Title

The nucleosome as a barrier for transcription and control of gene expression

Author

Cheng, Enze

Publication Date

2024

Peer reviewed|Thesis/dissertation

The nucleosome as a barrier for transcription and control of gene expression

by

Enze Cheng

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Carlos Bustamante, Chair

Professor Hernan Garcia

Professor Susan Marqusee

Summer 2024

The nucleosome as a barrier for transcription and control of gene expression

Abstract

The nucleosome as a barrier for transcription and control of gene expression

by

Enze Cheng

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Carlos Bustamante, Chair

Gene expression in eukaryotic cells is an intricate, multi-step process that allows the information encoded in DNA to be translated into proteins, which perform most of the critical functions within the cell. This process is highly regulated at various levels to ensure precise control over which genes are expressed, when, where, and to what extent. The primary stages of gene expression encompass transcription, RNA processing, messenger RNA (mRNA) export, and translation. Among these, transcription within eukaryotic cells is also a complex and tightly regulated process which involves a sophisticated interplay between the transcription machinery, regulatory proteins, chromatin structure, and RNA processing. Gaining a deep understanding of these processes is crucial for unraveling the mechanisms of gene expression regulation and its implications for diseases such as cancer, where normal regulatory mechanisms are frequently disrupted.

In this work, we reconstituted nucleosomes and active elongation complex with purified proteins, revealing that eukaryotic RNA polymerase II (RNAP) alone struggles significantly with nucleosome clearance, a finding that contrasts with *in vivo* observations. To further explore this phenomenon, we employed cryogenic electron tomography (cryo-ET), an advanced imaging technique, to capture intermediate states of RNAP as it transcribes through a nucleosome. Lastly, we examined transcription across nucleosome positioning sequences (NPS) in vivo using two-color confocal microscopy in the fruit fly embryos and assessed nucleosome occupancy employing ATAC-seq (assay for transposase-accessible chromatin with sequencing), revealing that these sequences do not significantly slow down transcription nor position nucleosomes with the same accuracy in vivo.

Dedication

To mom and dad, without whom none of these would have been possible.

Acknowledgments

My PhD was a seven-year adventure, a period of both personal and academic growth. It is with deep gratitude that I acknowledge the invaluable support and contributions of a myriad of individuals without whom I wouldn't be able to survive this journey.

First and foremost, I extend my heartfelt thanks to my advisor, Carlos Bustamante, whose guidance, patience, and insight were instrumental in shaping me into the scientist I am today. All of this started with an email when I was an exchange student in Berkeley. Over the years, his generous mentorship and steadfast support enabled me to explore and achieve heights I once thought unattainable.

Hernan Garcia and Susan Marqusee were fantastic thesis committee members who provided invaluable constructive criticism, insightful feedback, and encouraging words that were so profoundly meaningful to me at every stage of my research. I started to collaborate with Hernan when I was a second-year graduate student. I learned so much both from him and his lab. Without whom, I wouldn't be able to study transcription in vivo with confocal microscopy. On a similar note, Ahmet Yildiz, a remarkable member of my qualifying exam committee, brought a wealth of expertise and critical insights that transformed the daunting qualifying exam into an immensely valuable experience.

A special note of thanks goes to the Physics department and its administrative staff, particularly Joelle Miles, Marjani Jones, and Donna Sakima, whose assistance and support facilitated a smooth academic and administrative journey. Their kindness and efficiency have not gone unnoticed.

The Bustamante lab was such an amazing group to work with, and I cannot get enough of doing science, and hanging out with them along the way. I first want to thank my mentor when I was an exchange student in the lab, Liang Meng Wee. We worked together for quite a while, and I learned so much from him on how to do rigorous science. I am also grateful for working with such an amazing team of scientists: Cristhian Cañari, César Díaz, Alex Tong, and Meng Zhang. Starting as a physicist, I had so little knowledge of biology. Wee, Cristhan and César guided me and trained me on conducting biochemistry experiments. Alex and Meng played an important role in the data analysis of various projects. We worked together on different types of data collected with multiple instruments and I learned a lot of statistics and programming from them.

As my graduate career progressed, my fascination with studying transcription within living cells deepened. I am immensely grateful for the opportunity to collaborate with the Garcia Lab, which significantly enhanced my understanding and skills in working with fruit flies and measuring transcription elongation in embryos, a central theme of Chapter 4. This project would not have been possible without the invaluable support and expertise of Prof. Hernan Garcia. His knowledge of fly embryogenesis and live cell

imaging was crucial in designing the experiments. I extend my deepest thanks to Prof. Garcia for allowing me to collaborate closely with his students, from whom, especially Jonathan Liu and Jiaxi Zhao, I learned essential techniques in fly handling and live cell imaging.

My exploration into transcription in fly embryos led me to appreciate the crucial role of genomic studies in deepening our understanding of transcription processes. This realization was greatly enhanced by the opportunity to collaborate with Prof. Shelby Blythe (Northwestern University). Under his guidance, detailed in Chapter 5, I learned the intricate processes of next-generation sequencing, including extracting genomic DNA from fly embryos and preparing ATAC-seq libraries. Prof. Blythe's enthusiasm and expertise were invaluable; he not only trained me in experimental techniques but also engaged in extensive discussions about Drosophila genomics and sequencing methods, significantly broadening my knowledge and skills in this field.

Grad school is one of the most challenging times in my life, but with the help from my lab colleagues, I am able to survive. I would like to thank them in no particular order: Hossein Amiri, Lisa Alexander, Amir Bitran, Francesca Burgos, Zhijie Chen, Varsha Desai, Alfredo Flores, Ronen Gabizon, Omar Herrera, Shantanu Kadam, Sudeesh Krishnamurthy, JeongHoon Kim, Chen Li, Antony Lee, Katherinne Requejo Roque, Wenxia Lin, Bibiana Onoa, Maurizio Righini, Rohit Satija, Alan Shaw, Shannon Yan, Mina Sun, Sara Tafoya, John Van Patten, Robert Sosa, Jason Wong, Errol Watson, Guillermo Chacaltana, Ledet Getachew, Pauline Cheng, and Natalie Thung.

Last but not least, I extend my deepest gratitude to my parents, Jinzhao and Haifeng. Being far from home has only deepened my appreciation for their presence in my life. Their unwavering support in every aspect of my journey has been invaluable. I could not have achieved any of this without their endless encouragement and love. And finally, thank you to you, the reader, who downloaded and read through this extensive document. Your interest and time are greatly appreciated.

Contents

List of Figures

Chapter 1

Introduction

If the 20th century was celebrated as the golden age of physics, the 21st century is shaping up to be the golden age of biology. However, the traditional qualitative approaches no longer meet the sophisticated demands of current research. Instead, quantitative studies are increasingly critical in advancing our understanding of biology. With groundbreaking developments such as super-resolution microscopy, clustered regularly interspaced short palindromic repeats (CRISPR), and next-generation sequencing techniques, we now possess unprecedented capabilities to visualize, manipulate, and engineer biological systems like never before.

From a physicist's perspective, biology offers incredibly rich systems for the application of quantitative methods to unveil the underlying mechanisms. Within this broad spectrum, transcription stands out as a particularly intriguing subject. As a crucial aspect of the central dogma, it represents a domain where a quantitative grasp of its highly regulated and complex process is essential.

1.1 Eukaryotic RNA polymerase II

The eukaryotic RNAP occupies a central role in the transcriptional machinery, tasked with the conversion of protein-coding genes into mRNA transcripts and non-coding genes into non-coding RNA (ncRNA). This enzyme is distinguished by its composition: a 12-subunit structure that includes a 10-subunit core along with a peripheral heterodimer made up of subunits Rpb4 and Rpb7. In contrast, RNA polymerase I has 14 subunits and RNA polymerase III has 17 subunits. This configuration is conserved across a wide range of eukaryotes and sets RNAP II apart from its counterparts, RNA polymerase I and III, by its unique DNA template specificity, regulatory factors, and operational mechanisms (Khatter et al., 2017; Engel er al., 2018). The groundbreaking discovery of three chromatographically distinct RNAP in eukaryotes (Roeder and Rutter, 1969), contrasted with the presence of only one RNA polymerase in prokaryotes, suggests that mechanisms and regulations of transcription might be fundamentally different between prokaryotes and eukaryotes. Notably, the largest subunit, RPB1, features the carboxy-terminal domain (CTD), a hallmark of RNAP essential for its functionality. The CTD is characterized by repeat sequences of seven amino acids (Tyr-Ser-Pro-Thr-Ser-Pro-Ser, or YSPTSPS) that are subject to extensive post-translational modifications, thereby regulating the enzyme's interactions with an array of transcriptional and RNA processing factors. The purification of RNA polymerase I, II,

and III to homogeneity in 1975 marked a turning point (Sklar et al., 1975), facilitating in-depth structural and functional analyses. However, the amino acid sequences of these enzymes remained elusive until the 1990s, when advancements in biochemistry, molecular biology, cloning, and genomic sequencing technologies brought them to light (Young, 1991). The structural elucidation of RNAP began with the isolation of the active enzyme from yeast, leading to the first high-resolution structure of the 10-subunit yeast core RNAP obtained through X-ray crystallography at a resolution of 3 angstroms (Cramer et al., 2000). Subsequently, a more detailed structure of the complete 12-subunit RNAP holoenzyme was resolved to 2.8 angstroms (Cramer et al, 2001), offering profound insights into its architecture and function, as illustrated in Figure 1.1.

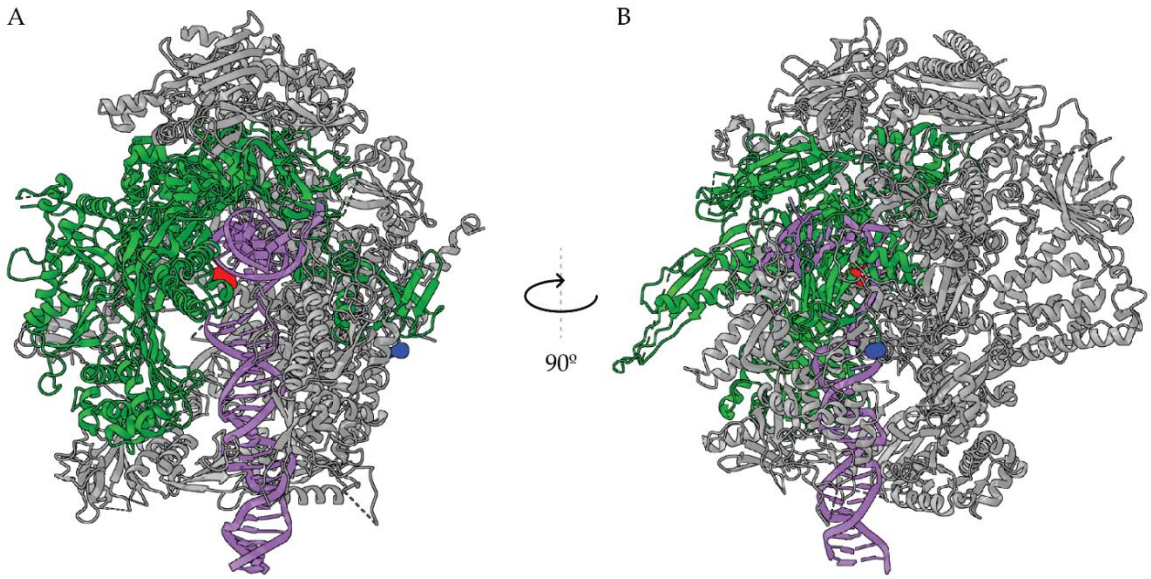

Figure 1.1: Three-dimensional structure of the yeast 12-subunit RNAP holoenzyme. (A) Surface view of the yeast RNAP holoenzyme. (B) Side view of the yeast RNAP holoenzyme. In both panels, the largest subunit, RPB1, is highlighted in green. The DNA template is shown in magenta. The red dot marks the catalytic site, while the blue dot indicates the start of the RPB1 CTD, an unstructured domain for which detailed structural information is not available.

1.2 Eukaryotic transcription regulation

The regulation of eukaryotic transcription is significantly more complex than that of prokaryotic transcription, involving a vast array of proteins. This complexity includes general transcription factors (GTFs), specific transcription factors, chromatin remodelers, histone modifying enzymes, the mediator complex, and enzymes that modify the RNAP CTD, such as kinases and phosphatases. Each of these components plays a critical role in finely tuning the expression of genes, reflecting the intricate regulatory mechanisms required to support the diverse functions and developmental processes of eukaryotic

cells. Eukaryotic transcription, similar to its prokaryotic counterpart (overviewed in Figure 1.2), is separated into three major phases: initiation, elongation, and termination. A schematic of eukaryotic transcription is overviewed in Figure 1.2.

Initiation begins with the assembly of the pre-initiation complex (PIC) at the promoter, followed by open complex formation and initiation. The recruitment of coactivators and GTFs to the promoter region, along with the modification of chromatin structure by chromatin remodeling proteins, sets the stage for productive transcription. These processes, which follow the establishment of a functional PIC, facilitate the transition from a paused RNAP state to a phase of productive initiation and subsequent elongation of the RNA transcript. In the absence of these steps, RNAP may enter abortive initiation, disengage from the DNA, and be recycled. Upon promoter clearance, the nascent mRNA is capped at the 5′ end. The cycle progresses to elongation, where RNAP transcribes the gene, and navigates splicing checkpoints and ultimately reaches termination upon completing the open reading frame (ORF). This leads to cleavage and polyadenylation of the pre-mRNA. Post-termination, the transcription machinery can be reassembled for subsequent rounds of transcription, with the assistance of factors like CPSF and CstF (Jong Chan Hong, 2016).

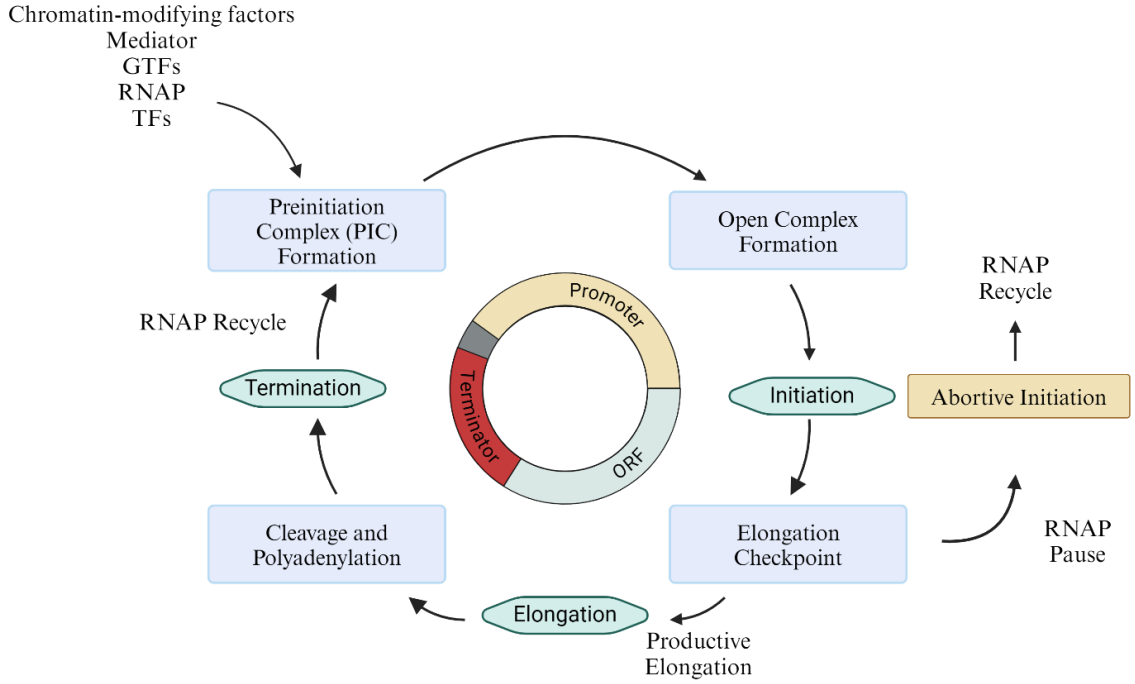

Figure 1.2: Eukaryotic transcription cycle. The main phases of transcription are highlighted in cyan and important regulation steps are indicated in blue. Additionally, abortive initiation is an alternative pathway of the initiation phase. The circle in the middle represents the location of events in relation to their position on the gene. Created with BioRender.com.

1.2.1 Initiation

During initiation, an activator identifies and attaches to an enhancer sequence located upstream of the promoter. This interaction facilitates the sequential recruitment of GTFs (Lin and Green, 1991; Kuras and Struhl, 1999; Li et al., 1999), including TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH, along with RNAP, in the following manner:

1) TFIID initially binds to the promoter region, predominantly via its TATA-binding protein (TBP) component, which specifically recognizes promoter element known as the TATA box (Nakajima et al., 1988; Buratowski et al., 1989).
2) TFIIA joins the complex and stabilizes TFIID's binding to the DNA (Lee et al., 1992; Imbalzano et al., 1994).
3) TFIIB is recruited next, binding to both TFIID and the DNA. It precisely defines the distance from the TATA box to the transcription start site (TSS) by facilitating the contact between the TATA DNA and RNAP. This interaction ensures that RNAP is positioned at the correct distance from the enzyme's active center, critical for initiating transcription effectively (Kerstin et al., 1996).
4) RNAP, guided by TFIIB, then attaches to the complex, positioned to begin RNA synthesis (Kerstin et al., 1996).
5) TFIIF is already associated with RNAP and helps stabilize its interaction with TFIIB and TFIID (Tan et al., 1994).
6) TFIIE and TFIIH are the last to join. TFIIE, through its direct interaction with RNAP (Kerstin et al., 1996), facilitates the recruitment of TFIIH (Maxon et al., 1994), which possesses helicase activity to unwind the DNA and kinase activity essential for phosphorylating RNAP CTD (Ohkuma et al., 1995), triggering the transition from initiation to elongation.

Together, these components assemble into the stable (closed) form of the PIC at the promoter. In the presence of ATP and other ribonucleoside triphosphates, TFIIH utilizes ATP hydrolysis to facilitate the melting of a ~10 base pair region upstream of the TSS. This action leads to the formation of an activated, yet unstable (open), PIC (Dvir et al., 1996; Holstege et al., 1996), which is then followed by the initiation of transcription and promoter clearance (Jiang et al., 1996).

After the formation of the open PIC and the start of RNA synthesis, RNAP must clear the promoter. Initially, during initiation, the CTD of RNAP is largely unphosphorylated. The escape of RNAP from promoter involves the phosphorylation of the CTD, particularly at serine 5 residues by TFIIH, which reduces its affinity for promoter-bound complexes (Dvir et al., 1997) and recruits capping enzyme to the nascent RNA (Schroeder et al., 2000; Komarnitsky et al., 2000).

