Decoding the impact of promoter sequences on transcriptional dynamics in vivo

Kervy Carola Fernandez Oliveros

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Génétique moléculaire

École doctorale : Sciences Chimiques et Biologiques pour la Santé (CBS2)

Unité de recherche : Institut de Génétique Moléculaire de Montpellier (IGMM UMR5535)

Decoding the impact of promoter sequences on transcriptional dynamics *in vivo*

Décodage de l'impact des séquences de promoteurs sur la dynamique transcriptionnelle *in vivo*

Présentée par **Kervy Carola Fernandez Oliveros**
Le **4 décembre 2019**

Sous la direction du **Dr. Mounia Lagha**

Devant le jury composé de

Pr. Nathalie DOSTATNI, *DR, Inserm – Paris*	*Rapporteur*
Pr. François PAYRE, *DR, CNRS –TOULOUSE*	*Rapporteur*
Pr. Ferenc MUELLER, *University Of Birmingham*	*Examinateur*
Dr. Eugenia BASYUK, *CR, CNRS – Montpellier*	*Examinateur*
Dr. Mounia Lagha, *CR, CNRS – Montpellier*	*Directeur de thèse*

UNIVERSITÉ DE MONTPELLIER

Résumé

La transcription chez les eucaryotes est un processus complexe qui nécessite une coordination précise entre les éléments trans-régulateurs et séquences cis-régulatrices nécessaires à la transcription. En utilisant des méthodes quantitatives d'imagerie sur embryons de *Drosophile* vivants, combinées à un modèle mathématique, nous évaluons l'influence des séquences promotrices minimales (entre -50 ; +50 bp du TSS) sur la dynamique transcriptionnelle *in vivo*. Plus précisément, nous avons quantifié deux aspects temporels de la transcription, à savoir : 1) la synchronie, définie comme le degré de coordination internucléaire lors de l'activation transcriptionnelle au sein d'un domaine spatial donné. 2) la cinétique des changements d'état du promoteur (actif et inactif), via les fluctuations d'intensité des sites de transcription (traces) dans chaque noyau du pattern. Grâce à l'utilisation d'un modèle mathématique de déconvolution, nous avons pu estimer le K_{on} (le taux de changement de l'état actif (ON) du promoteur), le k_{off} (taux de changement vers l'état (OFF))), ainsi que le taux d'initiation de la polymérase (K_{ini}).

Nous montrons ici que, les promoteurs minimaux possèdent la propriété de contrôler la synchronie de l'activation transcriptionnelle dans des embryons de *Drosophile* vivants et que les promoteurs minimaux (exemple *sna* et *kr*) pouvaient avoir des profils synchrones similaires. De plus, nous avons observé que certains promoteurs qui sont dans un état de « pause » dans un contexte endogène, ont tendance à avoir un profil synchrone (exemple : promoteurs *sna* et *kr*), mais que la synchronie d'activation internucléaire peut être obtenue en l'absence de « pause » comme pour le promoteur *wntD*, où la présence d'un motif TATA canonique au sein du promoteur pourrait être responsable de l'activation synchrone. La comparaison des cinétiques des promoteurs *sna* et *kr*, montrent qu'ils ne diffèrent pas dans le taux d'initiation des polymérases et ceci en dépit de leurs séquences promotrices divergentes. En revanche, ils diffèrent dans leur cinétique d'activité de leurs promoteurs respectifs (T_{on}, T_{off}).

La manipulation de la boîte TATA du promoteur transgénique du gène *snail,* montre que la boîte TATA affecte la synchronie et la durée OFF du gène en priorité. En revanche les mutations effectuées sur le motif INR du promoteur transgénique *krüppel* ne semblent pas affecter la synchronie d'activation, mais plutôt la durée de l'activité du promoteur, (T_{on}, T_{off}), avec des taux d'initiation similaires.

Ensemble ces résultats suggèrent que la composition des éléments promoteurs ont un impact sur l'initiation de la transcription principalement à deux niveaux (synchronie et activité du promoteur (états ON/OFF). Cette régulation jouerait un rôle dans l'adaptation du contrôle transcriptionnel dans un contexte développemental donné.

Abstract

Transcription in eukaryotes is a highly complex process that requires precise coordination in the assembly of trans-regulating factors through the recognition of *cis*-regulatory DNA sequences. Using quantitative live imaging methods combined with a mathematical model, we assess the influence of core promoter sequences (between -50; +50 bp of the TSS) on transcription dynamics *in vivo*. More precisely, we quantified two temporal aspects of transcription, namely: 1) Synchrony, defined as the degree of inter-nuclear coordination in transcription activation within a given spatial domain. 2) The kinetic parameters of promoter state, that is, the K_{on} (rate at which a gene changes from an inactive (OFF) to an active (ON) state), the K_{off} (rate of change from an ON to an OFF state) and the K_{ini} (rate of polymerase initiation when a promoter is in an ON state), defined using the fluctuations of the transcriptional site intensity over time in each nucleus of the acquired area.

Here, we show that minimal promoters control the synchrony of transcriptional activation in living *Drosophila* embryos. Moreover, we observed that endogenous paused promoters tend to have synchronous profiles (e.g. *sna, kr*). However, we also observe that synchrony profiles can be achieved in the absence of pausing (eg. *wntD)*, where the presence of a canonical TATA motif might be responsible for synchronous activation. The comparison of the kinetics of the *sna* and *kr* promoters shows that they do not differ in the initiation rate of polymerases, but in the time of promoter activity (duration of ON and OFF).

The manipulation of the TATA box of the *sna* transgenic promoter shows that the TATA box affects the synchrony and OFF duration primarily. On the other hand, the mutations performed on the INR motif of the *kr* transgenic promoter do not seem to affect the synchrony of nuclei activation or the initiation rates, but rather, promoter activity, through the modulation of the ON/OFF switch kinetics.

Together these results suggest that the composition of promoter elements has an impact on the initiation of transcription mainly at two levels (synchrony, and promoter activity (ON/OFF states). This regulation might play a role in adapting the transcriptional response in a given developmental context.

Acknowledgements

First of all, I would like to thank professor Dostatni, professor Payre, professor Mueller and Dr. Basyuk for agreeing to devote their time and expertise to evaluating my work. I understand the burden this represents and I am very grateful to you. I hope that, at less in compensation, that this thesis will stimulate your interest.

I would like to thank Dr Mounia Lagha for giving me the opportunity to work on an exciting project. Thanks, to all the members of the MLL team (former and current), for all the good exchanges, and time spent on my project and presentations, especially Anto (the software magician!) and Matthieu, for their valuable implication and help. I would like to thank Virginia for helping me to regain confidence, for her advice and help. Thank very much to François Juge for participating in our lab meetings and for his mediation and his useful advice.

Thanks, a lot, to our collaborators Pr. Ovidiu Radulescu and Dr Edouard Bertrand for their work, availability and helpful conversations.

Thank you to the MRI platform for their technical support, especially to Baptiste for all the time spent trying to find the good settings for my movies and smiFISH acquisitions.

Merci à tout l'équipe directive et administrative de l'IGMM, à Etienne Schwob et à Marc Piechaczyk pour avoir financé et soutenu la création de l'association des PhDs et post-docs, ainsi que les career seminars! Un grand merci à Françoise et à Sarah, pour toujours être à l'écoute des besoins administratifs et autres.

Merci à mon fils, merci pour tout son amour inconditionnel et sa patience. Merci à ma famille adoptive en France (les Barjonnet!) un merci énorme pour toute votre aide et soutien, tout ça n'aurait pas était possible sans vous, surtout sans Nico (je te dois plus qu'une fière chandelle). Merci à mes deux sœurs de cœur, Sabine et Janice pour votre amour, et votre soutien, pour me nourrir et me financer…merci aussi à Rachida pour tout son soutien, merci à mon incroyable coloc Mélanie et à Rémy (alias Mario bros) pour m'avoir encouragé et pour tous les gâteaux et pain aux chocolats (pas chocolatine hein !) qui ont maintenu mon niveau de glucose au top pour la rédaction !

Merci à tous mes amis (IGMM, IRIM, CEFE, CBS, CRBM), spécialement les labs DFL, JCL, MHL, EBL, MRL, MPL, and ESL. Merci à Saphia, Yousra (my tomato girl), Chloe, Lucie, Matthias, Fabi, Adham, Juan de Dios, David, Etienne, Léo, Géro, Annabela, Suzanna, Xavier, Iria, Encar y Ana, Ana Victoria (mi venezolana preferida, te quiero un monton!), Matthias (IRIM), Tanja, Jorge (y su tequila), Matt, Uriel, Ana, Jen, Nico, Pili y Sebas, Juan Pablo, Marco, Julian, Sergio, Baptiste, Olivier, Annika, Diwas… Elena (mi artista), Mike et Xavier (mon couple préféré)… ces 4 ans passes à Mtp ont été extraordinaires grâce à vous tous (et désolée si j'oublié des gens). Merci pour tout votre soutien et encouragements. Et bien sûr à mes « besties » Sarah Guizou et Pierre Gatel (todos para uno y uno para todos !!!). Love you guys, you are the best family I have ever had!!!!!!

Gracias a mis padres, por haberme dado el regalo de la vida. Gracias a mis abuelos por siempre haberme dejado y inculcado de ser yo misma y ser libre, y antes de todo ser una buena persona. Y gracias a todos aquellos que de una forma u otra contribuyeron a construir la persona que soy hoy.

Une dernière merci pour Virgina Despentes et ses podcast, et à Nietzsche pour me rappeler, pendant la rédaction de ce manuscrit que « Ce n'est pas le doute, c'est la certitude qui rend fou.. »

Table of contents

List of abbreviations

Short forms (acronyms)

BRE: B Recognition Element

ChIP: Chromatin Immuno-Precipitation

CTD: C-Terminal Domain

CUT: Cryptic Unstable Transcript

DCE: Downstream Core Element

DNA: Deoxyribose Nucleic Acid

DPE: Downstream core Promoter Element

DRE: Distal Regulatory Element

DSIF: DRB Sensitivity-Inducing Factor

EGFP: Enhanced Green Fluorescent Protein

eRNA: enhancer RNA

FRAP: Florescence Recovery After Photobleaching

GFP: Green Fluorescent Protein

GTF: General Transcription Factor

mRNA: messenger RNA

ncRNA : non-coding RNA

NELF: Negative Elongation Factor

NFR: Nucleosome Free Region

PAP: Poly A Polymerase

PAS: PolyAdenylation Signal

PIC: Pre-Initiation Complex

PROMPTs: Promoter upstream transcripts

pTEFb: Positive Transcription Elongation Factor beta

RNAP: RNA polymerase

rRNA: ribosomal RNA

-seq: Sequencing

smFISH: Single-molecule Fluorescence in situ Hybridization

shRNA: Short Harpin RNA

snoRNA: small nucleolar RNA

snRNA: small nuclear RNA

TAFs: TATA binding Associated Factors

TF: Transcription Factor

TSS: Transcription Start Site

UTR: Un-translated Region

List of figures

List of Tables and boxes

I - INTRODUCTION

The definition of Life has been a challenge for scientists and philosophers. This is partially because life is a process rather than a substance or an object. From a genetic point of view life could be seen as the translation and execution of the information contained in DNA to guide the cell to differentiate, grow and maintain itself until death.

The first step common for all organisms to express the information contained in their genome is transcription which results in the copying of a DNA sequence into a RNA transcript. Transcription is a broad and regulated process depending mainly on the assembly of one enzymatic complex, RNA polymerase.

In eukaryotes, there are three types of RNA polymerase, RNA pol I, RNA Pol II and RNA Pol III, discovered 50 years ago by (Roeder and Rutter, 1969), thanks to column chromatography separation assay. The nomenclature I, II and III corresponds to their order of elution from the column. Their distinct activities were also characterized due to their differential sensitivity to α-amanitin (Lindell *et al.*, 1970), a highly toxic bicyclic octapeptide naturally produced by some members of the *Amanita* mushrooms genus. RNAP I is completely insensitive to the peptide while RNAP III shows only partial inhibition of its activity at high concentrations. RNAP II is the most sensitive and can be completely blocked by using only a 1/40 fraction of the dose that achieves partial RNAP III inhibition (Kedinger *et al.*, 1970; Weinmann and Roeder, 1974).

This initial discovery made by Robert Roeder in 1969 and the subsequent work of Pierre Chambon, marked the beginning of a field of investigation, leading to a better understanding of the mechanisms underlying RNA polymerase function, structure and mechanisms to achieve gene expression in eukaryotic cells.

In this chapter, I will first introduce the transcription initiation mediated by the RNA Pol II, and then I will discuss its association with core promoters in a eukaryotic model organism.

1. The RNA Polymerase II Transcription

Transcription mediated by the RNAP II concern all the genes encoding proteins and many non-coding genes. It can be divided into 3 highly regulated steps: Initiation, Elongation and Termination. Each of one of these steps imply a considerable number of factors (Figure 1).

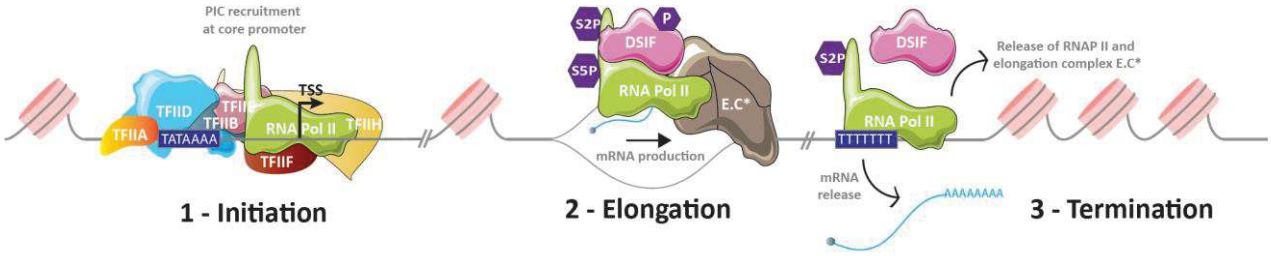


Figure 1. General steps of RNAP II transcription. First, during the Initiation the GTFs and the RNAP II are recruited at the core promoter. Second at the elongation step, RNAPII with the elongation complex machinery are into productive elongation given rise to pre-mRNA. Finally, the termination concludes with the release of the mRNA (caped and with a polyA tail) and the dissociation of the RNAP II from the DNA.

1.1. The RNAP II Initiation

The RNAPII initiation corresponds to the recognition and assembly of the pre-initiation complex (PIC) at the promoter (Figure 1, 2); with the final aim to enable the correct positioning of the RNAPII at the transcriptional start site (TSS). Early *in vitro* studies have described transcription as a succession of sequential events in which a tight interplay exists between GTFs, RNAPII and promoter DNA (Buratowski *et al.*, 1989; Zawel and Reinberg, 1992). The conventional pathway is depicted in figure 2. The local chromatin environment must adopt an "open" configuration to allow the recruitment of GTFs. This step can be achieved by certain chromatin remodelers and/or transcriptional factors (TF). Enabling the priming of enhancers, local nucleosome depletion and recruitment of the general transcription factors (GFTs) to bootstrap the recruitment of RNAP II at the core promoter. Biochemical and structural studies summarized in (Sainsbury, Bernecky and Cramer, 2015) proposed a sequential model where the first GTFs to recognize and bind core promoter

element is TFIID through one of its subunits, the Tata Binding Protein (TBP). This usually occurs around 25-31 base pairs (bp) upstream from the TSS, where depending on the promoter-type the TATA box element is located, and this interaction induces a unique distortion in the double helix DNA (Kim, Nikolov and Burley, 1993), facilitating the sequential recruitment of other complexes. After TFIID binding to the promoter it will follow the binding of TFIIA and TFIIB, next the recruitment of the complex RNAP II-TFIIF, and finally the binding of TFIIE and TFIIH. That will lead to the recruitment of the RNAP II, and then the formation of the Pre-Initiation Complex (PIC= GTFs + RNAP II). In the presence of ATP, the DNA is opened (Forming the 'transcription bubble') and RNA synthesis commences. In metazoan organisms, there is a supplementary step before effective elongation starts; call Promoter-proximal pausing or RNAPII pause. It occurs 30 to 60 bp downstream of the TSS (Core, Waterfall and Lis, 2008; Seila *et al.*, 2008) and is particularly robust at genes that are rapidly responsive to signaling pathways (Core *et al.*, 2012; Kwak *et al.*, 2013). RNAPII pause is induced by the binding of Negative Elongation Factor (NELF) and DRB sensitivity-inducing Factor (DSIF) (Yamaguchi, Shibata and Handa, 2013; Gaertner and Zeitlinger, 2014). Later through the action of activator, DSIF turns on to a positive elongation factor after its phosphorylation by pTEFb and dissociation of NELF (Gaertner and Zeitlinger, 2014), allowing the formation of the Pol II elongation complex to start effective elongation.

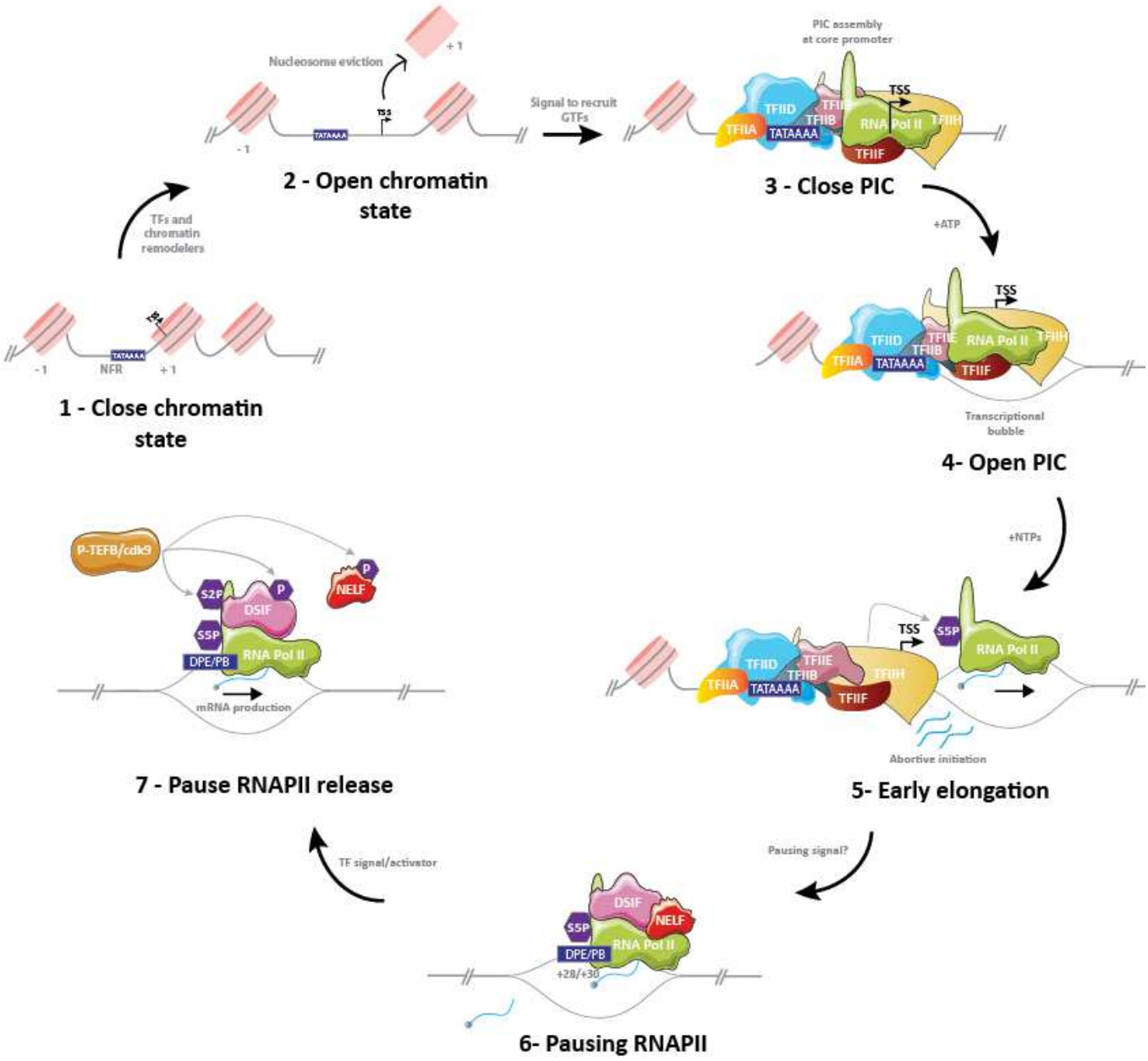


Figure 2. Stages of RNAP II initiation. 1, 2. The initiation of transcription needs the opening of the chromatin via Transcription Factors (TFs) and/or chromatin remodelers leading to the eviction of nucleosomes at the promoter region. 3. When the signal is given the GTFs are recruited at the promoter and subsequently the RNAP II, setting up the Pre-Initiation complex (PIC). 4. In the presence of ATP, the DNA is opened (forming the 'transcription bubble') 5. RNA synthesis commences with the dissociation of initiation factors. 6. At ~35 to 50 bp downstream of the TSS, the RNA synthesis is paused by the association of DSIF, NELF, to the RNAP II. 7. Activators give the signal to release the pause, P-TEFb via its cdk9 subunit phosphorylate DSIF, NELF and the RNAP II, enables the formation of the Pol II elongation complex, allowing RNA production to re-start. NFR, nucleosome free region; NTP, nucleoside triphosphate. DSIF, The DRB sensitivity-inducing factor; NELF, the Negative elongation factor; P-TEFb, the positive elongation factor b; cdk9, the cyclin-dependent kinase 9.

1.2. Key players of Transcription Initiation

RNAPII transcription cycle starts with the recruitment of the polymerase to promoters with an un-phosphorylated CTD (Lu *et al.*, 1991; Myers *et al.*, 1998). Although RNAPII can carry out the unwinding of DNA and polymerization of the RNA molecules on its own. *In vitro*, biochemical complementation assays showed that it needs other proteins, the General Transcription Factors (GTFs), to specifically recognize the promoters and start transcription (Orphanides, Lagrange and Reinberg, 1996).The complex formed by RNAPII and the GTFs is referred to as the Pre-Initiation Complex (PIC).

1.2.1 The RNA Polymerase II

RNAPII was initially assumed to solely transcribe mRNAs or stable RNAs. However, with the recent advances in high throughput genomic technologies, it was shown that RNAPII mediated transcription is not limited to mRNAs but also includes a fraction of small nuclear and nucleolar RNAs (snRNAs and snoRNAs, respectively), short Hairpin RNA (shRNA), enhancer RNAs (eRNAs), Cryptic Unstable Transcripts (CUTs) etc.(Cech and Steitz, 2014)
Concerning its structure, X-ray crystallography enabled the identification of 12 subunits in the human Pol II, named RBP1 to RBP12. The largest subunit RBP1 contains a C-terminal domain, hereafter referred to as the CTD, which is not present in the other 2 types of polymerases (Pol I and Pol III). This domain bears a heptapeptide sequence consisting of Y1-S2-P3-T4-S5-P6-S7 that is repeated in tandem (23 copies in *Dictyostelium*, 44 copies in *Drosophila*, 52 copies in *Zebrafish*, 52 copies in mammals). Most importantly, the post-translational modifications of the CTD were described to be crucial for the coordination of the entire transcription cycle, starting from the transcriptional initiation, pausing, elongation and possibly termination, to the processing of the nascent RNA (Eick and Geyer, 2013) (Figure 3).

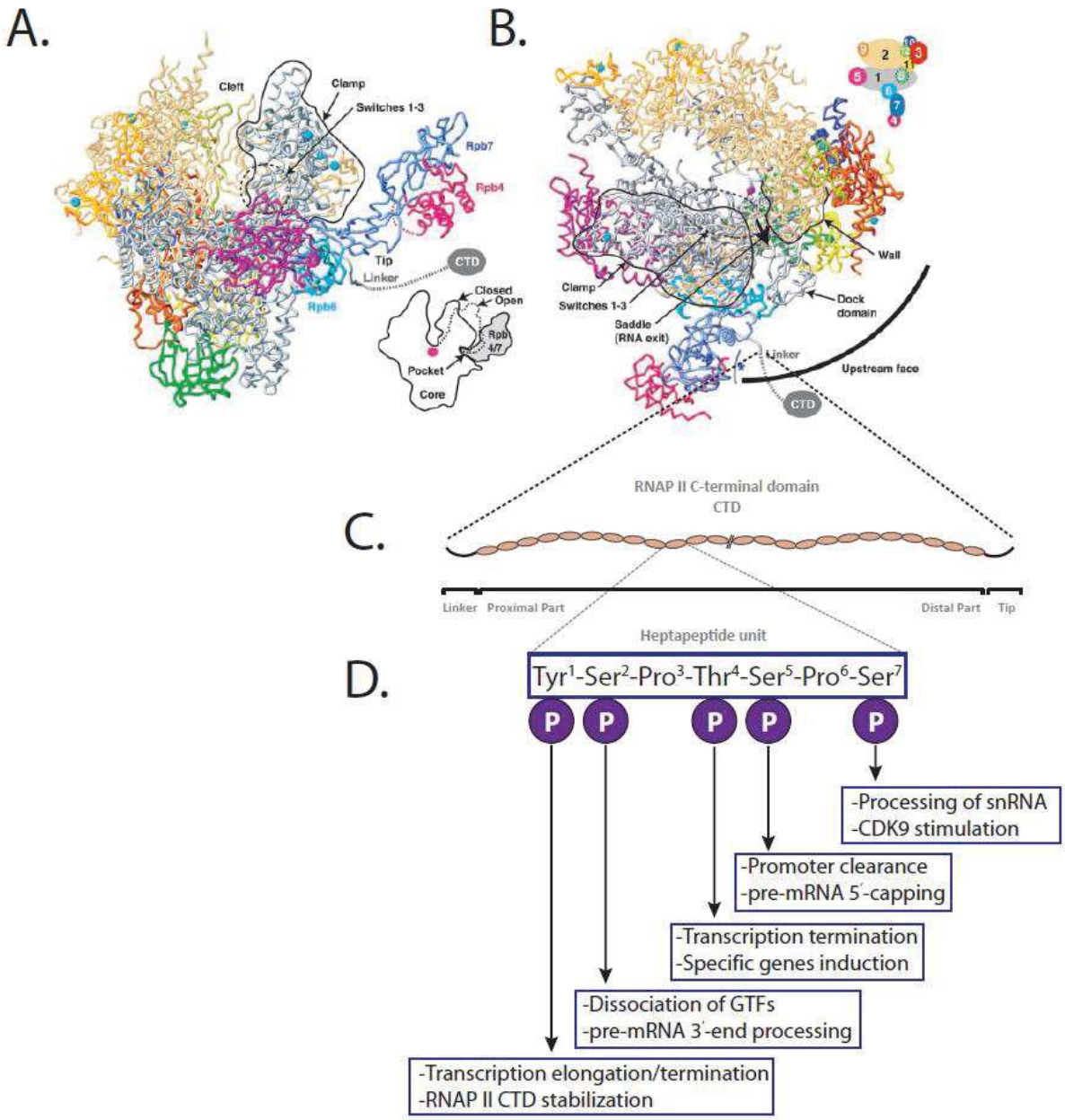


Figure 3. RNAP II structure and CTD functions. A. Ribbon presentation of the "front" view RNAP II (12 subunits). Right bottom corner representation of open and close configuration of RNAP II. B. Pol II upstream interaction face. Shown in a view of the model from the ''top''. The circle segment is centred at the active site and has a radius that corresponds to the minimal distance between the TATA box and the transcription start site (85 Å, _25 bp). The saddle between the wall and the clamp and the assumed direction of RNA exit are indicated. A blue asterisk indicates a potential RNA-binding face of Rpb7. A key to subunit colour is shown in the upper right corner. A and B are taken from (Armache et al., 2003). C. Cartoon of the CDT structure each oval circle represents an heptapeptide repetition. D. Constitution of the heptapeptide unit and its functions. Adapted from (Srivastava and Ahn., 2015).

1.2.1.1 Functions of the CTD

The CTD is dispensable for the activity of RNAPII *in vitro* (Zehring *et al.*, 1988; Kim and Dahmus, 1989) as the catalytic RNAP activity lies in the core of RPB1, but essential *in vivo* as deletions of the entire CTD are lethal in both yeast and mammalian cells in culture. Interestingly, the deletion of less than 50% of the CTD repeats is generally well tolerated (Chapman, Conrad and Eick, 2005) suggesting that there is a certain degree of redundancy in the essential functions that the CTD repeats ensure. It coordinates the precise spatiotemporal recruitment of the factors that are required for each step of the transcription cycle from promoter binding to transcription initiation, elongation, and termination. It also recruits factors that are responsible for co-transcriptional events such as nascent RNA processing (5'capping, splicing, polyAdenylation…) and chromatin remodeling (histones PTM's writers & erasers). Finally, the RNAPII CTD is important for post-transcriptional processes such as mRNA export (Meinel *et al.*, 2013), as well as transcription independent events like some forms of stress response (Heine, Horwitz and Parvin, 2008) and metaphase progression during the cell division (Hintermair *et al.*, 2016). The CTD Pol II reported functions are summarized in Figure 3, panel D.

1.2.2 The general transcription factors (GTFs)

Highly conserved from yeast to humans, the general transcription factors are necessary for the transcription of any gene by their specific polymerase, Pol I, Pol II and Pol III. There are six RNAP II GTFs, named TFIIA, B, D, E, F, and H. TFII stands for Transcription Factor of RNA Pol II, and A to H corresponds to the order of elution/discovery by separation using running column chromatography (Matsui *et al.*, 1980). Biochemical and structural studies summarized in (Sainsbury, Bernecky and Cramer, 2015) proposed a sequential model where the first GTFs to recognize and bind core promoter element is TFIID, followed by the binding of TFIIA and TFIIB, next the recruitment of the complex RNAP II-TFIIF, and finally the binding of TFIIE and TFIIH. This model was supported by single molecule imaging in living human cells (Zhang *et al.*, 2016). Known GTF and their functions are summarized in Table 1.

1.2.2.1 TFIID

The GTF TFII is a conserved protein complex of 750 KDa, present in Archaea and eukaryotes. The TFIID complex contains 15 subunits in yeast and 14 in humans and in *Drosophila* (Table 1), classified into The TATA Binding protein (TBP) and the TBP-associated factors (TAFs) respectively. TFIID complex is the first GTFs to bind the core promoter for nucleate the PIC. Thus, TFIID complex is involved in promoter recognition, PIC assembly and RNAP II recruitment. In yeast, 90% of promoters are regulated by TFIID and 10% by the SAGA complex (Huisanga and Pugh 2004). TFIID and SAGA are very similar in terms of structure and function. Furthermore they have common subunits. As well as TFIID, the SAGA complex allow the recruitment of TBP to the promoter by interacting with transcriptional activators.

The TATA Binding protein (TBP) is necessary for RNAP II transcription *in vitro* and in *vivo,* from promoters containing conventional TATA elements as well as functionally distinct promoters that lack TATA-like sequences (Horikoshi *et al.*, 1990; Cormack and Struhl, 1992). Initially, TBP was thought to bind specifically the TATA box sequence TATAWAWR (W= A or T; R= A or G), located in the core promoter between (-31 to -24 bp) in humans and drosophila (Haberle and Stark, 2018). However, the vast majority of the promoters do not contain canonical TATA box motif but are still bound by TBP. It has been shown indeed that TBP can

bind variants of TATA box DNA sequences (Coleman and Pugh, 1995). Structural analyses revealed TBP as a saddle-shape molecule, with a concave and convex surface. The concave surface of TBP is hydrophobic and interacts with the minor groove of the TATA box, inducing a ~ 90-degree bend in the DNA. The convex surface of TBP serves as a binding platform for several regulators, which can activate or repress TBP for DNA binding (Patel *et al.*, 2018; Bhuiyan and Timmers, 2019).

Sorting of cryo-EM images of promoter-bound human TFIID revealed five major states (canonical, extended, scanning, rearranged, engaged) that would correspond to different stages of DNA engagement by TBP (Patel et *al*. 2018). These observations led Nogales and coworkers to propose a scanning model for the "delivery" of TBP to DNA at the TATA box position. Timmers and coworkers (Bhuiyan and Timmers, 2019) propose that the TFIID recognition to the core promoter is independent to the TATA box. Rather they list four possibilities of TFIID recruitment on promoter: i) nucleosome histone modification, ii) promoter-DNA interaction with TAF1, TAF 2 and TAF4, iii) interactions with the Mediator complex and finally iv) recruitment of TFIID via activator-TAF interactions involving enhancer-promoter looping. In addition, they propose that the rate-limiting step in TFIID recruitment could be different for each RNAP II promoter. Nevertheless, full engagement of TBP with the TATA box is only achieved after TFIIA recruitment, likely via TAF4-12 interactions. Mutation of TATA box prevents progression to the DNA-engaged state (Patel *et al.*, 2018), suggesting that functional PIC formation on non-TATA promoters depends on additional factors (SAGA complex in yeast, TRF1 or TRF2 in Drosophila, Mediator complex etc) or that TFIID remains in the rearranged state to direct transcription initiation from non-TATA promoters *in vivo*. In bilateria three additional factors, having similar functions to the TBP has been described (Goodrich and Tjian, 2010; Akhtar and Veenstra, 2011), namely the TBP-related factor TRF1, TRF2 and TRF3.

The TRFs: the first identified TRF, was TRF1, discovered in *Drosophila* (Crowley *et al.*, 1993). So far, TRF1, it has only been found in insects. It can bind the TATA box with TFIIA and TFIIB and substitute TBP in the transcription of some promoters *in vitro* (Holmes and Tjian, 2000). TFR2 was found in bilateria, and unlike TBP or TRF 1 does not bind to the TATA box and does not appear to have any sequence specific DNA-binding activity (Dantonel *et al.*, 2000). This suggested that TFR2 might be involved in mediating transcription at TATA-less promoters

(Duttke *et al.*, 2014). The loss of TRF2 is embryonic lethal in *Drosophila* (Kopytova *et al.*, 2006), Zebrafish (Danio rerio) (Müller *et al.*, 2001)), and Xenopus (Veenstra, Weeks and Wolffe, 2000). In contrast, TRF2-deficient mice are viable but have a defect in spermiogenesis (Martianov, Viville and Davidson, 2002; Torres-Padilla and Tora, 2007)). TRF3 is found in vertebrates. It can bind to the TATA box. It interacts with TFIIA and TFIIB and it mediates transcription *in vitro* (Bártfai *et al.*, 2004). TRF3 is required for normal development in Zebrafish and Xenopus (Bártfai *et al.*, 2004). In mice the loss of TRF3 results in female sterility, due to the requirement of TRF3 for the differentiation of the female germ cells (Gazdag *et al.*, 2009).

The TBP-associated factors (TAFs): the TAFs are members of the TFIID complex. Early *in vitro* assays showed that some of the TAFs directly and specifically interact with the initiator and the downstream promoter elements (Burley and Roeder, 1996).They can stimulate or prevent the binding of TBP to the DNA. TAF1 can interact with acetylated histone via its bromodomain (BrDs) (Bhuiyan and Timmers, 2019) and TAF3 with the trimethylated lysine-4 of histone H3 (H3K4me3) via its PHD finger (Bhuiyan and Timmers, 2019). Recently TFIID has been implicated in the recognition of a novel chromatin mark, serotonylated glutamine-5 of histone H3 (H3Q5ser), which appears to enhance TFIID-H3K4me3 binding (Bhuiyan and Timmers, 2019). The extend of the enhanced H3K4me3 binding, the molecular mechanism, and the TAFs involved remain to be determined. The data suggested a role of TAFs in remodeling promoter nucleosomes to facilitate TFIID recruitment. TAFs can also interact with TFs to recruit the PIC at the promoter and activate transcription via their histone-fold domains. They have been involved in promoter selectivity by recognizing core promoters elements like the INR (TAF1/TAF2), DCE(TAF1), DPE (TAF6/TAF9) (Danino *et al.*, 2015; Vo Ngoc *et al.*, 2017), the MTE (TAF1/TAF2).

1.2.2.2 TFIIA

The role of TFIIA as a general transcription factor is debatable, as it was shown that promoters vary widely in terms of their requirement for TFIIA for transcriptional activation (Høiby *et al.*, 2007). TFIIA contains two conserved domains, a 4-helix bundle and a 12-stranded β-barrel (Sainsbury, Bernecky and Cramer, 2015). TFIIA has four principal functions on transcription initiation. The first function of TFIIA is its role of anti-repressor: TFIIA bind transcription inhibitors that mask the site of interaction of TBP-TATA box, such as NC2, TAF1, BTAF1. The second main role of TFIIA is to stabilize the complex TBP-DNA, via the TFIIA interaction with the convex surface of TBP, and a DNA region upstream of the TATA box. This increases the fraction of TBP-bound to the DNA and increase the probability of the TFIID complex to interact with the DNA (Cianfrocco *et al.*, 2013). Third TFIIA could facilitate the nucleation of the PIC via the activation of TFIIE and TFIIF (Langelier *et al.*, 2001). Finally, TFIIA plays a role as transcriptional co-activator by binding specific TF that will facilitate the PIC formation, mainly at TAF independent genes (Kobayashi, Boyer and Berk, 1995).

1.2.2.3 TFIIB

TFIIB is required for the recruitment of the RNAP II to the promoter (Ha *et al.*, 1993) and it facilitates TBP binding to DNA and DNA bending (Zhao and Herr, 2002). Highly conserved from archaea to eukaryotes, TFIIB is necessary for RNAP II transcription. It is a polypeptide of 33KDa, which contains four functional domains called, B-ribbon, B-reader, B-linker, and B-core. Each of them is in contact with the RNAP II. TFIIB works as a bridge to link TBP and the RNAPII (Chen and Hahn, 2004). The functions of TFIIB in the recruitment of RNAP II and in the interaction of TBP with the promoter are mediated by its N- and C-terminal domains. TFIIB interacts with the flanking regions of the TATA box that can contain TFIIB recognition elements (BREs) setting the orientation of the PIC nucleation (Deng and Roberts, 2005). In addition to orientating the PIC, TFIIB recruits the complex RNAP II-TFIIF and set the TSS position for transcription to start (Fishburn and Hahn, 2012). After recruitment of the RNAP II-TFIIH complex, TFIIB region connecting the B-ribbon and B-core domains traverse the RNAP II cleft to form two distinct elements, the B-reader and B-linker. The B-linker helix is involved in DNA opening (melting of DNA strand) and in the maintenance of the transcription bubble

(Kostrewa *et al.*, 2009). The B-reader loop binds to the template DNA for the recognition of the initiator sequence and start the RNA synthesis. The RNA synthesis appears stimulated by TFIIB by allosteric rearrangements of residues in the active site and the stabilisation of a close polymerase clamps (Sainsbury, Niesser and Cramer, 2013). TFIIB stabilizes an early initiation complex with a five-nucleotide RNA strand, while the B-reader loop blocks the path of the RNA beyond six nucleotides, helping DNA-RNA strand separation by directing the RNA to the exit tunnel (Sainsbury, Niesser and Cramer, 2013). Last, TFIIB is released from the complex when the RNA reaches a length of 12–13 nucleotides and clashes with the B-ribbon (Pal, Ponticelli and Luse, 2005). The interplay between TFIIB and nucleic acids is critical for the initiation-to-elongation transition.

1.2.2.4 TFIIF

TFIIF forms a stable complex with RNAP II before its integration into the PIC. TFIIF is a heterodimer of subunits TFIIFα (also known as RAP74) and TFIIFβ (also known as RAP30) or Tfg1 and Tfg2 in yeast (Chafin, Claussen and Price, 1991). Around ~50% of RNAPII is associated with TFIIF in yeast. TFIIF prevents non-specific interaction of RNAP II with DNA and stabilizes the PIC, in particular by stabilizing TFIIB within the PIC (Fishburn and Hahn, 2012). TFIIF also influences TSS selection, stimulates phosphodiester bond formation and early RNA synthesis (Ren, Lei and Burton, 1999)). TFIIF further contributes to the stabilization of the transcription bubble. Transcription can be initiated to some extent *in vitro* in absence of TFIIE and TFIIH but not if TFIIF is absent (Pan and Greenblatt, 1994).

1.2.2.5 TFIIE

TFIIE facilitates the recruitment of TFIIH to the PIC, and acts as a bridge between the RNAP II and TFIIH. It is a heterodimer composed of TFIIEα and TFIIEβ (Tfa1/2 in yeast) of 56 KDa and 34 KDa respectively. TFIIEα contains an N-terminal winged helix (WH) domain, a central zinc-finger domain and a C-terminal acidic domain, all highly conserved from bacteria to humans. They interact with the RNAP II but also with TBP, TFIID and TFIIH (Maxon, Goodrich and Tjian, 1994). The N-terminal half of TFIIEα is sufficient for interaction with TFIIEβ and transcription functions (Sainsbury, Bernecky and Cramer, 2015). TFIIEβ contains two WH domains and a basic C-terminal region only present in the eukaryotes. It works as a place for the interaction with the DNA and a plethora of proteins, mostly GTFs but also with TF such as Antenapedia and Abdominal-B in *Drosophila* (Zhu and Kuziora, 1996; Thomas and Chiang, 2006). Once recruited on the PIC, TFIIE stimulates the enzymatic activity of TFIIH subunits, necessaries to enter in competent phase of transcriptional elongation (Sainsbury, Bernecky and Cramer, 2015).

1.2.2.6 TFIIH

TFIIH is a protein complex, composed of 10 subunits. Three of these subunits have enzymatic properties, among which the ATPase XPB (known as Ssl2 in yeast) is required for promoter opening *in vitro* (Holstege, van der Vliet and Timmers, 1996) and *in vivo* (Guzmán and Lis, 1999)).

The six-subunit core module of TFIIH comprises XPD, another ATPase (Rad3 in yeast) that possess helicase activity and exhibit 3'–5' and 5'–3' directionality, p62 (Tfb1 in yeast), p52 (Tfb2 in yeast), p34 (Tfb4 in yeast), p8 (Tfb5 in yeast) and p44 (Ssl1 in yeast). The helicase function of XPD is required for DNA opening in the repair pathway (Coin, Oksenych and Egly, 2007). In addition, the three-subunits kinase module Cdk7-cyclin H-MAT1 (Kin28-Ccl1-Tfb3 in yeast), which targets the CDT domain of the RNAP II (phosphorylate the serine 5 of the CTD) are necessary for the clearing of the RNAP II from the promoter, the beginning of early elongation and the recruitment of capping factors (Sainsbury, Bernecky and Cramer, 2015).

Mutations in the genes encoding XPB and XPD are associated with the human diseases xeroderma pigmentosum, trichothiodystrophy and Cockayne syndrome (Egly, 2011).

Factor	Subunits	Function
RNAP II	RPB1 to RPB12 (12)	Transcribing enzyme
TFII A	α/β (2)	TBP stabilization and counteracts repressive effects of negative co-factors
TFIIB	TFIIB (1)	Pol II recruitment, TBP binding and TSS selection
TFIID	TBP; TAF 1 to 15* (14-15)	Pol II recruitment and promoter recognition
TFIIE	α/β (2)	recruitment of TFIIH and open DNA stabilization
TFIIF	α/β (2)	TSS selection and stabilization of TFIIB
TFIIH (core)	1 (p62); 2 (p44); 3 (p34); 4 (p52); 5 (p8); XPD subunit: ATPase; DNA repair; XPB subunit: ATPase; promoter opening (10)	Promoter opening and DNA repair
TFIIH (kinase module)	Cyclin H; CDK7; MAT1 (3)	CTD phosphorylation

Table 1. Sum up of the RNAP II and the GTFs components functions.

1.2.3 The Mediator complex

Subsequent research revealed numerous protein complexes that act as transcriptional co-regulators. One of the most studied co-regulators is the Mediator complex. It is an evolutionary conserved, multi-subunit, complex that is generally required for transcription by Pol II. It comprises up to 30 subunits in humans that are organized in 4 modules: the head, middle, tail and the kinase module which is not permanently associated to the rest of the complex (Table2) (Soutourina, 2018). The Mediator does not appear to have DNA-binding ability; it is rather recruited to chromatin via TFs bound to enhancers and, as implied by its name, serves as a link or a bridge between these TFs and the basal transcriptional machinery at the target gene promoter. A high enrichment of the Mediator was also noted at the so-called super-enhancers (Whyte *et al.*, 2013). The interaction between the Mediator and the TFs implicates, but is not restricted to, the tail module (Soutourina, 2018). Interestingly, electron microscopy studies revealed substantial structural shifts of this complex upon its binding to TFs, which spread throughout the whole complex without affecting its composition. The functional role of these structural changes are, however, still unclear. A model has been proposed whereby this complex is first recruited to the enhancers through

the TFs, then helps with the assembly of the transcriptional machinery (Jeronimo and Robert, 2014). Once RNAP II becomes phosphorylated on the S5 and escapes the promoter, it will dissociate from the transcription complex. The interaction with RNAP II involves both the head and the middle modules of the Mediator. Concerning the kinase module, it was shown in yeast to be recruited with the Mediator to enhancers. Subsequently, its dissociation from the enhancer complex is important to allow the association of the Mediator with RNAP II and the transcriptional complex, underlining repressive properties of this module. On the other hand, *in vitro* studies showed that Cdk8, belonging to the kinase module, is also capable of phosphorylating and activating RNAP II, providing evidence for a positive regulatory role of this module in transcription (Galbraith, Donner and Espinosa, 2010). Overall, these data point to the intricacy of studying this complex, due to its very dynamic and transient binding.

Factor	Subunits	Function
Mediator Head	Med6, 8, 11, 17, 18, 20, 22, 30	Promotes assembly and/or stabilization of PIC
Mediator Middle	Med1, 4, 7, 9, 10, 14, 19, 21, 31, 26	Interacts with Pol II together with the head Med1 is the target for numerous TFs, in particular nuclear receptors; Med14 acts as a scaffold for all three modules
Mediator Tail	Med29, 27, 24, 15, 16, 23	Connects core Mediator to DNA-binding TFs
Mediator Kinase	Med12L, 13L, CDK19/ CDK8, CycC	Reversibly associated with Mediator. Involved in transcriptional repression and activation.
Mediator Unassigned	Med25, 28	Location in the Mediator unknown. Interact with Tail and Head subunits

Table 2. Mediator subunits and functions.

1.3 Features of RNAP II Initiation

1.3.1 Focused versus Dispersed Initiation

Metazoan promoters either initiate transcription at precise positions (single sites or a narrow ~5nt), or at dispersed regions (over a 50-100 nt region), having direct consequences on gene expression (Schor *et al.*, 2017). Where RNAPII will start depends mainly on the sequence of the core promoter (Danino *et al.*, 2015; Haberle and Stark, 2018; Vo Ngoc, Kassavetis and Kadonaga, 2019).

Focused transcription initiation is associated with regulated genes and core promoter composition. Typically focused promoters often contain core promoter motifs such the TATA box, Inr, motif ten element (MTE) and or the DPE (Vo Ngoc *et al.*, 2017; Vo Ngoc, Kassavetis and Kadonaga, 2019). In contrast, dispersed promoters are associated with Ohler core promoter motifs 1,6,7 (Ohler *et al.*, 2002) and DNA replication-related element sequences (DRE). In *Drosophila*, developmental genes tend to exhibit a focused promoters while housekeeping genes are usually associated with a dispersed initiation (Ohler *et al.*, 2002; Haberle and Stark, 2018).

1.3.2 Bidirectional and divergent transcription

Bidirectional transcription is defined by the presence of two transcriptional events in both sense and antisense orientations within less than 1kb. It was detected in promoters as well as in enhancers and is pervasive across species including yeast, *C. elegants*, *M. musculus* and *H. sapiens* (Seila *et al.*, 2008), though far less common in *D. melanogaster* (Core *et al.*, 2012). However, two recent studies demonstrate that the proportion of divergent transcription in *D. melanogaster* is higher than reported before (Ibrahim *et al.*, 2018; Rennie *et al.*, 2018). For promoters, antisense transcripts are known as promoter upstream transcripts (PROMPTs) or upstream antisense RNAs (uaRNAs). Divergent transcription from core promoters is initiated from two differentially orientated TSSs (Figure 4). The PROMPTs are non-protein-coding and highly unstable. In the case of enhancers, transcription initiation is often bidirectional (Core *et al.*, 2014; Andersson and Sandelin, 2019), and also initiates from two distinct sites. However, unlike gene core promoters, enhancer produce short, unstable transcripts in both

senses. The function (if any) of divergent transcription is not yet fully understood. Some studies (Kim and Shiekhattar, 2015; Andersson and Sandelin, 2019) propose that divergent transcription at enhancers and promoters reflects a common architecture in DNA sequence for transcription initiation, but only promoters have evolved to produce stable transcripts. In contrast, Ibrahim and collaborators suggest that promoter and enhancer architecture is quite different based in their regulatory marks (e.g; H3K4me3 levels) and motif sequence predictions, making a clear-cut between both regions. So forth, arguing against a unified eukaryotic model.

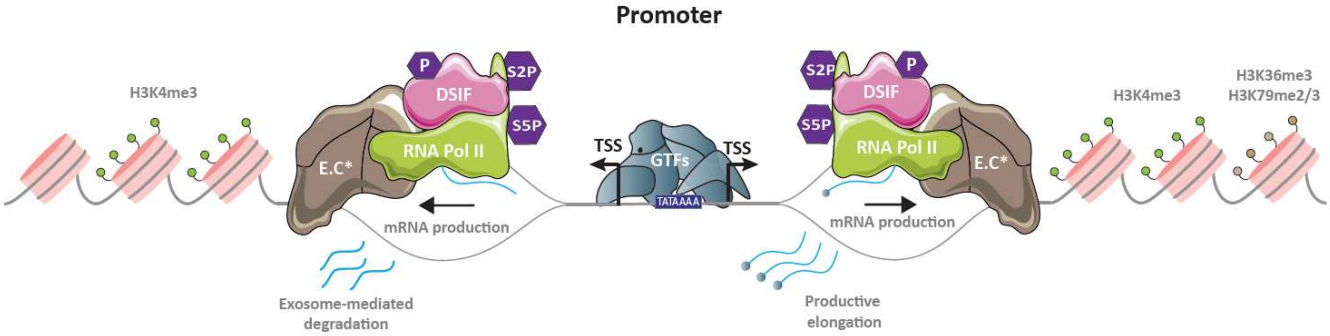


Figure 4. Divergent transcription at promoters. Sense and anti-sense transcription at promoters are mediated through the transcription machinery assembled independently at the same core promoter. H3K4me3 is highly enriched at these promoters. The ser-5P form of RNAP II is engaged in upstream anti-sense transcription, but it is not known whether ser-2P of RNAP II occurs during the anti-sens RNA elongation. In contrast, the anti-sense transcription produces RNAs that are degraded by the exosome. And mRNA synthesis from sense transcription is stabilized by the presence of 7-methyl guanosine cap structure (no shown here).

1.4. The RNAP II Elongation

For high eukaryotes, the RNAP II elongation process is divided in two phases: early elongation and productive elongation.

1. The early elongation consists in the production of a small RNA ending around +30 and +50 bp downstream of the TSS. It is followed by a more or less transient arrest of the elongation process, referred to as 'pausing'. (Adelman and Lis, 2012). During pausing, the C-terminal heptapeptide repeat domain (CTD) of its largest subunit (Rpb1) is phosphorylated on serine 5 residues by cdk7 kinase of the TFIIH complex. Pausing is triggered by two major factors: the DRB sensitivity-inducing factor (DSIF) and the Negative elongation factor (NELF) (Gaertner and Zeitlinger, 2014); (Figure 2). Using a variety of approaches, several studies have assessed the stability of paused Pol II in *Drosophila* or mammalian cells (Henriques *et al.*, 2013; Buckley *et al.*, 2014). These studies revealed that paused Pol II is relatively stable, with an average half-life of more than six minutes. Such a stably paused Pol II would be beneficial for synchronous induction of expression across many genes (Boettiger and Levine, 2009; Lagha *et al.*, 2013), and allow a temporal window to integrate signals and coordinate binding of signal-responsive transcription factors. Moreover, studies in *Drosophila* and mammalian cells have suggested that the intrinsic property of promoters might contribute to Pol II pausing. Promoter elements such as the GAGA motif, the downstream promoter element (DPE), the "pause button", and the TATA box have all been positively or negatively linked to Pol II pausing (Hendrix *et al.*, 2008; Nechaev *et al.*, 2010)

More recently, this classic view was challenged by two methods Chip Nexus (Shao and Zeitlinger, 2017) and single molecule footprint (Krebs *et al.*, 2017) which both revealed a much more dynamic and transient view of pausing. For example the half-life (or residence time) of RNAPII pausing at the Hsp70 gene was found to be <2.5 min in the basal state (no heat shock induction) (Krebs *et al.*, 2017). Faster than previously reported for this gene (Buckley *et al.*, 2014)

2. The productive elongation: during productive elongation, RNAPII is released from the paused state and proceeds to transcribe the body of the gene. This release requires the phosphorylation of DSIF, NELF and the RNAPII itself by the positive elongation factor b (P-TEFb). P-TEFb is recruited to the promoter by TF and cofactors, passing from an inactive configuration to an active one. P-TEFb is composed of different subunits, among which the cyclin-dependent kinase 9 (CDK9) which phosphorylates NELF, leading to its release from the elongation complex. At this step, DSIF becomes a positive elongation factor, and the RNAPII is phosphorylate at the serine 2 position in the CTD (Figure 2).

Elongation rates can vary between and within genes and it exists an important variety of rates across genes and for the same gene across various cell lines. (Jonkers and Lis, 2015; Tutucci, Livingston, *et al.*, 2018). On average, in *Drosophila* cells, Pol II elongation rate is in the range of~ 2.4-3.0 kb/min (Fukaya, Lim and Levine, 2017), while it has been estimated to be on the range of ~ 2-5 kb/min in mammalian cells (Saponaro *et al.*, 2014). Multiple factors impinge on this speed such as splicing factors, histones marks, and post-transcriptional regulations (Jonkers and Lis, 2015; Tutucci, Livingston, *et al.*, 2018).The resulting nascent transcripts need to be processed to prevent them from degradation. This achieved early on in the process of early elongation, through the 5’ capping complex and the spliceosome.

1.4.1 Promoter proximal pausing

RNAP II pausing was identified in *Drosophila* heat-shock and human c-myc genes (Gilmour and Lis, 1986; Krumm *et al.*, 1992). Although RNAP II pausing was originally considered to be restricted to few genes, nowadays it appears to be a common step in metazoan transcription (Muse *et al.*, 2007). Genome-wide RNAP II Chromatin immuno-precipitation (ChIP) studies revealed a concentrated RNAP II signal at around 20-60 bp downstream of the TSS, higher than that of RNAP II on the gene body. This accumulation, known as "RNAPII pausing" is observed in over 30% of all genes in metazoans (Adelman and Lis, 2012). This percentage however may largely vary between studies according to the pausing index (index= Pol II-promoter signal/ Pol II-gene body signal) defined (Chen et al., 2013).

RNAP II pausing primarily occurs at stimulus-responsive genes, such as those involved in development, stress and damage response and cell proliferation. The establishment of paused RNAP II requires two main factors: the DRB Sensitivity-Inducing Factor (DSIF) and the Negative Elongation Factor (NELF) complexes, which maintain the RNAP II at the pause site waiting for a signal to engage productive elongation. It was proposed that promoter proximal pausing serves to keep the promoters in an open conformation by competing with the promoter positioned nucleosomes (Gilchrist *et al.*, 2010). In the context of a developing embryo, pausing has been associated to coordinate gene expression across a field of expressing cells (Lagha et al,. 2013).

1.5 Termination

Once Polymerase reaches a polyA site, transcription stops: this is the termination process. During termination, Pol II and the nascent RNA are released from the DNA template. Definitive mechanisms of termination are still lacking, however two alternative models but not exclusive have been proposed.

The Allosteric model postulates that transcription through the polyadenylation signal (PAS) leads to a destabilization of the elongation complex and to the recruitment of termination factors, such as the cleavage and polyadenylation complex (CPA) (Proudfoot, 2016). The CPA complex is fairly large (>100 proteins) and its interaction with the nascent mRNA and the RNAP II, is thought to slows it down until termination without prior cleavage of the transcript. On the other hand, the Torpedo model proposes that once transcribed, the PAS induces downstream cleavage of the RNA. In this view transcription persists downstream of the PAS, but the Pol II speed decelerated in this 'termination zone' by the presence of G rich pause sites. An exonuclease, RAT1 in yeast, XRN2 in mammalians, recognizes the uncapped 5'end of the new transcript attached to the polymerase and degrades it, thanks to the lower transcription rate after PAS pause, the exonuclease eventually catches up with the transcribing complex inducing its dissociation (Rosonina, Kaneko and Manley, 2006). It is considered now, that what happens in terms of transcription termination is in some middle ground between both models, where termination by the torpedo model is concomitant with a conformational change of RNAP II triggering its release from the template (Rosonina, Kaneko and Manley, 2006).

2. *Cis*-regulatory DNA elements: Enhancers and the core promoters

In this chapter, I will focus in the features of core promoters, but first I will describe briefly, the definition of enhancers. Which coordinate activity with the core promoter is necessaire to achieve proper gene expression.

2.1. Enhancers

Enhancers are DNA sequences located often distally from promoters, and contain binding sites for transcription factors. In turn, transcription factors recruit cofactors, which modify the nearby chromatin and lead to transcriptional activation from a core promoter independent of their relative distance and orientation (Shlyueva, Stampfel and Stark, 2014).

