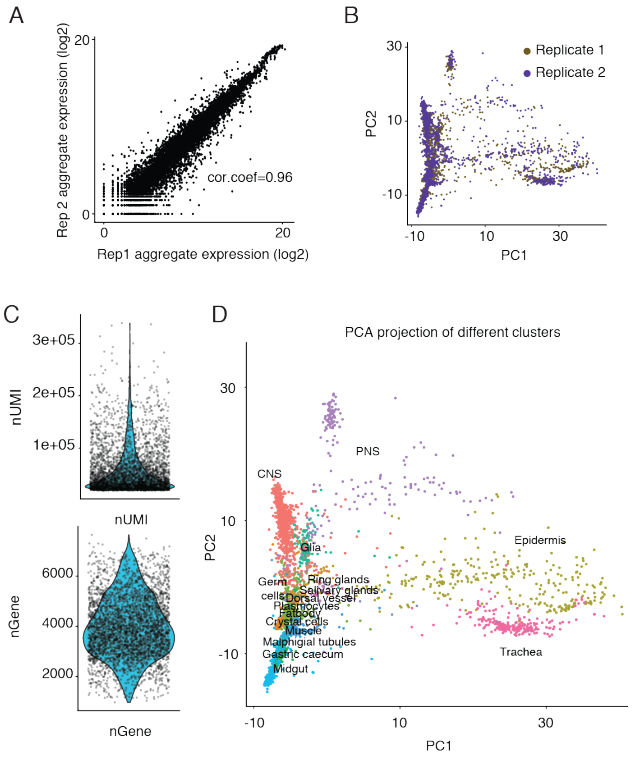

Figure 3.6: scRNA-seq experiments provide reproducible and high-quality gene expression profiles: A) Correlations of gene expression levels between single-cell replicates were calculated by computing the sum of read counts for each gene across all cells in both replicates, which shows very high similarity between the replicates. Pearson correlation was measured between the replicates. B) PCA projection of the scRNA-seq data from both the replicates show the consistency of the two replicates. C) The number of Unique Molecular identifiers (UMI) and the number of genes captured per cell are shown. On average, we captured about 4000 genes per cell. D) PCA projection of the scRNA-seq data shows tissues of similar origin clustering together, which indicates that the main sources of variations in the scRNA-seq data is biologically significant.

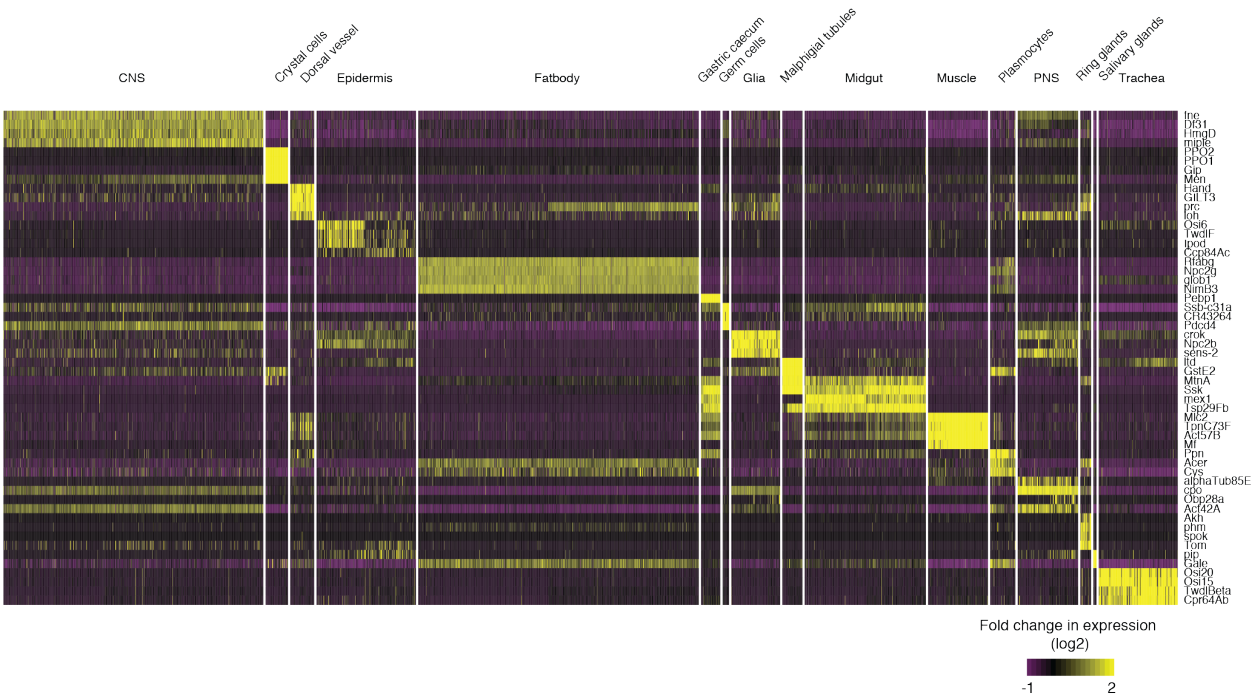

Figure 3.7: Expression profiles of marker genes of the clusters identified in the scRNA-seq data. Heat map of de-novo identified marker genes. Top four differentially expressed genes in each tissue (Wilcoxon rank sum test, expression in at least 5% of cells in at least one group) is shown.

3.2.3 Late genes differ in their Pol II occupancy pattern across tissues

To obtain the tissue-specific binding profile of Lola-I and Pol II, we isolated biotin-tagged nuclei from the tissue of interest with the help of streptavidin-coupled magnetic beads (Deal & Henikoff, 2011) from dissociated fixed embryos (14-17h) and used the chromatin as input ChIP-seq (Figure 3.9A). We tagged six different tissues: neurons (using elav-Gal4), glial cells (using repo-Gal4), muscle (using mef2-Gal4), trachea (using btl-gal4) and epidermis and gut (using enhancer trap-gal4 lines 7021 and 110394, respectively, see Methods) (Figure 3.9B).

Pol II ChIP-seq tracks from the six tissues confirm that the data are tissue-specific since each profile showed unique high Pol II occupancy at genes only expressed in that tissue (Figure 3.9B middle panel). For example, the tracheal gene *Osi20* and the muscle gene *Mlp60A* showed high Pol II occupancy in the trachea and muscle samples respectively, but not in the other tissues (Figure 3.9B middle panel). Furthermore, we compared the Pol II occupancy data with the expression programs obtained from the single-cell RNA-seq data, if we had a corresponding expression clus-

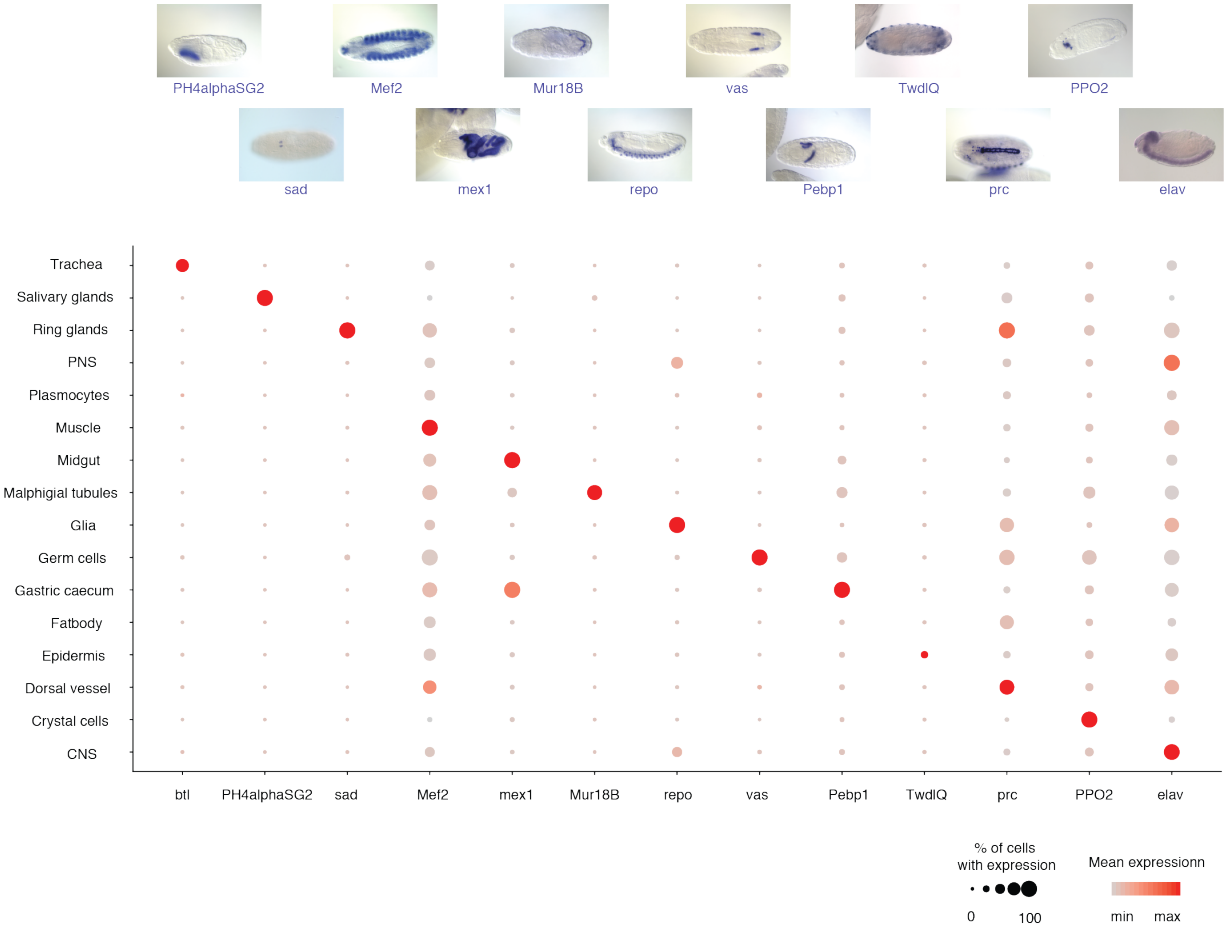

Figure 3.8: Comparison of scRNAseq data and the *in situ* hybridization data. Many of the known marker genes for the clusters based on previous studies and *in situ* hybridization from BDGP, have tissue specific expression in the scRNA-seq data, as shown in the Dot plot. The size of the dot represents the frequency of cells in a tissue with expression and the color intensity represents the mean expression in the tissue.

ter, and found an overall correlation between Pol II levels and gene expression (Figure 3.10). This led us to conclude that the Pol II profiles are indeed from the expected tissue and that obvious cross-contaminations were not observed.

In contrast to the tissue-specific genes with tissue-specific Pol II occupancy described above, we observed that other tissue-specific genes did not follow this pattern of Pol II occupancy. Some genes showed high Pol II occupancy in many or all tissues, despite being expressed in a very tissue-restricted fashion. For example, the Ace gene showed very high Pol II promoter occupancy in all tissues, although its expression is restricted to neuronal populations (Figure 3.9B right panel). This discrepancy can easily be explained by Pol II being in a paused state at these genes. Indeed, Pol II occupancy at the *Ace* gene peaks at the pausing position (30-50 bp downstream of the TSS) and is not detectable at substantial levels at the gene body (Figure 3.9B right panel). Thus, while the promoter has paused Pol II in all tissues, full-length transcripts are only made in specific tissues. This profile of Pol II pausing is not observed at the *Osi20* and *Mlp60A* genes, for which Pol II occupancy was tissue-specific. Rather than being highest at the pausing position, high Pol II occupancy is found throughout the body of the gene (Figure 3.9B middle panel).

This observation suggests that there are two types of tissue-specific promoters that are regulated in a fundamentally different fashion. At one type of promoter, Pol II is widely recruited and found paused across tissues, but productive Pol II elongation occurs in a tissue-specific manner. At the other promoter type, Pol II is recruited only in the tissue where it is expressed, and productive elongation may occur without detectable pausing.

3.2.4 The Pol II penetrance pattern across tissues separates paused and TATA genes

Since we observed such different Pol II behavior on different promoters, we wondered whether such behavior could be explained by differences in core promoter elements, such as for example by the presence of the TATA box or pausing elements. We therefore set out to systematically classify promoters based on their Pol II occupancy patterns. Pol II pausing is often measured with the pausing index (ratio of paused Pol II at promoter/elongating Pol II in the gene body), but this

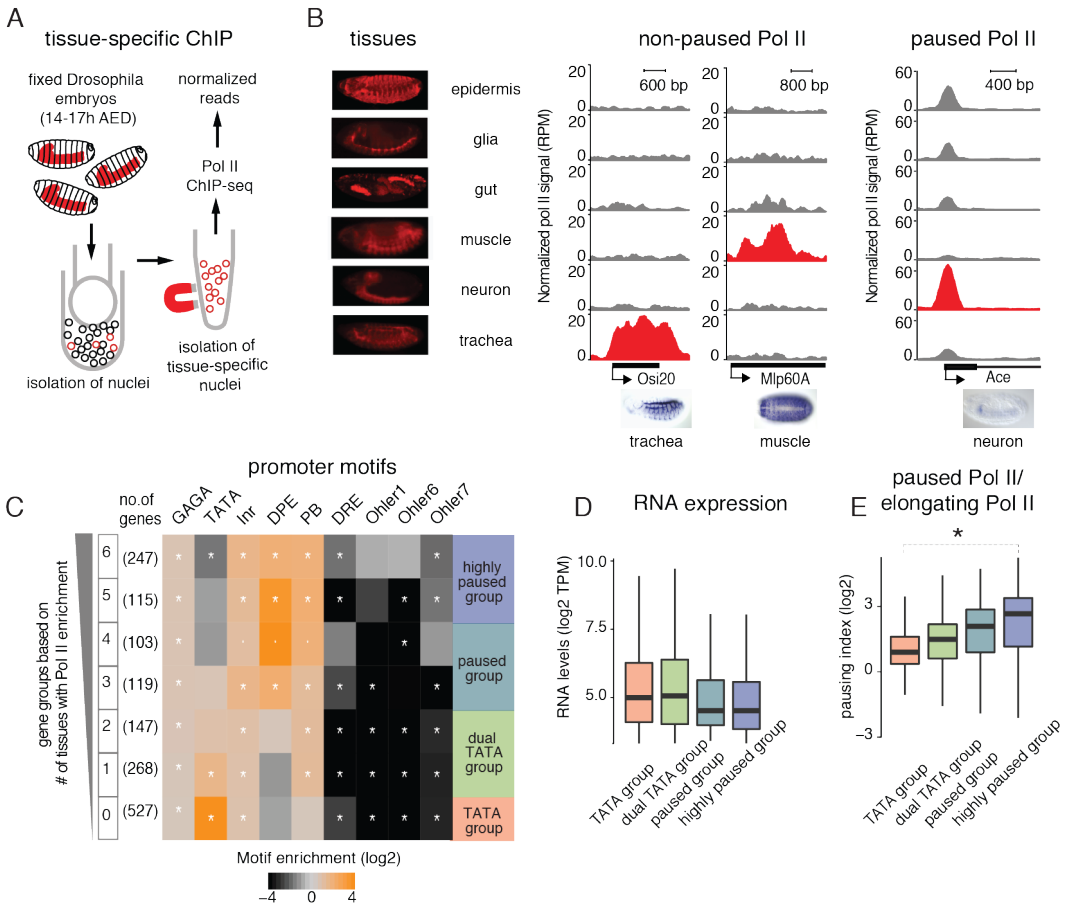

Figure 3.9: Tissue specific Pol II ChIP-seq shows differences in Pol II occupancy patterns at effector genes. A) Tissue specific ChIP-seq was done by isolating nuclei from specific tissues (shown in red) by expressing the Escherichia coli biotin ligase (BirA) and the biotin ligase recognition peptide (BLRP) fused with a nuclear envelope-targeting sequence in the tissue of interest. This allows the isolation of nuclei from the tissue of interest using streptavidin magnetic beads. B) Pol II ChIP-seq was done in six different tissues shown in the left panel. The middle and the right panel shows the read count normalized Pol II ChIP-seq tracks (RPM) from the six tissues at individual genes. Grey and red tracks indicate non-expressing tissues and expressing tissue respectively for a particular gene. Middle panel shows the Pol II profile at non-paused genes, which have Pol II only in the expressing tissues. Right panel shows the Pol II profile at a paused gene, which has Pol II in all the observed tissues. C) Identified effector genes were grouped in to seven groups based on the number of tissues in which they have above background Pol II enrichment. Pol II enrichment was calculated in a window starting from the Transcription Start Site (TSS) and ending 200bp downstream of the TSS. Distinct core promoter elements are differentially enriched in different groups (Fisher's exact test with multiple-testing correction, *P < 0.05). Only the groups with Pol II enrichment in 0 or 1 tissues are enriched for the TATA motif. D) RNA levels (log2 TPM) from 14-17 hrs embryos at the different effector gene groups are shown. The TATA genes are expressed at levels comparable to the paused genes. E) Pausing indices (log2), ratio of Pol II at the promoter vs the gene body, are shown for the different effector gene groups. The pausing indices of genes from the TATA group are significantly lower than that of the highly paused group (Wilcoxson two-sided test, *P $<10^{-15}$).

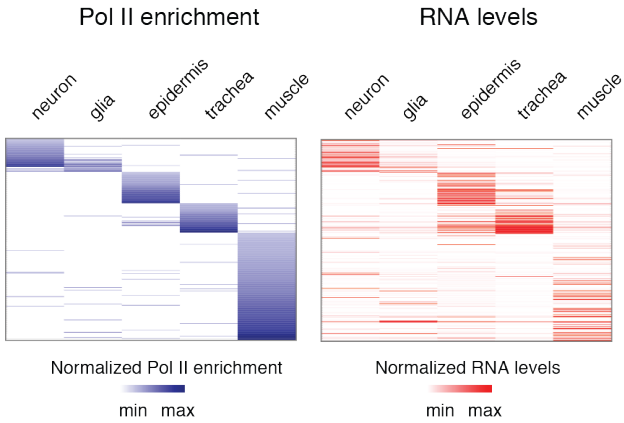

Figure 3.10: Tissue specific Pol II ChIP-seq shows specificity and correlates with the scRNA-seq expression profile: The specificity of the tissue specific Pol II ChIP-seq was evaluated by analyzing the correlation between the Pol II occupancy pattern for a gene and its corresponding expression profile in the scRNA-seq data. The analysis was restricted to genes with Pol II occupancy in a single tissue. The Pol II ChIP-seq signal was calculated around the promoter in a window starting from the TSS and ending 200bp downstream. There is a clear correspondence between the Pol II occupancy and the scRNA-seq gene expression profile.

measurement is not very reliable since it is highly sensitive to noise, dependent on transcription rates and cannot easily be applied to short genes. We were therefore intrigued by the idea that the pattern of Pol II across tissues could be used for classifying promoters.

To get a comprehensive view of Pol II occupancy profile in the late embryo, we first obtained the the general expression profile of all late genes by eliminating the developmental genes and ubiquitously expressed housekeeping genes. We selected all genes expressed in the late embryo and subtracted those genes that are expressed or have Pol II occupancy in the early embryo (2-4h). By subtracting genes with early Pol II occupancy, we also eliminated developmental genes that are not yet expressed in the very early embryo but have paused Pol II ready for activation during subsequent developmental stages (Gaertner et al., 2012). Many of these late genes, could play a role in tissue-specific functions and morphology in these differentiated tissues, and therefore many of them could be part of the previously categorized effector genes.

GO analysis found enrichment for effector function, such as synaptic transmission and chitin-based cuticle development (Figure 3.11A), which are tissue-specific biological functions. Promoter elements enrichment analysis revealed that sequence motifs typically found in paused pro-

moters such as DPE and PB was indeed significantly enriched in the promoter sequences, while sequence motifs found in housekeeping genes such as the Dref Response Element (DRE) were under-represented, as expected (Figure 3.11C). However, in addition paused promoters elements, TATA motifs were also significantly enriched among late effector genes, raising the possibility that multiple promoter types are used to induce genes in the late embryo (Figure 3.11C).

Based on the examples we analyzed, we hypothesized that the higher the disposition of a promoter for Pol II pausing, the more widespread the promoter is occupied across different tissues. Therefore, we grouped the effector genes based on their Pol II penetrance across tissues, i.e. the number of tissues (from 0 to 6 tissues) in which Pol II was detected around the transcription start site above background. We then analyzed the enrichment of core promoter elements in all seven groups (Figure 3.9C).

The results show that Pol II penetrance across tissues indeed correlates with differences in core promoter elements, with TATA genes and highly paused genes being at the two extreme ends. Promoters where Pol II was found across many tissues were under-enriched for TATA and were highly enriched for pausing elements such as the DPE and PB, consistent with them being highly paused. On the other hand, promoters with Pol II occupancy in very few tissues were highly enriched for TATA (Figure 3.9C). When Pol II was found in 1 or 2 tissues, the pausing element PB was also still significantly enriched, indicating that they may be dual promoters. Strikingly, the most canonical TATA promoters, which contain TATA in the absence of pausing elements, were found in the group with no apparent Pol II occupancy and yet the evidence suggests that these genes are transcribed (Figure 3.9D). Pol II may be hard to detect at these canonical TATA genes, presumably because Pol II does not pause, and thus significant levels of Pol II can only be detected with high levels of transcription. Although the penetrance of Pol II occupancy across tissues appeared to be a good measurement for Pol II pausing, we considered that it may be confounded by other factors such as expression levels. We might observe Pol II in fewer tissues simply because the expression levels are lower and thus these genes pass the threshold of Pol II detection in a smaller number of tissues. However, there was no positive correlation between Pol II penetrance and whole-embryo transcript

levels. Although TATA gene expression was initially lower in whole embryos at 14-14.5 h, their expression rose over time until their expression was overall higher than that of highly paused genes (Figure 3.9D and 3.17). Thus, genes with Pol II occupancy across many tissues were not expressed at higher levels than those with Pol II in only few tissues, consistent with paused Pol II being often associated with lower expression levels (Shao and Zeitlinger 2017). We also tested other measurements for Pol II pausing, including the pausing index, and found them to be consistent with the Pol II penetrance measurements (Figure 3.9E). This supports the idea that the penetrance of Pol II occupancy across tissues correlates with the degree of Pol II pausing and that TATA genes are those where Pol II is found in the fewest tissues and with the least amount of pausing.

We next asked whether these promoter groups differed in their function. While GO analysis suggested that there were some differences, all groups were enriched for expected effector gene functions (Figure 3.11B). This led us to conclude that all groups represent genes of similar function, but that they differ in the promoter type and mechanism by which they were expressed in a tissue-specific fashion. Thus, while we are interested in the expression profile of the Lola-I, we also decided to analyze the differences between paused and TATA groups among the late genes.

3.2.5 Lola-I establishes paused Pol II throughout the embryo

Pol II occupancy across different tissue types shows the presence two classes of late genes in general and the strong paused promoters have Pol II in all or most of the tissues that we analyzed independent of gene expression. If Lola-I binds to its targets and recruits paused Pol II in multiple tissues independent of expression, it would suggest that Lola-I is regulating the establishment of paused Pol II at its targets and is not a general activator. In order to differentiate between these roles of Lola-I, we first analyzed the expression of Lola-I. We found Lola-I to be expressed throughout the embryo using immunostaining and smFISH (Figure 3.12A and 3.13). We then analyzed the binding of Lola-I and Pol II in different tissues from the late stage embryo. We observed that Lola-I binds to its target promoters in all the observed tissues independent of gene expression (Figure 3.12B and C). This supports the idea that Lola-I might establish paused Pol II at these

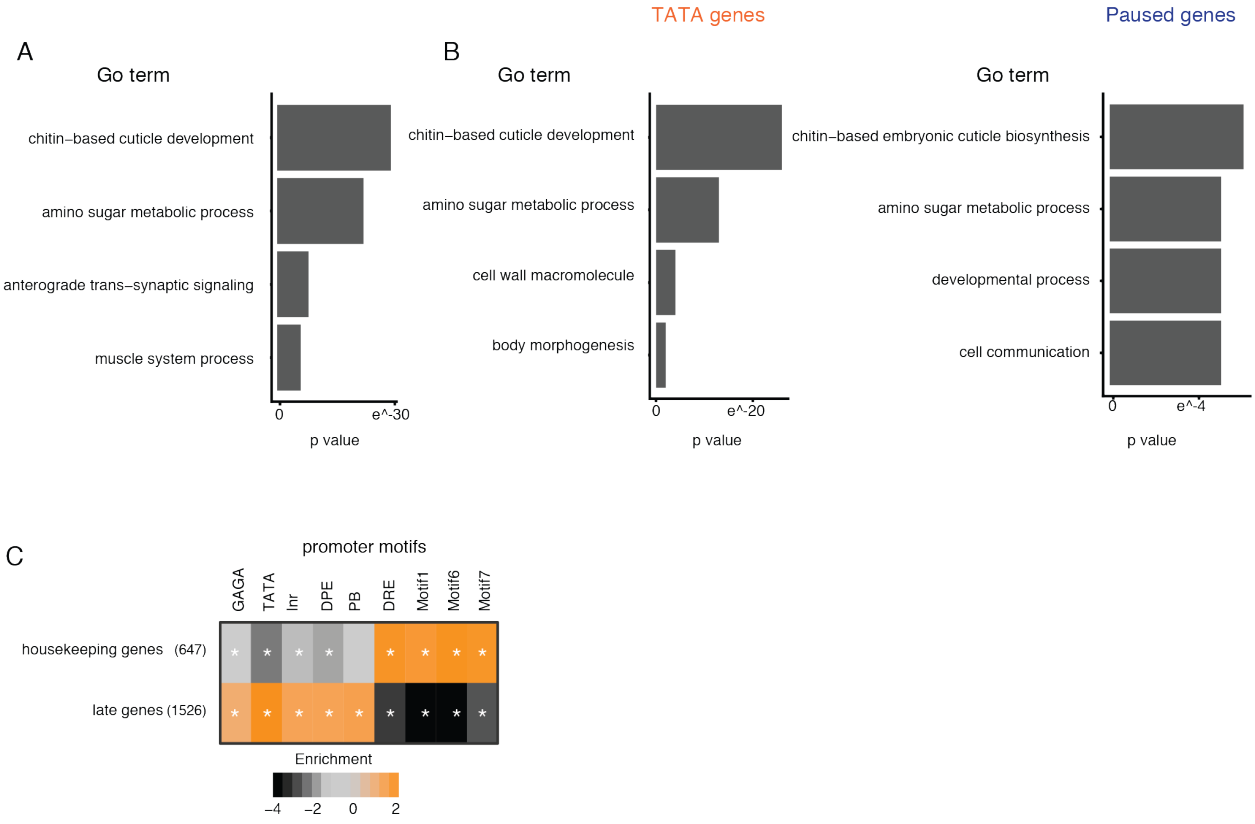

Figure 3.11: Functional categories and promoter elements of effector genes: (A & B) GO term enrichments were calculated for Biological Processes. For each gene group and each GO term, the enrichment was calculated as the fraction of genes in a group associated with a GO term element over the fraction of all genes with the same GO term. The statistical significance for the enriched GO terms were calculated using the Hypergeometric test. GO terms associated with less than five genes were not included in the analysis. Top representative terms are shown in the bar plot. A) Identified effector genes are enriched for expected GO terms. B) GO term enrichments for the TATA group and the Paused group effector genes were calculated separately. While there are some differences between the groups, both the gene groups have expected effector like functional categories. C) Promoter elements enriched in the identified effector genes are shown. Effector genes are enriched for the focused promoter elements and are depleted for motif generally associated with the housekeeping genes.

sites throughout the embryo. To further test this idea, we analyzed the changes in accessibility and Pol II binding at these sites in *lola*^{ORC4} embryos in different tissues, by recombining the *lola*^{ORC4} fly lines with the tissue specific INTACT lines. We observed that the accessibility and the Pol II recruitment is decreased in the absence of Lola at the Lola sites in both the tracheal and gut tissues that we analyzed (Figure 3.12D,E,F). This conclusively suggests that Lola-I establishes paused Pol II throughout the embryo at target promoters during the late stages of embryogenesis and is not a general activator and thus, Lola-I prepares genes for activation by pre-loading them with paused Pol II. Moreover, our data also suggest that Lola-I is a stage specific factor, since it is expressed only in the late stages, and not a tissue-specific factor, since it leads to ubiquitous Pol II recruitment.