1.2.2 Elongation

As RNAP moves away from the promoter and begins synthesizing RNA, there is a transition phase where initiation factors are released or exchanged for elongation factors (Roeder, 2005). Promoter-proximal pausing occurs in this transition, where RANP experiences a temporary halt (Rougvie and Lis, 1988; Rasmussen and Lis, 1993; Rougvie and Lis, 1990). This pausing serves as a crucial checkpoint, allowing for the integration of regulatory signals before proceeding with productive transcription elongation (Wen and Shatkin, 1999; Lindstrom et al., 2003; Mandal et al., 2004). The release of RANP from this paused state is primarily mediated by the Positive Transcription Elongation Factor b (P-TEFb), which phosphorylates the Ser2 residues on the CTD of RANP (Yang et al., 2001; Nguyen et al., 2001). P-TEFb also targets the negative elongation factors, Negative Elongation Factor (NELF) and DRB Sensitivity-Inducing Factor (DSIF), with phosphorylation converting DSIF into a factor that positively influences elongation (Peterlin and Price, 2006).

Various DNA-binding proteins can influence elongation by interacting directly with RNAP (Sims III et al., 2004) or by modifying the chromatin landscape (Bowman and Poirier, 2015). Transcription factors that bind to enhancer or silencer regions can recruit coactivators or corepressors, influencing RNAP's elongation efficiency (Lambert et al., 2018). The phosphorylation of Ser2 acts as a beacon, signaling the recruitment of a suite of elongation factors that enhance RANP's processivity (Ho and Shuman, 1999; Komarnitsky et al., 2000; Ahn et al., 2004; Bowman and Kelly, 2014; González-Jiménez et al., 2021). These factors play instrumental roles in modulating the chromatin landscape ahead of RNAP, ensuring efficient and accurate transcription. Among these, SPT6 (Endoh et al., 2004; Ardehali et al., 2009), TFIIS (Kulish and Struhl, 2001; Ishibashi et al., 2014; Schweikhard et al., 2014), and FACT (Orphanides et al., 1998; Hsieh et al., 2013; Jeronimo et al., 2021; Ehara et al., 2022) are key players that contribute to stabilizing RNAP on the DNA template, enabling smooth transcription through nucleosomal barriers, and rectifying errors in RNA synthesis, respectively.

As RANP embarks on the elongation phase, it encounters a further layer of regulatory complexity. The chromatin environment, marked by specific histone modifications and the strategic placement of nucleosomes, along with DNA-binding proteins, exerts a substantial influence on RNAP's transcriptional velocity (Knezetic and Luse, 1986; Lorch et al., 1987; Chen et al., 2019). During elongation, histone tails undergo various post-translational modifications, such as methylation, acetylation, phosphorylation, and ubiquitination (Bowman and Poirier, 2014). These modifications can either facilitate or hinder RNAP's progress by altering chromatin's compactness and accessibility. For example, acetylation of histone tails generally promotes a more open chromatin structure, enhancing transcriptional activity (Li et al., 2007). Nucleosomes can act as natural barriers to RNAP (Kireeva et al., 2005). Their strategic placement along the DNA and the remodeling of chromatin by ATP-dependent chromatin remodelers play critical roles in regulating transcription elongation (Singl et al., 2021; Singh and Mueller-Planitz, 2021). Remodelers can reposition nucleosomes, temporarily evict them, or facilitate the incorporation of histone variants, thus modulating RNAP's access to the DNA template (Becker and Workman, 2013; Li et al., 2015).

In essence, the transition of RNAP from initiation through promoter-proximal pausing to the elongation phase embodies a complex regulatory framework. This intricately orchestrated process ensures that transcription elongation is not only a matter of RANP traversing the DNA but is a finely tuned response to a plethora of cellular signals and chromatin states, culminating in the precise expression of genetic information.

1.2.3 Termination

Transcription termination is a critical phase in the process of gene expression, marking the conclusion of transcription by RANP. This phase ensures the proper release of the newly synthesized pre-messenger RNA (pre-mRNA) from the DNA template and the disengagement of RNAP from the DNA, preventing potential interference with the transcription of downstream genes (Greger et al., 2000) and allowing the enzyme to be recycled for future transcription events. Transcription termination is intricately regulated and requires a coordinated interplay of various molecular factors to achieve precise control over the end of transcription (Richard and Manley, 2009).

The termination of transcription in eukaryotes involves several key steps and mechanisms, distinguishing it from the simpler processes observed in prokaryotes (Kuehner et al., 2011; Ray-coni et al., 2016; Roberts, 2019). One central aspect of eukaryotic termination is the coupling with RNA processing events (Proudfoot et al., 2002; Buratowski, 2005), particularly cleavage (Dye and Proudfoot, 2001; Teixeira et al., 2004; Nabavi and Nazar, 2008)) and polyadenylation of the pre-mRNA (Whitelaw and Proudfoot, 1986; Logan et al., 1987; Connelly and Manley, 1988; Tian and Graber, 2011), which are crucial for the maturation of mRNA and its subsequent translation into proteins. This coupling ensures that the pre-mRNA is properly processed and polyadenylated before being released, contributing to the stability and translational efficiency of the mRNA.

Two main models describe the mechanisms of transcription termination in eukaryotes: the allosteric model and the torpedo model (reviewed in Rosonina et al., 2006). The allosteric model proposes that changes in the structure and composition of the transcription complex, triggered by specific sequences or processing events, lead to the termination of transcription. Conversely, the torpedo model suggests that the exonucleolytic degradation of the nascent RNA downstream of the cleavage site by specific exonucleases catches up to the slowing Pol II, promoting its release from the DNA.

The termination process also involves specific protein factors, such as the cleavage and polyadenylation specificity factor (CPSF) and the cleavage stimulation factor (CstF), which recognize signal sequences in the pre-mRNA and initiate the cleavage and polyadenylation process (Kaufmann et al., 2004; Lackford et al., 2014; Li et al., 2015; Takagaki et al., 1990; Takagaki and Manley, 1997; Perez Canadillas and Varani, 2003). Additionally, termination factors like Rat1/Xrn2 exonuclease in the torpedo model play

crucial roles in ensuring efficient termination and RNAP recycling (Kim et al., 2004; Luo et al., 2006; Chalamcharla et al., 2015; West et al., 2004; Brannan et al., 2012; Eaton et al., 2018).

1.3 Chromatin Structure

In eukaryotic cells, DNA is well protected and packaged into higher-order structure by the fundamental building block known as nucleosome. Previous research has revealed that nucleosome packaging of the DNA template influences every aspect of transcription, from initiation through elongation to termination (Workman and Kingston, 1998). The dynamic nature of chromatin structure allows for the intricate regulation of gene expression, ensuring that genes are expressed at the right times and in the correct cell types and also suggests that the principles and mechanisms of transcription on naked DNA, as understood from prokaryotic studies, may not be directly applicable in this context.

1.3.1 Nucleosome structure

At its simplest level, chromatin is organized into the nucleosome, the basic unit of chromatin structure (Figure 1.3). The nucleosome core is composed of ~ 147 base pairs of DNA wrapped in 1.65 left-handed turns around a histone octamer (two each of H2A, H2B, H3, and H4) known as 'beads-on-a-string' (Olins and Olins, 1974), establishing 14 contact points between histone proteins and DNA (Kornberg, 1974; Luger et al., 1997). A linker DNA of variable length, associated with the H1 histone, connects adjacent nucleosomes. Within the nucleosome, the positively charged lysine and arginine residues of histone proteins form electrostatic interaction with the negatively charged phosphate backbone of the DNA, occurring at intervals of 10 base pairs (Kujirai and Kurumizaka, 2020). The resulting nucleosome structure is regarded as the first level of hierarchical packing of the genome. This packaging of DNA protects it against damaging while simultaneously restricting the accessibility of the underlying DNA sequences. Previous biochemical and biophysical studies have demonstrated that a nucleosome downstream of the elongating RNAP stalls transcription extensively (Shaw et al., 1978; Izban and Luse, 1992; Bednar et al., 1999; Bondarenko et al., 2006; Hall et al., 2009; Chen et al., 2019). Contrastingly, both sequencing and imaging studies conducted in vivo reveal that the transcription of nucleosomal DNA proceeds at speeds similar to that of naked DNA, despite the presence of nucleosomal barriers observed in vitro (Singh and Padgett, 2009; Fuchs et al., 2014; Darzacq et al., 2007; Larson et al., 2011; Garcia et al., 2013; Fukaya et al., 2018; Liu et al., 2021).

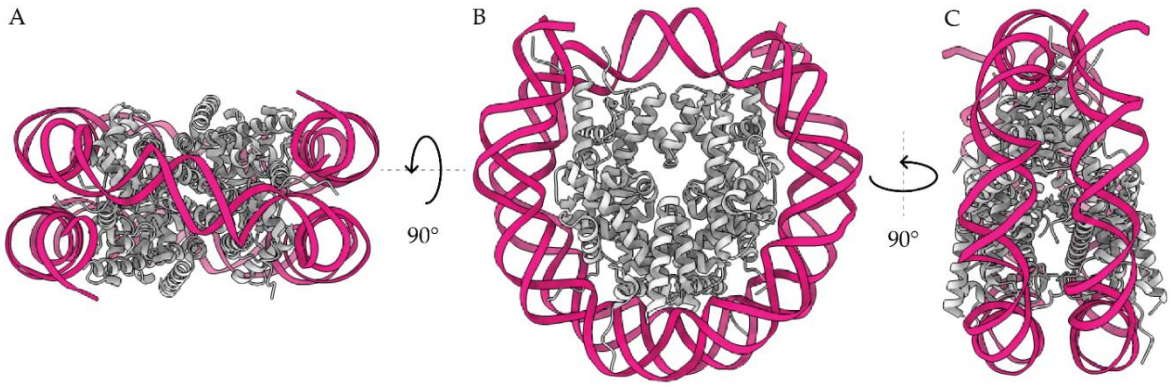

Figure 1.3: Three-dimensional structure of the nucleosome. (A) Top view of the nucleosome. (B) Front view of the nucleosome. (C) Side view of the nucleosome. In all panels, the histone proteins are highlighted in grey. The DNA is shown in magenta.

1.3.2 Post-translational modification of nucleosome

Post-translational modifications (PTMs) of nucleosomes play a critical role in the regulation of gene expression and the maintenance of genomic integrity (Millán-Zambrano et al., 2022). Nucleosomes, the fundamental units of chromatin, consist of DNA wrapped around histone protein octamers. These histone proteins, particularly their N-terminal tails, are subject to a diverse array of PTMs, including methylation (Zhang and Reinberg, 2001; Greer and Shi, 2012; Miller and Grant, 2013; Hyun et al., 2017), acetylation (Struhl, 1998; Roth et al., 2001; Eberharter and Becker, 2002; Di Cerbo et al., 2014), phosphorylation (Wei et al., 1999; Lo et al., 2005; Rossetto et al., 2012), ubiquitination (Zhang, 2003; Chandrasekharan et al., 2009; Cao and Yan , 2012), and sumoylation (Nathan et al., 2003; Shiio and Eisenman, 2003; Ryu and Hochstrasser, 2021). Each of these modifications can influence chromatin structure and function in nuanced ways, affecting the accessibility of DNA to transcription factors, the machinery for DNA replication, repair, and the overall chromatin dynamics.

Methylation and acetylation are among the most studied PTMs of histones. Methylation of lysine or arginine residues on histones can either activate (Vermeulen et al., 2007; Lauberth et al., 2013) or repress (Müller et al., 2002; Yamane et al., 2007) transcription, depending on the specific residue modified and the degree of methylation (Hyun et al., 2017). Acetylation of lysine residues, on the other hand, is generally associated with transcriptional activation by weakening the interaction between histones and DNA, thereby making the chromatin more accessible (Mutskov et al., 1998). Phosphorylation of histone tails, particularly during cell division and in response to DNA damage, marks significant chromosomal events and is crucial for chromosome condensation, segregation (Wei et al., 1999), and the DNA damage response (Foster and Downs, 2005). Ubiquitination, often linked with histone degradation and transcriptional regulation (Shmueli et al., 2022), and sumoylation, generally associated with transcriptional repression (Shiio and Eisenman, 2003), further exemplify the complexity

of nucleosome regulation by PTMs.

The combinatorial nature of these modifications, often described as the "histone code," dictates distinct chromatin states that determine gene expression patterns. For instance, specific patterns of PTMs recruit chromatin remodelers or histone modifiers to specific genomic loci, modulating the transcriptional outcome. These dynamic PTMs facilitate the cell's ability to rapidly respond to internal signals and external environmental cues by altering gene expression.

1.3.3 Chromatin higher-order structure

Nucleosomes organize into chromatin fibers, broadly classified into euchromatin and heterochromatin (Morrison and Thakur, 2021). Euchromatin, less condensed and gene-rich, facilitates gene expression by allowing easy access for the transcription machinery, responding quickly to cellular and environmental changes through histone modifications like methylation (Ng et al., 2003; Barski et al., 2007). In contrast, heterochromatin is more compact, restricting access to transcription factors, thus silencing transcription (Richards and Elgin, 2002; Saksouk et al., 2015; Penagos-Puig and Furlan-Magaril, 2020). Found in repetitive sequence regions and crucial for genome stability, heterochromatin's dense structure is reinforced by specific histone modifications, such as H3K9 methylation, playing a key role in processes like X-chromosome inactivation (Keniry et al., 2016).

Chromatin extends beyond nucleosomes into higher-order structures critical for gene regulation and spatial genome organization. The 30 nm fiber, a subject of ongoing research, potentially compacts chromatin further, though its exact role in cells remains debated (Chen et al., 2021). Chromatin loops, anchored by proteins like CTCF and cohesin, organize the genome into topologically associating domains (TADs), which regulate interactions between enhancers and promoters, thereby influencing gene expression (Wutz et al., 2017; Nuebler et al., 2018; Szabo et al., 2019). Additionally, chromosomes are organized within the nucleus into discrete, non-overlapping chromosome territories (Visser et al., 1999; Williams, 2003). These territories contribute to the regulation of gene expression by affecting the spatial arrangement of the genome. These structures not only contribute to the compact packaging of the genome within the limited confines of the nucleus (Zuleger et al., 2011) but also play vital roles in regulating gene expression through spatial and temporal control mechanisms (Branco and Pombo, 2006; Schneider and Grosschedl, 2007; Kalhor et al., 2012). As research progresses, understanding these higher-order chromatin organizations continues to provide insights into the intricate regulation of gene expression and its implications for development, health, and disease.

1.4 Experimental and computational tools

1.4.1 Cryogenic electronic tomography

Cryo-ET is a cutting-edge imaging technique that has revolutionized our understanding of cellular and molecular structures at near-atomic resolution in their native state. By rapidly freezing samples to cryogenic temperatures, cryo-ET preserves biological specimens in a close to natural, hydrated state, avoiding the artifacts introduced by traditional sample preparation methods like dehydration and staining. This technology provides a three-dimensional (3D) view of complex biological assemblies, enabling researchers to visualize the intricate details of cellular components, such as the organization of the cytoskeleton (Gui et al., 2023), the structure of organelles (Weber et al., 2019), and the interaction between macromolecules (Murata and Wolf, 2018).

The process involves tilting the frozen sample at various angles within an electron microscope to collect a series of two-dimensional (2D) images, which are then computationally reconstructed into a 3D volume, offering unprecedented insights into the spatial organization of biological structures. Cryo-ET bridges the gap between traditional electron microscopy, which offers high-resolution images of biological samples, and techniques like X-ray crystallography or NMR spectroscopy, which provide atomic resolution of purified proteins but lack the context of the cellular environment (Wang and Wang, 2017; Gauto et al., 2019).

As a powerful tool for structural biology, cryo-ET is instrumental in advancing our understanding of the mechanisms underlying cellular processes and the architecture of viral particles, membrane-bound organelles, and large protein complexes. Its ability to capture snapshots of biological processes in action opens new avenues for exploring the dynamic nature of life at the molecular level, with significant implications for drug discovery, the development of therapeutic interventions, and our overall understanding of life's complexity.

1.4.2 Live cell imaging with nascent RNA labeling

Live cell imaging with nascent RNA labeling has emerged as a transformative approach for studying RNA synthesis and dynamics within living cells, marking a significant leap from traditional, fixation-based methods. This technique harnesses the capabilities of fluorescence microscopy and RNA labeling to provide a real-time window into gene expression processes. By employing fluorescent markers to tag nascent RNA molecules, it enables direct observation of transcriptional activities (Wansink et al., 1993), RNA movement (Briley et al., 2015), and interactions within the cellular environment (Rackham and Brown, 2004; Huranová et al., 2009; Han et al., 2014; Alam et la., 2017).

Historically, the investigation of transcription in vivo heavily relied on cell fixation, a method that, while useful for capturing high-resolution snapshots of cellular states, freezes biological processes in time (Coons, 1961; Moter and Göbel, 2000; Dyba et al., 2003; Rust et al., 2006). This approach inherently lacks the capacity to track the unfolding

dynamics of transcription, limiting insights into the temporal aspects of RNA behavior and gene activity. The static nature of these traditional methods highlighted a critical gap in our ability to observe transcription as an active process, paving the way for the development of live cell imaging techniques.

The advent of systems like MS2 and PP7 stem-loops has been pivotal in advancing the study of transcription, particularly elongation, in the context of living cells. These systems employ RNA stem-loops that bind to bacteriophage-derived coat proteins tagged with fluorescent markers, making it possible to visualize nascent RNA transcripts in real-time (Johansson et al., 1997; Tutucci et al., 2018). This method allows for the engineering of gene constructs that include these stem-loop sequences, enabling the specific visualization of RNA as it is being transcribed. Figure 1.4A gives an overview of how a dual color labeling system works. In this configuration, a reporter gene is engineered to contain multiple copies of MS2 and PP7 stem-loop sequences. Taking MS2 as an illustrative example, during transcription, the emerging RNA adopts hairpin loop structures. These loops are specifically recognized and bound by a constitutively expressed coat protein, tailored to be complementary to the MS2 sequences and fused to the fluorescent protein mCherry. This interaction leads to an accumulation of the fluorescent protein at the site of active transcription, manifesting as a fluorescent punctum (Figure 1.4B). This punctum acts as a quantitative indicator, reflecting the abundance of nascent RNA transcripts generated from the gene. The intensity of fluorescence observed in a specific channel linearly correlates with the number of actively engaged RNAP molecules on the gene. By analyzing fluorescent signals from both channels, the transcription elongation rate can be determined. This is achieved by dividing the number of base pairs between the MS2 and PP7 stem loops by the time required for their transcription. Such a method provides a precise measure of transcription dynamics, allowing the quantification of elongation rates across different genomic contexts.