In multicellular organisms, enhancers are responsible for the precise control of spatiotemporal patterns of gene expression. Initially named "modulators" enhancer elements were discovered in the early 80s in functional test of sea urchin histone gene expression in the *Xenope oocyte,* where DNA sequences located upstream of the TATA box motif positive influence H2A gene transcription (Grosschedl and Birnstiel, 1980). Deletion of the modulator activity resulted in 15- to 20 fold decrease in H2A expression, and the activity of the modulator was retained even when its DNA sequence was inverted. Similarly, the tandem of 72 bp DNA repeats located upstream of viral SV40 gene were found to be indispensable for SV40 gene expression (Benoist and Chambon, 1981). The first mammalian cellular enhancer identified was the enhancer from the efficient expression of the immunoglobulin (Ig) heavy-chain (Banerji, Olson and Schaffner, 1983; Gillies *et al.*, 1983; Neuberger and Calabi, 1983), these studies provided the first evidence of tissue or cell specific activity of enhancers. In yeast, transcriptional activation was found also to be mediated by enhancer-like sequences, known as upstream activation sequences (UASs), although their distances from the core promoter are much shorter than the distances found in mammals (Guarente, 1988). A series of studies on the SV40 enhancers established the conceptual framework for defining enhancers and their properties as follow: (1) Enhancers increase transcription of a linked gene from its correct initiation site specified by the core promoter, (2) enhancer activity is independent of orientation relative to its target gene, (3) enhancers can function independent of their

position relative to the target genes, and also over long distances, (4) enhancers can function with a heterologous promoter, (5) enhancers exhibit DNase I hypersensitivity (HS), which reflects a less compacted chromatin state as a result of the binding of various transcription factors. Although these properties were defined more than three decades ago, they are still widely used to classify enhancers (Banerji, Rusconi and Schaffner, 1981; Moreau *et al.*, 1981; Fromm and Berg, 1982; Khoury and Gruss, 1983; Atchison and Perry, 1988).

In the recent years, the dichotomy of enhancers and promoters has been challenged by the recruitment of RNAP II to gene enhancers and the bidirectional transcription at promoters (Haberle and Stark, 2018). Furthermore, CapStarr-seq experiments in several mammalian cell lines showed that 2-3% of coding-gene promoters have enhancer-like functions in a given cell line (Dao *et al.*, 2017). On the basis on these broad similarities, a unifying model has been proposed, in which promoters and enhancers are considered as a single class of regulatory elements, with a common architecture for transcription initiation (Core *et al.*, 2014; Andersson, Sandelin and Danko, 2015) whatever not all scientific community are in agreement with this model (Ibrahim *et al.*, 2018).

Nevertheless, unlike gene core promoters, transcripts from enhancers (eRNA) or from (PROMPTs) or upstream antisense RNAs (uaRNA), are often short and unstable leading to a fast degradation by the exosome. Suggesting that the difference of core promoters and enhancer rely in the capacity of core promoters to produce stable mRNA transcripts.

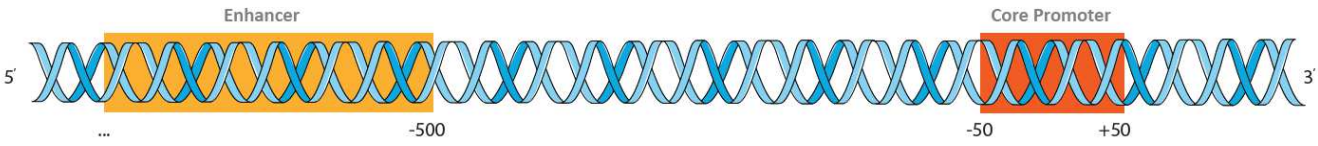


Figure 5. *Cis*-regulatory DNA modules. Core promoter are located around the TSS encompassing ~100 pb. Positioning of enhancers is less clear, they could be located near or far of their target gene.

2.2. The core promoter

The core promoter is defined as a specialized short DNA sequence (-50; +50) at transcription start sites (TSS) of protein coding and non-coding genes. It serves as a platform to assemble the PIC. They are composed of various combinations of sequences motifs highly conserved across evolution and can integrate specific regulatory cues from enhancers and regulatory proteins to control the transcription process. Core promoters are sufficient to direct transcription initiation but have low basal activity, figure 5.

Promoter: The promoter is defined as a genomic region encompassing a gene core promoter and an upstream proximal promoter, which autonomously could drive transcription (Haberle and Stark, 2018).

Proximal promoter: Is a transcription-activating sequence immediately upstream of the core promoter (typically up to 250 bp upstream of the TSS) that contains binding sites for sequence-specific transcription factors and functions like an enhancer (Haberle and Stark, 2018).

2.3. Enhancer- core promoter specificity

Enhancer-promoter pairs are commonly engaged by enhancer's looping, which physically brings these regulatory elements into proximity, through recruitment of multiple proteins (activators, co-activators, Mediator complex, cohesin and the PIC). Studies in recent years, employing advanced global methodologies such as chromatin conformation capture (3C), its derivatives (4C, 5C, Hi-C) and ChIA-PET, have led to the discovery of both intrachromosomal and interchromosomal physical contacts with promoters. While multiple enhancers can interact with multiple promoters (Arnold *et al.*, 2013), specificity between certain enhancers and promoters has been observed. The mechanisms that determine enhancer-promoter specificity are still poorly understood, but they are thought to include biochemical compatibility, constraints imposed by the three-dimensional architecture of chromosomes, insulator elements, effects of local chromatin environment and core promoter elements composition (van Arensbergen, van Steensel and Bussemaker, 2014).

The compatibility of enhancer-promoter interactions has mostly been studied in *Drosophila*. One of the early studies analysing the compatibility between enhancer-promoter pairs

examined the expression of the neighbouring gooseberry (gsb) and gooseberry neuro (gsbn) genes (Li and Noll, 1994). Swapping experiments revealed that although both enhancers (GsbE and GsbnE) are located between the two TSSs of the two genes (and thus cross activation could potentially occur), the GsbE could only activate the gsb promoter, while the GsbnE could only activate the gsbn promoter. Another study showed compatibility between the decapentaplegic (dpp) promoter and its enhancer, which only activates the dpp gene, but not other genes that are located closer to it (Merli *et al.*, 1996). High-throughput imaging of thousands of transparent transgenic zebrafish embryos (which were injected with about two hundred combinations of enhancer-core promoter pairs driving the expression of the GFP reporter gene), demonstrated the specificity of individual enhancer-promoter interactions and highlighting the importance of the core promoter sequence in these interactions (Gehrig *et al.*, 2009). Enhancer-promoter specificity mediated by promoter composition, was first demonstrated in transgenic *Drosophila* sister lines that contain a DPE- or a TATA-dependent reporter gene at precisely the same genomic position relative to the enhancer (Butler and Kadonaga, 2001). This study, identified enhancers that can discriminate between core promoters that are dependent on a TATA or a DPE motif. Also in *Drosophila* a study in promoter competition revealed that both the AE1 enhancer from the *Drosophila* Antennapedia gene complex and the IAB5 enhancer from the Bithorax gene complex preferentially activate TATA-containing promoters when challenged with linked TATA-less promoters (Ohtsuki and Levine, 1998). Nevertheless, both enhancers were able to activate transcription from a TATA-less promoter in reporters that lacked a linked TATA-containing promoter. Furthermore, Caudal a key regulator of the *Drosophila* Hox gene network, activates transcription with a preference for a DPE motif relative to the TATA-box (Juven-Gershon, Hsu and Kadonaga, 2008). Another study analysed the *Drosophila* dorsal-ventral developmental gene network, regulated by the transcription factor Dorsal, and discovered that the majority of Dorsal target genes contain DPE sequence motifs (Zehavi *et al.*, 2014). The DPE motif is functional in multiple Dorsal target genes, and mutation of the DPE leads to a loss of transcriptional activity. Moreover, the analysis of hybrid enhancer-promoter constructs of Dorsal targets reveals that the core promoter plays a pivotal role in the transcriptional output (Zehavi et al., 2014). Lately, the Stark lab developed a genome wide screen called STARR-seq (self-transcribing active regulatory region sequencing), where they identified thousands of

different enhancers, by testing their ability of activate transcription of a synthetic promoter containing four core promoter elements (TATA box, Inr, MTE and DPE) (Arnold et al., 2013). Following this study, the Stark lab analysed the compatibility between thousands of enhancers and two kind of core promoters, one from a gene ubiquitously expressed (Ribosomal protein gene 12 (RpS12)) and another from a developmentally regulated transcription factor (even skipped transcription factor) in *Drosophila melanogaster* S2 and ovarian somatic cells (OSCs). They show that enhancers show a marked specificity to one of two core promoters. Housekeeping enhancers are active across two cell types, while developmental enhancers exhibit strong cell-type specificity (Zabidi *et al.*, 2015). Moreover, core promoters sequences analysis showed a difference in the core promoters elements composition of developmental and housekeeping genes (Zabidi *et al.*, 2015), indicating a likely role of these motifs in the promoter-enhancer affinity. Another important question is how enhancer-promoter interaction can modulate the promoter activity, as different enhancers can activate transcription using the same promoter (Arnold et al., 2013). In order to address this question, the Stark lab developed the STAP-seq method. Basically, the authors cloned reported plasmids, carrying various selected core promoters (from developmental and housekeeping genes) of around 200 bp fragments size, within a strong developmental enhancer (the transcription factor Zn finger homeodomain 1 (zfh1) enhancer) and a protein-coding open reading frame. Then they transfected the plasmid library in *Drosophila melanogaster* S2 cells and quantified the transcripts that initiated from each candidate by deep sequencing (Arnold *et al.*, 2017). They demonstrate that the zfh1 developmental enhancer has a preference for developmental promoters rather than housekeeping promoters. In addition, TSSs of genes with five or more developmental enhancers (in endogenous context) were significantly more inducible than those with only one or two developmental enhancers. When they restricted the analysis to core promoters that contain only TATA box, Inr, MTE, or DPE motifs (i.e., those that preferentially function with developmental enhancers), they found that the most responsive core promoters were enriched near genes coding for transcription factors, whereas weak ones were predominantly near genes for cell-type-specific enzymes. This suggests that highly responsive non-housekeeping core promoters might regulate genes that require rapid induction (e.g., transcription factors), whereas weakly responsive ones could be employed at genes with

potentially lower transcription kinetics (e.g., enzymes). Together, these results suggest that core promoters with different levels of enhancer responsiveness are employed for the transcription of genes with different functions and different regulatory characteristics.
Furthermore, these results also, demonstrate distinct compatibilities of enhancers to their cognate promoters and the importance of the core promoters and its elements in the regulation of enhancer-promoter interactions, and the consequent transcriptional activity.

2.4. Identification of the core promoter

The identification of the core promoter region depends upon the correct identification of the TSS or the TSSs. Promoter elements can be identified by comparing evolutionarily distant genomes looking for regions of conservation upstream of annotated genes but this do not describe the precise activation point or TSS (Mouse Genomic Sequencing Consortium, 2002). For the moment, new technologies continue to improve the accuracy to detect TSS at the single base pair level (Box 1, no exhaustive list). Up to now, the best approach for the mapping of TSSs in cells involve the determination of the 5' ends of capped nascent transcripts (Shiraki *et al.*, 2003; Nechaev *et al.*, 2010; Kwak *et al.*, 2013; Sloutskin *et al.*, 2015). For example, Shiraki and collaborators at the RIKEN genomic science center, developed the cap analysis gene expression (CAGE), which allows high-throughput identification of sequence tags corresponding to 5'ends of mRNA at the cap sites and the identification of the TSS. The method essentially uses cap trapper full-length cDNAs, to the 5'ends of which linkers are attached. This is followed by the cleavage of the first 20 base pairs by restriction enzymes, PCR, concatamerization, and cloning of the CAGE tags. CAGE tags are sequenced and the libraries mapped to the genome for TSS identification at the single nucleotide resolution. The main advantage of this technique is that it is low cost in comparison with full-length cDNA library sequencing it has a much throughput of identified tags. This allowed a higher coverage in different studies providing a better view of promoters and TSS distributions.

Box 1|**Experimental methods to determine TSS position**

The conventional hallmark of TSSs in most eukaryotes is addition of a 7-methyl guanosine cap structure to the 5'-triphosphate of the first base transcribed by RNAP II. This unique feature of the transcription initiation nucleotide is the basis of several methods aiming to identify and map the exact positions in the genome in which RNAP II transcription starts. The derivative techniques differ in terms of resolution and throughput but share the common principles, as explain below.

Global Run-on Sequencing (GRO-Seq) active RNAP II is incubated with 5-bromouridine 5'-triphosphate (Br-UTP). RNA molecules that have incorporated BrUTP can be affinity purified by antibodies against bromodeoxyuridine (anti-BrdU). After cap removal and end repair, the eluted RNA is reverse-transcribed to cDNA. Deep sequencing of the cDNA identifies RNAs that are actively transcribed by RNAPII, in both senses. Major limitations of GRO-seq are the amount of starting material (the number of cells that are required lies in the 10ˆ7 range), resolution is only 30–100 nt, requires nascent RNAs of at least 18 nt (Gardini, 2017). **5'-GRO-Seq** is a modified version of the GRO-seq where only capped RNAs are selected, however, this method is restricted to cell culture due to the requirement for incubation in the presence of labelled nucleotides (Lam et al., 2013).

Precision Nuclear Run-on Sequencing for RNA Polymerase II Start Sites (PRO-Cap) maps RNAPII initiation sites during RNA transcription with base-pair resolution. A nuclear run-on reaction with biotin-NTP and sarkosyl is carried out on nuclear lysates. Incorporation of the first biotin-NTP halts further elongation of nascent RNA strands by RNAPII. The RNA strands are extracted and purified through streptavidin pull-down. Next, 3' adapters are ligated directly to the purified sample before another streptavidin purification step. The 5' ends are repaired and ligating 5' adapters. The adapter-flanked RNA fragments are enriched through another streptavidin pull-down process before RT and PCR amplification. The resultant cDNA strands are sequenced from the 5'end, and RNAPII pause sites are mapped. This technic is limited to *in vitro* reactions.

Paired-end Analysis of Transcription Start Sites (PEAT) poly(A) RNAs are enriched from total RNA and the caps are removed, The 5' ends of uncapped mRNAs are ligated to chimeric linkers containing MmeI restriction endonuclease sites prior to RT. The RT primers also contain an MmeI site, resulting in single-stranded cDNA flanked by MmeI sites. The fragments are PCR-amplified and circularized into circular single-stranded cDNA, which is amplified further by rolling-circle amplification. MmeI is used to cut circular cDNA at the 2 MmeI sites to create linear, double-stranded cDNA fragments that are 93–95 bp long. The fragments are ligated to paired-end adapters, amplified, and sequenced. Improved accuracy and alignment yield compared to older, single end TSS mapping strategies, however, does not distinguish between capped and non capped RNA (NI et al., 2010).

The cap analysis gene expression (CAGE) is the most commonly used and exploits the 2',3'-diol structure of the cap nucleotide, which is only present in only one other place on an RNA molecule besides the cap - its extreme 3' end. The diol structure is susceptible to a specific chemical oxidation, which can be followed by biotinylation, enabling selection of capped messages by immunoprecipitation with streptavidin. The enriched capped RNA fraction is then converted into cDNAs that span the entire lengths of the capped RNA molecules, which could be next converted into short DNA tags derived from their 5' ends.

These small cDNA are subsequently inserted into a plasmid, cloned and sent for sequencing (Shiraki et al., 2003). The advantage of the CAGE is the reduction of the sequencing cost that allows a higher genome coverage, but also the nucleotide base pair precision in contrast to GRO-seq assays (~30bp).The major disadvantage with this method lies in the large amount of material needed, precluding its application for small input material such as embryos. Only works on mature RNA. **Deep CAGE** is a derivate of CAGE combined with next-generation high-throughput sequencers allowing an accurate estimation of transcript abundance (Kurosawa et al.,2011).

5'-end serial analysis of gene expression (5' SAGE) and robust analysis of 5'-transcript ends (5'-RATE) based on the **CAGE** procedure, both take advantage of the fact that the 5' cap is resistant to phosphatase treatment, which removes mono-, di- or triphosphates from cleaved or degraded RNA. Subsequent removal of the cap using tobacco acid pyrophosphatase leaves a 5'-monophosphate, which is amenable to ligation with a specific linker nucleotide that marks the position of the native 5' end of RNA and can later be used to select and sequence the 5' ends of capped cDNAs. (Suzuki et al., 2004; Gowda et al,. 2006).

Native Elongating Transcript Sequencing (NET-Seq) In this method, the RNAP II elongation complex is immunoprecipitated, and the RNA is extracted and reverse transcribed to cDNA. Deep sequencing allows for 3’-end sequencing of nascent RNA providing nucleotide resolution mapping of the transcripts. The advantage of this method is specifically mapping of the RNA-bound protein. Nevertheless, this method requires nascent RNA of at least 18 pb, as well, high specific antibodies to avoid the precipitation of nonspecific complexes (Harlen et al., 2016; Mayer and Churchman., 2016).

2.5. Properties of core promoters

The mapping of the TSS at the single base resolution shed light about the properties of core promoters, leading to a "general" classification accordingly with their shape, their sequence composition, their chromatin configuration, the function of their gene and their initiation pattern (Haberle *et al.*, 2014; Schor *et al.*, 2017).

According to the shape or from where the gene start to transcribe, two kinds of promoters were proposed: the focus, sharp or narrow promoters, that have a single predominant TSS, are associated with core promoter elements such the TATA box or the Initiator (Inr) and the dispersed or broad promoters, that have several TSS, used more or less at the same frequency across the gene, associated with ubiquitously expressed genes, and that do not contain a TATA box (FitzGerald *et al.*, 2006) (Figure 6). They are often associated with CpG island in vertebrates, but not in *Drosophila,* where CpG island are common (Rach *et al.*, 2009). In *Drosophila,* dispersed promoters are associated with promoter elements that often co-occur in a specific order and orientation: Ohler1, DNA replication element (DRE), Ohler6 and Ohler7 (Rach *et al.*, 2009; Hoskins *et al.*, 2011) (Figure 6). This dichotomy of promoter shape is conserved across species suggesting a functional importance not yet understood. For example, during the *Zebrafish* early embryonic development, Haberle and collaborators showed that TSS from maternal expressed genes differ from the zygotic ones, suggesting different mechanism for TSS selection within the same promoter before and during zygote genome activation (ZGA). In *Drosophila Melanogaster*, Rach et al showed that core elements in broad or dispersed promoters had lower levels of conservation than the peaked promoters and different spatiotemporal patterns of activity. Genes maternally expressed were found to have alternative promoters used in the later stages of development. Moreover, they showed differences in term of core promoter composition between maternal inherent and zygotic expressed genes. Genome-wide studies in human and *Drosophila* (Lenhard et al., 2012) classify core promoters in three types (Figure 6). Type I promoters contain TATA boxes, Inr motifs and focussed TSS; They also lack CpG islands and have unprecisely positioned nucleosomes (Rach et al.,2011). They are generally associated with tissue-specific expression in adult tissues and acquire histone H3 Lys 4 trimethylation (H3K4me3) and H3 Lys acetylation (H3K27ac). Type II promoters contain CpG island and dispersed TSSs, in mammals they lack

TATA boxes, and in *Drosophila* they contain DRE and Ohler motifs. They have precisely positioned nucleosome positions encircle a well-defined nucleosome free region at the promoter, marked by H3K4me3 and H3K27ac (Rach et al.,2011). Type III promoters are associated with developmentally regulated genes involved in patterning and morphogenesis, which in *Drosophila* contain combination of Initiator and DPE motifs. In mammals, they contain large CpG islands and they resemble housekeeping genes core promoters, which in embryonic stem cells are marked bivalently with both H3K3me3 and H3K27me3 (repressive mark) (Bernstein *et al.*, 2006).

2.5.1 Core promoter and pausing

Promoters of key developmental genes in the early *Drosophila* embryo play a critical role on RNAP II pausing. They have a direct impact on the "time of synchrony", which refers to the time it takes to achieve coordinate gene expression over 50% of the nuclei (Lagha et al., 2013). The substitution of paused promoters, which show rapid and synchronous gene activation, with non-paused promoters, results in slow and stochastic activation of gene expression. In the case of the *snail* gene, perturbing the synchronous activation by replacing the promoter with a less synchronous promoter impair gastrulation. Thus there is a positive correlation between pausing, synchrony and gene expression controlled by promoter sequences rather than enhancers.

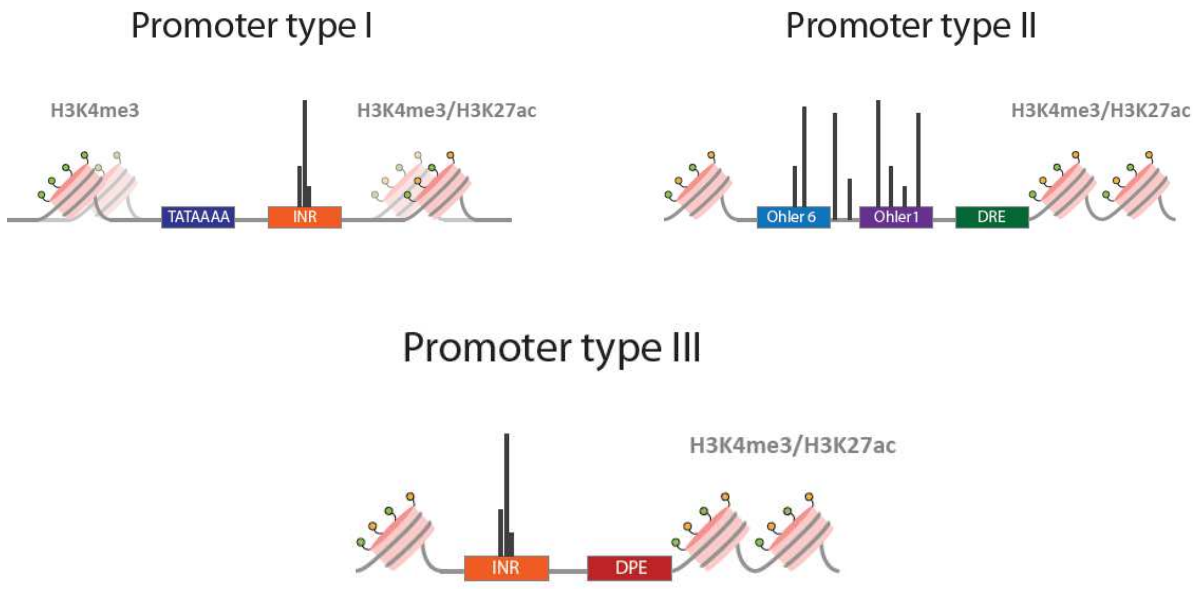


Figure 6. Promoters types in *Drosophila*. According to the position of the TSS, to the nucleosome organization and to the core promoter motifs; promoters can be classified into 3 types. Promoter type I: with focused TSS initiation, unprecise nucleosome position around the TSS, and enriched for TATA boxes and Inr motifs. Promoter type II, with broad TSS initiation, well defined nucleosome organization and enriched in Ohler and DRE motifs. Promoter type III, contain focused TSS initiation, defined nucleosome organisation and enriched in Inr and DPE motifs, associated with developmental and paused genes.

2.6. Core promoter elements

The analysis of core promoters at the level of the sequences, unveiled the presence of small DNA motifs (~3-8 bp) that are well conserved across species (from bacteria to humans) summarized in table 3. These small sequences are known as core promoter elements and may confer distinct properties to the core promoter (Figure 7). It might be noted that, there are no universal core promoter elements that are found in all promoters.

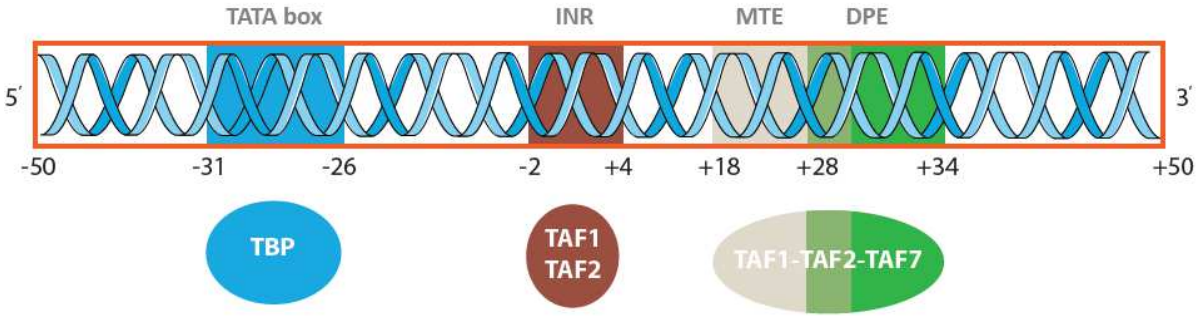


Figure 7. Core promoter motifs and their readers.

2.6.1 The TATA box and BRE Motifs

The first of the core promoter element to be identified was the TATA box (Goldberg, 1979 (PhD Thesis)); (Mathis and Chambon, 1981). In metazoans the TATA canonical consensus is TATAWAAR (W= A or T; R= A or G), with the 5'T generally located at ~ (-30; -31 pb) from the +1 of the TSS and is found in only about 10% to 20% of the core promoters (Cianfrocco *et al.*, 2013). The TATA motif is bound by the TBP subunit of TFIID and both TATA box and TBP are conserved from archaebacterial to humans (Reeve, 2003). Functional studies show implications of the TATA box in the regulation of transcription activity, RNAPII orientation, stability, pausing, and transcriptional noise. (Wang, Trivedi and Johnson, 1997), using unfractionated *Drosophila* nuclear extract showed the implication of the TATA box in promoting transcription, punctual or complete mutations of the TATA box motif lead to a dramatic diminution of the promoter activity.

In vitro transcription using HeLa nuclear extracts (Patwardhan *et al.*, 2009) showed that mutations where disrupting the AT that defines the TATA box of the promoter of the human

cytomegalovirus CMV promoter (TATATA, −28 to −23) led to a clear drop in transcriptional efficiency. In the same study, substitutions of C→T at −32 increased transcriptional efficiency of the promoter from the human beta globin gene (*HBB*) that has a no canonical TATA box (CATAAA, −32 to −27), potentially secondary to the formation of a more optimal TATA box with respect to distance from the TSS. In yeast, mutations of the TATA box from the PHO5 promoter displayed decreasing noise strength with a decreasing rate of gene expression (Raser, 2004) interesting, they show by comparing mutations from the TATA box and from the UAS sequences that two promoters can produce the same mean mRNA population with different noise characteristics. In *Dictyostelium* cells, point mutations in the TATA box of the *act5* promoter, T→A (first T of TATAAA) or A→C (first A TATAAA), suggested that the TATA box does not affect the duration or frequency of gene expression but rather modulates the initiation rates (Corrigan *et al.*, 2016). In the HIV-1 virus, mutations of the TATA box were reported to decrease the HIV-1 expression and the affinity of TBP for the promoter (van Opijnen *et al.*, 2004; Savinkova *et al.*, 2013). Furthermore (Tantale *et al.*, 2016) by recording long movies of transcriptional activity showed that the TATA box control the long permissive and non-permissive periods of transcriptional activity, likely due to the affinity and stability of TBP/ TATA box interaction, indicating that the activity of the HIV-1 promoter correlates the affinity of TBP for the TATA box. Recently work suggested the TATA box could affect the stability of the pause RNAPII pausing, however, the extend of it may depend on the whole promoter context (Shao et al., 2019).

The TFIIB recognition elements (BREs), which are bound by the TFIIB basal transcription factor, were first observed in archaeal genes through mutational analysis. Structural analysis of TBP-TFIIB-DNA as well as functional studies identified a 7 bp sequence, that is conserved from archaea to humans. These motifs can be present on either side of the TATA box, with the upstream BRE (BREu) at −38 to −32 (sequence: G/CG/CG/ACGCC, present in ~ 25% of the eukaryotic core promoters, however more prevalent in TATA-less promoters (28.1%) than in TATA-containing promoters (11.8%) (Gershenzon and Ioshikhes, 2005). And the downstream BRE (BREd) at −23 to −17 (sequence: G/ATT/AT/GT/GT/GT/G). Both can act in conjunction with the TATA box, to either stimulate or prevent transcription in a context-dependent manner (Smale and Kadonaga, 2003). For example, the BREu motif was found to inhibit the ability of the *Drosophila* Caudal protein to activate transcription from TATA-dependent

promoters (Juven-Gershon et al, 2008). Functions of the BREs are not well understood and further analysis are need it for determine their role in transcription regulation. In addition, the BRE consensus sequence in *Drosophila* has not yet been precisely determined.

2.6.2 The Initiator (INR)

First described by the Chambon's laboratory (Corden *et al.*, 1980). The initiator (INR) encompasses the +1 TSS, is probably the most prevalent core promoter motif in focused core promoters (FitzGerald et al, 2006). It is bound by the TAF1 and TAF2 subunits of TFIID (Kaufmann and Smale, 1994). The mammalian INR consensus sequence is YYA+1NWYY (Y=C or T; N= A or T or C or G; W= A or T) and the *Drosophila* consensus is TCA+1KTY (K=G or T; Y=C or T) (Table 3). Initially it was propose that the function of the INR was to initiate basal transcription in the absence of the TATA box (Smale and Baltimore, 1989). Nowadays the INR element is better known to correlate with focus transcription. This is likely by stabilizing the interaction between TFIID and the DNA via TAF1/TAF2 subunits (Sainsbury, Cramer et al 2015). Recently the Zeitlinger's laboratory (Shao et al, 2019) demonstrate the implication of the INR element in the pausing stability. By using a reporter-ChIP-nexus assay in *Drosophila cells*, the authors showed that a single mutation of the canonical INR sequence (the G at +2 position of the TSS) of highly pause genes, can decrease dramatically the stability of the RNAP II pausing.

2.6.3 The Downstream core Promoter Element (DPE)

The DPE was discovered in Drosophila by the analysis of TATA-less promoters (Burke and Kadonaga, 1996). The consensus for the Drosophila DPE is RGWYV (R= A or G; W= A or T; Y= C or T; V= A or C or G) from +28 to +32 relative to the +1 TSS (Vo Ngoc et al. 2017). It is estimated that ~40% of Drosophila promoters contain a DPE motif (Kutach and Kadonaga, 2000). The DPE is also present in humans, although rare, it is recognized by the human basal transcriptional machinery (Burke and Kadonaga, 1996; Juven-Gershon, Hsu and Kadonaga, 2006). A variant of the Drosophila DPE is "the motif 9" (Ohler et al, 2002). The motif 9 sequence has combined features of the Motif Ten Element (MTE) and the DPE, and is sometimes referred to as the "Ohler DPE." The DPE functions in cooperation with the INR element for the binding of TFIID, as well as for transcription activity (Burke and Kadonaga, 1996). The mutation of either element results in a loss of TFIID binding and promoter activity. Indeed, there is a precise spacing requirement between the DPE and INR, a single nucleotide increase or decrease in the spacing between the two elements can result in a several-fold decrease in transcriptional activity (Burke and Kadonaga, 1997; Kutach and Kadonaga, 2000). The DPE was reported to precisely align with the peak of RNAP II pausing (Nechaev et al,2010). In Drosophila one-fifth of paused promoters are enriched for DPE. This may indicate that DPE, as opposed to the TATA box could contribute to RNAP II pausing (Hendrix et al, 2008). Recently, (Shao et al, 2019) showed that the DPE motif in concert with the Pause Button (PB) and the Motif Ten Element (MTE) in Drosophila, plays a role in the stability of the pause RNAP II, and in this study, the authors refer to these elements as pausing elements. In terms of interactions, cryo-EM structure of the Pre-Initiation Complex, showed that the TAF1 and the TAF2 subunits of the TFIID are in contact with the DPE (Louder et al., 2016). However, photo-crosslinking experiments with purified TFIID indicated that the DPE is rather in contact with the TAF6 and TAF9 subunits of TFIID (Burke and Kadonaga, 1997). The basis for this difference is not known, but could be due to alternate conformations of promoter-bound TFIID or possibly to the presence of multiple TFIID complexes at the promoter in the photo-crosslinking experiments.

2.6.4 The Motif Ten Element (MTE)

The MTE was discovered in *Drosophila*, it was first identified by computational analysis by (Ohler et al, 2002). The consensus for the *Drosophila* MTE is CSARCSSAACGS (S = C or G; R = A or G). It is located immediately upstream of the DPE (+18 to +27). Like the DPE, the MTE works cooperatively with the INR for TFIID binding and transcriptional activity in TATA less promoters (Lim *et al.*, 2004). In addition, there is a synergy on transcription activity between the MTE and the DPE (Lim *et al.*, 2004; Theisen, Lim and Kadonaga, 2010). However, although the majority of the MTE content promoters contain DPE, the MTE motif functions independently of DPE. Photo-crosslinking data with purified TFIID showed that TAF6 and TAF9 are in close proximity to the MTE (Theisen et al, 2010). The photo-crosslinking results with the MTE are similar to those seen with the DPE (Burke and Kadonaga, 1997), but differ from the structural studies of TFIID in the PIC (Louder et al, 2016). An analysis of the downstream region of *Drosophila* core promoters suggested that the MTE might be considered to be a single functional unit with multiple contact points with TFIID that promotes the binding of the TFIID to the core promoter region (Theisen et al, 2010).

2.6.5 The pause Button (PB)

The Pause button was identified in *Drosophila* by the Levine lab (Hendrix et al, 2008). In this study the authors compare promoter sequences from a region spanning +1 to +60 bp downstream of the TSS, from highly stalled (or poised) genes versus constitutive promoters. They found that stalled promoters contain a significantly higher GC and CpG content than constitutive promoters. In addition, the authors found that stalled promoters are enrich with a 7pb motif: KCGRWCG (K = G or T; R = A or G; W = A or T) located between +25 and +35 bp from the TSS, called by authors the pause button. This motif is similar and occurs to overlap to that of the DPE. Over one-fifth of the paused *Drosophila* promoters are enriched for the DPE, the MTE and the PB motifs, all of which are located close to the pause site. Notably, 75% of the genes in the dorsal-ventral network were identified as paused genes (Hendrix et al, 2008). Over two thirds of Dorsal target genes contain a DPE motif (Zehavi et al, 2014). These correlations, in addition to the fact that PB and DPE are GC-rich, both motifs share the 'GGWC' sub-consensus (W = A or T) , they overlap with the paused Pol II, and precisely align with the

peak of RNAP II pausing (Nechaev et al, 2010), may indicate these motifs could contribute to Pol II pausing.

2.6.6 The polypyrimidine initiator (TCT)

It is a rare but biologically important core promoter motif in bilateria. In *Drosophila*, the TCT motif is present in only ~120 core promoters, including the ribosomal protein gene promoters (Parry *et al.*, 2010), as well as in some other genes that encode factors that are involved in translation (Parry et al, 2010). The TCT consensus sequence in *Drosophila* is YYC+1TTTYY (Y = C or T), which is similar to the TCT consensus in humans (YC+1TYTYY) (Parry et al. 2010). Although very similar to the *Drosophila* INR motif, the two elements are functionally distinct. The TCT is likely to be important for the coordination of the expression of the ribosomal protein genes. The TCT motif in *Drosophila* is mediated by TBP-related factor 2 (TRF2) instead of TBP (Wang et al, 2014). These findings indicate that there is a specialized transcription system involving TRF2 and the TCT motif for the expression of the ribosomal protein genes in *Drosophila*. This TCT-based system complements the RNA polymerase I and RNA polymerase III transcription systems for the synthesis of the components of the ribosome.

2.6.7 Other promoter elements

Computational sequences analysis, reported other motifs which may serve as targets of the general transcription machinery. (Ohler et al., 2002) determined 10 most over-represent motif in a cluster of 1941 TSS in *Drosophila.* This analysis described sequence motifs located in the promoter such as, the DNA replication related element (DRE) which is a target of the DNA replication-related-element binding factor (DREF). DREF, which was discovered in *Drosophila* and was later found to have orthologues in many other species (including humans), is involved in transcriptional regulation of proliferation related genes (MATSUKAGE *et al.*, 2008). Other sequences motif unveil by this analysis are the Ohler motif 1, 6, 7. These motifs were found to be enriched in housekeeping genes and they seem to be associated with dispersed transcriptional initiation.

During my thesis, we put a particular attention to the TATA box, the INR, and the DPE motifs. I will discuss the effect on transcription dynamics of these core promoter motifs in the second part of the manuscript.

Core promoter motif	Consensus sequence	position relative to TSS	Bound by
TATA box	TATAWAWR	-31 to -24	TBP
Inr	TCAGTY	-5 to -2	TAF 1/TAF 2
DPE	RGWCGTG	+28 to +34	TAF1 and TAF2, possibly TAF7
	RGWYVT	+28 to +33	
	GCGWKCGGTTS	+24 to +32	
MTE	CSARCSSAACGS	+18 to +29	Possibly TAF1 and TAF2
Ohler 1	YGGTCACACTR	-60 to -1	M1BP
Ohler 6	KTYRGTATWTTT	-100 to -1	N.A
Ohler 7	KNNCAKCNCTRNY	-60 to +20	N.A
DRE	WATCGATW	-100 to -1	DREF
TCT	YYCTTTYY	-2 to +6	N.A
BRE^{u}	SSRCGCC	-38 to -32	TFIIB
BRE^{d}	RTDKKKK	-23 to -17	TFIIB
Pause button	KCGRWCG	+25 to +35	N.A

Table 3. Core promoter motifs, its canonical sequences and the protein interactions in *Drosophila*. Adapted from (Haberle and Stark, 2018). For the consensus sequences, D = A or G or T; K = G or T; M = A or C; R = A or G; S = C or G; V = A or C or G; W = A or T; Y = C or T.

3. *Drosophila* early embryogenesis

The *Drosophila* early embryogenesis is incredibly rapid and precise. Transcriptional patterns of *Drosophila* embryo development have been intensively studied over the past decades, and new imaging methods allow the visualization and quantification of gene expression. Thanks to the facility of collection, ease of genetic manipulations, and an extensive knowledge of the genomic landscape (thanks to ENCODE consortium for example), the *Drosophila* embryo has become an organism of choice for the study of transcription in a multicellular organism.

The *Drosophila* embryo starts its life as an ellipsoid egg of approximately 180 microns in diameter and 510 microns in length (MARKOW, BEALL and MATZKIN, 2009). During the first two hours after fertilization, the zygotic nucleus undergoes 13 rapid division cycles (8 to 15 min) without cytokinesis and these nuclei migrate to the periphery of the egg around nuclear cycle (nc) 8th. Until the nc 14th, the embryo is a syncytium with no membranes separating the nuclei, which allows proteins to freely diffuse between neighbouring nuclei. Most of the important patterning events of the embryo occur during nc 14^{th} (which lasts about 55 minutes), when the mid blastula transition (MBT) and the zygote genome activation (ZGA) co-occur. At the end of nc14, the first major morphogenetic movement occurs: the process of gastrulation (Figure 8).

3.1. The early *Drosophila* development

In only 2 hours (at 21°C), a typical drosophila embryo undergoes 14 nuclear divisions, leading to a large cell with thousands of nuclei. This remarkable speed is due to the absence of gap phases and fast replication of the DNA. By the nc 8^{th}, 256 nuclei are produced and they start to migrate towards the periphery of the egg, where the mitoses continue, albeit at a progressively slower rate. During the 9^{th} division cycle, about five nuclei reach the surface of the posterior pole of the embryo. These nuclei become enclosed by cell membranes and generate the pole cells that will give rise to the gametes of the adult. Most of the other nuclei arrive at the periphery of the embryo at nc 10^{th} and then undergo four more divisions at progressively slower rates. During these stages of nuclear division, the embryo is called a **syncytial blastoderm**, meaning that all the cleavage nuclei are contained within a common cytoplasm (Figure 8).

At fertilization, the embryonic genome is nearly quiescent and maternally loaded gene products direct development. As the embryo progresses, zygotic gene products become required for developmental events and cell cycle progression; thus, control of development and the cell cycle is handed off from the maternal genome to the zygotic genome, often called the maternal to zygotic transition (MZT).

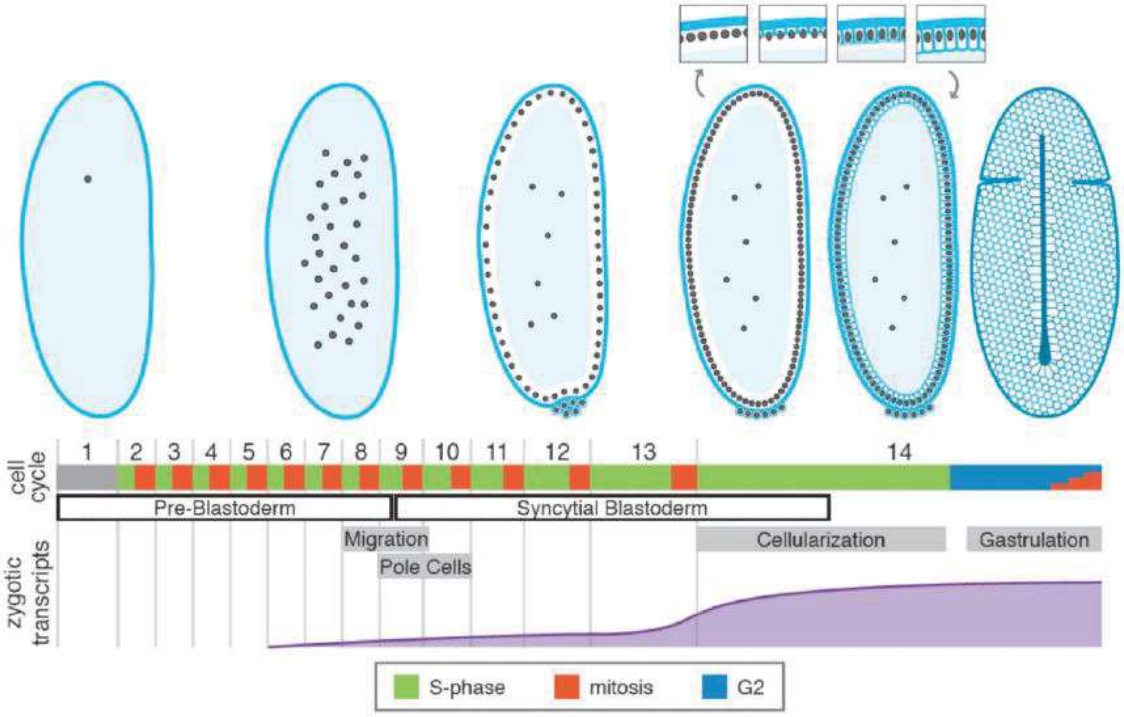


Figure 8. Early *D*rosophila development. From (Farrell and O'Farrell, 2014). A diagram of the first 14 cycles of *Drosophila* development with notable morphological stages illustrated at the top. Note that while most embryos are displayed as sections through the middle of the embryo with the ventral side to the right, the final illustration is a surface view, with the ventral side up. The process of cellularization is diagrammed in more detail in the insets. The duration of each phase of the cell cycle is below: S phase (green), mitosis (red), and G2 (blue). Mitosis 14 is represented as a series of small bars because the embryo is no longer synchronous at this time and individual groups of cells enter mitosis at different times according to a developmentally programmed schedule. The timing of notable morphological events is demarcated in grey boxes: the migration of the nuclei to the blastoderm, the insulation of the germline by cellularization of the pole cells, the cellularization of the blastoderm nuclei, and the onset of the first gastrulation movement—ventral furrow formation. Below this is diagrammed the approximate number of genes for which zygotic transcripts have been detected over time.

3.2 Maternal to zygotic transition (MZT)

This important developmental transition comprises two major processes: 1-maternal mRNA clearance (or maternal RNA decay) and 2- the new synthesis of zygotic products (Figure 9). In *Drosophila,* the maternal mRNA clearance occurs in two phases, an initial early one upon egg activation, followed by a phase that requires zygotically synthesized products (Bashirullah *et al.*, 1999; Bashirullah, Cooperstock and Lipshitz, 2001; Tadros *et al.*, 2003; Tadros, Westwood and Lipshitz, 2007). About 25% of the cleared transcripts are degraded strictly by the maternal machinery, 35% strictly through the zygotic machinery, while 40% show mixed decay effected by both maternal and zygotic mechanisms (De Renzis *et al.*, 2007; Tadros, Westwood and Lipshitz, 2007; Thomsen *et al.*, 2010). The scale and dynamics of these phases vary across species, (Vastenhouw, Cao and Lipshitz, 2019); (Figure 9).

The second step of the MZT is the onset of zygotic genome activation (ZGA), a period over which transcription is gradually activated (Schulz and Harrison, 2019; Vastenhouw, Cao and Lipshitz, 2019); (Figure 9). However, this progressive period of zygotic activation has been traditionally separated into two transcriptional waves: a minor wave that occurs during the cleavage divisions; and a major wave that, in many species, coincides with the lengthening of the cell cycle. The timing of these waves and the number of cell divisions vary across species. Rapid developing species such as worms (*Caenorhabditis elegans*), frogs (*Xenopus laevis*), fish (*Danio rerio*) and flies (*Drosophila melanogaster*) complete the MZT and enter gastrulation only a few hours after fertilization. By contrast, in slower developing mammals such as mice (*Mus musculus*) and humans, the MZT takes one or more days (Schulz and Harrison, 2019). In the case of the *Drosophila melanogaster,* the major wave co-occurs with the **Mid-blastula transition (MBT)** described as the specific stage during the development of the embryo, which is marked by lengthening and desynchronization of the cell cycles (Vastenhouw et al, 2019). Chen and colleagues (Chen *et al.*, 2013) showed that genes expressed before the MBT or preMBT genes, tend to have particularly low levels of H3K4me3, with no notable enrichment of Pol II at the pause site in most of them. In addition, non- paused preMBT promoters are highly enriched in TATA and INR motifs, in contrast to paused preMBT promoters that are rather enriched in GAGA, INR and PB. This is in stark contrast with MBT genes which are transcribed during the MBT. These genes show high levels of H3K4me3 at their +1

nucleosome, most of them have high Pol II occupancy at the pausing site, and their core promoters are often enriched in INR, DPE, MTE and PB motifs and lack TATA motifs (Figure 9), showing that core promoters correlate with pausing profiles.

Given the association of particular core promoter motifs to early developmental genes, and given that these genes are expressed in a very particular context (fast development, short interphase), one can question whether these promoters exhibit transcriptional properties, which are adapted to this particular developmental context. This is the working hypothesis that guided our interest towards core promoters or pre-MBT genes.

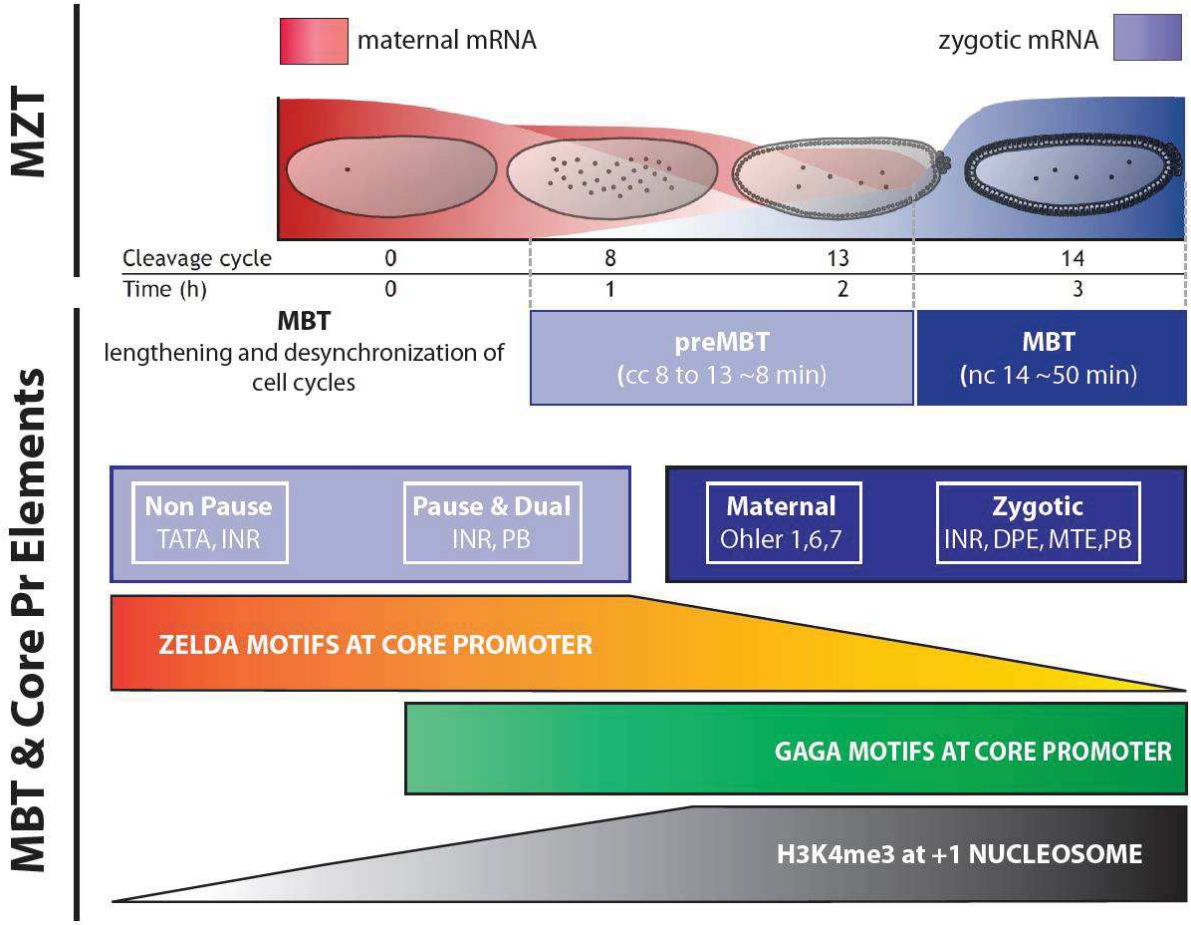


Figure 9. Relationship between the MZT, the MBT and the core promoter elements in *Drosophila*. Adapted from (Vastenhouw, Cao and Lipshitz, 2019)**. Note:** the Zelda and GAGA motifs enrichment were reported in Chen et al., 2013, where they performed motif analysis on 200bp centred on the transcriptional site.

3.3. Patterning of the early Drosophila embryo

The early embryo is divided in two axes, the anterior/posterior (AP) and the dorsal/ventral (DV) axes, thanks to maternal supplies accumulated during oogenesis. The AP axis is under the control of the maternal morphogen Bicoid, a homeodomain transcription factor and the maternal factor Nanos, an RNA-binding protein involved in translational repression. During oogenesis, mRNAs of Bicoid are accumulated in the anterior pole of the embryo. After fertilization, Bicoid mRNAs are translated and the protein diffuses in the embryo from the anterior to the posterior creating a gradient (Driever and Nüsslein-Volhard, 1988). The embryo cells have the ability to measure the concentration of Bicoid protein and activate the *gap* genes responsible for the head and thorax development (Figure 10).

The RNA of Nanos is localized to the posterior pole of the embryo, where Nanos protein is required for abdomen formation. In the egg, at first, the distribution of the Nanos mRNA is homogeneous. Then, when the Smaug protein is available, 96% of Nanos mRNA is degraded (Gavis and Lehmann, 1992). Only the Nanos mRNA in the posterior pole is maintained through the interaction with the Oskar protein. After fecundation, the Nanos mRNA is translated and forms a posterior-anterior gradient. Nanos inhibit the translation of the Hunchback mRNA, a gap activated by the morphogen Bicoid and necessary for the anterior segments (Figure 10).

The Dorsal-ventral (DV) patterning of the *Drosophila* embryo is specified by the morphogen Dorsal, a sequence-specific transcription factor related to mammalian NF-κB (Roth, Stein and Nüsslein-Volhard, 1989; Rushlow *et al.*, 1989). The Dorsal protein is distributed in a nuclear gradient along the DV axis in the shape of a Gaussian distribution, with peak levels present in most ventral nuclei, where it controls the activation of the genes *Twist and Snail,* both involved in mesoderm fate and gastrulation (Figure 10). The Dorsal gradient initiates DV patterning by regulating 50-60 target genes in a concentration-dependent fashion through enhancer sequences (Stathopoulos *et al.*, 2002; Zeitlinger *et al.*, 2007). In the most dorsal regions of the embryo the Dorsal gradient is shallow and contains limited spatial information. In this region of the embryo, much of the spatial information is contained in the decapentaplegic (Dpp) gradient, which is indirectly specified by the Dorsal gradient (Podos and Ferguson, 1999). The Dpp gradient regulates a number of genes in a concentration-

dependent fashion including *Pannier (pnr)* and *Tailup (tup)*, both involved in dorsal cell fate and nervous system development (Figure 10).

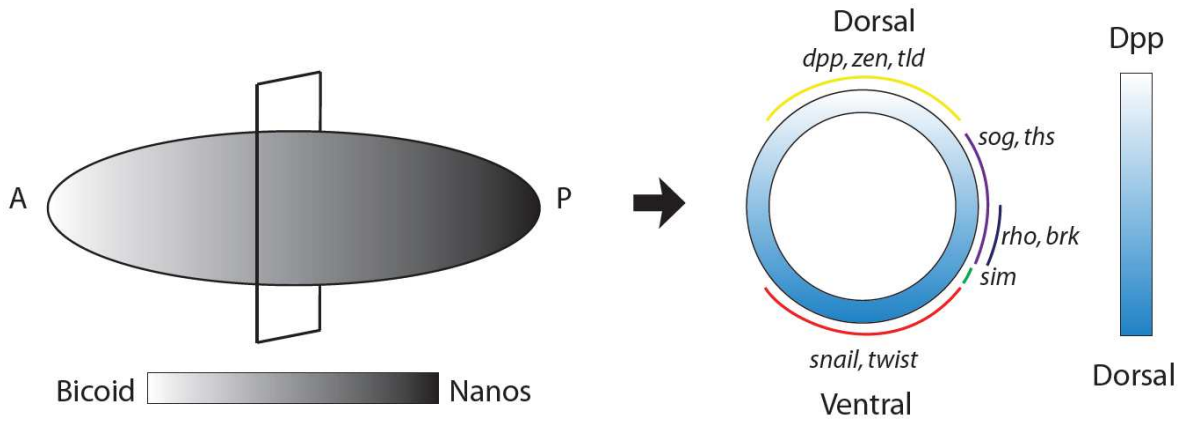


Figure 10. Antero-Posterior and Dorso-Ventral patterning in *Drosophila* embryos. In the left, schematic representation of the gradient for the main proteins establishing the A/P axis (bicoid and nanos). In the right, schematic representation of the gradient of the morphogen dorsal, which generates distinct patterns of gene expression (High levels of dorsal are necessary for the *snail* and *twist* expression. In contrast, low levels are needed for the expression of *sog* and *ths*. *Dpp* is expressed at the most dorsal region of the embryo and is repressed by Dorsal).

4. Imaging gene expression dynamics

In recent years, major advances in microscopy and in DNA/RNA/protein labelling methods enabled the visualisation and quantification of transcriptional dynamics in living single cells (Pichon *et al.*, 2018). The vast majority of these studies were performed on cultured cells and they provided key insights on transcriptional kinetics. However, even though, studies in cultured cells have greats advantages in terms of manipulation and visualisation, they do not reflect the spatial and temporal constrains of a multicellular organism in its natural context.

4.1 Insights from RNA FISH

Studies in cultured cells, have shown the heterogeneity in gene expression for cells shared the same environment and genetic background, termed noise. Noise in gene expression is defined as the standard deviation divided by the mean of the distribution of mRNA concentration, and it contains extrinsic and intrinsic components (Raser, 2005). The first component is produced by the fluctuations of the proteins controlling gene expression (i.e., transcription factors, polymerases, ribosomes, etc.), while the second component depends on the stochastic nature of each step during gene expression (i.e., promoter activation, transcription, translation, and mRNA and protein decay) (Elowitz *et al.*, 2002; Swain, Elowitz and Siggia, 2002). Noise is gene specific and specific combinations of regulatory sequences (promoter, enhancers, insulators) and transcriptional regulators (TF expression, frequency, or amplitude) influence the ratio between intrinsic and extrinsic noise and consequently the mRNAs levels. In eukaryotic cells, a substantial contribution to variability is attributable to transcriptional bursting. smFISH or live-imaging experiments in bacteria, yeast, or mammalian cells were extensively used to estimate the mean and the variance of mRNAs in a population of cells (Levsky *et al.*, 2002; Raj *et al.*, 2006; Zenklusen, Larson and Singer, 2008). This approach calculates the size and frequency of transcription bursts and found a positive correlation between the burst size and the noise amplitude. Moreover, in yeast TATA motif has been correlated with noise, mutations of the TATA motif lead to a decrease of noise strength (Raser, 2004; Blake *et al.*, 2006). A recent study in Drosophila performed by the Furlong lab showed, how promoter variants have a direct impact on transcriptional noise (Schor *et al.*, 2017). In this study natural polymorphisms in the promoter region of the same

gene can lead to an increase of the expression noise that not always accompanied by an increase in gene expression suggesting that elevated noise levels are not merely a consequence of higher expression.

In fixed embryos single-molecule fluorescence in situ hybridization (smFISH) has long been the standard technique for localizing and quantifying individual mRNAs (Femino *et al.*, 2003). Furthermore, the combination of biochemical studies, computational tools, and mathematical models provides information about absolute levels of gene expression, cell-to-cell variability and assumptions in kinetics of transcription initiation. However, transcription is a dynamic process and lack of temporal resolution from the smFISH studies limits the comprehension of the events govern gene expression.

Live imaging technics allow the study of the mRNA in real time on live organisms. Although low throughput, this provides the best spatial information and can be quantitative when single transcripts can be detected (H. G. Garcia *et al.*, 2013; Lenstra *et al.*, 2016; Tutucci, Vera, *et al.*, 2018). Direct visualization of transcriptional activities of particular targets has been achieved by imaging nascent transcripts containing RNA stem-loops (such as MS2 and PP7) (H. G. Garcia *et al.*, 2013; Lucas *et al.*, 2013; Fernandez and Lagha, 2019) bound to GFP-tagged stem-loop binding proteins (Figure 11).