3.2.6 Lola-I binding leads to nucleosome depletion

The DNA sequences at the paused promoters intrinsically favor strong nucleosome occupancy (Tillo et al. (2010), Xi et al. (2010)). A strong promoter nucleosome has been observed at the paused promoters but only in the absence of paused Pol II (Gaertner et al. (2012), Gilchrist et al. (2010)). We hypothesized that the Pol II occupancy at the paused promoters might be regulated by the removal of the promoter nucleosome. To understand if Lola-I regulates the promoter nucleosome occupancy at the paused promoter, we analyzed the nucleosome occupancy changes at late genes and specifically at the Lola-I target promoters using MNAse-seq (Micrococcal Nuclease digestion followed by sequencing). We found that the binding of Lola-I and Pol II recruitment correlates with decreased promoter nucleosome occupancy (Figure 3.14A and 3.14B). If Lola-I binding is required for the decreased promoters nucleosome occupancy at these promoters, loss of Lola-I should lead to increased nucleosome occupancy. Our analysis from the *lola*^{ORC4} mutants revealed that the nucleosome occupancy in the *lola*^{ORC4} mutants at the target promoters is increased compared to the *lola*^{wt} embryos (Figure 3.14E). These results suggest that Lola-I could regulate the nucleosome occupancy at paused promoters. However, paused Pol II itself has been shown to be important for the maintenance of nucleosome free regions at promoters (Gilchrist et al., 2010).

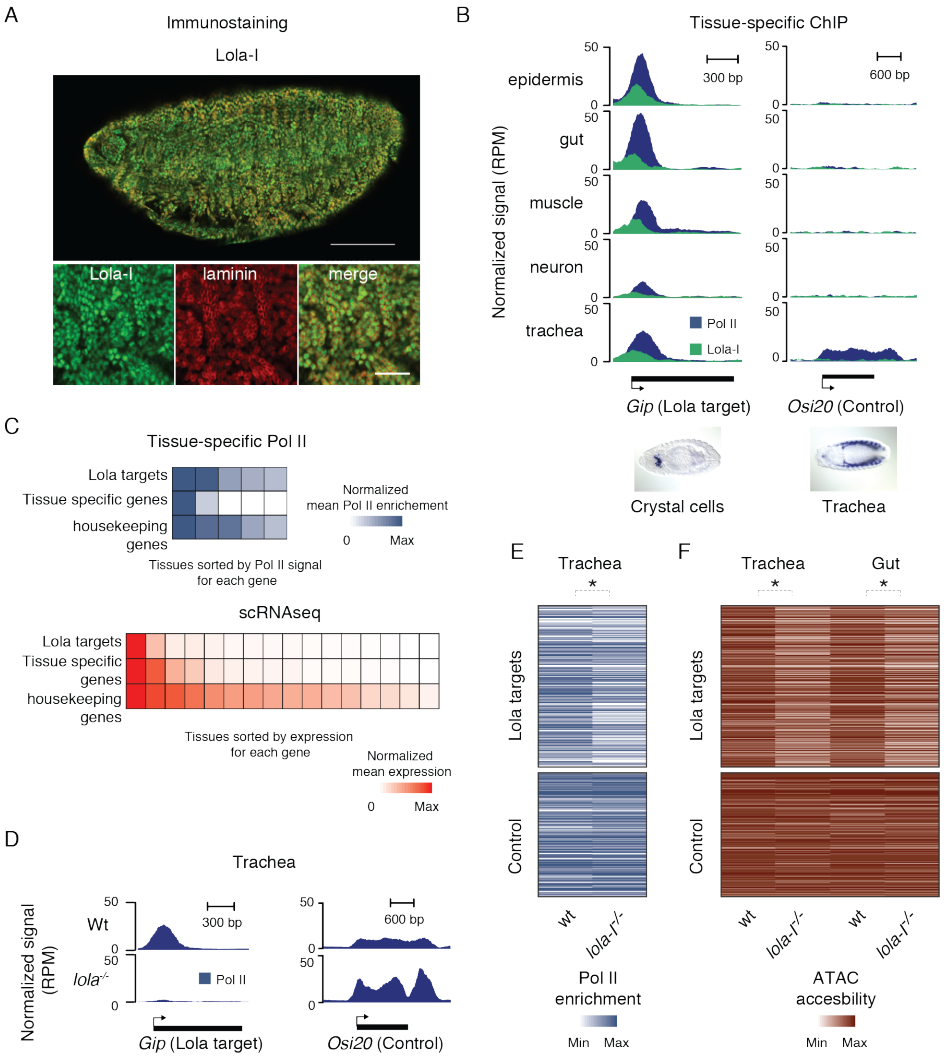

Figure 3.12: Lola-I establishes paused Pol II throughout the embryo. A) Immunostaining using the Lola-I antibody shows ubiquitous the expression of Lola-I. B) Tissue-specific ChIP-seq data for Lola-I and Pol II is shown. The read count normalized Pol II ChIP-seq and Lola-I tracks (RPM) from the five tissues at individual genes (*Gip* - Lola-I target which is expressed in crystal cells and Osi-20 - non Lola-I target which is expressed in Trachea) are shown. Single gene example shows specific binding of Lola-I and Pol II at the Lola-I target promoter *Gip*, in all the five tissues. C) Average Pol II occupancy and the expression profiles of different groups of genes are shown. Lola-I targets have broad Pol II occupancy like the housekeeping genes, but are expressed in a restricted fashion like the tissue-specific genes. Pol II and expression values for each gene is sorted from high to low, normalized to the highest value. Normalized mean profile for each group is obtained by averaging the sorted normalized profiles of all genes in a group. D) Single gene example shows specific loss of Pol II binding at the Lola-I target promoter *Gip*, in the tracheal tissue, in the *lola*^{ORC4} mutant. E) Heatmap shows specific loss of Pol II binding at the Lola-I target promoters in the tracheal tissue, in the *lola*^{ORC4} mutant (Wilcoxson two-sided test, *P $<10^{-4}$). F) Heatmap shows specific loss of ATAC accessibility at the Lola-I target promoters in the gut and tracheal tissues, in the *lola*^{ORC4} mutant (Wilcoxson two-sided test, *P $<10^{-12}$).

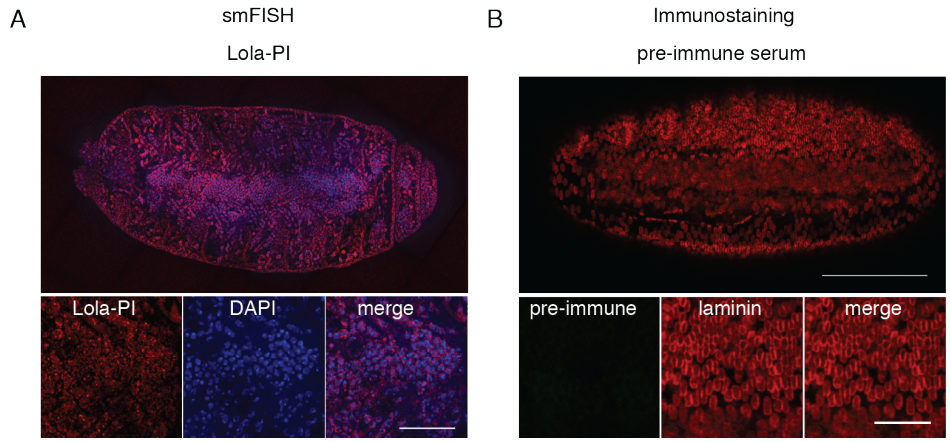

Figure 3.13: Lola-I is expressed ubiquitously. A) single-molecule FISH shows ubiquitous expression of *lola-I*. B) Control Immunostaining with pre-immune serum shows no immunostainig for Lola-I

Therefore, it is possible that the increases in promoter nucleosome occupancy in the Lola-PI mutant is an indirect effect due to the reduction in paused polymerase. To test whether Lola-I can directly establish nucleosome-free state at nearby regions, we analyzed the nucleosome occupancy changes at non-TSS distal regions bound by Lola-I which lack Pol II. While the majority of the Lola-I binding sites are at the TSS, a significant proportion of the binding sites lie at non-TSS distal regions (around 40%). We found that the Lola-I motifs at these sites were also conserved (Figure 3.2B). We compared the nucleosome occupancy changes at these distal Lola-I target sites to the promoter binding sites. The nucleosome occupancy is dramatically increased in the absence of Lola-I both in the early embryos and in the Lola-I mutant embryos at the distal sites similar to what we observed at the promoter sites (Figure 3.14C). In addition, we also analyzed the changes in accessibility at these sites using ATAC-seq. The accessibility of these sites is also decreased similar to the Lola-I promoters. These results indicate that while paused Pol II might contribute the depletion of the promoter nucleosome, Lola-I can establish nucleosome-free regions, at least partially, independent of paused Pol II.

We then asked whether Lola-I binding at distal site also might regulate Pol II recruitment at a nearby promoter by perhaps a non-local mechanism. While the binding of Lola-I at both the TSS and non-TSS regions correlate with decreased nucleosome occupancy at the local binding site

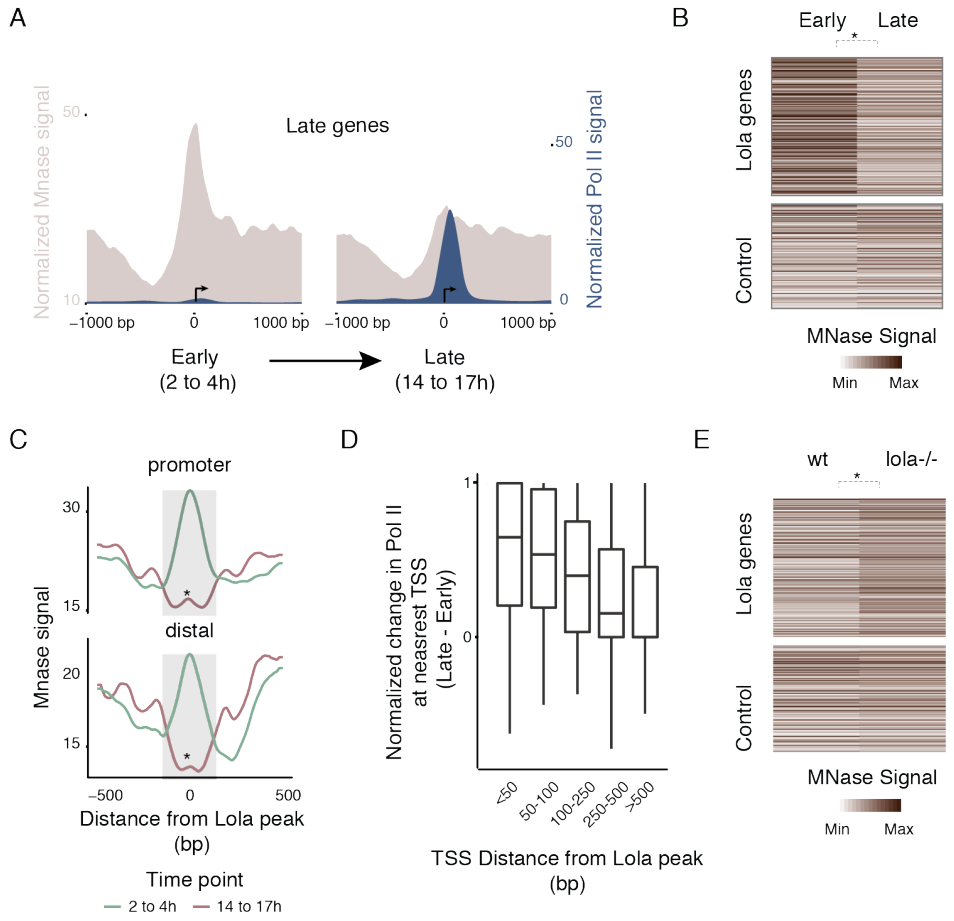

Figure 3.14: Nucleosome occupancy changes at Lola-I targets. A) Metagene plots show anti-correlation between nucleosome occupancy measured by MNAse-seq and the Pol II occupancy at the late genes. B) Heatmap shows specific loss of nucleosome occupancy at the Lola-I target promoters during the late stages of *Drosophila* embryogenesis (Wilcoxson two-sided test, *P $<10^{-8}$). C) Metagene plots show similar loss of nucloeome occpuancy at both the promoter proximal and distal Lola-I binding sites. D) Changes in the normalized Pol II binding between the early and late stages at the Lola-I target promoters are measured and are binned based on the distance between the Lola-I binding site and the nearby promoter were the Pol II changes are measured. Boxplots show that Pol II binding occupancy changes only when the Lola-I binding is nearby. E) Heatmap shows increaed nucleosome occupancy at the Lola-I target promoters upon the loss of Lola-I binding in the *lola*^{ORC4} mutants (Wilcoxson two-sided test, *P $<10^{-7}$).

(Figure 3.14C), only when Lola-I was bound near the TSS it leads to Pol II recruitment (Figure 3.14D). Therefore, we concluded that Lola regulates Pol II recruitment by depleting nucleosomes locally. Further analysis revealed that the binding of Lola not only leads to depletion of nucleosomes at the Lola-I binding site, but also leads to phasing of nucleosomes around the binding site (Figure 3.14C). We can eliminate that the phasing is because of transcription as it is also observed at the distal sites (Figure 3.14C). In addition, it is also observed at the TSS sites only if it is centered at the Lola binding site. This lead us to speculate that Lola-I might bind strongly to its binding sites to allow other remodelers to phase nucleosomes using Lola as an anchor. It is also possible that Lola directly recruits nucleosome remodelers to displace nucleosomes. These results clearly suggest that Lola-I regulates the Pol II occupancy at the promoters by regulating the nucleosome occupancy indicating that Lola-I might be a pioneer factor.

3.2.7 Lola-I binds to the nucleosome edge

Binding of transcription factors including the pioneer factors to nucleosome bound DNA is significantly affected by the location of the transcription factor binding sites (TFBSs) within a nucleosome (Sekiya et al. (2009), Soufi et al. (2015), Zhu et al. (2018)). The binding is influenced by both how far the motif is from the edge to the center of the nucleosome (translational setting) and whether it is inward or outward facing with respect to the nucleosome core (rotational setting). To analyze if Lola-I prefers a particular motif localization, we plotted the nucleosome occupancy around the Lola-I binding sites using the MNAse-seq data (Figure 3.15A). We found preferential localization of Lola-I binding sites to edges of the nucleosomes. The Lola-I binding sites are localized at the edges of the nucleosomes in the early embryo when Lola-I is absent (Figure 3.15A). To corroborate these results, we analyzed the binding preference of purified Lola-I protein to nucleosome bound DNA using an *in vitro*-binding assay. Using the technique we developed before (Yu & Buck (2019)), we combined the traditional TF-nucleosome binding assay with high throughput sequencing, so that we can analyze multiple positions on the nucleosome simultaneously through just one binding reaction (Figure 3.15B,C). In this assay, we incubated Lola-PI to purified nu-

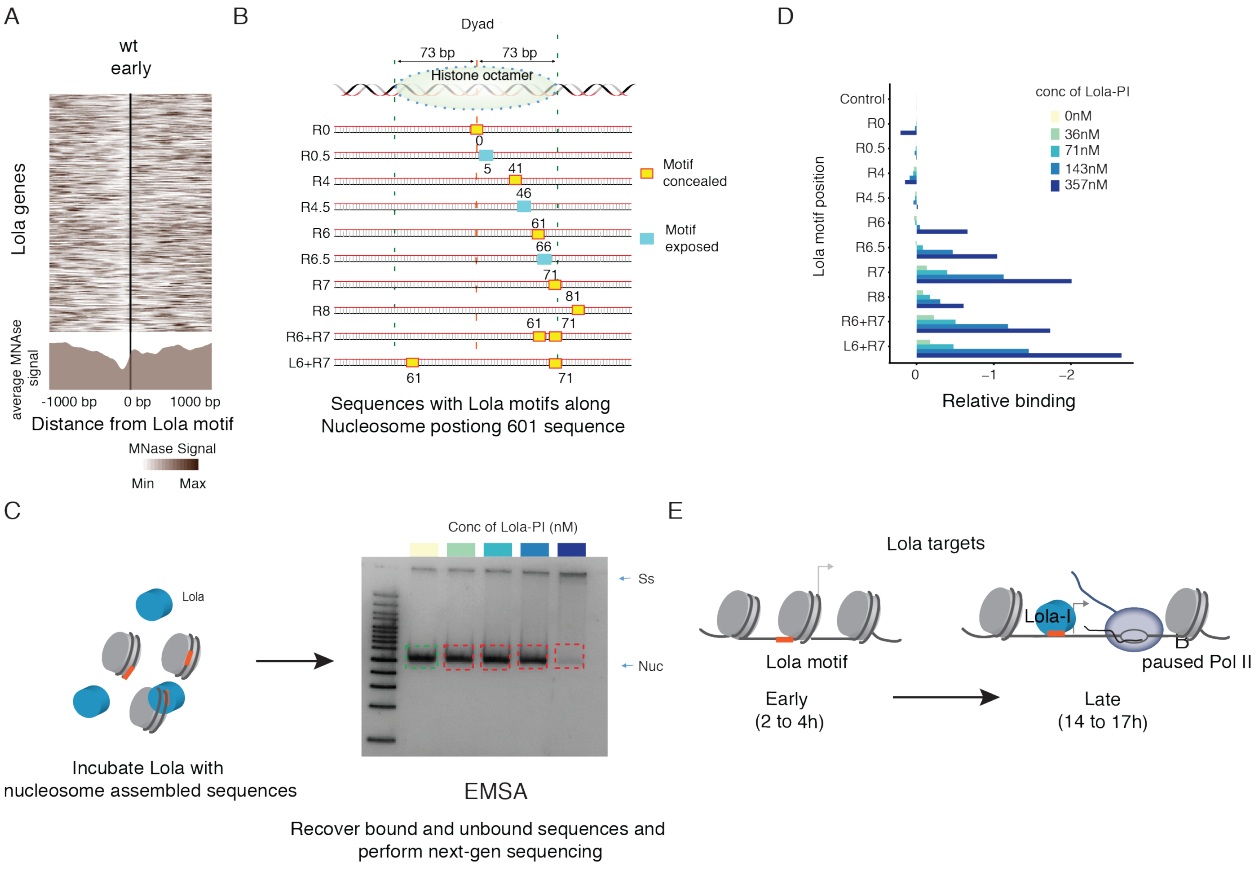

Figure 3.15: Lola-I binds to the nucleosome edge. A) Nucleosome occupancy around the Lola-I binding motifs are plotted. MNAse signal is flipped in some cases such that the motifs on either sides of nucleosome are aligned. Heatmap and metagene plots show the positioning of Lola-I motif along the nucleosomal edge. B) *In vitro* binding assay design is shown. Lola-I motifs are inserted at different position along the well-known nucleosome positioning 601-Widom sequence. C) The gel-shift assay shows the binding of Lola-I to nucleosome bound 601 sequences with differing Lola-I motifs. Both the nucleosomal and the super-shifted fractions are purified and sequenced to measured the relative affinity of Lola-I to different Lola-601 sequences. D) Relative binding of Lola-I to different 601-Lola sequences at different concentrations are shown. Lola-I binds only when the motif is located along the nucleosomal edge. E) Schematic explains the action of Lola-I during embryogenesis. Binding of Lola-I to its motifs located along the nucleosome edge during the late stages of *Drosophila* embryogenesis leads to decreased nucleosome occupancy and followed by the establishment of paused Pol II.

cleosomes bound to a strong nucleosome positioning sequence, 601-Widom sequence (Lowary & Widom (1998)), with or without the binding site and at different locations along the 601 sequence (Figure 3.15B,C). We measured the relative binding of Lola-I to different sequences by sequencing the super-shifted and unbound bands from a native polyacrylamide gel. As the concentration of Lola-PI increased, intensity of the band close to the well (Ss) gradually increased, while the nucleosome-only band (Nuc) almost disappeared (Figure Figure 3.15C). Our results demonstrate that nucleosome sequences containing a Lola-PI TFBS located near nucleosome edges (R6, R6.5, R7) are bound first at the lowest concentrations (Figure 3.15D). While the nucleosomes containing a Lola-PI TFBS located near the dyad (R0, R0.5, R4, R4.5) are not specifically bound as compared to the control nucleosomes (Figure 3.15D). Moreover, the template with TFBS at the linker region (R8) is not bound significantly more than the boundary regions, it only displays a comparable bound frequency as R6 (Figure 3.15D). Similar results were obtained by analyzing the nucleosome-only bands by measuring how many templates were bound by Lola-I and shifted away.

We also tested if there are any cooperative interactions involved in Lola-I binding, by positioning multiple Lola-I binding sites. Different from what we expected, binding strength of R6+R7 and L6+R7 looks pretty similar to the sum of R6 and R7, indicating that two neighboring sites might not produce strong enough synergistic effect to enhance Lola-PI binding to the nucleosomes at least at the distances tested (Figure 3.15D).

Based on our results we propose a mechanism of promoter opening and Pol II establishment by Lola-I. The expression of Lola-I during late stages of *Drosophila* embryogenesis leads to the binding of Lola-I to its motifs located along the nucleosome edge. Binding of Lola-I leads to decreased nucleosome occupancy, which in turn leads to the recruitment of Pol II, followed by the pausing of Pol II at the site, throughout the embryo at the Lola-I target promoters (Figure 3.15E). Our results thus provide the first example of how paused Pol II might be established de-novo during development.

3.2.8 Paused and TATA genes have different expression characteristics

Gene activation is inherently stochastic. The randomness involved in the multi-step assembly of multi-component transcriptional machinery is major source of this stochasticity. Preassembly of transcriptionally competent Pol II at the paused promoters, has been proposed to reduce the noise in gene expression (Boettiger et al. (2011)). Thus, we wanted to test whether the two promoter types, non-paused TATA promoters and the paused promoters, that we identified among the late genes may confer different expression characteristics. When we analyzed the single-cell expression data of TATA genes and highly paused genes within all tissues clusters, we made two observations (Figure 3.16). First, we found that TATA genes were expressed in a more tissue-specific fashion, with very little background in the remaining tissues. Thus, TATA genes are often undetectable in the tissue where they are not expressed, while highly paused genes typically had detectable and overall much higher levels of background expression. Second, the expression of TATA genes within a tissue cluster was often more variable between the cells, while highly paused genes within a cluster were more consistently expressed at the same level. This difference in background is well illustrated by the TATA gene *Ccp84Aa*, which is expressed in the epidermis, and the highly paused gene *Gip*, which is expressed in crystal cells (Figure 3.16A) and is schematically illustrated in Figure 3.16E.

To test whether the difference in background expression is statistically significant between all TATA and all highly paused promoters, we scored the percentage of cells within each cluster that have detectable expression for each gene. We then selected clusters with expression (cluster with the highest median expression) and those without (clusters with lowest median expression, bottom five clusters) and analyzed how many cells within the cluster showed detectable expression. The results show that paused genes are indeed more consistently expressed within a cluster of cells that show expression, but also have higher background expression when not expressed (Figure 3.16 C,D).

A simple explanation for this trend could be that paused genes have overall higher expression levels and therefore are generally detectable at higher levels. We found that at the embryonic

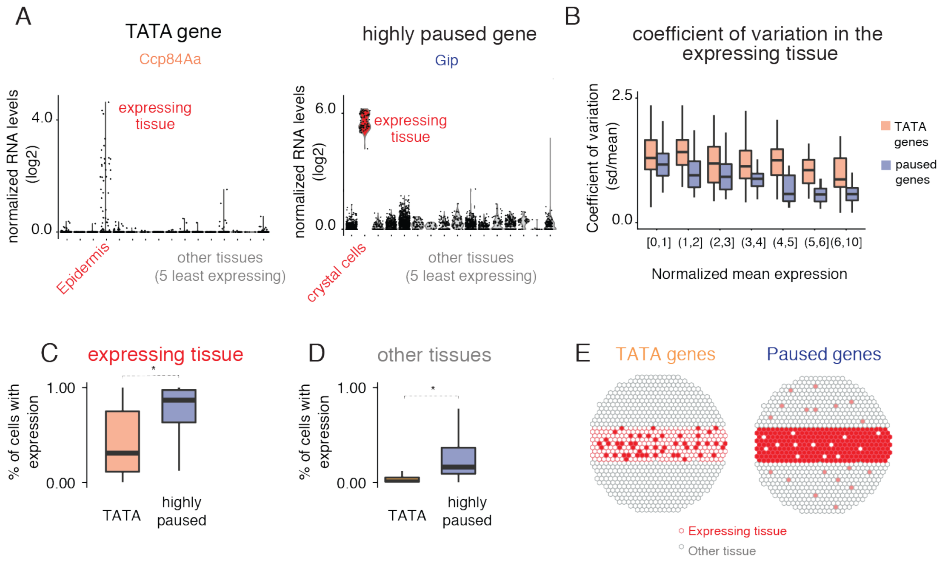

Figure 3.16: scRNA-seq shows differences in expression characteristics of TATA and Paused genes. (a) Normalized gene expression levels (read count normalized for each cell, log2) in different tissues, from the scRNA-seq experiment, for a TATA group gene and a Paused group gene are shown. The TATA gene, *Ccp84Aa*, shows noisy expression in Epidermis, without any detectable expression in many of the cells. The background expression for the TATA gene in other non-expressing tissues is very low. The Paused gene, *Gip*, shows very robust expression in Crystal cells, with high expression in all the Crystal cells, albeit with a high background expression in the non-expressing tissues. (b) The differences in cell-to-cell gene expression variability for the TATA and the Paused effector gene groups were analyzed using the scRNA-seq data. The coefficient of variation (standard deviation/mean) of gene expression was calculated for all genes in the tissue with the highest expression for each gene. The coefficient of variation was on average lower for the Paused genes compared to the TATA genes. (c and d) The frequency of cells with any detectable expression was calculated in tissues with the highest expression for each gene (expressing tissue) and in five other tissues with the least expression for each gene (other tissues). (c) The frequency of cells with detectable expression in the expressing tissues, a measure of expression robustness, is lower for the TATA genes compared to the Paused genes. (d) The frequency of cells with detectable expression in the other tissues, a measure of background expression, is also lower for the TATA genes compared to the Paused genes. (e) A schematic of the expression characteristics of TATA and Paused genes is shown. Grey hexagons represent the non-expressing tissue and red hexagons represent the expressing tissue.