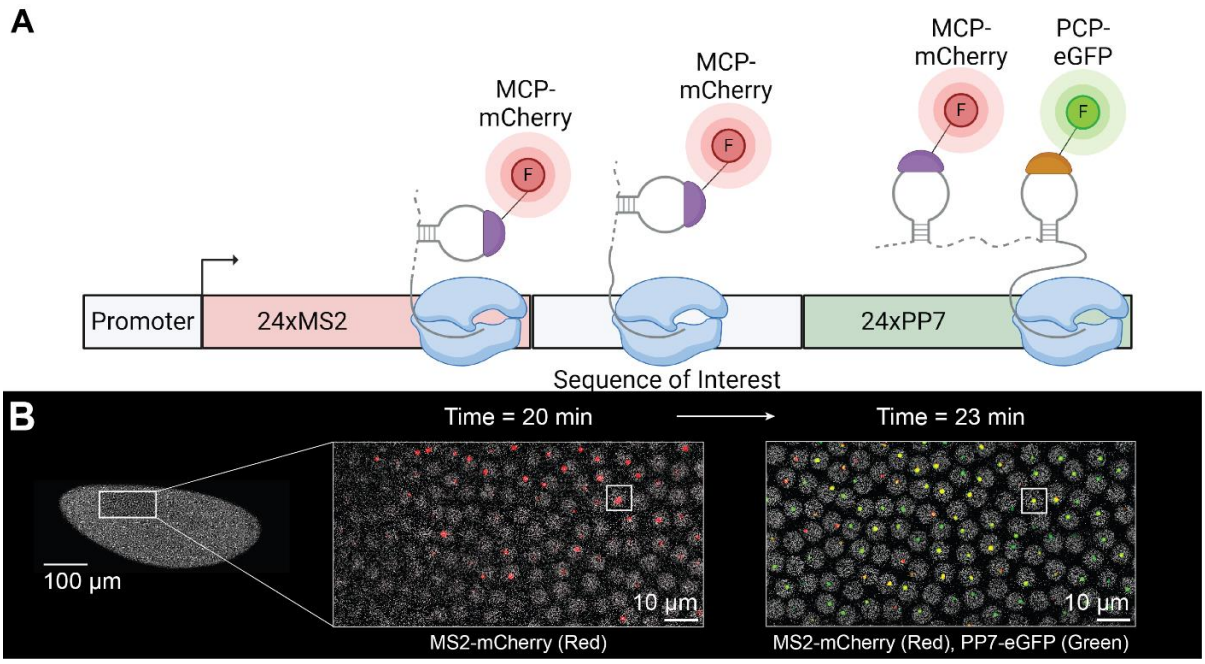

Figure 1.4: Overview of nascent RNA labeling through MS2 and PP7 stem-loop system. (A) Stem loop sequences (e.g., 24 copies of MS2 or PP7 before and after sequence of interest) are inserted into the body of a gene, producing hairpin loops in the nascent RNA transcripts. Complementary RNA binding proteins fused with fluorescent proteins then detect and bind to these RNA hairpin loops, resulting in the localization of fluorescence at the transcriptional locus. Created with BioRender.com (B) Exemplar fluorescence data on transcription elongation dynamics in live cell *Drosophila* embryos. When a tagged gene is active, fluorescent puncta will appear in a microscope's field of view in a sequential manner, red for the MS2 stem loop and green for the PP7. Created with BioRender.com.

1.4.3 Next generation sequencing and ATAC-seq

Next-Generation Sequencing (NGS) represents a suite of advanced technologies that have revolutionized genomic research by enabling the rapid sequencing of DNA and RNA at unprecedented scales and costs (Niedringhaus et al., 2012; Behjati and Tarpey, 2013; Liu et al., 2014; Slatko et al., 2018). Emerging in the early 21st century, NGS technologies have facilitated a wide range of applications, from whole-genome sequencing (Ng and Kirkness, 2010) and targeted gene panels to transcriptome analysis (Mutz et al., 2013) and epigenetic studies (Meaburn and Schulz, 2012). By allowing scientists to sequence millions of fragments simultaneously, NGS provides a comprehensive snapshot of the genetic landscape, uncovering variations and mutations, elucidating gene functions, and exploring the complexities of gene expression and regulation (Vogelstein et al., 2013; Hussmann et al., 2021).

Among the diverse applications of NGS, Assay for Transposase-Accessible

Chromatin using sequencing (ATAC-seq) stands out as a powerful technique designed to probe chromatin accessibility across the genome. ATAC-seq leverages the precision of NGS to map open chromatin regions and identify DNA sequences that are not tightly packed within nucleosomes, indicating active regulatory regions such as enhancers, promoters, and insulators. This is achieved through the use of a hyperactive Tn5 transposase that inserts sequencing adapters into accessible regions of DNA. The DNA is then sequenced, revealing areas of the genome that are open and likely involved in transcriptional regulation (Buenrostro et al., 2015).

The integration of ATAC-seq with NGS technology has dramatically enhanced our understanding of the epigenetic landscape and its impact on gene expression (Luo et al., 2022), cellular differentiation (Ranzoni et al., 2021), and the development of diseases (Wang et al., 2018). ATAC-seq requires relatively low input material compared to other chromatin accessibility assays, making it particularly suitable for limited or precious samples, including single-cell analyses (Beak and Lee, 2020). This has opened new avenues for studying cellular heterogeneity and the dynamics of chromatin accessibility in complex tissues and developmental processes.

1.5 Overview of dissertation

This dissertation is organized into four major chapters that delve into the complexities of transcription through nucleosomal DNA. Chapter 2 focuses on using purified proteins to reconstitute the transcription elongation complex in vitro. Here, we demonstrate that nucleosomes act as significant barriers to transcription, with RNAP alone unable to efficiently bypass these obstacles. In Chapter 3, we utilize cryo-ET to obtain detailed structural insights into the interactions between RNAP and nucleosomes during transcription, providing visual evidence of the transcriptional challenges posed by nucleosomes.

Chapter 4 transitions to an in vivo setting within *Drosophila* embryos to explore how transcription is influenced by nucleosome positioning sequences previously studied in vitro. Surprisingly, we find that transcription elongation rates are not significantly impacted by the presence of these sequences, suggesting a disparity between *in vitro* and *in vivo* transcription dynamics.

Finally, Chapter 5 extends the investigation into nucleosome positioning by employing ATAC-seq to analyze chromatin accessibility. This analysis reveals that, contrary to *in vitro* findings, the underlying DNA sequence does not significantly influence nucleosome positioning within the cellular context. This suggests that additional cellular factors may mitigate the impact of nucleosome positioning on transcription *in vivo*.

Chapter 2
Characterization of transcription on nucleosomal DNA with purified components

2.1 Introduction

The characterization of transcription on nucleosomal DNA using purified components is a critical area of study that delves into the fundamental mechanisms by which the transcription machinery interacts with chromatin. This line of research provides insights into the biophysical and biochemical aspects of how nucleosomes, the basic units of chromatin composed of DNA wrapped around histone proteins, affect transcriptional processes.

In vitro systems that use purified components, such as RNAP, general transcription factors, and reconstituted nucleosomes, allow for controlled studies of transcriptional initiation (Nogales et al., 2017), elongation (Palangat et al., 2012), and termination (Artsimovitch and Henkin, 2009). By employing these purified elements, researchers can meticulously analyze how specific histone modifications, the positioning of nucleosomes along the DNA, and the presence of various histone variants influence the efficiency and fidelity of transcription.

One key focus is on understanding how RNAP navigates through the nucleosomal barrier. Nucleosomes can impede the progress of RNAP (Studitsky et al., 2017), making it essential to study how transcription factors and chromatin remodelers facilitate transcriptional elongation through these structures. The dynamic removal and reassembly of nucleosomes during transcription, a process assisted by chromatin remodelers and histone chaperones, are crucial for regulating gene expression and ensuring genomic stability (Becker and Workman, 2013; Lai and Pugh, 2017).

Furthermore, these studies often aim to replicate the natural chromatin environment as closely as possible, although the complexity of *in vivo* conditions can never be fully duplicated *in vitro*. Nonetheless, these experiments are invaluable for dissecting the mechanical and structural challenges that the transcription machinery encounters and for identifying the strategies it employs to overcome these challenges.

Overall, the characterization of transcription on nucleosomal DNA using purified components helps to bridge our understanding of molecular biology from basic biophysical interactions to complex cellular functions, shedding light on the regulatory strategies cells use to control gene expression.

In this chapter, we explore the assembly of transcription elongation complex (TEC) using yeast RNAP, focusing on templates that incorporate extensive upstream and downstream DNA with and without nucleosomes. We compared two traditional *in vitro* methods: the sequential assembly method and the mismatch bubble assembly method. By employing mismatch bubble method, we can better understand how RNAP alone can overcome the nucleosome barrier *in vitro*. This approach allows for a more accurate simulation of the transcriptional process, providing insights into the mechanics of RNAP progression in environments structured similarly to natural chromatin. Through detailed experimentation, we aim to elucidate the mechanics behind RNAP's ability to navigate complex chromatin landscapes, thereby contributing to our understanding of transcriptional regulation in a chromatin context.

2.2 Results

2.2.1 Sequential TEC assembly is effective in assembling TECs, but has non-specific ligation products with downstream component

Sequential TEC assembly involves annealing RNA to the template DNA strand, the incorporation of RNAP into the DNA-RNA hybrid, followed by addition of non-template DNA strand (Palangat et al., 2012).

To study RNAP transcription elongation on various DNA templates, the entire DNA construct needs to be several hundred base-pairs long to allow the assembly of nucleosome on the downstream DNA component. Nonetheless, the efficiency of TECs decreases with the increased length of the template and non-template DNA strands used in sequential assembly. In other words, the downstream components must be ligated to the TECs only after they are fully assembled, and the options for further purification of the ligated complex are quite restricted. Therefore, we initially investigate the ligation between the double-stranded bubble DNA, utilized in TEC assembly, and the downstream nucleosome, using either *E. coli* DNA ligase or T4 DNA ligase.

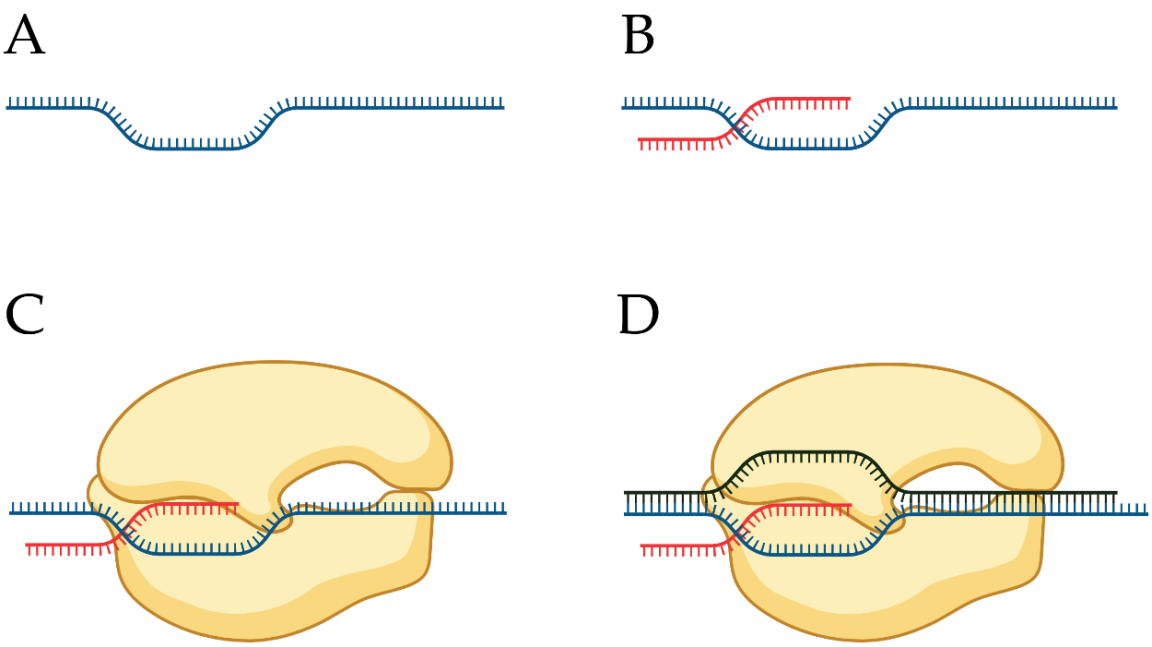

Figure 2.1 Schematic of sequential TEC assembly. (A) The template DNA strand colored in blue, and bubble sequence is indicated by the curved region. (B) Annealing the RNA oligo, colored in red, to the template DNA. (C) Binding the RNAP, colored in yellow, to the pre-annealed RNA-DNA hybrids. (D) Addition of the non-template DNA strand, colored in black, to the RNA-DNA-RNAP complex and the 3′ DNA overhang created during this process is capable of being ligated to the downstream components. Created with BioRender.com.

As shown in Figure 2.2, neither *E. coli* DNA ligase nor T4 DNA ligase can efficiently ligate the nucleosome loaded on a Widom 601 NPS (Lowary and Widom, 1998; Thåström et al., 1999) to the bubble DNA, even when nucleosomes are provided in excess of more than 3-fold. In the ligation experiment, the presence of multiple bands observed in native polyacrylamide gel electrophoresis (PAGE) of the purified nucleosome sample indicates that the sample comprises a mixture of bare DNA, hexasomes, and octasomes. This diversity in nucleosome species adds an extra layer of complexity to the ligation reaction, as each type of nucleosome can differently affect the efficiency and outcome of the ligation process. This variation suggests that the sequential assembly may not be suitable for our subsequent experiments due to the unpredictable nature of ligation outcomes with varied nucleosome compositions. In light of these complexities, we have decided to shift our approach to the mismatch bubble assembly method. This alternative technique addresses the challenges associated with ligation variability by first ligating the mismatch bubble DNA to the NPS before nucleosome assembly. This method ensures a more controlled and consistent nucleosome configuration, facilitating more reliable outcomes in our experimental processes and a more homogeneous sample preparation for our cryo-ET experiments, will be discussed in chapter 3. Further details on the preparation of nucleosomes using this method will be elaborated in the following

section.

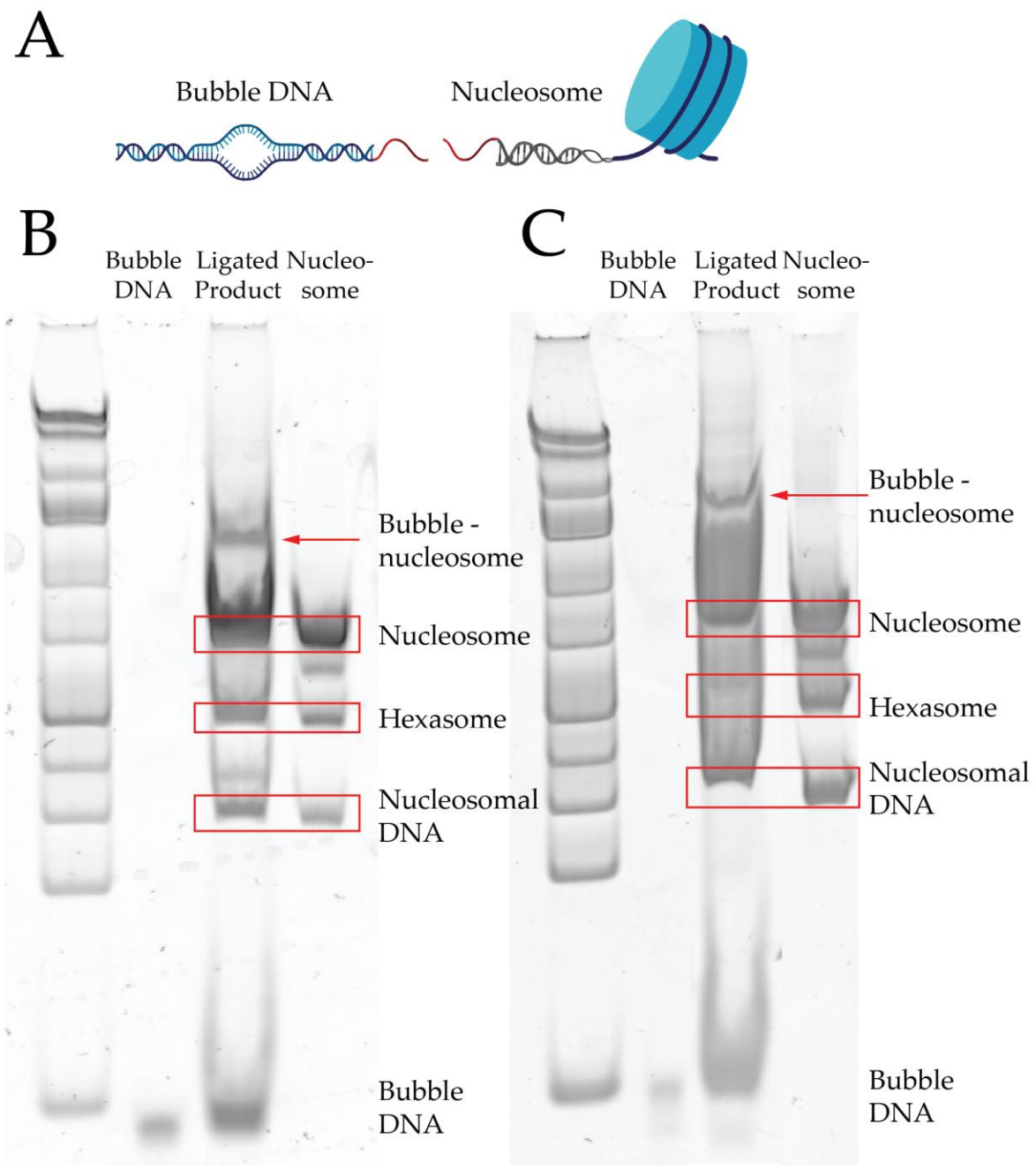

Figure 2.2 Inefficient ligation of double-stranded bubble DNA, used in sequential TEC assembly, to a purified nucleosome when using neither *E.coli* DNA ligase nor T4 DNA ligase. (A) Schematic of ligation setup. Double-stranded bubble DNA is colored in blue and black, and the overhang generated during TEC assembly is color in red. In this ligation test, RNA and RNAP are omitted during TEC assembly, only template and non-template DNA strands are used. Nucleosome is loaded on Widom 601 sequence, colored in black and grey, and DNA overhang on the 5′ end is colored red. The

ligation is set up by mixing bubble DNA, nucleosome, and T4 DNA ligase or *E.coli* DNA ligase. (B) Ligation with *E.coli* DNA ligase and (C) with T4 DNA ligase are both inefficient, even nucleosome sare 3-fold in excess than the bubble DNA used in reactions. Both ligation experiments were assessed by a 4% Native PAGE with SYBR-safe staining the DNA. Created with BioRender.com.

2.2.2 Mismatch TEC assembly enables homogeneous sample preparation

In contrast to sequential TEC assembly, mismatch TEC assembly does not require the addition of the non-template DNA strand to the DNA-RNA hybrid. Instead, the transcription bubble is created through a 9 base-pair noncomplementary region within the double-stranded DNA and ligation to the downstream component can be performed prior to the assembly of RNAP (Kujirai et al., 2018; zhang et al., 2021). With mismatch TEC assembly, after ligating the double-strand DNA containing the mismatch bubble, further purification can be conducted on both the DNA itself and the DNA loaded with histone octamers. The preparation of the DNA and loading of the nucleosome will be discussed in the Material and Method section. We then evaluate the efficiency of RNAP loading on the DNA, as demonstrated in Figure 2.3B. Despite RNAP being provided in excess (4-fold compared to DNA template), only a portion of DNA successfully has RNAP loaded onto it, shown as the RNAP Loaded band in Figure 2.3B. Those DNA without RNAP loaded are represented by the Bubble DNA band. This loading inefficiency will not impact our bulk biochemistry experiments. However, it will introduce additional complexity into our system for cryo-ET studies. Consequently, we will incorporate further selection steps during sample preparation for cryo-ET to mitigate this issue.