4.2. Insights from live imaging

The major contribution of *in vivo* mRNA imaging studies concerns the study of transcriptional bursts. Indeed, eukaryotes transcription is predominantly bursty, but all genes seem to be characterized by different bursting properties (frequency, amplitude, duration). In *Drosophila* (Fukaya, Lim and Levine, 2016) using MS2 and PP7 system showed that the enhancer position, strength and interaction with the promoters influences burst characteristics. Also, (Tantale *et al.*, 2016) showed using MS2 system in mammals cells that transcriptional burst are generated by group of closely spaced polymerase convoys. Moreover, in this study the authors, showed that burst can occur at different time-scales controlled by different mechanisms, short burst

(lasting minutes) are under the mediator control, and long burst (lasting hours) are controlled by the TATA-binding protein (TBP).

These studies in live and fixed embryos show how single molecules imaging technics can contribute to our knowledge on transcription dynamics, however, additional technologies are need it, to allow the simultaneous visualization of the nascent mRNA and the molecular mechanisms modulating the transcription initiation process on time and on space.

4.2.1 Bursting

Transcription in higher eukaryotes is a stochastic process involving extended periods of inactivity interspersed with bursts of RNA synthesis (Suter *et al.*, 2011; Senecal *et al.*, 2014; Lenstra *et al.*, 2016). Bursting properties are influenced by promoter architecture, the chromatin landscape, the genomic position, and the transcription factor binding dynamics (Lenstra *et al.*, 2016; Nicolas, Phillips and Naef, 2017). However, how *cis*-regulatory elements, such as promoters and enhancers, affect transcription burst size (number of polymerases loaded), burst duration and frequency is unclear.

Several studies, so far, reveal that enhancers predominantly modulate frequency of transcription bursts (Senecal *et al.*, 2014; Fukaya, Lim and Levine, 2016; Larsson *et al.*, 2019), whereas core promoters affect their size. (Raj *et al.*, 2006; Suter *et al.*, 2011; Larsson *et al.*, 2019). The role of enhancers in modulating the frequency through the number, availability, and the affinity of TF sites has been intensively studied (Fukaya et al, 2016; Nicolas et al,2017). However, how core promoters regulate the burst size has remained elusive. In part because of the different combination of elements that a core promoter could harbour, making difficult to decode their direct implication in transcription initiation and in bursting dynamics. Using single cell allele specific RNA-seq in primary mouse fibroblasts (Larsson *et al.*, 2019) showed that genes with a TATA box in their core promoter had significantly larger burst size (here burst size means number of RNAP II loaded). This study also revealed that when TATA-containing core promoters are associated with an INR motif the burst size was found to be significantly boosted. Thus, the authors conclude that a synergy between the TATA box and

the INR element might exist. However, the INR element alone seems no to have an effect on burst size.

The finding that transcription of developmental genes is bursty is unexpected. Indeed, in the context of the rapid and robust development of *Drosophila*, gene expression should in principle be tightly controlled to avoid any source of noise and ensure the right production of mRNAs in time and space. Single molecule FISH measurement at several *Drosophila* genes, show larger variations than can be explained by random noise, suggestive of bursting transcription (Blake *et al.*, 2006; Paré *et al.*, 2009).

All this suggest that *cis*-regulatory sequences may play a pivotal role in the regulation of transcription initiation events, by using different strategies likely adapted to each gene context.

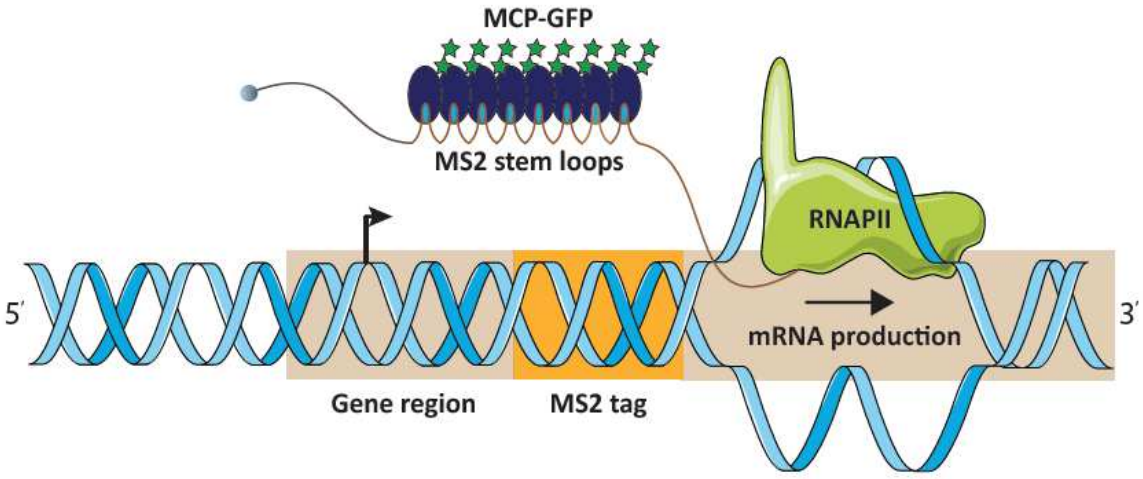


Figure 11. The MS2 tagging system. The system relies on the labelling of RNA by addition of multiple stem loops to the DNA sequence (orange box). When a stem loop sequence is transcribed a MS2 coat protein (dark blue) fused to a fluorescent protein would bind as a dimer (green stars). When several stem loops are bound by the MCP-GFP (example here) the accumulation of fluorescent proteins would allow the visualization of transcription of a target gene in real-time.

5. Aim of the present work

The aim of my thesis is to dissect the role of core promoter sequences on transcription initiation dynamics. In particular, the link between specific core promoter composition and transcriptional kinetics.

To assess this problematic, we use transgenic *Drosophila* lines where plasmids with the same enhancer, MS2 sequence and gene reporter but with a series of minimal promoters (100pb) were inserted in the same genomic location. The MS2 sequence allows the fluorescent labelling of transcripts through the binding of the mRNA MS2 loops (24X) to the MS2 coat protein fused to GFP. This technique combined with Fast scanner Confocal allows to record the transcription on live embryos with a high temporal resolution (4s).

II-RESULTS

1. Lightening up gene activation in living Drosophila embryos (review)

When I started my PhD I started acquisitions using a classical confocal microscope, that allowed to record transcription dynamics on live embryos using a temporal resolution of 20 sec. Then during my 3rd year of PhD the core facilities of the institute acquires a new confocal microscope carrying an AiryScan module. This module allows to increase the temporal resolution of the data (4s) keeping the same resolution. This technical review explains the details and main points to take into account to perform live acquisition from live embryos.

Contribution of this manuscript:

With the help of the MRI imagery platform, I set up the protocol explained in this review to perform the live imaging at 4 second time resolution. I produced the data shown in this technical review.

The Dr. Lagha and I wrote the manuscript.

2. Decoding the impact of promoter sequences on transcriptional dynamics *in vivo*

Here are the main results concerning the promoter quantification. The results concerning kinetics estimates obtained by the deconvolution method are still preliminary. Indeed this summer before starting the redaction of this manuscript, my collaborators realized that there was a problem in the image acquisitions that I made. This problem concerned approximately 50 movies, which were processed and analysed. Therefore, since then, with the help of Antonello Trullo, Matthieu Dejean and Virginia Pimmett, we are currently re-acquiring those movies. The result presented in this version of the manuscript are based on the last reliable acquisitions that we made.

Contributions of this manuscript:

Here I described the contributions that I made for this manuscript:

1. Generation of *Drosophila* transgenic lines:

   I used the SnaE<SnaPr<MS2(24X)<Yellow plasmid (already in the laboratory) as a reference to clone the promoter used in this study. Amplification of promoters were performed either by PCR or by gene synthesis.

   Lines that I created for this study:

   -SnaE<KrPr<MS2(24X)<Yellow

   -SnaE<KrINRmut<MS2(24X)<Yellow

   -SnaE<KrINRmut G to T< MS2(24X)<Yellow

   -SnaE<SnaTATAlightMutPr<MS2(24X)<Yellow

   -SnaE<SnaTATAStrongMutPr<MS2(24X)<Yellow.

   -SnaE<RhoPr<MS2(24X)<Yellow

   -SnaE<Sna+INRPr<MS2(24X)<Yellow

   -SnaE<SnaTATAlightMut+INRPr<MS2(24X)<Yellow

   -SnaE<KrPr+TATA<MS2(24X)<Yellow

   -SnaE<KrPr+TATA+INR<MS2(24X)<Yellow

2. Live Imaging acquisition:

   Here, I put the number of acquisitions performed for this manuscript and the contribution of each person. In total: 52 movies were made. I made 34 acquisitions, M. Dejean made 15 acquisitions, and V. Pimmett made 3 acquisitions.

Movies 4Sec/ MCP-GFP Hist RFP Homozygous	
Line	**Made by**
Kr	Matthieu
Kr_em1	Carola
Kr_em2	Carola
Kr_em3	Carola
Sna_em1	Carola
Sna_em2	Carola
Sna_em3	Carola
SnaTATALight Mut_em1	Carola

SnaTATALight Mut_em2	Carola
SnaTATALight Mut_em3	Carola
SnaTATALight Mut_em4	Carola
SnaTATALight Mut_Emb1	Matthieu
SnaTATALight Mut_Emb5	Carola
SnaTATAStrong Mut_emb2	Virginia
SnaTATAStrong Mut_emb2	Virginia
SnaTATAStrong Mut_emb3	Virginia
SnaTATAStrong Mut_emb1	Carola
KrINRmut G to T em1	Carola
KrINRmut G to T em2	Carola
KrINRmut G to T em3	Carola
KrINRmut G to T em4	Carola
krINRmut_em1	Carola
krINRmut_em2	Carola
krINRmut_em3	Carola
krINRmut_em4	Carola
krINRmut_em5	Carola
krINRmu_em6	Carola
Ilp4_emb1	Matthieu
Ilp4_emb2	Matthieu
Ilp4_emb3	Matthieu
Ilp4_emb4	Matthieu
Ilp4+INR_emb1	Matthieu
Ilp4+INR_emb2	Matthieu
Ilp4+INR_emb3	Matthieu
Ilp4+INR_emb4	Matthieu
WntD_emb1	Matthieu
WntD_emb2	Matthieu
WntD_emb3	Matthieu
Sna+INR_em1	Carola
Sna+INR_emb1	Matthieu
Sna+INR_em5	Carola
Sna+INR_em6	Carola
Sna+INR_em2	Carola
Sna+INR_em3	Carola
Sna+INR_em4	Carola
KrTATA_em1	Carola
KrTATA_em2	Carola
KrTATA-INR_emb1	Matthieu
KrTATA-INR_em1	Carola
KrTATA-INR_em2	Carola
Brk_Emb1	Carola
Brk_Emb2	Carola

3. Live imaging settings and analysis:

With the help of the MRI platform I set the Airy Scan confocal microscope LSM 880 live imaging settings at 4 seconds resolution. However, M. Dejean under the supervision of Dr. Dufourt and Dr. Lagha modified the previous settings. Indeed, two problems needed to be solved:

1. The number of z-stacks was not enough to capture the TS fluorescence dynamics through interphase 14.

2. During the first trimester of 2019, the microscope laser power dropped by almost 80%, therefore the acquisitions made during this time were not reliable. Unfortunately, I did not systematically measure the laser power. Thus, I did not realize this important laser drop. However, the MRI facility did not realize it either.

To solve this, M. Dejean modified two parameters:

1. The Z-stacks went from 25 to 30.

2. The average from 2 to 1, in order to keep the same temporal resolution (4 sec).

Concerning the laser power, we verified systematically before start any acquisition.

The software for live imaging analysis was developed in the laboratory by Dr. A. Trullo. He developed a GUI that allowed us to track and quantify TS intensities in each nucleus. The analysis of the movies was a collective work. Indeed, M. Dejean developed a R© script to fast sort the data, to have an overview of synchronies and intensities and to verify my analysis. On my side, I made the data analysis using two statistical software: R© studio and Graph Prism©. False-color movies were generated by Dr A. Trullo. I made the figures and graphs presented in this manuscript under the supervision of Dr. M. Lagha, with the exception of the supplementary table 2, a courtesy of M. Dejean.

4. SmiFISH

I adapted the smiFISH protocol from (Tsanov et al., 2016) to the fly embryo with the advice of Dr. Marion Peter from E. Bertrand's laboratory.

The protocol is described in page 127, supplementary protocol. Design of the primary probes were done using a software developed by Dr. Thierry Gostan (Biostatistical platform IGMM).

With the help of the MRI platform I set up the settings for the smiF acquisition, as well as a pipeline for data quantification using the commercial *Imaris* software. M. Dejean under the advice of Dr. Dufourt and Dr. Lagha, verified the microscope settings which he found were not optimal. The pipeline of analysis and the data quantification were validated by Dr Lagha and M Dejean. I performed all figures and statistical analysis in this manuscript under the supervision of M. Lagha.

Of note, after comparing the smiFISH signal with that obtained from a Stellaris smFISH, we realized that the smiFISH labelling/imaging was not optimal. Thus, more work will need to be done to optimize this part of the project.

5. Mathematical Model:

The mathematical model was developed by Pr. O. Radulescu and Dr. E. Bertrand. The preliminary analysis and data sorting were performed by Dr. M.Lagha and myself.

6. Manuscript:

Under the supervision of Dr. M. Lagha, I wrote this manuscript. A. Trullo aided me in writing the live imaging analysis section in material and methods.

1. **Lightening up gene activation in living Drosophila embryos (Technical Review)**

Title: Lightening up gene activation in living Drosophila embryos

Carola Fernandez and Mounia Lagha*

Affiliation :

Institut de Génétique Moléculaire de Montpellier, University of Montpellier, CNRS, Montpellier, France

*Correspondence: mounia.lagha@igmm.cnrs.fr

Chapter 5

Lighting Up Gene Activation in Living *Drosophila* Embryos

Carola Fernandez and Mounia Lagha

Abstract

With its rapid development, ease of collection, and the presence of a unique layer of nuclei located close to the surface, the *Drosophila* syncytial embryo is ideally suited to study the establishment of gene expression patterns during development. Recent improvements in RNA labeling technologies and confocal microscopy allow for visualizing gene activation and quantifying transcriptional dynamics in living *Drosophila* embryos. Here we review the available tools for mRNA fluorescent labeling and detection in live embryos and precisely describe the overall procedure, from design to mounting and confocal imaging.

Key words Live imaging, Transcription, Embryo, *Drosophila*, mRNA, MS2/MCP system

1 Introduction

> Progress in science depends on new techniques, new discoveries and new ideas, probably in that order
> —Sidney Brenner

Two major recently developed technological breakthroughs, CRISPR/Cas9 gene editing and the deployment of fast imaging microscopy techniques have opened up new avenues for the study of gene expression during the development of multicellular organisms. Owing to decades of genetic manipulations, whole genome profiling and large scale in situ hybridization experiments, the *Drosophila* embryonic blastoderm embryo provides among the best-characterized gene expression patterns. However, the majority of these studies have been performed on dead embryos, fixed at specific development stages, thus lacking the temporal aspects of transcription.

By adapting the MS2/MCP mRNA labeling system to the *Drosophila* embryo [1, 2], it is now possible to visualize transcriptional activation from endogenous loci (or transgenes), with a high temporal resolution (in the order of seconds), in an live developing embryo.

Yaron Shav-Tal (ed.), *Imaging Gene Expression: Methods and Protocols*, Methods in Molecular Biology, vol. 2038,
https://doi.org/10.1007/978-1-4939-9674-2_5,

In this technical review, we precisely describe the protocol to image nascent mRNAs in living early *Drosophila* embryos, from the creation of transgenic flies to image acquisition and analysis.

2 Materials

In developing *Drosophila* embryos, transcription is usually monitored using a two-component tagging strategy: mRNAs containing multimerized tags and a fluorescently labeled detector. By binding several fluorescent detectors, mRNAs containing repeated tags will appear as diffraction-limited spots. Multiple tag/detector couples have been used in cell culture and yeast [13–16] but so far only the MS2/MCP and PP7/PCP systems, both adapted from bacteriophages, have been implemented in *Drosophila* embryos [1–12].

2.1 Tags Used in Drosophila

Two types of vectors are engineered; depending on whether the goal is to monitor transcription from a transgene or from an endogenous locus.

To monitor nascent transcription from a transgene, an enhancer-promoter-reporter construct is inserted into an appropriate receiving plasmid for either targeted genomic or random insertion [17, 18].

To monitor transcription from an endogenous locus, gene editing by CRISPR/Cas9 has become the tool of choice [4, 11, 12]. In addition to one or two guide RNAs containing plasmids, a donor plasmid with homology arms fused to MS2 or PP7 tags is generated and co-injected.

A key consideration in the design strategy is the location of the tag: 5′UTR, introns or 3′UTR (Fig. 1). The three locations have pros and cons (*see* **Note 1**) [19].

Another important choice is the number of repeats. Multimerizing the tags significantly enhances the signal but may potentially impeding RNA biogenesis (*see* **Note 2**).

1. MS2 tag: the MS2 tag is the most popular system used to visualize transcriptional dynamics in real time. A typical strategy is to clone a 24X repeated sequence (~1.3 kb) in the 5′UTR region of a reporter gene [1, 5–11] (*see* **Note 3**).
2. PP7 tag: PP7 repeats are often used as an orthogonal system, which when combined to MS2, allows for the simultaneous tagging of two parts of an mRNA [9] or of two different mRNAs [10].

In Drosophila, 12X PP7 repeats have been successfully used to image mRNA in ovaries [20] but its detection in embryos is typically achieved using 24X PP7 repeats [4, 10–12].

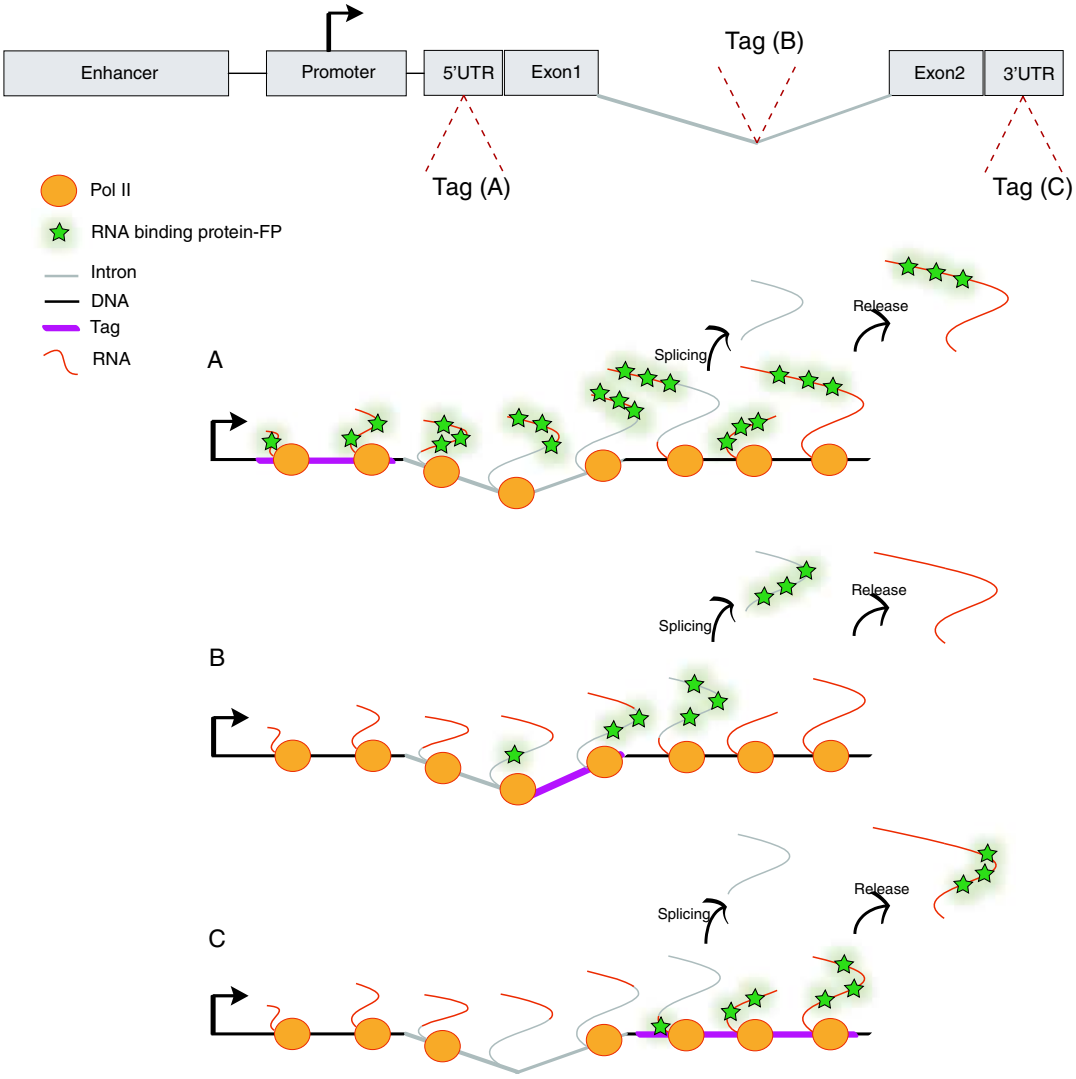


Fig. 1 Three possible locations for insertion of the stem loop repeats (mRNA tag). Schematic representation of an enhancer-promoter-reporter construct with three possible locations of the tag: in the 5′UTR (**a**), in an intron (**b**) and in the 3′UTR (**c**). As exemplified with multiple elongating RNA Pol II (orange circles), the position of the tag (purple box) will influence the signal intensity and its persistence

Importantly, increasing the length of the reporter gene will increase the signal persistence. In the example schematized in Fig. 1, the coding sequence of the *yellow* gene (not expressed during early fly embryogenesis) is used because it contains a 2.7 kb long intronic sequence, which has recently been further extended by inserting a 1 kb long sequence [10].

2.2 Detectors

MS2 coat protein (MCP) and PP7 coat protein (PCP) are the RNA-binding proteins that specifically recognize MS2 and PP7

Table 1
Fluorescent MS2 and PP7 detectors in *Drosophila*

Detector	Reference
hsp83-MCP-GFP	Forrest and Galvis (2003) [21]
nos-MCP-GFP-NoNLS	Garcia et al. (2013) [1]
pNOS-NoNLS-MCP-mCherry	Huang et al. (2017) [22]
NLS-PCP-GFP	Halstead et al. (2015) [20]
NLS-MCP-RFP	Halstead et al. (2015) [20]
nos-SV40NLS-tdTomato-PCP	Fukaya et al. (2016) [9]
nos-SV40NLS-mCherry-PCP,His2AV-eBFP2	Lim et al. (2017) [11]
nos-MCP-GFP,His2Av-mRFP	Lim et al. (2018) [12]
nos-NLS-PCP-(mKate2)$_3$	Chen et al. (2018) [4]
vas-NLS-MCP-(mTagBFP2)$_3$	Chen et al. (2018) [4]

loops contained in mRNA respectively. They can be fused to a variety of fluorescent proteins listed in Table 1 [1, 4, 9, 10, 12, 20–22].

Until 2013, the MCP detector was found to generate aggregates, which appeared as bright nuclear spots even in the absence of MS2 containing mRNAs [23]. This artifact significantly delayed the wide deployment of MS2 type of technologies for imaging transcription in living fly embryos. Moreover, free unbound detector molecules always result in significant background signal.

These two important challenges were unraveled by playing with the localization of free diffusing fluorescently-tagged MCP [13] and by tuning its expression level [24]. For example, Garcia et al., [1] created an MCP-GFP plasmid without a nuclear localization signal (NLS) under the control of the *nanos (nos)* enhancer and promoter sequences. When expressed in a transgenic fly, this MCP-GFP does not form aggregates in embryos and results in a low nuclear background signal. However, if the goal is to monitor cytoplasmic mRNAs, fluorescently-tagged MCP must be trapped in the nucleus to allow a minimum amount of free cytoplasmic MCP using an NLS [20].

2.3 Transgenic Fly Lines

To image nascent mRNAs in living *Drosophila* blastoderm embryos, a minimum of three fly stocks are typically required:

1. A stock expressing the tagged gene of interest (e.g., *MS2-reporter*).

2. A stock containing the maternally expressed fluorescent detector (e.g., *nos-MCP-GFP*), necessarily deposited by the mother since the MCP needs to be present in the freshly laid embryos.
3. Optional, but highly recommended, a transgene for nuclei detection, typically a fluorescently tagged histone or nucleoporin transgene (e.g., *H2av-mRFP/eBFP2*) [1, 11] or (e.g., *mRFP-Nup107*) [2, 25].

2.4 Embryo Collection and Mounting

2.4.1 Embryo Collection

1. Apple juice agar plates (35 mm).
2. Small cages (35 mm).
3. Active dry yeast.
4. Paintbrush.

2.4.2 Mounting

1. Breathable Biofoil film.
2. Heptane-glue.
3. Double-sided tape.
4. 10S Oil.
5. Cover glass (20 × 20 mm).
6. Plastic embryo holder (mounting slide).

3 Methods

3.1 Embryo Collection and Mounting

1. Fly stocks and crosses: virgin females carrying a fluorescently tagged MCP and a nuclear marker (e.g., *nos-MCP-GFP, His2Av-mRFP*) are crossed with males carrying an integrated reporter-transgene of interest (Enhancer > promoters > MS2/PP7 stem loops > reporter gene) or an endogenous locus edited with MS2/PP7 repeats. Since fluorescent detectors (and nuclear marker) are expressed from a maternal promoter, it is essential to perform the cross in this order. The cross is performed in a normal feeding tube (Fig. 2a left), at least 1 day in advance, to allow fertilization of all females. The optimal temperature is between 21 °C and 25 °C but should be kept consistent between experiments [8].
2. To collect embryos, 1 day before the collection, transfer the cross into a collection cage (Fig. 2a right) covered with an apple juice agar plate containing fresh yeast.
3. For embryo mounting, prepare a mounting slide with a breathable Biofoil film (Fig. 2b). Identify the hydrophobic side of the film, on which the embryo will be mounted. Note that the hydrophobic side of the membrane will resist to labeling with a permanent marker, while the hydrophilic side will be easily labeled. Mount the membrane into the membrane holder (Fig. 2b).

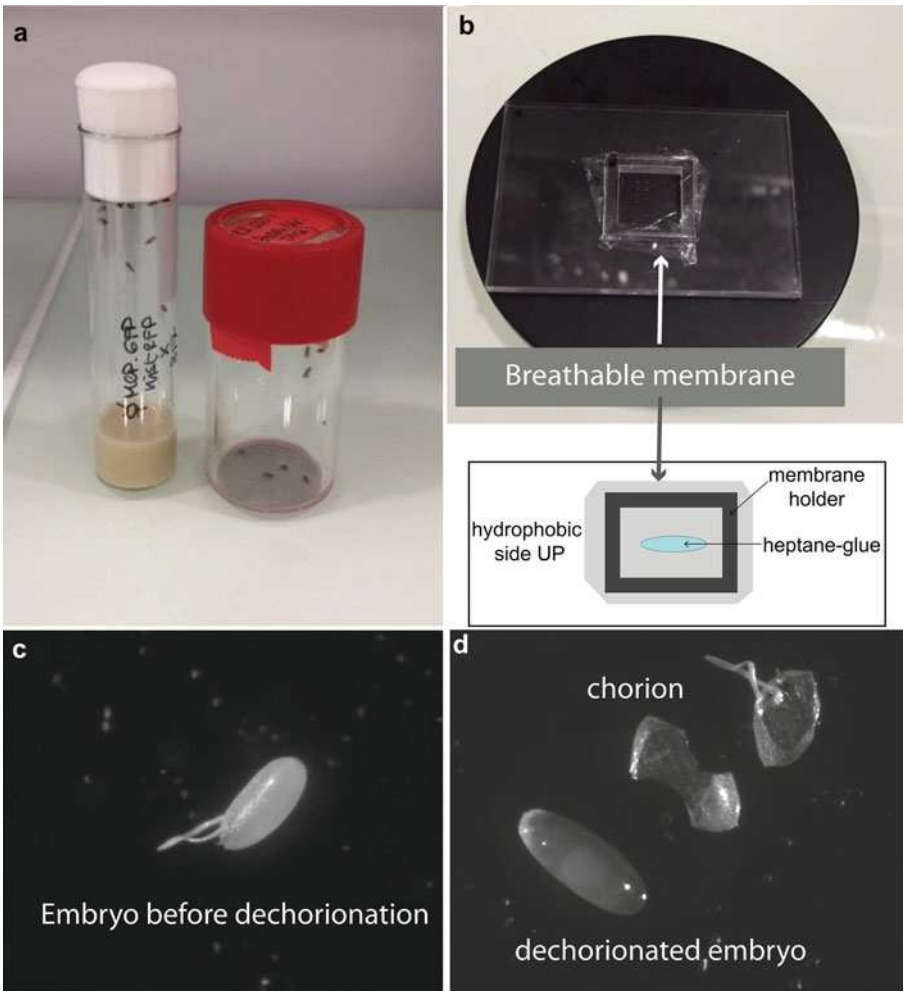


Fig. 2 Procedure for collecting and mounting *Drosophila* embryos for live imaging. (**a**) Collection tube (left) to set up the cross and collection cage (right). (**b**) Image of a mounting slide and schematic representation of mounting set-up. (**c**, **d**) Images of a *Drosophila* embryo before (**c**) and after (**d**) manual dechorionation with double-sided tape

With a Pasteur pipette, put a few drops of heptane-glue (~25 μL, *see* **Note 4**) on the hydrophobic side of the mounting slide. While waiting for the heptane-glue to dry, proceed with the embryo manual dechorionation procedure.

4. Embryo dechorionation is performed under a binocular magnifier (Nikon SMZ745) 2–5X, preferentially under a nonheating light source to avoid embryo dehydration. After the proper timing of egg-laying, replace the plate containing the freshly laid eggs with a new plate. Transfer several embryos gently with a paintbrush to a dechorionation slide, which consists of a microscope slide where a piece of double-sided tape has been attached.

With another small piece of double-sided tape, roll the embryo on the tape to manually remove the chorion, (this step must be performed rapidly to avoid embryo dehydration) (Fig. 2d). It is also possible to remove the chorion chemically with bleach.

5. For mounting and orienting, transfer each dechorionated embryo immediately to the mounting slide; place them on top of the heptane-glue surface (which should be dry) (*see* **Note 5**). Each embryo must be oriented according to the region of expression of interest (*see* **Note 6**). Proceed with the next embryo, trying to mount it in a line. Aligning and orienting the mounted embryos (horizontally or vertically) will facilitate setting up the positions on the microscope.

   Put a drop of 10S Oil on the surface of a coverslip (~80 μL) then invert the coverslip over the mounting slide to cover the embryos and carefully press the coverslip over the mounting slide to slightly flatten embryos without moving them.

3.2 Imaging

1. Imaging is performed using laser scanning confocal microscopy (*see* **Note 7**). The latest generation of confocal microscope allows for a significant gain in temporal resolution. For example, with the Zeiss LSM 880 Fast AiryScan microscope, it is possible to image up to four times faster than a classical confocal while gaining a factor of 1.3 in resolution. We typically image a square region of 70 × 70 μm with 25 z-stacks spaced by 0.5 μm, with a zoom of 3X (Fig. 3a–d). This area corresponds to about 15% of the whole embryo (Fig. 3). With the settings described, it is possible to reach a 4 s temporal resolution per frame.
2. We use a magnification of 40X.
3. A lens with a 1.3 Numerical Aperture is used. The NA needs to be high enough to improve the spatial resolution. Indeed, the size of the finest detail that can be resolved is proportional to $\lambda/2\mathrm{NA}$.
4. Excitation wavelength used is 488 nm for the green channel, but depends on the dye used (here GFP) and 561 nm for the red channel (nuclei labeling).
5. The excitation power of the 488 nm laser is maintained around 0.8% and the 561 nm laser at 0.7%. When the signal is strong (for example with 5′ tags), we recommend choosing a low laser power to avoid photobleaching, phototoxicity and signal saturation (*see* **Note 8**).
6. Resolution: the pixel size is 120 nm.
7. Detector Gain: master gain = 850; digital gain = 1.5. The gain should be carefully fixed; if it is too high, the background will

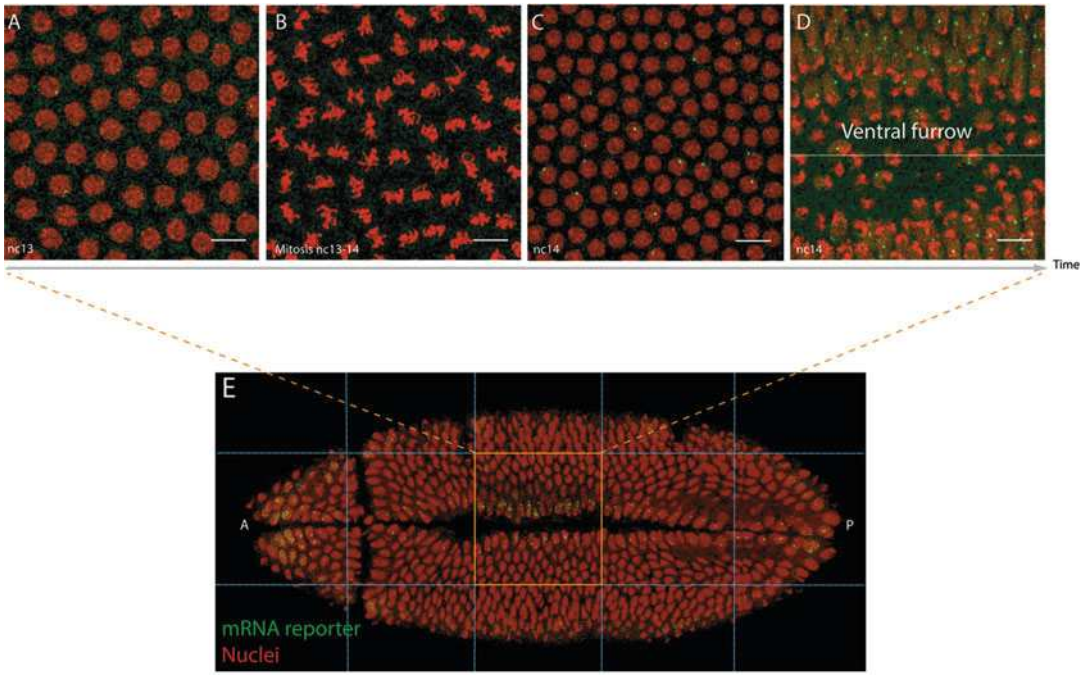


Fig. 3 Live imaging of nascent mRNA in an early *Drosophila* embryo. Maximum intensity projected (MIP) image extracted from a typical movie showing the transcriptional foci in green (MCP-GFP) and the nuclei in red (His2Av-mRFP). (**a**) Snapshot of a MIP Z-stack image at nuclear cycle (nc) nc13. (**b**) Snapshot of a MIP Z-stack image at mitosis between nc13 and nc14. (**c**) Snapshot of a MIP Z-stack image at mid nc14. (**d**) Snapshot of a MIP Z-stack image at the end of nc14 when gastrulation occurs (the ventral furrow is indicated with a dashed line). Gastrulation is used as a landmark to determine the dorsoventral (D/V) axis. (**e**) MIP Z-stack tile scan at the end of the nc14 (ventral view). The typical area of acquisition is indicated in orange, while the blue lines indicate the various tiles (*A* anterior, *P* posterior)

be amplified. If modifications have to be made, first lower the digital gain.

8. At the end of each acquisition, it is important to locate the position of the imaging area with respect to the whole embryo. To this end, we recommend recording a tile scan of the embryo after gastrulation, centered on the imaged area, with typically five horizontal and three vertical tiles and a 10% overlap. With this information, it is possible to extract precise anteroposterior and dorsoventral coordinates (Fig. 3e).

3.3 Data Analysis

3.3.1 Nuclei Segmentation and Tracking

In general, nuclei are automatically segmented using either DNA labeling (*Histone-RFP/Histone-BFP* transgenes) or nuclear envelope labeling (*Nup-RFP*). While this segmentation can be automatically achieved in early interphases using several commercial software (e.g., Imaris) or already developed routines (e.g., ImageJ plugins), it is more challenging when nuclear density becomes important in nc14 or during mitosis, when nuclei move and change

their shape. Recently, we have developed a user-friendly open access software (coded in Python) named *MitoTrack*, able to automatically segment and track nuclei during mitosis [26].

3.3.2 2D vs 3D Spot Segmentation

Two types of quantitative analyses can be envisaged. If the objective is to record the timing of first detection of the transcription site for each nucleus, spots segmentation can be performed in 2D. In this case, custom-made or commercial software will segment and track nuclei and transcription spots in maximum intensity projected images.

However, if the aim is to quantify fluctuations intensities of transcription spots to infer promoter dynamics, it is then essential to analyze the data involving all the z-planes in order to determine intensities of the spots in 3D [27] (*see* **Notes 8** and **9**).

4 Notes

1. To monitor transcriptional dynamics, three different locations of tag insertion are possible: 5′UTR, introns and 3′UTR (Fig. 2).

   5′UTR insertions maximize signal intensity, as fluorescence will be produced by all elongating polymerases along the whole length of the transcribed locus. Fluorescence appearance will closely follow transcriptional initiation and will faithfully reflect the initiation of a burst. However, fluorescence signal will vanish only after all elongating polymerases will reach the 3′ end and be released. Thus, the signal will persist for some time, even if the promoter is in an inactive (off) state (Fig. 2a). Moreover, 5′-tagging can perturb mRNA biogenesis.

   As an alternative to a 5′ insertion, multimerized tags can be inserted into intronic sequences. If splicing is cotranscriptional, signal persistence will be rather short and intensity will depend on intron length. If splicing is post-transcriptional, signal persistence at the transcription site will depend on un-spliced mRNA retention. The main advantage of intronic tagging is its compatibility with endogenous key developmental genes targeting. Since the tag will be spliced, it is unlikely to perturb mRNA export and translation (Fig. 2b). However, care should be taken not to perturb intronic enhancers and splicing donor and acceptor sites.

   Finally, 3′UTR tag insertions allow for a more accurate estimation of promoter activities. The main inconvenience of 3′UTR tagging is that tagged mRNAs signals are generally weak, as the fluorescence has less time to accumulate (Fig. 2c). This tagging strategy can perturb mRNA stability.

2. Depending on the aims in terms of signal amplification, different numbers of tag multimers can be used. In *Drosophila* embryos the current popular strategy is to insert a 24X tag repetition. This signal amplification permits the visualization of transcriptional sites (TS) but is not able to reach single molecule sensitivity in the early *Drosophila* embryo. However recent improvements of the MS2 tag, with multimerization up to 128 repeats (packed into a 3 kb sequence) opens new opportunities for single molecule visualization [28].
3. When constructing the tagged reporter, it is important to check that the multimerized tags (MS2 or PP7) are devoid of transcription factor binding sites that operate during the embryonic stage and the spatial domain of interest. For example, it was recently reported that the initial version of the 24X MS2 repeats contained several binding sites for the transcription factor Zelda [3].
4. The surface of heptane glue on the hydrophobic side of the membrane must be even to correctly flatten the embryos. It is essential not to use too much heptane-glue as it will create an uneven surface and will take longer to dry.
5. When mounting the embryos, do not align more than 20 embryos. Otherwise they tend to dehydrate, which usually leads to embryonic lethality.
6. Some confocal microscopes (e.g., LSM880 Fast Scan mode) do not allow stage rotation, therefore, is important to correctly orient each embryo during the mounting step.
7. In principle, imaging could be achieved with other technologies than confocal microscopy. For example, imaging with a light sheet microscope certainly provides better illumination of the acquired area [30], nevertheless, *Drosophila* embryo orientation remains difficult. When imaging with a light sheet microscope, each embryo is mounted into an agarose block. Once mounted, the overall block can be rotated to correctly orient the embryo, however, the embryo itself cannot be moved directly.
8. In order to perform quantitative analyses and compare various acquisitions, it is important to record the laser power after each acquisition with a power-meter (ie: thorlab PM100A). Indeed, lasers always display intrinsic fluctuations.
9. With current labeling methods, single molecule detection has not yet been reached in living *Drosophila* embryos. Each transcriptional dot therefore contains several nascent mRNAs. In order to estimate the average number of fluorescent signals corresponding to a single mRNA molecule, it is possible to calibrate the live data with information from single-molecule fluorescence in situ hybridization (smFISH) experiments

[29]. Given the intensity of a single molecule, extracted from smFISH experiments, one can estimate the average number of nascent mRNAs at the transcription site. With this information, it is possible to calibrate fluorescent signals from live imaging data and infer the average fluorescent signal corresponding to a unique molecule of tagged mRNA [1].

Acknowledgments

We are very grateful to J. Dufourt for critical reading and insightful suggestions on the manuscript. We thank M. Bellec, M. Dejean, and A. Trullo for helpful discussions. We acknowledge the imaging facility MRI, member of the national infrastructure France-BioImaging supported by the French National Research Agency (ANR-10-INBS-04, "Investments for the future"). The ERC Sync-Dev and HFSP-CDA grants supported this work.

References

1. Garcia HG, Tikhonov M, Lin A et al (2013) Quantitative imaging of transcription in living Drosophila embryos links polymerase activity to patterning. Curr Biol 23:2140
2. Lucas T, Ferraro T, Roelens B et al (2013) Live imaging of Bicoid-dependent transcription in Drosophila embryos. Curr Biol 23:2135
3. Lucas T, Tran H, Perez Romero CA et al (2018) 3 minutes to precisely measure morphogen concentration. PLoS Genet 14: e1007676
4. Chen H, Levo M, Barinov L et al (2018) Dynamic interplay between enhancer–promoter topology and gene activity. Nat Genet 50:1296–1303
5. Bothma JP, Garcia HG, Esposito E et al (2014) Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living Drosophila embryos. Proc Natl Acad Sci U S A 111:10598
6. Esposito E, Lim B, Guessous G et al (2016) Mitosis-associated repression in development. Genes Dev 30(13):1503–1508
7. Ferraro T, Esposito E, Mancini L et al (2016) Transcriptional memory in the Drosophila embryo. Curr Biol 26:212–218
8. Dufourt J, Trullo A, Hunter J et al (2018) Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs. Nat Commun 9:5194
9. Fukaya T, Lim B, Levine M (2017) Rapid rates of pol II elongation in the Drosophila embryo. Curr Biol 27:1387–1391
10. Fukaya T, Lim B, Levine M (2016) Enhancer control of transcriptional bursting. Cell 166:358
11. Lim B, Heist T, Levine M et al (2018) Visualization of transvection in living Drosophila embryos. Mol Cell 70:287–296.e6
12. Lim B, Fukaya T, Heist T et al (2018) Temporal dynamics of pair-rule stripes in living *Drosophila* embryos. Proc Natl Acad Sci 115 (33):10430
13. Bertrand E, Chartrand P, Schaefer M et al (2017) Localization of *ASH1* mRNA particles in living yeast. Mol Cell 2:437–445
14. Larson DR, Zenklusen D, Wu B et al (2011) Real-time observation of transcription initiation and elongation on an endogenous yeast gene. Science 332:475–478
15. Urbanek MO, Galka-marciniak P, Olejniczak M et al (2014) RNA imaging in living cells methods and applications. RNA Biol 11 (8):1083–1095
16. Pichon X, Lagha M, Mueller F et al (2018) A growing toolbox to image gene expression in single cells: sensitive approaches for demanding challenges. Mol Cell 71(3):468–480
17. Groth AC, Fish M, Nusse R, Calos MP (2004) Construction of transgenic Drosophila by using the site-specific integrase from phage φC31. Genetics 166:1775–1782
18. Ren X, Sun J, Housden BE et al (2013) Optimized gene editing technology for Drosophila melanogaster using germ line-specific Cas9. Proc Natl Acad Sci U S A 110:19012

19. Ferraro T, Lucas T, Clémot M et al (2016) New methods to image transcription in living fly embryos: the insights so far, and the prospects. Wiley Interdiscip Rev Dev Biol 5 (3):296–310
20. Halstead JM, Lionnet T, Wilbertz JH et al (2015) Translation. An RNA biosensor for imaging the first round of translation from single cells to living animals. Science 347:1367–1671
21. Forrest K, Gavis E (2003) Live imaging of endogenous RNA reveals a diffusion and entrapment mechanism for nanos mRNA localization in Drosophila. Curr Biol 13:1159–1168
22. Huang A, Amourda C, Zhang S et al (2017) Decoding temporal interpretation of the morphogen bicoid in the early drosophila embryo. elife 6:1–21
23. Weil TT, Parton RM, Davis I (2010) Making the message clear: visualizing mRNA localization. Trends Cell Biol 20:380–390
24. Fusco D, Accornero N, Lavoie B et al (2003) Single mRNA molecules demonstrate probabilistic movement in living mammalian cells. Curr Biol 13:161
25. Katsani KR, Karess RE, Nathalie Dostatni VD (2008) In vivo dynamics of Drosophila nuclear envelope components. Mol Biol Cell 19:3652–3666
26. Trullo et al., in preparation
27. Tran H, Perez-romero CA, Ferraro T et al (2018) LiveFly: a toolbox for the analysis of transcription dynamics in live Drosophila embryos. Methods Mol Biol 1863:183–195
28. Tantale K, Mueller F, Kozulic-Pirher A et al (2016) A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting. Nat Commun 7:12248
29. Femino AM, Fay FS, Fogarty K et al (1998) Visualization of single RNA transcripts in situ. Science 280:585–590
30. Keller PJ, Schmidt AD, Santella A et al (2010) Fast, high-contrast imaging of animal development with scanned light sheet-based structured-illumination microscopy. Nat Methods 7:637–642

2. Decoding the impact of promoter sequences on transcriptional dynamics *in vivo*

Title: Decoding the impact of promoter sequences on transcriptional dynamics in vivo

Matthieu Dejean[1], Carola Fernandez[1], Virginia Pimmett[1], Antonello Trullo[1], Edouard Bertrand[1], Ovidiu Radulescu[2] and Mounia Lagha[1]

Affiliation :

[1]Institut de Génétique Moléculaire de Montpellier, University of Montpellier, CNRS, Montpellier, France

[2]DIMNP, UMR CNRS 5235, University of Montpellier, Montpellier, France

The order of the authors will be determined in the future as current members of the Lagha lab are performing experiments to complement this work (Chip-qPCR, Cage analysis and live imaging and analysis of 6 new genotypes with snaE and other Enhancer/Promoter combinations). Moreover, data form a Sna-MS2 crispr line will probably be added to the final manuscript.

Decoding the impact of promoter sequences on transcriptional dynamics *in vivo*

Summary

Transcription initiation is a multistep process requiring the coordination between different molecular events to occur, how core promoters integrate the molecular signaling to direct transcriptional initiation dynamics is unclear. Here we employed quantitative live imaging (MS2/MCP RNA labeling combined with high speed confocal microscopy) to extract transcriptional initiation properties of various *Drosophila* developmental promoters harboring different core promoter motifs. We have developed an image analysis pipeline, embedded into a user-friendly Graphical User Interface, for nuclei tracking and 3D detection of the associated mRNA molecules. Using an innovative mathematical approach, we are able to deconvolve MS2/MCP fluorescent traces to position each individual polymerase initiation event for each nucleus *in vivo*. By combining data from hundreds of transcribing nuclei located in a defined spatial domain within an embryo, we can infer promoter-switching rates for each promoter sequence.

We show that core promoters control the synchrony of gene activation, and that synchrony and mRNA levels are not always correlated. Furthermore, mutations in the TATA box of transgenic promoters, strongly affect synchrony while INR mutations let it almost unchanged. We found that the TATA containing *sna* promoter and the INR containing *kr* promoter, initiate at a similar rate (in the range of 1 Pol II/ 8 seconds). However, these promoters differ by the duration of their ON state and the presence of one (a short lived, like for *sna* promoter) or two OFF states (a short and a longer lived OFF state like for *kr* promoter). Interestingly, mutating the INR sequence of the *kr* promoter is sufficient to eliminate this long lived OFF state and converts the *krINR mut* synthetic promoter to a *sna*-like switching rate. Conversely, adding an ectopic INR sequence at the *sna* synthetic promoter leads to the stabilization of a long metastable OFF state (and a 3-state modeling is required to infer its kinetic parameters). Remarkably, none of the INR manipulation seems to affect Pol II initiation rate. Finally, mutations in the TATA box have an impact primary in the OFF duration (increase ~4X), without adding a supplementary OFF state or refractory period.

Altogether, our results provide evidence of the temporal regulation of the promoter core sequences, suggesting the existence of different modes of regulation mediated by particular canonical motifs during the early *Drosophila* development in a transgenic context.

Key words: *Drosophila* embryo, core promoters, synchrony, transcription kinetics, mRNA levels, core promoter motifs.

Introduction

Transcription is a highly complex process that requires precise assembly of *trans*-acting factors through the recognition of specific regulatory DNA sequences. The advent of new technologies in the last decades has enabled to decipher key steps controlling gene expression in eukaryote organisms. Nevertheless, the precise nature of interactions regulating this process is not yet well understood.

Enhancers and promoters represent DNA regulatory regions responsible for ensuring proper spatiotemporal expression patterns of eukaryotic genes. Enhancers have the ability to increase transcription from a specific gene core promoter, independently of their distance and orientation. These *cis*-regulatory modules have been the subject of numerous studies, in which their role in modulating transcription through the number, availability, and the affinity of TF sites has been intensively studied. However, the role of core promoter sequences on the modulation of transcription is much less clear.

The core promoter has been defined as a specialized short DNA sequence (~100 pb) around the transcription start sites (TSS) of protein coding and non-coding genes, serving as a platform for the assembly of the pre-initiation complex (PIC) with the final aim of recruiting RNAP II. Lately, this definition has been challenged by the recruitment of RNAP II to enhancers (Santa *et al.*, 2010; Koch *et al.*, 2011; Kim and Shiekhattar, 2015). Nevertheless, unlike mRNA produced from core promoters, transcripts from enhancers (eRNA) are short and unstable leading to their fast degradation (Haberle et al., 2018). Thus, a major difference between promoters and enhancers consists in the capacity of promoters to produce stable mRNA transcripts. Promoter sequences harbour conserved elements (from bacteria to human), the role of which has not clearly been elucidated yet. Remarkably, there are no universal core promoter elements that are found in all promoters. In the work by Chen et al., (Chen *et al.*,

2013) Pol II ChIP-Seq experiments in time resolved *Drosophila* embryos revealed that promoter architecture correlated with waves of genome activation during the maternal to zygotic transition. Indeed, promoters of early expressed zygotic genes are enriched with specific core promoter motifs (TATA and INR), in contrast with zygotic genes expressed later, suggesting a dynamic usage of core promoters according to the developmental context.

Furthermore, promoter-swapping experiments of key developmental genes in the *Drosophila* embryo revealed that core promoter sequences play a critical role on RNAP II pausing which in turns influence the synchrony of gene activation. In this study, synchrony was defined as the time needed to achieve coordinate gene expression over 50% of the pattern. Substitutions of paused promoters, which show rapid and synchronous activity, with non-paused promoters, result in slow and stochastic activation of gene expression. In the case of the *snail* gene the loss of synchronous activation by swapping its core promoter with one less synchronous, leads to an abnormal mesoderm invagination (Lagha *et al.*, 2013).

Recording transcription in living cells has taught us that genes are often transcribed in a pulsatile manner, with extended periods of inactivity interspersed with bursts of RNA synthesis (Raj and van Oudenaarden, 2008; Suter *et al.*, 2011; H. G. Garcia *et al.*, 2013; Senecal *et al.*, 2014). This phenomenon is known as transcriptional bursting and it is characterized by three properties, namely: the frequency (how often bursts occur), the amplitude (how strong transcription is during a burst) and the duration (for how long in time a burst last). So far, studies to characterize transcriptional bursting showed that enhancers and transcriptional factors predominantly modulate frequency of transcriptional bursts (Senecal *et al.*, 2014; Fukaya, Lim and Levine, 2016; Larsson *et al.*, 2019), whereas promoters affect their size (amplitude and duration) (Raj *et al.*, 2006; Suter *et al.*, 2011; Nicolas, Phillips and Naef, 2017). In the context of developing organisms, most studies focused on the impact of enhancer sequences on transcriptional bursts, however, how core promoters regulate the burst size aspects are not well characterized.

Taken together, these observations raise questions about the role of promoters on the control of transcriptional initiation. These sequences do not seem to merely serve as a platform for RNAP II recruitment but are likely to act as a key regulatory step to precisely tune levels of gene expression. Therefore, in this study, we aimed at decoding the role of core

promoter sequences on transcriptional dynamics. Our goal is to understand the influence of core promoters in regulating key distinct limiting steps of transcription initiation *in vivo.* Specifically, we studied the impact of promoter sequences and their motifs on transcriptional coordination within a defined pattern (synchrony) and quantified bursting features. To this end, we employed a synthetic controlled approach that combines high-speed microscopy and quantitative methods. To disentangle the contribution of core promoter motifs such as TATA and INR, we quantified various promoter mutant versions and questioned the combined effect of these two motifs. This allowed us to infer key kinetic parameters controlled by core developmental promoters.

Results

A synthetic imaging platform to image promoter dynamics

To address the impact of the core promoter sequences, we undertook a synthetic approach where core promoters of developmental genes were isolated from their native context and inserted with a mini-gene into a defined landing site. With this approach, differences in transcriptional output between the transgenic lines can be solely attributed to the sequence of the promoter. Such comparisons would be impossible if we were to study promoters in their endogenous context. Our mini-genes are controlled by a unique enhancer, from the *snail* distal enhancer and contain a unique reporter (from the *yellow* gene that contains a large intron) assuming that by this way the elongation speed would be the same for the tested transgenes. The 24 MS2 loops have been inserted in the 5'UTR of the reported gene (Ferraro *et al.*, 2016). In this assay the enhancer is located immediately upstream of the minimal promoter to avoid looping (Figure 1A).

To select the core promoters to study, we took into consideration known core promoter motifs (Supplementary Table 1) and the particular developmental context of their cognate genes. Two of the selected core promoters are TATA-containing promoters, namely *sna,* and *wntD.* The *kr* and the *ilp4* promoters contain a canonical INR motif and a bridge motif in the case of *ilp4* (Figure 1A, Supplementary table 2). Even though *kr* is not a TATA-containing promoter, it was shown to be bound by TBP in a chromatin immunoprecipitation assay (Chen et al, 2013). These core promoters are endogenously expressed during the interphase 14 and TSS positions for each promoter were mapped using published CAGE data (Supplementary Figure 1). Currently, we are analysing two more developmental promoters, *rhomboid (rho)* that contains a canonical TATA box and INR, and *brinke*r (*brk*) that is devoid of canonical core promoter motifs.

Imaging and quantifying transcription at high temporal resolution

Given that the recruitment of Pol II occurs in the order of seconds in cultured cells (Steurer *et al.*, 2018), we decided to image our MS2 transgenic *Drosophila* embryos at an unprecedented temporal resolution of 4 sec, possible thanks to a confocal microscope complemented with an Airyscan module. In this study, promoter constructs are controlled by the minimal *snail* distal enhancer activated by the dorsal protein, which is distributed in a ventral-dorsal gradient, with peaks levels in ventral regions (H. Garcia *et al.*, 2013; Dufourt *et al.*, 2018). To ensure that we are comparing nuclei with non-limiting levels of dorsal activator, we use gastrulation as a landmark to define precise dorso-ventral coordinates (Figure 1B, right panel), and select nuclei (~80) in a region of 50µm centred around the ventral furrow where the levels of dorsal protein are not limiting. The fluorescence signal is recorded during the interphase 13 and 14. But, only single nuclear traces during the interphase 14 are analysed using the mitosis 13 to 14 as time zero.
Here is an example of the raw data:
https://www.dropbox.com/s/zw0nnxj7siyo9v5/KrPrem1CF.avi?dl=0

Nuclei position is followed using a nuclear marker *His2Av-mRFP* (in red- Figure 1B). The spot intensity traces from an individual locus were analysed using a user-friendly software developed in the lab, which follows the fluctuation of intensity over time in 3D, and connects each fluorescent spot to its corresponding nucleus, extracting information for each transcribing nucleus during the interphase 14. Each spot intensity is divided by the surrounding nuclei background, and detected intensities coming from sister chromatids are added by the software assuming that replication occurs early during the nc14, with a similar dynamic for all nuclei of a given movie (Trullo et al., *in preparation*) (Supplementary Figure 2).

Core developmental promoters modify the synchrony of nuclei activation

Quantitative *in situ* hybridisation studies detecting nascent transcription showed that genes could activate nuclei expression almost simultaneously or in a synchronous manner, whereas other genes display uncoordinated or stochastic nuclei expression (Boettiger and Levine, 2009; Lagha *et al.*, 2013). By monitoring the time of the first nuclei activation within the acquisition area (Figure 1B), we verified that developmental core promoters control the synchrony of nuclei activation within the pattern of gene expression. Indeed, we observed that *sna*, *kr,* and *wntD* minimal promoters (in this order) show fast and synchronous nuclei activation. Around 10 min into interphase 14, nuclei activation of the *sna* promoter has reached 50% of the pattern (*kr*~ 14min; *wntD*~ 16) and by 30 min, almost 100% of the pattern is filled for the three promoters. In contrast with the *ilp4* promoter which nuclei activation is slower; 50% of the pattern is filled around 27 min and does not reach 100% of pattern activation at 40 min into interphase 14 (Figure 1 C, D). This difference of synchrony from the *ilp4* could be attributed to the limit of signal detection from the microscope. Therefore, to confirm the synchrony pattern of *ilp4* and other promoters we applied a complementary approach. We used single molecule *in situ* hybridation (smFISH). This method is sensitive to signal detection given its ability to detect a single transcript, but it lacks the temporal resolution that live imaging provides as it requires fixed embryos. To quantify the data, we staged the embryos using the invagination of the membrane for cellularization during the interphase 14 (Foe and Alberts, 1983; Farrell and O'Farrell, 2014), and selected embryos at mid interphase 14 (~25-35 min). Analysis is ongoing.