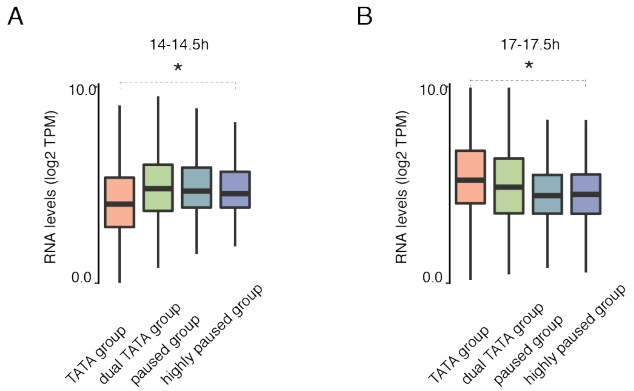

Figure 3.17: TATA gene expression increases over time: RNA levels (log2 TPM) from (a) 14-14.5 h and (b) 17-17.5 h embryos at the different effector gene groups are shown. (a) The TATA genes are expressed at levels lower than the paused genes at 14-14.5 h. (b) By 17-17.5 h, the TATA genes are expressed at levels higher than the paused genes.

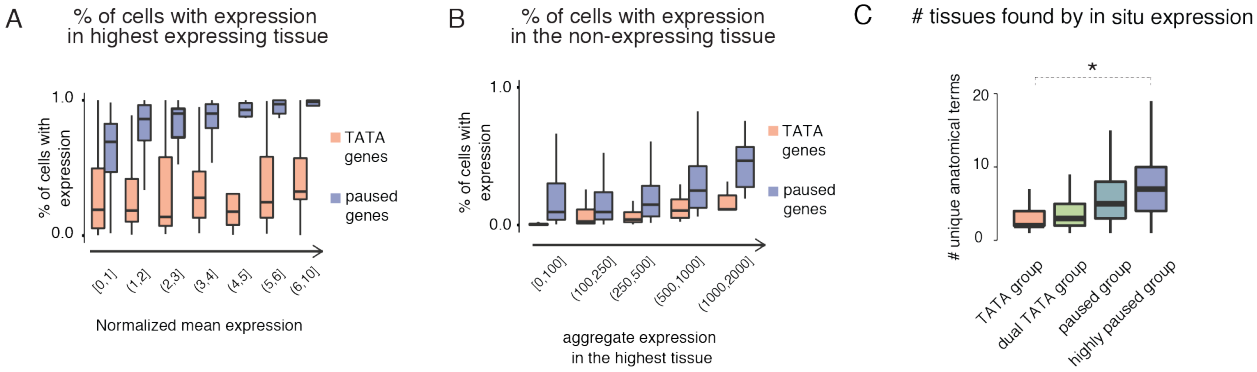

Figure 3.18: Differences in the expression profiles of effector gene groups: (a) The differences in cell-to-cell gene expression variability between the TATA and the Paused effector gene groups in the highest expressing tissue was analyzed using the scRNA-seq data. The coefficient of variation (standard deviation/mean) of gene expression was calculated for all genes in the different effector gene groups, stratified by mean gene expression levels. The coefficient of variation was on average lower for the Paused genes compared to the TATA genes. (b) The frequency of cells with detectable expression in the five least expressing tissues (different for each gene) was calculated for all genes in the different effector gene groups, stratified by total expression in the expressing tissues for each gene. The frequency of expressing cells was on average higher for the Paused genes compared to the TATA genes. (c) Number of annotation terms associated with each gene in the BDGP insitu database is calculated. It is a measure of whether the expression of a gene is restricted to specific subsets of tissue.

stage (14-14.5 h) we performed our scRNAseq analysis, the TATA genes were slightly at lower level than the paused genes but show higher expression at later stages (3.17). To rule out the effect of expression differences on our results, we therefore analyzed the relationship between median expression and percentage of cells with detectable expression (Figure 3.18A). Consistent with the whole embryo RNA-seq data at 14-14.5h, TATA genes were not yet as highly expressed, and the lower expression of TATA genes could have contributed to the observed differences to some extent. However, it was also clear that this was not a sufficient explanation for the observed differences. Many TATA promoters with high levels of transcripts in a cluster nevertheless had a high fraction of cells with undetectable levels of transcripts, and as a result, the TATA promoters overall followed a different distribution than the paused genes (Figure 3.18A). In addition, the relationship between median expression and the coefficient of variance also show higher variation for TATA genes compared to the paused genes (Figure 3.16B). On of the reasons why expression of TATA genes appear noisier could be because they are expressed in sub-populations of cells even within a single cluster. Analysis of the traditional *in situ* data from BDGP supports this idea, since less number of tissue types are associated with the TATA genes (Figure 3.17C). However, previous results suggest that in addition to this effect, TATA genes might indeed be inherently noisier than the paused genes. Similarly, even after accounting for the expression differences, the paused genes show higher background expression than the TATA genes. These results suggest that the paused genes have more robust expression when expressed, but also have higher background expression in tissues where these genes are not expressed. TATA genes, on the other hand, are highly tissue-specific with little background expression, but result in larger variations among the cells that show expression.

3.2.9 Expression of the target genes is reduced in the absence of Lola-I

Lola-I regulates the loading of paused Pol II to a specific set of late promoters. This allows us to utilize the Lola-I mutants and evaluate the causal role of paused Pol II and promoter accessibility on gene expression and in reducing the noise in gene expression. We first performed whole embryo

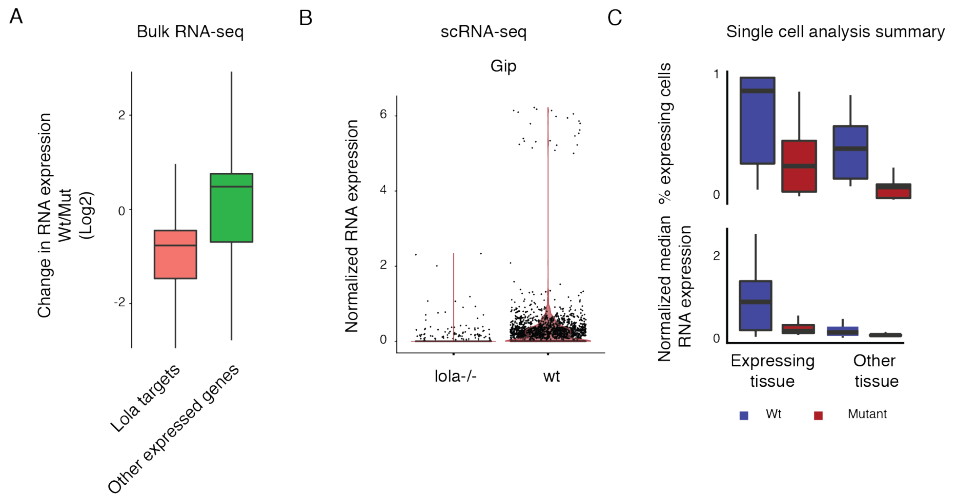

Figure 3.19: Changes in expression of Lola-I targets. A) Boxplot shows the changes in bulk RNA levels between the *lola*^{wt} and *lola*^{ORC4} embryos in Log2. Lola-I target genes are specifically downregulated in the *lola*^{ORC4} embryos. B) scRNAseq expression data of the Lola-I target gene *Gip* is shown. *Gip* shows very high expression in the few crystal cells, but also have a background expression in other cells. In the *lola*^{ORC4} embryos both the specific expression and the basal expression is reduced. C) Boxplot shows downregulation of both the specific expression and the basal expression of the Lola-I targets are reduced in the *lola*^{ORC4} embryos compared to the *lola*^{wt} embryos. Expression is measured using two metrics: frequency of cells with detectable expression and the median expression.

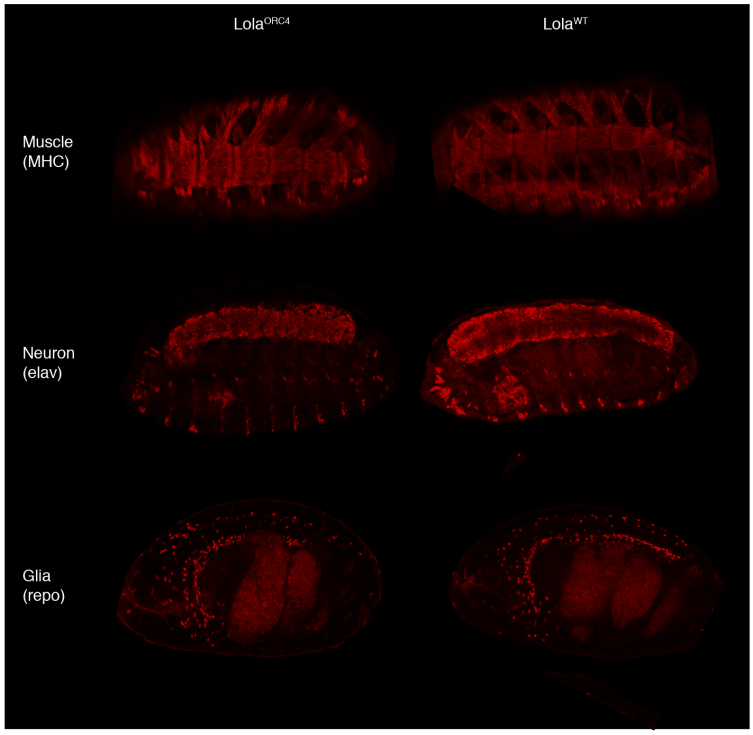

Figure 3.20: Lola-I mutants do not show any gross visible phenotypic defects. Visualization of Muscle, Neurons and Glia by immuno-staining using antibodies against MHC, elav and repo respectively.

RNA-seq in the *lola*^{wt} and *lola*^{ORC4} embryos. We observed that the expression of Lola-I targets is reduced in the Lola-I mutant emrbyos (Figure 3.19A).

We further analyzed the changes in gene expression of Lola-I target genes in different tissues using single cell RNA-seq. We performed Drop-Seq experiments in 14-14.5hr *lola*^{wt} and *lola*^{ORC4} embryos. We specifically analyzed the changes in basal expression of Lola-I target genes and the expression of genes in the tissue where they are active. The Lola-I target gene, *Gip*, which known to be expressed in the crystal cells, shows high expression in crystal cells but also shows basal expression in other tissues (Figure 3.19 B). Our results show that both the basal expression and the expression in the crystal cells are reduced in the Lola-I mutant embryos (Figure 3.19 B). We also observed similar changes in both the basal expression and active expression of Lola-I targets genome-wide to be also reduced in the Lola-I mutants (Figure 3.19C). It is important to also note that only moderate reduction in gene expression is observed in many cases. This suggest a model where gene activation happens in two stages, with establishment of paused Pol II followed by the enhancer mediated pause-release. In the absence of Lola-I mediated Pol II recruitment, enhancers could still activate gene expression, albeit less efficiently.

We then analyzed the phenotypic consequences of Lola-I mediated Pol II recruitment and the resulting reduction in gene expression. The *lola*^{ORC4} embryos die at the end of embryogenesis. However, the *lola*^{ORC4} embryos do not show any gross visible phenotype (Figure 3.20). We analyzed the neuronal, muscle and glial structures (Figure 3.20). We could not observe any gross phenotypic differences. This is consistent with Lola-I targets being late genes and the formation of these structures is primarily driven by early patterning events. Thus, the embryonic lethality observed in the *lola*^{ORC4} could be result of rather functioning defects in these tissues.

3.2.10 Paused promoters have higher chromatin accessibility

The expression characteristics we observed for TATA genes here are reminiscent of TATA genes in yeast, which are often induced in response to environmental stimuli and which tend to show high expression variability. This expression variability has been linked to the presence of a pro-

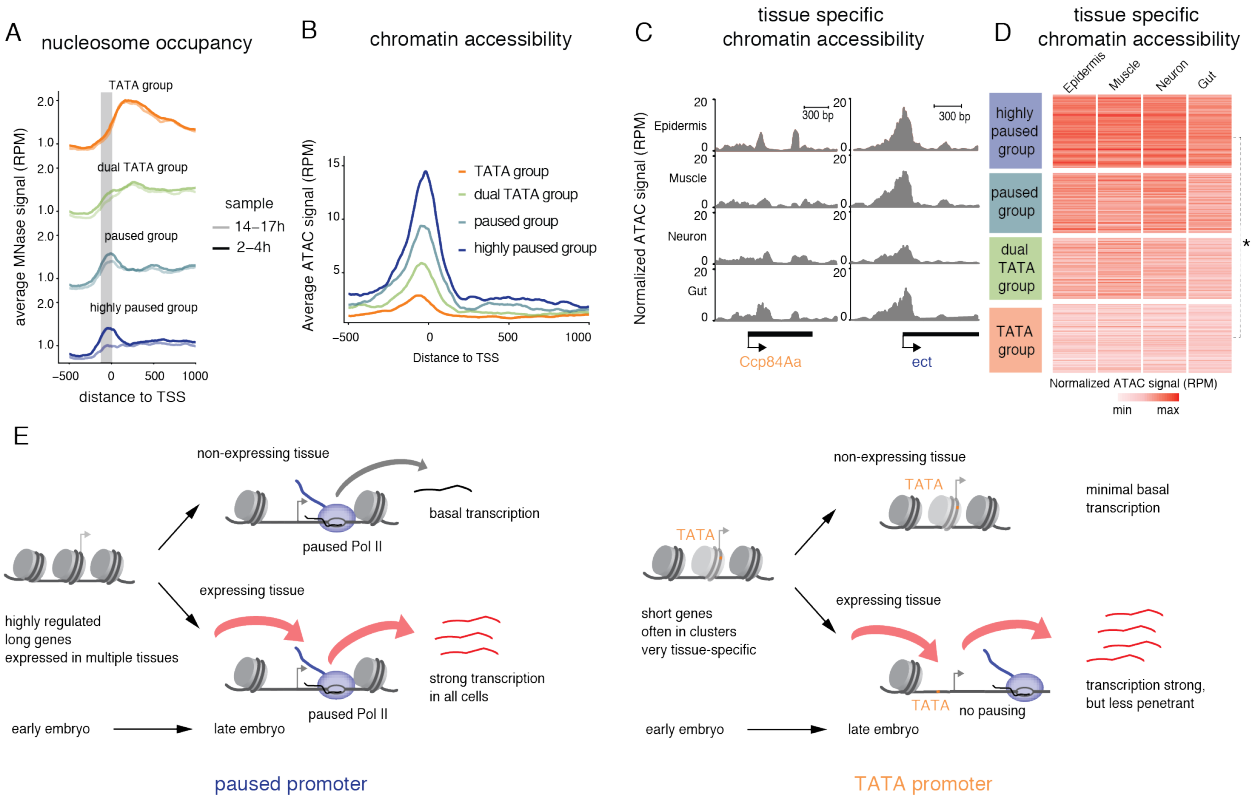

Figure 3.21: Paused genes are more accessible than the TATA genes. (a) Average read-count normalized MNase signal (RPM) from 2-4h and 14-17h at different groups of effector genes are shown. Grey area highlights the changes in nucleosome occupancy at the Paused genes. (b) Average read-count normalized ATAC-seq signal (RPM) from 14-17h at different groups of effector genes are shown. Paused genes show higher accessibility at the promoter region than the TATA genes. (c & d) Tissue specific ATAC-seq measures the chromatin accessibility in different tissues from 14-17 h embryo. (c) Read-count normalized ATAC (RPM) signal at individual genes from the TATA group (*Ccp84Aa*) and Paused gene group (ect) are shown. (d) Read-count normalized ATAC signals (RPM) from different tissues were measured from 150bp upstream of the TSS to the TSS. Paused genes show higher average accessibility across all tissues compared to the TATA genes (Wilcoxson two-sided test, *P $<10^{-15}$). (e) A model figure explaining the different characteristics of the TATA and Paused genes.

moter nucleosome, which needs to be removed during activation. Less variability is observed at promoters where the promoter nucleosome is depleted. We have previously observed that paused and TATA promoters in the *Drosophila* embryo have different nucleosome configurations and that the highly paused promoters showed strong nucleosome depletion when paused Pol II was present (Gaertner et al. (2012), Chen et al. (2013)). When we analyzed the promoter groups among our effector genes, we found the same pattern. We found that the higher the pausing, the more the promoters became nucleosome-depleted in the late embryo as compared to the early embryo, while TATA promoters did not show such change (Figure 4.6A). Thus, it is possible that the expression characteristics involve different nucleosome configurations. If the promoter nucleosome is a barrier to transcriptional activation, one might expect that paused genes have a lower activation barrier because the nucleosome is already depleted. However, from our nucleosome data, it was unclear whether TATA promoters are nucleosome-depleted. To test whether TATA promoters are more accessible than the paused promoters, we performed ATAC-seq experiments in 14-17h embryos (Figure 4.6B,C,D). We found that TATA promoters showed the least accessibility, and Pol II pausing correlated with higher average promoter accessibility (Figure 4.6B). This accessibility pattern for each promoter varied little between tissues, indicating that it reflects the basal state of the promoter and is not strongly influenced by activation levels (Figure 4.6C). Overall, TATA promoters had significantly lower accessibility than highly paused promoters. This is consistent with the hypothesis that the highly paused promoters are nucleosome depleted, while TATA promoters, like in yeast, are occluded by a promoter nucleosome. We propose that differences in the nucleosome occupancy between the non-paused TATA and paused promoters could contribute to the expression differences between these promoter types (Figure 3.21E).

3.2.11 Lola-I poises genes for activation

We wanted to validate our scRNA-seq experiments further and gain insights into how Lola-I mediated pre-loading of paused Pol II affects the transcriptional dynamics. Single-molecule FISH analysis can provide quantitative measurements such as the number of transcripts per cell, the

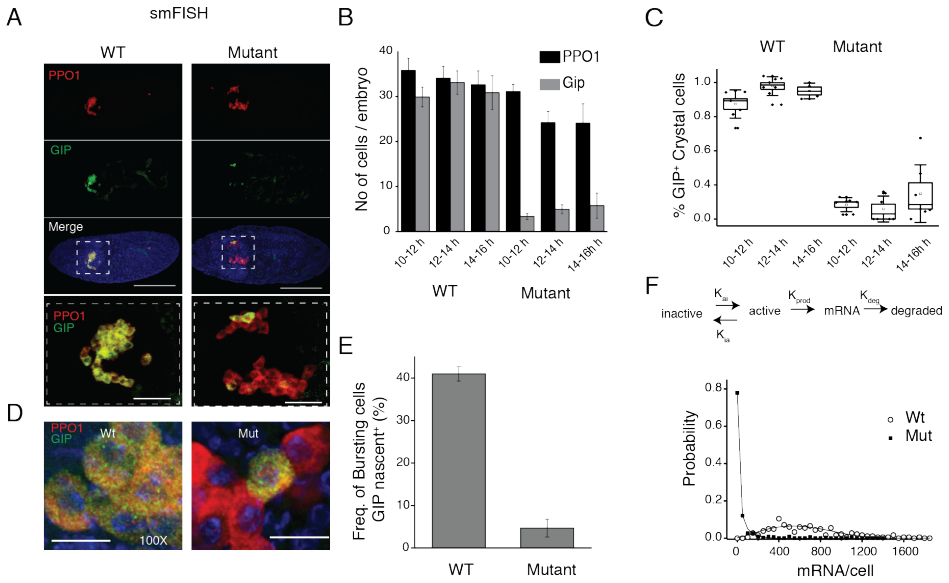

Figure 3.22: smFISH shows changes in activation dynamics. A) smFISH data of *Gip*, displayed in green, and *PPO2* and PPO2 genes displayed in red. See methods for details on smFISH probe generation, labelling and purification, staining protocol, imaging, and analysis. Overviews (top 3 rows of panels, scale bar = 100 microns) and zoomed regions (bottom panel, scale bar = 20 microns) are shown for data acquired with 40x magnification. B) Quantification of number of cells per embryo expressing high levels of *PPO2* and PPO2 vs. *Gip* in wt and mutant (see Methods) C) Percentage of Crystal cells (defined by high expression of *PPO2* and PPO2) that have visibly high levels of *Gip* expression. D) Zoomed region of Crystal cells in wt and mutant, taken with 100x magnification (see Methods) (scale bar = 10 microns). E) Frequency of bursting cells manually tabulated as the percentage of *PPO2* positive cells with a visible nascent transcript of *Gip* expression (see Methods for details of determination of nascent transcripts). F) Probability of mRNA/cell calculated from 100x magnification data (dots) and fits (lines) according to the model indicated (Raj et al. (2006)). The ratio of Kai and Kia was fixed based on the nascent transcript count from See Methods for a description of the model and conversation of intensity per cell to mRNA per cell.

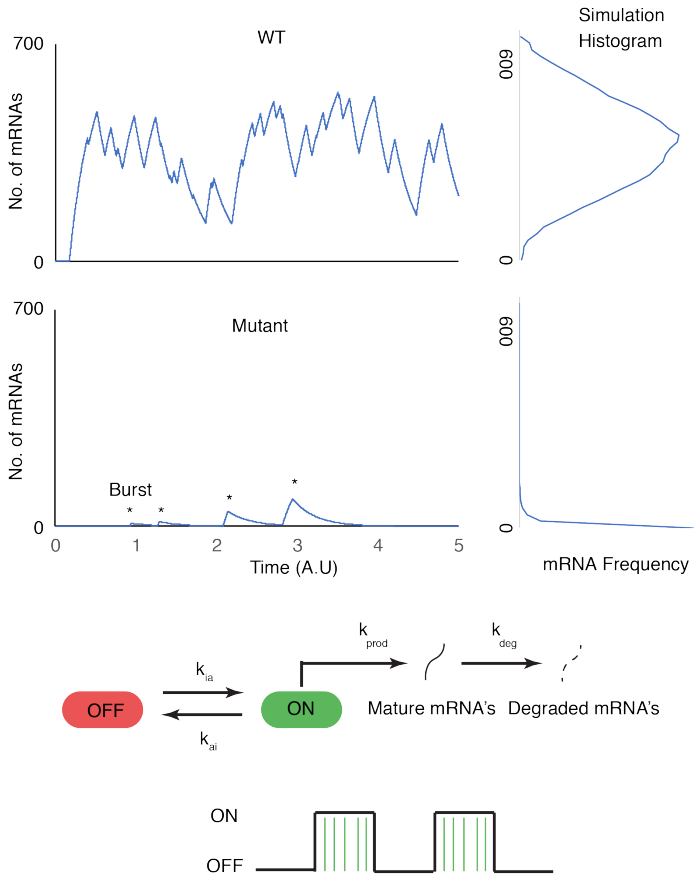

Figure 3.23: smFISH simulation shows changes in activation dynamics. Simulation of gene expression in the Lola WT and mutant embryos shows that the number of transcriptional bursts are reduced in the Lola-I mutant embryos. This is primarily caused by the reduction of the activation rate.

number of bursting cells in a population, the size of the transcriptional burst (Raj et al. (2006)). This could be utilized to model the expression dynamics in the Lola-I wt and mutant embryos and test how the promoter opening by Lola-I affects the gene expression dynamics in detail. We analyzed the expression profiles of the Lola-I target, *Gip*, which is expressed in the crystal cells (Figure 3.22A,B,C). We chose to analyze *Gip*, because we could easily identify crystal cells, where it is expressed, with the marker genes *PPO2* and PPO2. These marker genes are expressed exclusively in the crystal cells, and smFISH probes for the marker genes were available. Our smFISH analysis showed that *Gip* is expressed at a high level in almost all the *PPO2* and PPO2 positive crystal cells in the $lola^{\mathrm{wt}}$ embryos (Figure 3.22A,B,C), consistent with our scRNA-seq data (Figure 3.19B). In the Lola-I mutant $lola^{\mathrm{ORC4}}$ embryos, the expression of the *Gip* gene is reduced but not uniformly in all the crystal cells (Figure 3.22A,B,C). Most of the crystal cells showed no detectable expression for *Gip*, but in a few crystal cells, *Gip* was detected at almost wildtype levels (Figure 3.22A,B,C). Since only a limited number of crystal cells were captured in the scRNAs-seq analysis, it is possible that we did not observe these high-*Gip* cells in the mutants in our scRNAseq data. We eliminated the possibility that the observed differences are due to slight developmental delays in the Lola-I mutants since even the 14-16 h mutant embryos had less percentage of *Gip* expressing crystal cells than the earlier stage wildtype embryos(10-12 h and 12-4 h) (Figure 3.22B,C). Intuitively, this could be explained by the *Gip* promoters switching to the active state less frequently in the mutant embryos, but while the promoter is in the active state produces a similar amount of transcripts. To test this model, we analyzed our *Gip*-smFISH data at high resolution. Consistent with our hypothesis, we observed that the frequency of bursting cells, with observable nascent expression in the nucleus, was decreased in the mutant (Figure 3.22D,E). We further fitted the observed mRNA distribution in the wildtype and mutant embryos to a simple two-state expression model (Figure 3.22F). Our model fitted to the observed values very well (Figure 3.22F, 3.23). Our modeling results suggest that the main difference between the wildtype and mutant embryos is the activation frequency. We also observed slight decrease in the production rate and slight increase in the activation rate. Moreover, since paused Pol II itself has been proposed to

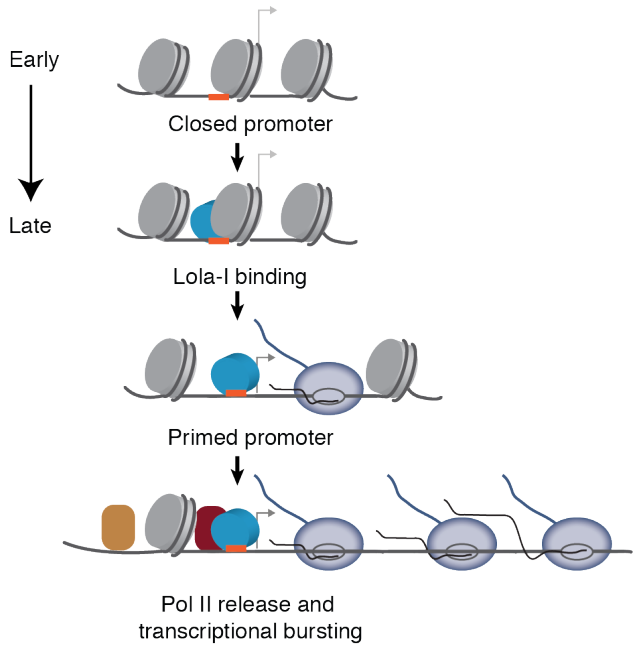

Figure 3.24: Schematic explains the action of Lola-I during *Drosophila* embryogenesis. The Lola-I target promoters are closed by the presence of nucleosomes during the early stages. Binding of Lola-I to its motifs located along the nucleosome edge during the late stages of embryogenesis leads to decreased nucleosome occupancy and followed by the establishment of paused Pol II. Further gene expression is triggered in specific tissues by the recruitment of pause-release inducing factors.

help keep the promoters open, it is not clear once the promoter becomes active, whether Lola-I is required for the subsequent burst. Stable pausing of Pol II has been proposed to happen during the inter-transcriptional burst period. This would predict that we would observe more number of *Gip* positive cells in the Lola-I mutant embryos over time (Figure 3.22C). But our results argue that Lola-I might be required to keep the promoter open even after the first transcriptional burst (Figure 3.22C). However, at other promoters with more stable paused Pol II, Lola-I might be playing a role only during the first round of transcriptional burst.