2.2.3 Transcription on bare or nucleosomal DNA with mismatch bubble initiation

To explore the impact of nucleosomes on transcription in vitro, we employed fluorescently labeled RNA, enabling us to track the elongation of nascent RNA. For this purpose, we utilized RNAP purified from cultured budding yeast, along with nucleosomes assembled using native human histone octamers. By adopting the mismatch bubble assembly technique for RNAP, as detailed in section 2.2.2, we were able to efficiently load RNAP onto either bare DNA or DNA wrapped in nucleosomes.

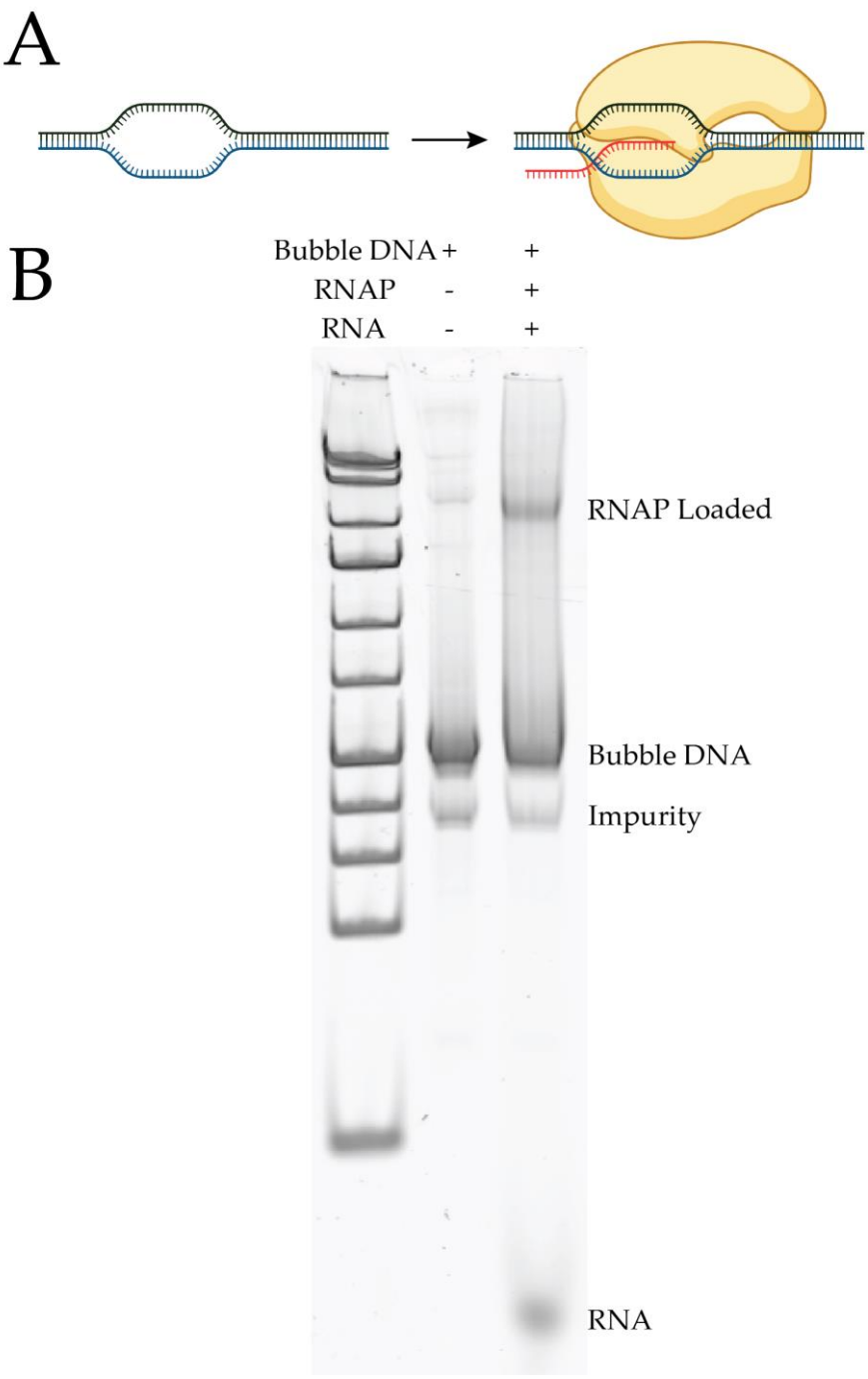

Figure 2.3 Mismatch TEC assembly demonstrates inefficient loading of RNAP onto DNA, yet effectively minimizes the formation of nonspecific products. (A) Schematic of mismatch TEC assembly. The assembly is conducted by mixing DNA containing a 9-nt noncomplementary region, RNA primer, and RNAP. (B) Loading of RNAP onto purified bubble DNA featuring a 9-nt mismatch and a downstream Widom 601 sequence. The upper most band on the top of the third lane indicates bubble DNA with RNAP successfully loaded and the bottom most band is the free RNA primer. The

loading was assessed by 4% Native PAGE with SYBR-safe staining DNA. Created with BioRender.com.

Fluorescently labeled RNA was used to follow transcription, as illustrated in Figure 2.4. The bands in this figure represent RNA transcripts of varying lengths: lower bands correspond to shorter RNA transcripts, while upper bands correspond to longer RNA transcripts. The upper most band, run-off, corresponds to the longest RNA transcripts that were synthesized. The TECs are assembled by combining the DNA template, with or without nucleosome, RNA, and RNAP. This mixture is then incubated at 37°C for an hour to facilitate more complete assembly of the TECs. The transcription is reinitiated by the addition of all four ribonucleotides (rNTPs), each at 1 mM, at room temperature and the reactions are quenched at various time points by the introduction of formamide loading buffer. In the absence of rNTPs, RNAP is unable to extend the RNA, resulting in the RNA maintaining its original length. When transcribing on the bare DNA template, RNAP encountered a significant pausing site immediately adjacent to the mismatch bubble region, denoted as initial pauses. Once these pauses were overcome, RNAP proceeded to transcribe to the end of the DNA template smoothly, completing the transcription of entire template in less than 10 minutes. These initial pauses will be further discussed in the subsequent section. When transcribing on a nucleosomal DNA template, RNAP not only encountered the same initial pauses observed with the bare DNA template but also faced multiple pausing sites within the NPS region. Over time, RNAP is capable of clearing the nucleosome and this clearance is evidenced by increasing intensity of the run-off over time. However, a significant portion of transcribing RNAP remains hindered by the nucleosome, even after 10 minutes of transcription, as shown by the bands in the NPS region. This observation indicates that RNAP alone struggles to efficiently overcome even a single nucleosome barrier. In contrast, within a cellular context where the gene body is densely packed with multiple nucleosomes, RNAP can still achieve transcription elongation rate of approximately 30 base-pairs per second, equating to 2 kilobase-pairs per minute. This discrepancy highlights the critical role of transcription factors in vivo, which significantly mitigate the nucleosome barrier and enhance transcription efficiency.

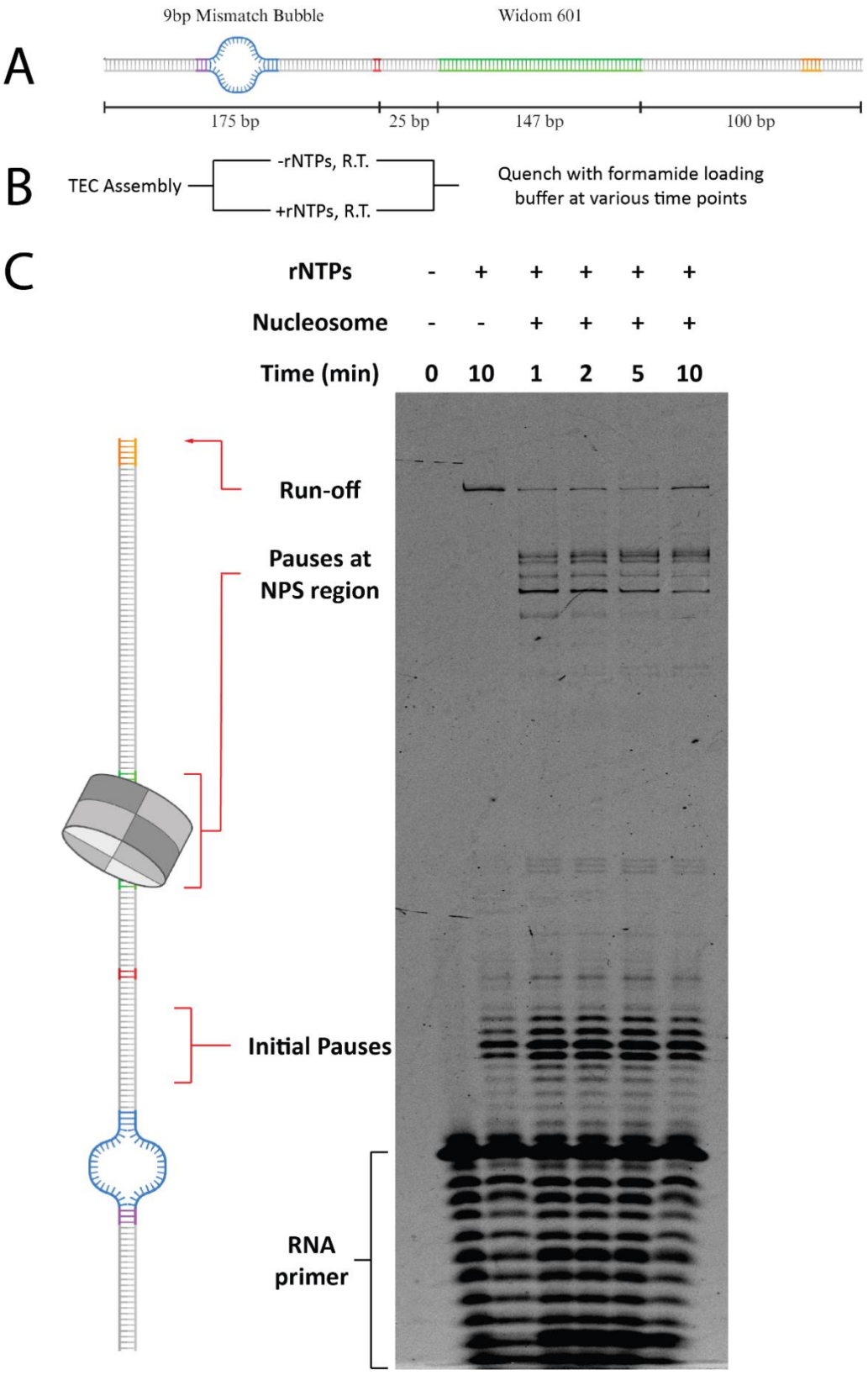

Figure caption in the following page

Figure 2.4 Transcription on bare or nucleosomal DNA with and without rNTPs. (A) Schematic of DNA construct. A Widom 601 sequence for nucleosome loading is located on the downstream of a 9-nt mismatch bubble (B) Transcription reaction procedure. TECs are assembled by mixing either bare DNA or the same DNA pre-loaded with human nucleosome with an RNA primer and RNAP, followed by incubation at 37 ℃ for an hour. Subsequently, either 1mM of each rNTPs or reaction buffer is added, and the mixture is incubated at room temperature (R.T.). The reactions are quenched at various time points by the addition of formamide loading buffer. (C) Denature PAGE to check the transcription product. With 1 mM rNTPs, RNAP transcribing on bare DNA reaches run-off without any pause at the NPS region. Conversely, RNAP transcribing on nucleosomal DNA experiences multiple pauses at the NPS region, yet it gradually reaches run-off with extended reaction time. Created with BioRender.com.

2.3 Conclusions

In this chapter, we focused on the efficient assembly of TEC using yeast RNAP across various DNA templates, particularly those that incorporate nucleosomes. We performed a comparative analysis of two primary in vitro assembly methods: the sequential assembly method and the mismatch bubble assembly method. The findings indicate that while both methods are effective in assembling TECs, the sequential assembly method often results in suboptimal ligation with downstream components, which can limit the functionality of the complexes in subsequent experiments. On the other hand, the mismatch bubble assembly method provides a more robust alternative, particularly useful for studies aimed at understanding how RNAP navigates nucleosomal barriers. This method not only enhances the ligation efficiency but also enables a more controlled study of transcription through nucleosomal DNA by circumventing the issues encountered with sequential assembly.

The experiments conducted reveal that RNAP, when assembled using the mismatch bubble method, can successfully overcome nucleosome barriers in vitro, albeit with some difficulty, highlighting the challenges RNAP faces in a chromatin context. This setup allows for a more accurate simulation of the transcriptional process, reflecting more closely the natural chromatin environment.

In conclusion, we enhance our understanding of the mechanical challenges inherent in transcription through chromatin. It provides valuable insights into the dynamics of RNAP interaction with nucleosomes and the effectiveness of different assembly methods in replicating these complex biological processes. The findings underscore the importance of choosing appropriate assembly techniques for studying transcriptional mechanics in vitro and contribute significantly to our broader understanding of transcription regulation within chromatin.

2.4 Acknowledgements

We thank Prof. Craig Kaplan for providing us with the wild-type yeast RNAP. We thank Dr. César Díaz Celis for the human histone octamers.

2.5 Material and Methods

All DNA-modifying enzymes used in this study were purchased from NEB (New England Biolabs, Ipswich, Massachusetts). Oligonucleotides were purchased from IDT (Integrated DNA Technologies, Coralville, Iowa), while nucleotide triphosphates were purchased from Thermo Fisher Scientific (Waltham, Massachusetts). Standard salts and buffer components were purchased from Sigma Aldrich (Burlington, Massachusetts), unless specified otherwise in the text.

2.5.1 Protein preparation

The yeast RNAP used in this study was prepared as described previously (Kaplan et al., 2008) and was a generous gift of Prof. Craig Kaplan. The RNAP was expressed at its endogenous level and purified using a tandem-affinity epitope tag on the Rpb3 subunit. The purification process involved initial binding of RNAP to an IgG column, followed by release through overnight cleavage with TEV protease. The eluate from the IgG column was subsequently passed through a Hi-Trap SP column (GE Healthcare, Chicago, Illinois) and then subjected to buffer exchange into Uno-Q buffer (25 mM Tris-Acetate, pH 7.5 at room temperature, 5% glycerol, 1 mM EDTA). The RNAP was then loaded onto an Uno-Q column (Bio-Rad, Hercules, California), and the fractions containing RNAP were eluted using a gradient of (NH4)2SO4. The eluted fractions were pooled, concentrated, and the buffer was exchanged into in vitro transcription (TB) buffer (20 mM Tris pH 8.0, 40 mM KCl, 5 mM MgCl2, 2 mM DTT, 5% glycerol). Aliquots at a concentration of 1 mg/mL were snap-frozen in liquid nitrogen and stored at -80°C until needed for experiments.

Histone octamer were prepared following the method outlined by Meng et al. (2022). Human histone proteins H2A, H2B, H3, H4, were purchased from The Histone Source. To reconstitute the wild-type histone octamer, lyophilized histones were initially dissolved in unfolding buffer (20 mM Tris-HCl pH 7.5, 7 M guanidine hydrochloride, 10 mM DTT). These were then mixed in a molar ratio of H2A:H2B:H3:H4 = 1.2:1.2:1:1 and dialyzed against 1 L of refolding buffer (10 mM Tris-HCl pH 8.0; 2 M NaCl; 1 mM EDTA; 5 mM DTT). The solution underwent four dialysis sessions, including three buffer exchanges over 48 hours. After dialysis, the refolded octamer was centrifuged and concentrated to approximately 0.5 mL, then applied to a Superdex 200 Increase 10/300

GL column (Cytiva, Marlborough, Massachusetts) pre-equilibrated with refolding buffer. This gel filtration step effectively separated the histone octamers from aggregates, tetramers, and dimers. The fractions were analyzed using 15% SDS-PAGE, and those containing the four histones in equimolar quantities—as determined by Coomassie blue staining—were pooled and concentrated to about 10 mg/mL. These pooled fractions were then stored at -80°C.

2.5.2 Synthesis of DNA templates for nucleosome reconstitution

A tandem-A sequence was incorporated into the pGEM-3z/601 (Addgene, #26656) plasmid, which was then amplified in E. coli strain DH5α. The Widom 601 NPS fragments were PCR amplified from this plasmid and subsequently purified using a HiTrap-Q column (Bio-Rad, Hercules, California). Mismatch bubble DNA fragments were prepared by annealing two complementary single-stranded DNA oligonucleotides. These purified Widom 601 fragments underwent digestion with the BsaI-HF restriction enzyme and were then ligated to the mismatch bubble fragments using T4 DNA ligase. The ligation products underwent a two-step purification process: initially using a HiTrap-Q column (Bio-Rad, Hercules, California), followed by 5% preparative PAGE (59:1 acrylamide:bisacrylamide) via a Prep Cell apparatus (Bio-Rad, Hercules, California) to ensure selection based on correct fragment length. The final DNA template combines the Widom 601 NPS with the mismatch bubble, structured as follows:
non-template strand:
5′ –
CGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCGA
ATTCTGGCCATCTTTGTGTTTGGTGTGTTTGGGCTTCTGTTTTCTCTTGCTTGCTTG
CCTGCTGGTCCGGGTTCCTTTTGTCCTTGGTTTGCTTGTTTCTTCAAAAAGAGTTC
ATCCCTTATGTGATGGACCCTATACGCGGCCGCCCTGGAGAATCCCGGTGCCGA
GGCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTACGCG
CTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCTCCAGGCACGT
GTCAGATATATACATCCTGTGCATGTATTGAACAGCGACCTTGCCGGTGCCAGT
CGGATAGTGTTCCGAGCTCCCACTCTAGAGGATCCCCGGCAGTGGGTGTACAG
AACGTCCAGTGAGATGCAT – 3′

template strand:
5′ –
ATGCATCTCACTGGACGTTCTGTACACCCACTGCCGGGGATCCTCTAGAGTGGG
AGCTCGGAACACTATCCGACTGGCACCGGCAAGGTCGCTGTTCAATACATGCA
CAGGATGTATATATCTGACACGTGCCTGGAGACTAGGGAGTAATCCCCTTGGC
GGTTAAAACGCGGGGGACAGCGCGTACGTGCGTTTAAGCGGTGCTAGAGCTGT
CTACGACCAATTGAGCGGCCTCGGCACCGGGATTCTCCAGGGCGGCCGCGTAT
AGGGTCCATCACATAAGGGATGAACTCTTTTTGAAGAAACAAGCAAACCAAGG
ACAAAAGGAACCCGGACCAGCAGGCAAGCAAGCAAGAGAAAACAGAAGCCC

AAACACACCAAACACAAGAGCTAATTGAATTCGCTGTTTCCTGTGTGAAATTGT
TATCCGCTCACAATTCCACACAACATACG – 3′

2.5.3 Loading and purification of nucleosome

Human histone octamers and DNA templates featuring Widom 601 NPS and mismatch bubbles were combined in equimolar ratios in a high-salt buffer (10 mM Tris-HCl pH 8.0; 2 M NaCl; 1 mM EDTA; 0.5 mM DTT; 1 mM PMSF) at a final DNA concentration of 100 ng/μL. This assembly mixture was first dialyzed using a 3.5 kDa dialysis membrane at 4°C against 500 mL of the same high-salt buffer for 1 hour. This was followed by a gradual 36-hour lineal gradient dialysis against 2 L of low-salt buffer (10 mM Tris-HCl pH 8.0; 1 mM EDTA; 0.5 mM DTT; 1 mM PMSF), with continuous stirring. A final 3-hour dialysis was then performed against 500 mL of low-salt buffer. The success of the nucleosome reconstitution was assessed via 4% native PAGE (59:1 acrylamide:bisacrylamide) using 0.2X TBE buffer. Mononucleosomes were further purified from hexasomes and bare DNA through 4% preparative PAGE (59:1 acrylamide:bisacrylamide) using a Prep Cell apparatus (Bio-Rad, Hercules, California).