Synchrony and mRNA levels

As promoters are known for controlling levels of gene expression, we tested if synchrony directly correlates with promoter strengths. First, we integrated the area under the curve of intensity traces for each nucleus during interphase 14 and called it integral amplitude (Figure 2A, 2B). Intensity traces correspond to the production of nascent transcripts at the TS. Thus, this approach allows to "estimate" the total promoter activity recorded during interphase 14 for each nucleus. When we compare the amount of total intensity produced for *sna* and *kr* promoters, we observe that it diverges for the two promoters: the *sna* promoter shows

significantly higher integral amplitude than the *kr* promoter (Figure 2C, Mann Whitney test, p-value< 0.0001), even though the *kr* promoter shows higher intensity dispersion. The *ilp4* promoter has very low values of integral amplitude (Figure 2C), and signal intensities for this promoter are very low, close to the limits of our threshold for signal detection (Figure 2D). In conclusion, these two promoters (*sna* and *kr*) activated transcription in a similar temporal manner. However, they differ in their promoter activity. As smFISH detects single transcription, we verified the previous observation by quantifying the levels of mRNA of these promoters at steady state, which is defined here as the timing during which nuclei reach ~ 100% of pattern activation in live imaging data (started at ~25 min for both promoters).

To label our transgenic mRNAs, we used an inexpensive version of the smFISH, named smiFISH, initially developed in cell culture (Tsanov *et al.*, 2016), and we adapted it to the fly embryo. When compared to the expensive Stellaris smFISH method, it is as accurate in detecting single molecules but smiFISH labelling shows more background (Supplementary Figure 3). To circumvent this background issue, we established a rigorous imaging and quantification pipeline. First, we systematically took images for each embryo at the border and at the center of the *sna* pattern (Figure 3A-white square). The border image was used to extract the background signal and to set up a threshold for detection of single molecules (Figure 3B, 3C). The density of single molecules at the border is weaker, which enables a better quantification of single molecule intensities. Indeed, within the central part of the pattern, mRNA density is higher, often resulting in an overlap between two molecules. Figure 3G shows the intensity distribution of single mRNA molecules detected by the commercial *Imaris®* software from raw images, the median intensity value (black line) is used to quantify the amount of single mRNA molecules in the center image (Figure 3E). Transcriptional sites (TS) are also sorted. (Little, Tikhonov and Gregor, 2013), observed in homozygous embryos that during the interphase 13, up to four TS can be detectable, but as sister chromatid loci remain in close proximity during the interphase and because transcription sites occasionally occupy overlapping focal volumes, the number of active loci in a nucleus is challenging to discern. To overcome this issue, the authors used total fluorescence of all transcription sites in a nucleus to measure the instantaneous transcriptional activity. In this study, we used heterozygous embryos to simplify the counting of TS, and quantified the instantaneous

transcriptional activity per nucleus as described in (Little, Tikhonov and Gregor, 2013) (Figure 3F).

The *sna* and *kr* promoters have a similar average of RNA production per TS (*sna* ~29 mol/TS and *kr* ~32mol/TS, Mann Whitney test, *p-value*: 0.64), but *kr* promoter shows a bigger dispersion (higher noise expression) as observed in the live data. As the same reporter gene is used for each transgenic line, it is unlikely that the rate of mRNA degradation changes for both promoters. Differences observed between live and fixed in terms of mRNA quantification are likely to come from the timing of the smFISH acquisition. Figure 2D shows differences in terms of intensity for *sna* and *kr* at the beginning of the nc 14 (~8-23 min) but after this timing, difference is smooth between both promoters producing similar amounts of intensity. This observation is confirmed in the smiFISH data, but does not allow to conclude that the *sna* promoter produces higher mRNA levels than the *kr* promoter.

Therefore, we decided to use another approach to test how similar or different in terms of transcriptional dynamics, the *sna* and *kr* promoters are.

Inferring the kinetics of transcription dynamics of *sna* and *kr* promoters

The fluorescent signal that we measure with live imaging is a combination of several engaged polymerases, giving rise to several newly synthesized transcripts. To estimate the GFP signal produced by a single initiation event, we used the smiFISH based mRNA quantification to calibrate live data and convert MCP-GFP intensities into number of polymerases. Briefly, TS intensities from smiFISH data at the steady state, were converted in number of mRNA molecules (TS instantaneous activity). As one mRNA molecule is transcribed by one single polymerase, we could infer the instantaneous number of polymerases at one TS during the steady state on interphase 14. Taking this value as reference during the steady state we inferred the number of polymerases on interphase 14 in live imaging acquisitions.

Live imaging intensity traces from each promoter were analysed using a deconvolution approach developed by E.Bertrand and O.Radulescu (Tantale et al *in preparation)*. First, single nuclei traces are filtered to detect transcriptional steady state. This refers to the time when the promoters is in full permissive state. We make the hypothesis that right after mitosis the promoter is not in full permissive state and the probability to have abortive initiation events

is higher. To avoid the misleading of the kinetic parameters, analysed traces are taken into consideration after the first building phase (initial intensity slope) (Figure 4A). Second single nuclei polymerase traces go through a numerical deconvolution (Figure 4B), to place each single polymerase initiation event (Figure 4C). Third, a multi-exponential regression fitting is applied to the distribution of waiting times defined as the time between successive initiation events on pooled nuclei data (Figure 4D, 4E). Finally based on the fitting, kinetic parameters are estimated: the *Kon* (rate at which a gene switches from an inactive (OFF) to an active state (ON)), the *Koff* (rate of ON to OFF state) and the *Kini* (rate of polymerase initiation when a promoter is in an ON state) (Figure 4F). Estimation of these parameters allows for inferring promoter kinetics without the arbitrary burst calling. Moreover, here we can distinguish two quantitative inputs affecting the burst size: the ON duration and the rate of Pol II loading. Using this approach[1]*, we observe that the RNAP II initiation rate is similar for *sna* and *kr* promoters. However, *sna* and *kr* promoter cannot be fit using the same mathematical model. Indeed, a two-state model is enough to recapitulate waiting time distributions for *sna*, but this is not the case for *kr* promoter, where a third-state is necessary to capture transcription dynamics (Figure G).

Thus, despite significant sequence divergence, (*sna* has a strong TATA box and no INR, while *kr* has a light TATA motif and a canonical INR element), these two sequences are able to load polymerases at similar rates. The main difference between both promoters rely in the ON and OFF switching rates, as the same mathematical model cannot fit them. We wanted to question why by performing mutation analysis.

To address the impact of *sna* and *kr* specific core promoter elements, namely TATA box and the INR motif, we generated a series of TATA and INR mutants derived from both promoters. (Figure 5A, 6B).

[1] The data obtained using the deconvolution method is preliminary, here, we show the first results obtained so far.

Canonical TATA box controls ON duration potentially through RNAP II recruitment

To assess the role of the TATA box in the *sna* promoter transcription dynamics, two mutations were performed: the TATA *light* mutation where the last 3 nucleotides of the sequence were replaced by the pseudo TATA motif coming from the *kr* promoter (TATAAAA -> TATA**GTT**), and the TATA *strong* mutation where the first 4 nucleotides were replaced (TATAAAA -> **GCAC**AAA).

Synchrony is affected in both mutants (*sna> snaTATA Light mut> snaTATA strong mut*), with a drastic effect for the TATA strong mutation where only ~20% of nuclei within the pattern succeed to activate transcription at 30 min into nc14 (Figure 5B). Average fluorescence levels produced across interphase 14 for the mutants are decreased too (*sna> snaTATA Light mut> snaTATA strong mut*), with bigger fluctuations for the *snaTATA strong mutation*, showing that even though nuclei activation is compromised, the promoter activity is not completely abolished (Figure 5C, 5D). The smiFISH data follow the same tendency as the live results in terms of RNA production, however, the ratio of nascent molecules numbers between promoters is smaller (*sna> snaTATA Light mut* ~1.3 X; (*sna> snaTATA strong mut* ~2x) (Supplementary Figure 4A, 4B). Due to the low percentage of nuclei activation in the *snaTATA strong mut* (Figure 5B), we compared kinetics of the *sna* vs the *snaTATA Light mutation* promoters. Interestingly, a two-state model was enough to recapitulate transcription dynamics from both promoters. Deconvolution approach for single nuclei traces showed that *TATA light mutation* increases Pol II initiation time (~1.5X), decrease the ON duration (~1.3X). But, the strongest effect of *TATA light mutation* was that on the duration of the OFF state (~4X) (Figure 5E, 5F, 5H). We interpret the effect on the OFF duration as a role for the TATA box sequence in stabilizing the TBP protein, which in turn will stabilize the Pre-Initiation Complex (PIC) to recruit the RNAP II.

The INR element affects ON and OFF kinetics rates without changing the synchrony of nuclei activation

To assess the role of the INR motif we made two types of mutations. First, based on the recent study of the Zeiltinger laboratory (Shao, Alcantara and Zeitlinger, 2019). This study showed that stably paused promoters correlate with the presence of an INR sequence containing a G in position +2. In cultured cells, using 3 model promoters in a plasmid, mutating this G into an

A or T drastically reduced Pol II pausing (Shao, Alcantara and Zeitlinger, 2019). Therefore, we tested the influence of the INR motif on transcriptional initiation kinetics by replacing the G+2 base in the *kr* transgenic promoter for a T (*krINRmut G to T*). In the second mutation, we changed the whole INR sequence of the *kr* promoter with the TSS sequence of *sna* promoter (*krINRmut*) a non-INR-containing promoter (Figure 6A). Synchrony did not change for these two types of INR mutant promoters in comparison with the *kr* intact promoter (Figure 6B). Thus, contrary to the TATA box, the INR does not seem to affect synchrony. However, the average fluorescence levels produced across interphase 14 for the *krINRmut* promoter is higher than *kr* and *krINRmut G to T* promoters (in this order) (Figure 6C). Point out that promoter with similar synchronies could display different promoter activities (or mRNA levels).

After applying the deconvolution method on pooled nuclei from several movies of these three genotypes (*kr*, *krINRmut G to T and krINRmut)* we obtained preliminary promoter switching rates, summarized in Figure 6E, 6F, 6G. The interpretation of these parameters is still at its early phase and thus this paragraph should not be considered as definitive.

Dynamics of transcription activation could not be recapitulate using a two-model state for *kr* promoter. However, a two-state model was enough for capture *krINRmut.* Suggesting a role of the INR sequence on the control of ON and OFF duration during the transcriptional activity. In the case of *krINRmut G to T* a two-state model seems enough to recapitulate its dynamics. However, we cannot explain, so far, the dynamics observed for this promoter in terms of switch ON and OFF rates. Therefore, I will focus, on the results coming from the *kr* vs *krINRmut* promoters.

The juxtaposition of the kinetic parameters for the *kr* and from the *krINRmut* promoter, showed similar RNAP II initiation rates (one initiation every 7.5 vs 7.3 sec respectively). Both promoters differ in their ON, OFF durations, and this difference seems to rely on the presence of an INR motif. To test, weather the INR motif is sufficient to modify the ON, OFF durations, we decided to add an INR motif to the *sna* promoter which dynamics can be capture using a two-state model.

The *sna+INR* promoter show a similar initiation rate to those of *sna*, *kr* and *krINRmut* promoters. However, its dynamics could not be capture using a two-state model, these results indicate that the INR is adding a supplementary state on the control of transcription dynamics. So far, the INR motif has been correlated with stable RNAP II pausing promoters (Shao, Alcantara and Zeitlinger, 2019), therefore one can hypothetize that the presence of INR might correspond to a pausing state. Nevertheless to test this hypothesis, complementary approches must to be applied.

To conclude, we cannot explain why these mutations of the INR behave in opposite directions in term of promoter activity/intensity while having similar synchrony profiles. Therefore, it is difficult to understand how these parameters could account for the clear differences in mRNA distribution we see (Figure 6C). Thus, it is difficult to conclude on a role for INR. To clarify the role of the INR, we are currently testing other mutations. Namely, we are testing the effect of the addition of the INR in the in the *snaTATALight mut* promoter (*snaTATALight+INR mut)*. And, in the same trend, we are testing the addition of a canonical TATA box to the *kr* transgenic promoter (*kr+TATA*), to inquire if there is a synergy between the TATA and the INR, and the *kr+TATA_INR mut* to disentangle the contribution of each motif in this context.

Core promoter sequences outside element motifs

To characterize the importance of sequences surrounding core promoter motifs. We took advantage of two mutant constructions that harbour the same promoter elements but differ by the rest of their sequence, Namely, the *snaTATA Light mut* construct, which contain the pseudo TATA motif from the *kr* promoter, and the *krINR mut* that contains the TSS from the *sna* promoter (Supplementary figure 6). In terms of synchrony, we observed that 50% of the pattern is filled in ~12min for the *krINR mut,* and in ~25min in for the *snaTATA Light mut* (Supplementary Figure 3B). It is worth noticing that the spacing in base pair (bp) between the pseudo TATA motif and the TSS of the *snaTATA Light mut* motif (22bp), differs from the *krINR mut* (20bp). Therefore, the higher synchrony of the *krINR mut* could be a result of a better communication between the pseudo TATA and the TSS, or rely on intrinsic/discrete features of the *kr* sequence. Therefore, it might be interesting to test whether the spacing between the pseudo TATA motif and the TSS plays a role in the synchrony, or whether the effect of

replacing the TSS from *kr* promoter by the *sna* one, in an endogenous context, would lead to a higher synchrony as we observe in our context.

Discussion

Using a combination of quantitative imaging and a mathematical deconvolution model, we show that core promoter sequences control the temporal aspects of nuclei activation (referred to as synchrony). Here we obtained evidence that core promoter elements impact differently various steps of transcriptional initiation. Manipulation of the TATA box in the *sna* transgenic promoter, showed its influence to foster nuclei activation within the pattern and control the OFF durations. In contrast, the INR mutations in the *kr* transgenic promoter did not change the synchrony of nuclei activation, neither the RNAP II initiation rate but rather have an influence on the states controlling the ON and OFF switching dynamics during the transcription process.

Promoter elements and transcriptional synchrony

Timing and position of gene expression are essential to specificity tissues and their function in a developing organism. Here we show that, similarly to enhancers, minimal promoters exhibit the intrinsic property to control the timing of transcriptional activation, referred to as synchrony, as suggested by fixed approaches (Lagha *et al.*, 2013). Moreover, we observed that endogenous paused promoters tend to have synchronous profiles (*sna, kr*). However, we also observe that synchrony profiles can be achieved in the absence of pausing like for *wntD*, where the presence of a canonical TATA motif might be responsible for synchronous activation. Synchrony seems to be essential in *Drosophila* development. Indeed (Lagha *et al.*, 2013) showed by swapping developmental promoters with different levels of synchrony could affect the proper mesoderm invagination. Thus, core promoter sequences seem to have the capacity to regulate transcription at different levels. Nevertheless, with our approach, we did not test enough combinations to be able to 'predict' from a given promoter sequence its synchrony profile. In any case, by characterizing new promoters, one can test the importance of synchrony in endogenous conditions. For example, it could be interesting to see if, in its

natural context, the *ilp4* promoter shows the same stochastic profile, and the consequences of changing it for a synchronous promoter.

Core promoter architecture during early development

The awakening of the zygote genome, even though progressive, is marked by two waves of gene activation: a minor wave (~nc 8-13), where ~100 zygotic genes are activated in rapid nuclear cycles. And a major wave (at nc 14) that co-occurs with the mid-blastula transition where the cell cycle is lengthening, and cellularization starts (Chen *et al.*, 2013; Schulz and Harrison, 2019; Vastenhouw, Cao and Lipshitz, 2019).

Chen et al., 2013; show how differential RNAP II occupancy profiles correlated with temporal constraints during the two waves of activation, and how this is correlated with different promoter architectures that seem to be differentially used during these two waves of ZGA. Genes expressed during the minor wave (or before the MBT) are enriched in TATA box motifs (e.g. *sna*), whereas those expressed during the major wave (or during the MBT) are enriched with INR, DPE, and MTE elements (e.g. *kr*).

When we compared the dynamics of two promoters of these two categories (*sna and kr*), we did not notice significant differences in terms of synchrony. However, quantification of their promoter kinetics revealed that, while they share similar rates of RNAP II initiation, the *kr* promoter exhibits different ON and OFF kinetics dynamics. Whether long ON time durations can be generalized to more MBT promoters remains to be proved, however, it is tempting to speculate that short interphases prior to MBT impose particular constraints to promoter architecture allowing them to transcribe more efficiently, possibly via a reduced timing of pausing. Remarkably, the replacement of the TSS of the *kr* promoter with that of the *sna* promoter leads to a significant increase in its promoter strength (without affecting synchrony, thus exemplifying one more time the distinction between synchrony and levels). This increase in mRNA production could not be explained by a more efficient rate of RNAP II initiation as they have similar rates (1 Polymerase every ~8 seconds). Rather, mRNA upregulation seems

to rely on a better synergy between the non-canonical TATA motif from *kr* and the TSS from the *sna* promoter to stabilize TFIID.

Core promoters control burst size

A recent study using single allele RNA sequencing in 7.186 genes from 224 individual primary mouse fibroblast cells, showed that core promoter elements affect burst size. The authors showed that genes with TATA elements in their core promoter had significantly larger bursts than genes without TATA elements (Larsson *et al.*, 2019). By burst size, we refer to the observed number of transcribed polymerases in a burst, multiplied by the ON duration.

Consistent with results from the (Larsson *et al.*, 2019) study, our data show that minimal promoter sequences affect burst size. When we mutated the TATA box in the *sna* promoter, the ON time and the RNAP II initiation rate of the traces are affected, showing a decrease of the burst size. Here, the deconvolution of intensity traces enables to distinguish between two main controllers of burst size: the ON duration and the rate of RNAP II initiation. In *Dictyostelium,* the mutation of the TATA box does not affect the ON duration but rather the RNAP II initiation rate (Corrigan *et al.*, 2016). This suggests different strategies between organisms to regulate the burst size. In our data, in the case of the sna promoter, the distribution of the waiting times can be fitted with a three-parameter bi-exponential (*Kini, Kon and Koff),* thus, a two-state model is sufficient to capture the kinetics operating in early *Drosophila* embryo in our synthetic reporter context. This is not the case for the *eve* promoter, where live-cell measurement showed larger distribution of the burst sizes that did not fit a two state model, but rather a multi-step model, suggesting more complicated regulation (Bothma *et al.*, 2014). Moreover, in our case we did not observe long OFF times (ON durations > OFF durations) like in other systems (Suter *et al.*, 2011; Tantale *et al.*, 2016), offering little possibility to have different refractory steps during the OFF durations.

Interplay between core promoter elements and trans-regulating factors

Very few studies have questioned the impact of the stability and turnover of trans-regulating factors at the core promoter on transcription dynamics. This lack of knowledge may be due to the absence of available tools to monitor at the same time, the assembly of trans-regulating factors and transcription dynamics at one specific locus. Core promoter elements seem to play an essential role to guide the assembly of the General Transcription Factors (GTFs) for the recruitment and well positioning of the RNAP II at the TSS of a gene (Sainsbury, Bernecky and Cramer, 2015; Louder *et al.*, 2016; Patel *et al.*, 2018). The TATA box motif is known to be recognized by the TBP protein, which is an essential subunit of the GTF TFIID. *In vitro* assays have shown that TATA-containing promoters have a high affinity and stabilize the binding of TFIID (Wong and Bateman, 1994; Coleman and Pugh, 1995; Kolesnikova *et al.*, 2018; Bhuiyan and Timmers, 2019). TFIID is essential for the assembly of the PIC and, in consequence, the RNAP II recruitment. Therefore, reducing the TFIID affinity and/or stability could explain why mutations of the TATA box can affect the burst size.

Moreover, TFIID also contacts the INR core element through two other subunits, the TAF 1 and TAF2 (Chalkley, 1999; Sainsbury, Bernecky and Cramer, 2015; Louder *et al.*, 2016; Patel *et al.*, 2018). TAF 1 could act as an activator but also as a repressor of transcription initiation (Kokubo *et al.*, 1994; Struhl and Moqtaderi, 1998; Sainsbury, Bernecky and Cramer, 2015). In our data, the mutation in the INR sequence on the *kr* transgenic leads to opposite results. On one hand, the single mutation of the *INR* sequence at the +2 position (G to T) leads to a decrease of the burst size and the initiation rate (a decrease of the initiation rate corresponds to the increase of time between initiations), suggesting perhaps a destabilization of the TFIID and, in consequence, the RNAP II recruitment, or a destabilization of the RNAP II at the TSS after recruitment. On the other hand, the replacement of the INR sequence with the TSS of the *sna* promoter (*krINRmut*) shows an increase in the burst size, with no changes in the initiation rate. Thus, it is possible that the TSS from the *sna* promoter allows a better communication with the no-canonical TATA box from the *kr* promoter, resulting in a stabilization of TFIID complex. Another possibility is that the absence of the INR motif in this TSS suppresses a negative control in the RNAP II loading.

In conclusion, core promoters tightly tune gene expression levels in time and space, via different combinations of motifs in their sequences. It might be that these motifs influence the mRNA levels produced, by regulating essentially the recruitment and the stability of trans-regulating factors to the core promoter.

Ongoing work

Currently we are testing other developmental core promoters to characterize their temporal and kinetics properties, namely *rhomboid* (*rho*), a TATA-, INR-, DPE-containing promoter, and *brinker (brk)* that does not harbour any canonical promoter motifs.

As the mutations of the INR motif (located at the TSS of the *kr* promoter) do not affect the synchrony, we performed a supplementary mutation to study the role of the *sna* TSS on synchrony. We replaced the TSS of the *sna* promoter by the TSS of the *kr* promoter (TCACAG->TCAGTC).

In addition, we replaced the TSS of the *snaTATA Light mut* for the *kr* TSS. The idea behind this is to compare it with the *kr* promoters and observe the impact of the sequences surrounding the core elements.

Also, we replaced the pseudo TATA motif of the *krINR mut* promoter by the *sna* TATA motif, to test the synergy of the TATA with the TSS in a *kr* sequence context (TATAGTTA->TATAAAAA)

To better characterize the effect of the single mutation of the INR +2 position on transcriptional dynamics, we generated a transgenic line taking advantage of the natural occurrence of T in position +2 in the ilp4 TSS and replaced it by a G (TCA**T**TT-> TCA**G**TT).

Additionally, we mutated the TSS of the brk promoter to transform it in an INR-containing promoter.

All of these transgenic lines are currently under analysis.

Perspectives

The characterization of the role of the core promoter elements using a synthetic approach gives useful information, to understand key steps in transcriptional regulation *in vivo*, nevertheless, the role of core promoters needs to be assessed in an endogenous context with natural enhancers. The advancement of genomic engineering with the CRISPR technology makes it now possible to complement transgenic assays with endogenous data, in which the local chromatin environment, the enhancer-promoter specifications, and the developmental context will also play a role. Thus, in complement, a *sna-MS2-crispr* line was generated and preliminary results show similar initiation rates with the *sna* transgenic promoter.

Material and Methods

Drosophila stocks

For live imaging: virgin females expressing *MCP-GFP-His2Av-mRFP* were crossed with males carrying an integrated reporter-transgene of interest *(SnaE<CorePr<24XMS2<Yellow)*.
For smiFISH: yellow-white virgin females where crossed with males carrying an integrated reporter-transgene of interest *(SnaE<CorePr<24XMS2<Yellow)*.

Cloning and transgenesis

The *snail* enhancer>*sna* core promoter transgene was described in Ferraro et al14. The 24X MS2 tag was inserted immediately upstream of the yellow reporter gene coding sequence, and downstream of the *sna* core promoter. Developmental core promoters used in this assay were cloned from genomic DNA by PCR and inserted into the final plasmid using restriction sites placed on both sides of the core promoter (XhoI and BamHI). All transgenic MS2 flies were inserted in the same landing site (AttP2 landing site on chromosome 3LT) via PhiC31 integrated-mediated transgenesis.

Live Imaging

Embryos were dechorionated with tape and mounted between a hydrophobic membrane and a coverslip as described in (Fernandez and Lagha, 2019). Imaging is performed using laser scanning confocal microscopy Zeiss LSM 880 supplemented by an AiryScan module, with the following settings: GFP and RFP proteins were excited using a 488 nm and 561 laser respectively. The excitation power for the GFP with the Argon laser is set up at 3.8µW before each live movie acquisition. A GaAsP-PMT array of an Airyscan detector was used to detect the GFP fluorescence. The image area covers 540 × 540 pixels; pixel size 0.131; pixel dwell time 0.73 µsec. Images were acquired using a 40x oil objective, 30z stacks 0.5µm apart, with zoom 3 and bi-directional scanning. Under these conditions, the time resolution is of 3.86 sec. Raw data is processed using Airyscan 3D Zen software (Zeiss).

Live imaging analysis

Briefly, processed Airyscan data are analysed with a custom-made software developed in Python® and using PyQt5 library for the graphical user interface (GUI). Typical raw data are multi z-stacks time series of early drosophila embryos, organized in two channels: one for the GFP (MS2-transcriptional sites) the other for RFP (histones-nuclei). The aim of the analysis is to follow nuclei and monitor their transcriptional activity, storing information about transcriptional site intensity, volume, among other features and crossed them with spatio-temporal information. To cover all the interphase 14, we have to work on several raw data files, which we concatenate and maximum intensity project both channels for visualization in the GUI. Nuclei MIP time series is pre-smoothed and then thresholded with Otsu-thresholding algorithm. The resulting binary image is labelled and treated with watershed algorithm to split touching nuclei; finally, the labelled nuclei are tracked following a minimum distance criterion over their center of mass. Spots are instead segmented in 3D, frame by frame, with a blob detection algorithm. Segmented spots are first filtered to remove fake detection and then associated (in 2D) to be connected with the closest nuclei and taking their label. As a result, the information for each nucleus frame by frame, of the 3D volume and the total intensity of its transcriptional site is obtained. This information is spatially organized: we collected nuclei information only in a region of 50µm centred on the gastrulation line (manually set with a pop-up tool). Intensity values of transcriptional sites are corrected for the background: in each time step, the intensity of each spot is divided by the average intensity of the pixels surrounding the spot itself. In this way, we correct the intensity for several perturbations (photo-bleaching, different z value of spots). The developed software provides tools for visualization in every step of analysis and tools for manual corrections, where needed.

SmiFISH

Probes

smiFISH primary probes (recognizing the yellow mRNA) were designed as previously described (Tsanov *et al.*, 2016). Primary and secondary probes (FLAP-Y) were produced by Integrated DNA Technologies (IDT), with the following production details. The primary probes are produced using high-throughput oligonucleotides synthesis in 96-well plates. The total length of primary probes (transcript binding + FLAP-binding) should not exceed 60 nucleotides

for cheaper synthesis. The secondary probes are conjugated to Cy3, through 5' and 3' amino modifications.

Protocol and Image analysis

smiFISH was performed according to the Supplementary Protocol. Imaging stacks were acquired using laser scanning confocal microscopy Zeiss LSM 880 supplemented by an AiryScan module, with the following settings: Cy3 and DAPI were excited using a laser at 555 nm and 415 nm respectively. The pixel size is (XYZ): 0.04× 0.04x 0.20 µm, images were acquired using a 40x oil objective, z spacing 200 nm, with zoom 3, with Airyscan in super resolution mode. We typically obtained stacks representing 20 mm in total axial thickness starting at the embryo surface. Image analysis was performed using the commercial *Imaris*® software (module XT, version 9.2.2).

Deconvolution Method

This method was developed by E.Bertrand and O.Radulescu (Tantale et al *in preparation).* Briefly, the model is defined by 3 parameters or constants: the k1p constant (transition between OFF to ON promoter activity), the k1m constant (transition ON to OFF) and the k2 constant (transcription initiation rate during the ON state). The whole procedure to estimate these three parameters is summarized in three steps: first, the deconvolution of the signal to place each polymerase position enabling the estimation of the distribution of the waiting times between successive initiation events. The second step is a bi-exponential fit with three independent parameters of the distribution function. The last step consisted in the calculations of the kinetic parameters (k1p, k1m, k2) from the parameters of fit. For this purpose, analytical formulas are used: 1/ k1p= T_{off} ; 1/ k1m= T_{on} ; 1/k2= Initiation rate. This procedure is tested using a model simulation to verify the precision of the fitting.

Figure 1.

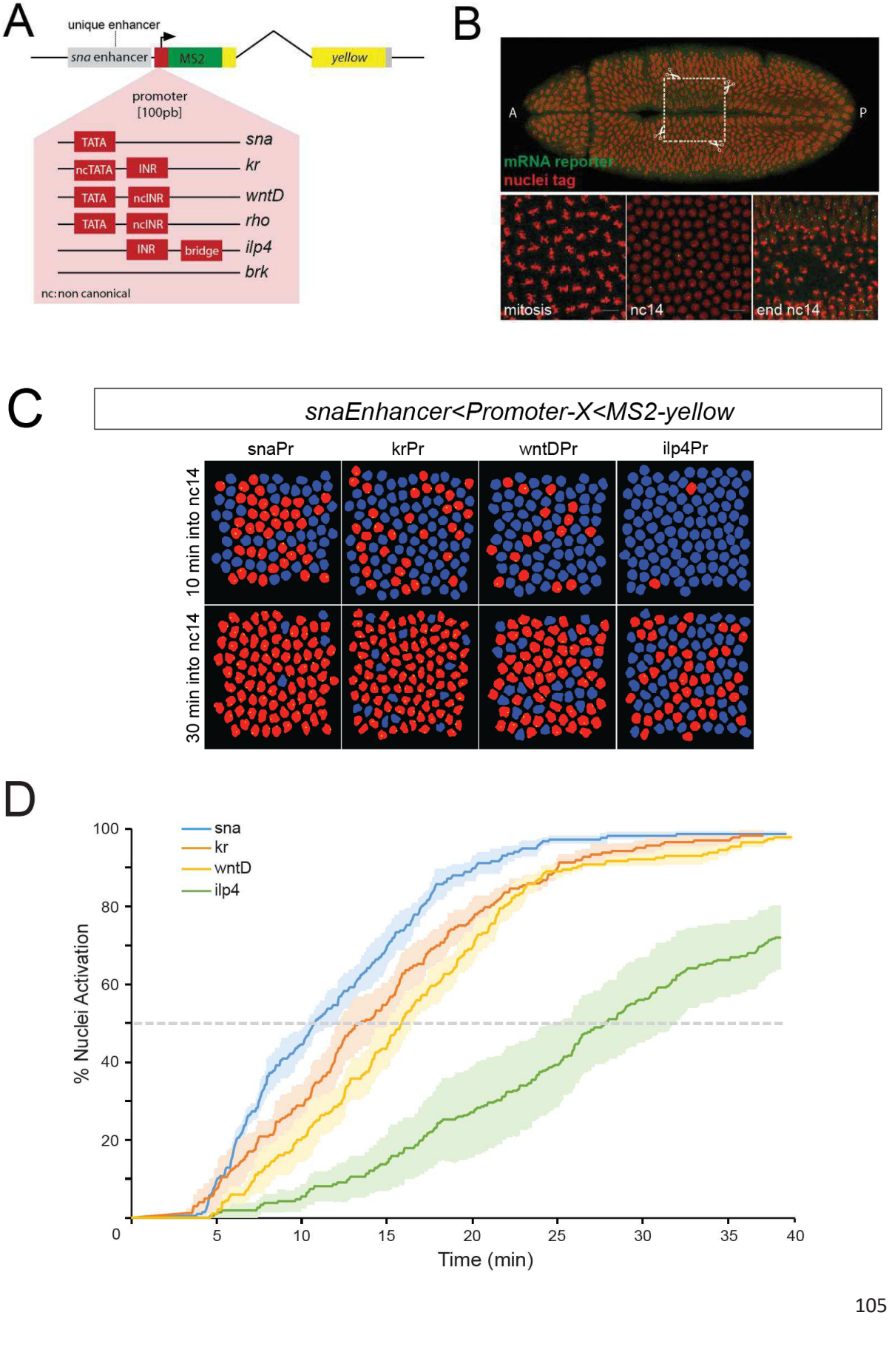
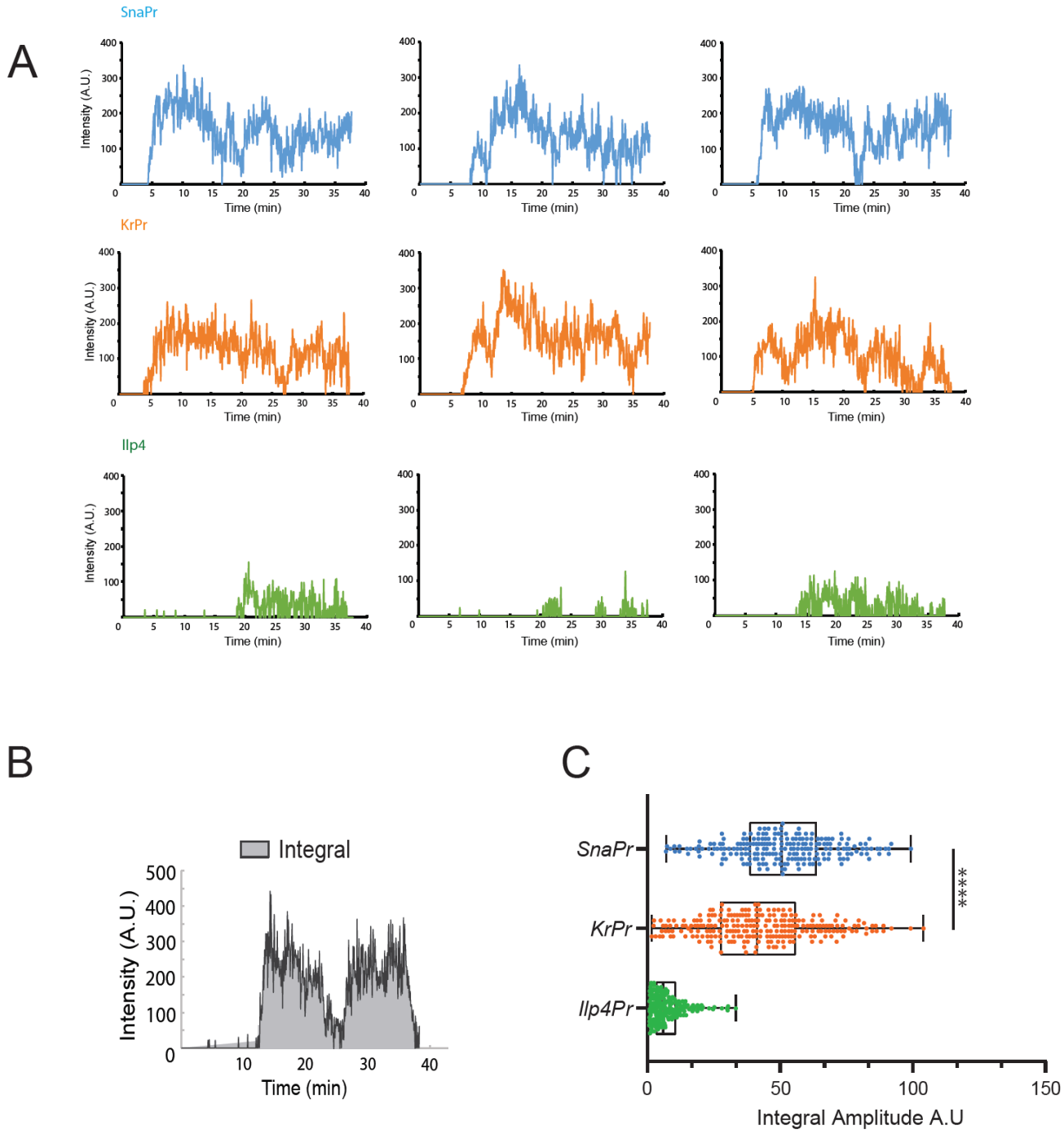

A
unique enhancer
sna enhancer
MS2
yellow
promoter
[100pb]
TATA
sna
ncTATA
INR
kr
TATA
ncINR
wntD
TATA
ncINR
rho
INR
bridge
ilp4
brk
nc:non canonical
B
A
P
mRNA reporter
nuclei tag
mitosis
nc14
end nc14
C
snaEnhancer<Promoter-X<MS2-yellow
snaPr
krPr
wntDPr
ilp4Pr
10 min into nc14
30 min into nc14
D
sna
kr
wntD
ilp4
% Nuclei Activation
Time (min)
0
5
10
15
20
25
30
35
40
20
40
60
80
100

Figure 1. Minimal developmental promoters control the temporal coordination in gene activation within a pattern. (A) Schematic representation of *yellow* reporter gene containing a series of 100 bp minimal promoters (light red box), 24× MS2 RNA stem loops in 3' of the promoter and 538 bp of the *snail* shadow enhancer located in 5' of the promoter. Bright red boxes indicated the core element composition for each tested promoter gene, namely *sna, kr, wntD, rho* and *ilp4*. (B) Upper panel, maximum intensity projected Z-stack tile scan from live imaging movie, representing nuclei labelled with His2Av-mRFP transgene at the end of nc 14. The white square represents the acquisition area (70x70 µm) visualized in the bottom panels. Bottom panels, snapshots of maximum intensity projected Z-stack from live imaging movie, during the mitosis 13-14, mid nc 14, end of the nc 14. (MCP-GFP signal coming from transcribing foci is visualized in the form of green dots). (C) Snapshot of false-colored image, representing the segmented transcriptionally inactive nuclei (blue), active nuclei (red) and associated transcriptional foci (yellow dot). Profiles of instantaneous minimal promoter activation show how the pattern is filled at ~10 and ~30 minutes into nc 14, for each tested promoter. (D) Temporal coordination profiles during nc14. Synchrony curves were obtained by quantifying transcriptional activation during nc14 in a region of 50µm around the ventral furrow: *sna* (3 movies, n =216 nuclei) (blue curve), *kr* (4 movies, n =247 nuclei) (orange curve), *wntD* (3 movies, n = 199 nuclei) (yellow curve), and *ilp4* (4 movies, n = 187) (green curve). Dashed grey line represents 50% of activation. The time origin is the end of the nc13 mitosis. Error bars represent SEM.

Figure 2.


A
SnaPr
KrPr
Ilp4
Intensity (A.U.)
Time (min)
B
Integral
Intensity (A.U.)
Time (min)
C
SnaPr
KrPr
Ilp4Pr
****
Integral Amplitude A.U

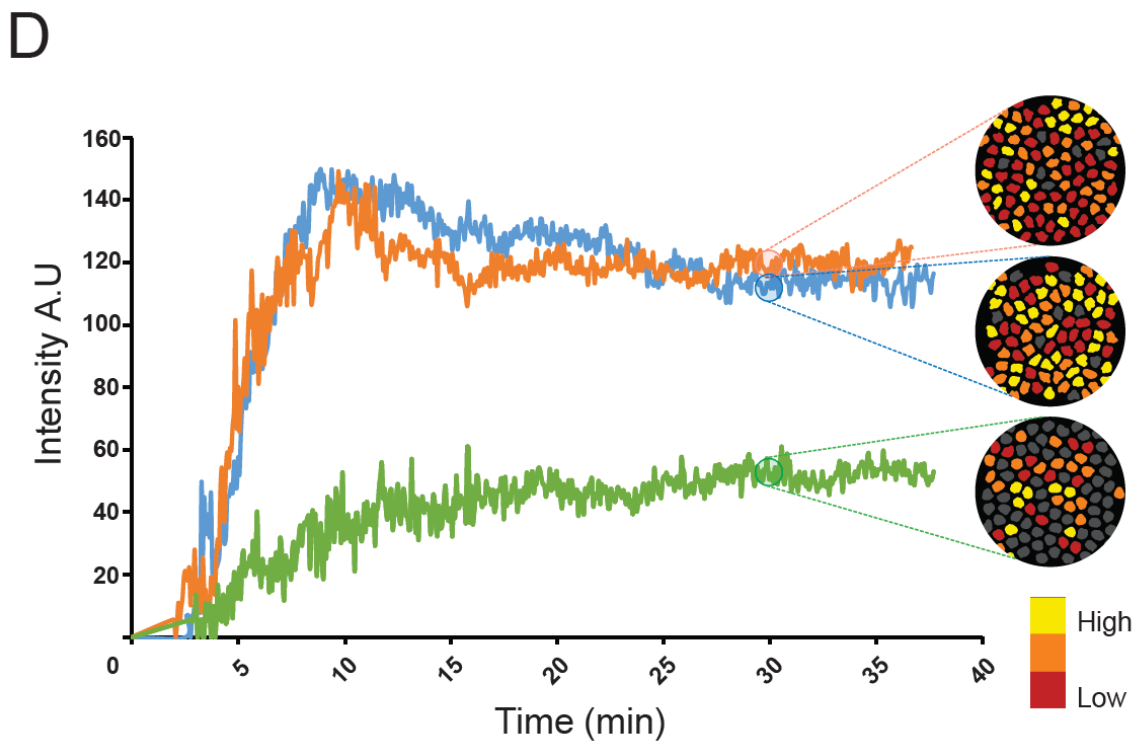


Figure 2. Transcriptional dynamics and intensity levels. (A) Intensity traces for single transcriptional foci through interphase 14 (~38 min) from sna (bleu), *kr* (orange) and *Ilp4* (green) promoters. (B) Schematic representation of how integral amplitude was calculated; the grey surface represents the area under the curve. (C) Boxplot showing the distribution values of integral amplitude for each promoter. Each dot represents the integral amplitude coming from a single nucleus (*sna*, n= 216; *kr*, n=247; ilp4, n=187). Integral amplitude comparison between *sna* vs *kr* promoter, p-value <0.001; Mann Whitney test α= 0.05. (D) Average Intensities in A.U after background division for the *sna*, *kr* and *ilp4* promoters across interphase 14 (~30 min after mitosis). On the right, in a circle, is a heatmap snapshot of the intensities for each promoter at t=~30 min (high intensities are in yellow, low intensities in red, no signal in grey).

Figure 3.

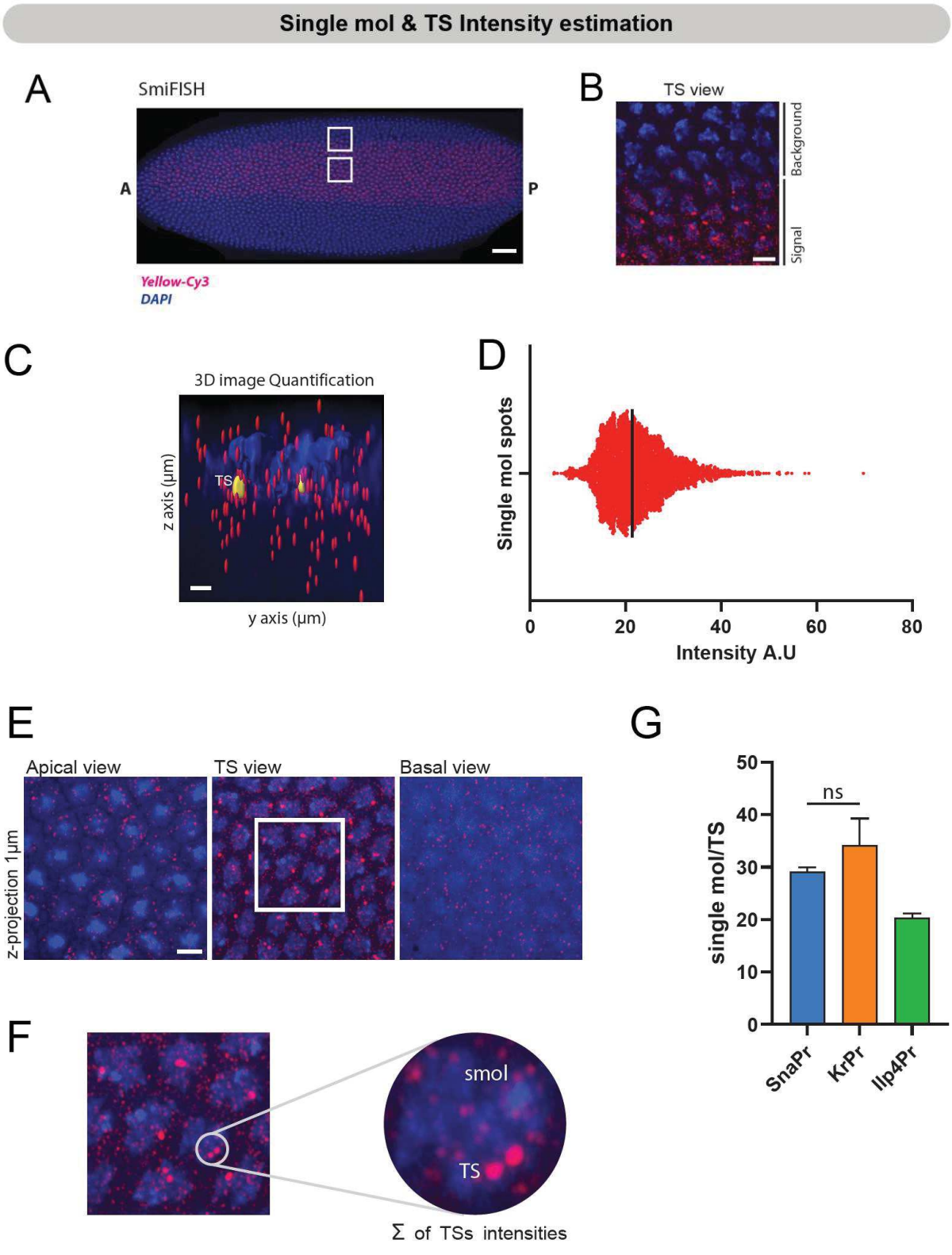
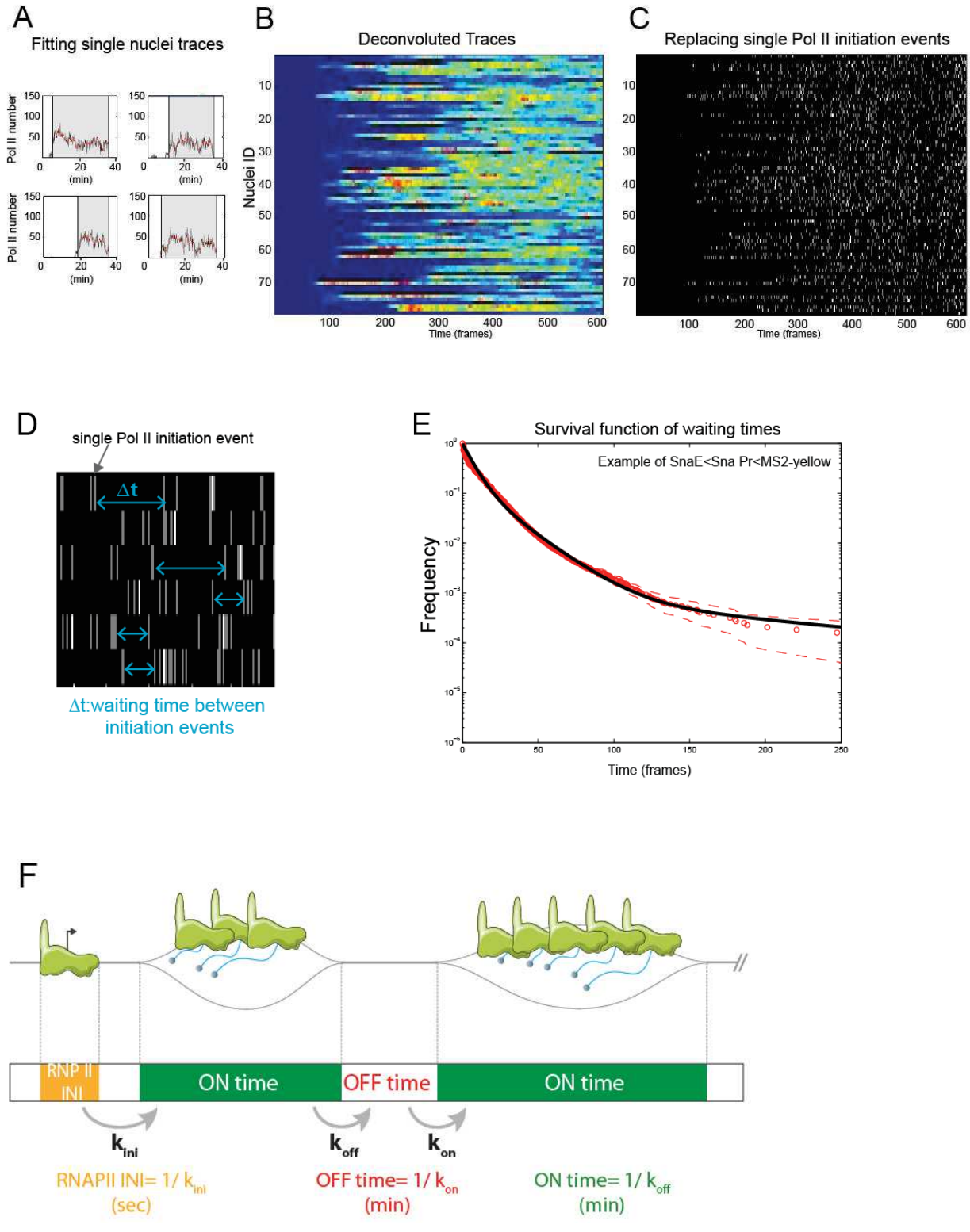

Single mol & TS Intensity estimation
A
SmiFISH
A
P
Yellow-Cy3
DAPI
B
TS view
Background
Signal
C
3D image Quantification
TS
z axis (µm)
y axis (µm)
D
Single mol spots
0
20
40
60
80
Intensity A.U
E
Apical view
TS view
Basal view
z-projection 1µm
F
smol
TS
Σ of TSs intensities
G
single mol/TS
50
40
30
20
10
0
ns
SnaPr
KrPr
Ilp4Pr

Figure 3. Promoter activity by single molecule FISH.(A) Max projected tile scan acquisition of a whole embryo at mid interphase 14, ventral view, labelled with 24X fluorescent oligonucleotide probes against *yellow* (red), and Dapi staining for the nuclei (bleu). The white square representing the image on B at the top and in E at the bottom, scale bar represents 50µm. (B) Single plane at the transcriptional site (TS), of a border image of the pattern, scale bar represents 5µm. (C) 3D graphical representation of detected signal in the *Imaris®* software. Transcription sites are false coloured in yellow, nuclei in bleu and single RNA molecules in red. (D) Horizontal violin plot of the intensity values for each single molecule, the black line represents the median. (E) 1µm Z-projection of the acquired region at top/center/bottom in Z. Single molecule signal shows a higher density at the TS level. The white square representing the image on F. (F) Zoom at the TS level to show detection of sister chromatids during the mid-interphase 14. (G) Histogram showing the average number of RNA molecules per TS for *sna* (bleu) and *kr* (orange); *ilp4* (green), (n= 5 embryos for each tested promoter, at mid interphase 14).

Figure 4.


A
Fitting single nuclei traces
Pol II number
(min)
B
Deconvoluted Traces
Nuclei ID
Time (frames)
C
Replacing single Pol II initiation events
Time (frames)
D
single Pol II initiation event
Δt
Δt:waiting time between initiation events
E
Survival function of waiting times
Example of SnaE<Sna Pr<MS2-yellow
Frequency
Time (frames)
F
RNP II INI
ON time
OFF time
ON time
k_ini
k_off
k_on
RNAPII INI= 1/ k_ini (sec)
OFF time= 1/ k_on (min)
ON time= 1/ k_off (min)

G

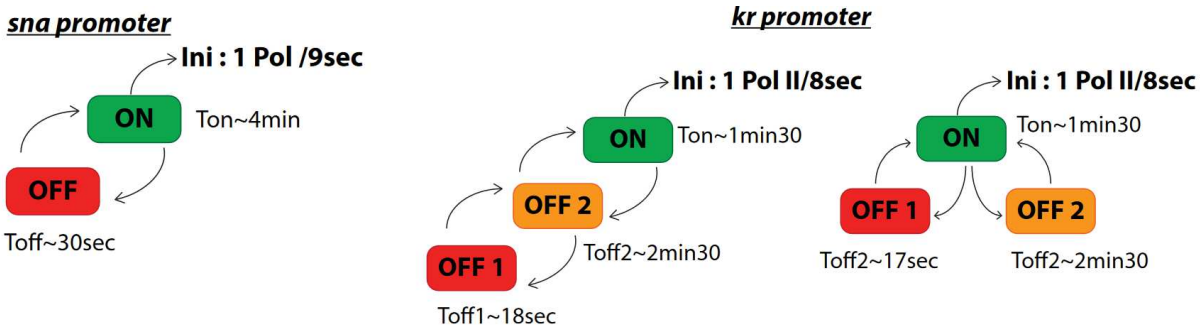


Figure 4. Inferring transcription kinetics. (A) Single nuclei traces calibrated in number of transcribing polymerases, and filtered to detect steady state (grey square). (B) Heat-map showing polymerase density at single locus resolution, during interphase 14 (~30 min), (high density in red, low or no density in bleu). (C) Graphic representation placing each single polymerase initiation event during the interphase 14 (~30 min), each polymerase is represented by a white vertical line. (D) Graphic representation of the waiting times between two initiation events. (E) Multi-exponential fit of the waiting times enables to infer promoter state kinetics (k_{on}, k_{off} and k_{ini}), black line corresponds to the fitting simulation, small red circles correspond to distribution times extracted from raw data. (F) Cartoon describing the switch between ON (green) and OFF (red) states of gene activation, the start of a new initiation event is represented in orange. The kinetic parameters (k_{on}, k_{off} and k_{ini}) enable to infer the average time of a burst duration, the off time and the rate of RNAP II initiation. (G) Schematic representation of estimated kinetic parameters for the *sna* promoter (left) with a two-state model, and for the *kr* promoter (right) which could be explained by two different three-state model.

Figure 5.