Our results support the model that Lola-I is priming the target promoters for activation. We propose that the removal of the promoter nucleosome is a major rate-limiting step in gene activation. Priming of promoters by Lola-I by removing the promoter nucleosome and establishing paused Pol II allows frequent and faster switching to the active state (Figure 3.24).

Chapter 4

Discussion

During gene activation, enhancers transition from an inactive closed state through an open poised state before becoming fully active. This transition is initiated by the binding of pioneer factors and histone modifications (Zaret & Carroll (2011), Spitz & Furlong (2012)). We propose that similar to the intermediary poised state of the enhancers, promoter-proximal pausing of RNA polymerase II could be an intermediate state that promoters transition through during gene activation. While promoter-proximal pausing of RNA polymerase II was discovered a while back, it has not been clear whether Pol II recruitment to the paused promoters occur by default or is regulated. Before the beginning of my thesis, it has been shown that at a group of *Drosophila* genes, paused Pol II occupancy changes during *Drosophila* embryonic development, suggesting that the Pol II occupancy at the paused promoters might be regulated (Gaertner et al. (2012)). This lead us to search for factors that could regulate the loading of paused Pol II at these genes. Here, we have shown that the loading of paused Pol II at promoters is regulated. We have also identified how paused Pol II is established at a group of late *Drosophila* genes by a promoter pioneer factor Lola-I and provided mechanistic insights into this process. Our discovery that the loading of paused Pol II is regulated implies that the activation of these genes can be regulated in a combinatorial fashion. A straightforward model would be that activation happens in two stages: The first step is the establishment of paused Pol II, which leads to a competent promoter that is responsive to further activation signals. The second step is the binding of other transcriptional activators at enhancers, which stimulates the gene expression by pause release. Active enhancers might still be able to act on closed promoters, albeit less efficiently.

Many enhancers are primed for activation in a restricted fashion by the lineage/tissue-specific expression of pioneer factors (PU.1, Pax7, FoxA, etc) (Heinz et al. (2010), Budry et al. (2012), Iwafuchi-Doi & Zaret (2014)). For example, specific enhancers are primed by the pioneer factor FoxA in the foregut, PU.1 in the macrophage-lymphoid-lineage, etc. This is followed by the acti-

vation of enhancers by transcription factors such as ER, GR, C/EBP, etc., which are expressed in a further restricted manner (Heinz et al. (2010), Lupien et al. (2008), Iwafuchi-Doi et al. (2016)). Necessitating prior action of pioneer factors, before the transcriptional activators can act on an enhancer, allows combinatorial logics to be employed at these enhancers and achieve precise gene expression. In the case of Lola-I, its expression is restricted to the late stages of *Drosophila* embryogenesis; thereby, these promoters are primed only at the late stages. Interestingly, unlike many of the other pioneer factors, Lola-I is instead a stage-specific pioneer than a lineage-specific pioneer factor.

4.1 Mechanisms of action by Lola-I

Nucleosome bound promoters are not accessible to many transcription factors, including the transcription machinery. However, it has been proposed that the presence of specialized DNA binding domains allow pioneer factors to access the nucleosome bound DNA better. For example, the winged-helix domains of FoxA1 mimic H1 linker interactions and allow binding to nucleosomal DNA (Cirillo & Zaret (1999), Cirillo et al. (2002)). Given the diversity of pioneer factor binding domains, they may employ different mechanisms to access their nucleosome occluded target site. Thus, the exact molecular basis for the ability of pioneer factors to access their target DNA sequences in closed chromatin remains obscure.

Binding of pioneer factors to nucleosome bound DNA has been proposed to have two components: sequence-driven binding that is mediated by the DNA-binding domain, and non-specific interactions with nucleosomes that is mediated by the additional regions of the protein (Cirillo & Zaret (1999)). In this study, Lola-I binding to the nucleosome is stronger when the Lola-I motif is at the edge than when it is in the linker region. This suggests that the binding of Lola-I could be mediated by both the sequence driven component and non-specific interactions with nucleosomes.

Following the binding of pioneer factors to DNA, the molecular strategies employed by them to elicit chromatin accessibility is also not clear. In some cases, cooperative binding of several factors could displace the nucleosomes without any remodelers. While in other cases, the chro-

matin opening could be achieved together with the action of chromatin remodelers (Tsukiyama et al. (1994)). Lola-I could also make regions accessible by recruiting chromatin remodelers like some of the other BTB domain-containing factors (Tsukiyama et al. (1994)). It is surprising that the exact mechanisms of pioneering activity are not known for many pioneer factors. In our case, our preliminary experiments suggested that the absence of any pioneering activity following the overexpression of Lola-I in the early embryo. This presents an opportunity to look at the intermediate state between the binding of a pioneer factor and the complete remodeled accessible state. This could shed some light on the pioneering process of Lola-I.

Our preliminary pull-down experiments suggested that the Lola-I does not interact directly with Pol II. Thus the mechanisms of action of Lola-I might be through the opening of the promoter region rather than by direct recruitment of Pol II. It is not clear if the basal recruitment of Pol II to the open promoters is enough to establish high Pol II occupancy at these promoters. During gene activation, enhancers might also stimulate this basal rate of recruitment in addition to stimulating the Pol II release from the paused site.

While there is an apparent increase in the expression of Lola-I during *Drosophila* embryogenesis, it is not clear if it is the only driver of the opening of the late promoters. There could also be other concurrent changes than happen during embryogenesis which are also required for the opening of the Lola-I target promoters. Our preliminary results using the overexpression of Lola-I during the early stages suggest that it is not just the expression of Lola-I, which drives the opening of the Lola-I target promoters. One possibility is that there are chromatin remodelers that interact with Lola-I, which are missing in the early stages. It is also possible that there are repressive factors that prevent the action of Lola-I during the early stages.

The different splice isoforms made from the *lola* locus, in addition to the same N-terminal BTB domain, and different C termini zinc finger domains that bind to diverse sets of DNA motifs (Goeke et al. (2003)), also are composed mostly of low complexity domains. Such low-complexity domains (LCDs), also known as intrinsically disordered regions (IDRs), are thought to facilitate the formation of phase-separated domains (Kato & McKnight (2018); Brangwynne et al. (2009)).

There is growing evidence that interactions leading to phase transitions play an important role in transcriptional regulation (Chong et al. (2018); Cho et al. (2018) Boehning et al. (2018); Strom et al. (2017); Kato & McKnight (2018)), although the precise nature of this role remains unclear. One hypothesis is that domains formed by homo- and heterotypic interactions between LCDs serve to locally enrich transcription factors, potentially in the vicinity of their targets (Tsai et al. (2017)), thereby altering their local concentration and modulating their binding dynamics. Such a role has been proposed for Zelda (Dufourt et al. (2018), Yamada et al. (2019)). In the case of HP1, HP1 mediated phase transition has been proposed to be necessary for the compaction of the HP1 regulated regions (Larson et al. (2017)). These interactions has also been proposed to modify the nucleosome structure (Sanulli et al. (2019)). It is not known if there are LCD mediated interactions between Lola-I proteins or if such interactions play a role in the pioneering function of Lola-I.

4.2 Role of different Lola isoforms

It is conceivable that the various isoforms of Lola are also involved in the establishment of paused Pol II. The different isoforms of Lola have different sequence-specific DNA binding domains and have diverse expression patterns during development (Goeke et al. (2003), Enuameh et al. (2013)). While Lola-I establishes Pol II throughout the embryo, other isoforms of Lola might recruit Pol II in a more restrictive way. There are cases of tissue-specific recruitment of paused Pol II in *Drosophila*. For example, the *sim* gene which plays a significant role in the development of the ventral midline, (Nambu et al. (1991) Wheeler et al. (2009)) contains paused Pol II in the ectoderm, but not in the mesoderm where the gene is never expressed (Levine (2011)). Also, there are isoforms of Lola with the same binding domains/ exon as Lola-I but with different 5' UTR. The functions of these isoforms are also not known. Moreover, one of the isoforms of Lola, Lola-G, has the same DNA binding motif as Lola-I, but the DNA binding domain is different than Lola-I (Giniger et al. (1994), Goeke et al. (2003)).

Lola-I mediated opening of promoter and establishment of paused Pol II happens at a fraction of late paused genes. It is still not clear how the paused Pol II is established at the remaining

fraction of the late paused genes. Speculatively, the other Lola isoforms might be involved in the opening of these promoters. Moreover, there could also be stage or tissue-specific establishment of paused Pol II at many promoters in larval, pupal, or other stages.

Lola was first identified in *Drosophila* as a putative transcription factor required for axon guidance in the nervous system. The name Longitudinals lackings come from the role of Lola in guiding axons away from the embryonic midline (Seeger et al. (1993), Giniger et al. (1994)). Several other functions of Lola were later identified, including its role in ISNb motorneuron innervation of ventrolateral musculature (Madden et al. (1999), Goeke et al. (2003)), neuronal stem cell differentiation (Southall et al. (2014)), programmed cell death in nurse cells during oogenesis (Bass et al. (2007)), etc. Lola-I isoform specifically lacks a strong neuronal phenotype, with only minor defects in motor neuron intervention (Goeke et al. (2003)). Interestingly, some of the studies suggest that the some Lola isoforms could also mediate chromatin condensation (Zhang et al. (2003)).

4.3 Are there other promoter pioneers that establish paused Pol II?

GAGA factor is a well-known pioneer factor in Drosophila. It is necessary for the establishment of open-chromatin both at the promoter and at distal regions (Fuda et al. (2015)). GAF has also shown to be necessary for the paused Pol II occupancy (Fuda et al. (2015)). GAF can interact with NELF and, therefore, can also affect the paused Pol II stability (Li & Gilmour (2013)). However, GAF is expressed throughout embryogenesis. Therefore it is not clear if it plays a role in regulating the timing of paused Pol II occupancy at certain regions. It is still possible that the changes in concentration of GAF could regulate, to some extent, the binding of the GAF to particular targets. GAF binding could also be regulated depending on the number of binding sites, the location of binding sites within a nucleosome or interaction with certain other factors (Espinás et al. (1999), Katsani et al. (1999)). Thereby, GAF could still regulate the timing of paused Pol II occupancy at certain regions during development. Many of the Lola- I target promoters have GAF motif. Thus, Lola-I can cooperatively interact with GAF to establish paused Pol II (Lomaev et al. (2017)). This could be the reason why some promoters do not show any or little reduction in Pol II occupancy

when Lola-I is mutated. The role of Lola-I at these sites might be redundant. Other compensatory mechanisms, including that through GAF, might be involved in establishing paused Pol II at these sites. This idea of cooperativity between different pioneer factors has been observed with other pioneer factors (FoxA, GATA, GR, etc.,) (Jin et al. (2014), Cirillo et al. (2002), Hebbar & Archer (2007)).

PHA-4 is another pioneer factor which has been shown to important for the recruitment of Pol II at the foregut genes during *C.elegens* embryogenesis (Hsu et al. (2015)). However, it is not clear whether the Pol II established by PHA-4 is the bonafide paused Pol II observed in other metazoans. The Pol II observed at the PHA-4 sites is upstream of the TSS (Hsu et al. (2015), Maxwell et al. (2014)), not the typical 30-50 bp downstream peak observed in other eukaryotes (Core et al. (2008), Nechaev et al. (2010)). Additionally, NELF, which plays a significant role in establishing paused Pol II, is not present in *C.elegens* (Narita et al. (2003)). It is still fascinating to see other forms of Pol II loading at promoters, which might also prime genes for activation. It is also interesting to note that PHA-4 binding has been proposed to first bring Pol II to its target sites before the changes inaccessibility. This is unlike the case of Lola-I, where we have observed that opening of promoters by Lola-I can occur independent of Pol II recruitment and, therefore, likely precedes the Pol II recruitment at the Lola-I promoters.

4.4 Establishment of paused Pol II in mammals

The importance and prevalence of Pol II pausing in mammalian development is evident. However, it is not clear whether promoter opening is used as a general regulatory mechanism of gene activation, given that the promoter is generally more accessible in mammals compared to *Drosophila*. There are, however, quite a few examples of promoters being primed for activation at specific stages of development (Lu et al. (2016)). While Lola itself is not present in mammals, there could be other factors, including the many BTB-domain factors found in mammals, that are involved in the regulation of paused Pol II recruitment in mammals.

4.5 BTB domain factors

The BTB domain (Broad Complex, Tramtrack, and Bric-a-Brac) (Zollman et al. (1994)), also known as the POZ domain (Poxvirus and Zinc finger) (Bardwell & Treisman (1994)), was initially identified in *Drosophila* and poxvirus proteins. It is a highly conserved protein-protein interaction, and multimerization domain (Stogios et al. (2005)). It is found across eukaryotes and is present in more than 200 proteins in humans (Shiro & Natalie (2010)). Structurally, the core of the BTB domain consists of five α-helices, and three β-strands (Ahmad et al., 1998). In addition to the conserved core, protein-specific extensions lead to specific interactions (Ahmad et al. (1998)). At many BTB domain-containing proteins, BTB mediated interactions are crucial for the function. BTB domain is not only involved in the homo and hetero-dimer formation between the BTB family proteins (Ahmad et al. (1998), Bonchuk et al. (2011)), but is also important for the formation of large multimeric complexes (Bonchuk et al. (2011), Espinás et al. (1999)) and interactions with non-BTB domain proteins (Huynh & Bardwell (1998), Furukawa et al. (2003), Xu et al. (2003), Pintard et al. (2003), Geyer et al. (2003), Huynh & Bardwell (1998)). In the case of GAGA factor, oligomerization has been shown to be essential for its cooperative binding (Espinás et al. (1999)). It has been proposed that the binding of the oligomeric GAF complex to multiple motif sites binds the DNA in a way that it keeps the GAF target sites in an open conformation. It remains to be seen if Lola-I forms such complexes and the formation of large oligomers play a role in its pioneer activity.

The BTB/POZ has been shown to involved in a broad range of processes during development in numerous studies done in *Drosophila* and humans (Psq, Ttk, Broad, BCL-6, etc.) (Ferres-Marco et al. (2006), Xiong & Montell (1993), Lai & Li (1999), Li et al. (1997), Karim et al. (1993), Kojima et al. (2001)). The mutations in the BTB domain proteins often lead to cancers (Kelly & Daniel (2006), van Roy & McCrea (2005)). PLZF and BCL-6 are some of the most studied human BTB domain factors. Presence of BTB domain across proteins of vastly different functions suggest that it might be only involved in protein-protein interactions. This suggests that while the BTB domain might help Lola-I in bringing chromatin remodelers to the target site or keep it open

by binding as an oligomer, it might not play a role in the initial binding of Lola-I to the nucleosome bound target sites.

4.6 Enrichment of TATA genes at tissue-specific effector genes

TATA box-containing promoters are often associated with tissue-specific effector genes, genes that are important for the structure and function of differentiated tissues. The underlying reason for this promoter preference is unclear. Here we analyzed the relationship between promoter types and gene expression across tissues, using single-cell RNA-seq performed on differentiated cells from the late Drosophila embryo. Our analysis confirmed the usage of TATA-containing promoters by the tissue-specific effector genes. In addition, they are also induced by the paused promoter type. These TATA depleted paused promoters have high levels of paused Pol II throughout the embryo, even in tissues where the genes are not expressed. The two promoter types show differences in promoter accessibility, gene expression noise, and background expression. We propose that the TATA-containing promoters have a higher activation barrier but preferentially evolve in tissue-specific genes due to their strong activation potential and higher evolvability.

Tissue-specific genes can be roughly divided into those that regulate the tissue's development and those known as effector genes, which regulate the tissue's form and function in its differentiated state. Developmental genes are required from the earliest embryonic stages on to specify the tissue's location, pattern, and cell identities. While the effector genes are expressed later, once the tissue is specified, to give the tissue its characteristic biological function (Barsi et al., 2015). Although effector genes are critical for the fitness of an organism throughout its adult life, effector genes are surprisingly poorly characterized Barsi et al. (2015). For example, TATA-box containing promoters are preferentially found among the effector genes (Schug et al. (2005); Lenhard et al. (2012); FitzGerald et al. (2006); Carninci et al. (2006); Engström et al. (2007)), yet the reason for this has not been unclear. It was not known whether the TATA promoter is a typical characteristic of effector genes, whether it confers particular expression properties, and whether other promoter types may also be prevalent among effector genes. We found that while both the TATA and paused

genes are prevalent among the effector genes, the most tissue-specific genes are regulated by TATA promoters. We speculate that the presence of a labile nucleosome confer the TATA promoter more tissue-restricted expression.

The gap in knowledge of effector genes was likely for historical reasons. The main tool for the discovery of gene functions has been genetics. Unlike mutations in developmental genes, which often produce clear patterning defects, mutations in effector genes may result in phenotypes that are very subtle and harder to identify, let alone be classifiable into groups or pathways. With the advent of genomics techniques such as RNA-seq and ChIP-seq, it is now possible to analyze the expression of effector genes more systematically. Single-cell RNA-seq is an excellent way to more completely capture the heterogeneity of differentiated tissues (Wagner et al. (2016); Tanay & Regev (2017); Rosenberg et al. (2018); Klein et al. (2015); Satija et al. (2015); Macosko et al. (2015); Zheng et al. (2017); Shekhar et al. (2016); Zeisel et al. (2015); Grün et al. (2015); Villani et al. (2017); Marques et al. (2016); Halpern et al. (2017); La Manno et al. (2016)). This allows the identification of tissues and their expression profiles in an unbiased manner. In addition to providing expression differences between tissues for each gene, single-cell RNA-seq also allowed us to analyze the cell-to-cell variation within a tissue. Our analysis revealed that the TATA promoters tended to be expressed in a much more tissue-specific fashion with little background expression in other tissues, but their expression tended to be more variable than that of paused genes. These results provide the first insights into different groups of effector genes, which has implications for their function and evolution in metazoans.

In the early embryo, developmental genes have highly paused promoters where Pol II frequently pauses 30-50 bp downstream of the transcription start site (TSS) before being released into productive elongation. The developmental genes also tend to be expressed highly synchronously (Boettiger & Levine (2009)), and without the synchronous activation, coordinated cellular behaviors such as gastrulation may not proceed properly (Lagha et al. (2013)). Thus, having high levels of paused Pol II appears to be a mechanism that prepares genes for fast and synchronous activation during development. In contrast, genes with a TATA box tend to have transiently paused Pol II

(Chen et al. (2013), Gaertner et al. (2012), Shao & Zeitlinger (2017)). For example, many of the very first genes expressed in the zygotic genome have TATA elements, and Pol II ChIP-seq data suggest that Pol II does not pause during their initial transcription (Chen et al., 2013). While the benefits of a strongly paused promoter have been demonstrated, the benefits of a TATA promoter that mediate the expression of the very first genes without Pol II pausing is not clear. Likewise, it had not been clear unclear why tissue-specific effector genes often have TATA box-containing promoters. Our results provide some insights into this process. TATA-containing promoters might be utilized preferentially in very tissue-specific genes due to their less-leaky expression patterns. The noise expression profiles from the TATA promoters might also be beneficial in some cases such as in the immune system.

4.7 Why do some genes contain paused Pol II?

Classical studies of the heat shock genes and recent studies involving swapping of paused promoters with non-paused promoters suggest that paused Pol II significantly accelerates the timing of gene induction and synchronicity(Lagha et al. (2013), Wilkins & Lis (1997)). In the case of heat shock genes and the developmental genes, the importance of rapid or synchronous expression is evident. However, many of the targets of Lola-I are not the typical developmental genes. Our GO term enrichment analysis suggests that many of them are metabolic genes and might be required in all the cells to some extent. This goes well with the view that pausing allows fine-tuning of expression. The presence of paused Pol II could allow Lola-I targets to be expressed at a basal level throughout the embryo, and in specific tissues, the expression could be boosted by further activation signals. However, we cannot rule out the importance of precise regulation of the expression dynamics at these genes.

4.8 Concluding remarks

Identification of the three eukaryotic RNA Polymerases, fifty years ago, and their biochemical purification immediately afterward, marked a significant shift in the field of eukaryotic transcription. The numerous biochemical experiments that followed have characterized the many aspects of eukaryotic transcription in great detail. In addition to the biochemical experiments, the developments in molecular biology, genome-wide studies, and the recent advancements in the field of live imaging and single-cell technologies have also contributed to our understanding of eukaryotic transcription today. Similarly, our current understanding of paused Pol II is based on the insights provided by the convergence of these and other technologies.

Paused Pol II was discovered about forty years ago. The biochemical studies on paused Pol II helped to characterize the various aspects of Pol II pausing, including the factors involved in pausing and their roles. However, only with the genome-wide studies later, which showed the high prevalence of Pol II pausing across many genes in different organisms, the importance of Pol II pausing as a general regulatory mechanism was appreciated. The recent single-molecule and live imaging studies are not only helping us understand the importance of paused Pol II for gene expression, but also understand the transcriptional process in never-before details. In this study, we have combined the various genomics technologies to understand how Pol II recruitment to the paused promoters might be regulated and how the paused Pol II affects the transcriptional activation dynamics. In the future, the ability to image paused Pol II in live-cell, together with various transcriptional factors, pause-inducing factors, etc., could provide significant insights into the transcriptional process. Similarly, improvements in the single-molecule and single-cell studies could allow us to better correlate the binding of various factors across the genome with different transcriptional steps and infer about the transcriptional process. In addition, the vast amounts of data brought by these technologies have to be handled by effective machine learning strategies.

References

Adelman, K. & Lis, J. T. (2012). Promoter-proximal pausing of rna polymerase ii: emerging roles in metazoans. *Nature Reviews Genetics*, 13, 720 EP –.

Ahmad, K. F., Engel, C. K., & Privé, G. G. (1998). Crystal structure of the btb domain from plzf. *Proc Natl Acad Sci U S A*, 95(21), 12123–8.

Akalin, A., Fredman, D., Arner, E., Dong, X., Bryne, J. C., Suzuki, H., Daub, C. O., Hayashizaki, Y., & Lenhard, B. (2009). Transcriptional features of genomic regulatory blocks. *Genome Biol*, 10(4), R38.

Allison, L. A., Moyle, M., Shales, M., & Ingles, C. J. (1985). Extensive homology among the largest subunits of eukaryotic and prokaryotic rna polymerases. *Cell*, 42(2), 599–610.

Amleh, A., Nair, S. J., Sun, J., Sutherland, A., Hasty, P., & Li, R. (2009). Mouse cofactor of brca1 (cobra1) is required for early embryogenesis. *PLoS One*, 4(4), e5034.

Anderson, J. D. & Widom, J. (2000). Sequence and position-dependence of the equilibrium accessibility of nucleosomal dna target sites. *J Mol Biol*, 296(4), 979–87.

Arner, E., Daub, C. O., Vitting-Seerup, K., Andersson, R., Lilje, B., Drabløs, F., Lennartsson, A., Rönnerblad, M., Hrydziuszko, O., Vitezic, M., Freeman, T. C., Alhendi, A. M. N., Arner, P., Axton, R., Baillie, J. K., Beckhouse, A., Bodega, B., Briggs, J., Brombacher, F., Davis, M., Detmar, M., Ehrlund, A., Endoh, M., Eslami, A., Fagiolini, M., Fairbairn, L., Faulkner, G. J., Ferrai, C., Fisher, M. E., Forrester, L., Goldowitz, D., Guler, R., Ha, T., Hara, M., Herlyn, M., Ikawa, T., Kai, C., Kawamoto, H., Khachigian, L. M., Klinken, S. P., Kojima, S., Koseki, H., Klein, S., Mejhert, N., Miyaguchi, K., Mizuno, Y., Morimoto, M., Morris, K. J., Mummery, C., Nakachi, Y., Ogishima, S., Okada-Hatakeyama, M., Okazaki, Y., Orlando, V., Ovchinnikov, D., Passier, R., Patrikakis, M., Pombo, A., Qin, X.-Y., Roy, S., Sato, H., Savvi, S., Saxena, A., Schwegmann, A., Sugiyama, D., Swoboda, R., Tanaka, H., Tomoiu, A., Winteringham, L. N.,

Wolvetang, E., Yanagi-Mizuochi, C., Yoneda, M., Zabierowski, S., Zhang, P., Abugessaisa, I., Bertin, N., Diehl, A. D., Fukuda, S., Furuno, M., Harshbarger, J., Hasegawa, A., Hori, F., Ishikawa-Kato, S., Ishizu, Y., Itoh, M., Kawashima, T., Kojima, M., Kondo, N., Lizio, M., Meehan, T. F., Mungall, C. J., Murata, M., Nishiyori-Sueki, H., Sahin, S., Nagao-Sato, S., Severin, J., de Hoon, M. J. L., Kawai, J., Kasukawa, T., Lassmann, T., Suzuki, H., Kawaji, H., Summers, K. M., Wells, C., FANTOM Consortium, Hume, D. A., Forrest, A. R. R., Sandelin, A., Carninci, P., & Hayashizaki, Y. (2015). Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science*, 347(6225), 1010–4.

Arnold, C. D., Gerlach, D., Stelzer, C., Boryń, Ł. M., Rath, M., & Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by starr-seq. *Science*, 339(6123), 1074–1077.

Banerji, J., Rusconi, S., & Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote sv40 dna sequences. *Cell*, 27(2, Part 1), 299 – 308.

Barberis, A., Müller, C. W., Harrison, S. C., & Ptashne, M. (1993). Delineation of two functional regions of transcription factor tfiib. *Proc Natl Acad Sci U S A*, 90(12), 5628–32.

Bardwell, V. J. & Treisman, R. (1994). The poz domain: a conserved protein-protein interaction motif. *Genes Dev*, 8(14), 1664–77.

Barsi, J. C., Tu, Q., Calestani, C., & Davidson, E. H. (2015). Genome-wide assessment of differential effector gene use in embryogenesis. *Development*, 142(22), 3892–901.

Bartman, C. R., Hsu, S. C., Hsiung, C. C.-S., Raj, A., & Blobel, G. A. (2016). Enhancer regulation of transcriptional bursting parameters revealed by forced chromatin looping. *Mol Cell*, 62(2), 237–247.

Basehoar, A. D., Zanton, S. J., & Pugh, B. F. (2004). Identification and distinct regulation of yeast tata box-containing genes. *Cell*, 116(5), 699–709.

Bass, B. P., Cullen, K., & McCall, K. (2007). The axon guidance gene lola is required for programmed cell death in the drosophila ovary. *Dev Biol*, 304(2), 771–85.

Bentley, D. L. (2014). Coupling mrna processing with transcription in time and space. *Nat Rev Genet*, 15(3), 163–75.

Bernstein, B. E., Liu, C. L., Humphrey, E. L., Perlstein, E. O., & Schreiber, S. L. (2004). Global nucleosome occupancy in yeast. *Genome Biol*, 5(9), R62.

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S. L., & Lander, E. S. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, 125(2), 315–26.