2.5.4 In vitro transcription setup

The transcription reaction was conducted by combining 0.1 μM nucleosome, 0.08 μM RNAP, and 0.2 μM RNA primer (5′-FAM-AUAAUUAGCUC-3′) in a 10 μL reaction mixture (20 mM Tris-HCl, pH 8 at R.T., 40 mM KCl, 5 mM MgCl2, 10 mM DTT). This mixture was incubated at 30°C for one hour. Subsequently, all four NTPs—ATP, UTP, GTP, and CTP—were added at a concentration of 1 mM each. The reaction was allowed to proceed at room temperature and was quenched at specific time points by adding an equal volume of formamide loading buffer (95% Formamide, 10 mM EDTA, pH 8.0). The RNA products were then analyzed using a 10% urea PAGE (29:1 acrylamide:bisacrylamide) in 0.5X TBE buffer and visualized on a Typhoon TRIO imager (GE Healthcare, Chicago, Illinois).

Chapter 3
Structures of RNAP transcribing through a nucleosome

3.1 Introduction

In eukaryotic cells, DNA is tightly packed into chromatin, a structure that limits accessibility to genetic material and influences all DNA-related processes, including transcription. The nucleosome, the fundamental structural unit of chromatin, consists of approximately 147 DNA base pairs wrapped around a core of histone proteins in 1.7 left-handed superhelical turns. This nucleosome core particle (NCP) comprises a central tetramer of histones H3 and H4 ($[H3\ H4]_2$), flanked by H2A-H2B histone dimers (Bilokapic et al., 2018). The histone tails, which extend from this core, are sites for various epigenetic modifications that regulate DNA accessibility (Kouzarides, 2007).

While the nucleosome efficiently organizes DNA within the nucleus, shields it from damage, and ensures equal genetic material distribution during cell division, it also poses a significant transcriptional barrier, as established by early biochemical studies (Morse, 1989; Kirov et al., 1992; Kireeva et al., 2002). The challenge of how RNAP navigates this obstacle at high speeds in vivo (Veloso et al., 2014) remains a critical question in molecular biology, driving research to uncover the mechanisms that facilitate RNAP's movement through chromatin during transcription—a process vital for gene expression and cellular functionality. To mitigate the nucleosome barrier, eukaryotic cells have evolved several strategies. Many promoters and transcription start sites, for example, feature sequences that are less prone to bending and thus less likely to form nucleosomes (Tirosh and Barkai, 2008; Bai et al., 2010; Rando and Winston, 2012; Zaugg and Luscombe, 2012). These nucleosome-depleted regions increase the accessibility of cis-regulatory elements to transcription factors and RNAP. Additionally, certain transcription factors are capable of binding to nucleosomal DNA and can trigger nucleosome remodeling, either through their intrinsic properties or by recruiting chromatin remodeling factors. As RNAP proceeds with transcription, it traverses an organized array of nucleosomes. The elucidation of this complex interaction has benefitted greatly from recent advances in cryo-EM, providing unprecedented insights into the transcription process.

High resolution structures of a transcribing RNAP-nucleosome complex have been solved using cryo-EM single particle analysis (SPA) at notable pause sites, such as super helical locations (SHL)-1, -2, -5, and -6 (Gaykalova et al., 2015; Kujirai et al., 2018; Vos et al., 2018). Despite achieving detailed snapshots at these sites, the ensemble averaging method limits the ability to capture the dynamic process in its entirety, especially beyond

the nucleosome dyad where structural information remains incomplete. Current models suggest RNAP maintains close contact with nucleosomes at major pauses, which might complement the Ø-loop model (Studitsky et al., 1995; Kulaeva et al, 2013) where a step-wise DNA unwrapping upstream of the nucleosome facilitates RNAP progress to the dyad. This suggests that RNAP does not constantly peel DNA off the nucleosome surface; instead, it navigates by moving on unwrapped DNA, rectifying nucleosome fluctuations. Upon reaching SHL-1, it is hypothesized that the DNA behind RNAP forms a transient intranucleosomal loop upon rewrapping onto the nucleosome entry site that stabilizes interactions between the histone core and DNA. Recent structural studies have highlighted the dynamic nature of the interactions between RNAP and nucleosomes during transcription. These studies reveal that the rewrapping of upstream DNA back onto the nucleosome entry site can occur in various orientations (Gaykalova et al., 2015; Kulaeva et al, 2013), suggesting a more flexible and complex relationship between RNAP and the nucleosome than previously appreciated. This flexibility points to the existence of multiple potential states or configurations during transcription, with varying distances and interactions between RNAP and the nucleosome. To better understand the molecular choreography of nucleosomal DNA as RNAP passes, we plan to employ cryo-ET, a method designed to capture and reconstruct individual instances of RNAP-nucleosome complexes without the need for averaging, preserving rare structural states.

Cryo-ET is a sophisticated imaging technique that allows for the visualization of biomolecules at molecular resolution within their natural, hydrated state (Lučić et al., 2008). This method captures a series of images of a specimen at various tilt angles using a transmission electron microscope. These images are then computationally reconstructed to create a detailed three-dimensional model. While cryo-ET provides three-dimensional insights into the arrangement and interactions of biomolecules within their native environments, it generally achieves lower resolutions compared to some other structural determination methods such as single particle cryo-EM.

In recent years, machine learning has become crucial in enhancing the analysis of cryo-ET data, particularly in addressing challenges posed by the technique's inherent limitations. The application of machine learning algorithms aids in improving the efficiency and effectiveness of various aspects of cryo-ET data processing, including image denosing and three-dimensional reconstruction. One significant challenge in cryo-ET is the high noise levels in tomograms, which result from the need to avoid radiation damage of the samples. To mitigate radiation damage, images are taken at low electron doses, making it difficult to discern features clearly (Hattne et al., 2018). Machine learning can assist by enhancing image quality and distinguishing biological structures from noise beyond what is possible with the naked eye. Another issue arises from the physical constraints during data collection. Tilt-series images in cryo-ET can typically only be collected within a limited angular range of about ± 60° due to the design of the specimen holder. This limitation leads to incomplete data in the Fourier space, known as the "missing wedge," which causes artifacts in the reconstructed volume. The most noticeable impact of the missing wedge is anisotropic resolution, where objects in the tomogram appear elongated along the beam axis (Z-direction) (Moebel and Kervrann,

2020). This elongation distorts the electron microscopy density in both 3D volumes and 2D slices related to the Z-plane, complicating accurate interpretation and analysis. Machine learning algorithms are particularly adept at addressing these issues by improving the interpolation of missing data, enhancing feature recognition despite the presence of noise and missing wedge artifacts, and facilitating more accurate 3D segmentation. These tools allow for more precise modeling of the underlying biological structures, rendering them invaluable for advancing our understanding of cellular architecture at molecular resolution.

In this chapter, we are developing a cryo-ET analysis pipeline enhanced by machine learning for structural analysis, aiming to elucidate a comprehensive mechanism involving the unwrapping and rewrapping of upstream and downstream nucleosomal DNA during RNAP passage. Visualizing nucleosome structures before and after RNAP passage will make it possible to determine if an H2A-H2B dimer was ejected or if the nucleosome remained intact (Lai and Pugh, 2017; Venkatesh and Workman, 2015).

3.2 Results

3.2.1 Cryo-ET Construct design

One of the initial hurdles encountered in cryo-ET experiments is the preparation of samples, specifically RNAP-nucleosome complexes. The traditional method, commonly used in biochemistry, of sequential TEC assembly, which involves ligating DNA loaded with RNAP and DNA loaded with a nucleosome, has proven to be inefficient (Figure 2.2). This inefficiency results in a low yield of the desired RNAP-nucleosome complexes, posing significant challenges for detailed structural analysis. However, an alternative method using mismatch TEC assembly has demonstrated more promising results. This approach involves first ligating DNA containing a mismatch bubble sequence with DNA containing an NPS. Subsequently, the nucleosome is loaded onto this prepared DNA, followed by purification of the nucleosome bound to full-length DNA. This method significantly enhances the efficiency of preparing RNAP-nucleosome complexes, thereby improving the yield and quality of the samples for cryo-ET analysis.

In our efforts to optimize the assembly of RNAP-nucleosome complexes, we introduced several features to the DNA design, as shown in Figure 3.1A. First, an EcoRI restriction site was strategically positioned upstream of the 9-nucleotide (nt) mismatch region, which will be blocked when a RNAP is loaded but unable to restart. Additionally, we incorporated a thymine-less (T-less) cassette downstream of this mismatch and about 25 base-pairs before the NPS. This modification is intended to enrich the active RNAP in the sample. Lastly, a deactivated Cas9 (dCas9) binding site was added to the end of the DNA sequence. This site allows for the binding of dCas9 protein, which serves as a barrier to prevent RNAP from dissociating from the DNA after transcription through the nucleosome. It also serves as a marker to indicate the directionality of the DNA within

the tomogram. By implementing these features, we aimed to enrich our samples with complexes that contain RNAP successfully restarted and retained on the DNA, thereby improving both the efficiency and purity of the samples for detailed structural analysis.

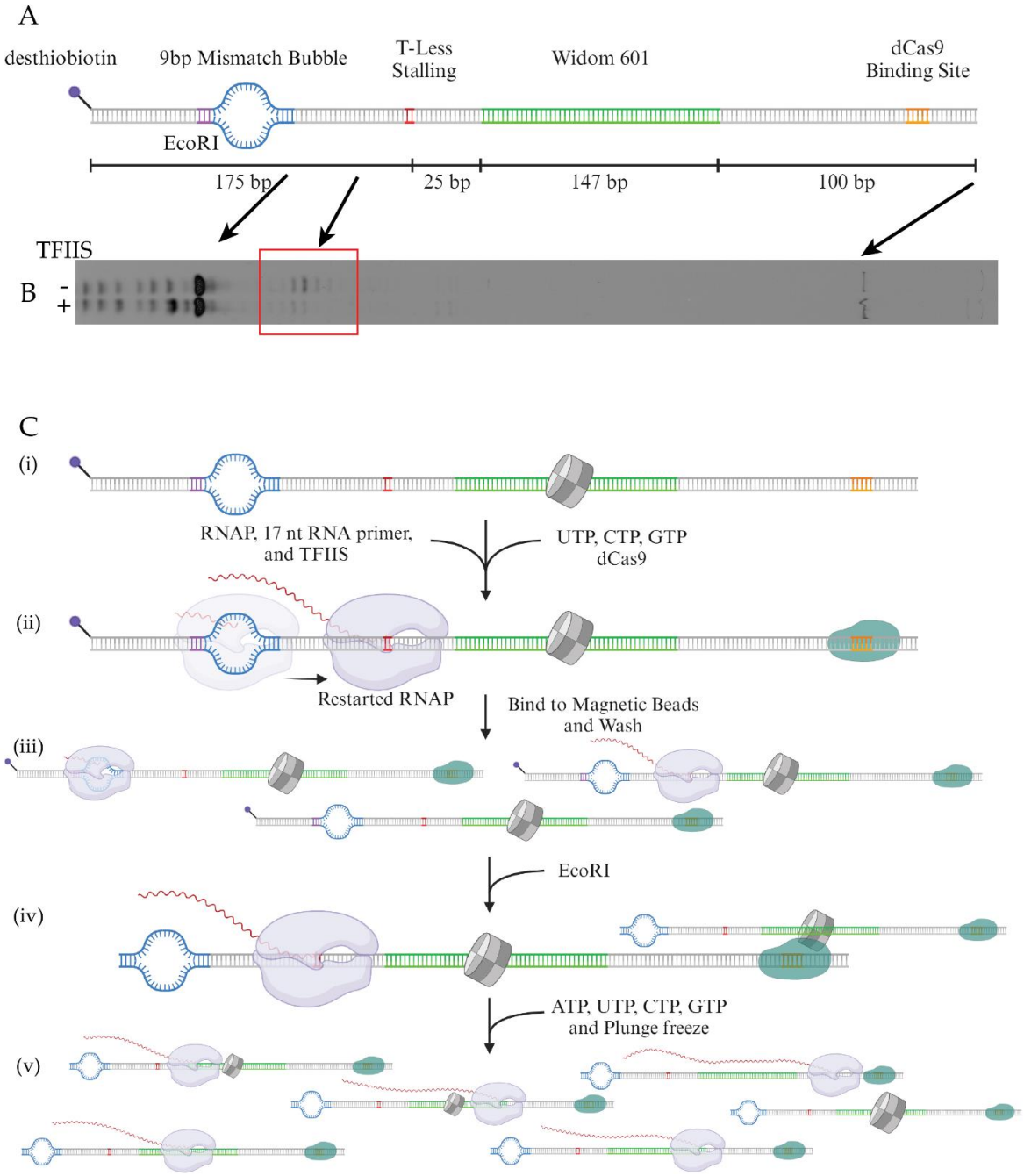

Figure 3.1 Cryo-ET sample preparation. (A) Sequence design of the DNA template. (B) TFIIS facilitates transcription in the linker region. (C) Schematic of sample preparation pipeline. Created with BioRender.com.

In Chapter 2, we conducted experiments to examine transcription on DNA preloaded with a human nucleosome, as depicted in Figure 2.4C. We observed that RNAP, in the absence of TFs, encountered significant stalling at two specific regions: one in the linker region upstream of the NPS, and the other directly within the NPS itself. To address the stalling in the linker region, we introduced Transcription Factor IIS (TFIIS) into the system. TFIIS has the capability to cleave RNA that protrudes from the RNAP catalytic site, effectively rescuing RNAP from its backtracked state (Jeon et al, 1994; Awrey et al., 1998; Ishibashi et al, 2014; Schweikhard et al., 2014). In vitro transcription assays were conducted on bare DNA to visualize RNA products using Urea-PAGE, as detailed in section 2.5.4. This experiment demonstrated that RNAP, with the assistance of transcription factor TFIIS, encountered fewer pauses in the linker region. This is evidenced by bands of lower intensity within the red box in assays incorporating TFIIS, as shown in Figure 3.1B. However, to ensure that TFIIS did not interfere with transcription through the nucleosome, we implemented additional steps to remove TFIIS once RNAP reached the T-less cassette.

The sample preparation pipeline for cryo-ET involves several meticulous steps to ensure the accurate assembly and purification of complexes for imaging. Magnetic streptavidin beads are employed to facilitate the preparation, starting with DNA that has a desthiobiotin tag at its 5′ end. The detailed procedures (Figure 3.1C) are as follows: (i) This DNA is first loaded with human nucleosomes and then purified to remove any free DNA or DNA with incomplete NCPs. (ii) The purified DNA, now properly assembled with histone octamers (hereafter referred to as nucleosomes), is mixed with RNAP, 17-nt RNA primers, TFIIS, and three of the rNTPs—UTP, CTP, and GTP—at a saturating concentration (1 mM). This mixture is incubated at 30°C for an hour to allow RNAP to restart. (iii) Following incubation, the mixture is added to magnetic streptavidin beads. This step is crucial for binding the desthiobiotin-labeled DNA to the beads, allowing for a subsequent wash to remove excess RNAP, RNA primers, TFIIS, and NTPs. Only nucleosomes, inactive RNAP-nucleosome complexes, and active RNAP-nucleosome complexes remain attached to the beads. (iv) EcoRI restriction enzyme is then applied to the bead-bound complexes to release the nucleosomes and nucleosomes with restarted RNAP. (v) Prior to plunge freezing for cryo-ET imaging, a complete set of all four rNTPs (each at 1 mM) is added to allow for a brief 10-minute transcription at room temperature. This final transcription step is crucial as it positions RNAP at various points along the DNA template, resulting in a sample that contains nucleosomes both with and without associated RNAP, capturing a snapshot of transcription in action.

3.2.2 Particle recognition with template matching to detect NCP, RNAP, and dCas9

In the sample prepared for cryo-electron tomography, four major components are crucial for analysis: DNA, NCP, RNAP, and dCas9. The EcoRI restriction enzyme, although utilized during sample preparation and present in the final sample, is too small

to be detected in the tomogram. DNA, due to its elongated structure and extremely flexible nature, poses identification challenges with conventional imaging methods; this issue will be explored further in subsequent sections. In contrast, NCP, RNAP, and dCas9 are all globular proteins, which are more readily identifiable within the tomograms. These components can be distinguished using the conventional 2D template matching or 2D convolutional neural network (2D-CNN) recognition techniques. These methods rely on comparing observed structures to pre-existing or trained models to identify and position molecules within the tomogram.

Tomograms were first aligned using the patch-tracking method along the edge of the carbon area (Figure 3.2A, bottom region below the dark edge), as it provides a large-scale, low-resolution signal for aligning the entire image. Subsequently, the 3D reconstruction software EMAN2 was used to generate raw density maps. Examination of the z-slices in these maps revealed distinct globular shapes corresponding to NCP, RNAP, and dCas9 particles, as shown in Figure 3.2A. For the training of the 2D-CNN, initially developed by Chen et al. (2019), we manually labeled a small subset of particles to establish a training dataset, where blue circles represent the particles of interest and red circles denote the background, as shown in Figure 3.2A. Detailed illustrations of all selected particles are provided in Figure 3.2B, and the areas chosen as background are displayed in Figure 3.2C. The trained network successfully identified all globular-shaped particles, marked by blue circles in Figure 3.2D. Particles on the carbon film were then identified (Figure 3.2E, red dots) and excluded from further processing to enhance data accuracy and relevance. Following this, good particles were processed through 3D classification and subtomogram averaging (STA) using the e2tomo protocol to discern types, relative orientations, and molecular details of the particles in ice. Our results demonstrate that averaged maps derived from all classified particle groups achieved resolutions of 11 Å for dCas9, 12 Å for NCP, and 10 Å for RNAP, determined by Fourier shell correlation (FSC) (Figure 3.3 B, D, and F), revealing detailed structural features of the particles. The X-axis in these figures represents the spatial frequency, which corresponds to different levels of detail in the 3D reconstruction. The Y-axis represents the FSC coefficient, which quantifies the similarity between the two independent 3D reconstructions at each spatial frequency. When the FSC curve intersects the 0.143 level on the Y-axis, it indicates the spatial frequency at which the correlation between the two reconstructions drops below a reliable level.