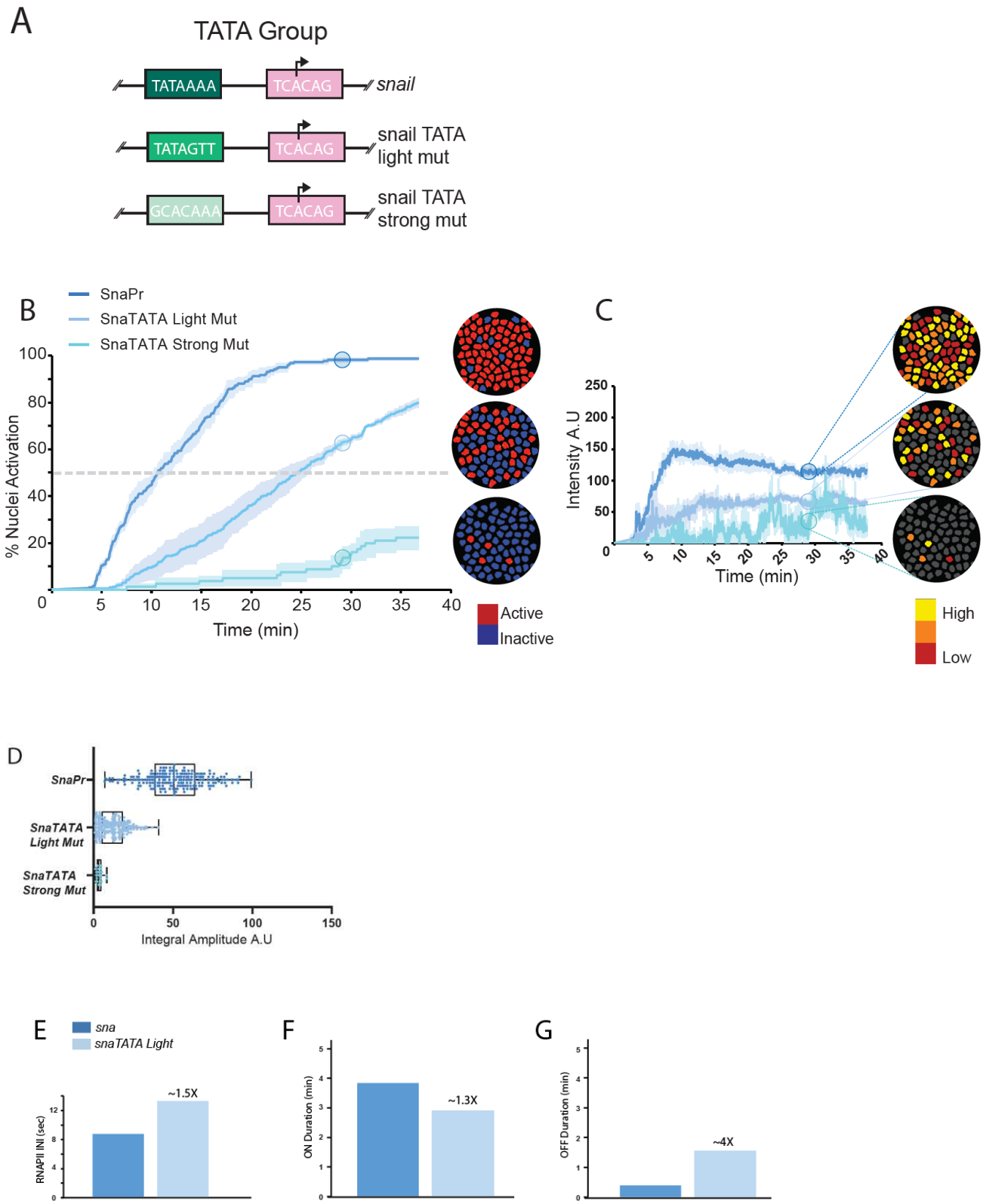
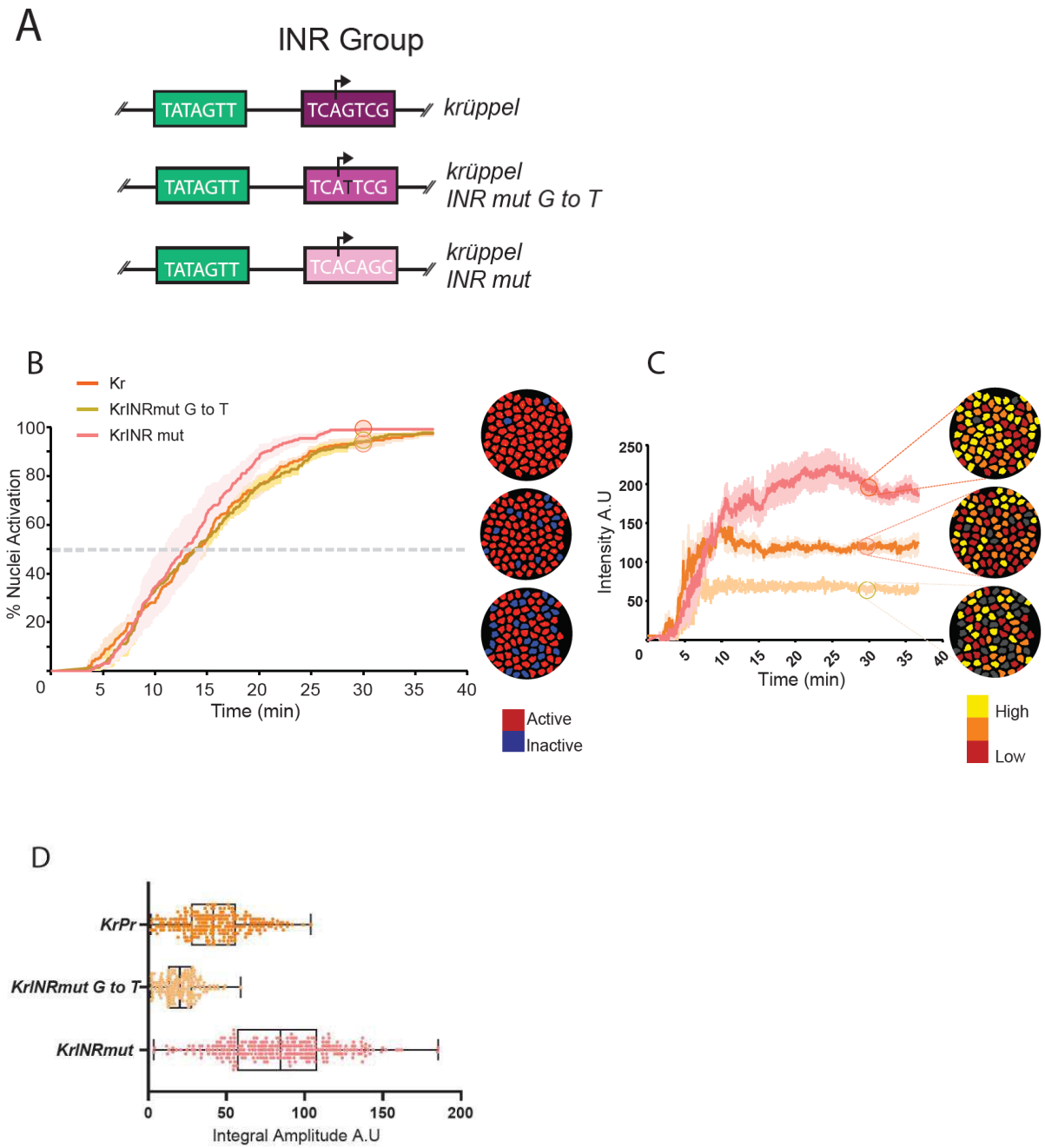

A
TATA Group
TATAAAA
TCACAG
snail
TATAGTT
TCACAG
snail TATA
light mut
GCACAAA
TCACAG
snail TATA
strong mut
B
SnaPr
SnaTATA Light Mut
SnaTATA Strong Mut
% Nuclei Activation
Time (min)
Active
Inactive
C
Intensity A.U
Time (min)
High
Low
D
SnaPr
SnaTATA
Light Mut
SnaTATA
Strong Mut
Integral Amplitude A.U
E
sna
snaTATA Light
RNAPII INI (sec)
~1.5X
F
ON Duration (min)
~1.3X
G
OFF Duration (min)
~4X

Figure 5. Effect of the TATA box element on transcription dynamics. (A) Cartoon representing mutations of the *sna* TATA-containing promoter, in the *TATA light* mutation, the last 3 nucleotides of the sequence were replaced by the pseudo TATA motif coming from the *kr* promoter (TATAAAA -> TATA**GTT**), for *TATA strong* mutation the first four nucleotides were replaced (TATAAAA -> **GCAC**AAA). (B) Synchrony profiles for the TATA group promoters, *sna* (3 movies, n =216 nuclei) (dark blue curve); *snaTATA Light mut* (3 movies, n= 202) (light blue curve); *snaTATA Strong mut* (4 movies, n= 29) (cyan colour curve). Dashed grey line represents 50% of activation. The time origin is the end of the nc13 mitosis. Error bars in lighter colours represent SEM. On the right, in circles: snapshot of the active nuclei (false-colour in red), at t=~30 min, inactive nuclei are false-colour in bleu (C) Average intensities in A.U across the interphase 14 (~38min after mitosis) for the TATA group. On the right, in circles: a heatmap snapshot of the intensities for each promoter at t=~30 min (high intensities are in yellow, low intensities in red, no signal in grey). (D) Distribution of the integral amplitude values in A.U for the TATA group, each dot represents the integral for one nucleus (some nuclei data might overlap). *sna* (3 movies, n =216 nuclei) (dark blue dots) *snaTATA Light mut* (3 movies, n= 202) (light blue dots); *snaTATA Strong mut* (4 movies, n= 29) (cyan colour dots). (E)(F)(G) Kinetic parameters estimated with the deconvolution method using a two-state model, for the *sna* and *snaTATA Light mut* promoter. RNAP II initiation rate time (1.5X) and OFF duration (4X) time are increased in the *snaTATA Light mut* promoter, ON duration time is decreased (1.3X) in the mutant.

The estimations for the *snaTATA Strong mut,* are not shown because the number of active nuclei is too low to represent the population. Stats: s*na*=216 nuclei, *snaTATA Light mut*=353.

Figure 6.


A
INR Group
TATAGTT
TCAGTCG
krüppel
TATAGTT
TCATTCG
krüppel
INR mut G to T
TATAGTT
TCACAGC
krüppel
INR mut
B
Kr
KrINRmut G to T
KrINR mut
% Nuclei Activation
Time (min)
Active
Inactive
C
Intensity A.U
Time (min)
High
Low
D
KrPr
KrINRmut G to T
KrINRmut
Integral Amplitude A.U

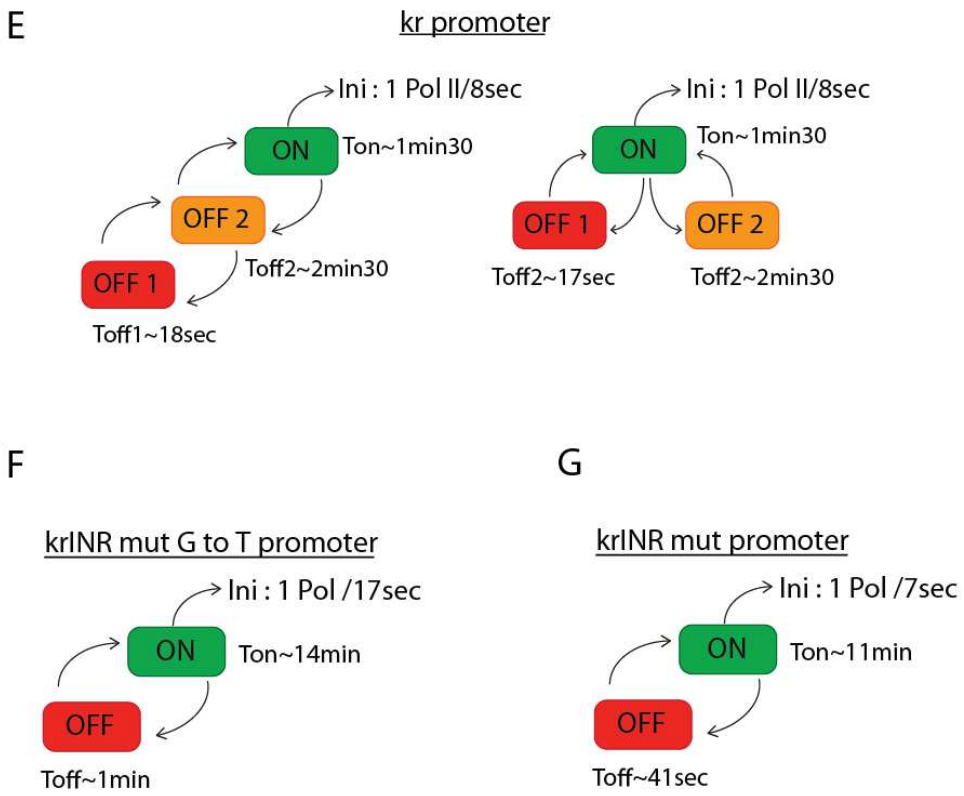


Figure 6. Effect of the INR element on transcription dynamics. (A) Cartoon representing mutations of the *kr* INR-containing promoter, in the *krINRmut G to T* mutation, the G in position +2 has been replaced by a T (TCA**G**TC -> TCA**T**TC). In the *krINRmut* the TSS of the *kr* promoter has been replaced by the TSS of the *sna* promoter (TCAGTC-> TCACAG). (B) Synchrony profiles for the INR group promoters, *kr* (3 movies, n =247 nuclei) (orange curve); *krINRmut G to T* (3 movies, n= 169) (light orange); *krINRmut* (5 movies, n= 343) (rose curve). Dashed grey line represents 50% of activation. The time origin is the end of the nc13 mitosis. Error bars in lighter colors represent SEM. On the side, in circles: snapshot of the active nuclei (false-colour in red), at t=~30 min, inactive nuclei are false-colour in bleu (C) Average intensities in A.U across the interphase 14 (~38min after mitosis) for the INR group. On the side, in circles, a heatmap snapshot of the intensities for each promoter at t=~30 min (high intensities are in yellow, low intensities in red, no signal in grey). (D) Distribution of the integral amplitude values in A.U for the INR group, each dot represents the integral for one nucleus (some nuclei data might overlap). *kr* (3 movies, n =247 nuclei) (orange curve); *krINRmut G to T* (3 movies, n= 169) (light orange); *krINRmut* (5 movies, n= 343) (rose curve).

(E)(F)(G) Estimated kinetic parameters for INR Group. Dynamics for the *kr* promoter are estimated with a three-model state. For *krINR mut G to T*, and *KrINR mut*, a two-state model is enough to capture their dynamics.

References

Adelman, K. and Lis, J. T. (2012) 'Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans', *Nature Reviews Genetics*, 13(10), pp. 720–731. doi: 10.1038/nrg3293.

Akhtar, W. and Veenstra, G. J. C. (2011) 'TBP-related factors: a paradigm of diversity in transcription initiation', *Cell & Bioscience*, 1(1), p. 23. doi: 10.1186/2045-3701-1-23.

Andersson, R. and Sandelin, A. (2019) 'Determinants of enhancer and promoter activities of regulatory elements', *Nature Reviews Genetics*. Nature Publishing Group, pp. 1–17. doi: 10.1038/s41576-019-0173-8.

Andersson, R., Sandelin, A. and Danko, C. G. (2015) 'A unified architecture of transcriptional regulatory elements', *Trends in Genetics*, 31(8), pp. 426–433. doi: 10.1016/j.tig.2015.05.007.

van Arensbergen, J., van Steensel, B. and Bussemaker, H. J. (2014) 'In search of the determinants of enhancer–promoter interaction specificity', *Trends in Cell Biology*, 24(11), pp. 695–702. doi: 10.1016/j.tcb.2014.07.004.

Arnold, C. D. *et al.* (2013) 'Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq', *Science*, 339(6123), pp. 1074–1077. doi: 10.1126/science.1232542.

Arnold, C. D. *et al.* (2017) 'Genome-wide assessment of sequence-intrinsic enhancer responsiveness at single-base-pair resolution', *Nature Biotechnology*, 35(2), pp. 136–144. doi: 10.1038/nbt.3739.

Atchison, M. L. and Perry, R. P. (1988) 'Complementation between two cell lines lacking kappa enhancer activity: implications for the developmental control of immunoglobulin transcription.', *The EMBO journal*, 7(13), pp. 4213–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2854057 (Accessed: 15 October 2019).

Banerji, J., Olson, L. and Schaffner, W. (1983) 'A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes', *Cell*, 33(3), pp. 729–740. doi: 10.1016/0092-8674(83)90015-6.

Banerji, J., Rusconi, S. and Schaffner, W. (1981) 'Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences.', *Cell*, 27(2 Pt 1), pp. 299–308. doi: 10.1016/0092-8674(81)90413-x.

Bártfai, R. *et al.* (2004) 'TBP2, a Vertebrate-Specific Member of the TBP Family, Is Required in Embryonic Development of Zebrafish', *Current Biology*. Cell Press, 14(7), pp. 593–598. doi: 10.1016/J.CUB.2004.03.034.

Bashirullah, A. *et al.* (1999) 'Joint action of two RNA degradation pathways controls the timing of maternal transcript elimination at the midblastula transition in Drosophila melanogaster.', *The EMBO journal*. European Molecular Biology Organization, 18(9), pp. 2610–20. doi: 10.1093/emboj/18.9.2610.

Bashirullah, A., Cooperstock, R. L. and Lipshitz, H. D. (2001) 'Spatial and temporal control of RNA stability', *Proceedings of the National Academy of Sciences*, 98(13), pp. 7025–7028. doi: 10.1073/pnas.111145698.

Benoist, C. and Chambon, P. (1981) 'In vivo sequence requirements of the SV40 early promoter region', *Nature*. Nature Publishing Group, 290(5804), pp. 304–310. doi: 10.1038/290304a0.

Bernstein, B. E. *et al.* (2006) 'A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells', *Cell*, 125(2), pp. 315–326. doi: 10.1016/j.cell.2006.02.041.

Bhuiyan, T. and Timmers, H. T. M. (2019) 'Promoter Recognition: Putting TFIID on the Spot', *Trends in Cell Biology*. doi: 10.1016/j.tcb.2019.06.004.

Blake, W. J. *et al.* (2006) 'Phenotypic Consequences of Promoter-Mediated Transcriptional Noise', *Molecular Cell*. doi: 10.1016/j.molcel.2006.11.003.

Boettiger, A. N. and Levine, M. (2009) 'Synchronous and stochastic patterns of gene activation in the Drosophila embryo.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 325(5939), pp. 471–3. doi: 10.1126/science.1173976.

Bothma, J. P. *et al.* (2014) 'Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living Drosophila embryos', *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1410022111.

Buckley, M. S. *et al.* (2014) 'Kinetics of promoter Pol II on Hsp70 reveal stable pausing and key insights into its regulation', *Genes and Development*. doi: 10.1101/gad.231886.113.

Buratowski, S. *et al.* (1989) 'Five intermediate complexes in transcription initiation by RNA polymerase II', *Cell*. Cell Press, 56(4), pp. 549–561. doi: 10.1016/0092-8674(89)90578-3.

Burley, S. K. and Roeder, R. G. (1996) 'Biochemistry and Structural Biology of Transcription Factor IID (TFIID)', *Annual Review of Biochemistry*, 65(1), pp. 769–799. doi: 10.1146/annurev.bi.65.070196.004005.

Butler, J. E. and Kadonaga, J. T. (2001) 'Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs.', *Genes & development*. Cold Spring Harbor Laboratory Press, 15(19), pp. 2515–9. doi: 10.1101/gad.924301.

Cech, T. R. and Steitz, J. A. (2014) 'The noncoding RNA revolution-trashing old rules to forge new ones.', *Cell*, 157(1), pp. 77–94. doi: 10.1016/j.cell.2014.03.008.

Chafin, D. R., Claussen, T. J. and Price, D. H. (1991) 'Identification and purification of a yeast protein that affects elongation by RNA polymerase II.', *The Journal of biological chemistry*, 266(14), pp. 9256–62. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1851172 (Accessed: 15 October 2019).

Chalkley, G. E. (1999) 'DNA binding site selection by RNA polymerase II TAFs: a TAFII250-TAFII150 complex recognizes the Initiator', *The EMBO Journal*, 18(17), pp. 4835–4845. doi: 10.1093/emboj/18.17.4835.

Chapman, R. D., Conrad, M. and Eick, D. (2005) 'Role of the mammalian RNA polymerase II C-terminal domain (CTD) nonconsensus repeats in CTD stability and cell proliferation.', *Molecular and cellular biology*, 25(17), pp. 7665–74. doi: 10.1128/MCB.25.17.7665-7674.2005.

Chen, H.-T. and Hahn, S. (2004) 'Mapping the Location of TFIIB within the RNA Polymerase II Transcription Preinitiation Complex', *Cell*, 119(2), pp. 169–180. doi: 10.1016/j.cell.2004.09.028.

Chen, K. *et al.* (2013) 'A global change in RNA polymerase II pausing during the Drosophila midblastula transition', 2, p. 861. doi: 10.7554/eLife.00861.001.

Cianfrocco, M. A. *et al.* (2013) 'Human TFIID Binds to Core Promoter DNA in a Reorganized Structural State', *Cell*, 152(1–2), pp. 120–131. doi: 10.1016/j.cell.2012.12.005.

Coin, F., Oksenych, V. and Egly, J.-M. (2007) 'Distinct Roles for the XPB/p52 and XPD/p44 Subcomplexes of TFIIH in Damaged DNA Opening during Nucleotide Excision Repair', *Molecular Cell*, 26(2), pp. 245–256. doi: 10.1016/j.molcel.2007.03.009.

Coleman, R. A. and Pugh, B. F. (1995) 'Evidence for Functional Binding and Stable Sliding of the TATA Binding Protein on Nonspecific DNA', *Journal of Biological Chemistry*, 270(23), pp. 13850–13859. doi: 10.1074/jbc.270.23.13850.

Corden, J. *et al.* (1980) 'Promoter sequences of eukaryotic protein-coding genes', *Science*, 209(4463), pp. 1406–1414. doi: 10.1126/science.6251548.

Core, L. J. *et al.* (2012) 'Defining the Status of RNA Polymerase at Promoters', *Cell Reports*. The Authors, 2(4), pp. 1025–1035. doi: 10.1016/j.celrep.2012.08.034.

Core, L. J. *et al.* (2014) 'Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers', *Nature Genetics*, 46(12), pp. 1311–1320. doi: 10.1038/ng.3142.

Core, L. J., Waterfall, J. J. and Lis, J. T. (2008) 'Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters.', *Science (New York, N.Y.)*, 322(5909), pp. 1845–8. doi: 10.1126/science.1162228.

Cormack, B. P. and Struhl, K. (1992) 'The TATA-binding protein is required for transcription by all three nuclear RNA polymerases in yeast cells.', *Cell*. Elsevier, 69(4), pp. 685–96. doi: 10.1016/0092-8674(92)90232-2.

Corrigan, A. M. *et al.* (2016) 'A continuum model of transcriptional bursting.', *eLife*. eLife Sciences Publications, Ltd, 5. doi: 10.7554/eLife.13051.

Crowley, T. E. *et al.* (1993) 'A new factor related to TATA-binding protein has highly restricted expression patterns in Drosophila.', *Nature*, 361(6412), pp. 557–61. doi: 10.1038/361557a0.

Danino, Y. M. *et al.* (2015) 'The core promoter: at the heart of gene expression', *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*. doi: 10.1016/j.bbagrm.2015.04.003.

Dantonel, J.-C. *et al.* (2000) 'TBP-like Factor Is Required for Embryonic RNA Polymerase II Transcription in C. elegans', *Molecular Cell*. Cell Press, 6(3), pp. 715–722. doi: 10.1016/S1097-2765(00)00069-1.

Dao, L. T. M. *et al.* (2017) 'Genome-wide characterization of mammalian promoters with distal enhancer functions', *Nature Genetics*, 49(7), pp. 1073–1081. doi: 10.1038/ng.3884.

Deng, W. and Roberts, S. G. E. (2005) 'A core promoter element downstream of the TATA box that is recognized by TFIIB', *Genes & Development*, 19(20), pp. 2418–2423. doi: 10.1101/gad.342405.

Driever, W. and Nüsslein-Volhard, C. (1988) 'The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner', *Cell*, 54(1), pp. 95–104. doi: 10.1016/0092-8674(88)90183-3.

Dufourt, J. *et al.* (2018) 'Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs', *bioRxiv*. Springer US, p. 282426. doi: 10.1101/282426.

Duttke, S. H. C. *et al.* (2014) 'TRF2 and the evolution of the bilateria', *Genes & Development*, 28(19), pp. 2071–2076. doi: 10.1101/gad.250563.114.

Egly, J.-M. (2011) '[DNA repair mechanisms: from basic biology to genetic pathology].', *Bulletin de l'Academie nationale de medecine*, 195(7), pp. 1689–90. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22812171 (Accessed: 15 October 2019).

Eick, D. and Geyer, M. (2013) 'The RNA Polymerase II Carboxy-Terminal Domain (CTD) Code', *Chemical Reviews*, 113(11), pp. 8456–8490. doi: 10.1021/cr400071f.

Elowitz, M. B. *et al.* (2002) 'Stochastic Gene Expression in a Single Cell', *Science*, 297(5584), pp. 1183–1186. doi: 10.1126/science.1070919.

Farrell, J. A. and O'Farrell, P. H. (2014) 'From Egg to Gastrula: How the Cell Cycle Is Remodeled During the *Drosophila* Mid-Blastula Transition', *Annual Review of Genetics*. Annual Reviews , 48(1), pp. 269–294. doi: 10.1146/annurev-genet-111212-133531.

Femino, A. M. *et al.* (2003) 'Visualization of single molecules of mRNA in situ.', *Methods in*

enzymology, 361, pp. 245–304. doi: 10.1016/s0076-6879(03)61015-3.

Fernandez, C. and Lagha, M. (2019) 'Lightening Up Gene Activation in Living Drosophila Embryos', *Methods in molecular biology*, 2038. Available at: https://doi.org/10.1007/978-1-4939-9674-2_5.

Ferraro, T. *et al.* (2016) 'New methods to image transcription in living fly embryos: The insights so far, and the prospects', *Wiley Interdisciplinary Reviews: Developmental Biology*. doi: 10.1002/wdev.221.

Fishburn, J. and Hahn, S. (2012) 'Architecture of the Yeast RNA Polymerase II Open Complex and Regulation of Activity by TFIIF', *Molecular and Cellular Biology*, 32(1), pp. 12–25. doi: 10.1128/MCB.06242-11.

FitzGerald, P. C. *et al.* (2006) 'Comparative genomics of Drosophila and human core promoters.', *Genome Biology*, 7(7), p. R53. doi: 10.1186/gb-2006-7-7-r53.

Foe, V. E. and Alberts, B. M. (1983) 'Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in Drosophila embryogenesis.', *Journal of cell science*, 61, pp. 31–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6411748 (Accessed: 14 October 2019).

Fromm, M. and Berg, P. (1982) 'Deletion mapping of DNA regions required for SV40 early region promoter function in vivo.', *Journal of molecular and applied genetics*, 1(5), pp. 457–81. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6296253 (Accessed: 15 October 2019).

Fukaya, T., Lim, B. and Levine, M. (2016) 'Enhancer Control of Transcriptional Bursting.', *Cell*. Elsevier, 166(2), pp. 358–368. doi: 10.1016/j.cell.2016.05.025.

Fukaya, T., Lim, B. and Levine, M. (2017) 'Rapid Rates of Pol II Elongation in the Drosophila Embryo.', *Current biology : CB*. NIH Public Access, 27(9), pp. 1387–1391. doi: 10.1016/j.cub.2017.03.069.

Gaertner, B. and Zeitlinger, J. (2014) 'RNA polymerase II pausing during development', *Development*. doi: 10.1242/dev.088492.

Galbraith, M. D., Donner, A. J. and Espinosa, J. M. (2010) 'CDK8', *Transcription*, 1(1), pp. 4–12. doi: 10.4161/trns.1.1.12373.

Garcia, H. G. *et al.* (2013) 'Quantitative Imaging of Transcription in Living Drosophila Embryos Links Polymerase Activity to Patterning', *Current Biology*. doi: 10.1016/j.cub.2013.08.054.

Garcia, M. *et al.* (2013) 'Size-dependent regulation of dorsal–ventral patterning in the early Drosophila embryo', *Developmental Biology*, 381(1), pp. 286–299. doi: 10.1016/j.ydbio.2013.06.020.

Gavis, E. R. and Lehmann, R. (1992) 'Localization of nanos RNA controls embryonic polarity', *Cell*, 71(2), pp. 301–313. doi: 10.1016/0092-8674(92)90358-J.

Gazdag, E. *et al.* (2009) 'TBP2 is essential for germ cell development by regulating transcription and chromatin condensation in the oocyte.', *Genes & development*. Cold Spring Harbor Laboratory Press, 23(18), pp. 2210–23. doi: 10.1101/gad.535209.

Gehrig, J. *et al.* (2009) 'Automated high-throughput mapping of promoter-enhancer interactions in zebrafish embryos', *Nature Methods*, 6(12), pp. 911–916. doi: 10.1038/nmeth.1396.

Gershenzon, N. I. and Ioshikhes, I. P. (2005) 'Promoter classifier: software package for promoter database analysis.', *Applied bioinformatics*, 4(3), pp. 205–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16231962 (Accessed: 15 October 2019).

Gilchrist, D. A. *et al.* (2010) 'Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation.', *Cell*, 143(4), pp. 540–51. doi: 10.1016/j.cell.2010.10.004.

Gillies, S. D. *et al.* (1983) 'A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene', *Cell*, 33(3), pp. 717–728. doi: 10.1016/0092-8674(83)90014-4.

Gilmour, D. S. and Lis, J. T. (1986) 'RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells.', *Molecular and cellular biology*. American Society for Microbiology (ASM), 6(11), pp. 3984–9. doi:

10.1128/mcb.6.11.3984.

Goodrich, J. A. and Tjian, R. (2010) ‘Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation’, *Nature Reviews Genetics*, 11(8), pp. 549–558. doi: 10.1038/nrg2847.

Grosschedl, R. and Birnstiel, M. L. (1980) ‘Spacer DNA sequences upstream of the T-A-T-A-A-A-T-A sequence are essential for promotion of H2A histone gene transcription in vivo.’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 77(12), pp. 7102–6. doi: 10.1073/pnas.77.12.7102.

Guarente, L. (1988) ‘UASs and enhancers: common mechanism of transcriptional activation in yeast and mammals.’, *Cell*, 52(3), pp. 303–5. doi: 10.1016/s0092-8674(88)80020-5.

Guzmán, E. and Lis, J. T. (1999) ‘Transcription Factor TFIIH Is Required for Promoter Melting In Vivo’, *Molecular and Cellular Biology*, 19(8), pp. 5652–5658. doi: 10.1128/MCB.19.8.5652.

Ha, I. *et al.* (1993) ‘Multiple functional domains of human transcription factor IIB: distinct interactions with two general transcription factors and RNA polymerase II.’, *Genes & Development*, 7(6), pp. 1021–1032. doi: 10.1101/gad.7.6.1021.

Haberle, V. *et al.* (2014) ‘Two independent transcription initiation codes overlap on vertebrate core promoters’, *Nature*, 507(7492), pp. 381–385. doi: 10.1038/nature12974.

Haberle, V. and Stark, A. (2018) ‘Eukaryotic core promoters and the functional basis of transcription initiation’, *Nature Reviews Molecular Cell Biology*. Europe PMC Funders, pp. 621–637. doi: 10.1038/s41580-018-0028-8.

Heine, G. F., Horwitz, A. A. and Parvin, J. D. (2008) ‘Multiple Mechanisms Contribute to Inhibit Transcription in Response to DNA Damage’, *Journal of Biological Chemistry*, 283(15), pp. 9555–9561. doi: 10.1074/jbc.M707700200.

Hendrix, D. A. *et al.* (2008) ‘Promoter elements associated with RNA Pol II stalling in the Drosophila embryo’, *Proceedings of the National Academy of Sciences*, 105(22), pp. 7762–7767. doi: 10.1073/pnas.0802406105.

Henriques, T. *et al.* (2013) ‘Stable Pausing by RNA Polymerase II Provides an Opportunity to

Target and Integrate Regulatory Signals', *Molecular Cell*, 52(4), pp. 517–528. doi: 10.1016/j.molcel.2013.10.001.

Hintermair, C. *et al.* (2016) 'Specific threonine-4 phosphorylation and function of RNA polymerase II CTD during M phase progression.', *Scientific reports*, 6(1), p. 27401. doi: 10.1038/srep27401.

Høiby, T. *et al.* (2007) 'A facelift for the general transcription factor TFIIA', *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*. Elsevier, 1769(7–8), pp. 429–436. doi: 10.1016/J.BBAEXP.2007.04.008.

Holmes, M. C. and Tjian, R. (2000) 'Promoter-selective properties of the TBP-related factor TRF1.', *Science (New York, N.Y.)*, 288(5467), pp. 867–70. doi: 10.1126/science.288.5467.867.

Holstege, F. C., van der Vliet, P. C. and Timmers, H. T. (1996) 'Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH.', *The EMBO journal*, 15(7), pp. 1666–77. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8612591 (Accessed: 15 October 2019).

Horikoshi, M. *et al.* (1990) 'Analysis of structure-function relationships of yeast TATA box binding factor TFIID.', *Cell*. Elsevier, 61(7), pp. 1171–8. doi: 10.1016/0092-8674(90)90681-4.

Hoskins, R. A. *et al.* (2011) 'Genome-wide analysis of promoter architecture in Drosophila melanogaster', *Genome Research*, 21(2), pp. 182–192. doi: 10.1101/gr.112466.110.

Ibrahim, M. M. *et al.* (2018) 'Determinants of promoter and enhancer transcription directionality in metazoans.', *Nature communications*. Nature Publishing Group, 9(1), p. 4472. doi: 10.1038/s41467-018-06962-z.

Jeronimo, C. and Robert, F. (2014) 'Kin28 regulates the transient association of Mediator with core promoters', *Nature Structural & Molecular Biology*, 21(5), pp. 449–455. doi: 10.1038/nsmb.2810.

Jonkers, I. and Lis, J. T. (2015) 'Getting up to speed with transcription elongation by RNA polymerase II', *Nature Reviews Molecular Cell Biology*. NIH Public Access, pp. 167–177. doi: 10.1038/nrm3953.

Juven-Gershon, T., Hsu, J.-Y. and Kadonaga, J. T. (2008) 'Caudal, a key developmental regulator, is a DPE-specific transcriptional factor', *Genes & Development*, 22(20), pp. 2823–2830. doi: 10.1101/gad.1698108.

Kaufmann, J. and Smale, S. T. (1994) 'Direct recognition of initiator elements by a component of the transcription factor IID complex.', *Genes & development*, 8(7), pp. 821–9. doi: 10.1101/gad.8.7.821.

Kedinger, C. *et al.* (1970) 'α-Amanitin: A specific inhibitor of one of two DNA-dependent RNA polymerase activities from calf thymus', *Biochemical and Biophysical Research Communications*, 38(1), pp. 165–171. doi: 10.1016/0006-291X(70)91099-5.

Khoury, G. and Gruss, P. (1983) 'Enhancer elements', *Cell*, 33(2), pp. 313–314. doi: 10.1016/0092-8674(83)90410-5.

Kim, J. L., Nikolov, D. B. and Burley, S. K. (1993) 'Co-crystal structure of TBP recognizing the minor groove of a TATA element.', *Nature*, 365(6446), pp. 520–7. doi: 10.1038/365520a0.

Kim, T.-K. and Shiekhattar, R. (2015) 'Architectural and Functional Commonalities between Enhancers and Promoters.', *Cell*. Elsevier, 162(5), pp. 948–59. doi: 10.1016/j.cell.2015.08.008.

Kim, W. Y. and Dahmus, M. E. (1989) 'The major late promoter of adenovirus-2 is accurately transcribed by RNA polymerases IIO, IIA, and IIB.', *The Journal of biological chemistry*, 264(6), pp. 3169–76. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2914948 (Accessed: 15 October 2019).

Kobayashi, N., Boyer, T. G. and Berk, A. J. (1995) 'A class of activation domains interacts directly with TFIIA and stimulates TFIIA-TFIID-promoter complex assembly.', *Molecular and Cellular Biology*, 15(11), pp. 6465–6473. doi: 10.1128/MCB.15.11.6465.

Koch, F. *et al.* (2011) 'Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters', *Nature Structural & Molecular Biology*, 18(8), pp. 956–963. doi: 10.1038/nsmb.2085.

Kokubo, T. *et al.* (1994) 'Interaction between the N-terminal domain of the 230-kDa subunit and the TATA box-binding subunit of TFIID negatively regulates TATA-box binding.',

Proceedings of the National Academy of Sciences of the United States of America, 91(9), pp. 3520–4. doi: 10.1073/pnas.91.9.3520.

Kolesnikova, O. *et al.* (2018) 'Molecular structure of promoter-bound yeast TFIID', *Nature Communications*, 9(1), p. 4666. doi: 10.1038/s41467-018-07096-y.

Kopytova, D. V. *et al.* (2006) 'Two Isoforms of Drosophila TRF2 Are Involved in Embryonic Development, Premeiotic Chromatin Condensation, and Proper Differentiation of Germ Cells of Both Sexes', *Molecular and Cellular Biology*, 26(20), pp. 7492–7505. doi: 10.1128/MCB.00349-06.

Kostrewa, D. *et al.* (2009) 'RNA polymerase II–TFIIB structure and mechanism of transcription initiation', *Nature*, 462(7271), pp. 323–330. doi: 10.1038/nature08548.

Krebs, A. R. *et al.* (2017) 'Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters.', *Molecular cell*, 67(3), pp. 411-422.e4. doi: 10.1016/j.molcel.2017.06.027.

Krumm, A. *et al.* (1992) 'The block to transcriptional elongation within the human c-myc gene is determined in the promoter-proximal region.', *Genes & Development*, 6(11), pp. 2201–2213. doi: 10.1101/gad.6.11.2201.

Kwak, H. *et al.* (2013) 'Precise maps of RNA polymerase reveal how promoters direct initiation and pausing.', *Science (New York, N.Y.)*, 339(6122), pp. 950–3. doi: 10.1126/science.1229386.

Lagha, M. *et al.* (2013) 'XPaused Pol II coordinates tissue morphogenesis in the drosophila embryo', *Cell*. doi: 10.1016/j.cell.2013.04.045.

Langelier, M. F. *et al.* (2001) 'Structural and functional interactions of transcription factor (TF) IIA with TFIIE and TFIIF in transcription initiation by RNA polymerase II.', *The Journal of biological chemistry*. PMC Canada manuscript submission, 276(42), pp. 38652–7. doi: 10.1074/jbc.M106422200.

Larsson, A. J. M. *et al.* (2019) 'Genomic encoding of transcriptional burst kinetics', *Nature*. Nature Publishing Group, 565(7738), pp. 251–254. doi: 10.1038/s41586-018-0836-1.

Lenstra, T. L. *et al.* (2016) 'Transcription Dynamics in Living Cells', *Annual Review of Biophysics*. doi: 10.1146/annurev-biophys-062215-010838.

Levsky, J. M. *et al.* (2002) 'Single-Cell Gene Expression Profiling', *Science*, 297(5582), pp. 836–840. doi: 10.1126/science.1072241.

Li, X. and Noll, M. (1994) 'Compatibility between enhancers and promoters determines the transcriptional specificity of gooseberry and gooseberry neuro in the Drosophila embryo.', *The EMBO Journal*. John Wiley & Sons, Ltd, 13(2), pp. 400–406. doi: 10.1002/j.1460-2075.1994.tb06274.x.

Lim, C. Y. *et al.* (2004) 'The MTE, a new core promoter element for transcription by RNA polymerase II', *Genes & Development*, 18(13), pp. 1606–1617. doi: 10.1101/gad.1193404.

Lindell, T. J. *et al.* (1970) 'Specific Inhibition of Nuclear RNA Polymerase II by agr-Amanitin', *Science*, 170(3956), pp. 447–449. doi: 10.1126/science.170.3956.447.

Little, S. C., Tikhonov, M. and Gregor, T. (2013) 'Precise developmental gene expression arises from globally stochastic transcriptional activity', *Cell*. doi: 10.1016/j.cell.2013.07.025.

Louder, R. K. *et al.* (2016) 'Structure of promoter-bound TFIID and model of human pre-initiation complex assembly', *Nature*, 531(7596), pp. 604–609. doi: 10.1038/nature17394.

Lu, H. *et al.* (1991) 'The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex.', *Proceedings of the National Academy of Sciences of the United States of America*, 88(22), pp. 10004–8. doi: 10.1073/pnas.88.22.10004.

Lucas, T. *et al.* (2013) 'Live Imaging of Bicoid-Dependent Transcription in Drosophila Embryos', *Current Biology*. doi: 10.1016/j.cub.2013.08.053.

MARKOW, T. A., BEALL, S. and MATZKIN, L. M. (2009) 'Egg size, embryonic development time and ovoviviparity in Drosophila species', *Journal of Evolutionary Biology*, 22(2), pp. 430–434. doi: 10.1111/j.1420-9101.2008.01649.x.

Martianov, I., Viville, S. and Davidson, I. (2002) 'RNA Polymerase II Transcription in Murine Cells Lacking the TATA Binding Protein', *Science*, 298(5595), pp. 1036–1039. doi: 10.1126/science.1076327.

Mathis, D. J. and Chambon, P. (1981) 'The SV40 early region TATA box is required for accurate in vitro initiation of transcription', *Nature*. Nature Publishing Group, 290(5804), pp. 310–315. doi: 10.1038/290310a0.

Matsui, T. *et al.* (1980) 'Multiple factors required for accurate initiation of transcription by purified RNA polymerase II.', *The Journal of biological chemistry*, 255(24), pp. 11992–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7440580 (Accessed: 15 October 2019).

MATSUKAGE, A. *et al.* (2008) 'The DRE/DREF transcriptional regulatory system: a master key for cell proliferation', *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1779(2), pp. 81–89. doi: 10.1016/j.bbagrm.2007.11.011.

Maxon, M. E., Goodrich, J. A. and Tjian, R. (1994) 'Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance.', *Genes & Development*, 8(5), pp. 515–524. doi: 10.1101/gad.8.5.515.

Meinel, D. M. *et al.* (2013) 'Recruitment of TREX to the transcription machinery by its direct binding to the phospho-CTD of RNA polymerase II.', *PLoS genetics*. Edited by G. P. Copenhaver, 9(11), p. e1003914. doi: 10.1371/journal.pgen.1003914.

Merli, C. *et al.* (1996) 'Promoter specificity mediates the independent regulation of neighboring genes.', *Genes & Development*, 10(10), pp. 1260–1270. doi: 10.1101/gad.10.10.1260.

Moreau, P. *et al.* (1981) 'The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants', *Nucleic Acids Research*, 9(22), pp. 6047–6068. doi: 10.1093/nar/9.22.6047.

Müller, F. *et al.* (2001) 'TBP is not universally required for zygotic RNA polymerase II transcription in zebrafish', *Current Biology*, 11(4), pp. 282–287. doi: 10.1016/S0960-9822(01)00076-8.

Muse, G. W. *et al.* (2007) 'RNA polymerase is poised for activation across the genome', *Nature Genetics*, 39(12), pp. 1507–1511. doi: 10.1038/ng.2007.21.

Myers, L. C. *et al.* (1998) 'The Med proteins of yeast and their function through the RNA polymerase II carboxy-terminal domain', *Genes & Development*, 12(1), pp. 45–54. doi:

10.1101/gad.12.1.45.

Nechaev, S. *et al.* (2010) 'Global Analysis of Short RNAs Reveals Widespread Promoter-Proximal Stalling and Arrest of Pol II in Drosophila', *Science*, 327(5963), pp. 335–338. doi: 10.1126/science.1181421.

Neuberger, M. S. and Calabi, F. (1983) 'Reciprocal chromosome translocation between c-myc and immunoglobulin γ2b genes', *Nature*, 305(5931), pp. 240–243. doi: 10.1038/305240a0.

Nicolas, D., Phillips, N. E. and Naef, F. (2017) 'What shapes eukaryotic transcriptional bursting?', *Mol. BioSyst.* doi: 10.1039/C7MB00154A.

Ohler, U. *et al.* (2002) 'Computational analysis of core promoters in the Drosophila genome.', *Genome biology*. BioMed Central, 3(12), p. RESEARCH0087. doi: 10.1186/gb-2002-3-12-research0087.

Ohtsuki, S. and Levine, M. (1998) 'GAGA mediates the enhancer blocking activity of the eve promoter in the Drosophila embryo.', *Genes & development*. Cold Spring Harbor Laboratory Press, 12(21), pp. 3325–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9808619 (Accessed: 22 July 2016).

van Opijnen, T. *et al.* (2004) 'The Human Immunodeficiency Virus Type 1 Promoter Contains a CATA Box Instead of a TATA Box for Optimal Transcription and Replication', *Journal of Virology*, 78(13), pp. 6883–6890. doi: 10.1128/JVI.78.13.6883-6890.2004.

Orphanides, G., Lagrange, T. and Reinberg, D. (1996) 'The general transcription factors of RNA polymerase II.', *Genes & Development*, 10(21), pp. 2657–2683. doi: 10.1101/gad.10.21.2657.

Pal, M., Ponticelli, A. S. and Luse, D. S. (2005) 'The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II.', *Molecular cell*, 19(1), pp. 101–10. doi: 10.1016/j.molcel.2005.05.024.

Pan, G. and 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.', *The Journal of biological chemistry*, 269(48), pp. 30101–4. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/7982911 (Accessed: 15 October 2019).

Paré, A. *et al.* (2009) 'Visualization of Individual Scr mRNAs during Drosophila Embryogenesis Yields Evidence for Transcriptional Bursting', *Current Biology*. doi: 10.1016/j.cub.2009.10.028.

Parry, T. J. *et al.* (2010) 'The TCT motif, a key component of an RNA polymerase II transcription system for the translational machinery.', *Genes & development*, 24(18), pp. 2013–8. doi: 10.1101/gad.1951110.

Patel, A. B. *et al.* (2018) 'Structure of human TFIID and mechanism of TBP loading onto promoter DNA', *Science*, 362(6421), p. eaau8872. doi: 10.1126/science.aau8872.

Patwardhan, R. P. *et al.* (2009) 'High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis.', *Nature biotechnology*. NIH Public Access, 27(12), pp. 1173–5. doi: 10.1038/nbt.1589.

Pichon, X. *et al.* (2018) 'A Growing Toolbox to Image Gene Expression in Single Cells: Sensitive Approaches for Demanding Challenges', *Molecular Cell*. doi: 10.1016/j.molcel.2018.07.022.

Podos, S. D. and Ferguson, E. L. (1999) 'Morphogen gradients: new insights from DPP.', *Trends in genetics : TIG*, 15(10), pp. 396–402. doi: 10.1016/s0168-9525(99)01854-5.

Proudfoot, N. J. (2016) 'Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut.', *Science (New York, N.Y.)*. Europe PMC Funders, 352(6291), p. aad9926. doi: 10.1126/science.aad9926.

Rach, E. A. *et al.* (2009) 'Motif composition, conservation and condition-specificity of single and alternative transcription start sites in the Drosophila genome.', *Genome biology*, 10(7), p. R73. doi: 10.1186/gb-2009-10-7-r73.

Raj, A. *et al.* (2006) 'Stochastic mRNA Synthesis in Mammalian Cells', *PLoS Biology*. Edited by U. Schibler, 4(10), p. e309. doi: 10.1371/journal.pbio.0040309.

Raj, A. and van Oudenaarden, A. (2008) 'Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences', *Cell*, 135(2), pp. 216–226. doi:

10.1016/j.cell.2008.09.050.

Raser, J. M. (2004) ‘Control of Stochasticity in Eukaryotic Gene Expression’, *Science*, 304(5678), pp. 1811–1814. doi: 10.1126/science.1098641.

Raser, J. M. (2005) ‘Noise in Gene Expression: Origins, Consequences, and Control’, *Science*, 309(5743), pp. 2010–2013. doi: 10.1126/science.1105891.

Reeve, J. N. (2003) ‘Archaeal chromatin and transcription.’, *Molecular microbiology*, 48(3), pp. 587–98. doi: 10.1046/j.1365-2958.2003.03439.x.

Ren, D., Lei, L. and Burton, Z. F. (1999) ‘A region within the RAP74 subunit of human transcription factor IIF is critical for initiation but dispensable for complex assembly.’, *Molecular and cellular biology*, 19(11), pp. 7377–87. doi: 10.1128/mcb.19.11.7377.

Rennie, S. *et al.* (2018) ‘Transcription start site analysis reveals widespread divergent transcription in D. melanogaster and core promoter-encoded enhancer activities’, *Nucleic Acids Research*, 46(11), pp. 5455–5469. doi: 10.1093/nar/gky244.

De Renzis, S. *et al.* (2007) ‘Unmasking activation of the zygotic genome using chromosomal deletions in the Drosophila embryo.’, *PLoS biology*. Edited by T. Kornberg, 5(5), p. e117. doi: 10.1371/journal.pbio.0050117.

Roeder, R. G. and Rutter, W. J. (1969) ‘Multiple Forms of DNA-dependent RNA Polymerase in Eukaryotic Organisms’, *Nature*. Nature Publishing Group, 224(5216), pp. 234–237. doi: 10.1038/224234a0.

Rosonina, E., Kaneko, S. and Manley, J. L. (2006) ‘Terminating the transcript: breaking up is hard to do.’, *Genes & development*. Cold Spring Harbor Laboratory Press, 20(9), pp. 1050–6. doi: 10.1101/gad.1431606.

Roth, S., Stein, D. and Nüsslein-Volhard, C. (1989) ‘A gradient of nuclear localization of the dorsal protein determines dorsoventral pattern in the Drosophila embryo’, *Cell*, 59(6), pp. 1189–1202. doi: 10.1016/0092-8674(89)90774-5.

Rushlow, C. A. *et al.* (1989) ‘The graded distribution of the dorsal morphogen is initiated by selective nuclear transport in Drosophila’, *Cell*, 59(6), pp. 1165–1177. doi: 10.1016/0092-

8674(89)90772-1.

Sainsbury, S., Bernecky, C. and Cramer, P. (2015) 'Structural basis of transcription initiation by RNA polymerase II', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 16(3), pp. 129–143. doi: 10.1038/nrm3952.

Sainsbury, S., Niesser, J. and Cramer, P. (2013) 'Structure and function of the initially transcribing RNA polymerase II–TFIIB complex', *Nature*, 493(7432), pp. 437–440. doi: 10.1038/nature11715.

Santa, F. De *et al.* (2010) 'A Large Fraction of Extragenic RNA Pol II Transcription Sites Overlap Enhancers', *PLoS Biology*. Public Library of Science, 8(5), p. e1000384. doi: 10.1371/JOURNAL.PBIO.1000384.

Saponaro, M. *et al.* (2014) 'RECQL5 Controls Transcript Elongation and Suppresses Genome Instability Associated with Transcription Stress', *Cell*, 157(5), pp. 1037–1049. doi: 10.1016/j.cell.2014.03.048.

Savinkova, L. *et al.* (2013) 'An Experimental Verification of the Predicted Effects of Promoter TATA-Box Polymorphisms Associated with Human Diseases on Interactions between the TATA Boxes and TATA-Binding Protein', *PLoS ONE*. Edited by M. Gasset, 8(2), p. e54626. doi: 10.1371/journal.pone.0054626.

Schor, I. E. *et al.* (2017) 'Promoter shape varies across populations and affects promoter evolution and expression noise', *Nature Genetics*. Nature Publishing Group, 49(4), pp. 550–558. doi: 10.1038/ng.3791.

Schulz, K. N. and Harrison, M. M. (2019) 'Mechanisms regulating zygotic genome activation', *Nature Reviews Genetics*. Nature Publishing Group, pp. 221–234. doi: 10.1038/s41576-018-0087-x.

Seila, A. C. *et al.* (2008) 'Divergent Transcription from Active Promoters', *Science*, 322(5909), pp. 1849–1851. doi: 10.1126/science.1162253.

Senecal, A. *et al.* (2014) 'Transcription factors modulate c-Fos transcriptional bursts', *Cell Reports*. The Authors, 8(1), pp. 75–83. doi: 10.1016/j.celrep.2014.05.053.

Shao, W., Alcantara, S. G.-M. and Zeitlinger, J. (2019) 'Reporter-ChIP-nexus reveals strong contribution of the Drosophila initiator sequence to RNA polymerase pausing', *eLife*, 8. doi: 10.7554/eLife.41461.

Shao, W. and Zeitlinger, J. (2017) 'Paused RNA polymerase II inhibits new transcriptional initiation', *Nature GeNetics VOLUME*, 49(7). doi: 10.1038/ng.3867.

Shiraki, T. *et al.* (2003) 'Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage.', *Proceedings of the National Academy of Sciences of the United States of America*, 100(26), pp. 15776–81. doi: 10.1073/pnas.2136655100.

Shlyueva, D., Stampfel, G. and Stark, A. (2014) 'Transcriptional enhancers: from properties to genome-wide predictions', *Nature Reviews Genetics*, 15(4), pp. 272–286. doi: 10.1038/nrg3682.

Sloutskin, A. *et al.* (2015) 'ElemeNT: a computational tool for detecting core promoter elements', *Transcription*, 6(3), pp. 41–50. doi: 10.1080/21541264.2015.1067286.

Smale, S. T. and Baltimore, D. (1989) 'The "initiator" as a transcription control element', *Cell*, 57(1), pp. 103–113. doi: 10.1016/0092-8674(89)90176-1.

Smale, S. T. and Kadonaga, J. T. (2003) 'The RNA polymerase II core promoter.', *Annual review of biochemistry*, 72(1), pp. 449–79. doi: 10.1146/annurev.biochem.72.121801.161520.

Soutourina, J. (2018) 'Transcription regulation by the Mediator complex', *Nature Reviews Molecular Cell Biology*, 19(4), pp. 262–274. doi: 10.1038/nrm.2017.115.

Stathopoulos, A. *et al.* (2002) 'Whole-genome analysis of dorsal-ventral patterning in the Drosophila embryo.', *Cell*, 111(5), pp. 687–701. doi: 10.1016/s0092-8674(02)01087-5.

Steurer, B. *et al.* (2018) 'Live-cell analysis of endogenous GFP-RPB1 uncovers rapid turnover of initiating and promoter-paused RNA Polymerase II', *Proceedings of the National Academy of Sciences*. National Academy of Sciences, p. 201717920. doi: 10.1073/PNAS.1717920115.

Struhl, K. and Moqtaderi, Z. (1998) 'The TAFs in the HAT', *Cell*, 94(1), pp. 1–4. doi:

10.1016/S0092-8674(00)81213-1.

Suter, D. M. *et al.* (2011) 'Mammalian genes are transcribed with widely different bursting kinetics', *Science*. doi: 10.1126/science.1198817.

Swain, P. S., Elowitz, M. B. and Siggia, E. D. (2002) 'Intrinsic and extrinsic contributions to stochasticity in gene expression', *Proceedings of the National Academy of Sciences*, 99(20), pp. 12795–12800. doi: 10.1073/pnas.162041399.

Tadros, W. *et al.* (2003) 'Regulation of maternal transcript destabilization during egg activation in Drosophila.', *Genetics*, 164(3), pp. 989–1001. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12871909 (Accessed: 15 October 2019).

Tadros, W., Westwood, J. T. and Lipshitz, H. D. (2007) 'The Mother-to-Child Transition', *Developmental Cell*, 12(6), pp. 847–849. doi: 10.1016/j.devcel.2007.05.009.

Tantale, K. *et al.* (2016) 'A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting', *Nature Communications*. doi: 10.1038/ncomms12248.

Theisen, J. W. M., Lim, C. Y. and Kadonaga, J. T. (2010) 'Three key subregions contribute to the function of the downstream RNA polymerase II core promoter.', *Molecular and cellular biology*, 30(14), pp. 3471–9. doi: 10.1128/MCB.00053-10.

Thomas, M. C. and Chiang, C.-M. (2006) 'The general transcription machinery and general cofactors.', *Critical reviews in biochemistry and molecular biology*, 41(3), pp. 105–78. doi: 10.1080/10409230600648736.

Thomsen, S. *et al.* (2010) 'Genome-wide analysis of mRNA decay patterns during early Drosophila development', *Genome Biology*, 11(9), p. R93. doi: 10.1186/gb-2010-11-9-r93.

Torres-Padilla, M. E. and Tora, L. (2007) 'TBP homologues in embryo transcription: who does what?', *EMBO reports*. European Molecular Biology Organization, 8(11), pp. 1016–8. doi: 10.1038/sj.embor.7401093.

Tsanov, N. *et al.* (2016) 'SmiFISH and FISH-quant - A flexible single RNA detection approach with super-resolution capability', *Nucleic Acids Research*, 44(22). doi: 10.1093/nar/gkw784.

Tutucci, E., Vera, M., *et al.* (2018) 'An improved MS2 system for accurate reporting of the mRNA life cycle.', *Nature methods*. NIH Public Access, 15(1), pp. 81–89. doi: 10.1038/nmeth.4502.

Tutucci, E., Livingston, N. M., *et al.* (2018) 'Imaging mRNA In Vivo, from Birth to Death', *Annual Review of Biophysics*. Annual Reviews , 47(1), pp. 85–106. doi: 10.1146/annurev-biophys-070317-033037.

Vastenhouw, N. L., Cao, W. X. and Lipshitz, H. D. (2019) 'The maternal-to-zygotic transition revisited', *Development (Cambridge, England)*, 146(11). doi: 10.1242/dev.161471.

Veenstra, G. J., Weeks, D. L. and Wolffe, A. P. (2000) 'Distinct roles for TBP and TBP-like factor in early embryonic gene transcription in Xenopus.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 290(5500), pp. 2312–5. doi: 10.1126/science.290.5500.2312.

Vo Ngoc, L. *et al.* (2017) 'The punctilious RNA polymerase II core promoter.', *Genes & development*, 31(13), pp. 1289–1301. doi: 10.1101/gad.303149.117.

Vo Ngoc, L., Kassavetis, G. A. and Kadonaga, J. T. (2019) 'The RNA Polymerase II Core Promoter in *Drosophila*', *Genetics*, 212(1), pp. 13–24. doi: 10.1534/genetics.119.302021.

Wang, H. D., Trivedi, A. and Johnson, D. L. (1997) 'Hepatitis B virus X protein induces RNA polymerase III-dependent gene transcription and increases cellular TATA-binding protein by activating the Ras signaling pathway.', *Molecular and Cellular Biology*, 17(12), pp. 6838–6846. doi: 10.1128/MCB.17.12.6838.

Weinmann, R. and Roeder, R. G. (1974) 'Role of DNA-Dependent RNA Polymerase III in the Transcription of the tRNA and 5S RNA Genes', *Proceedings of the National Academy of Sciences*, 71(5), pp. 1790–1794. doi: 10.1073/pnas.71.5.1790.

Whyte, W. A. *et al.* (2013) 'Master transcription factors and mediator establish super-enhancers at key cell identity genes.', *Cell*, 153(2), pp. 307–19. doi: 10.1016/j.cell.2013.03.035.

Wong, J. M. and Bateman, E. (1994) 'TBP-DNA interactions in the minor groove discriminate between A:T and T:A base pairs', *Nucleic Acids Research*, 22(10), pp. 1890–1896. doi:

10.1093/nar/22.10.1890.

Yamaguchi, Y., Shibata, H. and Handa, H. (2013) 'Transcription elongation factors DSIF and NELF: promoter-proximal pausing and beyond.', *Biochimica et biophysica acta*, 1829(1), pp. 98–104. doi: 10.1016/j.bbagrm.2012.11.007.

Zabidi, M. A. *et al.* (2015) 'Enhancer–core-promoter specificity separates developmental and housekeeping gene regulation', *Nature*. Nature Publishing Group, 518(7540), pp. 556–559. doi: 10.1038/nature13994.

Zawel, L. and Reinberg, D. (1992) 'Advances in RNA polymerase II transcription', *Current Opinion in Cell Biology*, 4(3), pp. 488–495. doi: 10.1016/0955-0674(92)90016-6.

Zehavi, Y. *et al.* (2014) 'The core promoter composition establishes a new dimension in developmental gene networks', *Nucleus*, 5(4), pp. 298–303. doi: 10.4161/nucl.29838.

Zehring, W. A. *et al.* (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 of the United States of America*, 85(11), pp. 3698–702. doi: 10.1073/pnas.85.11.3698.

Zeitlinger, J. *et al.* (2007) 'Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the Drosophila embryo', *Genes & Development*, 21(4), pp. 385–390. doi: 10.1101/gad.1509607.

Zenklusen, D., Larson, D. R. and Singer, R. H. (2008) 'Single-RNA counting reveals alternative modes of gene expression in yeast', *Nature Structural & Molecular Biology*, 15(12), pp. 1263–1271. doi: 10.1038/nsmb.1514.

Zhang, Z. *et al.* (2016) 'Rapid dynamics of general transcription factor TFIIB binding during preinitiation complex assembly revealed by single-molecule analysis.', *Genes & development*, 30(18), pp. 2106–2118. doi: 10.1101/gad.285395.116.