Berroteran, R. W., Ware, D. E., & Hampsey, M. (1994). The sua8 suppressors of saccharomyces cerevisiae encode replacements of conserved residues within the largest subunit of rna polymerase ii and affect transcription start site selection similarly to sua7 (tfiib) mutations. *Mol Cell Biol*, 14(1), 226–37.

Bintu, L., Ishibashi, T., Dangkulwanich, M., Wu, Y.-Y., Lubkowska, L., Kashlev, M., & Bustamante, C. (2012). Nucleosomal elements that control the topography of the barrier to transcription. *Cell*, 151(4), 738–749.

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. *Nature Structural & Molecular Biology*, 25(9), 833–840.

Boettiger, A. N. (2013). Analytic approaches to stochastic gene expression in multicellular systems. *Biophys J*, 105(12), 2629–40.

Boettiger, A. N. & Levine, M. (2009). Synchronous and stochastic patterns of gene activation in the drosophila embryo. *Science*, 325(5939), 471–3.

Boettiger, A. N., Ralph, P. L., & Evans, S. N. (2011). Transcriptional regulation: effects of promoter proximal pausing on speed, synchrony and reliability. *PLoS Comput Biol*, 7(5), e1001136.

Boija, A., Mahat, D. B., Zare, A., Holmqvist, P.-H., Philip, P., Meyers, D. J., Cole, P. A., Lis, J. T., Stenberg, P., & Mannervik, M. (2017). Cbp regulates recruitment and release of promoter-proximal rna polymerase ii. *Mol Cell*, 68(3), 491–503.e5.

Bonchuk, A., Denisov, S., Georgiev, P., & Maksimenko, O. (2011). Drosophila btb/poz domains of "ttk group" can form multimers and selectively interact with each other. *J Mol Biol*, 412(3), 423–36.

Bonn, S., Zinzen, R. P., Girardot, C., Gustafson, E. H., Perez-Gonzalez, A., Delhomme, N., Ghavi-Helm, Y., Wilczyński, B., Riddell, A., & Furlong, E. E. M. (2012a). Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat Genet*, 44(2), 148–56.

Bonn, S., Zinzen, R. P., Perez-Gonzalez, A., Riddell, A., Gavin, A.-C., & Furlong, E. E. M. (2012b). Cell type-specific chromatin immunoprecipitation from multicellular complex samples using bits-chip. *Nat Protoc*, 7(5), 978–94.

Bortvin, A. & Winston, F. (1996). Evidence that spt6p controls chromatin structure by a direct interaction with histones. *Science*, 272(5267), 1473–6.

Bossard, P. & Zaret, K. S. (1998). Gata transcription factors as potentiators of gut endoderm differentiation. *Development*, 125(24), 4909–17.

Boyle, A. P., Davis, S., Shulha, H. P., Meltzer, P., Margulies, E. H., Weng, Z., Furey, T. S., & Crawford, G. E. (2008). High-resolution mapping and characterization of open chromatin across the genome. *Cell*, 132(2), 311–22.

Bradsher, J. N., Jackson, K. W., Conaway, R. C., & Conaway, J. W. (1993a). Rna polymerase ii transcription factor siii. i. identification, purification, and properties. *J Biol Chem*, 268(34), 25587–93.

Bradsher, J. N., Tan, S., McLaury, H. J., Conaway, J. W., & Conaway, R. C. (1993b). Rna polymerase ii transcription factor siii. ii. functional properties and role in rna chain elongation. *J Biol Chem*, 268(34), 25594–603.

Brahma, S. & Henikoff, S. (2019). Rsc-associated subnucleosomes define mnase-sensitive promoters in yeast. *Mol Cell*, 73(2), 238–249.e3.

Brangwynne, C. P., Eckmann, C. R., Courson, D. S., Rybarska, A., Hoege, C., Gharakhani, J., Jülicher, F., & Hyman, A. A. (2009). Germline p granules are liquid droplets that localize by controlled dissolution/condensation. *Science*, 324(5935), 1729–32.

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic rna-seq quantification. *Nat Biotechnol*, 34(5), 525–7.

Brookes, E., de Santiago, I., Hebenstreit, D., Morris, K. J., Carroll, T., Xie, S. Q., Stock, J. K., Heidemann, M., Eick, D., Nozaki, N., Kimura, H., Ragoussis, J., Teichmann, S. A., & Pombo, A. (2012). Polycomb associates genome-wide with a specific rna polymerase ii variant, and regulates metabolic genes in escs. *Cell Stem Cell*, 10(2), 157–70.

Browning, D. F. & Busby, S. J. W. (2004). The regulation of bacterial transcription initiation. *Nature Reviews Microbiology*, 2(1), 57–65.

Budry, L., Balsalobre, A., Gauthier, Y., Khetchoumian, K., L'honoré, A., Vallette, S., Brue, T., Figarella-Branger, D., Meij, B., & Drouin, J. (2012). The selector gene pax7 dictates alternate pituitary cell fates through its pioneer action on chromatin remodeling. *Genes Dev*, 26(20), 2299–310.

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, dna-binding proteins and nucleosome position. *Nat Methods*, 10(12), 1213–8.

Buratowski, S., Hahn, S., Guarente, L., & Sharp, P. A. (1989). Five intermediate complexes in transcription initiation by rna polymerase ii. *Cell*, 56(4), 549–61.

Buratowski, S. & Sharp, P. A. (1990). Transcription initiation complexes and upstream activation with rna polymerase ii lacking the c-terminal domain of the largest subunit. *Molecular and Cellular Biology*, 10(10), 5562–5564.

Buratowski, S. & Zhou, H. (1993). Functional domains of transcription factor tfiib. *Proc Natl Acad Sci U S A*, 90(12), 5633–7.

Burke, T. W. & Kadonaga, J. T. (1996). Drosophila tfiid binds to a conserved downstream basal promoter element that is present in many tata-box-deficient promoters. *Genes Dev*, 10(6), 711–24.

Burke, T. W. & Kadonaga, J. T. (1997). The downstream core promoter element, dpe, is conserved from drosophila to humans and is recognized by tafii60 of drosophila. *Genes Dev*, 11(22), 3020–31.

Byers, S. A., Price, J. P., Cooper, J. J., Li, Q., & Price, D. H. (2005). Hexim2, a hexim1-related protein, regulates positive transcription elongation factor b through association with 7sk. *J Biol Chem*, 280(16), 16360–7.

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(6), 626–35.

Casas-Vila, N., Bluhm, A., Sayols, S., Dinges, N., Dejung, M., Altenhein, T., Kappei, D., Altenhein, B., Roignant, J.-Y., & Butter, F. (2017). The developmental proteome of drosophila melanogaster. *Genome Res*, 27(7), 1273–1285.

Catarino, R. R., Neumayr, C., & Stark, A. (2017). Promoting transcription over long distances. *Nature Genetics*, 49, 972 EP –.

Cavarec, L., Jensen, S., Casella, J. F., Cristescu, S. A., & Heidmann, T. (1997). Molecular cloning and characterization of a transcription factor for the copia retrotransposon with homology to the btb-containing lola neurogenic factor. *Mol Cell Biol*, 17(1), 482–94.

Chalkley, G. E. & Verrijzer, C. P. (1999). Dna binding site selection by rna polymerase ii tafs: a taf(ii)250-taf(ii)150 complex recognizes the initiator. *EMBO J*, 18(17), 4835–45.

Chapman, R. D., Heidemann, M., Hintermair, C., & Eick, D. (2008). Molecular evolution of the rna polymerase ii ctd. *Trends Genet*, 24(6), 289–96.

Chasman, D. I., Flaherty, K. M., Sharp, P. A., & Kornberg, R. D. (1993). Crystal structure of yeast tata-binding protein and model for interaction with dna. *Proc Natl Acad Sci U S A*, 90(17), 8174–8.

Chen, F., Gao, X., & Shilatifard, A. (2015a). Stably paused genes revealed through inhibition of transcription initiation by the tfiih inhibitor triptolide. *Genes Dev*, 29(1), 39–47.

Chen, F. X., Woodfin, A. R., Gardini, A., Rickels, R. A., Marshall, S. A., Smith, E. R., Shiekhattar, R., & Shilatifard, A. (2015b). Paf1, a molecular regulator of promoter-proximal pausing by rna polymerase ii. *Cell*, 162(5), 1003 – 1015.

Chen, F. X., Woodfin, A. R., Gardini, A., Rickels, R. A., Marshall, S. A., Smith, E. R., Shiekhattar, R., & Shilatifard, A. (2015c). Paf1, a molecular regulator of promoter-proximal pausing by rna polymerase ii. *Cell*, 162(5), 1003–15.

Chen, F. X., Xie, P., Collings, C. K., Cao, K., Aoi, Y., Marshall, S. A., Rendleman, E. J., Ugarenko, M., Ozark, P. A., Zhang, A., Shiekhattar, R., Smith, E. R., Zhang, M. Q., & Shilatifard, A. (2017). Paf1 regulation of promoter-proximal pause release via enhancer activation. *Science*, 357(6357), 1294–1298.

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.

Cheng, B., Li, T., Rahl, P. B., Adamson, T. E., Loudas, N. B., Guo, J., Varzavand, K., Cooper, J. J., Hu, X., Gnatt, A., Young, R. A., & Price, D. H. (2012). Functional association of gdown1 with rna polymerase ii poised on human genes. *Mol Cell*, 45(1), 38–50.

Chereji, R. V., Ocampo, J., & Clark, D. J. (2017). Mnase-sensitive complexes in yeast: Nucleosomes and non-histone barriers. *Mol Cell*, 65(3), 565–577.e3.

Cho, W.-K., Spille, J.-H., Hecht, M., Lee, C., Li, C., Grube, V., & Cisse, I. I. (2018). Mediator and rna polymerase ii clusters associate in transcription-dependent condensates. *Science*, 361(6400), 412–415.

Choi, W. S., Yan, M., Nusinow, D., & Gralla, J. D. (2002). In vitro transcription and start site selection in schizosaccharomyces pombe. *J Mol Biol*, 319(5), 1005–13.

Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G. M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X., & Tjian, R. (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science*, 361(6400).

Choukrallah, M.-A., Song, S., Rolink, A. G., Burger, L., & Matthias, P. (2015). Enhancer reper-

toires are reshaped independently of early priming and heterochromatin dynamics during b cell differentiation. *Nat Commun*, 6, 8324.

Chung, W.-H., Craighead, J. L., Chang, W.-H., Ezeokonkwo, C., Bareket-Samish, A., Kornberg, R. D., & Asturias, F. J. (2003). Rna polymerase ii/tfiif structure and conserved organization of the initiation complex. *Mol Cell*, 12(4), 1003–13.

Churchman, L. S. & Weissman, J. S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature*, 469(7330), 368–73.

Cirillo, L. A., Lin, F. R., Cuesta, I., Friedman, D., Jarnik, M., & Zaret, K. S. (2002). Opening of compacted chromatin by early developmental transcription factors hnf3 (foxa) and gata-4. *Mol Cell*, 9(2), 279–89.

Cirillo, L. A., McPherson, C. E., Bossard, P., Stevens, K., Cherian, S., Shim, E. Y., Clark, K. L., Burley, S. K., & Zaret, K. S. (1998). Binding of the winged-helix transcription factor hnf3 to a linker histone site on the nucleosome. *EMBO J*, 17(1), 244–54.

Cirillo, L. A. & Zaret, K. S. (1999). An early developmental transcription factor complex that is more stable on nucleosome core particles than on free dna. *Mol Cell*, 4(6), 961–9.

Coleman, R. A., Taggart, A. K., Burma, S., Chicca, 2nd, J. J., & Pugh, B. F. (1999). Tfiia regulates tbp and tfiid dimers. *Mol Cell*, 4(3), 451–7.

Conaway, J. W., Bradsher, J. N., Tan, S., & Conaway, R. C. (1993). Transcription factor siii: a novel component of the rna polymerase ii elongation complex. *Cell Mol Biol Res*, 39(4), 323–9.

Conaway, J. W. & Conaway, R. C. (1989a). A multisubunit transcription factor essential for accurate initiation by rna polymerase ii. *J Biol Chem*, 264(4), 2357–62.

Conaway, R. C. & Conaway, J. W. (1989b). An rna polymerase ii transcription factor has an associated dna-dependent atpase (datpase) activity strongly stimulated by the tata region of promoters. *Proc Natl Acad Sci U S A*, 86(19), 7356–60.

Conaway, R. C. & Conaway, J. W. (1993). General initiation factors for rna polymerase ii. *Annu Rev Biochem*, 62, 161–90.

Corden, J. L. (1990). Tails of rna polymerase ii. *Trends Biochem Sci*, 15(10), 383–7.

Corden, J. L., Cadena, D. L., Ahearn, Jr, J. M., & Dahmus, M. E. (1985). A unique structure at the carboxyl terminus of the largest subunit of eukaryotic rna polymerase ii. *Proc Natl Acad Sci U S A*, 82(23), 7934–8.

Core, L. J., Martins, A., Danko, C. G., Waters, C. T., Siepel, A., & Lis, J. T. (2014a). Analysis of nascent rna identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nature Genetics*, 46, 1311 EP –.

Core, L. J., Martins, A. L., Danko, C. G., Waters, C. T., Siepel, A., & Lis, J. T. (2014b). Analysis of nascent rna identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet*, 46(12), 1311–20.

Core, L. J., Waterfall, J. J., & Lis, J. T. (2008). Nascent rna sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*, 322(5909), 1845–8.

Cornman, R. S. (2009). Molecular evolution of drosophila cuticular protein genes. *PLoS One*, 4(12), e8345.

Cortes, P., Flores, O., & Reinberg, D. (1992). Factors involved in specific transcription by mammalian rna polymerase ii: purification and analysis of transcription factor iia and identification of transcription factor iij. *Mol Cell Biol*, 12(1), 413–21.

Coulombe, B., Killeen, M., Liljelund, P., Honda, B., Xiao, H., Ingles, C. J., & Greenblatt, J. (1992). Identification of three mammalian proteins that bind to the yeast tata box protein tfiid. *Gene Expr*, 2(2), 99–110.

Cramer, P., Bushnell, D. A., & Kornberg, R. D. (2001). Structural basis of transcription: Rna polymerase ii at 2.8 angstrom resolution. *Science*, 292(5523), 1863–76.

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., Boyer, L. A., Young, R. A., & Jaenisch, R. (2010). Histone h3k27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A*, 107(50), 21931–6.

Dao, L. T. M., Galindo-Albarrán, A. O., Castro-Mondragon, J. A., Andrieu-Soler, C., Medina-Rivera, A., Souaid, C., Charbonnier, G., Griffon, A., Vanhille, L., Stephen, T., Alomairi, J., Martin, D., Torres, M., Fernandez, N., Soler, E., van Helden, J., Puthier, D., & Spicuglia, S. (2017). Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nature Genetics*, 49, 1073 EP –.

Darzacq, X., Yao, J., Larson, D. R., Causse, S. Z., Bosanac, L., de Turris, V., Ruda, V. M., Lionnet, T., Zenklusen, D., Guglielmi, B., Tjian, R., & Singer, R. H. (2009). Imaging transcription in living cells. *Annu Rev Biophys*, 38, 173–96.

Day, D. S., Zhang, B., Stevens, S. M., Ferrari, F., Larschan, E. N., Park, P. J., & Pu, W. T. (2016). Comprehensive analysis of promoter-proximal rna polymerase ii pausing across mammalian cell types. *Genome Biol*, 17(1), 120.

De Kumar, B., Parker, H. J., Parrish, M. E., Lange, J. J., Slaughter, B. D., Unruh, J. R., Paulson, A., & Krumlauf, R. (2017). Dynamic regulation of nanog and stem cell-signaling pathways by hoxa1 during early neuro-ectodermal differentiation of es cells. *Proc Natl Acad Sci U S A*, 114(23), 5838–5845.

Deal, R. B. & Henikoff, S. (2011). The intact method for cell type-specific gene expression and chromatin profiling in arabidopsis thaliana. *Nat Protoc*, 6(1), 56–68.

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(1), 96 – 105.

Deweindt, C., Albagli, O., Bernardin, F., Dhordain, P., Quief, S., Lantoine, D., Kerckaert, J. P., & Leprince, D. (1995). The laz3/bcl6 oncogene encodes a sequence-specific transcriptional

inhibitor: a novel function for the btb/poz domain as an autonomous repressing domain. *Cell Growth Differ*, 6(12), 1495–503.

Diao, Y., Fang, R., Li, B., Meng, Z., Yu, J., Qiu, Y., Lin, K. C., Huang, H., Liu, T., Marina, R. J., Jung, I., Shen, Y., Guan, K.-L., & Ren, B. (2017). A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nature Methods*, 14, 629 EP –.

Dorer, D. R., Rudnick, J. A., Moriyama, E. N., & Christensen, A. C. (2003). A family of genes clustered at the triplo-lethal locus of drosophila melanogaster has an unusual evolutionary history and significant synteny with anopheles gambiae. *Genetics*, 165(2), 613–21.

Dorighi, K. M., Swigut, T., Henriques, T., Bhanu, N. V., Scruggs, B. S., Nady, N., Still, 2nd, C. D., Garcia, B. A., Adelman, K., & Wysocka, J. (2017). Mll3 and mll4 facilitate enhancer rna synthesis and transcription from promoters independently of h3k4 monomethylation. *Mol Cell*, 66(4), 568–576.e4.

Dufourt, J., Trullo, A., Hunter, J., Fernandez, C., Lazaro, J., Dejean, M., Morales, L., Nait-Amer, S., Schulz, K. N., Harrison, M. M., Favard, C., Radulescu, O., & Lagha, M. (2018). Temporal control of gene expression by the pioneer factor zelda through transient interactions in hubs. *Nat Commun*, 9(1), 5194.

Dvir, A. (2002). Promoter escape by rna polymerase ii. *Biochim Biophys Acta*, 1577(2), 208–223.

Dvir, A., Conaway, R. C., & Conaway, J. W. (1997). A role for tfiih in controlling the activity of early rna polymerase ii elongation complexes. *Proc Natl Acad Sci U S A*, 94(17), 9006–10.

Dvir, A., Garrett, K. P., Chalut, C., Egly, J. M., Conaway, J. W., & Conaway, R. C. (1996). A role for atp and tfiih in activation of the rna polymerase ii preinitiation complex prior to transcription initiation. *J Biol Chem*, 271(13), 7245–8.

Dyer, P. N., Edayathumangalam, R. S., White, C. L., Bao, Y., Chakravarthy, S., Muthurajan, U. M.,

& Luger, K. (2004). Reconstitution of nucleosome core particles from recombinant histones and dna. *Methods Enzymol*, 375, 23–44.

Ebright, R. H. (2000). Rna polymerase: structural similarities between bacterial rna polymerase and eukaryotic rna polymerase ii. *J Mol Biol*, 304(5), 687–98.

Egyházi, E., Ossoinak, A., Pigon, A., Holmgren, C., Lee, J. M., & Greenleaf, A. L. (1996). Phosphorylation dependence of the initiation of productive transcription of balbiani ring 2 genes in living cells. *Chromosoma*, 104(6), 422–33.

Elgin, S. C. (1996). Heterochromatin and gene regulation in drosophila. *Curr Opin Genet Dev*, 6(2), 193–202.

Emami, K. H., Jain, A., & Smale, S. T. (1997). Mechanism of synergy between tata and initiator: synergistic binding of tfiid following a putative tfiia-induced isomerization. *Genes Dev*, 11(22), 3007–19.

ENCODE Project Consortium, Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigó, R., Gingeras, T. R., Margulies, E. H., Weng, Z., Snyder, M., Dermitzakis, E. T., Thurman, R. E., Kuehn, M. S., Taylor, C. M., Neph, S., Koch, C. M., Asthana, S., Malhotra, A., Adzhubei, I., Greenbaum, J. A., Andrews, R. M., Flicek, P., Boyle, P. J., Cao, H., Carter, N. P., Clelland, G. K., Davis, S., Day, N., Dhami, P., Dillon, S. C., Dorschner, M. O., Fiegler, H., Giresi, P. G., Goldy, J., Hawrylycz, M., Haydock, A., Humbert, R., James, K. D., Johnson, B. E., Johnson, E. M., Frum, T. T., Rosenzweig, E. R., Karnani, N., Lee, K., Lefebvre, G. C., Navas, P. A., Neri, F., Parker, S. C. J., Sabo, P. J., Sandstrom, R., Shafer, A., Vetrie, D., Weaver, M., Wilcox, S., Yu, M., Collins, F. S., Dekker, J., Lieb, J. D., Tullius, T. D., Crawford, G. E., Sunyaev, S., Noble, W. S., Dunham, I., Denoeud, F., Reymond, A., Kapranov, P., Rozowsky, J., Zheng, D., Castelo, R., Frankish, A., Harrow, J., Ghosh, S., Sandelin, A., Hofacker, I. L., Baertsch, R., Keefe, D., Dike, S., Cheng, J., Hirsch, H. A., Sekinger, E. A., Lagarde, J., Abril, J. F., Shahab, A., Flamm, C., Fried, C., Hackermüller, J., Hertel, J., Lindemeyer, M., Missal, K., Tanzer, A., Washietl, S.,

Korbel, J., Emanuelsson, O., Pedersen, J. S., Holroyd, N., Taylor, R., Swarbreck, D., Matthews, N., Dickson, M. C., Thomas, D. J., Weirauch, M. T., Gilbert, J., Drenkow, J., Bell, I., Zhao, X., Srinivasan, K. G., Sung, W.-K., Ooi, H. S., Chiu, K. P., Foissac, S., Alioto, T., Brent, M., Pachter, L., Tress, M. L., Valencia, A., Choo, S. W., Choo, C. Y., Ucla, C., Manzano, C., Wyss, C., Cheung, E., Clark, T. G., Brown, J. B., Ganesh, M., Patel, S., Tammana, H., Chrast, J., Henrichsen, C. N., Kai, C., Kawai, J., Nagalakshmi, U., Wu, J., Lian, Z., Lian, J., Newburger, P., Zhang, X., Bickel, P., Mattick, J. S., Carninci, P., Hayashizaki, Y., Weissman, S., Hubbard, T., Myers, R. M., Rogers, J., Stadler, P. F., Lowe, T. M., Wei, C.-L., Ruan, Y., Struhl, K., Gerstein, M., Antonarakis, S. E., Fu, Y., Green, E. D., Karaöz, U., Siepel, A., Taylor, J., Liefer, L. A., Wetterstrand, K. A., Good, P. J., Feingold, E. A., Guyer, M. S., Cooper, G. M., Asimenos, G., Dewey, C. N., Hou, M., Nikolaev, S., Montoya-Burgos, J. I., Löytynoja, A., Whelan, S., Pardi, F., Massingham, T., Huang, H., Zhang, N. R., Holmes, I., Mullikin, J. C., Ureta-Vidal, A., Paten, B., Seringhaus, M., Church, D., Rosenbloom, K., Kent, W. J., Stone, E. A., NISC Comparative Sequencing Program, Baylor College of Medicine Human Genome Sequencing Center, Washington University Genome Sequencing Center, Broad Institute, Children's Hospital Oakland Research Institute, Batzoglou, S., Goldman, N., Hardison, R. C., Haussler, D., Miller, W., Sidow, A., Trinklein, N. D., Zhang, Z. D., Barrera, L., Stuart, R., King, D. C., Ameur, A., Enroth, S., Bieda, M. C., Kim, J., Bhinge, A. A., Jiang, N., Liu, J., Yao, F., Vega, V. B., Lee, C. W. H., Ng, P., Shahab, A., Yang, A., Moqtaderi, Z., Zhu, Z., Xu, X., Squazzo, S., Oberley, M. J., Inman, D., Singer, M. A., Richmond, T. A., Munn, K. J., Rada-Iglesias, A., Wallerman, O., Komorowski, J., Fowler, J. C., Couttet, P., Bruce, A. W., Dovey, O. M., Ellis, P. D., Langford, C. F., Nix, D. A., Euskirchen, G., Hartman, S., Urban, A. E., Kraus, P., Van Calcar, S., Heintzman, N., Kim, T. H., Wang, K., Qu, C., Hon, G., Luna, R., Glass, C. K., Rosenfeld, M. G., Aldred, S. F., Cooper, S. J., Halees, A., Lin, J. M., Shulha, H. P., Zhang, X., Xu, M., Haidar, J. N. S., Yu, Y., Ruan, Y., Iyer, V. R., Green, R. D., Wadelius, C., Farnham, P. J., Ren, B., Harte, R. A., Hinrichs, A. S., Trumbower, H., Clawson, H., Hillman-Jackson, J., Zweig, A. S., Smith, K., Thakkapallayil, A., Barber, G., Kuhn, R. M., Karolchik, D., Armengol, L., Bird, C. P., de Bakker, P. I. W.,

Kern, A. D., Lopez-Bigas, N., Martin, J. D., Stranger, B. E., Woodroffe, A., Davydov, E., Dimas, A., Eyras, E., Hallgrímsdóttir, I. B., Huppert, J., Zody, M. C., Abecasis, G. R., Estivill, X., Bouffard, G. G., Guan, X., Hansen, N. F., Idol, J. R., Maduro, V. V. B., Maskeri, B., McDowell, J. C., Park, M., Thomas, P. J., Young, A. C., Blakesley, R. W., Muzny, D. M., Sodergren, E., Wheeler, D. A., Worley, K. C., Jiang, H., Weinstock, G. M., Gibbs, R. A., Graves, T., Fulton, R., Mardis, E. R., Wilson, R. K., Clamp, M., Cuff, J., Gnerre, S., Jaffe, D. B., Chang, J. L., Lindblad-Toh, K., Lander, E. S., Koriabine, M., Nefedov, M., Osoegawa, K., Yoshinaga, Y., Zhu, B., & de Jong, P. J. (2007). Identification and analysis of functional elements in 1% of the human genome by the encode pilot project. *Nature*, 447(7146), 799–816.

Engström, P. G., Ho Sui, S. J., Drivenes, O., Becker, T. S., & Lenhard, B. (2007). Genomic regulatory blocks underlie extensive microsynteny conservation in insects. *Genome Res*, 17(12), 1898–908.

Enuameh, M. S., Asriyan, Y., Richards, A., Christensen, R. G., Hall, V. L., Kazemian, M., Zhu, C., Pham, H., Cheng, Q., Blatti, C., Brasefield, J. A., Basciotta, M. D., Ou, J., McNulty, J. C., Zhu, L. J., Celniker, S. E., Sinha, S., Stormo, G. D., Brodsky, M. H., & Wolfe, S. A. (2013). Global analysis of drosophila cys2-his2 zinc finger proteins reveals a multitude of novel recognition motifs and binding determinants. *Genome Res*, 23(6), 928–40.

Espinás, M. L., Jiménez-García, E., Vaquero, A., Canudas, S., Bernués, J., & Azorín, F. (1999). The n-terminal poz domain of gaga mediates the formation of oligomers that bind dna with high affinity and specificity. *J Biol Chem*, 274(23), 16461–9.

Fakhouri, T. H. I., Stevenson, J., Chisholm, A. D., & Mango, S. E. (2010). Dynamic chromatin organization during foregut development mediated by the organ selector gene pha-4/foxa. *PLoS Genet*, 6(8).

Fang, H., Wei, S., Lee, T.-H., & Hayes, J. J. (2016). Chromatin structure-dependent conformations of the h1 ctd. *Nucleic Acids Res*, 44(19), 9131–9141.