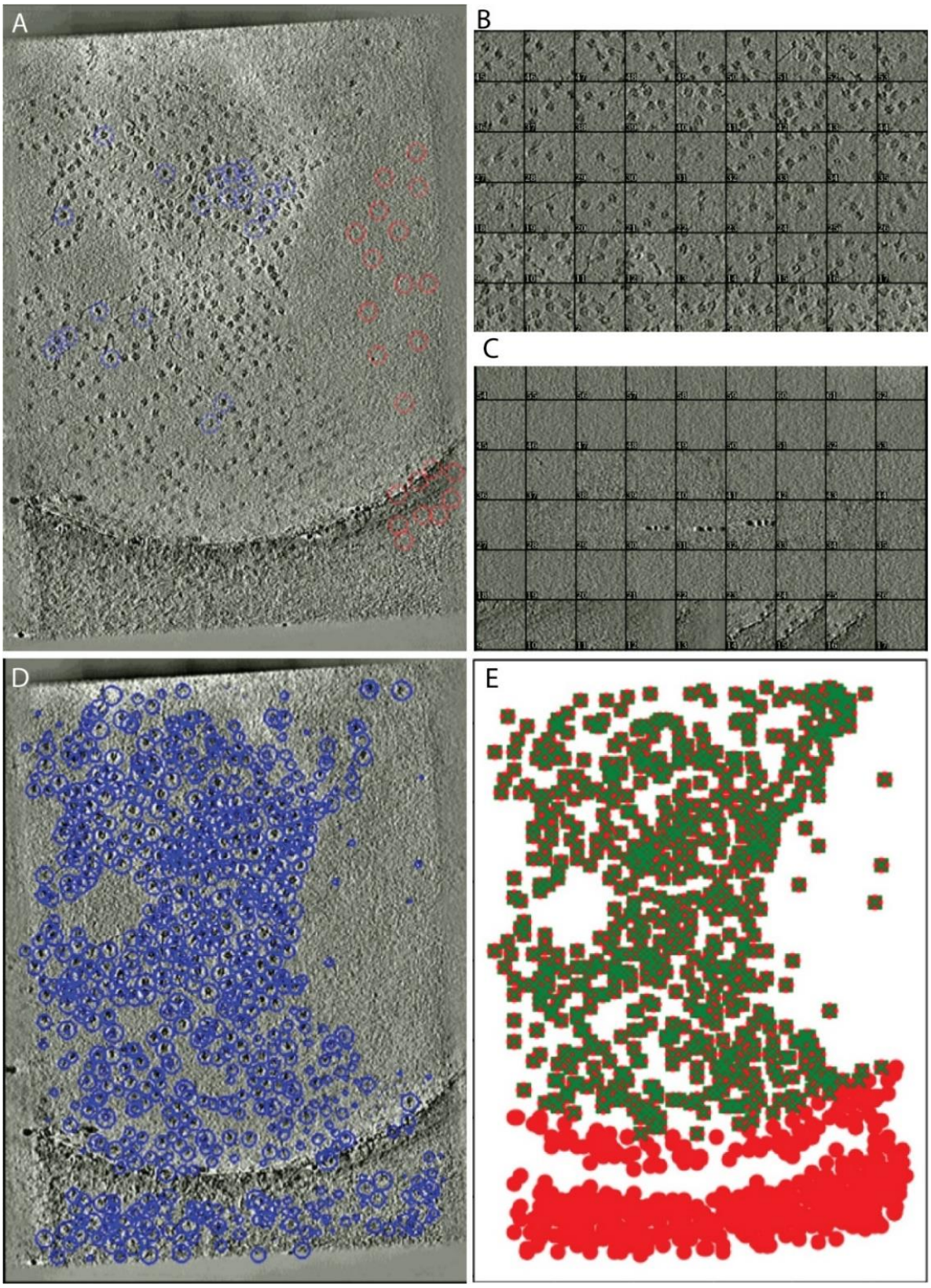

Figure caption in the following page.

Figure 3.2 Application of 2D-CNN to detect globular proteins. (A) Manual Labeling for Training Sets: dozens of target particles (encircled in blue) and background regions (marked in red) were selected from a representative 2D slice of a tomogram containing the TEC complex (B) Regions with particles of interest. (C) Regions as background. (D) Prediction of trained 2D-CNN with detected particles encircled in blue. (E) Separating particles on the carbon area (red) from those in ice (green) as predicted by the 2D-CNN. Particles on the carbon area exhibit lower contrast and may interact with the carbon surface. Only particles in the ice were submitted for further analysis.

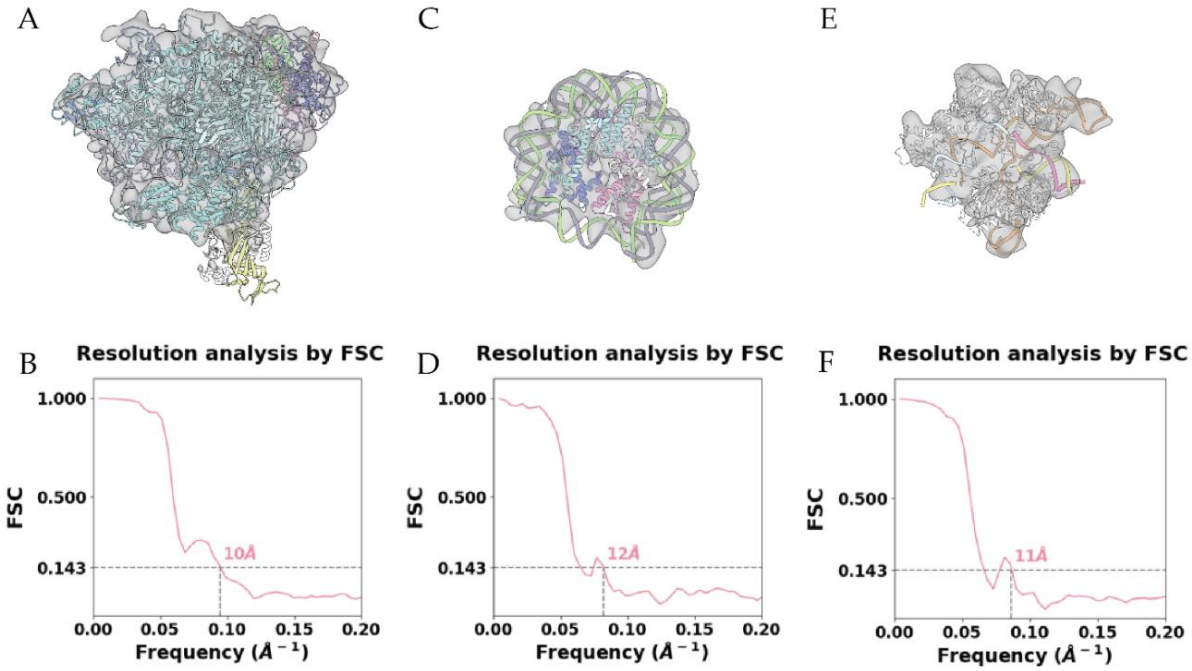

Figure 3.3 Sub-tomogram averaging of RNAP, NCP, and dCas9 particles and the determination of reconstruction resolution. (A, C, E) Subtomogram-averaged maps and the fitted structural models of RNAP, NCP, dCas9, respectively (B, D, F) Map resolution determination of RNAP, NCP, and dCas9, respectively, through FSC.

Given the orientation and types of each particle, we mapped the particles back to their corresponding tomogram (Figure 3.4A), revealing their relative orientation and interactions in z-dimensional slice (Figure 3.4B) and 3D volume slab (Figure 3.4C). As we did not involve purification in this round of preparation, excess RNAP still remained in the sample. This was further improved in subsequent preparations.

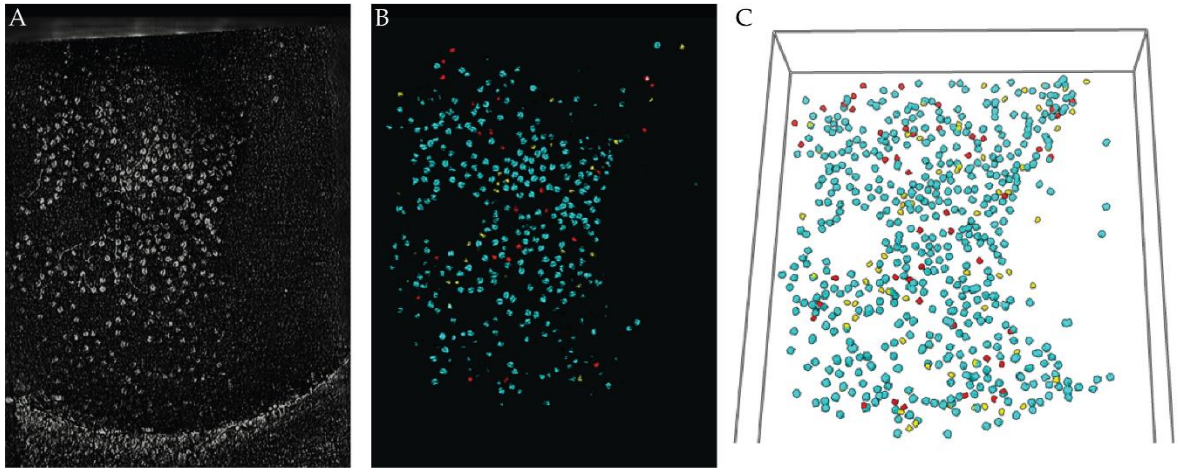

Figure 3.4. Back-mapping of RNAP, NCP, and dCas9 particles to the tomogram. (A) A representative z-dimensional tomographic slice of the raw tomogram after E2TOMO 3D reconstruction (thickness = 0.667 nm). (B) A representative z-dimensional slice of the synthetic tomogram after mapping the subtomogram averaged structures of RNAP, NCP, and dCas9 back to their corresponding positions (thickness = 0.667 nm). (C) 3D representation of the synthetic tomogram in B, displaying the distribution of the three types of particles, with RNAP, NCP, and dCas9 colored in cyan, red, and yellow, respectively.

3.2.3 DNA detection within 3D tomogram using the 3D-convolutional neuron network (3D-CNN)

Recognizing DNA in tomograms presents significant challenges due to its flexible, unstructured nature and the small diameter of the DNA double helix. These characteristics hinder the effective use of STA for further refinement, as STA relies on identifying and averaging multiple instances of similar structures to enhance signal-to-noise ratios (SNR) and resolution. DNA's inherent structural variability and its slim profile complicate these processes, making it difficult to achieve the precision required for detailed visualization and analysis in cryo-electron tomography studies. Accurate segmentation of DNA from the noisy tomogram is pivotal for correctly assembling it with proteins in subsequent analysis steps. Initial experiments using a 2D-CNN to identify DNA fibers resulted in enhanced visualization of DNA fiber signals (as shown in Figure 3.5A-B). To generate the training dataset, approximately hundreds of DNA z-dimension slice images were cropped from the real tomogram followed by manually annotated DNA trajectory. These images pairs serving as the training set and submitted to the EMAN2 workflow.

However, this detection process, performed slice-by-slice along the z-dimension of the tomogram, encountered issues of differential detection capability. DNA fibers lying in the x-y plane were better detected but appeared elongated in the z-direction (indicated by thicker fiber widths in Figure 3.5A-B, blue circles) due to the missing wedge artifact. Conversely, DNA oriented along the z-direction often went undetected after the prediction process (as shown in Figure 3.5A-B, red circles), appearing merely as "dots" in the z-dimension which the trained 2D CNN could not recognize. This limitation highlights the challenges of using 2D CNN for detecting structures in three-dimensional data where orientation and artifact-induced distortions affect the accuracy of segmentation.

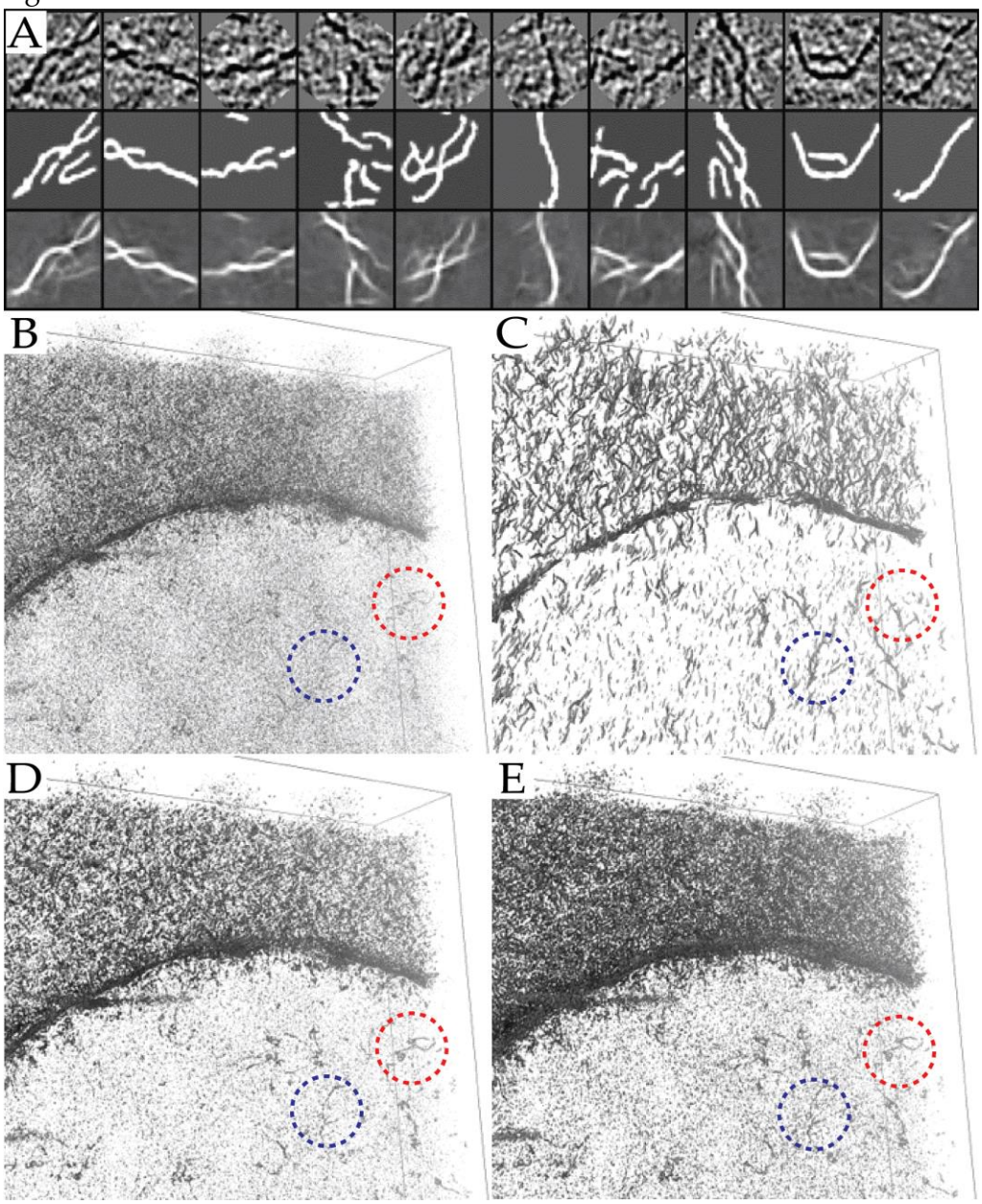

Figure caption in the following page.

Figure 3.5 Comparison of different denoising methods with regard to DNA recognition. (A) A representative 2D-CNN training dataset showing the real data (top row), manual annotation (middle row), and the network prediction (bottom row). (B) Raw tomogram. (C) Denoised tomogram through 2D-CNN. (D) Denoised tomogram through Low-pass filter. (E) Denoised tomogram through IsoNet. In all four panels, representative DNA lying in the x-y plane is encircled in blue, DNA oriented along z-direction is encircled in red.

To tackle the challenges of the missing wedge artifact and anisotropic detection of DNA signals, we implemented a strategy using two 3D CNN: IsoNet and REST. IsoNet is designed to correct the missing wedge artifact by restoring isotropic resolution in tomograms. It leverages a deep learning-based approach that incorporates prior structural knowledge to fill in missing information, enhancing the accuracy and quality of 3D reconstructions. This improvement ensures that DNA fibers are more uniformly represented across all orientations, effectively reducing elongation artifacts commonly seen in the z-direction (as shown in Figure 3.5C-D). REST focuses on denoising the tomograms. It employs iterative training, beginning with lower noise conditions and incrementally introducing higher noise levels. This strategy enables the network to distinguish more effectively between signal and noise, thus enhancing the visibility and clarity of DNA signals.

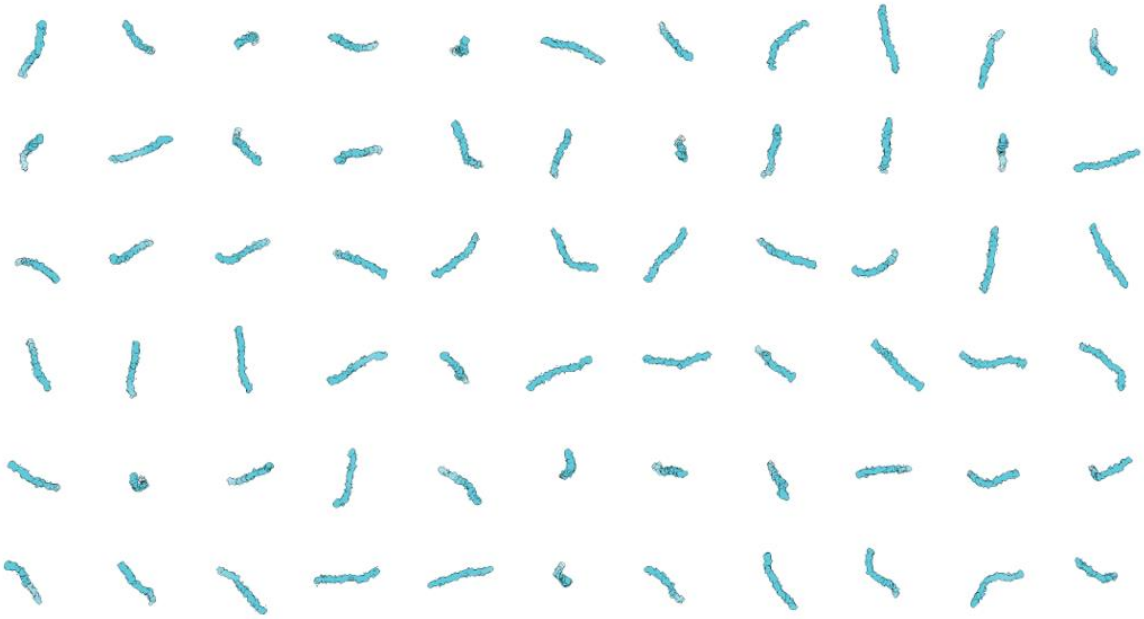

Figure 3.6. Sixty representative 40-base pair DNA particles displaying various conformations. The DNA particles were generated through relaxation of a linear double-strand B-type DNA using molecular dynamics simulation from OxDNA.

To implement the REST network effectively on our data, we created 3000 synthetic volume cubes of 40 bp linear B-DNA following molecular dynamics (MD) relaxation,

illustrated in Figure 3.6. These cubes were exposed to various noise levels, with SNR ranging from 2.5 to 0.003, to generate a series of training pairs, illustrated in the top and middle rows of Figures 3.7. Our initial training efforts using data at an SNR of 0.01 did not successfully identify DNA features within the low SNR volumes. To overcome this, we adopted an iterative training strategy, starting with datasets with high SNR and progressively moving to low SNR. This method enabled the REST network to gradually enhance its capability to differentiate DNA signals from the background noise. This progressive training approach helped the network to adjust and maintain effectiveness across a spectrum of realistic noise levels and made reliable prediction, as illustrated in the bottom row of Figure 3.7. Although the predictive capability of the networks diminished at an SNR of 0.003 due to significant noise interference, applying the models trained at this SNR could successfully restore DNA signals from volume cubes at an SNR of 0.01. These in silico experiments demonstrate the efficacy of using 3D CNNs to enhance signal detection within challenging noise environments, paving the way for more precise DNA segmentation in subsequent structural and functional analyses.