Zhao, X. and Herr, W. (2002) 'A regulated two-step mechanism of TBP binding to DNA: a solvent-exposed surface of TBP inhibits TATA box recognition.', *Cell*, 108(5), pp. 615–27. doi: 10.1016/s0092-8674(02)00648-7.

Zhu, A. and Kuziora, M. A. (1996) 'Homeodomain Interaction with the β Subunit of the General Transcription Factor TFIIE', *Journal of Biological Chemistry*, 271(35), pp. 20993–20996. doi: 10.1074/jbc.271.35.20993.

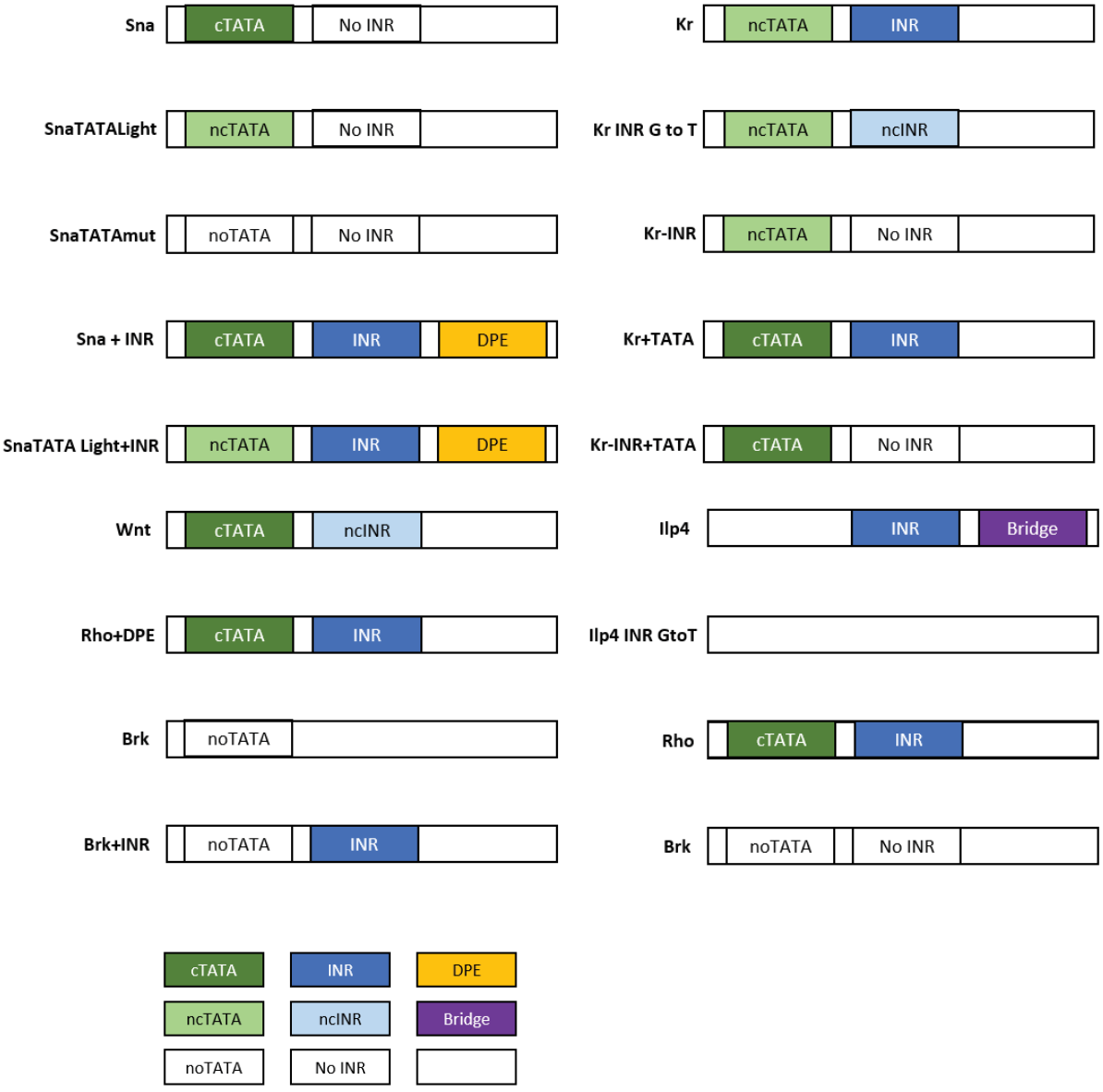
Supplementary Table 1. Core promoter elements annotations and its interaction. As described (Sloutskin *et al.*, 2015; Haberle and Stark, 2018)

Core promoter motif	Sequence logo	Position relative to the TSS	Bound by
TATA-box		-30/-31 to -23/-24	TBP
Inr		-2 to +4	TAF1 and TAF2
BREu		–38 to –32	TFIIB
BREd		–23 to –17	TFIIB
DPE*		+28 to +33	TAF6 and TAF9 and possibly TAF1
MTE*		+18 to +29	Possibly TAF1 and TAF2
Bridge*		+18 to +22; +30 to +33	-
dTCT		-2 to +6	-

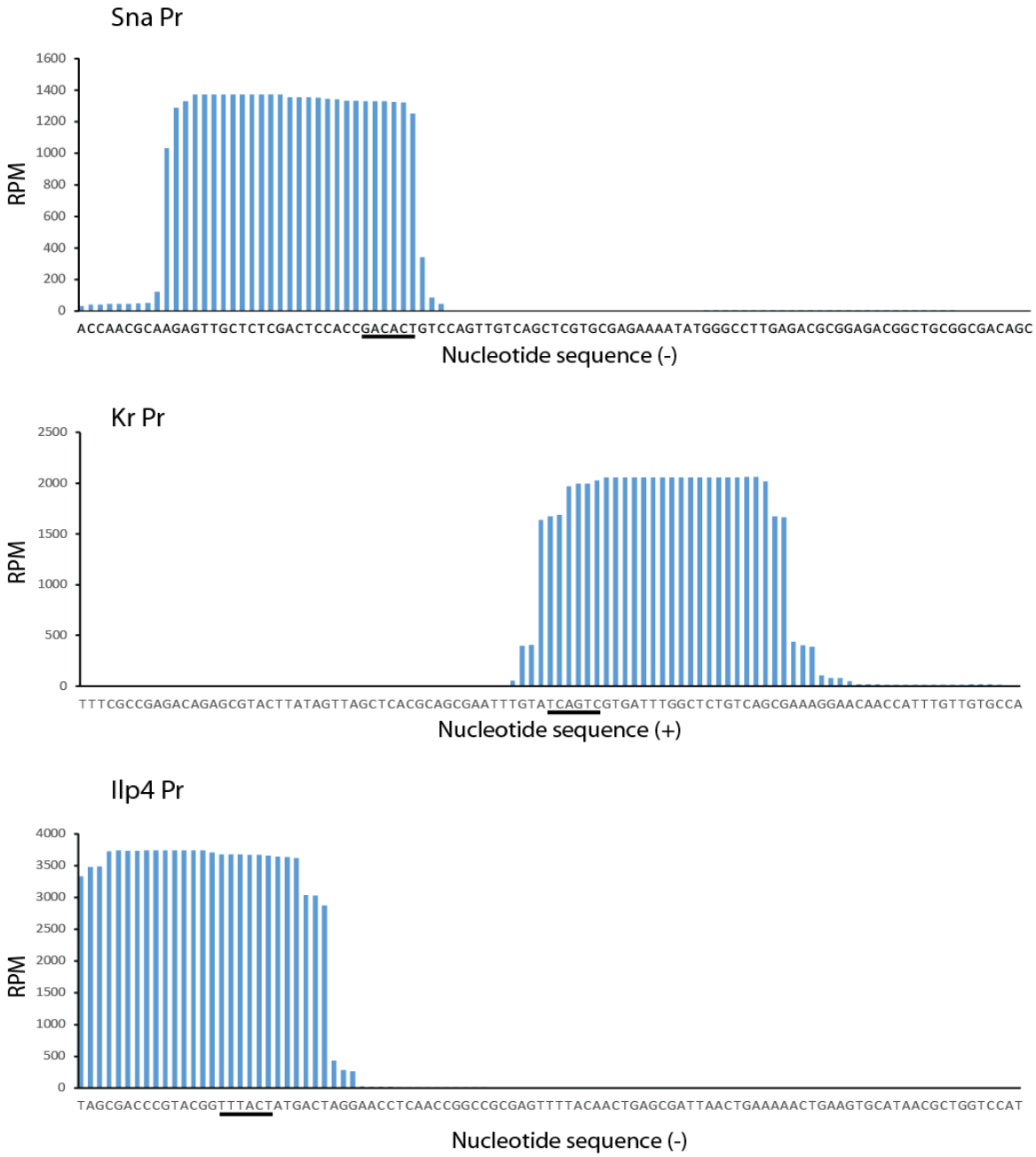
*Inr dependent

Supplementary Table 2. Developmental core promoter and their element motifs

Courtesy of Matthieu Dejean

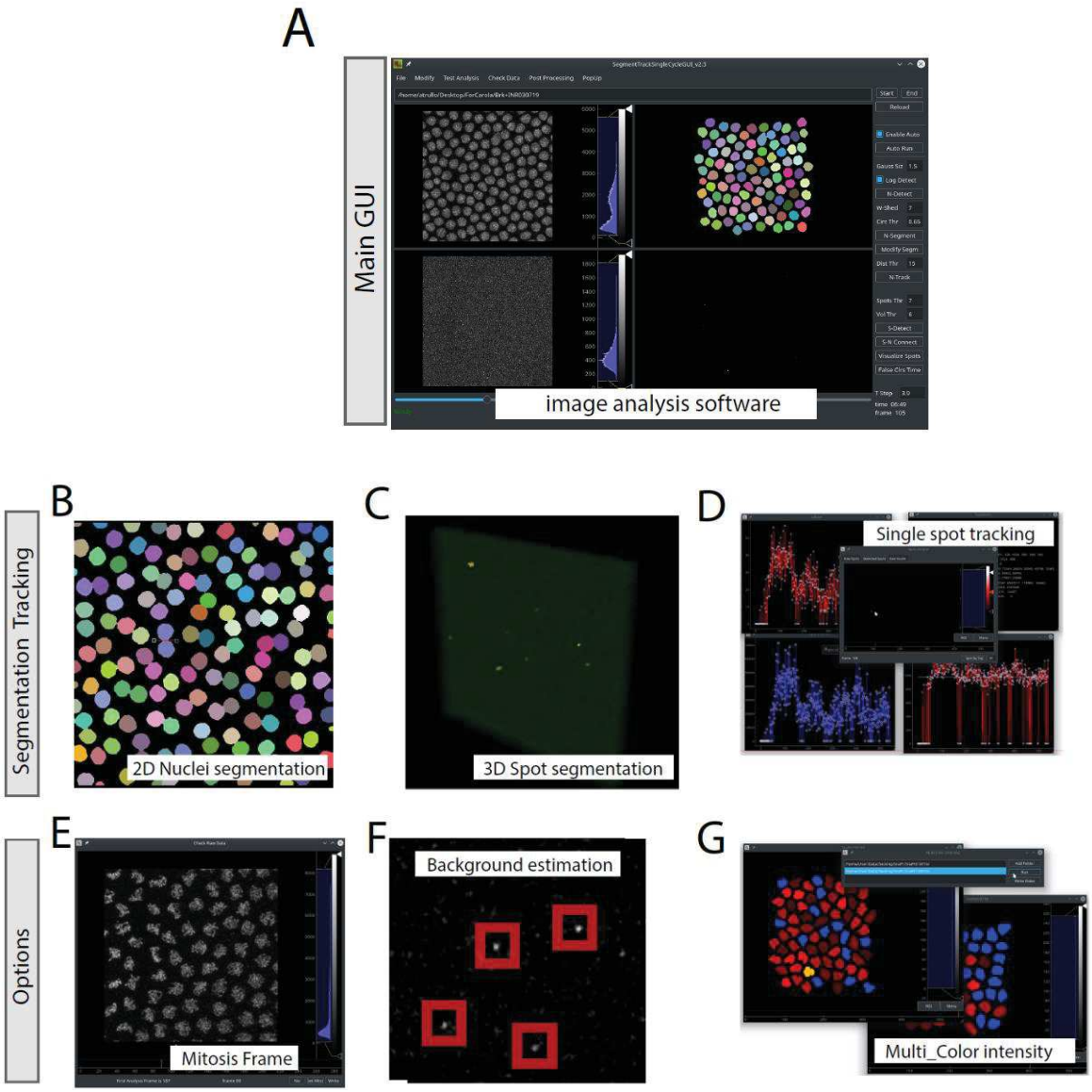


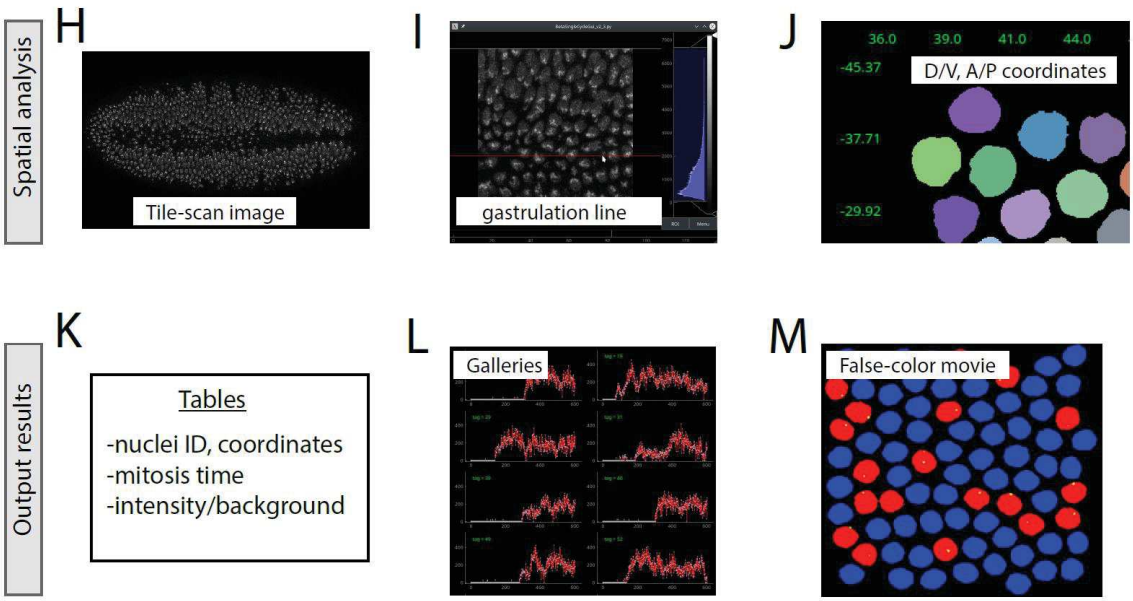
Supplementary Figure 1.



Example of CAGE data to identify TSS. Analysis of the CAGE data in Drosophila embryos (yellow cinnabar brown peck strain), between 2-4h after egg laying. Data were downloaded from modENCODE projet (modENCODE_5344). The analysis was performed by the bio-computational platform of the IGMM institute; by Amal Makrini. The alignment of the data was done using dm3 genome version from flyBase®. Histograms represent the enrichment of reads (rpm) on the *sna, kr* and *ilp4* promoter, indicative of the TSS position for these promoters.

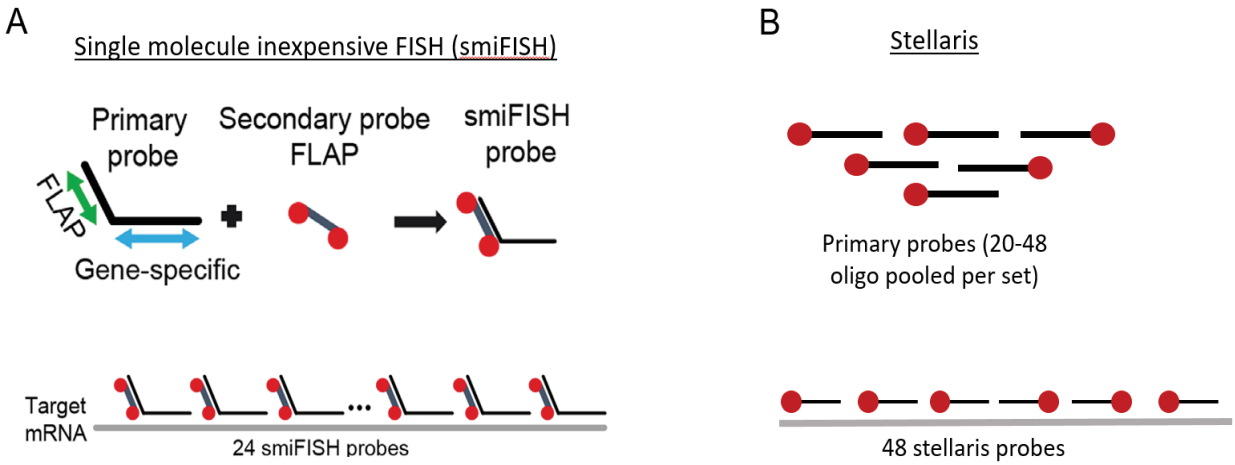
Supplementary Figure 2.





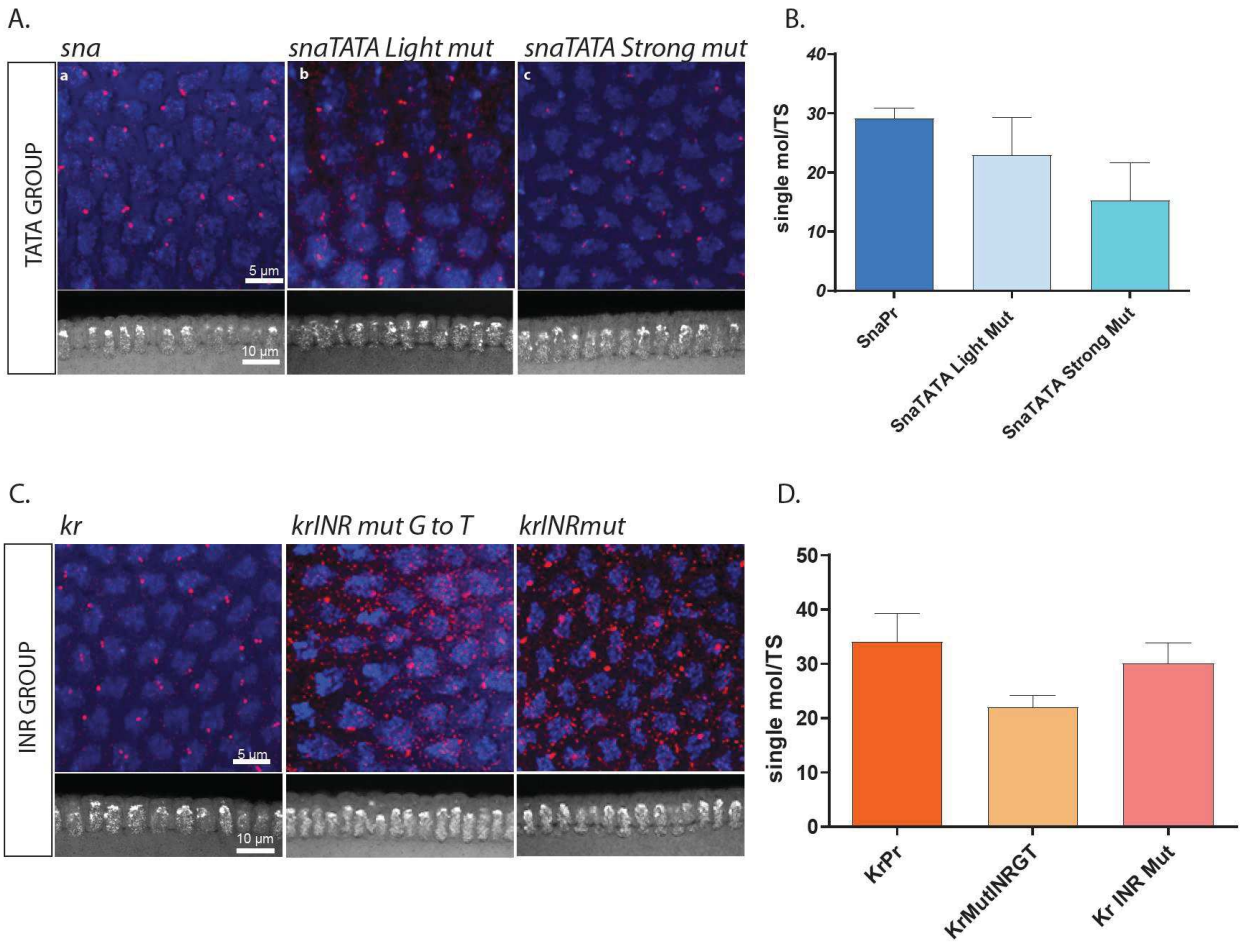
Analysis of the live imaging data. Developed by (Trullo et al., *in preparation)*. (A) Snapshot of the graphical user interphase, where the live imaging AiryScan processed data is loaded to extract quantitative and qualitative information. (B) Snapshot of the segmentation and tracking of the nuclei, before segmentation the nuclei are maximum projected in 2D. A manual tool allows the correction of not well-segmented nuclei. (C) Snapshot of the 3D spot segmentation. (D) Snapshot of a tool allowing the visualisation and the tracking for single spot traces on time. (E)(F) Snapshot of a tool allowing to place as starting time (t=0) the mitosis, extraction of the surrounding background on the green channel (spot channel), respectively. (G) Snapshot of a tool that enables to false colour the nuclei according to the level of intensity produced frame by frame, the final output is in video format. (H)(I)(J) Snapshot to allow spatial detection, the tile scan is used to set the coordinates of the image acquisition, the last frame where the gastrulation is recorded serves to fix the analysed area (50µm around the gastrulation line), finally we obtained A/P and D/V coordinates for each nuclei. (K) (L) (M) examples of the information obtained for the processed data.

Supplementary figure 3.



Single inexpensive FISH vs Stellaris FISH principle. (A) Figure from (Tsanov *et al.*, 2016), the primary probes are composed of a sequence combining specific probes (26-32 bp) for the gene of interest to a sequence called FLAP (28bp), this are unlabelled probes. The secondary probe is composed of a complementary sequence of the FLAP that is labelled in 5' and in 3' by a selected fluorescence molecule (Cy3, Cy5, Alexa-488…). The smiFISH probe is obtained by PCR hybridization of the primary and secondary probe (supplemental protocol1). Therefore, primary probes could be labelled with different fluorescence proteins without need to re-synthetize a new probe set. (B) Stellaris technique principle, here primary probes are directly labelled in the 5' sequence. Thus, primary probes do not allow multicolour labelling as for the smiFISH, increasing the cost of purchasing.

Supplementary figure 4.



Estimation of the mRNA levels by SmiFISH. (A) Example of the raw data images for the TATA group. Heterozygous embryos were labelled with 24X cy3-yellow probes. The staging of the embryos on the interphase 14 were done using the membrane invagination (bottom, square). (B) Histogram of the number of single molecules per TS for the TATA group (*sna*, n=5 embryos; *snaTATA Light mut*, n=2 embryos, *snaTATA mut*, n=5 embryos). (C) Raw data image for the INR group (with the same setting as described with the TATA group) (D) Histogram of the number of single molecules per TS for the INR group (kr, n=5; *krINRmut G to T*, n=3 embryos, *krINRmut*, n=3 embryos).

Supplementary Protocol.

smiFISH protocol for *Drosophila* embryos, adapted from (Tsanov et al, 2016).

1. Probe set Preparation
   Primary probes are produced in 96-well plates in Tris-EDTA pH 8.0 (TE) Buffer, at final concentration of 100µM.
   Prepare an equimolar mixture of 24 primary probes, and dilute the mixture 5 times in TE buffer 1x = Probeset* (final concentration of individual probes is 0.833µM)

2. Flap preparation
   Suspend the lyophilized FLAP in TE buffer at the final concentration of 100µM. make aliquots and stored at -20°C in the dark.

3. FLAP hybridization reaction
   By PCR
   Prepare the reaction mix (final volume 10µl)

Element	Volume	Amount
Probeset*	2µl	40 pmol
100µM FLAP	1µl	100 pmol
10X NEB 3	1µl	
H2O	6µl	

PCR programme (smiFISH)

1 cycle	85°C	3min
1 cycle	65°C	3min
1 cycle	25°C	5min

Lid: 99°C
Put the sample tubes on ice until use. For long term storage, keep hybridized duplexes at -20°C in the dark.

4. Embryos Preparation:
   Day 1
   Rock embryos in 1:1, MeOH: EtOH 5min
   Rince 2x with EtOH
   Rock in EtOH 5min
   Rince 2x EtOH

Rock in EtOH 5min

Rince 2x MeOH
Rock in MeOH 5min
Rinse 2x with PBT
Rock 4x in PBT 15min
Rock in 15% formamide (deionize) 1X SSC 15min
(Prepare Mix 1 & Mix 2 on ice, in fold, during last 15 min)

Element	Volume for 250µl (mix1 + mix2)
Mix 1	
20X SSC	12.5
20µg/µl *E.coli* tRNA	4.25
100% formamide (deionize)	37.5
Flap-probes duplexes	5
H20	65.75

Mix 2	
20mg/ml RNAse-free BSA	2.5
200mM VRC*	2.5
40% dextran sulphate	66.25
H2O	53.75

VRC* vortex before use

Vortex thoroughly Mix2. Then add Mix1 to Mix2 and vortex again.
Remove 15% formamide (deionize) 1X SSC from embryos, then add 250µl from (mix1 +mix2).
Incubate overnight at 37°C in the dark.

Day 2 *(*in the dark)
Rinse 2x in freshly prepared 15% formamide (deionize) 1X SSC at 37°C.
Rinse twice in PBT and remove
Rock in 250µl of 1:50000 DAPI, 20 min.
Rinse 3x with PBT
Mounting.

Note
For 5 ml of 15% formamide (deionize) 1X SSC.

20x SCC	250µl
100% Formamide deionize	750µl
H2O	4ml

III. APPENDICES

I collaborated to the work of the Dr Dufourt and Dr Trullo, to test the effect of the pioneer factor Zelda on transcriptional dynamics and memory in Drosophila embryos.

I contributed by generating drosophila transgenics used in this study. In addition, I performed some of the live imaging acquisition and was involved in the scientific discussions of the manuscript.

nature COMMUNICATIONS

Corrected: Publisher correction

ARTICLE

DOI: 10.1038/s41467-018-07613-z OPEN

Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs

Jeremy Dufourt[1], Antonio Trullo[1], Jennifer Hunter[1], Carola Fernandez[1], Jorge Lazaro[1], Matthieu Dejean[1], Lucas Morales[1], Saida Nait-Amer[1], Katharine N. Schulz[2], Melissa M. Harrison[2], Cyril Favard[3], Ovidiu Radulescu[4] & Mounia Lagha[1]

Pioneer transcription factors can engage nucleosomal DNA, which leads to local chromatin remodeling and to the establishment of transcriptional competence. However, the impact of enhancer priming by pioneer factors on the temporal control of gene expression and on mitotic memory remains unclear. Here we employ quantitative live imaging methods and mathematical modeling to test the effect of the pioneer factor Zelda on transcriptional dynamics and memory in *Drosophila* embryos. We demonstrate that increasing the number of Zelda binding sites accelerates the kinetics of nuclei transcriptional activation regardless of their transcriptional past. Despite its known pioneering activities, we show that Zelda does not remain detectably associated with mitotic chromosomes and is neither necessary nor sufficient to foster memory. We further reveal that Zelda forms sub-nuclear dynamic hubs where Zelda binding events are transient. We propose that Zelda facilitates transcriptional activation by accumulating in microenvironments where it could accelerate the duration of multiple pre-initiation steps.

[1] Institut de Génétique Moléculaire de Montpellier, University of Montpellier, CNRS-UMR 5535, 1919 Route de Mende, Montpellier 34293 cedex 5, France. [2] Department of Biomolecular Chemistry, University of Wisconsin School of Medicine and Public Health Madison, 6204B Biochemical Sciences Building 440 Henry Mall, Madison 53706 WI, USA. [3] Institut de Recherche en Infectiologie de Montpellier, CNRS, University of Montpellier UMR 9004, 1919 route de Mende, Montpellier 34293 cedex 5, France. [4] DIMNP, UMR CNRS 5235, University of Montpellier, Place E. Bataillon - Bât. 24 cc 107, Montpellier 34095 cedex 5, France. These authors contributed equally: Jeremy Dufourt, Antonio Trullo. Correspondence and requests for materials should be addressed to M.L. (email: mounia.lagha@igmm.cnrs.fr)

During the first stages of metazoan embryogenesis, the zygotic genome is largely quiescent, and development relies on maternally deposited mRNAs and proteins. Transcriptional activation of the zygotic genome occurs only hours after fertilization and requires specific transcription factors. In *Drosophila*, the transcriptional activator Zelda (Zld) plays an essential role in zygotic genome activation[1].

Zelda exhibits several pioneer factor properties such as (i) priming *cis*-regulatory elements prior to gene activation, which establishes competence for the adoption of cell fates, and (ii) opening chromatin, which facilitates subsequent binding by classical transcription factors. Indeed, hours before their activation, Zelda binds to thousands of genomic regions corresponding to enhancers of early-activated genes[2,3]. Zelda binding is associated with modifications of the associated chromatin landscape and overall chromatin accessibility[2–5].

Moreover, this ubiquitously distributed activator is able to potentiate the action of spatially restricted morphogens (*e.g* Bicoid and Dorsal)[6,7]. Consequently, target gene response is strengthened by Zelda binding, both spatially and temporally, for developmental enhancers[8] as well as for synthetic enhancers where input parameters are tightly controlled[9].

In addition to these properties, classical pioneer factors bind nucleosomes and are typically retained on the chromosomes during mitosis[10,11]. Mitotic retention of these transcription factors places them as ideal candidates for the transmission of chromatin states during cellular divisions, through a mitotic bookmarking mechanism.

However, the putative function of Zelda in mitotic memory has thus far not been examined. Moreover, all studies concerning Zelda's role as a quantitative timer have been performed on fixed embryos with limited temporal resolution. Thus, the effects of Zelda on the temporal dynamics of transcriptional activation and on transmission of active states through mitosis remain to be addressed.

Here we employ quantitative live imaging methods to directly measure the impact of Zelda on transcriptional dynamics in vivo. We show that increasing the number of Zelda-binding sites at an enhancer fosters temporal coordination in de novo gene activation (synchrony), regardless of past transcriptional states. By monitoring transcription in Zelda-maternally depleted embryos, we reveal that Zelda is dispensable for mitotic memory. On the contrary, we found that Zelda allows the temporal disadvantage of arising from a transcriptionally inactive mother to be bypassed. Using a mathematical modeling framework, we propose that mitotic memory requires long-lasting transitions, which are accelerated by Zelda, thus overriding mitotic memory of silent states. By analyzing Zelda protein kinetics during many cell cycles, here we report that while it is not retained on chromosomes during mitosis, this transcription factor exhibits a highly dynamic behavior. We observe that intra-nuclear distribution of Zelda is not homogeneous, and that it accumulates in local hubs, where its residence time is surprisingly transient (in the second range). Taken together, our results provide insights into Zelda-mediated temporal control of transcription during zygotic genome activation in vivo.

Results and Discussion

Zelda fosters temporal transcriptional coordination. In the blastoderm embryo, two enhancers control *snail* (*sna*) gene expression, a proximal (primary) and a distal (shadow enhancer) [12,13]. Both enhancers are bound by Zelda (Zld) at early nuclear cycles (nc)[2,3]. When driven by its native enhancers, endogenous expression of *snail* is extremely rapid, precluding analysis of the impact of Zelda on fine-tuning the timing of its activation. We therefore used a previously described truncated version of the *sna* shadow enhancer (*snaE*) that leads to a stochastic activation, compatible with the tracking of transcriptional memory across mitosis[14].

To follow transcriptional dynamics in a quantitative manner, we created a series of *enhancer* < *promoter* < *MS2-yellow* transgenes, with a unique minimal promoter (*sna* promoter) and an *MS2-yellow* reporter (Fig. 1a). Upon transcription, MS2 stem-loops are rapidly bound by the maternally provided MCP-GFP, which allows tracking of transcriptional activation in living embryos[15]. To decipher the role of Zelda in *cis*, a variable number of canonical Zld-binding sites (CAGGTAG)[16] were added to the *snaE* in three locations, either close to the promoter (Zld 3'), in the middle (Zld mid) or at a distance (Zld 5') (Fig. 1a).

Using fluorescent in situ hybridization, we confirmed that the pattern of expression of *snaE-extraZld* transgenes was dictated by the regulatory logic of the *sna* enhancer, and expression was indeed evident in the presumptive mesoderm, similarly to the endogenous *sna* pattern (Fig. 1b–g). To precisely measure temporal dynamics, we took advantage of our MCP/MS2 system to track transcriptional activation in living embryos.

First, we implemented an automatic image analysis pipeline that segments nuclei in nc14 (Fig. 1h, i) and tracks the time of transcriptional activation (Fig. 1j, k). In the early fly embryo, a subset of transcriptional activators, including Dorsal, control gene expression in a graded manner[17]. It was therefore important to quantify temporal dynamics of gene activation in a spatially controlled domain. For this purpose, we used the ventral furrow (site of invagination during gastrulation process) as a landmark to define dorso-ventral coordinates (Fig. 1l, m and Supplementary Fig. 1a). Unless otherwise indicated, we studied temporal dynamics of gene activation in a region of 50µm centered around the ventral furrow to ensure non-limiting levels of the *sna* activators Dorsal (Supplementary Fig. 1b) and Twist[18].

We initially examined the impact of Zelda on the timing of gene activation. As identified with static approaches[6,7,16], our dynamic data showed that increasing the number of Zld-binding sites resulted in precocious transcriptional activation (Fig. 1n–q Supplementary Fig. 1c, d and Supplementary Movie 1-4). Zelda-binding impacted not only the onset of gene activation but also the temporal coordination among a spatially defined domain (*i.e* the presumptive mesoderm, 50µm around the ventral furrow), referred to as synchrony[19] (Fig. 1r and Supplementary Fig. 1d). For example, at precisely 10 min into interphase 14, expression of the *snaE* transgene was not coordinated (≈16% of active nuclei), whereas the number of active nuclei at this stage increased upon the addition of extra Zld-binding sites (Fig. 1n–q). The precise kinetics of gene activation, *i.e* synchrony curves, was quantified as a percentage of active nuclei within a defined spatial domain for each transgene during the first 30 min of nc14. *SnaE* generated a slow activation kinetic, where half of the domain is activated at t50≈18 min (Fig. 1r). Adding only a single CAGGTAG sequence, 5' to the enhancer accelerated this kinetic (t50≈13 min). With two extra Zld-binding sites synchrony is further increased (t50≈9.5 min). The activation kinetics of two extra canonical Zld sites mimicked the dynamics generated by the intact long shadow enhancer (t50≈9.5 min) (Fig. 1r). The most dramatic synchrony shift was caused by the addition of a single CAGGTAG sequence close to the promoter (+1Zld 3') (t50≈6.5 min) (Supplementary Fig. 1d). This behavior was not changed when two additional Zelda sites were added, suggesting that promoter proximal Zelda binding causes a strong effect, that cannot be further enhanced by the addition of more distant sites (Fig. 1r and Supplementary Fig. 1d). By moving the CAGGTAG site away from the promoter, to the middle of the enhancer, synchrony was decreased (t50≈7.5 min) (Supplementary Fig. 1d)

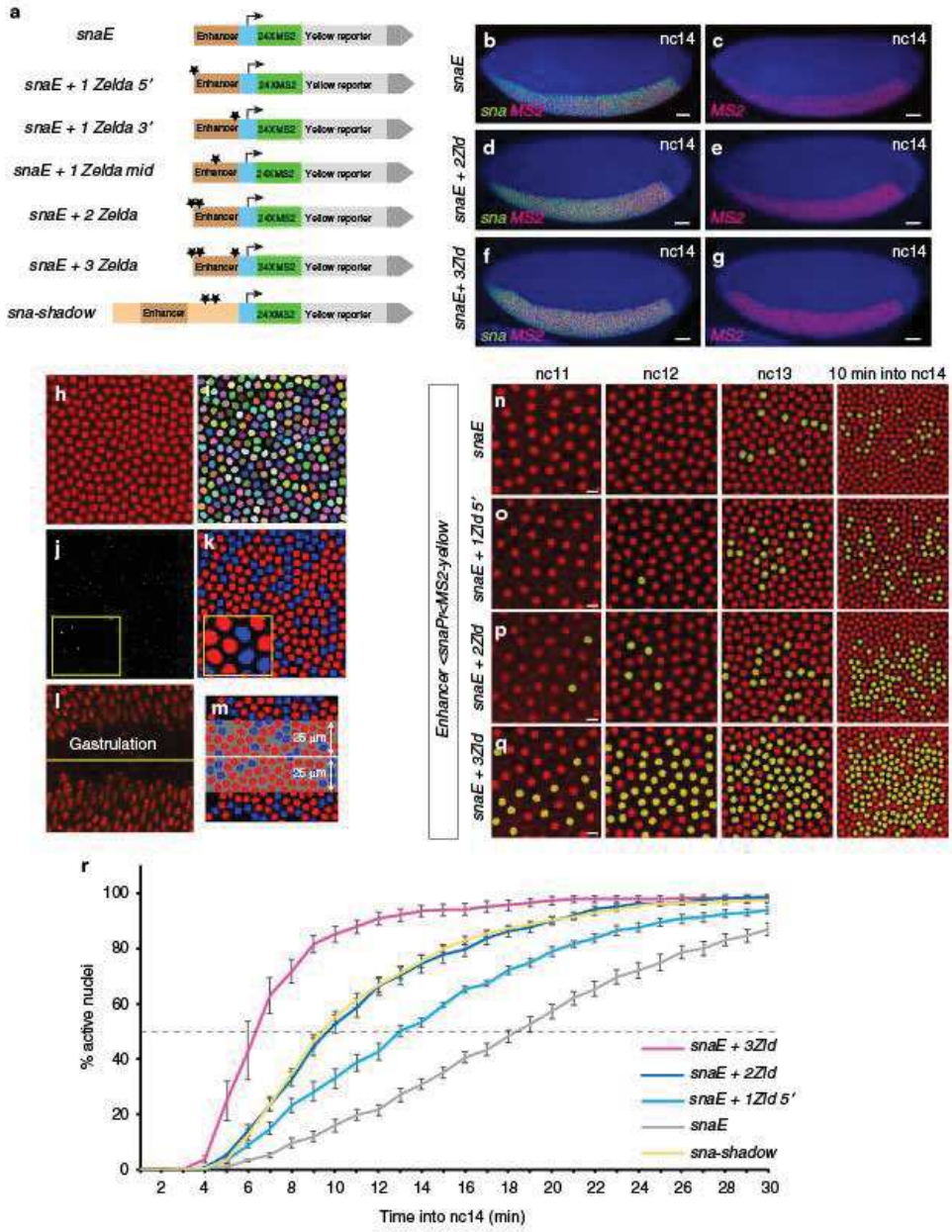

a
snaE
snaE + 1 Zelda 5′
snaE + 1 Zelda 3′
snaE + 1 Zelda mid
snaE + 2 Zelda
snaE + 3 Zelda
sna-shadow
Enhancer
24XMS2
Yellow reporter
b
c
d
e
f
g
nc14
sna MS2
MS2
snaE
snaE + 2Zld
snaE+ 3Zld
h
j
k
l
Gastrulation
m
25 µm
nc11
nc12
nc13
10 min into nc14
n
o
p
q
Enhancer <snaPr<MS2-yellow
snaE
snaE + 1Zld 5′
snaE + 2Zld
snaE + 3Zld
r
% active nuclei
Time into nc14 (min)
snaE + 3Zld
snaE + 2Zld
snaE + 1Zld 5′
snaE
sna-shadow

Fig. 1 Zelda-binding sites provide quantitative temporal control of gene expression. **a** Schematic representation of the truncated version of the *sna* shadow enhancer (*snaE*) and the variants carrying different numbers and locations of extra Zelda (Zld)-binding sites (black stars represent canonical Zld-binding sites: CAGGTAG). Each enhancer controls a common minimal promoter (blue box, *sna* promoter) that transcribes a *yellow* gene with 24 *MS2* repeats located at its 5′ end. **b–g** Transgenic embryos stained with a *MS2* probe (red, panels **b–g**) and with a *snail* probe (green, **b**, **d**, **f**). Dorsal views at nc14, anterior to the left, scale bars represent 30µm. **h–m** Workflow for automatic tracking of single nucleus transcriptional activity. **h**, **i** Snapshot of a maximum intensity projected Z-stack from live imaging movie representing nuclei labeled with *His2Av-mRFP* transgene (**h**) and associated automatic segmentation, where each tracked nucleus is given a random color (**i**). **j** Maximum intensity projected Z-stack image extracted from a typical movie showing the transcriptional foci (white dot). **k** False-colored image, representing the segmented transcriptionally inactive nuclei (blue), active nuclei (red) and associated transcriptional foci (yellow dot). **l** Maximum intensity projected Z-stack image representing a typical gastrulation, used as a landmark to determine dorso-ventral (D/V) coordinates (yellow line). **m** Representative image exhibiting the spatial domain (grey, here 25µm surrounding the ventral furrow) defined by precise D/V coordinates. **n–q** Live imaging of transcriptional activity driven by the *snaE* upon increasing numbers of Zld-binding sites. Nuclei that displayed transcriptional activity were false-colored in yellow. Snapshots from maximum intensity projected Z-stack of live imaging movies are shown from nc11 to nc14. Scale bars represent 10µm. **r** Temporal coordination profiles during nc14. Synchrony curves were obtained by quantifying transcriptional activation during nc14 in a region of 50µm around the ventral furrow: *snaE* (6 movies, $n = 970$ nuclei) (grey curve), *snaE* + *1Zld5′* (4 movies, $n = 700$ nuclei) (light blue curve), *snaE* + *2Zld* (5 movies, $n = 723$ nuclei) (dark blue curve), *snaE* + *3Zld* (4 movies, $n = 824$ nuclei (purple curve)) and *snaE-shadow* (4 movies, $n = 612$) (yellow curve). Dashed line represents 50% of activation. The time origin is the end of mitosis (start of nc14). Error bars represent SEM

suggesting a stronger role for Zelda when positioned close to the promoter.

In this system, addition of a single CAGGTAG was sufficient to increase the activation kinetics. This is in sharp contrast to what has been reported for a synthetic enhancer, whereby activation only occurred when three or more Zelda-binding sites were added[9]. This difference is likely due to the ability of Zelda to function cooperatively with other activators of *sna*, which are not present in the synthetic enhancer previously examined[9].

While our initial analysis focused on the region surrounding the ventral furrow, live imaging data also provided insights into the temporal activation in the entirety of the *sna* pattern. By temporally tracking activation along the dorso-ventral axis, we could observe the consequences of the Dorsal nuclear spatial gradient[20] (Supplementary Fig. 1e). Indeed, the first activated nuclei were located where activators are at their peaks levels, whilst the late activated nuclei were closer to the mesodermal border, as previously suggested by analysis of fixed embryos and modeling[7,21]. By comparing activation timing between 0-50µm from the ventral furrow, we show that adding extra Zld-binding sites led to more rapid activation across the entire spatial domain (Supplementary Fig. 1e). These temporal data are consistent with the idea that Zelda potentiates the binding of spatially restricted morphogens like Bicoid or Dorsal[6–8,22] and thus support the previously described pioneering activities of Zelda[5].

Zelda bypasses transcriptional memory. Our ability to track transcriptional activation in live embryos provided the opportunity to determine activity through multiple nuclear cycles. Using the sensitized *snaE* transgene, we recently documented the existence of a transcriptional mitotic memory, whereby the transcriptional status of mother nuclei at nc13 influences the timing of activation of the descendants in the following cycle[14]. To address the challenge of following rapid movements during mitosis, we developed a semi-automatic mitotic nuclei tracking software (Fig. 2a). By combining lineage information with transcriptional status, we quantified the timing of activation of hundreds of nuclei in nc14 in a spatially defined domain. This allowed us to distinguish those transcriptionally active nuclei that arose from an active mother nucleus (green curves, Fig. 2b–d and Supplementary Fig. 1f, g) from those coming from an inactive mother nucleus (red curves, Fig. 2b–d and Supplementary Fig. 1f, g). Differences between the kinetics of activation of these two populations provide an estimation of transcriptional mitotic memory (Fig. 2b).

Based on the ability of Zelda to define hundreds of cis-regulatory regions through the establishment or maintenance of accessible chromatin[4,5], we reasoned that Zelda could be in part responsible for the establishment of transcriptional mitotic memory during early development. Thus, we leveraged our live imaging system to directly test the role of additional Zld-binding sites on transcriptional activation through mitosis. Contrary to our expectation, namely that extra Zelda transgenes would exhibit enhanced memory; adding extra Zld-binding sites in *cis* augmented the speed of activation in both nuclei derived from active mothers and those from inactive mothers (Fig. 2c, d and Supplementary Fig. 1f, g). Increasing the number of Zld-binding sites reduced the differences in the timing of activation for descendants from active mothers and their neighboring nuclei, arising from inactive mothers. Our results suggest that extra Zld-binding sites accelerate transcriptional dynamics regardless of the past transcriptional status. Thus, Zelda seems capable of bypassing transcriptional memory by compensating for the negative effect of having an inactive mother nucleus.

Zelda is dispensable for transcriptional mitotic memory. We then sought to test whether Zelda was required for this memory. Activation kinetics for descendants from active and inactive nuclei were quantified after reducing maternal Zelda expression using RNAi. Validating this approach, germline expression of *zld*-RNAi[7,23] reduced the levels of *zld* mRNA to 2–4% of that found in control *white*-RNAi embryos (Supplementary Fig. 3a).

Reduction of maternal Zelda decreased the synchrony of activation of our transgenes with additional Zld-binding sites (Fig. 3a–j, Supplementary Fig. 3b–d and Supplementary Movie 5, 6). In *zld*-RNAi embryos, temporal coordination was reduced, and all transgenes showed a t50≈16–19 min (Fig. 3i, j, Supplementary Fig. 3b-d and Supplementary Movie 5, 6). This confirms that the previously identified acceleration of transcriptional activation upon addition of extra Zld-binding sites is due to Zelda activity on these enhancers. Further confirming the role of maternally contributed Zelda in potentiating Dorsal-dependent gene expression, transcriptional activation was restricted to a reduced region of about ≈25 µm around the pseudo-furrow in Zelda maternally depleted embryos (Supplementary Movie 6). Together, these data demonstrate that expression of our *snaE* transgenes reflect the general properties of Zld-mediated gene expression.

We then quantified the level of memory in the *zld*-RNAi embryos with three *snaE < MS2* transgenes containing varying number of Zld-binding sites. In spite of the maternal reduction of

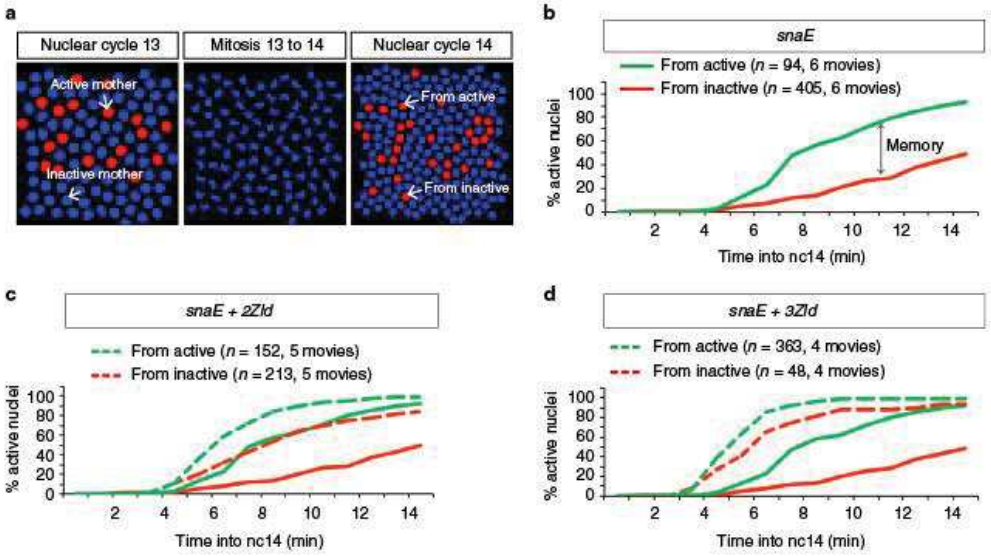


Fig. 2 Zelda accelerates gene activation regardless of transcriptional history **a** Representative images from the Graphical User Interface implemented for automatic tracking of nuclei across multiple divisions. Chromosomal splitting at the end of anaphase is automatically detected for each nucleus. **b–d** Quantification of transcriptional memory in various *snaE < MS2* embryos. At each time point in nc14, the cumulative number of activated nuclei, expressed as a percentage of the total number of nuclei, is plotted after segregation into two populations: descendants of active mothers (green) and descendants of inactive mothers (red). The timing of activation of each nucleus has been normalized to its specific timing of mitosis. Temporal dynamics from *snaE* are depicted using solid curves, while those from transgenes with extra Zld-binding sites are represented as dashed curves (**c**, **d**)

Zelda, a strong memory bias was still observed (Fig. 3k, l and Supplementary Fig. 3e). We conclude that although possessing several key properties of a pioneer factor, Zelda is not necessary to elicit heritability of active transcriptional states through mitosis.

Altogether our results show that Zelda accelerates transcription and fosters synchrony, despite playing no role in transcriptional memory maintenance.

To further examine whether Zelda binding to an enhancer was sufficient to trigger a memory bias, temporal activation from a second regulatory element (*DocE*) was examined (Supplementary Fig. 2). This region is located in a gene desert within the *Doc1/2* locus and is characterized by high Zelda occupancy in early embryos (Supplementary Fig. 2a, b)[3]. We created an *MS2* reporter transgene with this putative enhancer and found that it drives expression in the dorsal ectoderm in a pattern overlapping endogenous *Doc1* mRNA expression (Supplementary Fig. 2c, d). Upon *zld*-RNAi, expression from this enhancer was strongly reduced, confirming a role for Zelda in driving expression from the *DocE* element (Supplementary Movie 7 and 8).

During nc13, our *DocE* transgene was stochastically activated in three distinct dorsal domains, in which we could track transcriptional mitotic memory (Supplementary Fig. 2c). None of the three domains (anterior, central, posterior) revealed a bias in the timing of activation in nc14 for descendants of active mothers compared to those arising from inactive mothers (Supplementary Fig. 2e–h). We therefore conclude that this *DocE* element does not trigger memory, despite being bound by Zelda. Thus, in a model system that shows stochastic activation and that uses Zelda, we did not observe any mitotic memory, further confirming the lack of involvement of Zelda in memory. Moreover these results show that experiencing transcription at a given cycle does not necessarily lead to a rapid *post*-mitotic reactivation in the following cycle.

Modeling the impact of Zelda on synchrony and memory. To gain insights into the role of Zelda on transcriptional dynamics, we developed a simple mathematical framework that describes the waiting times prior to the first detected transcriptional initiation event in nc 14, as a sequence of discrete transitions (see Methods and Supplementary Methods). This model considers that the activation of transcription follows a stepwise progression through a series of nonproductive OFF events that precede activation (ON event) (Fig. 4a). Given the various well characterized steps required prior to productive transcriptional elongation (e.g. promoter opening, transcription factor binding, pre-initiation complex recruitment…), it is reasonable to consider a series of transitions that a promoter must travel through prior to activation, with an allocated duration[11,24].

In this model, the average number of rate-limiting transitions is provided by the parameter *'a'* while the duration of the transitions by parameter *'b'* (expressed in seconds and considered the same for all transitions, a more complex model with heterogeneous *'b'* is discussed in the Supplementary Methods) (Fig. 4a). In the model, the distribution of waiting times prior to activation for a given gene in a set of nuclei can be described as a mixture of gamma distributions. Gamma distributions are frequently used in statistics for modeling waiting times. These distributions depend on two parameters, the shape parameter *'a'* and the scale parameter *'b'*. When, like in our model, the waiting time is the sum of a number of independent, exponentially

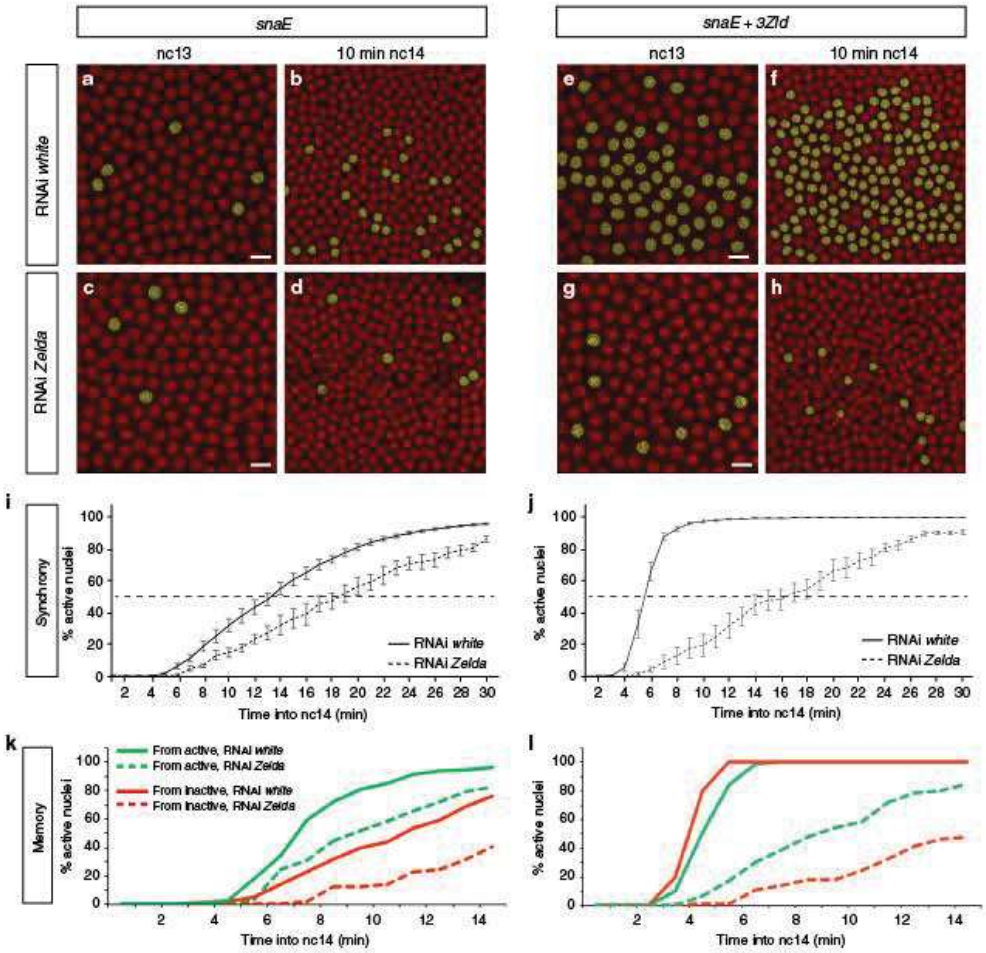


Fig. 3 Transcriptional memory persists in the absence of Zelda. **a–h** Snapshots of maximum intensity projected Z-stack of *Drosophila* embryo movies expressing *snaE* (**a–d**) or *snaE + 3Zld* (**e–h**) at nc13 (**a**, **c**, **e**, **g**) and nc14 (**b**, **d**, **f**, **h**), in a *white*-RNAi (**a**, **b**, **e**, **f**) or in a *zld*-RNAi genetic background (**c**, **d**, **g**, **h**). Nuclei exhibiting a spot of active transcription are false-colored in yellow. **i**, **j** Temporal coordination of all activated nuclei in nc14 for *snaE* (**i**) and *snaE + 3Zld* transgene (**j**), in a *white*-RNAi (solid curves) and in a *zld*-RNAi genetic background (dashed curves). Statistics: *snaE white*-RNAi: 9 movies, $n = 972$; *snaE zld*-RNAi: 4 movies $n = 156$; *snaE + 3Zld white*-RNAi: 6 movies $n = 541$; *snaE + 3Zld zld*-RNAi: 5 movies $n = 160$. Error bars represent SEM. **k**, **l** Kinetics of activation during the first 15 min of nc14 driven by the *snaE* transgene or the *snaE + 3Zld* transgene in *white*-RNAi (solid curves) or in *zld*-RNAi embryos (dashed curves). The kinetics of nuclei coming from transcriptionally active mothers in nc13 (green curves) is compared to those arising from inactive mothers (red curves). Statistics: *snaE white*-RNAi: 6 movies, $n = 302$; *snaE zld*-RNAi: 4 movies $n = 86$; *snaE + 3Zld white*-RNAi: 6 movies $n = 273$; *snaE + 3Zld zld*-RNAi: 5 movies $n = 111$). In Fig. 3, only nuclei located within a 25 µm rectangle centered around the ventral furrow/pseudo-furrow for *zld*-RNAi are considered. Scale bars represent 10 µm

distributed steps of equal mean duration, '*a*' is the number of transitions (steps) while '*b*' is the mean duration. Thus, '*a*' = *1* corresponds to the exponential distribution. A mixture of gamma distributions covers the case when the number of transitions (parameter '*a*') is random. The cumulative distribution function for the mixture is formulated in Fig. 4a and in the Supplementary Methods. Using the model, the average number of steps (the mean parameter '*a*' in the mixture) can be computed from the first two moments of the distribution of waiting times (Supplementary Equation 3 Supplementary Methods). This rough but direct estimate shows that '*a*' is less than three for all genotypes. Thus, we have restricted our model to three limiting transitions and consequently to three nonproductive states (OFF1, OFF2 and OFF3) (Fig. 4a).