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., 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(7493), 462–70.

Feaver, W. J., Gileadi, O., & Kornberg, R. D. (1991). Purification and characterization of yeast rna polymerase ii transcription factor b. *J Biol Chem*, 266(28), 19000–5.

Ferres-Marco, D., Gutierrez-Garcia, I., Vallejo, D. M., Bolivar, J., Gutierrez-Aviño, F. J., & Dominguez, M. (2006). Epigenetic silencers and notch collaborate to promote malignant tumours by rb silencing. *Nature*, 439(7075), 430–6.

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(11), e1000216.

Filion, G. J., van Bemmel, J. G., Braunschweig, U., Talhout, W., Kind, J., Ward, L. D., Brugman, W., de Castro, I. J., Kerkhoven, R. M., Bussemaker, H. J., & van Steensel, B. (2010). Systematic protein location mapping reveals five principal chromatin types in drosophila cells. *Cell*, 143(2), 212–24.

Fishburn, J., Tomko, E., Galburt, E., & Hahn, S. (2015). Double-stranded dna translocase activity of transcription factor tfiih and the mechanism of rna polymerase ii open complex formation. *Proc Natl Acad Sci U S A*, 112(13), 3961–6.

FitzGerald, P. C., Sturgill, D., Shyakhtenko, A., Oliver, B., & Vinson, C. (2006). Comparative genomics of drosophila and human core promoters. *Genome Biol*, 7(7), R53.

Flores, O., Lu, H., Killeen, M., Greenblatt, J., Burton, Z. F., & Reinberg, D. (1991). The small subunit of transcription factor iif recruits rna polymerase ii into the preinitiation complex. *Proc Natl Acad Sci U S A*, 88(22), 9999–10003.

Flores, O., Lu, H., & Reinberg, D. (1992). Factors involved in specific transcription by mammalian rna polymerase ii. identification and characterization of factor iih. *J Biol Chem*, 267(4), 2786–93.

Flores, O., Maldonado, E., & Reinberg, D. (1989). Factors involved in specific transcription by mammalian rna polymerase ii. factors iie and iif independently interact with rna polymerase ii. *J Biol Chem*, 264(15), 8913–21.

Fortini, M. E. & Rubin, G. M. (1990). Analysis of cis-acting requirements of the rh3 and rh4 genes reveals a bipartite organization to rhodopsin promoters in drosophila melanogaster. *Genes Dev*, 4(3), 444–63.

Fraser, N. W., Sehgal, P. B., & Darnell, J. E. (1978). Drb-induced premature termination of late adenovirus transcription. *Nature*, 272(5654), 590–593.

Fu, T. J., Peng, J., Lee, G., Price, D. H., & Flores, O. (1999). Cyclin k functions as a cdk9 regulatory subunit and participates in rna polymerase ii transcription. *J Biol Chem*, 274(49), 34527–30.

Fuda, N. J., Guertin, M. J., Sharma, S., Danko, C. G., Martins, A. L., Siepel, A., & Lis, J. T. (2015).

Gaga factor maintains nucleosome-free regions and has a role in rna polymerase ii recruitment to promoters. *PLoS Genet*, 11(3), e1005108.

Fujinaga, K., Irwin, D., Huang, Y., Taube, R., Kurosu, T., & Peterlin, B. M. (2004). Dynamics of human immunodeficiency virus transcription: P-tefb phosphorylates rd and dissociates negative effectors from the transactivation response element. *Mol Cell Biol*, 24(2), 787–95.

Fukaya, T., Lim, B., & Levine, M. (2016). Enhancer control of transcriptional bursting. *Cell*, 166(2), 358 – 368.

Furukawa, M., He, Y. J., Borchers, C., & Xiong, Y. (2003). Targeting of protein ubiquitination by btb-cullin 3-roc1 ubiquitin ligases. *Nat Cell Biol*, 5(11), 1001–7.

Gaertner, B., Johnston, J., Chen, K., Wallaschek, N., Paulson, A., Garruss, A. S., Gaudenz, K., De Kumar, B., Krumlauf, R., & Zeitlinger, J. (2012). Poised rna polymerase ii changes over developmental time and prepares genes for future expression. *Cell Rep*, 2(6), 1670–83.

Gariglio, P., Bellard, M., & Chambon, P. (1981a). Clustering of rna polymerase b molecules in the 5' moiety of the adult beta-globin gene of hen erythrocytes. *Nucleic Acids Res*, 9(11), 2589–98.

Gariglio, P., Bellard, M., & Chambon, P. (1981b). Clustering of RNA polymerase B molecules in the 5 prime moiety of the adult beta-globin gene of hen erythrocytes. *Nucleic Acids Research*, 9(11), 2589–2598.

Garriga, J. & Graña, X. (2004). Cellular control of gene expression by t-type cyclin/cdk9 complexes. *Gene*, 337, 15–23.

Gaudet, J. & Mango, S. E. (2002). Regulation of organogenesis by the caenorhabditis elegans foxa protein pha-4. *Science*, 295(5556), 821–5.

Geiger, J. H., Hahn, S., Lee, S., & Sigler, P. B. (1996). Crystal structure of the yeast tfiia/tbp/dna complex. *Science*, 272(5263), 830–6.

Gerard, M., Fischer, L., Moncollin, V., Chipoulet, J. M., Chambon, P., & Egly, J. M. (1991). Purification and interaction properties of the human rna polymerase b(ii) general transcription factor btf2. *J Biol Chem*, 266(31), 20940–5.

Gerber, H.-P., Hagmann, M., Seipel, K., Georgiev, O., West, M. A. L., Litingtung, Y., Schaffner, W., & Corden, J. L. (1995). Rna polymerase ii c-terminal domain required for enhancer-driven transcription. *Nature*, 374(6523), 660–662.

Geyer, R., Wee, S., Anderson, S., Yates, J., & Wolf, D. A. (2003). Btb/poz domain proteins are putative substrate adaptors for cullin 3 ubiquitin ligases. *Mol Cell*, 12(3), 783–90.

Ghavi-Helm, Y., Klein, F. A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., & Furlong, E. E. M. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. *Nature*, 512(7512), 96–100.

Giardina, C., Pérez-Riba, M., & Lis, J. T. (1992). Promoter melting and tfiid complexes on drosophila genes in vivo. *Genes Dev*, 6(11), 2190–200.

Gilchrist, D. A., Dos Santos, G., Fargo, D. C., Xie, B., Gao, Y., Li, L., & Adelman, K. (2010). Pausing of rna polymerase ii disrupts dna-specified nucleosome organization to enable precise gene regulation. *Cell*, 143(4), 540–51.

Gilchrist, D. A., Fromm, G., dos Santos, G., Pham, L. N., McDaniel, I. E., Burkholder, A., Fargo, D. C., & Adelman, K. (2012). Regulating the regulators: the pervasive effects of pol ii pausing on stimulus-responsive gene networks. *Genes Dev*, 26(9), 933–44.

Gilmour, D. S. & Lis, J. T. (1986). Rna polymerase ii interacts with the promoter region of the noninduced hsp70 gene in drosophila melanogaster cells. *Mol Cell Biol*, 6(11), 3984–9.

Giniger, E., Tietje, K., Jan, L. Y., & Jan, Y. N. (1994). lola encodes a putative transcription factor required for axon growth and guidance in drosophila. *Development*, 120(6), 1385–98.

Goeke, S., Greene, E. A., Grant, P. K., Gates, M. A., Crowner, D., Aigaki, T., & Giniger, E. (2003). Alternative splicing of lola generates 19 transcription factors controlling axon guidance in drosophila. *Nat Neurosci*, 6(9), 917–24.

Goldberg, M. L. (1979). Sequence analysis of drosophila histone genes. *Ph.D. Thesis, Stanford University, CA : xiv + 159pp.*

Goodrich, J. A. & Tjian, R. (1994). Transcription factors iie and iih and atp hydrolysis direct promoter clearance by rna polymerase ii. *Cell*, 77(1), 145–56.

Goodrich, J. A. & Tjian, R. (2010). Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. *Nat Rev Genet*, 11(8), 549–58.

Graña, X., De Luca, A., Sang, N., Fu, Y., Claudio, P. P., Rosenblatt, J., Morgan, D. O., & Giordano, A. (1994). Pitalre, a nuclear cdc2-related protein kinase that phosphorylates the retinoblastoma protein in vitro. *Proc Natl Acad Sci U S A*, 91(9), 3834–8.

Grant, P. A., Duggan, L., Côté, J., Roberts, S. M., Brownell, J. E., Candau, R., Ohba, R., Owen-Hughes, T., Allis, C. D., Winston, F., Berger, S. L., & Workman, J. L. (1997). Yeast gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an ada complex and the saga (spt/ada) complex. *Genes Dev*, 11(13), 1640–50.

Grewal, S. I. S. & Elgin, S. C. R. (2002). Heterochromatin: new possibilities for the inheritance of structure. *Curr Opin Genet Dev*, 12(2), 178–87.

Grün, D., Lyubimova, A., Kester, L., Wiebrands, K., Basak, O., Sasaki, N., Clevers, H., & van Oudenaarden, A. (2015). Single-cell messenger rna sequencing reveals rare intestinal cell types. *Nature*, 525(7568), 251–5.

Grünberg, S., Warfield, L., & Hahn, S. (2012). Architecture of the rna polymerase ii preinitiation complex and mechanism of atp-dependent promoter opening. *Nat Struct Mol Biol*, 19(8), 788–96.

Gualdi, R., Bossard, P., Zheng, M., Hamada, Y., Coleman, J. R., & Zaret, K. S. (1996). Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. *Genes Dev*, 10(13), 1670–82.

Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R., & Young, R. A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. *Cell*, 130(1), 77–88.

Guo, S., Yamaguchi, Y., Schilbach, S., Wada, T., Lee, J., Goddard, A., French, D., Handa, H., & Rosenthal, A. (2000). A regulator of transcriptional elongation controls vertebrate neuronal development. *Nature*, 408(6810), 366–9.

Ha, I., Lane, W. S., & Reinberg, D. (1991). Cloning of a human gene encoding the general transcription initiation factor iib. *Nature*, 352(6337), 689–95.

Halpern, K. B., Shenhav, R., Matcovitch-Natan, O., Toth, B., Lemze, D., Golan, M., Massasa, E. E., Baydatch, S., Landen, S., Moor, A. E., Brandis, A., Giladi, A., Avihail, A. S., David, E., Amit, I., & Itzkovitz, S. (2017). Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. *Nature*, 542(7641), 352–356.

Hammonds, A. S., Bristow, C. A., Fisher, W. W., Weiszmann, R., Wu, S., Hartenstein, V., Kellis, M., Yu, B., Frise, E., & Celniker, S. E. (2013). Spatial expression of transcription factors in drosophila embryonic organ development. *Genome Biol*, 14(12), R140.

Hansen, S. K., Takada, S., Jacobson, R. H., Lis, J. T., & Tjian, R. (1997). Transcription properties of a cell type-specific tata-binding protein, trf. *Cell*, 91(1), 71–83.

Harrison, M. M., Li, X.-Y., Kaplan, T., Botchan, M. R., & Eisen, M. B. (2011). Zelda binding in the early drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-zygotic transition. *PLoS Genet*, 7(10), e1002266.

Hawkes, N. A. & Roberts, S. G. (1999). The role of human tfiib in transcription start site selection in vitro and in vivo. *J Biol Chem*, 274(20), 14337–43.

Hayes, J. J. & Lee, K. M. (1997). In vitro reconstitution and analysis of mononucleosomes containing defined dnas and proteins. *Methods*, 12(1), 2–9.

He, N., Liu, M., Hsu, J., Xue, Y., Chou, S., Burlingame, A., Krogan, N. J., Alber, T., & Zhou, Q. (2010). Hiv-1 tat and host aff4 recruit two transcription elongation factors into a bifunctional complex for coordinated activation of hiv-1 transcription. *Mol Cell*, 38(3), 428–38.

He, Y., Fang, J., Taatjes, D. J., & Nogales, E. (2013). Structural visualization of key steps in human transcription initiation. *Nature*, 495, 481 EP –.

Hebbar, P. B. & Archer, T. K. (2007). Chromatin-dependent cooperativity between site-specific transcription factors in vivo. *J Biol Chem*, 282(11), 8284–91.

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., Ye, Z., Lee, L. K., Stuart, R. K., Ching, C. W., Ching, K. A., Antosiewicz-Bourget, J. E., Liu, H., Zhang, X., Green, R. D., Lobanenkov, V. V., Stewart, R., Thomson, J. A., Crawford, G. E., Kellis, M., & Ren, B. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, 459(7243), 108–12.

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E., & Ren, B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet*, 39(3), 311–8.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H., & Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and b cell identities. *Mol Cell*, 38(4), 576–89.

Heitz, E. (1928). Das heterochromatin der moose. *Jehrb Wiss Botanik, 69:762-818*.

Hendrix, D. A., Hong, J.-W., Zeitlinger, J., Rokhsar, D. S., & Levine, M. S. (2008). Promoter elements associated with rna pol ii stalling in the drosophila embryo. *Proc Natl Acad Sci U S A*, 105(22), 7762–7.

Henriques, T., Scruggs, B. S., Inouye, M. O., Muse, G. W., Williams, L. H., Burkholder, A. B., Lavender, C. A., Fargo, D. C., & Adelman, K. (2018). Widespread transcriptional pausing and elongation control at enhancers. *Genes Dev*, 32(1), 26–41.

Hernandez, N. (1993). Tbp, a universal eukaryotic transcription factor? *Genes Dev*, 7(7B), 1291–308.

Herr, A. J., Jensen, M. B., Dalmay, T., & Baulcombe, D. C. (2005). Rna polymerase iv directs silencing of endogenous dna. *Science*, 308(5718), 118–120.

Hewitt, S. C., Li, R., Adams, N., Winuthayanon, W., Hamilton, K. J., Donoghue, L. J., Lierz, S. L., Garcia, M., Lydon, J. P., DeMayo, F. J., Adelman, K., & Korach, K. S. (2019). Negative elongation factor is essential for endometrial function. *FASEB J*, 33(2), 3010–3023.

Hisatake, K., Malik, S., Roeder, R. G., & Horikoshi, M. (1991). Conserved structural motifs between xenopus and human tfiib. *Nucleic Acids Res*, 19(23), 6639.

Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K., & Sharp, P. A. (2017). A phase separation model for transcriptional control. *Cell*, 169(1), 13 – 23.

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(6914), 439–45.

Hodges, C., Bintu, L., Lubkowska, L., Kashlev, M., & Bustamante, C. (2009). Nucleosomal fluctuations govern the transcription dynamics of rna polymerase ii. *Science*, 325(5940), 626–8.

Holstege, F. C., Fiedler, U., & Timmers, H. T. (1997). Three transitions in the rna polymerase ii transcription complex during initiation. *EMBO J*, 16(24), 7468–80.

Hope, I. A., Mahadevan, S., & Struhl, K. (1988). Structural and functional characterization of the short acidic transcriptional activation region of yeast gcn4 protein. *Nature*, 333(6174), 635–640.

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 Res*, 21(2), 182–92.

Hsin, J.-P. & Manley, J. L. (2012). The rna polymerase ii ctd coordinates transcription and rna processing. *Genes Dev*, 26(19), 2119–37.

Hsu, H.-T., Chen, H.-M., Yang, Z., Wang, J., Lee, N. K., Burger, A., Zaret, K., Liu, T., Levine, E., & Mango, S. E. (2015). Transcription. recruitment of rna polymerase ii by the pioneer transcription factor pha-4. *Science*, 348(6241), 1372–6.

Huynh, K. D. & Bardwell, V. J. (1998). The bcl-6 poz domain and other poz domains interact with the co-repressors n-cor and smrt. *Oncogene*, 17(19), 2473–84.

Imbalzano, A. N., Zaret, K. S., & Kingston, R. E. (1994). Transcription factor (tf) iib and tfiia can independently increase the affinity of the tata-binding protein for dna. *J Biol Chem*, 269(11), 8280–6.

Inostroza, J., Flores, O., & Reinberg, D. (1991). Factors involved in specific transcription by mammalian rna polymerase ii. purification and functional analysis of general transcription factor iie. *J Biol Chem*, 266(14), 9304–8.

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lönnerberg, P., & Linnarsson, S. (2014). Quantitative single-cell rna-seq with unique molecular identifiers. *Nat Methods*, 11(2), 163–6.

Iwafuchi-Doi, M. (2019). The mechanistic basis for chromatin regulation by pioneer transcription factors. *Wiley Interdiscip Rev Syst Biol Med*, 11(1), e1427.

Iwafuchi-Doi, M., Donahue, G., Kakumanu, A., Watts, J. A., Mahony, S., Pugh, B. F., Lee, D., Kaestner, K. H., & Zaret, K. S. (2016). The pioneer transcription factor foxa maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Mol Cell*, 62(1), 79–91.

Iwafuchi-Doi, M. & Zaret, K. S. (2014). Pioneer transcription factors in cell reprogramming. *Genes Dev*, 28(24), 2679–92.

Jacq, X., Brou, C., Lutz, Y., Davidson, I., Chambon, P., & Tora, L. (1994). Human tafii30 is present in a distinct tfiid complex and is required for transcriptional activation by the estrogen receptor. *Cell*, 79(1), 107–17.

Jang, M. K., Mochizuki, K., Zhou, M., Jeong, H.-S., Brady, J. N., & Ozato, K. (2005). The bromodomain protein brd4 is a positive regulatory component of p-tefb and stimulates rna polymerase ii-dependent transcription. *Mol Cell*, 19(4), 523–34.

Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B., & Smale, S. T. (1994). Dna sequence requirements for transcriptional initiator activity in mammalian cells. *Mol Cell Biol*, 14(1), 116–27.

Jennings, B. H., Shah, S., Yamaguchi, Y., Seki, M., Phillips, R. G., Handa, H., & Ish-Horowicz, D. (2004). Locus-specific requirements for spt5 in transcriptional activation and repression in drosophila. *Curr Biol*, 14(18), 1680–4.

Ji, X., Zhou, Y., Pandit, S., Huang, J., Li, H., Lin, C. Y., Xiao, R., Burge, C. B., & Fu, X.-D. (2013). Sr proteins collaborate with 7sk and promoter-associated nascent rna to release paused polymerase. *Cell*, 153(4), 855–68.

Jin, H.-J., Zhao, J. C., Wu, L., Kim, J., & Yu, J. (2014). Cooperativity and equilibrium with foxa1 define the androgen receptor transcriptional program. *Nat Commun*, 5, 3972.

Kao, S. Y., Calman, A. F., Luciw, P. A., & Peterlin, B. M. (1987). Anti-termination of transcription within the long terminal repeat of hiv-1 by tat gene product. *Nature*, 330(6147), 489–93.

Kapranov, P., Cheng, J., Dike, S., Nix, D. A., Duttagupta, R., Willingham, A. T., Stadler, P. F., Hertel, J., Hackermüller, J., Hofacker, I. L., Bell, I., Cheung, E., Drenkow, J., Dumais, E., Patel, S., Helt, G., Ganesh, M., Ghosh, S., Piccolboni, A., Sementchenko, V., Tammana, H., & Gingeras, T. R. (2007). Rna maps reveal new rna classes and a possible function for pervasive transcription. *Science*, 316(5830), 1484–8.

Karaiskos, N., Wahle, P., Alles, J., Boltengagen, A., Ayoub, S., Kipar, C., Kocks, C., Rajewsky, N., & Zinzen, R. P. (2017). The drosophila embryo at single-cell transcriptome resolution. *Science*, 358(6360), 194–199.

Karim, F. D., Guild, G. M., & Thummel, C. S. (1993). The drosophila broad-complex plays a key role in controlling ecdysone-regulated gene expression at the onset of metamorphosis. *Development*, 118(3), 977–88.

Kato, M. & McKnight, S. L. (2018). A solid-state conceptualization of information transfer from gene to message to protein. *Annu Rev Biochem*, 87, 351–390.

Katsani, K. R., Hajibagheri, M. A., & Verrijzer, C. P. (1999). Co-operative dna binding by gaga transcription factor requires the conserved btb/poz domain and reorganizes promoter topology. *EMBO J*, 18(3), 698–708.

Kaufmann, J. & Smale, S. T. (1994). Direct recognition of initiator elements by a component of the transcription factor iid complex. *Genes Dev*, 8(7), 821–9.

Kedinger, C., Gniazdowski, M., Mandel, Jr, J. L., Gissinger, F., & Chambon, P. (1970). Alpha-amanitin: a specific inhibitor of one of two dna-pendent rna polymerase activities from calf thymus. *Biochem Biophys Res Commun*, 38(1), 165–71.

Kelly, K. F. & Daniel, J. M. (2006). Poz for effect–poz-zf transcription factors in cancer and development. *Trends Cell Biol*, 16(11), 578–87.

Killeen, M., Coulombe, B., & Greenblatt, J. (1992). Recombinant tbp, transcription factor iib, and rap30 are sufficient for promoter recognition by mammalian rna polymerase ii. *J Biol Chem*, 267(14), 9463–6.

Killeen, M. T. & Greenblatt, J. F. (1992). The general transcription factor rap30 binds to rna polymerase ii and prevents it from binding nonspecifically to dna. *Mol Cell Biol*, 12(1), 30–7.

Kim, J. L., Nikolov, D. B., & Burley, S. K. (1993a). Co-crystal structure of tbp recognizing the minor groove of a tata element. *Nature*, 365(6446), 520–7.

Kim, T. K., Ebright, R. H., & Reinberg, D. (2000). Mechanism of atp-dependent promoter melting by transcription factor iih. *Science*, 288(5470), 1418–22.

Kim, T.-K., Hemberg, M., Gray, J. M., Costa, A. M., Bear, D. M., Wu, J., Harmin, D. A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., Markenscoff-Papadimitriou, E., Kuhl, D., Bito, H., Worley, P. F., Kreiman, G., & Greenberg, M. E. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature*, 465, 182 EP –.

Kim, T.-K. & Shiekhattar, R. (2015). Architectural and functional commonalities between enhancers and promoters. *Cell*, 162(5), 948 – 959.

Kim, Y., Geiger, J. H., Hahn, S., & Sigler, P. B. (1993b). Crystal structure of a yeast tbp/tata-box complex. *Nature*, 365(6446), 512–20.

Kim, Y. J., Björklund, S., Li, Y., Sayre, M. H., & Kornberg, R. D. (1994). A multiprotein mediator of transcriptional activation and its interaction with the c-terminal repeat domain of rna polymerase ii. *Cell*, 77(4), 599–608.

Klein, A. M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz,

D. A., & Kirschner, M. W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*, 161(5), 1187–1201.

Knezetic, J. A. & Luse, D. S. (1986). The presence of nucleosomes on a dna template prevents initiation by rna polymerase ii in vitro. *Cell*, 45(1), 95–104.

Kobayashi, N., Boyer, T. G., & Berk, A. J. (1995). A class of activation domains interacts directly with tfiia and stimulates tfiia-tfiid-promoter complex assembly. *Mol Cell Biol*, 15(11), 6465–73.

Koenecke, N., Johnston, J., He, Q., Meier, S., & Zeitlinger, J. (2017). Drosophila poised enhancers are generated during tissue patterning with the help of repression. *Genome Res*, 27(1), 64–74.

Kojima, S., Hatano, M., Okada, S., Fukuda, T., Toyama, Y., Yuasa, S., Ito, H., & Tokuhisa, T. (2001). Testicular germ cell apoptosis in bcl6-deficient mice. *Development*, 128(1), 57–65.

Kokubo, T., Swanson, M. J., Nishikawa, J. I., Hinnebusch, A. G., & Nakatani, Y. (1998). The yeast taf145 inhibitory domain and tfiia competitively bind to tata-binding protein. *Mol Cell Biol*, 18(2), 1003–12.

Koleske, A. J. & Young, R. A. (1994). An rna polymerase ii holoenzyme responsive to activators. *Nature*, 368(6470), 466–9.

Koleske, A. J. & Young, R. A. (1995). The rna polymerase ii holoenzyme and its implications for gene regulation. *Trends Biochem Sci*, 20(3), 113–6.

Kornberg, R. D. (1974). Chromatin structure: a repeating unit of histones and dna. *Science*, 184(4139), 868–71.

Kornberg, R. D. (1977). Structure of chromatin. *Annu Rev Biochem*, 46, 931–54.

Kornberg, R. D. & Thomas, J. O. (1974). Chromatin structure; oligomers of the histones. *Science*, 184(4139), 865–8.

Krumm, A., Meulia, T., Brunvand, M., & Groudine, M. (1992). The block to transcriptional elongation within the human c-myc gene is determined in the promoter-proximal region. *Genes Dev*, 6(11), 2201–13.

Ku, M., Koche, R. P., Rheinbay, E., Mendenhall, E. M., Endoh, M., Mikkelsen, T. S., Presser, A., Nusbaum, C., Xie, X., Chi, A. S., Adli, M., Kasif, S., Ptaszek, L. M., Cowan, C. A., Lander, E. S., Koseki, H., & Bernstein, B. E. (2008). Genomewide analysis of prc1 and prc2 occupancy identifies two classes of bivalent domains. *PLoS Genet*, 4(10), e1000242.

Kubik, S., Bruzzone, M. J., Jacquet, P., Falcone, J.-L., Rougemont, J., & Shore, D. (2015). Nucleosome stability distinguishes two different promoter types at all protein-coding genes in yeast. *Mol Cell*, 60(3), 422–34.

Kuras, L., Kosa, P., Mencia, M., & Struhl, K. (2000). Taf-containing and taf-independent forms of transcriptionally active tbp in vivo. *Science*, 288(5469), 1244–8.

La Manno, G., Gyllborg, D., Codeluppi, S., Nishimura, K., Salto, C., Zeisel, A., Borm, L. E., Stott, S. R. W., Toledo, E. M., Villaescusa, J. C., Lönnerberg, P., Ryge, J., Barker, R. A., Arenas, E., & Linnarsson, S. (2016). Molecular diversity of midbrain development in mouse, human, and stem cells. *Cell*, 167(2), 566–580.e19.

Lagha, M., Bothma, J. P., Esposito, E., Ng, S., Stefanik, L., Tsui, C., Johnston, J., Chen, K., Gilmour, D. S., Zeitlinger, J., & Levine, M. S. (2013). Paused pol ii coordinates tissue morphogenesis in the drosophila embryo. *Cell*, 153(5), 976–87.

Lai, F., Gardini, A., Zhang, A., & Shiekhattar, R. (2015). Integrator mediates the biogenesis of enhancer rnas. *Nature*, 525, 399 EP –.

Lai, Z. C. & Li, Y. (1999). Tramtrack69 is positively and autonomously required for drosophila photoreceptor development. *Genetics*, 152(1), 299–305.

Landry, C. R., Lemos, B., Rifkin, S. A., Dickinson, W. J., & Hartl, D. L. (2007). Genetic properties influencing the evolvability of gene expression. *Science*, 317(5834), 118–21.

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(3), R25.

Larson, A. G., Elnatan, D., Keenen, M. M., Trnka, M. J., Johnston, J. B., Burlingame, A. L., Agard, D. A., Redding, S., & Narlikar, G. J. (2017). Liquid droplet formation by hp1-alpha suggests a role for phase separation in heterochromatin. *Nature*, 547(7662), 236–240.