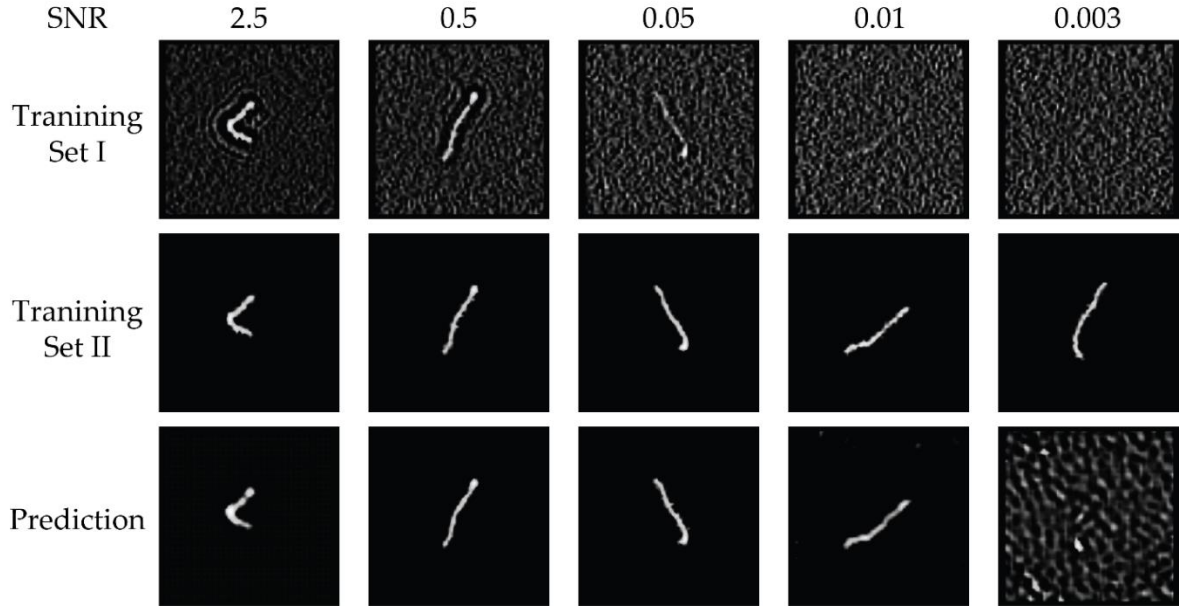

Figure 3.7 Step-wise training and prediction of a 3D-CNN for recognizing DNA signals across varying SNR levels. Training Set I was composed of simulated DNA models obtained through MD relaxation, subjected to varying levels of Gaussian noise. Training Set II consisted of noise-free simulated DNA models.

For real data, the lack of ground truth makes it challenging to accurately measure the SNR. As a test, we cropped a local area from a large tomographic dataset containing a piece of DNA. Figure 3.8 displays the 2D projected images of the raw cube, the low-pass filtered cube, and the cubes processed by IsoNet and a 2D-CNN. Although the 2D-CNN cannot isotropically enhance the DNA signal, it is included in the test as a type of volume source, clearly indicating the DNA location within the noisy real data for comparison. Compared with the simulated data, the SNR of the raw, low-pass filtered, and IsoNet

processed data (Figure 3.7) ranges between 0.01 and 0.003, as indicated by the level of distinction from the background. By applying all five trained 3D-CNN networks to the four types of real data, with or without processing, we observed differences in the outcomes of denoising. Besides the 2D-CNN control group (Figure 3.9), the IsoNet processed volume displayed the most standout DNA signal and continuity compared to the other groups. Therefore, we chose to use the 0.003 networks to denoise the IsoNet maps for the rest of the data analysis.

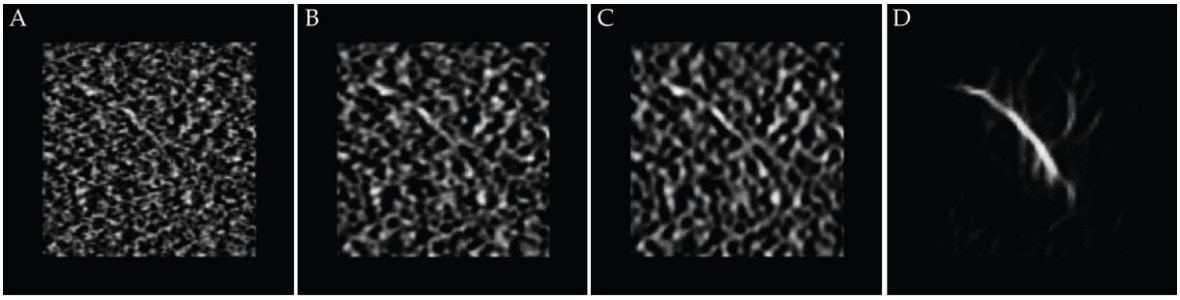

Figure 3.8 Comparison of different preprocessing methods in reshaping the data distribution. Images are displayed in the top panel and corresponding intensity distributions are displayed in the bottom panel (A) 2D-projection of a raw data cube (B) Data enhanced using a low-pass filter. (C) Data enhanced using IsoNet. (D) Data enhanced by a 2D-CNN.

By integrating IsoNet and REST, we can correct for the missing wedge artifact while simultaneously reducing noise, thereby achieving a more isotropic and clear representation of the DNA fibers (Figure 3.10A and B). This dual-network approach significantly enhances the detection capability of the DNA signal across all dimensions, facilitating more accurate assembly of DNA with proteins in subsequent analysis steps.

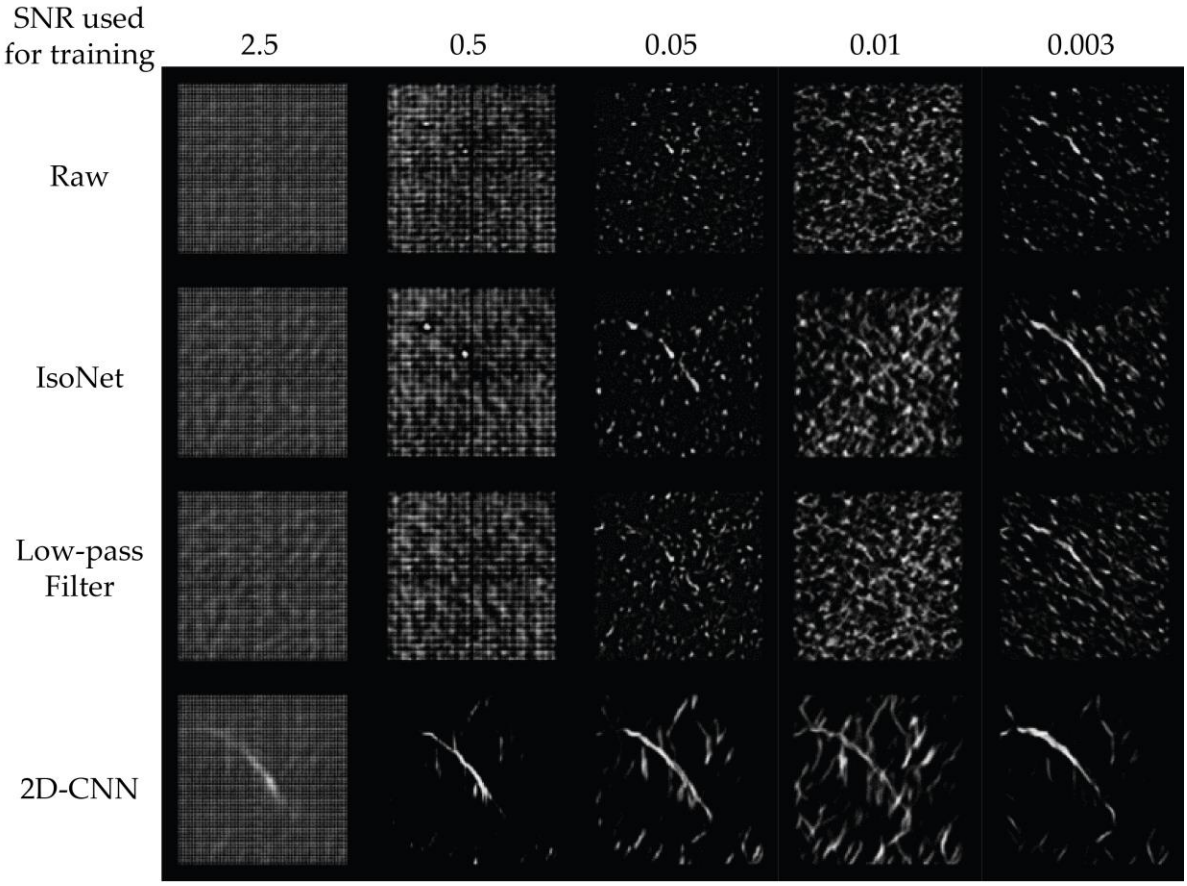

Figure 3.9 Application of five 3D-CNN trained with synthetic data of various SNR on raw data with and without preprocessing. A comparison matrix was constructed using combinations of four types of input (raw data, IsoNet processed data, low-pass filtered 10 Å data, and 2D-CNN predicted data) and five types of networks (with SNRs of 2.5, 0.5, 0.05, 0.01, and 0.003). The prominence of the DNA fibers indicated the level of denoising noise strength.

3.2.4 Reconstitute TEC from tomogram

By integrating the DNA maps from the 3D-CNN network with the locations of globular proteins identified through template matching (Figure 3.10C), we explored the complex interactions among RNAP, NCP, and dCas9 along the DNA. We manually crop the regions where DNA connectivity among these proteins was apparent, enhancing visualization by selectively hiding unrelated background DNA densities (Figure 3.10D-I). This analysis revealed that RNAP and nucleosomes are often closely associated, exhibiting a variety of rotational orientations relative to each other, suggesting direct interactions with nucleosome-bound DNA. Nonetheless, the overall frequency of these complexes was relatively low, suggesting that RNAP frequently traverses and potentially disrupts nucleosomes, leading to their detachment from DNA. This result

highlights the dynamic nature of transcription through nucleosomes, where RNAP not only engages with but also alters the nucleosome landscape.

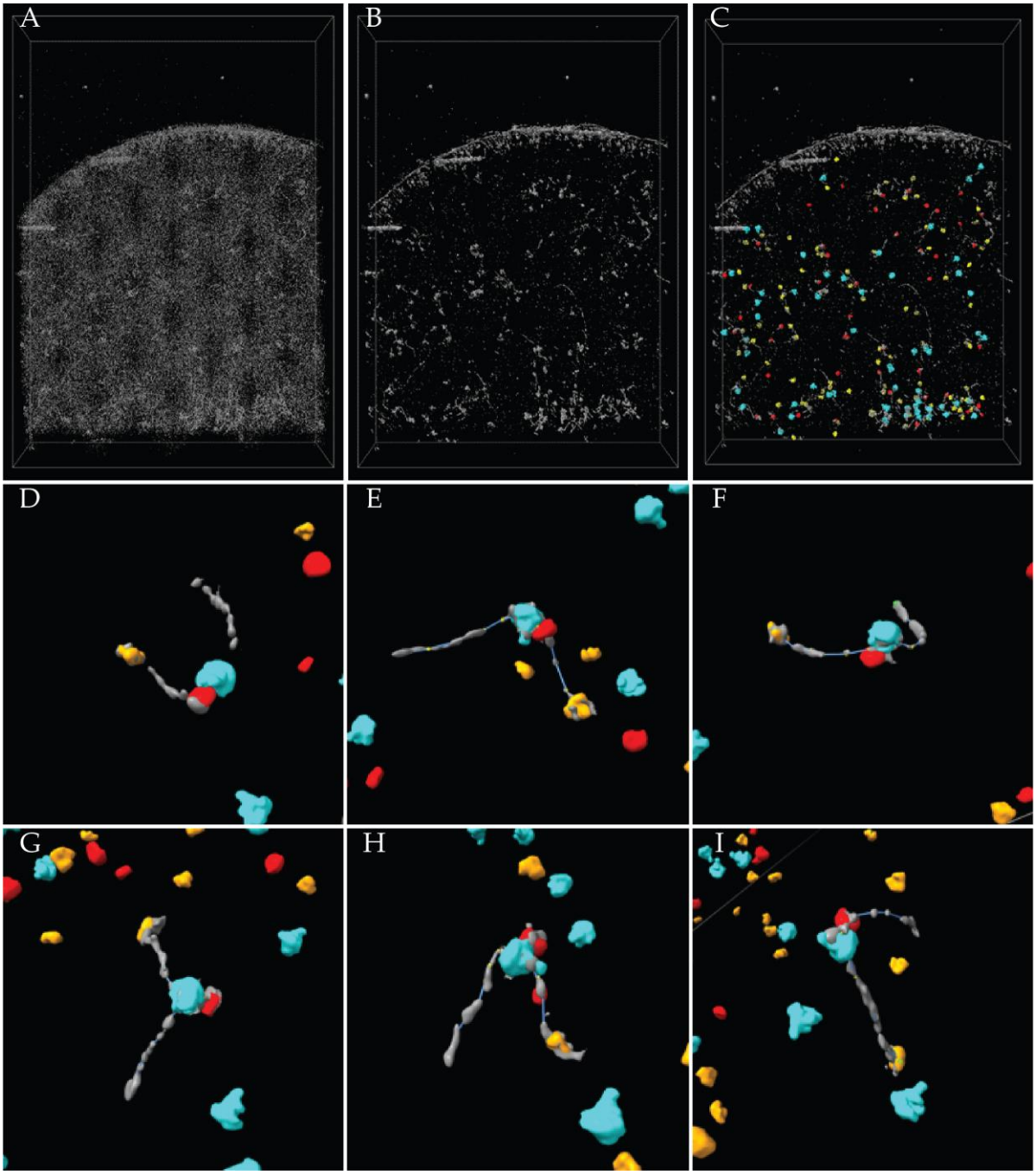

Figure 3.10 Segmentation of TEC after reassembling DNA with globular particles within tomograms. (A) Raw tomogram displaying original imaging data. (B) Image refined by a 3D-CNN trained on data with an SNR of 0.003 and enhanced by IsoNet, illustrating improved feature clarity. (C) Synthetic tomogram with superimposed denoised map from (B), showcasing the integrated structures of RNAP, NCP, and dCas9.

(D, E, F, G, H, I) Highlighted examples of complexes featuring RNAP, NCP, and dCas9, colored in cyan, red, and yellow respectively.

3.3 Conclusions

In this study, we fine-tuned the sample preparation procedures for cryo-ET and successfully enriched TEC in our samples. We also established a comprehensive data analysis pipeline. Our template matching results achieved a resolution of approximately 10 Å for the three types of particles—RNAP, NCP, and dCas9—within the tomograms. This resolution was sufficient to discern the relative orientations between RNAP and nucleosomes. However, while template matching provided valuable insights, it occasionally resulted in misfit particles, suggesting a need for additional refinement, potentially through the application of our 3D-CNN denoising strategy to enhance particle orientation accuracy before another round of sub-tomogram averaging.

Moreover, we explored the utilization of 3D-CNN for detecting and segmenting flexible DNA within cryo-electron tomograms, a task traditionally challenged by DNA's flexible and unstructured nature. Our approach, leveraging advanced machine learning techniques including the IsoNet and REST networks, significantly enhanced DNA signal detection. These networks corrected artifacts and denoised the tomograms, yielding a clearer and more isotropic representation of DNA fibers. This improvement not only facilitates more accurate modeling of biological structures but also enhances our understanding of DNA-protein interactions within their native cellular environments.

Our successful capture of the ternary complex involving RNAP, dCas9, and NCPs highlights the dynamic nature of transcription through chromatin by illustrating various rotational orientations and intricate DNA trajectories. These insights are crucial for understanding how nucleosomal DNA conformationally adapts as RNAP transcribes through it, further illuminating the complexities of gene regulation.

However, to fully elucidate the process of RNAP traversing the nucleosome barrier, more extensive analysis is essential. It is critical to investigate additional RNAP-nucleosome contact conformations and their relative orientations. Future efforts will aim at expanding the collection and analysis of data to capture a wider array of nucleosome configurations and transcription factors, enriching our understanding of the transcription landscape and its regulatory mechanisms. In particular, an important goal will be to characterize the conformation of the upstream DNA and its possible looping conformation.

3.4 Acknowledgements

We thank Prof. Craig Kaplan for providing us with the wild-type yeast RNAP and Dr. César Díaz Celis for the human histone octamers. Additionally, we thank Dr. Meng

Zhang for his contribution in collecting the cryo-ET data, analyzing the results, and editing the draft.

3.5 Material and method

3.5.1 TEC preparation

The RNAP used in this study was kindly provided by Prof. Craig Kaplan, and its purification procedure is detailed in section 2.5.1. Nucleosomal DNA templates were prepared using the methodologies outlined in sections 2.5.1, 2.5.2, and 2.5.3. For the assembly of the TEC, the following components were mixed: 0.33 μM nucleosome, 0.17 μM RNAP, 0.27 μM RNA primer (5′-FAM-AUAAUUAGCUC-3′), 0.2 μM yeast TFIIS, and 1 mM each of UTP, GTP, and CTP, in a 30 μL reaction mixture (20 mM Tris-HCl, pH 8 at room temperature, 40 mM KCl, 5 mM MgCl2, 1 mM DTT). This mixture was then added to 5 μL of Pierce Streptavidin Magnetic Beads (Thermo Fisher Scientific, Waltham, Massachusetts) and incubated at 30°C for an hour with agitation to allow RNAP to load onto the mismatch bubble, restart transcription to reach the T-less cassette, and enable complex binding to the magnetic beads.

Subsequently, the beads were washed three times with 1x TB(40) Buffer (20 mM Tris-HCl, pH 8 at R.T., 40 mM KCl, 5 mM MgCl2, 1 mM DTT) to remove unbound nucleosomes and excess components, including RNAP, TFIIS, RNA, and nucleotide triphosphates. The dCas9 mixture, comprising 2.5 μM dCas9 and 2.5 μM sgRNA (ECR7, sequence listed in the end), was prepared in a 20 μL reaction (20 mM Tris-HCl, pH 8 at room temperature, 40 mM KCl, 5 mM MgCl2, 1 mM DTT) and added to the beads, followed by a 30-minute incubation at 30°C with agitation. Afterward, three washes with 1x TB(40) Buffer were performed to remove excess dCas9 complexes.

The final TEC was released from the beads using 0.1 U/μL EcoRI-HF in a 10 μL reaction (20 mM Tris-HCl, pH 8 at room temperature, 40 mM KCl, 5 mM MgCl2, 1 mM DTT, 3% trehalose) at 37°C for 30 minutes with agitation. To restart transcription prior to plunge freezing, a 1 mM rNTPs mixture (1 mM ATP, 1 mM UTP, 1 mM GTP, 1 mM CTP, 20 mM Tris-HCl, pH 8 at room temperature, 40 mM KCl, 5 mM MgCl2, 1 mM DTT, 3% trehalose) was added to the eluate and incubated at room temperature for 10 minutes.