Taking advantage of the significant number of nuclei tracked in this study, we could fit this model to our data and estimate more

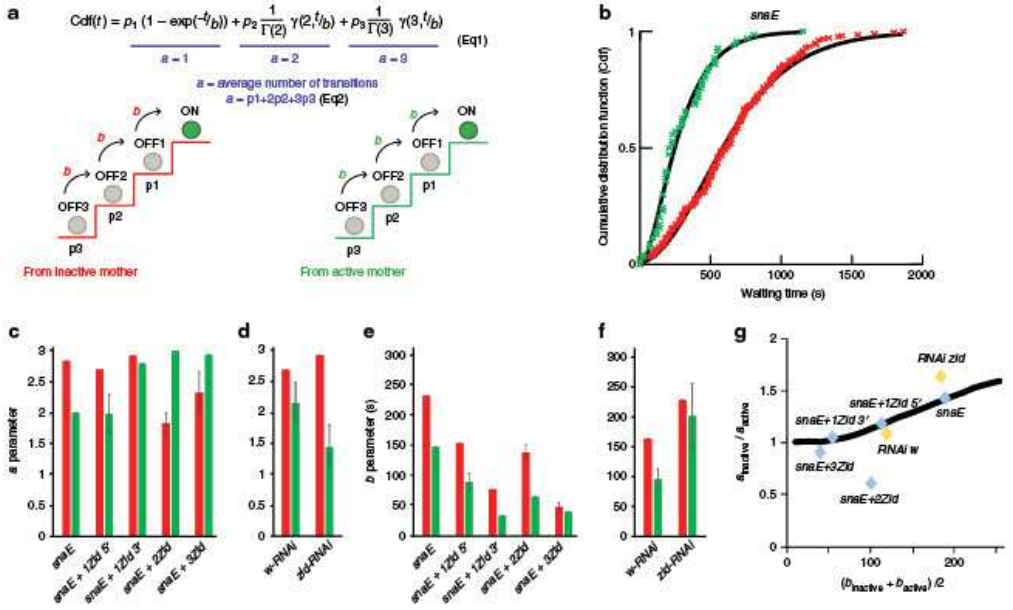


Fig. 4 Mathematical modeling of transcriptional memory. **a** Schematic representation of the discrete states (OFF1, OFF2 and OFF3) required to reach transcription (ON state), when coming from an inactive (red stairs) or an active (green stairs) mother nucleus. After mitosis, the probability for a nucleus to initiate the transcriptional activation process from a given state is provided by the parameter *p*, (p_1 = p(OFF1), p_2 = p(OFF2) and p_3 = p(OFF3)). The distribution of the random time to transcription (t) is modeled as a mixture of gamma distributions, according to the displayed Equation 1, where Cdf is the cumulative distribution function, Γ and γ are the complete and incomplete gamma functions, respectively. The average number of transitions '*a*' is provided by the sum of weighted probabilities (Equation 2). The time of each jump from one state to another is provided by the parameter '*b*'. Parameters of this model are predicted separately for nuclei coming from active and inactive mothers. The parameter fitting is based on Cdf of the random time to transcription in nc14 (t) (see also Supplementary Equation 1 Supplementary Methods). **b** Cdf of the random time to transcription in nc14. The origin of time is the end of mitosis nc13 to nc14, (determined by our automatic software, see "Methods" section) added to the time required to detect the first activation (proper to each genotype). Data (cross) are compared to predictions (solid curves) for nuclei, resulting from active (green) and inactive (red) mothers. **c**, **d** Mean '*a*' parameter; error intervals correspond to the variation among optimal and best suboptimal fits (see Supplementary Methods and Supplementary Data 1). **e**, **f** Mean '*b*' parameter; error intervals correspond to variation among optimal and best suboptimal fits (see Supplementary Methods and Supplementary Data 1). **g** The ratio of parameter '*a*' in subpopulations from inactive and active mothers is positively correlated (Spearman correlation coefficient = 0.82 $p = 0.034$) to the average '*b*' parameter. Predictions of the mathematical model are represented as a black line

accurately the parameters '*a*' and '*b*', from populations of nuclei (descendants of active and descendants of inactive nuclei) (Supplementary Data 1).

The distribution of waiting times for descendants of active nuclei and inactive nuclei were clearly distinct, indicating distinguishable scale and average shape parameters for these two populations (Fig. 4b). For *snaE*, the temporal behavior of descendants of active and inactive mothers differed in both parameters '*a*' and '*b*', with both being lower for descendants of active nuclei (Fig. 4c, e). Thus, our model suggests that descendants of active nuclei have fewer and shorter transitions prior to activation than descendants of inactive nuclei.

For transgenes with additional Zld-binding sites, the predictions for '*a*' are not obviously affected (Fig. 4c). This suggests that Zelda binding does not dramatically alter the number of transitions required for activation. By contrast, the estimates for '*b*' are systematically lower for descendants of active mothers and inactive mothers upon the addition of Zld-binding sites (Fig. 4e).

Thus, while the accelerated transcriptional activation observed with increased numbers of Zld-binding sites could be due to either a decrease in the duration of the transitions '*b*' or to the number of transitions '*a*', our modeling suggests that Zelda accelerates transcriptional activation primarily by speeding the lifetime of the transitions. Consistent with these findings, upon maternal reduction of Zelda, the '*b*' parameter is augmented, when compared to estimates in *white RNAi* controls (Fig. 4f).

In contrast to nuclei expressing the *snaE* transgene, the relation between estimates of '*a*' for descendants of active mothers and those of inactive mothers is more complex for extra-Zld transgenes (Fig. 4c, d). In order to gain some insights, we have used numerical simulations of an extended version of our model that includes modified transition dynamics during mitosis (Supplementary Fig. 4). In this version, we consider that at the beginning of mitosis, the states of active and inactive mother nuclei are OFF1 and OFF3, respectively. During mitosis, nuclei can undergo reversible (upward and backward) transitions. After mitosis, the resulting daughter nuclei follow the irreversible transition scheme represented in Fig. 4a. The simulations predict that the bias in '*a*' values (difference in the number of steps to reach active state, evaluated by $a_{inactive}$/ a_{active} and referred to as

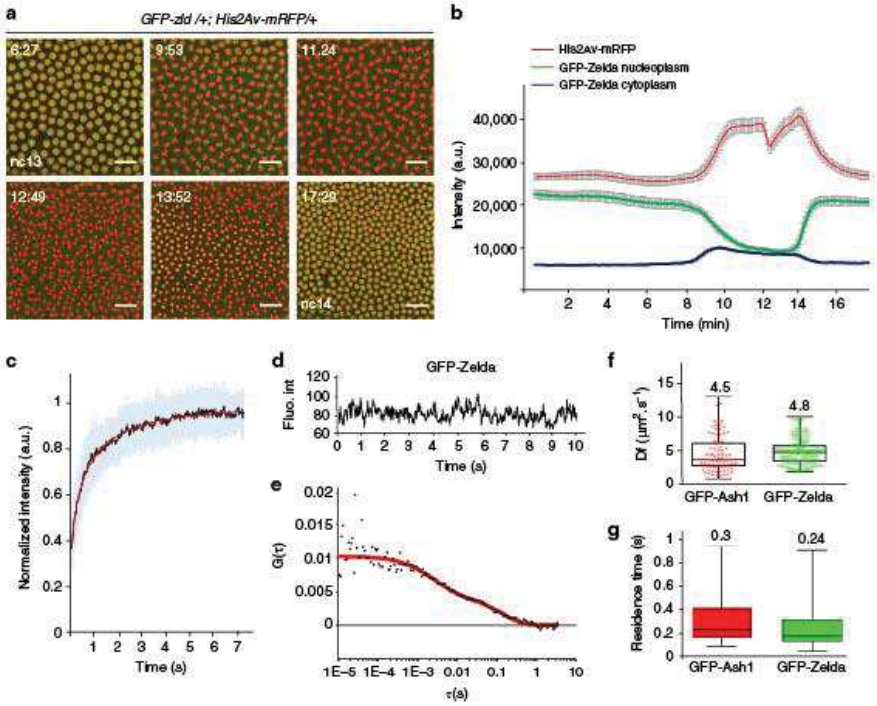


Fig. 5 Zelda global kinetic properties. **a** Living *GFP-zld/ + ;His2Av-mRFP/ +* embryo imaged by confocal microscopy from interphase of nc13 to early interphase of nc14. Successive representative maximum intensity projected Z-stack images are shown at the indicated timings (in minutes) (see also Supplementary Movie 9). Scale bars represent 20 µm. **b** Average intensity profiles for histones (red), nucleoplasmic Zelda (green) and cytoplasmic Zelda (blue) measured from a nc13 embryo transitioning into nc14. An automatic tracking of fluorescence from a minimum of 114 nuclei and 10 cytoplasmic ROI generated these profiles. Error bars represent SD. **c** FRAP mean curve (black) and the mean of all the fits (red curve) using reaction-diffusion models determined at the bleached spot for 25 nuclei from nc14 developing *GFP-zld* embryos. Error bars represent SD from different nuclei (light blue bars). **d** Example of a time trace obtained by FCS from a *GFP-zld* nc14 embryo showing no bleaching. **e** Example of autocorrelation function (black dots) related to **d** (red curve represents fitting using a reaction-diffusion model). **f–g** Kinetic parameters for Zelda and Ash1 extracted after fitting FCS data with a reaction-diffusion model. **f** Box plot representing estimated diffusion coefficients (D_f). Numbers above each plot represent the mean. Centered lines represent the median and whiskers represent min and max. **g** Box plot representing estimated residence times ($1/k_{off}$). Numbers above each plot represent the mean. Centered lines represent the median and whiskers represent min and max

memory bias) is correlated to the '*b*' values (Supplementary Methods). Remarkably, we could obtain this correlation with our temporal data (Fig. 4g). Hence, enhancers with large '*b*' values (having slow transitions) tend to have a strong $a_{inactive}/a_{active}$ ratio, whereas enhancers with small '*b*' values have a vanishing memory bias (Fig. 4g and Supplementary Methods). If we interpret transcriptional memory as the heritable reduction in the number of transitions required to reach the active state, this result suggests that stronger memory occurs when transitions are slower.

In summary, using a modeling approach, we converted our measurements of the timing of transcriptional activation in hundreds of single nuclei into discrete metastable states preceding gene activation. Although purely phenomenological, this model suggests that transcriptional memory is supported by a sequence of relatively long-lived metastable states preceding activation, which could be maintained by mitotic bookmarking. Zelda potentiates transcription by accelerating these states and is thus incompatible with mitotic memory, consistent with our observations in Zelda depleted embryos.

Zelda dynamic behavior. Based on the characteristics of pioneer factors, we had expected a role for Zelda in retaining transcriptional memory through mitosis. However, our genetic data and modeling indicate that Zelda was not the basis of memory. To better understand this result, we examined Zelda dynamics in living embryos particularly during mitosis (Fig. 5a, b and Supplementary Movie 9). We took advantage of an endogenously tagged fluorescent version of Zelda[25]. The *GFP-zld* flies are homozygous viable, thereby showing that the GFP-tagged version of Zelda retains wild-type physiological properties.

GFP-Zld localization with chromatin was examined through the mitotic cycle (Fig. 5, Supplementary Fig. 5 and Supplementary Movie 9) and demonstrated that Zelda is not detectably retained on the chromosomes during mitosis (Fig. 5a, b and Supplementary Movie 9), similar to what was shown in fixed embryos[26]. This is in comparison to the transcription factor GAF (GAGA Associated Factor) (Supplementary Fig. 5a). While most GFP-Zld molecules are localized in nuclei in interphase, they were redistributed to the cytoplasm during mitosis (Fig. 5b). Our live imaging approach revealed a highly dynamic behavior of Zelda,

whereby Zelda comes back to the nucleus rapidly at the end of mitosis prior to complete chromosome decompaction (Fig. 5a, b, Supplementary Fig. 5b and Supplementary Movie 9). When compared to Pol II-Ser5P by immunostaining on embryos exhibiting a mitotic wave, GFP-Zld signal filled the nuclei prior to polymerase (Supplementary Fig. 5b). By comparing Zelda cell cycle dynamics to Bicoid[27], we found that these two DNA binding proteins are similarly evicted during mitosis and are quickly re-localized to chromatin at the end of mitosis (Supplementary Fig. 5c and Supplementary Movie 10). Nonetheless, quantitative comparisons revealed that Zelda entry into the nucleus at the end of mitosis was faster than that of Bicoid (Supplementary Fig. 5d).

To better characterize Zelda kinetics, we performed intranuclear Fluorescence Recovery After Photobleaching (FRAP) experiments and determined that Zelda recovery was remarkably fast in interphase nuclei (Fig. 5c). Because FRAP is not well suited for fast moving proteins, we performed Fluorescence Correlation Spectroscopy (FCS) in living cycle14 embryos (Fig. 5d, e) and estimated the kinetic properties of Zelda (Fig. 5f, g). FCS experiments for Zelda were compared to Ash1, whose kinetics have been documented by both FRAP and FCS in early *Drosophila* embryos[28] and thus serves as a well-defined point of reference (Supplementary Fig. 5e–g). As expected for chromatin-binding proteins, auto-correlation curves for Zelda and Ash1 show more than one unique characteristic time (Fig. 5e and Supplementary Fig. 5f). To estimate chromatin-binding kinetics, FCS curves (86 curves for Ash1 and 109 curves for Zelda) were fitted with a reaction-diffusion model[29]. As previously revealed in Steffen et al.[28], we found that Ash1 is highly dynamic and estimate similar diffusion coefficients ($D_f \approx 4.5 \mu m^2 \cdot s^{-1}$ versus $D_{f(Steffen)} \approx 4.98\ \mu m^2 \cdot s^{-1}$) (Fig. 5f), thus validating our FCS experiments. Similarly to other transcription factors (e.g. Bicoid) we found that Zelda is a fast diffusing protein with apparent diffusion coefficient of $\approx 4.8\ \mu m^2 \cdot s^{-1}$ (Fig. 5f).

We also extracted the dissociation rate (k_{off}) (see Methods) and deduced the estimated residence time (RT) ($RT = 1/k_{off}$) (Fig. 5g). Our FCS results reveal that Zelda binding to chromatin is highly transient, on the order of hundreds of milliseconds (RT = 0.24 s). However, we note that this FCS-estimated RT likely represents only a lower-bound limit for Zelda binding, whereas a FRAP approach allows for upper-bound estimations (see next paragraph).

The low binding and fast diffusion rates for Zelda were unexpected given the large impact of a single extra Zld-binding site on accelerating the kinetics of gene activation (Fig. 1). However, similar transient chromatin bindings have been reported recently for other key transcription factors[8,30]. These transient bindings events seem to be compensated for by increased local concentrations, in particular nuclear microenvironments, referred to as subnuclear hubs[8]. We thus explored Zelda spatial distribution within nuclei in living *Drosophila* embryos.

Zelda accumulates into subnuclear hubs. Live imaging with GFP-Zld, revealed a heterogeneous distribution of Zelda proteins in the nucleus (Fig. 6a–d and Supplementary Movie 11–13). Thus, similarly to Bicoid[8], multiple Zelda proteins seem to cluster in local dynamic hubs that exhibit a diversity of sizes and lifetimes (Fig. 6a–b and Supplementary Movie 11–13). For example, Zelda large hubs are more visible at early cycles and at the beginning of interphases (Fig. 6a and Supplementary Movie 11, 13). These local inhomogeneities are unlikely GFP aggregate as we could observe them with a two-colored Zelda tagging approach (Fig. 6c). Consistent with this, Zelda hubs have been observed very recently with other methods[31].

These findings are consistent with a "phase separation" model of transcriptional regulation[32], which could promote Zelda cooperativity with other major transcriptional regulators, like Bicoid[8], which would then foster rapid and coordinated transcriptional initiation.

To decipher the link between Zelda hubs and transcription, we examined the spatial localization of nascent transcription with respect to Zelda hub positions.

Immuno-FISH and live imaging experiments revealed that only a subset of our MS2 reporter nascent RNAs were located within a Zelda dense region (Fig. 6d, Supplementary Fig. 6a, and Supplementary Movie 13). Furthermore, only limited co-localization was observed between dots of elongating Pol II (Ser2-P Pol II) and Zelda hubs (Supplementary Fig. 6b). Similar findings have been recently reported by following transcription from a *hunchback < MS2* transgene (*hb*)[31]. We therefore conclude that, at least with a synthetic transgene, long-lasting stable contacts between sites of transcription and Zelda dense regions are not detected. Given the important number of Zelda targets, rigorously connecting transcriptional activation to Zelda hubs would require a broader analysis with adapted imaging methods. However, these observations are reminiscent of a dynamic kissing model recently described between an active gene and mediator clusters in ES cells[33].

Having demonstrated that Zelda was not homogeneously distributed, we wanted to determine if the dynamic properties of Zelda within hubs differed from its global dynamics.

One possibility was that high local concentrations of Zelda would correspond to long-lasting chromatin-binding events that would result in increased residence time (long RT). Given the rapid movements of Zelda hubs (Supplementary Movie 11–14), capturing these long-lasting events by FCS was not feasible. Therefore, we performed FRAP experiments on long-lasting Zelda hubs (Fig. 6e and Supplementary Movie 14). We demonstrate that Zelda recovered as rapidly within long-lasting hubs as it did in bulk nuclei (Fig. 6f and Fig. 5c). By fitting our FRAP data with a diffusion-reaction model, we estimate a residence time on the order of seconds in both cases (hubs and global FRAP) (Fig. 6g). Fitting the FRAP experiments with alternative models, such as one binding reaction only (Supplementary Fig. 6c–e) or two distinct binding reactions (slow and fast) (Supplementary Fig. 6f–h) similarly supported an estimated residence time for Zelda on the order of seconds and that this is independent of its local concentration (Supplementary Fig. 6i). Of note, the two reactions model estimates two residence times, where the fastest estimate is similar to the FCS-extracted residence time (Fig. 5g and (Supplementary Fig. 6i)). Previous studies on Bicoid have also identified two populations: one with a short characteristic time (RT on the order hundreds of milliseconds), described as non-specific binding events and a population with longer characteristic times (RT on the order of seconds) corresponding to specific binding events[8,34].

From the combined data, we conclude that Zelda chromatin binding is transient, ranging from hundreds of milliseconds (likely unspecific binding) to seconds (likely specific binding) and this occurs regardless of its local concentration. How could Zelda exhibit a similarly transient chromatin binding at random positions and in regions where its concentration is high? An exciting possibility would be that dense regions could be enriched for Zelda target genes. Under this hypothesis, Zelda target genes could cluster together thereby increasing the probability to locally accumulate Zelda proteins. Alternatively, Zelda could spatially organize its targets in space through its pioneer factor properties and then maintain these clusters of targets possibly through its intrinsically disordered domains[35]. The latter scenario is

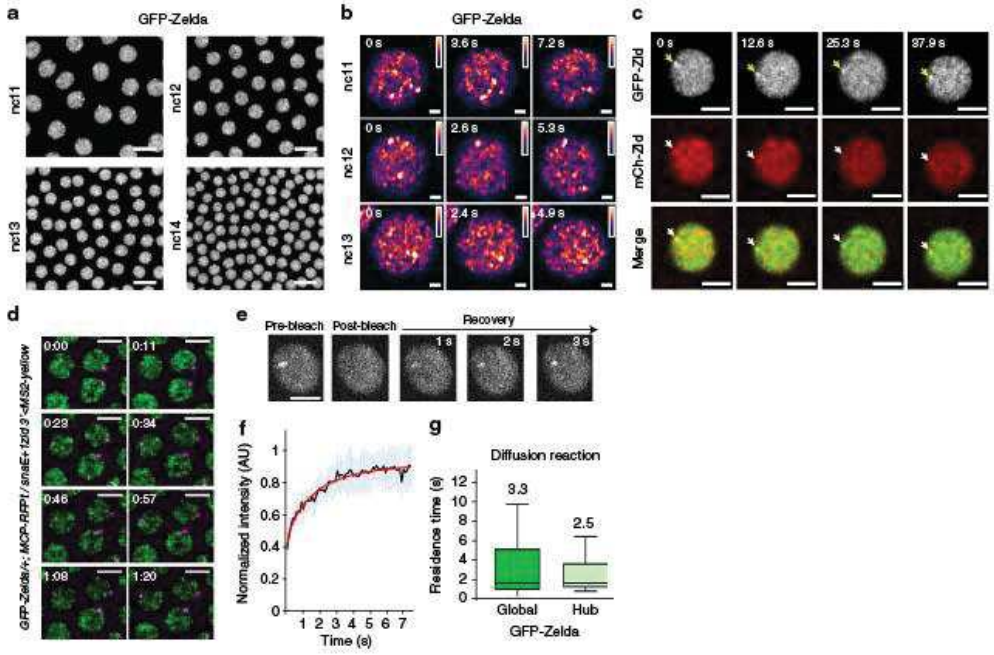


Fig. 6 Zelda hubs kinetic properties. **a** Living *GFP-zld* embryo imaged by confocal microscopy from interphase of nc11 to early interphase of nc14. Successive representative maximum intensity projected Z-stack images showing local inhomogeneities are shown at each cycle (see also Supplementary Movie 11). Scale bars represent 10 µm. **b** Nuclei zoom from living *GFP-zld* embryos imaged by confocal microscopy from interphase of nc11 to early interphase of nc13. Successive representative images showing local hubs are shown at each cycle. Scale bars represent 1 µm. **c** Nuclei zoom from living *GFP-zld/mCherry-Zld* embryos imaged by confocal microscopy at interphase of early nc11. Successive representative images showing Zelda subnuclear hubs are shown at each cycle (an example is indicated by an arrow). Scale bars represent 5µm. **d** Living *GFP-Zld/+;MCP-RFPt/+* embryo containing the *snaE + 1Zld3'* transgene (MS2 in magenta) imaged by confocal microscopy. Successive representative maximum intensity projected Z-stacks of slices containing the MS2 signal only (see also Supplementary Movie 13). Scale bars represent 5 µm. **e** Example of a FRAP experiment on a single hub. **f** FRAP mean curve (black) and the mean of all the fits (red curve) using diffusion reaction models determined at the bleached spot for 11 nuclei from nc10–13 developing *GFP-zld* embryos. Error bars represent SD from different nuclei (light blue bars). Scale bars represent 5 µm. **g** Box plot representing estimated residence times ($1/k_{off}$) extracted with diffusion reaction model of GFP-Zld in Global (dark green) compared to Hubs (light green). Numbers above each plot represent the mean RT. Centered lines represent the median and whiskers represent min and max

consistent with the finding that Zelda participates in the *Drosophila* 3D chromatin organization during the awakening of the zygotic genome[36,37].

Using quantitative imaging approaches in living embryos combined with mathematical modeling, we have investigated a role for the pioneer-like factor Zelda in regulating mitotic memory. We demonstrated that Zelda is dispensable for transcriptional mitotic memory. While our modeling suggests that memory may be potentiated by a reduction in the number of steps required for post-mitotic transcriptional activation, we propose that Zelda accelerates this activation by decreasing the time spent at each preceding step. Our data support a model whereby Zelda binds transiently to chromatin in localized nuclear microenvironments; to potentially accelerate the timing of the transitions required prior to transcriptional activation (e.g. local chromatin organization, recruitment of transcription factors, recruitment of Pol II and general transcription factors). These dynamic properties allow the pioneer-like factor Zelda to act as a quantitative timer for fine-tuning transcriptional activation during early *Drosophila* development.

Methods

***Drosophila* stocks and genetics**. The *yw* stock was used as a control. The germline driver *nos-Gal4:VP16* (BL4937) was recombined with an *MCP-eGFP-His2Av-mRFP* fly line (gift from T. Fukaya). RNAi were expressed following crosses between this recombinant and *UASp-shRNA-w* (BL35573) or *UASp-shRNA-zld*[4], virgin females expressing RNAi, *MCP-GFP-His2Av-mRFP* were crossed with *MS2* containing transgenic constructs males. The *GFP-zld and mCherry-zld* strains were obtained by CRISPR[25]. *GFP-ashf*[28] and *His2Av-mRFP* (BL23651) and *eGFP-bcd*[27].

Cloning and transgenesis. The *snaE* enhancer transgene was described in Ferraro et al[14]. The 24XMS2 tag was inserted immediately upstream of the yellow reporter gene coding sequence. We verified that our MS2 sequences do not contain any canonical binding site and we also compared synchrony of the *snaE* transgenes with our 24XMS2 to newly available MS2 loops, 24XMS2ΔZelda[38], kindly provided by Pr. Dostatni (Supplementary Fig. 7). Extra Zld-binding sites (consensus sequences CAGGTAG) were added to the *wt snaE* enhancer by PCR using the primers indicated in Supplementary Data 3. Insertion of an extra Zld-binding site in the middle of the *snaE* enhancer was done by directed mutagenesis using primers indicated in Supplementary Data 3. All transgenic MS2 flies were inserted in the same landing site (BL9750 line) using PhiC31 targeted insertion[39].

Live imaging. Embryos were dechorionated with tape and mounted between a hydrophobic membrane and a coverslip as described in[14]. All movies (except when specified) were acquired using a Zeiss LSM780 confocal microscope with the

following settings: GFP and RFP proteins were excited using a 488 nm and a 561 nm laser respectively. A GaAsP detector was used to detect the GFP fluorescence; a 40x oil objective and a 2.1 zoom on central ventral region of the embryo, 512 × 512 pixels, 16 bits/pixel, with monodirectional scanning and 21z stacks 0.5µm apart. Under these conditions, the time resolution is in the range of 22–25 s per frame. Two and one movie respectively for *snaE + 3Zld* in *white*-RNAi and *zld*-RNAi background were acquired on a Zeiss LSM880, keeping the time resolution is in the range of 21–22 s per frame. Image processing of LSM780 and LSM880 movies of the MS2-MCP-GFP signal were performed in a semi-automatic way using custom made software, developed in Python™. Live imaging of *GFP-zld/+; His2Av-mRFP/+* and *GFP-bal/+; His2Av-mRFP/+* embryos were performed using a Zeiss LSM780 confocal microscope with the following settings: GFP and RFP proteins were excited using a 488 nm and a 561 nm laser respectively. A GaAsP detector was used to detect the GFP fluorescence. 512 × 512 pixels, 16 bits/pixel, images were acquired using a 40x oil objective, with bidirectional scanning and 8 stacks 1µm apart. Under these conditions, the time resolution is in the range of 5–7 s per frame. Live imaging of *GFP-zld* embryos (related to Fig. 6b) was acquired using a Zeiss LSM880 confocal microscope with the following settings: GFP protein was excited using a 488 nm laser. A GaAsP-PMT array of an Airyscan detector was used to detect the GFP fluorescence. 568 × 568 pixels, 8 bits/pixel images were acquired using a 40x oil objective, 13–14z stacks 0.5µm apart, with a 4x zoom and bi-directional scanning, with fast Airyscan in optimal mode. Under these conditions and using a piezo Z, the time resolution is in the range of 2–4 s per frame.

Live imaging of *GFP-zld/mCherry-zld* embryos (related to Fig. 6c) was acquired using a Zeiss LSM880 confocal microscope with the following settings: GFP and mCherry proteins were excited using a 488 nm and a 561 nm laser respectively. A GaAsP detector was used to detect the GFP fluorescence and a GaAsP-PMT array of an Airyscan detector was used to detect mCherry. 512 × 512 pixels, 16 bits/pixel images were acquired using a 40x oil objective, 5z stacks 1µm apart, with a 3.6x zoom and bidirectional scanning. Under these conditions, the time resolution is in the range of 12–13 s per frame. Live imaging of *GFP-zld, MCP-RFPt* (stock details available upon request) embryos containing *snaE + 1Zld3'* (related to Supplementary Movie 13) was acquired using a Zeiss LSM880 confocal microscope with the following settings: GFP and MCP-RFPt proteins were excited using a 488 nm and a 561 nm laser respectively. A GaAsP detector was used to detect the GFP fluorescence. 512 × 512 pixels, 16 bits/pixel, images were acquired using a 40x oil objective, 12z stacks 0.5µm apart, with a 4x zoom. Under these conditions, the time resolution is in the range of 11–12 s per frame. Live imaging of *GFP-zld* embryos (related to Supplementary Movie 12) was acquired using a Zeiss LSM880 using a 40x /1.3 Oil objective. Images (256 × 256 pixels), 12bits/pixel, 5z stacks 1µm apart, zoom 8x, were acquired every ≈550 ms during 100 frames. GFP was excited with an Argon laser at 488 nm and detected between 496–565 nm.

Image analysis for transcriptional activation. Briefly, red (nuclei) and green (MS2 nascent RNA spots) channels were first maximum intensity projected. The analysis was divided into three parts: before mitosis (nc13), during mitosis and after mitosis (nc14). During interphases, nuclei were segmented (using mainly circularity arguments and water-shed algorithm) and tracked with a minimal distance criterion. Chromosomal splitting at the end of anaphase was automatically detected and used to define a nucleus-by-nucleus mitosis frame. During mitosis, nuclei were segmented by intensities and tracked by an overlap in consecutive frames criterion. By merging these three parts, we could give a label to each nc13 mother nucleus, their daughters in nc14 and an extra label to recognize one daughter from the other. MS2 spots were detected with a blob detection method with a user-defined threshold constant value. Detected spots were associated to the closest nuclei, inheriting their label. Finally, for each tracked nucleus, the timing of first transcriptional activation was recorded. Analyses were done with the activation timing of the two daughters for synchrony and the first active daughter for memory.

Manual tracking. Upon high reduction of maternal Zelda, zygotic activation is perturbed, and major developmental defects occur. Abnormal nuclear shape (example Supplementary Movie 6 and control Supplementary Movie 5) precludes their automatic segmentation. Kinetics of transcriptional activation was manually analyzed for *zld*-RNAi embryos. We implemented a spot detection algorithm to export files with the detected MS2 spot. The thresholding of the MS2 spot is consistent with all the other automatic analyses. A spatial domain was determined, taking the pseudo-furrow as a landmark and focusing on a 25µm region around it (note that activation does not seem to occur outside of this area). Nuclei in this region are visually tracked frame-by-frame to detect when activation in nc13 occurs, when mitosis occurs and to recover the first activation frame in nc14. Approximately 2 to 3-fold more nuclei from inactive mothers, compared to descendants of active mothers, were analyzed.

Immunostaining and RNA in situ hybridization. Embryos were dechorionated with bleach for 1–2 min and thoroughly rinsed with H_2O. They were fixed in fixation buffer (500 µl EGTA 0.5 M, 500 µl PBS 10x, 4 ml Formaldehyde MeOH free 10%, 5 ml Heptane) for 25 min on a shaker at 450 rpm; formaldehyde was replaced by 5 ml methanol and embryos were vortexed for 30 s. Embryos that sank to the bottom of the tube were rinsed three times with methanol. For immunostaining, embryos were rinsed with methanol and washed three times with PBT (PBS 1x 0.1% triton). Embryos were incubated on a wheel at room temperature twice for 30 min in PBT, once for 20 min in PBT 1% BSA, and at 4 °C overnight in PBT 1% BSA with primary antibodies. Embryos were rinsed three times, washed twice for 30 min in PBT, then incubated in PBT 1% BSA for 30 min, and in PBT 1% BSA with secondary antibodies for 2 h at room temperature. Embryos were rinsed three times then washed three times in PBT for 10 min. DNA staining was performed using DAPI at 0.5 µg.ml^{-1}. Primary antibody dilutions for immunostaining were mouse anti-GFP (Roche IgG1κ clones 7.1 and 13.1) 1:200; rabbit anti-GFP (Life technologies A11122) 1:100; mouse anti-Ser5-P Pol II (Covance H14, MMS-134R) 1:100; rat anti-Ser2-P Pol II (3E10) 1:5; rabbit anti-GAF (a gift from Dr. G Cavalli) 1:250; Secondary antibodies (anti-rabbit Alexa 488-conjugated (Life technologies, A21206); anti-mouse Alexa 488-conjugated (Life technologies, A21202); anti-mouse IgM Alexa 555-conjugated (Life technologies, A21426); anti-rabbit Alexa 555-conjugated (Life technologies, A31572); were used at a dilution 1:500. Fluorescent in situ hybridization was performed as described in[14]. A Dixogygenin-MS2 probe was obtained by in vitro transcription from a bluescript plasmid containing the 24-MS2 sequences, isolated with BamH1/BglII enzymes from the original Addgene MS2 plasmid (# 31865). *Snail* probe was described in[19]. Primary and Secondary antibody for Fluorescent in situ hybridization were sheep anti-Digoxigenin (Roche 11333089001) 1/375; mouse anti-Biotin (Life technologies, 03–3700) 1/375; anti-mouse Alexa 488-conjugated (Life technologies, A21202) and anti-sheep Alexa 555-conjugated (Life technologies, A21436) 1:500. Mounting was performed in Prolong Gold. Fluorescent in situ hybridization with *yellow* probes was performed as described in[40], probe sequences will be available upon request.

Fluorescence recovery after photobleaching. Fluorescence recovery after photobleaching (FRAP) in embryos at nc14 (Global) was performed with the following settings: a Zeiss LSM780 using a 40x/1.3 Oil objective and a pinhole of 66 µm. Images (512 × 32 pixels), 16bits/pixel, zoom 6x, were acquired every ≈20 ms during 400 frames. GFP was excited with an Argon laser at 488 nm and detected between 507–596 nm. Laser intensity was kept as low as possible to minimize unintentional photobleaching. A circular ROI (16 × 16 pixels) 0.07 µm/pixels, was bleached using two laser pulses at maximal power during a total of ≈130 ms after 50 frames. Hub FRAP was performed in embryos from nc10 to nc13 with the following settings: FRAP was performed on a Zeiss LSM880 using a 40 × /1.3 Oil objective and a pinhole of 83 µm. Images (256 × 256 pixels), 12bits/pixel, zoom 8x, were acquired every ≈104 ms during 100 frames. GFP was excited with an Argon laser at 488 nm and detected between 496–565 nm. Laser intensity was kept as low as possible to minimize unintentional photobleaching. A circular ROI (8 × 8 pixels) 0.1 µm/pixel, was bleached using four laser pulses at maximal power during a total of ≈114 ms after 3 frames. To discard any source of fluorescence intensity fluctuation other than molecular diffusion, the measured fluorescence recovery in the bleached ROI region (I_{bl}) was corrected by an unbleached ROI (I_{unbl}) of the same nucleus and another ROI outside of the nucleus (I_{out}) following the simple equation:

$$I_{bl_{corr}}(t) = \frac{I_{bl}(t) - I_{out}(t)}{I_{unbl}(t) - I_{out}(t)}. \quad (1)$$

The obtained fluorescence recovery was then normalized to the mean value of fluorescence before the bleaching i.e.,:

$$I_{bl_{norm}}(t) = \frac{I_{bl_{corr}}(t)}{\frac{1}{N}\sum_{n=1}^{-50} I_{bl}(n)}. \quad (2)$$

Fitting FRAP recovery curves. Three different analytical equations were used to fit the fluorescence recovery depending on the model of dynamics chosen.

For the pure reaction kinetics model, we started from the analytical expression of Sprague et al.[41] and decline it for the case of two exchanging populations, leading to the two following equations:

one exchanging population

$$F(t) = 1 - C_{eq} e^{-k_{off} t} \quad (3)$$

two exchanging populations

$$F(t) = 1 - C_{eq_1} e^{-k_{off_1} t} - C_{eq_2} e^{-k_{off_2} t} \quad (4)$$

With $C_{eq} = k^*_{on} / (k_{off} + k^*_{on})$ as in their case.

Finally, for the reaction-diffusion model, we started from the analytical expression developed in the supplemental (Supplementary Equation 35) of[29].

$$F(t) = F_{eq} F_D(t) + C_{eq} F_{exc}(t) \quad (5)$$

with C_{eq} defined as above and $F_{eq} = k_{off} / (k_{off} + k^*_{on})$. $F_D(t)$ is the fluorescence recovery due to diffusion and $F_{exc}(t)$ the fluorescence recovery due to exchange.

Since we used a Gaussian shape illumination profile, $F_D(t)$ is defined using a slightly modified version of the analytical equation of the 20th order limited development of the Axelrod model for Gaussian profile illumination and diffusion[42,43]:

$$F_D(t) = \frac{1-e^{-K}}{K}(1-M) + M\sum_{n=1}^{20}\frac{(-K)^n}{n!}\left(1+n+2n\frac{t}{\tau}\right)^{-1} \quad (6)$$

$$M = \frac{I_{(r>30r)} - I_0}{1-I_0}, \quad (7)$$

where K is a constant proportional to bleaching deepness, M is the mobile fraction and τ is the half time of recovery. To minimize the effect of mobile fraction on C_{eq}, M was kept between 0.9 and 1.1.

Diffusion coefficients of the different molecules were determined according to

$$D = \frac{\beta w^2}{4\tau} \quad (8)$$

with w the value of the radius at $1/e^2$ of the Gaussian beam (in our case, w = 0.56 µm or 0.4 µm) and β a discrete function of K tabulated in ref. [44].

$F_{exc}(t)$ is defined as in ref. [29], slightly modified with respect to the Gaussian illumination, leading to the following equation:

$$F_{exc}(t) = F_\infty - \left(\frac{1-e^{-K}}{K} - F_\infty\right)e^{-k_{off}t} \quad (9)$$

with K defined as previously.

Fluorescence correlation spectroscopy. FCS experiments were performed on a Zeiss LSM780 microscope using a 40x/1.2 water objective. GFP was excited using the 488 nm line of an Argon laser with a pinhole of 1 airy unit. Intensity fluctuation measured for 10 s were acquired and auto-correlation functions (ACFs) generated by Zen software were loaded in the PyCorrFit program[45]. Multiple measurements per nucleus, in multiple nuclei and embryos at 20 °C were used to generate multiple ACF, used to extract kinetic parameters. The FCS measurement volume was calibrated with a Rhodamine6G solution[46] using $D_f = 414\,\mu m^2.s^{-1}$. Each time series was fitted with the reaction dominant model[29]:

$$G_D(\tau) = \frac{1}{2^{3/2}N}F_{eq}\left(1+\frac{\tau}{\tau_{D_f}}\right)^{-1}\left(1+\frac{\tau}{\omega^2\tau_{D_f}}\right)^{-1/2} + \frac{1}{2^{3/2}N}C_{eq}e^{-k_{off}\tau} + G_\infty, \quad (10)$$

where $F_{eq} = k_{off}/(k_{off}+k^*_{on})$, $C_{eq} = k^*_{on}/(k_{off}+k^*_{on})$, $\tau_{Df} = w^2_{xy}/4\,D_f$ and $\omega = w_z/w_{xy}$.

For each correlogram the value of k^*_{on}, k_{off} and D_f (reaction constants and diffusion coefficient) were estimated using the PyCorrFit program implemented with Eq. 10.

Photophysics of the dye was neglected and we introduced a G_∞ value to account for long time persistent correlation during the measurements. Based on the average number of Zelda molecules (N), connected to the $G(0)$ value following the formula $G(0)=\frac{1}{2^{3/2}N}$ and on the illumination volume, we found that Zelda concentration is around $410^{+/-90}$ nM.

Mathematical model, general framework. We are interested in the time needed for post-mitotic transcription (re)activation. We model this time as the sum of two variables:

$$T_a = T_0 + T_r, \quad (11)$$

where T_0 is a deterministic incompressible lag time, the same for all nuclei, and T_r is a random variable whose value fluctuates from one nucleus to another. The decomposition in Eq. 11 can be justified by the experimental observation that all the reactivation curves (Fig. 1l) start with a nonzero length interval during which no nuclei are activated. Furthermore, T_r is defined such that it takes values close to zero with nonzero probability. This property allows us to set the time origin to the instant when the first nuclei initiates transcription, in order to determine T_r. For large populations of nuclei, this stems to setting the time origin to T_0 (see Supplementary Methods).

The stochastic part of the (re)activation time is modeled using a finite state, continuous time, Markov chain.

The states of the process are $A_1, A_2, \ldots, A_{n-1}, A_n$. The states A_i, $1 \le i \le n-1$ are metastable and OFF, i.e. not transcribing. The state A_n is ON, i.e. transcribing. Each metastable state has a given lifetime, defined as the waiting time before leaving the state and going elsewhere. For the purposes of this paper, we considered that each of the states has the same lifetime denoted τ. Also, the topology of the transitions is considered linear: in order to go to A_{i+1} one has to visit A_i.

We have also tested a more complex model, with uneven lifetimes and therefore, more parameters. However, the complex model did not provide a sensibly better fit with data. Furthermore, the determination of individual parameters of the more complex model is uncertain: many different combinations of parameter values lead to very close fit. Following Occam's razor principle, we based our analysis on the simplest, homogeneous jump model. For consistency, detailed models are described in Supplementary Methods.

Code availability. The image analysis software and associated graphical user interface developed for this study are available from the corresponding author upon request.

Data availability

All relevant data supporting the key findings of this study are available within the article and its Supplementary Information files or from the corresponding author upon reasonable request.

Received: 8 March 2018 Accepted: 9 November 2018
Published online: 05 December 2018

References

1. Liang, H.-L. et al. The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. *Nature* 456, 400–403 (2008).
2. Nien, C.-Y. et al. Temporal coordination of gene networks by Zelda in the early Drosophila embryo. *PLoS Genet.* 7, e1002339 (2011).
3. Harrison, M. M., Li, X.-Y., Kaplan, T., Botchan, M. R. & Eisen, M. B. Zelda binding in the early Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-zygotic transition. *PLoS Genet.* 7, e1002266 (2011).
4. Sun, Y. et al. Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophilazygotic genome activation. *Genome Res.* 25, 1703–1714 (2015).
5. Schulz, K. N. et al. Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early Drosophila embryo. *Genome Res.* 25, 1715–1726 (2015).
6. Xu, Z. et al. Impacts of the ubiquitous factor Zelda on Bicoid-dependent DNA binding and transcription in Drosophila. *Genes Dev.* 28, 608–621 (2014).
7. Foo, S. M. *et al.* Zelda potentiates morphogen activity by increasing chromatin accessibility. 24, 1341–1346 (2014).
8. Mir, M. et al. Dense Bicoid hubs accentuate binding along the morphogen gradient. *Genes Dev.* 31, 1784–1794 (2017).
9. Crocker, J., Tsai, A. & Stern, D. L. A fully synthetic transcriptional platform for a multicellular eukaryote. *Cell Rep.* 18, 287–296 (2017).
10. Caravaca, J. M. et al. Bookmarking by specific and nonspecific binding of FoxA1 pioneer factor to mitotic chromosomes. *Genes Dev.* 27, 251–260 (2013).
11. Bellec, M., Radulescu, O. & Lagha, M. Remembering the past: Mitotic bookmarking in a developing embryo. *Curr. Opin. Syst. Biol.* 11, 41–49 (2018).
12. Perry, M. W., Boettiger, A. N., Bothma, J. P. & Levine, M. Shadow enhancers foster robustness of Drosophila gastrulation. *Curr. Biol.* 20, 1562–1567 (2010).
13. Dunipace, L., Ozdemir, A. & Stathopoulos, A. Complex interactions between cis-regulatory modules in native conformation are critical for Drosophila snail expression. *Development* 138, 4566–4566 (2011).
14. Ferraro, T. et al. Transcriptional memory in the Drosophila embryo. *Curr. Biol.* 26, 212–218 (2016).
15. Pichon, X., Lagha, M., Mueller, F. & Bertrand, E. A growing toolbox to image gene expression in single cells: sensitive approaches for demanding challenges. *Mol. Cell* 71, 468–480 (2018).
16. Bosch, J. R. T. The TAGteam DNA motif controls the timing of Drosophila pre-blastoderm transcription. *Development* 133, 1967–1977 (2006).
17. Sandler, J. E. & Stathopoulos, A. Stepwise progression of embryonic patterning. *Trends Genet.* 32, 432–443 (2016).
18. Ip, Y. T., Park, R. E., Kosman, D., Yazdanbakhsh, K. & Levine, M. dorsal-twist interactions establish snail expression in the presumptive mesoderm of the Drosophila embryo. *Genes Dev.* 6, 1518–1530 (1992).
19. Lagha, M. et al. Paused Pol II coordinates tissue morphogenesis in the Drosophila embryo. *Cell* 153, 976–987 (2013).
20. Rushlow, C. A. & Shvartsman, S. Y. Temporal dynamics, spatial range, and transcriptional interpretation of the Dorsal morphogen gradient. *Curr. Opin. Genet. & Dev.* 22, 542–546 (2012).
21. Kanodia, J. S. et al. Pattern formation by graded and uniform signals in the early Drosophila embryo. *Biophys. J.* 102, 427–433 (2012).
22. Hannon, C. E., Blythe, S. A. & Wieschaus, E. F. Concentration dependent chromatin states induced by the bicoid morphogen gradient. *eLife* 6, 3165 (2017).

23. Ni, J.-Q. et al. A genome-scale shRNA resource for transgenic RNAi in Drosophila. *Nat. Methods* 8, 405–407 (2011).
24. Lionnet, T. & Singer, R. H. Transcription goes digital. *EMBO Rep.* 13, 313–321 (2012).
25. Hamm, D. C. et al. A conserved maternal-specific repressive domain in Zelda revealed by Cas9-mediated mutagenesis in Drosophila melanogaster. *PLoS Genet.* 13, e1007120 (2017).
26. Staudt, N., Fellert, S., Chung, H.-R., Jäckle, H. & Vorbrüggen, G. Mutations of the Drosophila zinc finger-encoding gene vielfältig impair mitotic cell divisions and cause improper chromosome segregation. *Mol. Biol. Cell.* 17, 2356–2365 (2006).
27. Gregor, T., Wieschaus, E. F., McGregor, A. P., Bialek, W. & Tank, D. W. Stability and nuclear dynamics of the bicoid morphogen gradient. *Cell* 130, 141–152 (2007).
28. Steffen, P. A. et al. Quantitative in vivo analysis of chromatin binding of Polycomb and Trithorax group proteins reveals retention of ASH1 on mitotic chromatin. *Nucleic Acids Res.* 41, 5235–5250 (2013).
29. Michelman-Ribeiro, A. et al. Direct measurement of association and dissociation rates of DNA binding in live cells by fluorescence correlation spectroscopy. *Biophys. J.* 97, 337–346 (2009).
30. Tsai, A. et al. Nuclear microenvironments modulate transcription from low-affinity enhancers. *eLife* 6, e28975–18 (2017).
31. Mir, M. *et al.* Dynamic multifactor hubs interact transiently with sites of active transcription in Drosophila embryos. Preprint at *bioRxiv* https://doi.org/10.1101/377812 (2018).
32. Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A Phase separation model for transcriptional control. *Cell* 169, 13–23 (2017).
33. Cho, W.-K. et al. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* 361, 412–415 (2018).
34. Abu-Arish, A., Porcher, A., Czerwonka, A., Dostatni, N. & Fradin, C. High mobility of bicoid captured by fluorescence correlation spectroscopy: implication for the rapid establishment of its gradient. *Biophys. J.* 99, L33–L35 (2010).
35. Hamm, D. C., Bondra, E. R. & Harrison, M. M. Transcriptional activation is a conserved feature of the early embryonic factor zelda that requires a cluster of four zinc fingers for DNA binding and a low-complexity activation domain. *J. Biol. Chem.* 290, 3508–3518 (2015).
36. Hug, C. B., Grimaldi, A. G., Kruse, K. & Vaquerizas, J. M. Chromatin architecture emerges during zygotic genome activation independent of transcription. *Cell* 169, 1–33 (2017).
37. Ogiyama, Y., Schuettengruber, B., Papadopoulos, G. L., Chang, J.-M. & Cavalli, G. Polycomb-dependent chromatin looping contributes to gene silencing during Drosophila Development. *Mol. Cell* 71, 1–22 (2018).
38. Lucas, T. et al. 3 minutes to precisely measure morphogen concentration. Preprint at *bioRxiv* https://doi.org/10.1101/305516 (2018).
39. Venken, K. J. T., He, Y., Hoskins, R. A. & Bellen, H. J. P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in D. melanogaster. *Science* 314, 1747–1751 (2006).
40. Tsanov, N. et al. smiFISH and FISH-quant – a flexible single RNA detection approach with super-resolution capability. *Nucleic Acids Res.* 44, e165–e165 (2016).
41. Sprague, B. L., Pego, R. L., Stavreva, D. A. & McNally, J. G. Analysis of binding reactions by fluorescence recovery after photobleaching. *Biophys. J.* 86, 3473–3495 (2004).
42. Escoffre, J. M., Hubert, M., Teissié, J., Rols, M. P. & Favard, C. Evidence for electro-induced membrane defects assessed by lateral mobility measurement of a GPI anchored protein. *Eur. Biophys. J.* 43, 277–286 (2014).
43. Axelrod, D., Koppel, D. E., Schlessinger, J., Elson, E. & Webb, W. W. Mobility measurement by analysis of fluorescence photobleaching recovery kinetics. *Biophysj* 16, 1055–1069 (1976).
44. Yguerabide, J., Schmidt, J. A. & Yguerabide, E. E. Lateral mobility in membranes as detected by fluorescence recovery after photobleaching. *Biophysj* 40, 69–75 (1982).
45. Müller, P., Schwille, P. & Weidemann, T. PyCorrFit-generic data evaluation for fluorescence correlation spectroscopy. *Bioinformatics* 30, 2532–2533 (2014).
46. Dertinger, T. et al. The optics and performance of dual-focus fluorescence correlation spectroscopy. *Opt. Express* 16, 14353–14368 (2008).

Acknowledgements

We are grateful to T.Fukaya for sharing the *MCP-eGFP-His2Av-mRFP* fly stock. We thank I.Ringrose for sharing *Ash1-GFP* fly stocks, C.Rushlow for *Zelda-RNAi* fly stocks and N. Dostatni for plasmids and eGFP-Bcd fly stocks. We thank G.Cavalli for providing the anti GAF antibody. We thank M. Bellec, L. Ciandrini, E. Bertrand, R. Feil and T. Forné for helpful discussions and critical reading of the manuscript. We acknowledge the imaging facility MRI, member of the national infrastructure France-BioImaging supported by the French National Research Agency (ANR-10-INBS-04, «Investments for the futures. A.T was a recipient of a FRM fellowship and is currently sponsored by the SyncDev ERC grant. Work in the lab of M.H. was supported by NIH R01GM 111694. This work was supported by the ERC SyncDev starting grant to M.L and a HFSP-CDA grant to M.L.

Author contributions

M.L conceived the project. M.L designed the experiment with the help from J.D. M.L, J. D, J.H, C.Fz, J.L, M.D, L.M, performed experiments. M.L, J.D and O.R analyzed the data and interpreted the results. A.T developed the software for image analysis. M.D and J.D performed fly handling. C.Fd advised J.D and A.T for FRAP and FCS experiments and performed the analysis. O.R performed mathematical modeling. M.H and K.S generated the *GFP-zld and mCherry-zld* lines and commented the data. M.L wrote the manuscript with help from J.D and O.R. All authors discussed, approved and reviewed the manuscript.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-07613-z.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

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

IV. DISCUSSION

1. Thesis summary: the main challenges

My thesis work is focused on the temporal regulation of gene expression, and more specifically in the contribution of core promoter sequences to regulate the synchrony of nuclei activation and the kinetic parameters of gene expression.

The studies in early *Drosophila* embryos where the notion of synchrony was assessed were performed in fixed embryos using fluorescent in situ hybridization (FISH). Although it is a very sensitive technique, temporal resolution is poor (at best 15-20 minutes) for a window of time (i.e.: interphase 14, that lasts around 50 min). This temporal resolution scale is an order of magnitude higher than a typical initiation time (on the order of seconds). Thus, we take advantage of a transgenic approach described in Lucas et al, 2013; Garcia et al, 2013; Fernandez and Lagha, 2019 to monitor transcriptional synchrony using the MS2 stem-loop/MCP system (Bertrand et al, 1998) in live *Drosophila* embryos. Our approach consisted of creating transgenic lines that are inserted in the same genomic location, depend on a unique enhancer, have the same reporter gene, and where only the core promoter sequence is modified to test its effect on temporal transcription dynamics.

Taking advantage of our transgenic approach, we also tested if core promoters could influence bursting parameters (frequency and burst size). Moreover, as core promoter sequences are composed by different combinations of highly conserved motifs, we would be able to disentangle the contribution of these particular sequences, namely the TATA box and the INR motifs, in controlling temporal dynamics of gene expression. Thus, I focused my attention in the transcriptional site (TS) raw traces to see if transcription from core promoters showed a bursty behavior and identify which parameters (frequency and/or burst size) were affected. Nevertheless, before starting the analysis, we faced two challenges. The first one concerned the analysis of the TS raw traces. Since I was the first person to be interested by bursting features, there was no available software in the lab that measured intensity fluctuations in 3 D (the main soft was developed for 2D spot detection). Thus, I had to wait until the biophysicist of the lab developed a 3D software. The second challenge we faced was the settings acquisitions for the live imaging, as we were focused on gene activation and the

first acquisitions were saturated to increase the signal visualization (or signal to noise ratio). Therefore, one important aspect of my thesis work consisted in finding the optimal settings to have a right compromise between signal to noise ratio (without saturation) and temporal resolution. Then, in 2017, our imaging facility acquired a new Zeiss LSM880 confocal microscope carrying an Airyscan module. Thanks to this technology, we improved the temporal resolution from 20 sec to 4 sec, without compromising on the spatial resolution.

Once we developed the appropriate microscopy parameters and the software allowing the 3D spots segmentation, we faced another problem concerning the definition of a burst event. In live Drosophila embryos, detection of single mRNA molecules using the MS2 system has not yet been reached. Therefore, setting up a threshold to define a burst event in our data could mislead the promoter effect on transcription kinetics. To overcome this, intensity traces from each promoter construct were analysed using a deconvolution approach developed by (E. Bertrand and O. Radulescu), where each single nucleus intensity trace goes through a numerical deconvolution, to place each single initiation without using a "bursting calling." However, to apply this method we needed to know the intensity produce for a single transcript to calibrate single nuclei traces in terms of the number of polymerases at each time point of the live acquisition. To face this, I performed single molecular FISH (smFISH) for each promoter and I developed a pipeline of analysis using the commercial software *Imaris*© with the aim of calibrating single intensity traces from the live acquisition, but also to compare average promoter activity (analysis is ongoing currently).

2. Main conclusions and Implications.

2.1.1 The question of synchrony

How transcription leads to precise and reproducible patterns of genes expression during development remains unclear so far. The development of new labeling methods to detect mRNA molecules showed that transcription is not a homogenous process. Indeed, neighboring cells could show different degrees of variability in terms of mRNAs production (Raj and van Oudenaarden, 2008, Suter et al 2011, Lenstra et al, 2016). In the context of a multicellular developing embryo, this variability is counterintuitive, if one posits that transcriptional precision would be required to produce reproducible and robust patterns of gene expression. Boettiger and Levine (2009) revealed the existence of two modes of transcription activation in the early *Drosophila* embryo: synchronous and stochastic. Some developmental genes are activated in a fast and synchronous manner, where virtually all the cells of the final pattern show simultaneous activation. In contrast, other genes exhibit an erratic pattern of activation, with different cells of the tissue turning on gene expression at different times. By using fixed *Drosophila* embryos, the authors also showed that promoter regions of synchronous genes show paused RNAPII, but not stochastic genes. In the same trend, Lagha et al (2013) used a promoter swapping strategy in fixed *Drosophila* embryos and discovered that core promoters of key developmental genes play a pivotal role in pausing, which in turn determines synchrony of gene activation. The authors demonstrate that substitutions of a paused minimal promoter (~100bp) with a non-paused promoter results in slow and stochastic activation of gene expression. Moreover, they revealed that the synchronous activation of the *snail* gene is essential for proper mesoderm invagination in the developing *Drosophila* embryo. Hence, it is the minimal promoter, and not the enhancer, that determines the levels of paused Pol II and the synchrony of gene activation.

These observations raised questions about the temporal regulation of gene expression, and more specifically if synchrony, controlled by core promoters, was one of the mechanisms leading to gene expression heterogeneity. Therefore, during my thesis I focused on how core promoter controls synchrony, and the relation of synchrony and RNA production, with the hypothesis that highly synchronous genes will be paused and would show robust and homogenous gene expression.

2.1.2 Minimal promoters control synchrony

The synchrony analysis for the live imaging acquisitions showed, as before by FISH (Boettiger and Levine, 2009; Lagha et al, 2013), that core promoter sequence control synchrony. However, in our study synchrony is not always correlated with promoter activity/strength, as exemplified by the INR group, where the three core promoters have similar synchronies (~13 min) but produce different levels of intensity (~2X difference between promoters) during the interphase 14. Another interesting observation was that synchrony did not automatically correlate with pausing. For example, the *ilp4* core promoter (which is endogenously paused), showed a poor synchrony (~27 min), in contrast with the paused promoter *kr* (~13 min). Opposingly, *wntD,* a non-paused promoter, shows a synchrony of ~15 min. This suggests that pausing is not the only mechanism to achieve synchrony. In the case of *wntD* the synchrony profile might rely on the presence of a canonical TATA box. Indeed, Chen et al (2013) associated the TATA box motif with rapid early transcription. Genes transcribed before interphase 14 are constrained by short interphases between mitosis (~8-15 min), and therefore the 'timing of transcription' would need to be optimized. These genes are highly enriched in TATA motifs, in contrast with genes expressed during the interphase 14 or later. This suggests that promoters are differentially used during the early phases of development, and this correlates with promoter architecture. Thus, to disentangle the distinct contributions of core promoter motifs on synchrony, we took two promoter that have similar synchrony and intensity production, but differ by their core promoter motifs. On one side we took the *sna* promoter to test the TATA box, and on the other *kr* promoter to test the INR motif.