Larsson, A. J. M., Johnsson, P., Hagemann-Jensen, M., Hartmanis, L., Faridani, O. R., Reinius, B., Segerstolpe, Å., Rivera, C. M., Ren, B., & Sandberg, R. (2019). Genomic encoding of transcriptional burst kinetics. *Nature*, 565(7738), 251–254.

Law, A., Hirayoshi, K., O'Brien, T., & Lis, J. T. (1998). Direct cloning of dna that interacts in vivo with a specific protein: application to rna polymerase ii and sites of pausing in drosophila. *Nucleic Acids Res*, 26(4), 919–24.

Lee, C., Li, X., Hechmer, A., Eisen, M., Biggin, M. D., Venters, B. J., Jiang, C., Li, J., Pugh, B. F., & Gilmour, D. S. (2008). Nelf and gaga factor are linked to promoter-proximal pausing at many genes in drosophila. *Mol Cell Biol*, 28(10), 3290–300.

Lee, C.-K., Shibata, Y., Rao, B., Strahl, B. D., & Lieb, J. D. (2004). Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat Genet*, 36(8), 900–5.

Lee, H., Kraus, K. W., Wolfner, M. F., & Lis, J. T. (1992). Dna sequence requirements for generating paused polymerase at the start of hsp70. *Genes Dev*, 6(2), 284–95.

Lenhard, B., Sandelin, A., & Carninci, P. (2012). Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat Rev Genet*, 13(4), 233–45.

LeRoy, G., Orphanides, G., Lane, W. S., & Reinberg, D. (1998). Requirement of rsf and fact for transcription of chromatin templates in vitro. *Science*, 282(5395), 1900–4.

Levine, M. (2011). Paused rna polymerase ii as a developmental checkpoint. *Cell*, 145(4), 502–11.

Li, B., Carey, M., & Workman, J. L. (2007). The role of chromatin during transcription. *Cell*, 128(4), 707–19.

Li, B., Howe, L., Anderson, S., Yates, 3rd, J. R., & Workman, J. L. (2003). The set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of rna polymerase ii. *J Biol Chem*, 278(11), 8897–903.

Li, J. & Gilmour, D. S. (2013). Distinct mechanisms of transcriptional pausing orchestrated by gaga factor and m1bp, a novel transcription factor. *EMBO J*, 32(13), 1829–41.

Li, S., Li, Y., Carthew, R. W., & Lai, Z. C. (1997). Photoreceptor cell differentiation requires regulated proteolysis of the transcriptional repressor tramtrack. *Cell*, 90(3), 469–78.

Li, X. Y., Bhaumik, S. R., & Green, M. R. (2000). Distinct classes of yeast promoters revealed by differential taf recruitment. *Science*, 288(5469), 1242–4.

Li, Y., Flanagan, P. M., Tschochner, H., & Kornberg, R. D. (1994). Rna polymerase ii initiation factor interactions and transcription start site selection. *Science*, 263(5148), 805–7.

Liang, H.-L., Nien, C.-Y., Liu, H.-Y., Metzstein, M. M., Kirov, N., & Rushlow, C. (2008). The zinc-finger protein zelda is a key activator of the early zygotic genome in drosophila. *Nature*, 456(7220), 400–3.

Lifton, R. P., Goldberg, M. L., Karp, R. W., & Hogness, D. S. (1978). The organization of the histone genes in drosophila melanogaster: functional and evolutionary implications. *Cold Spring Harb Symp Quant Biol*, 42 Pt 2, 1047–51.

Lin, C., Smith, E. R., Takahashi, H., Lai, K. C., Martin-Brown, S., Florens, L., Washburn, M. P., Conaway, J. W., Conaway, R. C., & Shilatifard, A. (2010). Aff4, a component of the ell/p-tefb elongation complex and a shared subunit of mll chimeras, can link transcription elongation to leukemia. *Mol Cell*, 37(3), 429–37.

Lindell, T. J., Weinberg, F., Morris, P. W., Roeder, R. G., & Rutter, W. J. (1970). Specific inhibition of nuclear rna polymerase ii by alpha-amanitin. *Science*, 170(3956), 447–9.

Liu, J. K., DiPersio, C. M., & Zaret, K. S. (1991). Extracellular signals that regulate liver transcription factors during hepatic differentiation in vitro. *Mol Cell Biol*, 11(2), 773–84.

Liu, X., Bushnell, D. A., Wang, D., Calero, G., & Kornberg, R. D. (2010). Structure of an rna polymerase ii-tfiib complex and the transcription initiation mechanism. *Science*, 327(5962), 206–9.

Lomaev, D., Mikhailova, A., Erokhin, M., Shaposhnikov, A. V., Moresco, J. J., Blokhina, T., Wolle, D., Aoki, T., Ryabykh, V., Yates, 3rd, J. R., Shidlovskii, Y. V., Georgiev, P., Schedl, P., & Chetverina, D. (2017). The gaga factor regulatory network: Identification of gaga factor associated proteins. *PLoS One*, 12(3), e0173602.

Lowary, P. T. & Widom, J. (1998). New dna sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol*, 276(1), 19–42.

Lu, F., Liu, Y., Inoue, A., Suzuki, T., Zhao, K., & Zhang, Y. (2016). Establishing chromatin regulatory landscape during mouse preimplantation development. *Cell*, 165(6), 1375–1388.

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 a resolution. *Nature*, 389(6648), 251–60.

Luo, Z., Lin, C., Guest, E., Garrett, A. S., Mohaghegh, N., Swanson, S., Marshall, S., Florens, L., Washburn, M. P., & Shilatifard, A. (2012a). The super elongation complex family of rna polymerase ii elongation factors: gene target specificity and transcriptional output. *Mol Cell Biol*, 32(13), 2608–17.

Luo, Z., Lin, C., & Shilatifard, A. (2012b). The super elongation complex (sec) family in transcriptional control. *Nat Rev Mol Cell Biol*, 13(9), 543–7.

Lupien, M., Eeckhoute, J., Meyer, C. A., Wang, Q., Zhang, Y., Li, W., Carroll, J. S., Liu, X. S., & Brown, M. (2008). Foxa1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell*, 132(6), 958–70.

MA, D., Olave, I., Merino, A., & Reinberg, D. (1996). Separation of the transcriptional coactivator and antirepression functions of transcription factor iia. *Proc Natl Acad Sci U S A*, 93(13), 6583–8.

Ma, J. & Ptashne, M. (1987). Deletion analysis of gal4 defines two transcriptional activating segments. *Cell*, 48(5), 847 – 853.

Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M., Trombetta, J. J., Weitz, D. A., Sanes, J. R., Shalek, A. K., Regev, A., & McCarroll, S. A. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, 161(5), 1202–1214.

Madden, K., Crowner, D., & Giniger, E. (1999). Lola has the properties of a master regulator of axon-target interaction for snb motor axons of drosophila. *Dev Biol*, 213(2), 301–13.

Makde, R. D., England, J. R., Yennawar, H. P., & Tan, S. (2010). Structure of rcc1 chromatin factor bound to the nucleosome core particle. *Nature*, 467(7315), 562–6.

Malik, S., Hisatake, K., Sumimoto, H., Horikoshi, M., & Roeder, R. G. (1991). Sequence of general transcription factor tfiib and relationships to other initiation factors. *Proc Natl Acad Sci U S A*, 88(21), 9553–7.

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A. F., Smith, A., & Stunnenberg, H. G. (2012). The transcriptional and epigenomic foundations of ground state pluripotency. *Cell*, 149(3), 590–604.

Marques, S., Zeisel, A., Codeluppi, S., van Bruggen, D., Mendanha Falcão, A., Xiao, L., Li, H., Häring, M., Hochgerner, H., Romanov, R. A., Gyllborg, D., Muñoz Manchado, A., La Manno,

G., Lönnerberg, P., Floriddia, E. M., Rezayee, F., Ernfors, P., Arenas, E., Hjerling-Leffler, J., Harkany, T., Richardson, W. D., Linnarsson, S., & Castelo-Branco, G. (2016). Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. *Science*, 352(6291), 1326–1329.

Marshall, N. F., Peng, J., Xie, Z., & Price, D. H. (1996). Control of rna polymerase ii elongation potential by a novel carboxyl-terminal domain kinase. *J Biol Chem*, 271(43), 27176–83.

Marshall, N. F. & Price, D. H. (1995). Purification of p-tefb, a transcription factor required for the transition into productive elongation. *J Biol Chem*, 270(21), 12335–8.

Martinez, E., Chiang, C. M., Ge, H., & Roeder, R. G. (1994). Tata-binding protein-associated factor(s) in tfiid function through the initiator to direct basal transcription from a tata-less class ii promoter. *EMBO J*, 13(13), 3115–26.

Martinez, E., Kundu, T. K., Fu, J., & Roeder, R. G. (1998). A human spt3-tafii31-gcn5-l acetylase complex distinct from transcription factor iid. *J Biol Chem*, 273(37), 23781–5.

Matsui, T., Segall, J., Weil, P. A., & Roeder, R. G. (1980). Multiple factors required for accurate initiation of transcription by purified rna polymerase ii. *J Biol Chem*, 255(24), 11992–6.

Mavrich, T. N., Jiang, C., Ioshikhes, I. P., Li, X., Venters, B. J., Zanton, S. J., Tomsho, L. P., Qi, J., Glaser, R. L., Schuster, S. C., Gilmour, D. S., Albert, I., & Pugh, B. F. (2008). Nucleosome organization in the drosophila genome. *Nature*, 453(7193), 358–62.

Maxon, M. E., Goodrich, J. A., & Tjian, R. (1994). Transcription factor iie binds preferentially to rna polymerase iia and recruits tfiih: a model for promoter clearance. *Genes Dev*, 8(5), 515–24.

Maxwell, C. S., Kruesi, W. S., Core, L. J., Kurhanewicz, N., Waters, C. T., Lewarch, C. L., Antoshechkin, I., Lis, J. T., Meyer, B. J., & Baugh, L. R. (2014). Pol ii docking and pausing at growth and stress genes in c. elegans. *Cell Rep*, 6(3), 455–66.

McChesney, P. A., Aiyar, S. E., Lee, O.-J., Zaika, A., Moskaluk, C., Li, R., & El-Rifai, W. (2006). Cofactor of brca1: a novel transcription factor regulator in upper gastrointestinal adenocarcinomas. *Cancer Res*, 66(3), 1346–53.

McNamara, R. P., McCann, J. L., Gudipaty, S. A., & D'Orso, I. (2013). Transcription factors mediate the enzymatic disassembly of promoter-bound 7sk snrnp to locally recruit p-tefb for transcription elongation. *Cell Rep*, 5(5), 1256–68.

Meers, M. P., Adelman, K., Duronio, R. J., Strahl, B. D., McKay, D. J., & Matera, A. G. (2018). Transcription start site profiling uncovers divergent transcription and enhancer-associated rnas in drosophila melanogaster. *BMC Genomics*, 19(1), 157.

Michels, A. A., Fraldi, A., Li, Q., Adamson, T. E., Bonnet, F., Nguyen, V. T., Sedore, S. C., Price, J. P., Price, D. H., Lania, L., & Bensaude, O. (2004). Binding of the 7sk snrna turns the hexim1 protein into a p-tefb (cdk9/cyclin t) inhibitor. *EMBO J*, 23(13), 2608–19.

Midorikawa, Y., Tsutsumi, S., Taniguchi, H., Ishii, M., Kobune, Y., Kodama, T., Makuuchi, M., & Aburatani, H. (2002). Identification of genes associated with dedifferentiation of hepatocellular carcinoma with expression profiling analysis. *Jpn J Cancer Res*, 93(6), 636–43.

Min, I. M., Waterfall, J. J., Core, L. J., Munroe, R. J., Schimenti, J., & Lis, J. T. (2011). Regulating rna polymerase pausing and transcription elongation in embryonic stem cells. *Genes Dev*, 25(7), 742–54.

Moreland, R. J., Tirode, F., Yan, Q., Conaway, J. W., Egly, J. M., & Conaway, R. C. (1999). A role for the tfiih xpb dna helicase in promoter escape by rna polymerase ii. *J Biol Chem*, 274(32), 22127–30.

Morse, R. H. (1993). Nucleosome disruption by transcription factor binding in yeast. *Science*, 262(5139), 1563–6.

Moshonov, S., Elfakess, R., Golan-Mashiach, M., Sinvani, H., & Dikstein, R. (2008). Links between core promoter and basic gene features influence gene expression. *BMC Genomics*, 9, 92.

Muniz, L., Egloff, S., Ughy, B., Jády, B. E., & Kiss, T. (2010). Controlling cellular p-tefb activity by the hiv-1 transcriptional transactivator tat. *PLoS Pathog*, 6(10), e1001152.

Muse, G. W., Gilchrist, D. A., Nechaev, S., Shah, R., Parker, J. S., Grissom, S. F., Zeitlinger, J., & Adelman, K. (2007). Rna polymerase is poised for activation across the genome. *Nat Genet*, 39(12), 1507–11.

Nambu, J. R., Lewis, J. O., Wharton, Jr, K. A., & Crews, S. T. (1991). The drosophila single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of cns midline development. *Cell*, 67(6), 1157–67.

Narita, T., Yamaguchi, Y., Yano, K., Sugimoto, S., Chanarat, S., Wada, T., Kim, D.-k., Hasegawa, J., Omori, M., Inukai, N., Endoh, M., Yamada, T., & Handa, H. (2003). Human transcription elongation factor nelf: identification of novel subunits and reconstitution of the functionally active complex. *Mol Cell Biol*, 23(6), 1863–73.

Nechaev, S., Fargo, D. C., dos Santos, G., Liu, L., Gao, Y., & Adelman, K. (2010). Global analysis of short rnas reveals widespread promoter-proximal stalling and arrest of pol ii in drosophila. *Science*, 327(5963), 335–8.

Ng, H. H., Dole, S., & Struhl, K. (2003). The rtf1 component of the paf1 transcriptional elongation complex is required for ubiquitination of histone h2b. *J Biol Chem*, 278(36), 33625–8.

Nguyen, V. T., Kiss, T., Michels, A. A., & Bensaude, O. (2001). 7sk small nuclear rna binds to and inhibits the activity of cdk9/cyclin t complexes. *Nature*, 414(6861), 322–5.

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(6545), 119–28.

Nikolov, D. B., Hu, S. H., Lin, J., Gasch, A., Hoffmann, A., Horikoshi, M., Chua, N. H., Roeder, R. G., & Burley, S. K. (1992). Crystal structure of tfiid tata-box binding protein. *Nature*, 360(6399), 40–6.

Ogryzko, V. V., Kotani, T., Zhang, X., Schiltz, R. L., Howard, T., Yang, X. J., Howard, B. H., Qin, J., & Nakatani, Y. (1998). Histone-like tafs within the pcaf histone acetylase complex. *Cell*, 94(1), 35–44.

Ohkuma, Y. & Roeder, R. G. (1994). Regulation of tfiih atpase and kinase activities by tfiie during active initiation complex formation. *Nature*, 368(6467), 160–3.

Ohkuma, Y., Sumimoto, H., Horikoshi, M., & Roeder, R. G. (1990). Factors involved in specific transcription by mammalian rna polymerase ii: purification and characterization of general transcription factor tfiie. *Proc Natl Acad Sci U S A*, 87(23), 9163–7.

Ohler, U., Liao, G.-c., Niemann, H., & Rubin, G. M. (2002). Computational analysis of core promoters in the drosophila genome. *Genome Biol*, 3(12), RESEARCH0087.

Okada, M. & Hirose, S. (1998). Chromatin remodeling mediated by drosophila gaga factor and iswi activates fushi tarazu gene transcription in vitro. *Mol Cell Biol*, 18(5), 2455–61.

Orphanides, G., Lagrange, T., & Reinberg, D. (1996). The general transcription factors of rna polymerase ii. *Genes Dev*, 10(21), 2657–83.

Orphanides, G., LeRoy, G., Chang, C. H., Luse, D. S., & Reinberg, D. (1998). Fact, a factor that facilitates transcript elongation through nucleosomes. *Cell*, 92(1), 105–16.

Orphanides, G., Wu, W. H., Lane, W. S., Hampsey, M., & Reinberg, D. (1999). The chromatin-specific transcription elongation factor fact comprises human spt16 and ssrp1 proteins. *Nature*, 400(6741), 284–8.

Pal, M., Ponticelli, A. S., & Luse, D. S. (2005). The role of the transcription bubble and tfiib in promoter clearance by rna polymerase ii. *Mol Cell*, 19(1), 101–10.

Palangat, M. & Landick, R. (2001). Roles of rna:dna hybrid stability, rna structure, and active site conformation in pausing by human rna polymerase ii. *J Mol Biol*, 311(2), 265–82.

Pan, G. & Greenblatt, J. (1994). Initiation of transcription by rna polymerase ii is limited by melting of the promoter dna in the region immediately upstream of the initiation site. *J Biol Chem*, 269(48), 30101–4.

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(7300), 956–60.

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(28), 18414–21.

Patikoglou, G. A., Kim, J. L., Sun, L., Yang, S. H., Kodadek, T., & Burley, S. K. (1999). Tata element recognition by the tata box-binding protein has been conserved throughout evolution. *Genes Dev*, 13(24), 3217–30.

Peng, J., Marshall, N. F., & Price, D. H. (1998a). Identification of a cyclin subunit required for the function of drosophila p-tefb. *J Biol Chem*, 273(22), 13855–60.

Peng, J., Zhu, Y., Milton, J. T., & Price, D. H. (1998b). Identification of multiple cyclin subunits of human p-tefb. *Genes Dev*, 12(5), 755–62.

Peterlin, B. M. & Price, D. H. (2006). Controlling the elongation phase of transcription with p-tefb. *Mol Cell*, 23(3), 297–305.

Phatnani, H. P. & Greenleaf, A. L. (2006). Phosphorylation and functions of the rna polymerase ii ctd. *Genes Dev*, 20(21), 2922–36.

Pintard, L., Willis, J. H., Willems, A., Johnson, J.-L. F., Srayko, M., Kurz, T., Glaser, S., Mains, P. E., Tyers, M., Bowerman, B., & Peter, M. (2003). The btb protein mel-26 is a substrate-specific adaptor of the cul-3 ubiquitin-ligase. *Nature*, 425(6955), 311–6.

Pinto, I., Ware, D. E., & Hampsey, M. (1992). The yeast sua7 gene encodes a homolog of human transcription factor tfiib and is required for normal start site selection in vivo. *Cell*, 68(5), 977–88.

Plet, A., Eick, D., & Blanchard, J. M. (1995). Elongation and premature termination of transcripts initiated from c-fos and c-myc promoters show dissimilar patterns. *Oncogene*, 10(2), 319–28.

Pointud, J. C., Larsson, J., Dastugue, B., & Couderc, J. L. (2001). The btb/poz domain of the regulatory proteins bric à brac 1 (bab1) and bric à brac 2 (bab2) interacts with the novel drosophila taf(ii) factor bip2/dtaf(ii)155. *Dev Biol*, 237(2), 368–80.

Price, D. H., Sluder, A. E., & Greenleaf, A. L. (1989). Dynamic interaction between a drosophila transcription factor and rna polymerase ii. *Mol Cell Biol*, 9(4), 1465–75.

Ptashne, M. & Gann, A. (1997). Transcriptional activation by recruitment. *Nature*, 386(6625), 569–577.

Pugh, B. F. & Tjian, R. (1990). Mechanism of transcriptional activation by sp1: evidence for coactivators. *Cell*, 61(7), 1187–97.

Pugh, B. F. & Tjian, R. (1991). Transcription from a tata-less promoter requires a multisubunit tfiid complex. *Genes Dev*, 5(11), 1935–45.

Purnell, B. A., Emanuel, P. A., & Gilmour, D. S. (1994). Tfiid sequence recognition of the initiator and sequences farther downstream in drosophila class ii genes. *Genes Dev*, 8(7), 830–42.

Rabenstein, M. D., Zhou, S., Lis, J. T., & Tjian, R. (1999). Tata box-binding protein (tbp)-related factor 2 (trf2), a third member of the tbp family. *Proc Natl Acad Sci U S A*, 96(9), 4791–6.

Rach, E. A., Yuan, H.-Y., Majoros, W. H., Tomancak, P., & Ohler, U. (2009). Motif composition, conservation and condition-specificity of single and alternative transcription start sites in the drosophila genome. *Genome Biol*, 10(7), R73.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., & Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, 470(7333), 279–83.

Raisner, R., Kharbanda, S., Jin, L., Jeng, E., Chan, E., Merchant, M., Haverty, P. M., Bainer, R., Cheung, T., Arnott, D., Flynn, E. M., Romero, F. A., Magnuson, S., & Gascoigne, K. E. (2018). Enhancer activity requires cbp/p300 bromodomain-dependent histone h3k27 acetylation. *Cell Rep*, 24(7), 1722–1729.

Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y., & Tyagi, S. (2006). Stochastic mrna synthesis in mammalian cells. *PLoS Biol*, 4(10), e309.

Rasmussen, E. B. & Lis, J. T. (1993). In vivo transcriptional pausing and cap formation on three drosophila heat shock genes. *Proc Natl Acad Sci U S A*, 90(17), 7923–7.

Reeve, J. N. (2003). Archaeal chromatin and transcription. *Mol Microbiol*, 48(3), 587–98.

Reinberg, D. & Roeder, R. G. (1987). Factors involved in specific transcription by mammalian rna polymerase ii. purification and functional analysis of initiation factors iib and iie. *J Biol Chem*, 262(7), 3310–21.

Richmond, T. J., Finch, J. T., Rushton, B., Rhodes, D., & Klug, A. (1984). Structure of the nucleosome core particle at 7 a resolution. *Nature*, 311(5986), 532–7.

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(11), 1647–1653.

Robzyk, K., Recht, J., & Osley, M. A. (2000). Rad6-dependent ubiquitination of histone h2b in yeast. *Science*, 287(5452), 501–4.

Roeder, R. G. (1996). The role of general initiation factors in transcription by rna polymerase ii. *Trends Biochem Sci*, 21(9), 327–35.

Roeder, R. G. & Rutter, W. J. (1969). Multiple forms of dna-dependent rna polymerase in eukaryotic organisms. *Nature*, 224(5216), 234–237.

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). Vsearch: a versatile open source tool for metagenomics. *PeerJ*, 4, e2584.

Rosenberg, A. B., Roco, C. M., Muscat, R. A., Kuchina, A., Sample, P., Yao, Z., Graybuck, L. T., Peeler, D. J., Mukherjee, S., Chen, W., Pun, S. H., Sellers, D. L., Tasic, B., & Seelig, G. (2018). Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science*, 360(6385), 176–182.

Rougvie, A. E. & Lis, J. T. (1988). The rna polymerase ii molecule at the 5' end of the uninduced hsp70 gene of d. melanogaster is transcriptionally engaged. *Cell*, 54(6), 795–804.

Rougvie, A. E. & Lis, J. T. (1990). Postinitiation transcriptional control in drosophila melanogaster. *Mol Cell Biol*, 10(11), 6041–5.

Saldi, T., Cortazar, M. A., Sheridan, R. M., & Bentley, D. L. (2016). Coupling of rna polymerase ii transcription elongation with pre-mrna splicing. *J Mol Biol*, 428(12), 2623–2635.

Sanulli, S., Trnka, M. J., Dharmarajan, V., Tibble, R. W., Pascal, B. D., Burlingame, A. L., Griffin, P. R., Gross, J. D., & Narlikar, G. J. (2019). Hp1 reshapes nucleosome core to promote heterochromatin phase separation. *Nature*.

Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., & Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol*, 33(5), 495–502.

Sauer, F., Hansen, S. K., & Tjian, R. (1995a). Dna template and activator-coactivator requirements for transcriptional synergism by drosophila bicoid. *Science*, 270(5243), 1825–8.

Sauer, F., Hansen, S. K., & Tjian, R. (1995b). Multiple tafiis directing synergistic activation of transcription. *Science*, 270(5243), 1783–8.

Scafe, C., Chao, D., Lopes, J., Hirsch, J. P., Henry, S., & Young, R. A. (1990). Rna polymerase ii c-terminal repeat influences response to transcriptional enhancer signals. *Nature*, 347(6292), 491–494.

Schug, J., Schuller, W.-P., Kappen, C., Salbaum, J. M., Bucan, M., & Stoeckert, Jr, C. J. (2005). Promoter features related to tissue specificity as measured by shannon entropy. *Genome Biol*, 6(4), R33.

Schwabish, M. A. & Struhl, K. (2006). Asf1 mediates histone eviction and deposition during elongation by rna polymerase ii. *Mol Cell*, 22(3), 415–22.

Seeger, M., Tear, G., Ferres-Marco, D., & Goodman, C. S. (1993). Mutations affecting growth cone guidance in drosophila: genes necessary for guidance toward or away from the midline. *Neuron*, 10(3), 409–26.

Segal, E. & Widom, J. (2009). Poly(da:dt) tracts: major determinants of nucleosome organization. *Curr Opin Struct Biol*, 19(1), 65–71.

Sekimizu, K., Kobayashi, N., Mizuno, D., & Natori, S. (1976). Purification of a factor from ehrlich ascites tumor cells specifically stimulating rna polymerase ii. *Biochemistry*, 15(23), 5064–70.

Sekiya, T., Muthurajan, U. M., Luger, K., Tulin, A. V., & Zaret, K. S. (2009). Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor foxa. *Genes Dev*, 23(7), 804–9.

Shah, N., Dorer, D. R., Moriyama, E. N., & Christensen, A. C. (2012). Evolution of a large, conserved, and syntenic gene family in insects. *G3 (Bethesda)*, 2(2), 313–9.

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.

Shao, W. & Zeitlinger, J. (2017). Paused rna polymerase ii inhibits new transcriptional initiation. *Nat Genet*, 49(7), 1045–1051.

Shekhar, K., Lapan, S. W., Whitney, I. E., Tran, N. M., Macosko, E. Z., Kowalczyk, M., Adiconis, X., Levin, J. Z., Nemesh, J., Goldman, M., McCarroll, S. A., Cepko, C. L., Regev, A., & Sanes, J. R. (2016). Comprehensive classification of retinal bipolar neurons by single-cell transcriptomics. *Cell*, 166(5), 1308–1323.e30.

Shilatifard, A., Lane, W. S., Jackson, K. W., Conaway, R. C., & Conaway, J. W. (1996). An rna polymerase ii elongation factor encoded by the human ell gene. *Science*, 271(5257), 1873–6.

Shiro, I. & Natalie, K., Eds. (2010). *Zinc finger proteins: from atomic contact to cellular functions*. Landes Bioscience.

Shopland, L. S., Hirayoshi, K., Fernandes, M., & Lis, J. T. (1995). Hsf access to heat shock elements in vivo depends critically on promoter architecture defined by gaga factor, tfiid, and rna polymerase ii binding sites. *Genes Dev*, 9(22), 2756–69.

Sigler, P. B. (1988). Acid blobs and negative noodles. *Nature*, 333(6170), 210–212.

Sikorski, T. W. & Buratowski, S. (2009). The basal initiation machinery: beyond the general transcription factors. *Current Opinion in Cell Biology*, 21(3), 344 – 351. Nucleus and gene expression.

Smale, S. T. & Baltimore, D. (1989). The "initiator" as a transcription control element. *Cell*, 57(1), 103–13.

Small, S., Blair, A., & Levine, M. (1992). Regulation of even-skipped stripe 2 in the drosophila embryo. *The EMBO Journal*, 11(11), 4047–4057.

Sobhian, B., Laguette, N., Yatim, A., Nakamura, M., Levy, Y., Kiernan, R., & Benkirane, M. (2010). Hiv-1 tat assembles a multifunctional transcription elongation complex and stably associates with the 7sk snrnp. *Mol Cell*, 38(3), 439–51.

Soufi, A., Garcia, M. F., Jaroszewicz, A., Osman, N., Pellegrini, M., & Zaret, K. S. (2015). Pioneer transcription factors target partial dna motifs on nucleosomes to initiate reprogramming. *Cell*, 161(3), 555–568.

Southall, T. D., Davidson, C. M., Miller, C., Carr, A., & Brand, A. H. (2014). Dedifferentiation of neurons precedes tumor formation in lola mutants. *Dev Cell*, 28(6), 685–96.

Spitz, F. & Furlong, E. E. M. (2012). Transcription factors: from enhancer binding to developmental control. *Nature Reviews Genetics*, 13, 613 EP –.

Stogios, P. J., Downs, G. S., Jauhal, J. J. S., Nandra, S. K., & Privé, G. G. (2005). Sequence and structural analysis of btb domain proteins. *Genome Biol*, 6(10), R82.

Strobl, L. J. & Eick, D. (1992). Hold back of rna polymerase ii at the transcription start site mediates down-regulation of c-myc in vivo. *EMBO J*, 11(9), 3307–14.

Strom, A. R., Emelyanov, A. V., Mir, M., Fyodorov, D. V., Darzacq, X., & Karpen, G. H. (2017). Phase separation drives heterochromatin domain formation. *Nature*, 547(7662), 241–245.

Struhl, K. (1999). Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell*, 98(1), 1–4.

Sun, J., Watkins, G., Blair, A. L., Moskaluk, C., Ghosh, S., Jiang, W. G., & Li, R. (2008). Deregulation of cofactor of brca1 expression in breast cancer cells. *J Cell Biochem*, 103(6), 1798–807.

Suter, D. M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., & Naef, F. (2011). Mammalian genes are transcribed with widely different bursting kinetics. *Science*, 332(6028), 472–474.

Svejstrup, J. Q., Li, Y., Fellows, J., Gnatt, A., Bjorklund, S., & Kornberg, R. D. (1997). Evidence for a mediator cycle at the initiation of transcription. *Proc Natl Acad Sci U S A*, 94(12), 6075–8.

Tadigotla, V. R., O Maoiléidigh, D., Sengupta, A. M., Epshtein, V., Ebright, R. H., Nudler, E., & Ruckenstein, A. E. (2006). Thermodynamic and kinetic modeling of transcriptional pausing. *Proc Natl Acad Sci U S A*, 103(12), 4439–44.

Tan, S., Hunziker, Y., Sargent, D. F., & Richmond, T. J. (1996). Crystal structure of a yeast tfiia/tbp/dna complex. *Nature*, 381(6578), 127–51.

Tanay, A. & Regev, A. (2017). Scaling single-cell genomics from phenomenology to mechanism. *Nature*, 541(7637), 331–338.

Tantin, D. & Carey, M. (1994). A heteroduplex template circumvents the energetic requirement for atp during activated transcription by rna polymerase ii. *J Biol Chem*, 269(26), 17397–400.

Tettey, T. T., Gao, X., Shao, W., Li, H., Story, B. A., Chitsazan, A. D., Glaser, R. L., Goode, Z. H., Seidel, C. W., Conaway, R. C., Zeitlinger, J., Blanchette, M., & Conaway, J. W. (2019). A role for fact in rna polymerase ii promoter-proximal pausing. *Cell Rep*, 27(13), 3770–3779.e7.

Thomas, S., Li, X.-Y., Sabo, P. J., Sandstrom, R., Thurman, R. E., Canfield, T. K., Giste, E., Fisher, W., Hammonds, A., Celniker, S. E., Biggin, M. D., & Stamatoyannopoulos, J. A. (2011). Dynamic reprogramming of chromatin accessibility during drosophila embryo development. *Genome Biol*, 12(5), R43.

Tillo, D., Kaplan, N., Moore, I. K., Fondufe-Mittendorf, Y., Gossett, A. J., Field, Y., Lieb, J. D., Widom, J., Segal, E., & Hughes, T. R. (2010). High nucleosome occupancy is encoded at human regulatory sequences. *PLoS One*, 5(2), e9129.

Tirode, F., Busso, D., Coin, F., & Egly, J. M. (1999). Reconstitution of the transcription factor tfiih: assignment of functions for the three enzymatic subunits, xpb, xpd, and cdk7. *Mol Cell*, 3(1), 87–95.

Tirosh, I., Weinberger, A., Carmi, M., & Barkai, N. (2006). A genetic signature of interspecies variations in gene expression. *Nat Genet*, 38(7), 830–4.

Tomancak, P., Beaton, A., Weiszmann, R., Kwan, E., Shu, S., Lewis, S. E., Richards, S., Ashburner, M., Hartenstein, V., Celniker, S. E., & Rubin, G. M. (2002). Systematic determination of patterns of gene expression during drosophila embryogenesis. *Genome Biol*, 3(12), RESEARCH0088.

Tomancak, P., Berman, B. P., Beaton, A., Weiszmann, R., Kwan, E., Hartenstein, V., Celniker, S. E., & Rubin, G. M. (2007). Global analysis of patterns of gene expression during drosophila embryogenesis. *Genome Biol*, 8(7), R145.

Tora, L. (2002). A unified nomenclature for tata box binding protein (tbp)-associated factors (tafs) involved in rna polymerase ii transcription. *Genes Dev*, 16(6), 673–5.

Tsai, A., Muthusamy, A. K., Alves, M. R., Lavis, L. D., Singer, R. H., Stern, D. L., & Crocker, J. (2017). Nuclear microenvironments modulate transcription from low-affinity enhancers. *Elife*, 6.

Tsukiyama, T., Becker, P. B., & Wu, C. (1994). Atp-dependent nucleosome disruption at a heat-shock promoter mediated by binding of gaga transcription factor. *Nature*, 367(6463), 525–32.

Tyree, C. M., George, C. P., Lira-DeVito, L. M., Wampler, S. L., Dahmus, M. E., Zawel, L., & Kadonaga, J. T. (1993). Identification of a minimal set of proteins that is sufficient for accurate initiation of transcription by rna polymerase ii. *Genes Dev*, 7(7A), 1254–65.

van Roy, F. M. & McCrea, P. D. (2005). A role for kaiso-p120ctn complexes in cancer? *Nat Rev Cancer*, 5(12), 956–64.

Verrijzer, C. P., Chen, J. L., Yokomori, K., & Tjian, R. (1995). Binding of tafs to core elements directs promoter selectivity by rna polymerase ii. *Cell*, 81(7), 1115–25.

Verrijzer, C. P. & Tjian, R. (1996). Tafs mediate transcriptional activation and promoter selectivity. *Trends Biochem Sci*, 21(9), 338–42.

Verrijzer, C. P., Yokomori, K., Chen, J. L., & Tjian, R. (1994). Drosophila tafii150: similarity to yeast gene tsm-1 and specific binding to core promoter dna. *Science*, 264(5161), 933–41.

Villani, A.-C., Satija, R., Reynolds, G., Sarkizova, S., Shekhar, K., Fletcher, J., Griesbeck, M., Butler, A., Zheng, S., Lazo, S., Jardine, L., Dixon, D., Stephenson, E., Nilsson, E., Grundberg, I., McDonald, D., Filby, A., Li, W., De Jager, P. L., Rozenblatt-Rosen, O., Lane, A. A., Haniffa, M., Regev, A., & Hacohen, N. (2017). Single-cell rna-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science*, 356(6335).

Vos, S. M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H., & Cramer, P. (2018a). Structure of activated transcription complex pol ii-dsif-paf-spt6. *Nature*, 560(7720), 607–612.

Vos, S. M., Farnung, L., Urlaub, H., & Cramer, P. (2018b). Structure of paused transcription complex pol ii-dsif-nelf. *Nature*, 560(7720), 601–606.

Vos, S. M., Pöllmann, D., Caizzi, L., Hofmann, K. B., Rombaut, P., Zimniak, T., Herzog, F., & Cramer, P. (2016). Architecture and rna binding of the human negative elongation factor. *Elife*, 5.

Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G. A., Winston, F., Buratowski, S., & Handa, H. (1998). Dsif, a novel transcription elongation factor that regulates rna polymerase ii processivity, is composed of human spt4 and spt5 homologs. *Genes Dev*, 12(3), 343–56.

Wagner, A., Regev, A., & Yosef, N. (2016). Revealing the vectors of cellular identity with single-cell genomics. *Nat Biotechnol*, 34(11), 1145–1160.

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 Dev*, 28(14), 1550–5.

Wang, Y. V., Tang, H., & Gilmour, D. S. (2005). Identification in vivo of different rate-limiting steps associated with transcriptional activators in the presence and absence of a gaga element. *Mol Cell Biol*, 25(9), 3543–52.

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. *Mol Cell*, 68(1), 118–129.e5.

Wasylyk, B. & Chambon, P. (1979). Transcription by eukaryotic rna polymerases a and b of chromatin assembled in vitro. *Eur J Biochem*, 98(2), 317–27.

Weber, C. M., Ramachandran, S., & Henikoff, S. (2014). Nucleosomes are context-specific, h2a.z-modulated barriers to rna polymerase. *Mol Cell*, 53(5), 819–30.

Weeks, J. R., Hardin, S. E., Shen, J., Lee, J. M., & Greenleaf, A. L. (1993). Locus-specific variation in phosphorylation state of rna polymerase ii in vivo: correlations with gene activity and transcript processing. *Genes Dev*, 7(12A), 2329–44.

Weil, P. A., Luse, D. S., Segall, J., & Roeder, R. G. (1979). Selective and accurate initiation of transcription at the ad2 major late promotor in a soluble system dependent on purified rna polymerase ii and dna. *Cell*, 18(2), 469–84.

Weinmann, R., Raskas, H. J., & Roeder, R. G. (1974). Role of dna-dependent rna polymerases ii and iii in transcription of the adenovirus genome late in productive infection. *Proc Natl Acad Sci U S A*, 71(9), 3426–39.

Weinmann, R. & Roeder, R. G. (1974). Role of dna-dependent rna polymerase 3 in the transcription of the trna and 5s rna genes. *Proc Natl Acad Sci U S A*, 71(5), 1790–4.

Wen, Y., Nguyen, D., Li, Y., & Lai, Z. C. (2000). The n-terminal btb/poz domain and c-terminal sequences are essential for tramtrack69 to specify cell fate in the developing drosophila eye. *Genetics*, 156(1), 195–203.

Werner, F. & Grohmann, D. (2011). Evolution of multisubunit rna polymerases in the three domains of life. *Nature Reviews Microbiology*, 9, 85 EP –.

West, M. L. & Corden, J. L. (1995). Construction and analysis of yeast rna polymerase ii ctd deletion and substitution mutations. *Genetics*, 140(4), 1223–33.

Wheeler, S. R., Stagg, S. B., & Crews, S. T. (2009). Midexdb: a database of drosophila cns midline cell gene expression. *BMC Dev Biol*, 9, 56.

Wilkins, R. C. & Lis, J. T. (1997). Dynamics of potentiation and activation: Gaga factor and its role in heat shock gene regulation. *Nucleic Acids Res*, 25(20), 3963–8.

Williams, L. H., Fromm, G., Gokey, N. G., Henriques, T., Muse, G. W., Burkholder, A., Fargo, D. C., Hu, G., & Adelman, K. (2015). Pausing of rna polymerase ii regulates mammalian developmental potential through control of signaling networks. *Mol Cell*, 58(2), 311–322.

Wolner, B. S. & Gralla, J. D. (2001). Tata-flanking sequences influence the rate and stability of tata-binding protein and tfiib binding. *J Biol Chem*, 276(9), 6260–6.

Workman, J. L. & Kingston, R. E. (1992). Nucleosome core displacement in vitro via a metastable transcription factor-nucleosome complex. *Science*, 258(5089), 1780–4.

Wright, T. J., Costa, J. L., Naranjo, C., Francis-West, P., & Altherr, M. R. (1999). Comparative analysis of a novel gene from the wolf-hirschhorn/pitt-rogers-danks syndrome critical region. *Genomics*, 59(2), 203–12.

Wu, C.-H., Lee, C., Fan, R., Smith, M. J., Yamaguchi, Y., Handa, H., & Gilmour, D. S. (2005). Molecular characterization of drosophila nelf. *Nucleic Acids Res*, 33(4), 1269–79.

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 Dev*, 17(11), 1402–14.

Wu, L., Li, L., Zhou, B., Qin, Z., & Dou, Y. (2014). H2b ubiquitylation promotes rna pol ii processivity via paf1 and ptefb. *Mol Cell*, 54(6), 920–931.

Xi, L., Fondufe-Mittendorf, Y., Xia, L., Flatow, J., Widom, J., & Wang, J.-P. (2010). Predicting nucleosome positioning using a duration hidden markov model. *BMC Bioinformatics*, 11, 346.

Xiong, W. C. & Montell, C. (1993). tramtrack is a transcriptional repressor required for cell fate determination in the drosophila eye. *Genes Dev*, 7(6), 1085–96.

Xu, J., Watts, J. A., Pope, S. D., Gadue, P., Kamps, M., Plath, K., Zaret, K. S., & Smale, S. T. (2009). Transcriptional competence and the active marking of tissue-specific enhancers by defined transcription factors in embryonic and induced pluripotent stem cells. *Genes Dev*, 23(24), 2824–38.

Xu, L., Wei, Y., Reboul, J., Vaglio, P., Shin, T.-H., Vidal, M., Elledge, S. J., & Harper, J. W. (2003). Btb proteins are substrate-specific adaptors in an scf-like modular ubiquitin ligase containing cul-3. *Nature*, 425(6955), 316–21.

Yamada, S., Whitney, P. H., Huang, S.-K., Eck, E. C., Garcia, H. G., & Rushlow, C. A. (2019). The drosophila pioneer factor zelda modulates the nuclear microenvironment of a dorsal target enhancer to potentiate transcriptional output. *Curr Biol*, 29(8), 1387–1393.e5.

Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T., & Handa, H. (2006). P-tefb-mediated phosphorylation of hspt5 c-terminal repeats is critical for processive transcription elongation. *Mol Cell*, 21(2), 227–37.

Yamaguchi, Y., Inukai, N., Narita, T., Wada, T., & Handa, H. (2002). Evidence that negative

elongation factor represses transcription elongation through binding to a drb sensitivity-inducing factor/rna polymerase ii complex and rna. *Mol Cell Biol*, 22(9), 2918–27.

Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., & Handa, H. (1999). Nelf, a multisubunit complex containing rd, cooperates with dsif to repress rna polymerase ii elongation. *Cell*, 97(1), 41–51.

Yang, Z., Yik, J. H. N., Chen, R., He, N., Jang, M. K., Ozato, K., & Zhou, Q. (2005). Recruitment of p-tefb for stimulation of transcriptional elongation by the bromodomain protein brd4. *Mol Cell*, 19(4), 535–45.

Yang, Z., Zhu, Q., Luo, K., & Zhou, Q. (2001). The 7sk small nuclear rna inhibits the cdk9/cyclin t1 kinase to control transcription. *Nature*, 414(6861), 317–22.

Yik, J. H. N., Chen, R., Nishimura, R., Jennings, J. L., Link, A. J., & Zhou, Q. (2003). Inhibition of p-tefb (cdk9/cyclin t) kinase and rna polymerase ii transcription by the coordinated actions of hexim1 and 7sk snrna. *Mol Cell*, 12(4), 971–82.

Yokoyama, A., Lin, M., Naresh, A., Kitabayashi, I., & Cleary, M. L. (2010). A higher-order complex containing af4 and enl family proteins with p-tefb facilitates oncogenic and physiologic mll-dependent transcription. *Cancer Cell*, 17(2), 198–212.

Young, R. A. (2011). Control of the embryonic stem cell state. *Cell*, 144(6), 940–54.

Yu, X. & Buck, M. J. (2019). Defining tp53 pioneering capabilities with competitive nucleosome binding assays. *Genome Res*, 29(1), 107–115.

Yuan, G.-C., Liu, Y.-J., Dion, M. F., Slack, M. D., Wu, L. F., Altschuler, S. J., & Rando, O. J. (2005). Genome-scale identification of nucleosome positions in s. cerevisiae. *Science*, 309(5734), 626–30.

Zabidi, M. A., Arnold, C. D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., & Stark, A.

(2014). Enhancer–core-promoter specificity separates developmental and housekeeping gene regulation. *Nature*, 518, 556 EP –.

Zaret, K. S. & Carroll, J. S. (2011). Pioneer transcription factors: establishing competence for gene expression. *Genes Dev*, 25(21), 2227–41.

Zehring, W. A., Lee, J. M., Weeks, J. R., Jokerst, R. S., & Greenleaf, A. L. (1988). The c-terminal repeat domain of rna polymerase ii largest subunit is essential in vivo but is not required for accurate transcription initiation in vitro. *Proceedings of the National Academy of Sciences*, 85(11), 3698–3702.

Zeisel, A., Muñoz-Manchado, A. B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., Rolny, C., Castelo-Branco, G., Hjerling-Leffler, J., & Linnarsson, S. (2015). Brain structure. cell types in the mouse cortex and hippocampus revealed by single-cell rna-seq. *Science*, 347(6226), 1138–42.

Zeitlinger, J., Stark, A., Kellis, M., Hong, J.-W., Nechaev, S., Adelman, K., Levine, M., & Young, R. A. (2007). Rna polymerase stalling at developmental control genes in the drosophila melanogaster embryo. *Nat Genet*, 39(12), 1512–6.

Zentner, G. E., Tesar, P. J., & Scacheri, P. C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res*, 21(8), 1273–83.

Zhang, W., Wang, Y., Long, J., Girton, J., Johansen, J., & Johansen, K. M. (2003). A developmentally regulated splice variant from the complex lola locus encoding multiple different zinc finger domain proteins interacts with the chromosomal kinase jil-1. *J Biol Chem*, 278(13), 11696–704.

Zhang, Z., English, B. P., Grimm, J. B., Kazane, S. A., Hu, W., Tsai, A., Inouye, C., You, C., Piehler, J., Schultz, P. G., Lavis, L. D., Revyakin, A., & Tjian, R. (2016). Rapid dynamics of general transcription factor tfiib binding during preinitiation complex assembly revealed by single-molecule analysis. *Genes Dev*, 30(18), 2106–2118.

Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., Gregory, M. T., Shuga, J., Montesclaros, L., Underwood, J. G., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P. W., Hindson, C. M., Bharadwaj, R., Wong, A., Ness, K. D., Beppu, L. W., Deeg, H. J., McFarland, C., Loeb, K. R., Valente, W. J., Ericson, N. G., Stevens, E. A., Radich, J. P., Mikkelsen, T. S., Hindson, B. J., & Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. *Nat Commun*, 8, 14049.

Zhu, F., Farnung, L., Kaasinen, E., Sahu, B., Yin, Y., Wei, B., Dodonova, S. O., Nitta, K. R., Morgunova, E., Taipale, M., Cramer, P., & Taipale, J. (2018). The interaction landscape between transcription factors and the nucleosome. *Nature*, 562(7725), 76–81.

Zhu, Y., Pe'ery, T., Peng, J., Ramanathan, Y., Marshall, N., Marshall, T., Amendt, B., Mathews, M. B., & Price, D. H. (1997). Transcription elongation factor p-tefb is required for hiv-1 tat transactivation in vitro. *Genes Dev*, 11(20), 2622–32.

Zollman, S., Godt, D., Privé, G. G., Couderc, J. L., & Laski, F. A. (1994). The btb domain, found primarily in zinc finger proteins, defines an evolutionarily conserved family that includes several developmentally regulated genes in drosophila. *Proc Natl Acad Sci U S A*, 91(22), 10717–21.

Appendix A

Supplementary Figures

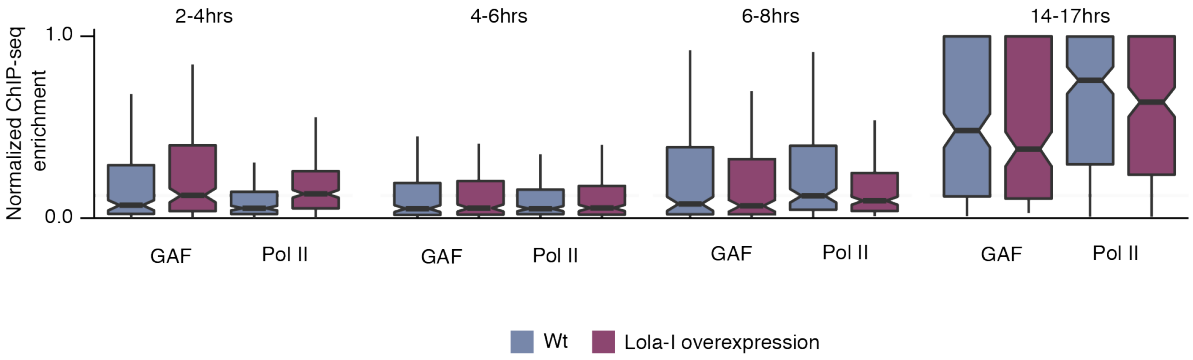

Figure A.1: Boxplot shows the differences in the normalized GAF and Pol II occupancy at the Lola-I target promoters by the Lola-I Wt and the Lola-I over-expression embryos. There are no clear changes in GAF and Pol II occupancy in the Lola-I over expression embryos. Over-expression lines were created by inserting full-length Lola-I cDNA construct driven by the actin-promoter into the attP40 locus.

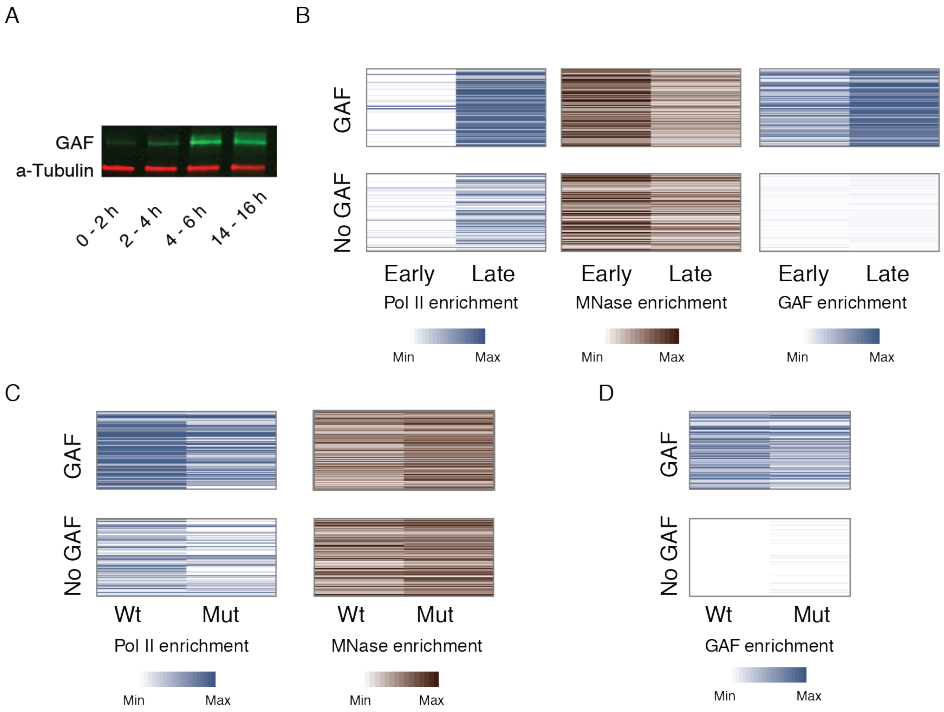

Figure A.2: Role of GAF in Pol II recruitment at the Lola-I target promoters. A) Western blot shows increase in the expression of GAF over *Drsohphila* embryonic development. While there is considerable expression at the early stages, the GAF expression increases over time. B) Heatmap shows the correlation between the recruitment of Pol II, the decrease in nucleosome occupancy, and the recruitment of GAF at regions with or without the GAGAG motif. There is no considerable amount of GAF binding at the regions without the GAGAG motif. C) Heatmap shows that both the Pol II occupancy and the MNase signal are reduced in the *lola*^{ORC4} mutant embryos. D) Heatmap shows that the GAF occupancy is decreased at the Lola-I target sites in the *lola*^{ORC4} mutant embryos.

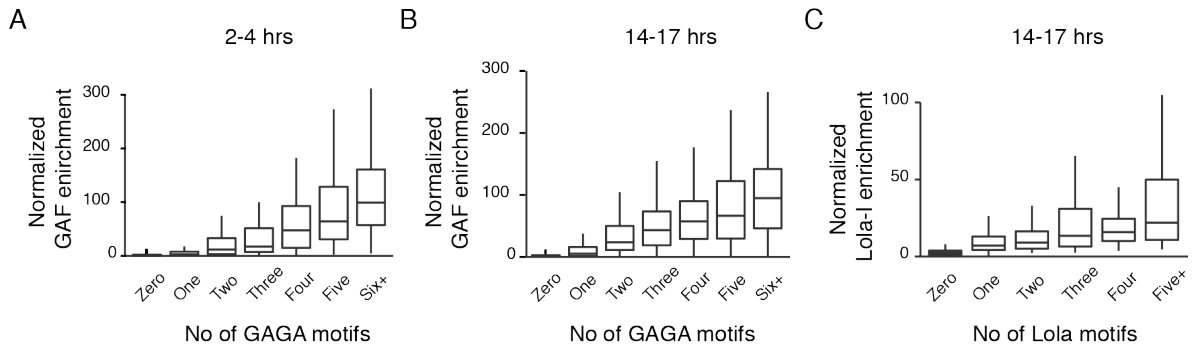

Figure A.3: A,B and C) Boxplot plot shows GAF and Lola-I signals at regions with different number of GAGAG motifs and Lola-I motifs. Strong cooperative binding can be observed for GAF especially at the 2-4 hrs. Binding of Lola-I is not super-additive with the increased number of Lola-I binding sites, suggesting no cooperative interactions in Lola-I binding.

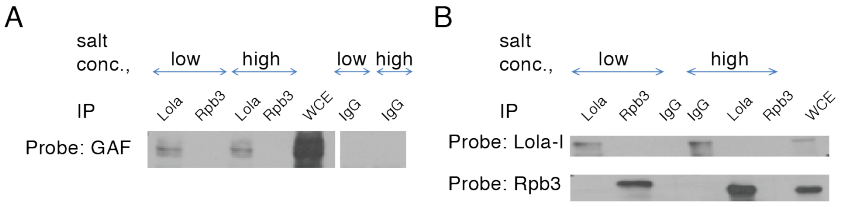

Figure A.4: Pull down experiments test interactions between Lola-I, GAF and RPB3. A) Western blot shows interactions between Lola-I and GAF at both the low salt wash and high salt wash conditions. B) Western blots show no interactions between Lola-I and RPB3 or GAF and RPB3.