The mutations of the TATA box showed a dramatically effect on synchrony and intensity production accordingly. This observation reinforces the idea that TATA containing promoters have high synchrony and robust expression, likely through the stabilization of the interaction with the TFIID complex. These results are consistent with the known properties of TATA-containing promoters. TATA is a strong core promoter element that efficiently supports transcription *in vitro* (Aso et al, 1994), and mediates efficient re-initiation in vitro (Yean and Gralla, 1997, 1999). Its presence *in vivo* correlates with 'bursts' of transcription that produce many transcripts within a short time (Zenklusen et al, 2008), and with genes that have significantly higher bust sizes (number of transcripts produced), (Larson et al, 2019).

The INR mutations did not affect synchrony, but rather the cumulative intensities. The replacement of the INR with the *sna* non-INR TSS leads to almost 2-fold more mRNA production; in contrast the mutation of a single base (the G in +2 position) results in almost 2-fold intensity decrease. The INR motif has been shown to correlate with a focused transcription initiation (Sainsbury, Cramer et al 2015) probably by stabilizing TFIID via TAF1 and TAF2 interactions, and with RNAPII pausing stability in the case of highly paused promoters (Shao et al, 2019). As synchrony seems not be affected for the INR mutants, we hypothesized that in the case of the mutant with the TSS of *sna* (a TATA containing promoter), this TSS allows a better communication with the non-canonical TATA motif present in the *kr* promoter. Nevertheless, in the single mutation of the INR, the interaction with the TFIID complex is probably destabilized leading to a low RNAPII recruitment.

Thus, from these observations, with the small number of sequences tested, it seems that the TATA motif affects synchrony while the INR does not.

2.1.3 Modeling promoter kinetics

Another notion that we sought to assess during my thesis work concerned the transcriptional bursting. Indeed, the visualization of the transcription process showed that transcription is not a continuous process but rather a discontinuous one, with periods of activity interspaced with periods of inactivity, better known as transcriptional bursting, adding a level of complexity to our understanding of regulation in gene expression. So far, the biological role of transcriptional bursting is unclear. However, several studies have been performed to understand the origins and consequences of transcriptional bursting and its contribution to gene expression heterogeneity (Lenstra et al, 2016, Nicolas et al, 2017).

The analysis of single TS traces using a mathematical model of deconvolution showed that dynamics for most tested developmental promoters were captured using a two-state model. In this model, a promoter switches between two states, ON and OFF, and transcription can take place only in the ON state, where polymerases transcribe with an initiation rate. In the two-state model, the ON and OFF times are assumed to be regulated by single rate-limiting steps, with exponentially distributed waiting times.

Interestingly, the two-state model was enough to capture the dynamics for the *sna* and the sna*TATAlight* mutant promoter, affecting the OFF time mainly (~4X), and to a lesser extent the initiation time (~1.5X) and the ON durations (~1.3X). The TATA motif is bound by the TBP subunit of TFIID (Reeve, 2003), and mutations of the TATA box were reported to decrease the HIV-1 expression and the affinity of TBP for the promoter (van Opijnen *et al.*, 2004; Savinkova *et al.*, 2013). Thus, is likely for the *sna* and *snaTATA light* promoters that the single rate-limiting step regulating promoter activity depends on the affinity and stability of TBP / TATA box interaction, which in turn affects the RNAPII recruitment, having a direct effect on synchrony and intensity levels. In the case of *kr,* an INR containing promoter, a two-model state was not enough to capture its dynamic, and therefore we considered a three-state model, with two OFF states and one ON state in which transcription can take place. Upon the mutation of the INR motif (*krINRmut*), *kr* dynamics can be capture by a two-state model. Moreover, the addition of an INR motif to the *sna* promoter, requires the use of a three-state model. It seems that here that the role of the INR consists of adding a refractory state between each transcription event. As the INR motif was correlated with RNAPII pausing

stability in high paused promoters (Shao et al, 2019), one can make the hypothesis that this refractory state corresponds to a paused state.

All together, these results suggest that TATA-enriched promoters and paused promoters have different transcription dynamics and might serve different purposes during development.

2.1.4 Pausing and promoter dynamics

Pausing is a hallmark of metazoan transcription. It occurs 30 to 60bp downstream of the TSS (Core et al, 2008) and is particularly robust at genes that are rapidly responsive to signalling pathways (Adelman and Lis, 2012; Kwak and Lis 2013). Pausing has been proposed as regulatory step to initiate transcription, preparing genes for rapid, synchronous and robust activation (Adelman and Lis, 2012; Saunders et al, 2013). However, it is unclear how pausing affects the initiation of new transcripts during consecutive rounds of transcription. Shao and Zeitlinger (2017) showed using a biochemical approach (ChIP-Nexus) that RNAPII pausing inhibits new initiation events, suggesting that paused RNAPII acts as a limiting step for transcription initiation at paused genes. Recently the vision of pause promoter, where the RNAPII is pre-loaded and stay 'stalled' until activation was challenged by recent studies concerning the dynamics of pause polymerase (Buckley et al, 2014; Shao and Zeitlinger, 2017; Krebs et al, 2017). Using different approaches, these studies showed that pausing could be highly dynamic. That is the case for the heat shock promoter *Hsp70*, known to be a highly paused gene. In the first study, using a tracked photoactivatable GFP-tagged Pol II at un-induced *Hsp70* gene on polytene chromosomes, the authors showed that Pol II is stably paused with a half-life of 5 min (Buckley et al, 2014). However, using two different biochemical approaches complemented by transcriptional perturbation assays with triptolide (Shao and Zeitlinger, 2017; Krebs et al, 2017) showed that the half-life of the RNAP II at the *hsp70* promoter was faster than this previous result (below 2.5 min) and this did not change when the gene was induced or not. Indeed, both studies demonstrated that upon triptolide treatment, paused genes have different degrees of pausing half–life variability. Thus, if pausing is a limiting-step to initiate new transcription events, two interesting question would be, first, the relation between pausing stability and core promoter elements, and second how

pause stability at promoters affects transcription kinetics (promoter switching ON/OFF states). In our case, INR containing promoters needed to be modelled with a three-state model to capture the full dynamics. As the INR motif was shown to be correlated with pausing stability (Shao et al, 2019) we hypothesize that the additional state observed for these promoters could be linked with pausing.

2.1.5 Promoter initiation rates

Initiation rates do not change dramatically between the different promoters (1 polymerase every ~ 8 sec), with the exception of the *ilp4* promoter (1 polymerase every ~ 20 sec). For the moment we cannot explain why *ilp4* had such low initiation rates, and therefore we have developed different mutations in its sequence to understand better its dynamics (ongoing). Although *sna* and *kr* present different transcriptional dynamics and core promoter architecture, their initiation rates were very close (~9 and ~8 sec, respectively). It seems that initiation rate is an independent parameter that might be regulated by other means, such as the enhancer (in terms of frequency of promoter activation) or the elongation speed (RNAPII spatial constraints).

Interestingly, the RNAPII initiation rate of the *sna* endogenous promoter (tagged by CRISPR-MS2 approach) is about, 1 polymerase every ~ 10 sec; very close to what we observe with the *sna* transgenic promoter of 1 polymerase every ~ 9 sec.

2.1.6 Promoter Dynamics : TATA box vs INR

In this study with a small set of genes, we observed that two distinct dynamics of transcription depending on core promoter architecture. On one side, TATA mutants have a strong effect on synchrony and intensity levels, however their dynamic of promoter activation can be captured by a two-state model. On the other side the INR mutants show no effect in synchrony, but rather affect the intensity levels. In terms of promoter dynamics, *kr* needed a three-state model to capture its dynamics, but when the INR motif is replaced, its dynamics can be capture with a two-state model. Based on the literature, I hypothesized that the supplementary state controlled by the INR in this context might be pausing.

Chen et al (2013) showed that early-activated genes (before the MBT that in *Drosophila* coincides with the Zygote Genome Activation (ZGA)) have no notable enrichment of RNAP II at the pause site (30–50 bp downstream from the TSS), in contrast with genes expressed during the MBT or at later stages that are highly occupied by the RNAPII at the pause site. Moreover, promoter architecture differs between early expressed genes (before the MBT) and MBT genes, suggesting that promoters are differentially used during the early phases of development. The observations made in this study seem to agree with this interpretation. It is possible that TATA promoters are preferentially used during the early gene transcription, where a fast and robust expression is needed because of the temporal constraint of mitosis cycles. Transcription of TATA dependent promoters might depend on one limiting step: the stability of interaction between TATA box and TBP, to ensure an optimal recruitment of the RNAPII to the promoter.

MBT genes have a supplementary step of regulation, which seems to be associated with the RNAPII pause, and in our case with the INR sequence. This suggests that core promoter activity could be more or less permissive for transcription in the context where transcription must occur, and that it is strongly associated with the core promoter architecture.

3. Remaining work & perspectives

3.1 Molecular Mechanisms

The next step consists of linking the observations made in this work with molecular mechanisms. In the case of the INR-containing promoters, we hypothesized that the supplementary refractory state observed in the modelling data might be related to RNAPII pausing. Thus, to test this hypothesis we need to verify weather is RNAPII pausing at these promoters. In the same way we need to verify the TBP occupancy at TATA-containing promoters and the TATA mutants, both to understand whether or not binding occurs but also to examine the relative frequency of interaction between TBP and the endogenous TATA box, *de novo* TATA box, and weakened TATA box. I plan to verify this by Chromatin Immunoprecipitation, followed by qPCR. (ChIP-qPCR).

To strengthen our observation concerning the roles of INR and TATA, we are making supplementary mutations (explained in the second manuscript: ongoing work section).

3.2 Enhancer-promoter selectivity

The current definition of an enhancer stipulates that a putative enhancer should work with heterologous promoters (Banerji, Rusconi and Schaffner, 1981; Moreau *et al.*, 1981; Fromm and Berg, 1982; Khoury and Gruss, 1983; Atchison and Perry, 1988). This concept is somewhat controversial because studies have shown that some enhancers display preferences for specific promoters (Arnold *et al.*, 2013, 2017; Shlyueva, Stampfel and Stark, 2014). It is unclear which case is biologically relevant, or even if these cases are mutually exclusive. Furthermore, what are the mechanisms behind this? For example, what could one expect in terms of synchrony if I test the same set of promoter with another developmental enhancer from either a stronger or a weaker promoter (i.e: *ilp4* enhancer)

3.3 Effect in endogenous conditions

To fully understand the impact of core promoter sequences during development, it is necessary to test them in an endogenous context with natural enhancers and temporal constraints. If particular core architectures are used preferentially during the development, changes in the core promoter element composition will lead to changes in the transcriptional response of early expressed genes. As the TATA box is associated with early expressed genes, it would be interesting to observe the consequences of TATA mutations in early gene expression, as well of the INR mutations.

V.REFERENCES

Adelman, K. and Lis, J. T. (2012) 'Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans', *Nature Reviews Genetics*, 13(10), pp. 720–731. doi: 10.1038/nrg3293.

Akhtar, W. and Veenstra, G. J. C. (2011) 'TBP-related factors: a paradigm of diversity in transcription initiation', *Cell & Bioscience*, 1(1), p. 23. doi: 10.1186/2045-3701-1-23.

Andersson, R. and Sandelin, A. (2019) 'Determinants of enhancer and promoter activities of regulatory elements', *Nature Reviews Genetics*. Nature Publishing Group, pp. 1–17. doi: 10.1038/s41576-019-0173-8.

Andersson, R., Sandelin, A. and Danko, C. G. (2015) 'A unified architecture of transcriptional regulatory elements', *Trends in Genetics*, 31(8), pp. 426–433. doi: 10.1016/j.tig.2015.05.007.

van Arensbergen, J., van Steensel, B. and Bussemaker, H. J. (2014) 'In search of the determinants of enhancer–promoter interaction specificity', *Trends in Cell Biology*, 24(11), pp. 695–702. doi: 10.1016/j.tcb.2014.07.004.

Arnold, C. D. *et al.* (2013) 'Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq', *Science*, 339(6123), pp. 1074–1077. doi: 10.1126/science.1232542.

Arnold, C. D. *et al.* (2017) 'Genome-wide assessment of sequence-intrinsic enhancer responsiveness at single-base-pair resolution', *Nature Biotechnology*, 35(2), pp. 136–144. doi: 10.1038/nbt.3739.

Atchison, M. L. and Perry, R. P. (1988) 'Complementation between two cell lines lacking kappa enhancer activity: implications for the developmental control of immunoglobulin transcription.', *The EMBO journal*, 7(13), pp. 4213–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2854057 (Accessed: 15 October 2019).

Banerji, J., Olson, L. and Schaffner, W. (1983) 'A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes', *Cell*, 33(3), pp. 729–740. doi: 10.1016/0092-8674(83)90015-6.

Banerji, J., Rusconi, S. and Schaffner, W. (1981) 'Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences.', *Cell*, 27(2 Pt 1), pp. 299–308. doi: 10.1016/0092-8674(81)90413-x.

Bártfai, R. *et al.* (2004) 'TBP2, a Vertebrate-Specific Member of the TBP Family, Is Required in Embryonic Development of Zebrafish', *Current Biology*. Cell Press, 14(7), pp. 593–598. doi: 10.1016/J.CUB.2004.03.034.

Bashirullah, A. *et al.* (1999) 'Joint action of two RNA degradation pathways controls the timing of maternal transcript elimination at the midblastula transition in Drosophila melanogaster.', *The EMBO journal*. European Molecular Biology Organization, 18(9), pp. 2610–20. doi: 10.1093/emboj/18.9.2610.

Bashirullah, A., Cooperstock, R. L. and Lipshitz, H. D. (2001) 'Spatial and temporal control of

RNA stability', *Proceedings of the National Academy of Sciences*, 98(13), pp. 7025–7028. doi: 10.1073/pnas.111145698.

Benoist, C. and Chambon, P. (1981) 'In vivo sequence requirements of the SV40 early promoter region', *Nature*. Nature Publishing Group, 290(5804), pp. 304–310. doi: 10.1038/290304a0.

Bernstein, B. E. *et al.* (2006) 'A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells', *Cell*, 125(2), pp. 315–326. doi: 10.1016/j.cell.2006.02.041.

Bhuiyan, T. and Timmers, H. T. M. (2019) 'Promoter Recognition: Putting TFIID on the Spot', *Trends in Cell Biology*. doi: 10.1016/j.tcb.2019.06.004.

Blake, W. J. *et al.* (2006) 'Phenotypic Consequences of Promoter-Mediated Transcriptional Noise', *Molecular Cell*. doi: 10.1016/j.molcel.2006.11.003.

Boettiger, A. N. and Levine, M. (2009) 'Synchronous and stochastic patterns of gene activation in the Drosophila embryo.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 325(5939), pp. 471–3. doi: 10.1126/science.1173976.

Bothma, J. P. *et al.* (2014) 'Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living Drosophila embryos', *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1410022111.

Buckley, M. S. *et al.* (2014) 'Kinetics of promoter Pol II on Hsp70 reveal stable pausing and key insights into its regulation', *Genes and Development*. doi: 10.1101/gad.231886.113.

Buratowski, S. *et al.* (1989) 'Five intermediate complexes in transcription initiation by RNA polymerase II', *Cell*. Cell Press, 56(4), pp. 549–561. doi: 10.1016/0092-8674(89)90578-3.

Burley, S. K. and Roeder, R. G. (1996) 'Biochemistry and Structural Biology of Transcription Factor IID (TFIID)', *Annual Review of Biochemistry*, 65(1), pp. 769–799. doi: 10.1146/annurev.bi.65.070196.004005.

Butler, J. E. and Kadonaga, J. T. (2001) 'Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs.', *Genes & development*. Cold Spring Harbor Laboratory Press, 15(19), pp. 2515–9. doi: 10.1101/gad.924301.

Cech, T. R. and Steitz, J. A. (2014) 'The noncoding RNA revolution-trashing old rules to forge new ones.', *Cell*, 157(1), pp. 77–94. doi: 10.1016/j.cell.2014.03.008.

Chafin, D. R., Claussen, T. J. and Price, D. H. (1991) 'Identification and purification of a yeast protein that affects elongation by RNA polymerase II.', *The Journal of biological chemistry*, 266(14), pp. 9256–62. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1851172 (Accessed: 15 October 2019).

Chalkley, G. E. (1999) 'DNA binding site selection by RNA polymerase II TAFs: a TAFII250-TAFII150 complex recognizes the Initiator', *The EMBO Journal*, 18(17), pp. 4835–4845. doi: 10.1093/emboj/18.17.4835.

Chapman, R. D., Conrad, M. and Eick, D. (2005) 'Role of the mammalian RNA polymerase II C-terminal domain (CTD) nonconsensus repeats in CTD stability and cell proliferation.', *Molecular and cellular biology*, 25(17), pp. 7665–74. doi: 10.1128/MCB.25.17.7665-

7674.2005.

Chen, H.-T. and Hahn, S. (2004) 'Mapping the Location of TFIIB within the RNA Polymerase II Transcription Preinitiation Complex', *Cell*, 119(2), pp. 169–180. doi: 10.1016/j.cell.2004.09.028.

Chen, K. *et al.* (2013) 'A global change in RNA polymerase II pausing during the Drosophila midblastula transition', 2, p. 861. doi: 10.7554/eLife.00861.001.

Cianfrocco, M. A. *et al.* (2013) 'Human TFIID Binds to Core Promoter DNA in a Reorganized Structural State', *Cell*, 152(1–2), pp. 120–131. doi: 10.1016/j.cell.2012.12.005.

Coin, F., Oksenych, V. and Egly, J.-M. (2007) 'Distinct Roles for the XPB/p52 and XPD/p44 Subcomplexes of TFIIH in Damaged DNA Opening during Nucleotide Excision Repair', *Molecular Cell*, 26(2), pp. 245–256. doi: 10.1016/j.molcel.2007.03.009.

Coleman, R. A. and Pugh, B. F. (1995) 'Evidence for Functional Binding and Stable Sliding of the TATA Binding Protein on Nonspecific DNA', *Journal of Biological Chemistry*, 270(23), pp. 13850–13859. doi: 10.1074/jbc.270.23.13850.

Corden, J. *et al.* (1980) 'Promoter sequences of eukaryotic protein-coding genes', *Science*, 209(4463), pp. 1406–1414. doi: 10.1126/science.6251548.

Core, L. J. *et al.* (2012) 'Defining the Status of RNA Polymerase at Promoters', *Cell Reports*. The Authors, 2(4), pp. 1025–1035. doi: 10.1016/j.celrep.2012.08.034.

Core, L. J. *et al.* (2014) 'Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers', *Nature Genetics*, 46(12), pp. 1311–1320. doi: 10.1038/ng.3142.

Core, L. J., Waterfall, J. J. and Lis, J. T. (2008) 'Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters.', *Science (New York, N.Y.)*, 322(5909), pp. 1845–8. doi: 10.1126/science.1162228.

Cormack, B. P. and Struhl, K. (1992) 'The TATA-binding protein is required for transcription by all three nuclear RNA polymerases in yeast cells.', *Cell*. Elsevier, 69(4), pp. 685–96. doi: 10.1016/0092-8674(92)90232-2.

Corrigan, A. M. *et al.* (2016) 'A continuum model of transcriptional bursting.', *eLife*. eLife Sciences Publications, Ltd, 5. doi: 10.7554/eLife.13051.

Crowley, T. E. *et al.* (1993) 'A new factor related to TATA-binding protein has highly restricted expression patterns in Drosophila.', *Nature*, 361(6412), pp. 557–61. doi: 10.1038/361557a0.

Danino, Y. M. *et al.* (2015) 'The core promoter: at the heart of gene expression', *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*. doi: 10.1016/j.bbagrm.2015.04.003.

Dantonel, J.-C. *et al.* (2000) 'TBP-like Factor Is Required for Embryonic RNA Polymerase II Transcription in C. elegans', *Molecular Cell*. Cell Press, 6(3), pp. 715–722. doi: 10.1016/S1097-2765(00)00069-1.

Dao, L. T. M. *et al.* (2017) 'Genome-wide characterization of mammalian promoters with distal enhancer functions', *Nature Genetics*, 49(7), pp. 1073–1081. doi: 10.1038/ng.3884.

Deng, W. and Roberts, S. G. E. (2005) 'A core promoter element downstream of the TATA box that is recognized by TFIIB', *Genes & Development*, 19(20), pp. 2418–2423. doi: 10.1101/gad.342405.

Driever, W. and Nüsslein-Volhard, C. (1988) 'The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner', *Cell*, 54(1), pp. 95–104. doi: 10.1016/0092-8674(88)90183-3.

Dufourt, J. *et al.* (2018) 'Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs', *bioRxiv*. Springer US, p. 282426. doi: 10.1101/282426.

Duttke, S. H. C. *et al.* (2014) 'TRF2 and the evolution of the bilateria', *Genes & Development*, 28(19), pp. 2071–2076. doi: 10.1101/gad.250563.114.

Egly, J.-M. (2011) '[DNA repair mechanisms: from basic biology to genetic pathology].', *Bulletin de l'Academie nationale de medecine*, 195(7), pp. 1689–90. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22812171 (Accessed: 15 October 2019).

Eick, D. and Geyer, M. (2013) 'The RNA Polymerase II Carboxy-Terminal Domain (CTD) Code', *Chemical Reviews*, 113(11), pp. 8456–8490. doi: 10.1021/cr400071f.

Elowitz, M. B. *et al.* (2002) 'Stochastic Gene Expression in a Single Cell', *Science*, 297(5584), pp. 1183–1186. doi: 10.1126/science.1070919.

Farrell, J. A. and O'Farrell, P. H. (2014) 'From Egg to Gastrula: How the Cell Cycle Is Remodeled During the *Drosophila* Mid-Blastula Transition', *Annual Review of Genetics*. Annual Reviews , 48(1), pp. 269–294. doi: 10.1146/annurev-genet-111212-133531.

Femino, A. M. *et al.* (2003) 'Visualization of single molecules of mRNA in situ.', *Methods in enzymology*, 361, pp. 245–304. doi: 10.1016/s0076-6879(03)61015-3.

Fernandez, C. and Lagha, M. (2019) 'Lightening Up Gene Activation in Living Drosophila Embryos', *Methods in molecular biology*, 2038. Available at: https://doi.org/10.1007/978-1-4939-9674-2_5.

Ferraro, T. *et al.* (2016) 'New methods to image transcription in living fly embryos: The insights so far, and the prospects', *Wiley Interdisciplinary Reviews: Developmental Biology*. doi: 10.1002/wdev.221.

Fishburn, J. and Hahn, S. (2012) 'Architecture of the Yeast RNA Polymerase II Open Complex and Regulation of Activity by TFIIF', *Molecular and Cellular Biology*, 32(1), pp. 12–25. doi: 10.1128/MCB.06242-11.

FitzGerald, P. C. *et al.* (2006) 'Comparative genomics of Drosophila and human core promoters.', *Genome Biology*, 7(7), p. R53. doi: 10.1186/gb-2006-7-7-r53.

Foe, V. E. and Alberts, B. M. (1983) 'Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in Drosophila embryogenesis.', *Journal of cell science*, 61, pp. 31–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6411748

(Accessed: 14 October 2019).

Fromm, M. and Berg, P. (1982) 'Deletion mapping of DNA regions required for SV40 early region promoter function in vivo.', *Journal of molecular and applied genetics*, 1(5), pp. 457–81. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6296253 (Accessed: 15 October 2019).

Fukaya, T., Lim, B. and Levine, M. (2016) 'Enhancer Control of Transcriptional Bursting.', *Cell*. Elsevier, 166(2), pp. 358–368. doi: 10.1016/j.cell.2016.05.025.

Fukaya, T., Lim, B. and Levine, M. (2017) 'Rapid Rates of Pol II Elongation in the Drosophila Embryo.', *Current biology : CB*. NIH Public Access, 27(9), pp. 1387–1391. doi: 10.1016/j.cub.2017.03.069.

Gaertner, B. and Zeitlinger, J. (2014) 'RNA polymerase II pausing during development', *Development*. doi: 10.1242/dev.088492.

Galbraith, M. D., Donner, A. J. and Espinosa, J. M. (2010) 'CDK8', *Transcription*, 1(1), pp. 4–12. doi: 10.4161/trns.1.1.12373.

Garcia, H. G. *et al.* (2013) 'Quantitative Imaging of Transcription in Living Drosophila Embryos Links Polymerase Activity to Patterning', *Current Biology*. doi: 10.1016/j.cub.2013.08.054.

Garcia, M. *et al.* (2013) 'Size-dependent regulation of dorsal–ventral patterning in the early Drosophila embryo', *Developmental Biology*, 381(1), pp. 286–299. doi: 10.1016/j.ydbio.2013.06.020.

Gavis, E. R. and Lehmann, R. (1992) 'Localization of nanos RNA controls embryonic polarity', *Cell*, 71(2), pp. 301–313. doi: 10.1016/0092-8674(92)90358-J.

Gazdag, E. *et al.* (2009) 'TBP2 is essential for germ cell development by regulating transcription and chromatin condensation in the oocyte.', *Genes & development*. Cold Spring Harbor Laboratory Press, 23(18), pp. 2210–23. doi: 10.1101/gad.535209.

Gehrig, J. *et al.* (2009) 'Automated high-throughput mapping of promoter-enhancer interactions in zebrafish embryos', *Nature Methods*, 6(12), pp. 911–916. doi: 10.1038/nmeth.1396.

Gershenzon, N. I. and Ioshikhes, I. P. (2005) 'Promoter classifier: software package for promoter database analysis.', *Applied bioinformatics*, 4(3), pp. 205–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16231962 (Accessed: 15 October 2019).

Gilchrist, D. A. *et al.* (2010) 'Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation.', *Cell*, 143(4), pp. 540–51. doi: 10.1016/j.cell.2010.10.004.

Gillies, S. D. *et al.* (1983) 'A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene', *Cell*, 33(3), pp. 717–728. doi: 10.1016/0092-8674(83)90014-4.

Gilmour, D. S. and Lis, J. T. (1986) 'RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells.', *Molecular and cellular*

biology. American Society for Microbiology (ASM), 6(11), pp. 3984–9. doi: 10.1128/mcb.6.11.3984.

Goodrich, J. A. and Tjian, R. (2010) 'Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation', *Nature Reviews Genetics*, 11(8), pp. 549–558. doi: 10.1038/nrg2847.

Grosschedl, R. and Birnstiel, M. L. (1980) 'Spacer DNA sequences upstream of the T-A-T-A-A-A-T-A sequence are essential for promotion of H2A histone gene transcription in vivo.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 77(12), pp. 7102–6. doi: 10.1073/pnas.77.12.7102.

Guarente, L. (1988) 'UASs and enhancers: common mechanism of transcriptional activation in yeast and mammals.', *Cell*, 52(3), pp. 303–5. doi: 10.1016/s0092-8674(88)80020-5.

Guzmán, E. and Lis, J. T. (1999) 'Transcription Factor TFIIH Is Required for Promoter Melting In Vivo', *Molecular and Cellular Biology*, 19(8), pp. 5652–5658. doi: 10.1128/MCB.19.8.5652.

Ha, I. *et al.* (1993) 'Multiple functional domains of human transcription factor IIB: distinct interactions with two general transcription factors and RNA polymerase II.', *Genes & Development*, 7(6), pp. 1021–1032. doi: 10.1101/gad.7.6.1021.

Haberle, V. *et al.* (2014) 'Two independent transcription initiation codes overlap on vertebrate core promoters', *Nature*, 507(7492), pp. 381–385. doi: 10.1038/nature12974.

Haberle, V. and Stark, A. (2018) 'Eukaryotic core promoters and the functional basis of transcription initiation', *Nature Reviews Molecular Cell Biology*. Europe PMC Funders, pp. 621–637. doi: 10.1038/s41580-018-0028-8.

Heine, G. F., Horwitz, A. A. and Parvin, J. D. (2008) 'Multiple Mechanisms Contribute to Inhibit Transcription in Response to DNA Damage', *Journal of Biological Chemistry*, 283(15), pp. 9555–9561. doi: 10.1074/jbc.M707700200.

Hendrix, D. A. *et al.* (2008) 'Promoter elements associated with RNA Pol II stalling in the Drosophila embryo', *Proceedings of the National Academy of Sciences*, 105(22), pp. 7762–7767. doi: 10.1073/pnas.0802406105.

Henriques, T. *et al.* (2013) 'Stable Pausing by RNA Polymerase II Provides an Opportunity to Target and Integrate Regulatory Signals', *Molecular Cell*, 52(4), pp. 517–528. doi: 10.1016/j.molcel.2013.10.001.

Hintermair, C. *et al.* (2016) 'Specific threonine-4 phosphorylation and function of RNA polymerase II CTD during M phase progression.', *Scientific reports*, 6(1), p. 27401. doi: 10.1038/srep27401.

Høiby, T. *et al.* (2007) 'A facelift for the general transcription factor TFIIA', *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*. Elsevier, 1769(7–8), pp. 429–436. doi: 10.1016/J.BBAEXP.2007.04.008.

Holmes, M. C. and Tjian, R. (2000) 'Promoter-selective properties of the TBP-related factor TRF1.', *Science (New York, N.Y.)*, 288(5467), pp. 867–70. doi: 10.1126/science.288.5467.867.

Holstege, F. C., van der Vliet, P. C. and Timmers, H. T. (1996) 'Opening of an RNA polymerase

II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH.', *The EMBO journal*, 15(7), pp. 1666–77. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8612591 (Accessed: 15 October 2019).

Horikoshi, M. *et al.* (1990) 'Analysis of structure-function relationships of yeast TATA box binding factor TFIID.', *Cell*. Elsevier, 61(7), pp. 1171–8. doi: 10.1016/0092-8674(90)90681-4.

Hoskins, R. A. *et al.* (2011) 'Genome-wide analysis of promoter architecture in Drosophila melanogaster', *Genome Research*, 21(2), pp. 182–192. doi: 10.1101/gr.112466.110.

Ibrahim, M. M. *et al.* (2018) 'Determinants of promoter and enhancer transcription directionality in metazoans.', *Nature communications*. Nature Publishing Group, 9(1), p. 4472. doi: 10.1038/s41467-018-06962-z.

Jeronimo, C. and Robert, F. (2014) 'Kin28 regulates the transient association of Mediator with core promoters', *Nature Structural & Molecular Biology*, 21(5), pp. 449–455. doi: 10.1038/nsmb.2810.

Jonkers, I. and Lis, J. T. (2015) 'Getting up to speed with transcription elongation by RNA polymerase II', *Nature Reviews Molecular Cell Biology*. NIH Public Access, pp. 167–177. doi: 10.1038/nrm3953.

Juven-Gershon, T., Hsu, J.-Y. and Kadonaga, J. T. (2008) 'Caudal, a key developmental regulator, is a DPE-specific transcriptional factor', *Genes & Development*, 22(20), pp. 2823–2830. doi: 10.1101/gad.1698108.

Kaufmann, J. and Smale, S. T. (1994) 'Direct recognition of initiator elements by a component of the transcription factor IID complex.', *Genes & development*, 8(7), pp. 821–9. doi: 10.1101/gad.8.7.821.

Kedinger, C. *et al.* (1970) 'α-Amanitin: A specific inhibitor of one of two DNA-dependent RNA polymerase activities from calf thymus', *Biochemical and Biophysical Research Communications*, 38(1), pp. 165–171. doi: 10.1016/0006-291X(70)91099-5.

Khoury, G. and Gruss, P. (1983) 'Enhancer elements', *Cell*, 33(2), pp. 313–314. doi: 10.1016/0092-8674(83)90410-5.

Kim, J. L., Nikolov, D. B. and Burley, S. K. (1993) 'Co-crystal structure of TBP recognizing the minor groove of a TATA element.', *Nature*, 365(6446), pp. 520–7. doi: 10.1038/365520a0.

Kim, T.-K. and Shiekhattar, R. (2015) 'Architectural and Functional Commonalities between Enhancers and Promoters.', *Cell*. Elsevier, 162(5), pp. 948–59. doi: 10.1016/j.cell.2015.08.008.

Kim, W. Y. and Dahmus, M. E. (1989) 'The major late promoter of adenovirus-2 is accurately transcribed by RNA polymerases IIO, IIA, and IIB.', *The Journal of biological chemistry*, 264(6), pp. 3169–76. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2914948 (Accessed: 15 October 2019).

Kobayashi, N., Boyer, T. G. and Berk, A. J. (1995) 'A class of activation domains interacts directly with TFIIA and stimulates TFIIA-TFIID-promoter complex assembly.', *Molecular and Cellular Biology*, 15(11), pp. 6465–6473. doi: 10.1128/MCB.15.11.6465.

Koch, F. *et al.* (2011) 'Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters', *Nature Structural & Molecular Biology*, 18(8), pp. 956–963. doi: 10.1038/nsmb.2085.

Kokubo, T. *et al.* (1994) 'Interaction between the N-terminal domain of the 230-kDa subunit and the TATA box-binding subunit of TFIID negatively regulates TATA-box binding.', *Proceedings of the National Academy of Sciences of the United States of America*, 91(9), pp. 3520–4. doi: 10.1073/pnas.91.9.3520.

Kolesnikova, O. *et al.* (2018) 'Molecular structure of promoter-bound yeast TFIID', *Nature Communications*, 9(1), p. 4666. doi: 10.1038/s41467-018-07096-y.

Kopytova, D. V. *et al.* (2006) 'Two Isoforms of Drosophila TRF2 Are Involved in Embryonic Development, Premeiotic Chromatin Condensation, and Proper Differentiation of Germ Cells of Both Sexes', *Molecular and Cellular Biology*, 26(20), pp. 7492–7505. doi: 10.1128/MCB.00349-06.

Kostrewa, D. *et al.* (2009) 'RNA polymerase II–TFIIB structure and mechanism of transcription initiation', *Nature*, 462(7271), pp. 323–330. doi: 10.1038/nature08548.

Krebs, A. R. *et al.* (2017) 'Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters.', *Molecular cell*, 67(3), pp. 411-422.e4. doi: 10.1016/j.molcel.2017.06.027.

Krumm, A. *et al.* (1992) 'The block to transcriptional elongation within the human c-myc gene is determined in the promoter-proximal region.', *Genes & Development*, 6(11), pp. 2201–2213. doi: 10.1101/gad.6.11.2201.

Kwak, H. *et al.* (2013) 'Precise maps of RNA polymerase reveal how promoters direct initiation and pausing.', *Science (New York, N.Y.)*, 339(6122), pp. 950–3. doi: 10.1126/science.1229386.

Lagha, M. *et al.* (2013) 'XPaused Pol II coordinates tissue morphogenesis in the drosophila embryo', *Cell*. doi: 10.1016/j.cell.2013.04.045.

Langelier, M. F. *et al.* (2001) 'Structural and functional interactions of transcription factor (TF) IIA with TFIIE and TFIIF in transcription initiation by RNA polymerase II.', *The Journal of biological chemistry*. PMC Canada manuscript submission, 276(42), pp. 38652–7. doi: 10.1074/jbc.M106422200.

Larsson, A. J. M. *et al.* (2019) 'Genomic encoding of transcriptional burst kinetics', *Nature*. Nature Publishing Group, 565(7738), pp. 251–254. doi: 10.1038/s41586-018-0836-1.

Lenstra, T. L. *et al.* (2016) 'Transcription Dynamics in Living Cells', *Annual Review of Biophysics*. doi: 10.1146/annurev-biophys-062215-010838.

Levsky, J. M. *et al.* (2002) 'Single-Cell Gene Expression Profiling', *Science*, 297(5582), pp. 836–840. doi: 10.1126/science.1072241.

Li, X. and Noll, M. (1994) 'Compatibility between enhancers and promoters determines the transcriptional specificity of gooseberry and gooseberry neuro in the Drosophila embryo.', *The EMBO Journal*. John Wiley & Sons, Ltd, 13(2), pp. 400–406. doi: 10.1002/j.1460-

2075.1994.tb06274.x.

Lim, C. Y. *et al.* (2004) 'The MTE, a new core promoter element for transcription by RNA polymerase II', *Genes & Development*, 18(13), pp. 1606–1617. doi: 10.1101/gad.1193404.

Lindell, T. J. *et al.* (1970) 'Specific Inhibition of Nuclear RNA Polymerase II by agr-Amanitin', *Science*, 170(3956), pp. 447–449. doi: 10.1126/science.170.3956.447.

Little, S. C., Tikhonov, M. and Gregor, T. (2013) 'Precise developmental gene expression arises from globally stochastic transcriptional activity', *Cell*. doi: 10.1016/j.cell.2013.07.025.

Louder, R. K. *et al.* (2016) 'Structure of promoter-bound TFIID and model of human pre-initiation complex assembly', *Nature*, 531(7596), pp. 604–609. doi: 10.1038/nature17394.

Lu, H. *et al.* (1991) 'The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex.', *Proceedings of the National Academy of Sciences of the United States of America*, 88(22), pp. 10004–8. doi: 10.1073/pnas.88.22.10004.

Lucas, T. *et al.* (2013) 'Live Imaging of Bicoid-Dependent Transcription in Drosophila Embryos', *Current Biology*. doi: 10.1016/j.cub.2013.08.053.

MARKOW, T. A., BEALL, S. and MATZKIN, L. M. (2009) 'Egg size, embryonic development time and ovoviviparity in Drosophila species', *Journal of Evolutionary Biology*, 22(2), pp. 430–434. doi: 10.1111/j.1420-9101.2008.01649.x.

Martianov, I., Viville, S. and Davidson, I. (2002) 'RNA Polymerase II Transcription in Murine Cells Lacking the TATA Binding Protein', *Science*, 298(5595), pp. 1036–1039. doi: 10.1126/science.1076327.

Mathis, D. J. and Chambon, P. (1981) 'The SV40 early region TATA box is required for accurate in vitro initiation of transcription', *Nature*. Nature Publishing Group, 290(5804), pp. 310–315. doi: 10.1038/290310a0.

Matsui, T. *et al.* (1980) 'Multiple factors required for accurate initiation of transcription by purified RNA polymerase II.', *The Journal of biological chemistry*, 255(24), pp. 11992–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7440580 (Accessed: 15 October 2019).

MATSUKAGE, A. *et al.* (2008) 'The DRE/DREF transcriptional regulatory system: a master key for cell proliferation', *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1779(2), pp. 81–89. doi: 10.1016/j.bbagrm.2007.11.011.

Maxon, M. E., Goodrich, J. A. and Tjian, R. (1994) 'Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance.', *Genes & Development*, 8(5), pp. 515–524. doi: 10.1101/gad.8.5.515.

Meinel, D. M. *et al.* (2013) 'Recruitment of TREX to the transcription machinery by its direct binding to the phospho-CTD of RNA polymerase II.', *PLoS genetics*. Edited by G. P. Copenhaver, 9(11), p. e1003914. doi: 10.1371/journal.pgen.1003914.

Merli, C. *et al.* (1996) 'Promoter specificity mediates the independent regulation of neighboring genes.', *Genes & Development*, 10(10), pp. 1260–1270. doi: 10.1101/gad.10.10.1260.

Moreau, P. *et al.* (1981) 'The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants', *Nucleic Acids Research*, 9(22), pp. 6047–6068. doi: 10.1093/nar/9.22.6047.

Müller, F. *et al.* (2001) 'TBP is not universally required for zygotic RNA polymerase II transcription in zebrafish', *Current Biology*, 11(4), pp. 282–287. doi: 10.1016/S0960-9822(01)00076-8.

Muse, G. W. *et al.* (2007) 'RNA polymerase is poised for activation across the genome', *Nature Genetics*, 39(12), pp. 1507–1511. doi: 10.1038/ng.2007.21.

Myers, L. C. *et al.* (1998) 'The Med proteins of yeast and their function through the RNA polymerase II carboxy-terminal domain', *Genes & Development*, 12(1), pp. 45–54. doi: 10.1101/gad.12.1.45.

Nechaev, S. *et al.* (2010) 'Global Analysis of Short RNAs Reveals Widespread Promoter-Proximal Stalling and Arrest of Pol II in Drosophila', *Science*, 327(5963), pp. 335–338. doi: 10.1126/science.1181421.

Neuberger, M. S. and Calabi, F. (1983) 'Reciprocal chromosome translocation between c-myc and immunoglobulin γ2b genes', *Nature*, 305(5931), pp. 240–243. doi: 10.1038/305240a0.

Nicolas, D., Phillips, N. E. and Naef, F. (2017) 'What shapes eukaryotic transcriptional bursting?', *Mol. BioSyst.* doi: 10.1039/C7MB00154A.

Ohler, U. *et al.* (2002) 'Computational analysis of core promoters in the Drosophila genome.', *Genome biology*. BioMed Central, 3(12), p. RESEARCH0087. doi: 10.1186/gb-2002-3-12-research0087.

Ohtsuki, S. and Levine, M. (1998) 'GAGA mediates the enhancer blocking activity of the eve promoter in the Drosophila embryo.', *Genes & development*. Cold Spring Harbor Laboratory Press, 12(21), pp. 3325–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9808619 (Accessed: 22 July 2016).

van Opijnen, T. *et al.* (2004) 'The Human Immunodeficiency Virus Type 1 Promoter Contains a CATA Box Instead of a TATA Box for Optimal Transcription and Replication', *Journal of Virology*, 78(13), pp. 6883–6890. doi: 10.1128/JVI.78.13.6883-6890.2004.

Orphanides, G., Lagrange, T. and Reinberg, D. (1996) 'The general transcription factors of RNA polymerase II.', *Genes & Development*, 10(21), pp. 2657–2683. doi: 10.1101/gad.10.21.2657.

Pal, M., Ponticelli, A. S. and Luse, D. S. (2005) 'The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II.', *Molecular cell*, 19(1), pp. 101–10. doi: 10.1016/j.molcel.2005.05.024.

Pan, G. and 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.', *The Journal of biological chemistry*, 269(48), pp. 30101–4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7982911 (Accessed: 15 October 2019).

Paré, A. *et al.* (2009) 'Visualization of Individual Scr mRNAs during Drosophila Embryogenesis Yields Evidence for Transcriptional Bursting', *Current Biology*. doi: 10.1016/j.cub.2009.10.028.

Parry, T. J. *et al.* (2010) 'The TCT motif, a key component of an RNA polymerase II transcription system for the translational machinery.', *Genes & development*, 24(18), pp. 2013–8. doi: 10.1101/gad.1951110.

Patel, A. B. *et al.* (2018) 'Structure of human TFIID and mechanism of TBP loading onto promoter DNA', *Science*, 362(6421), p. eaau8872. doi: 10.1126/science.aau8872.

Patwardhan, R. P. *et al.* (2009) 'High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis.', *Nature biotechnology*. NIH Public Access, 27(12), pp. 1173–5. doi: 10.1038/nbt.1589.

Pichon, X. *et al.* (2018) 'A Growing Toolbox to Image Gene Expression in Single Cells: Sensitive Approaches for Demanding Challenges', *Molecular Cell*. doi: 10.1016/j.molcel.2018.07.022.

Podos, S. D. and Ferguson, E. L. (1999) 'Morphogen gradients: new insights from DPP.', *Trends in genetics : TIG*, 15(10), pp. 396–402. doi: 10.1016/s0168-9525(99)01854-5.

Proudfoot, N. J. (2016) 'Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut.', *Science (New York, N.Y.)*. Europe PMC Funders, 352(6291), p. aad9926. doi: 10.1126/science.aad9926.

Rach, E. A. *et al.* (2009) 'Motif composition, conservation and condition-specificity of single and alternative transcription start sites in the Drosophila genome.', *Genome biology*, 10(7), p. R73. doi: 10.1186/gb-2009-10-7-r73.

Raj, A. *et al.* (2006) 'Stochastic mRNA Synthesis in Mammalian Cells', *PLoS Biology*. Edited by U. Schibler, 4(10), p. e309. doi: 10.1371/journal.pbio.0040309.

Raj, A. and van Oudenaarden, A. (2008) 'Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences', *Cell*, 135(2), pp. 216–226. doi: 10.1016/j.cell.2008.09.050.

Raser, J. M. (2004) 'Control of Stochasticity in Eukaryotic Gene Expression', *Science*, 304(5678), pp. 1811–1814. doi: 10.1126/science.1098641.

Raser, J. M. (2005) 'Noise in Gene Expression: Origins, Consequences, and Control', *Science*, 309(5743), pp. 2010–2013. doi: 10.1126/science.1105891.

Reeve, J. N. (2003) 'Archaeal chromatin and transcription.', *Molecular microbiology*, 48(3), pp. 587–98. doi: 10.1046/j.1365-2958.2003.03439.x.

Ren, D., Lei, L. and Burton, Z. F. (1999) 'A region within the RAP74 subunit of human transcription factor IIF is critical for initiation but dispensable for complex assembly.', *Molecular and cellular biology*, 19(11), pp. 7377–87. doi: 10.1128/mcb.19.11.7377.

Rennie, S. *et al.* (2018) 'Transcription start site analysis reveals widespread divergent transcription in D. melanogaster and core promoter-encoded enhancer activities', *Nucleic Acids Research*, 46(11), pp. 5455–5469. doi: 10.1093/nar/gky244.

De Renzis, S. *et al.* (2007) 'Unmasking activation of the zygotic genome using chromosomal deletions in the Drosophila embryo.', *PLoS biology*. Edited by T. Kornberg, 5(5), p. e117. doi: 10.1371/journal.pbio.0050117.

Roeder, R. G. and Rutter, W. J. (1969) 'Multiple Forms of DNA-dependent RNA Polymerase in Eukaryotic Organisms', *Nature*. Nature Publishing Group, 224(5216), pp. 234–237. doi: 10.1038/224234a0.

Rosonina, E., Kaneko, S. and Manley, J. L. (2006) 'Terminating the transcript: breaking up is hard to do.', *Genes & development*. Cold Spring Harbor Laboratory Press, 20(9), pp. 1050–6. doi: 10.1101/gad.1431606.

Roth, S., Stein, D. and Nüsslein-Volhard, C. (1989) 'A gradient of nuclear localization of the dorsal protein determines dorsoventral pattern in the Drosophila embryo', *Cell*, 59(6), pp. 1189–1202. doi: 10.1016/0092-8674(89)90774-5.

Rushlow, C. A. *et al.* (1989) 'The graded distribution of the dorsal morphogen is initiated by selective nuclear transport in Drosophila', *Cell*, 59(6), pp. 1165–1177. doi: 10.1016/0092-8674(89)90772-1.

Sainsbury, S., Bernecky, C. and Cramer, P. (2015) 'Structural basis of transcription initiation by RNA polymerase II', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 16(3), pp. 129–143. doi: 10.1038/nrm3952.

Sainsbury, S., Niesser, J. and Cramer, P. (2013) 'Structure and function of the initially transcribing RNA polymerase II–TFIIB complex', *Nature*, 493(7432), pp. 437–440. doi: 10.1038/nature11715.

Santa, F. De *et al.* (2010) 'A Large Fraction of Extragenic RNA Pol II Transcription Sites Overlap Enhancers', *PLoS Biology*. Public Library of Science, 8(5), p. e1000384. doi: 10.1371/JOURNAL.PBIO.1000384.

Saponaro, M. *et al.* (2014) 'RECQL5 Controls Transcript Elongation and Suppresses Genome Instability Associated with Transcription Stress', *Cell*, 157(5), pp. 1037–1049. doi: 10.1016/j.cell.2014.03.048.

Savinkova, L. *et al.* (2013) 'An Experimental Verification of the Predicted Effects of Promoter TATA-Box Polymorphisms Associated with Human Diseases on Interactions between the TATA Boxes and TATA-Binding Protein', *PLoS ONE*. Edited by M. Gasset, 8(2), p. e54626. doi: 10.1371/journal.pone.0054626.

Schor, I. E. *et al.* (2017) 'Promoter shape varies across populations and affects promoter evolution and expression noise', *Nature Genetics*. Nature Publishing Group, 49(4), pp. 550–558. doi: 10.1038/ng.3791.

Schulz, K. N. and Harrison, M. M. (2019) 'Mechanisms regulating zygotic genome activation', *Nature Reviews Genetics*. Nature Publishing Group, pp. 221–234. doi: 10.1038/s41576-018-0087-x.

Seila, A. C. *et al.* (2008) 'Divergent Transcription from Active Promoters', *Science*, 322(5909), pp. 1849–1851. doi: 10.1126/science.1162253.

Senecal, A. *et al.* (2014) 'Transcription factors modulate c-Fos transcriptional bursts', *Cell Reports*. The Authors, 8(1), pp. 75–83. doi: 10.1016/j.celrep.2014.05.053.

Shao, W., Alcantara, S. G.-M. and Zeitlinger, J. (2019) 'Reporter-ChIP-nexus reveals strong contribution of the Drosophila initiator sequence to RNA polymerase pausing', *eLife*, 8. doi: 10.7554/eLife.41461.

Shao, W. and Zeitlinger, J. (2017) 'Paused RNA polymerase II inhibits new transcriptional initiation', *Nature GeNetics VOLUME*, 49(7). doi: 10.1038/ng.3867.

Shiraki, T. *et al.* (2003) 'Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage.', *Proceedings of the National Academy of Sciences of the United States of America*, 100(26), pp. 15776–81. doi: 10.1073/pnas.2136655100.

Shlyueva, D., Stampfel, G. and Stark, A. (2014) 'Transcriptional enhancers: from properties to genome-wide predictions', *Nature Reviews Genetics*, 15(4), pp. 272–286. doi: 10.1038/nrg3682.

Sloutskin, A. *et al.* (2015) 'ElemeNT: a computational tool for detecting core promoter elements', *Transcription*, 6(3), pp. 41–50. doi: 10.1080/21541264.2015.1067286.

Smale, S. T. and Baltimore, D. (1989) 'The "initiator" as a transcription control element', *Cell*, 57(1), pp. 103–113. doi: 10.1016/0092-8674(89)90176-1.

Smale, S. T. and Kadonaga, J. T. (2003) 'The RNA polymerase II core promoter.', *Annual review of biochemistry*, 72(1), pp. 449–79. doi: 10.1146/annurev.biochem.72.121801.161520.

Soutourina, J. (2018) 'Transcription regulation by the Mediator complex', *Nature Reviews Molecular Cell Biology*, 19(4), pp. 262–274. doi: 10.1038/nrm.2017.115.

Stathopoulos, A. *et al.* (2002) 'Whole-genome analysis of dorsal-ventral patterning in the Drosophila embryo.', *Cell*, 111(5), pp. 687–701. doi: 10.1016/s0092-8674(02)01087-5.

Steurer, B. *et al.* (2018) 'Live-cell analysis of endogenous GFP-RPB1 uncovers rapid turnover of initiating and promoter-paused RNA Polymerase II', *Proceedings of the National Academy of Sciences*. National Academy of Sciences, p. 201717920. doi: 10.1073/PNAS.1717920115.

Struhl, K. and Moqtaderi, Z. (1998) 'The TAFs in the HAT', *Cell*, 94(1), pp. 1–4. doi: 10.1016/S0092-8674(00)81213-1.

Suter, D. M. *et al.* (2011) 'Mammalian genes are transcribed with widely different bursting kinetics', *Science*. doi: 10.1126/science.1198817.

Swain, P. S., Elowitz, M. B. and Siggia, E. D. (2002) 'Intrinsic and extrinsic contributions to stochasticity in gene expression', *Proceedings of the National Academy of Sciences*, 99(20), pp. 12795–12800. doi: 10.1073/pnas.162041399.

Tadros, W. *et al.* (2003) 'Regulation of maternal transcript destabilization during egg activation in Drosophila.', *Genetics*, 164(3), pp. 989–1001. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12871909 (Accessed: 15 October 2019).

Tadros, W., Westwood, J. T. and Lipshitz, H. D. (2007) 'The Mother-to-Child Transition', *Developmental Cell*, 12(6), pp. 847–849. doi: 10.1016/j.devcel.2007.05.009.

Tantale, K. *et al.* (2016) 'A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting', *Nature Communications*. doi: 10.1038/ncomms12248.

Theisen, J. W. M., Lim, C. Y. and Kadonaga, J. T. (2010) 'Three key subregions contribute to the function of the downstream RNA polymerase II core promoter.', *Molecular and cellular biology*, 30(14), pp. 3471–9. doi: 10.1128/MCB.00053-10.

Thomas, M. C. and Chiang, C.-M. (2006) 'The general transcription machinery and general cofactors.', *Critical reviews in biochemistry and molecular biology*, 41(3), pp. 105–78. doi: 10.1080/10409230600648736.

Thomsen, S. *et al.* (2010) 'Genome-wide analysis of mRNA decay patterns during early Drosophila development', *Genome Biology*, 11(9), p. R93. doi: 10.1186/gb-2010-11-9-r93.

Torres-Padilla, M. E. and Tora, L. (2007) 'TBP homologues in embryo transcription: who does what?', *EMBO reports*. European Molecular Biology Organization, 8(11), pp. 1016–8. doi: 10.1038/sj.embor.7401093.

Tsanov, N. *et al.* (2016) 'SmiFISH and FISH-quant - A flexible single RNA detection approach with super-resolution capability', *Nucleic Acids Research*, 44(22). doi: 10.1093/nar/gkw784.

Tutucci, E., Vera, M., *et al.* (2018) 'An improved MS2 system for accurate reporting of the mRNA life cycle.', *Nature methods*. NIH Public Access, 15(1), pp. 81–89. doi: 10.1038/nmeth.4502.

Tutucci, E., Livingston, N. M., *et al.* (2018) 'Imaging mRNA In Vivo, from Birth to Death', *Annual Review of Biophysics*. Annual Reviews , 47(1), pp. 85–106. doi: 10.1146/annurev-biophys-070317-033037.

Vastenhouw, N. L., Cao, W. X. and Lipshitz, H. D. (2019) 'The maternal-to-zygotic transition revisited', *Development (Cambridge, England)*, 146(11). doi: 10.1242/dev.161471.

Veenstra, G. J., Weeks, D. L. and Wolffe, A. P. (2000) 'Distinct roles for TBP and TBP-like factor in early embryonic gene transcription in Xenopus.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 290(5500), pp. 2312–5. doi: 10.1126/science.290.5500.2312.

Vo Ngoc, L. *et al.* (2017) 'The punctilious RNA polymerase II core promoter.', *Genes & development*, 31(13), pp. 1289–1301. doi: 10.1101/gad.303149.117.

Vo Ngoc, L., Kassavetis, G. A. and Kadonaga, J. T. (2019) 'The RNA Polymerase II Core Promoter in *Drosophila*', *Genetics*, 212(1), pp. 13–24. doi: 10.1534/genetics.119.302021.

Wang, H. D., Trivedi, A. and Johnson, D. L. (1997) 'Hepatitis B virus X protein induces RNA polymerase III-dependent gene transcription and increases cellular TATA-binding protein by activating the Ras signaling pathway.', *Molecular and Cellular Biology*, 17(12), pp. 6838–6846. doi: 10.1128/MCB.17.12.6838.

Weinmann, R. and Roeder, R. G. (1974) 'Role of DNA-Dependent RNA Polymerase III in the

Transcription of the tRNA and 5S RNA Genes', *Proceedings of the National Academy of Sciences*, 71(5), pp. 1790–1794. doi: 10.1073/pnas.71.5.1790.

Whyte, W. A. *et al.* (2013) 'Master transcription factors and mediator establish super-enhancers at key cell identity genes.', *Cell*, 153(2), pp. 307–19. doi: 10.1016/j.cell.2013.03.035.

Wong, J. M. and Bateman, E. (1994) 'TBP-DNA interactions in the minor groove discriminate between A:T and T:A base pairs', *Nucleic Acids Research*, 22(10), pp. 1890–1896. doi: 10.1093/nar/22.10.1890.

Yamaguchi, Y., Shibata, H. and Handa, H. (2013) 'Transcription elongation factors DSIF and NELF: promoter-proximal pausing and beyond.', *Biochimica et biophysica acta*, 1829(1), pp. 98–104. doi: 10.1016/j.bbagrm.2012.11.007.

Zabidi, M. A. *et al.* (2015) 'Enhancer–core-promoter specificity separates developmental and housekeeping gene regulation', *Nature*. Nature Publishing Group, 518(7540), pp. 556–559. doi: 10.1038/nature13994.

Zawel, L. and Reinberg, D. (1992) 'Advances in RNA polymerase II transcription', *Current Opinion in Cell Biology*, 4(3), pp. 488–495. doi: 10.1016/0955-0674(92)90016-6.

Zehavi, Y. *et al.* (2014) 'The core promoter composition establishes a new dimension in developmental gene networks', *Nucleus*, 5(4), pp. 298–303. doi: 10.4161/nucl.29838.

Zehring, W. A. *et al.* (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 of the United States of America*, 85(11), pp. 3698–702. doi: 10.1073/pnas.85.11.3698.

Zeitlinger, J. *et al.* (2007) 'Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the Drosophila embryo', *Genes & Development*, 21(4), pp. 385–390. doi: 10.1101/gad.1509607.

Zenklusen, D., Larson, D. R. and Singer, R. H. (2008) 'Single-RNA counting reveals alternative modes of gene expression in yeast', *Nature Structural & Molecular Biology*, 15(12), pp. 1263–1271. doi: 10.1038/nsmb.1514.

Zhang, Z. *et al.* (2016) 'Rapid dynamics of general transcription factor TFIIB binding during preinitiation complex assembly revealed by single-molecule analysis.', *Genes & development*, 30(18), pp. 2106–2118. doi: 10.1101/gad.285395.116.

Zhao, X. and Herr, W. (2002) 'A regulated two-step mechanism of TBP binding to DNA: a solvent-exposed surface of TBP inhibits TATA box recognition.', *Cell*, 108(5), pp. 615–27. doi: 10.1016/s0092-8674(02)00648-7.

Zhu, A. and Kuziora, M. A. (1996) 'Homeodomain Interaction with the β Subunit of the General Transcription Factor TFIIE', *Journal of Biological Chemistry*, 271(35), pp. 20993–20996. doi: 10.1074/jbc.271.35.20993.