ECR7: 5′ - mA* mC* mG*UUCUGUACACCCACUGCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC mU* mU* mU* U – 3′

m_* refers Phosphorothioated 2'-O-Methyl bases, e.g. mA*.

3.5.2 Cryo-EM Sample Preparation

The cryo-EM specimens were prepared using the conventional plunge freezing method. Briefly, an aliquot (~3 μL) of incubated sample at ~80 nM was placed onto a 200-mesh Quantifoil copper grid (Q210CR-06, Electron Microscopy Sciences), which had been glow-discharged for 15 seconds using a PELCO easiGlow™ Glow Discharge Cleaning System. After incubating for approximately 10 seconds, the grid was flash-frozen in liquid ethane at ~90% humidity and 8°C with a Leica EM GP rapid-plunging device (Leica Microsystems, Wetzlar, Germany) after blotting with filter paper for a controlled blotting time of 5 seconds. The flash-frozen grids were then transferred into liquid nitrogen for storage.

3.5.3 TEM data acquisition

Cryo-EM specimens were screened using a Titan Krios G2 TEM (Thermo Fisher Scientific, Waltham, Massachusetts) operated at 300 kV with a Gatan energy filter. Cryo-EM tilt image series of the samples were collected from -60° to +60° at 3° increments on a Titan Krios G3 TEM equipped with a Gatan energy filter and a K3 Summit direct electron detection camera. During data acquisition, the SerialEM 4.1 software was used to automatically track the specimen and maintain a defocus of ~2.5 μm. The acquired tilt image series at a magnification of ~50,000x (each pixel corresponds to 1.67 Å for the K3 camera) represents a total dose of ~183 e-/Å^2. For each tilt angle, a total of ~8 frames were collected with an exposure time of 0.25 seconds per frame.

3.5.4 Image preprocessing

The motion of the cryo-EM frames was corrected using MotionCor2. During data collection, a carbon area perpendicular to the tilt axis was included to aid in the detection of the contrast transfer function (CTF) when particles were scarce. The tilt series were initially aligned using IMOD with the patch tracing method. The aligned tilt series were submitted to E2TOMO for CTF correction and 3D reconstruction. Particle picking was performed using the automated procedure in E2TOMO, with positive and negative training set particles manually selected from the bin=4 tomogram. Particle extraction was performed with a box size of 324 pixels on the unbinned tilt series images at a pixel size of 1.67 Å. This particle stack was submitted for sub-tomogram averaging to identify particle positions and orientations. We carried out the 3D refinement routine with a single featureless discoidal-shaped initial model and calculated 3D averaging for a fraction of particles (pkeep=0.8). Map resolutions were estimated by measuring the FSC between two independently determined half-maps, with resolution defined at the point where the FSC fell below 0.143.

3.5.5 Missing wedge correction of 3D map

To reduce the tilt limitation that causes elongation artifacts in the 3D reconstruction, missing wedge correction was performed using the IsoNet method. Briefly, the E2TOMO 3D reconstructed tomogram at bin=4 with a pixel size of 6.68 Å was submitted to IsoNet. Since the whole tomogram was not CTF corrected in E2TOMO, the tomogram was first deconvolved and scaled with a low-frequency signal region (SNR falloff of 0.7). As particles were sparsely distributed, an initial mask at two standard deviations away from the mean was created to select the protein and DNA signals, and boxes of 64x64 pixels were generated within the mask. During the training process, noise was added at default iterations of 10, 15, 20, and 25 with noise levels of 0.05, 0.1, 0.15, and 0.2.

3.5.6 REST enhancement of DNA recognition using simulated data

The DNA model was created using OxDNA software, where a straight-line segment was converted into a 40-bp B-DNA strand PDB structure. This structure was converted into a coarse grained OxDNA format with its topology file. The structure was energy minimized for 2,000,000 steps with default parameter settings, and the last frame was used for a 3,000,000-step relaxation simulation at 277K. Sampling every 10,000 steps generated 300 structures with various conformations. These structures were converted back into PDB files and then submitted for Scipion3 HEMNMA_3D pipeline. In HEMNMA_3D, 10 ground truth volumes (with a voxel size of 6.68 Å in a box of 64x64x64 pixels) were generated for each of the 300 structures for sampling. The corresponding simulated subtomograms were created by adding different levels of noise (with a SNR of 2.5, 0.5, 0.1, 0.01, 0.003) with a constrained tilt range from -60° to +60° and a defocus value of 25,000 Å. For REST network training, the scipion3 created simulated source and target volume pairs were normalized to a range of 0 to 1, and epoch 2 step 1500 was used.

Chapter 4
Monitor transcription elongation in the *Drosophila* embryo

4.1 Introduction

Eukaryotic transcription elongation is a critical phase in the process of gene expression, where RNAP synthesizes RNA by adding ribonucleotides to the 3′ end of growing RNA chain. This phase follows the initiation of transcription and involves RNAP progressing along the DNA template, decoding genetic information into RNA sequences (Sims III et al., 2004).

Transcription elongation is highly regulated and involves numerous factors that influence the efficiency and fidelity of RNA synthesis. These factors include elongation factors that associate with RNAP, altering its processivity and ability to navigate through chromatin. Additionally, the phosphorylation of RNAP CTD is a pivotal event that regulates the interaction of RNAP with various proteins involved in elongation and RNA processing (Hsin and Manley, 2012).

The chromatin landscape poses significant challenges to RNAP during elongation (Chen et al., 2019; Bondarenko et al., 2006; Gaykalova et al., 2015). Histone modification and ATP-dependent chromatin remodeling are essential for facilitating RNAP's passage through nucleosomes, the basic units of chromatin (Zhang et al., 2015). This remodeling not only affects transcription speed and pause frequency but also plays a role in histone eviction and replacement (Schwabish and Struhl, 2006; Akai et al., 2009; Luk et al., 2010).

Furthermore, transcription elongation is closely coupled with RNA processing events. These include the addition of a 5′ cap to the nascent RNA (McCracken et al., 1997; Schroeder et al., 2000), splicing of introns (Fong et al, 2014; Saldi et al., 2021), and cleavage (Dye and Proudfoot, 2001; Doma and Parker, 2006) and polyadenylation of the transcript's 3′end gene (Whitelaw and Proudfoot, 1986; Logan et al., 1987). The coordinated regulation of these processes ensures that the mRNA is correctly processed and eventually transported to the cytoplasm for translation.

Researchers have utilized a diverse array of experimental approaches to study transcription elongation, ranging from reconstituted systems with purified components, as discussed in Chapter 2, to advanced single-molecule techniques. These methodologies have enabled detailed analysis of the roles of various elongation factors, the influence of chromatin structures, and the dynamics of RNA synthesis. However, the use of these simplified in vitro systems, while beneficial for controlled experimental study, may not fully capture the physiological relevance of the findings. In vitro conditions lack the complexity of cellular metabolism, the specificity of cell-type factors, and the long-term

effects and feedback mechanisms inherent to living organisms. Additionally, crucial aspects such as the spatial organization of the nucleus and interactions within cellular compartments, which are vital for a comprehensive understanding of transcription regulation, are missing from in vitro setups.

In vivo studies of transcription elongation provide unmatched insights into the dynamic regulation of gene expression within their natural cellular contexts. Unlike in vitro methods that simulate cellular processes in a controlled environment, in vivo approaches embrace the full complexity of the cellular environment, capturing how transcription is influenced by factors like cellular metabolism, chromatin structure, and endogenous regulatory networks. Among the various techniques employed in in vivo studies of transcription, live cell imaging (Stephens and Allan, 2003) and genome-wide sequencing (Goodwin et al., 2016) are particularly noteworthy. Live cell imaging allows researchers to observe transcription in real time (Tutucci et al., 2018), offering a dynamic view of how cells respond to stimuli (Akolpoglu et al. 2022) or drug treatments (Orth et al., 2011). On the other hand, genome-wide sequencing provides a comprehensive snapshot of genetic activity across the entire genome but offers a more static perspective (Meyer and Liu, 2014). Both approaches are crucial for elucidating the physiological relevance of transcription mechanisms. They provide insights into how transcription is regulated within the natural chromatin state, including interactions with cellular components and the spatial organization within the nucleus. In this chapter, we will focus on utilizing live cell imaging to monitor transcription in real-time, providing dynamic insights into the cellular processes as they occur. The subsequent chapter will delve into genome-wide studies, exploring how underlying genetic modification influence transcription.

The labeling strategy utilized in our research is the MS2 and PP7 stem-loop system, which is designed to visualize specific RNA molecules in real time through the high affinity binding of bacteriophage MS2 and PP7 coat proteins to their corresponding RNA stem-loop structures (Bertrand et al., 1998; Golding et al., 2005; Chao et al., 2008; Larson et al., 2011; Tutucci et al., 2018). By incorporating multiple copies of either the MS2 or PP7 stem-loop sequence into the RNA of interest, and coupling this with maternally deposited MS2 coat proteins fused to a fluorescent marker within the same cells, we are able to meticulously track the expression and spatial distribution of the tagged RNA with exceptional temporal and spatial resolution. This versatile system has been successfully implemented across a range of organisms, including yeast (Bertrand et al., 1998; Tutucci et al., 2018), humans (Vera et al., 2019; Braselmann et al., 2020; Park et al., 2020), *Drosophila* (Hoppe and Ashe, 2021; Jonathan et al., 2021), Dictyostelium (Chubb et al., 2006), and C. elegans (Lee et al., 2019; Hu et al., 2023), demonstrating its broad applicability and effectiveness in studying RNA dynamics in diverse biological contexts (Wells et al., 2016).

Drosophila melanogaster, commonly known as the fruit fly, is highly valued as a model organism due to its brief life cycle, ease of genetic manipulation, and the conservation of its developmental processes across species (Jennings, 2011). The transparency of *Drosophila* embryos is particularly advantageous, allowing for the direct observation and

real-time recording of cellular and molecular activities, making these embryos ideal subjects for live cell imaging (Garcia et al., 2013; Simpson et al., 2014; Greiss et al., 2016). In this study, we employ sophisticated imaging techniques, specifically confocal microscopy, paired with fluorescent markers like the MS2 stem-loop system, to monitor dynamic developmental processes.

In this chapter, we utilize the MS2 and PP7 stem-loop system within a *Drosophila* embryo to monitor transcription elongation in real-time. For this purpose, we designed an artificial reporter construct similar to that reported in Liu et al. (2021). The reporter gene contains 24 copies of MS2 stem loops at the 5' end and 24 copies of PP7 stem loops at 3' end, and placed it under the control of a regulatory DNA containing the hunchback (hb) P2 minimal enhancer and promoter, as illustrated in Figure 4.1A. The *lacZ* sequence from *E.coli* was placed as a natural spacer between the MS2 and PP7 stem loops. To study the effect of NPS in vivo, part of the *lacZ* is replaced by a 5sRNA sequence, a Widom 601 sequence, or 12 copies of Widom 601 sequences.

Incorporating the *lacZ* sequence as a natural spacer between the MS2 and PP7 loops provides a clear separation of two fluorescent signals that allows for measuring the average transcription elongation rate. To delve into the influence of NPS in a live setting, portions of the *lacZ* sequence were substituted with a 5sRNA sequence, a single Widom 601 sequence, or an array of 12 Widom 601 sequences. These modifications are designed to assess how natural and synthetic NPS impact the transcriptional machinery's ability to navigate chromatin within a live organism, thereby offering insights into chromatin structure's role in gene regulation. Building on our in vitro findings from Chapter 2, where RNAP exhibited more frequent stalling on nucleosomal DNA compared to bare DNA, we hypothesized that substituting NPS for the *lacZ* sequence in the reporter would result in increased nucleosome placement on the DNA. This could potentially slow down transcription due to the additional barriers created by these nucleosomes.

4.2 Results

4.2.1 Dual-color reporter for tracking transcription elongation

To effectively monitor transcription elongation using the MS2 and PP7 stem-loop system in *Drosophila* embryos, two critical components must be incorporated into the experimental setup. Firstly, a reporter gene is engineered to include arrays of both MS2 and PP7 stem-loops flanking the sequence of interest. This configuration is crucial for visualizing the transcription elongation process, as these loop structures serve as binding sites for the second necessary component: the MS2 and PP7 coat proteins. These proteins are usually fused to a fluorescent marker and are expressed in mother flies, allowing them to be maternally deposited into the embryos. During transcription of the reporter gene, the emerging RNA transcripts that contain the stem-loop sequences will bind to

these fluorescently tagged coat proteins, enabling real-time imaging of RNA synthesis and elongation within the living embryo.

To conduct our study, we have developed two reporter fly lines, each designed with a distinct sequence inserted between the MS2 stem-loops at the 5' end and the PP7 stem-loops at the 3' end of the reporter gene. In one line, a portion of the *lacZ* gene sequence serves as a spacer (Liu et al., 2021), providing a classic reporter module commonly used in genetic studies. In the other line, arrays of the Widom 601 sequence, which are interspersed with various linker sequences. These linkers are strategically designed to precisely match the sequences found in the corresponding positions of the *lacZ* construct. This deliberate alignment not only allows us to directly compare the effects of the Widom 601 nucleosome positioning sequence against the *lacZ* spacer, under identical spatial configurations but also benefit our forthcoming next-generation sequencing experiment which we will elaborate on chapter 5. The last fly line, referred to as "2-color fly" in subsequent discussions, is used to express the fluorescently labeled coat proteins, MS2-mCherry and PCP-eGFP. This line also expresses Histone-iRFP, serving as a marker to facilitate the determination of mitosis stage (Liu et al., 2021). Detailed information about these three fly lines can be found in the Material and Method section.

To set up the embryo collection cage, approximately 100 virgin females from the 2-color fly line are combined with 30 males from the reporter fly line. Ninety minutes before imaging, the plate in the cage should be replaced with a new one that contains fresh yeast paste to encourage egg laying. After the waiting period, the plate is replaced again to prepare for the next collection cycle. Embryos are then collected around nuclear cycle 13, which can be determined based on their appearance under a stereomicroscope. Approximately 20 embryos were collected and then dechorionated using bleach. After dechorionation, the embryos were carefully mounted between a semipermeable membrane and a coverslip, with halocarbon oil applied in between. This setup ensures appropriate optical clarity for imaging, facilitates oxygen exchange, and prevents dehydration of the embryos during the imaging process. Data collection for each individual embryo started once the embryo reached nuclear cycle 12 or later and continued for 25 minutes following the start of nuclear cycle 14. The timing is critical as the hb gene becomes transcriptionally active at the start of the nuclear cycle, followed by a gradual transition into a transcriptionally silent state (Garcia et al., 2013).

As individual RNAP transcribes through a sequence containing MS2 or PP7 stem loops, the maternally deposited MCP-mCherry and PCP-GFP fusion protein specifically bind to their respective stem loops. This interaction results in the formation of fluorescent puncta, which are visible under a laser-scanning confocal microscope (Figure 4.1B). The intensity of the fluorescent puncta observed in each color channel during imaging directly correlates with the number of actively transcribing RNAP that have elongated past MS2 or PP7 stem loop sequences incorporated into the RNA. This relationship is linear, meaning that as more RNAP pass these points, the fluorescence intensity increases accordingly. However, the fluorescence intensity measurements are expressed in arbitrary units which can vary between the different color channels due to differences in fluorescence properties and detection efficiencies. Detailed procedures of data collection

are described in the Material and Methods section.

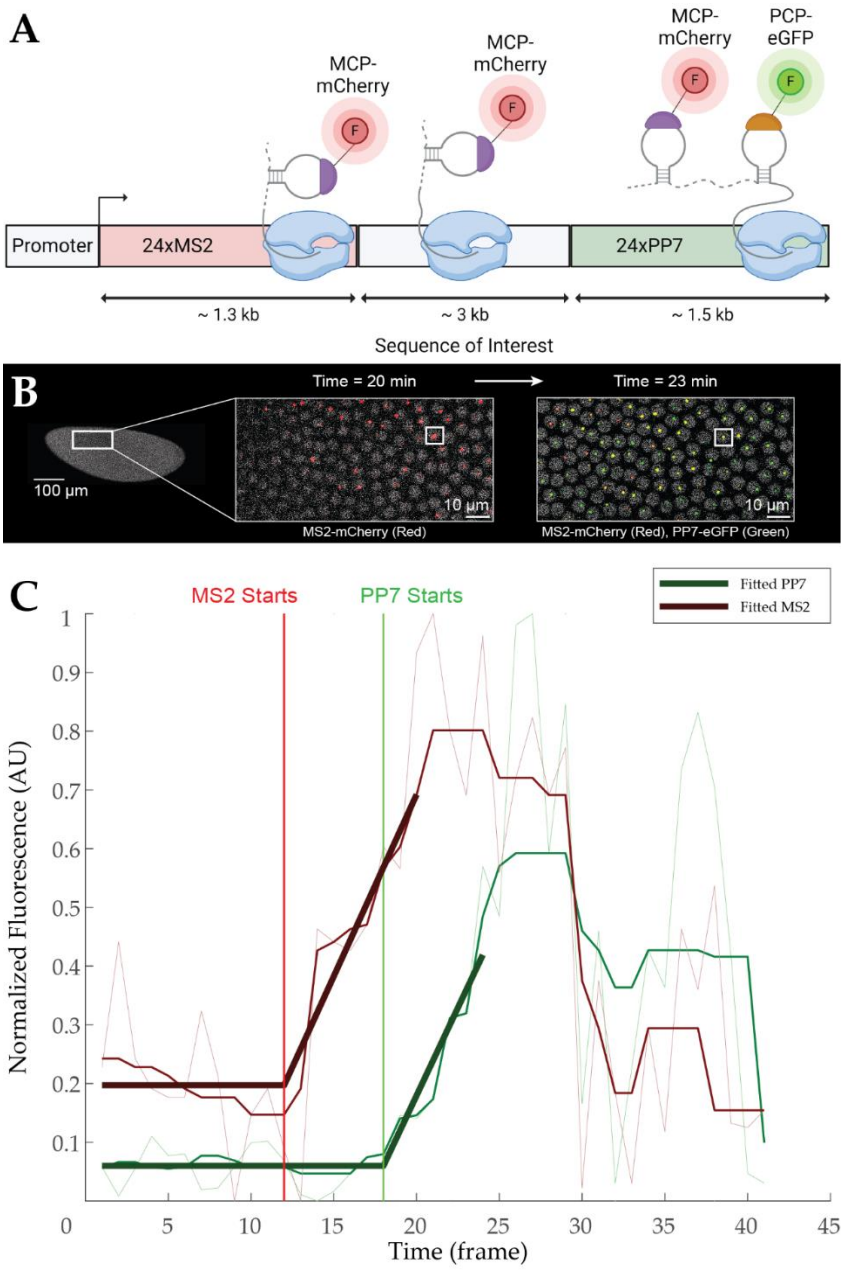

Figure 4.1 Schematic of reporter design and rising time detection. (A) Transcription of the stem loops in the reporter leads to the formation of fluorescent puncta, with the signal from the 5' mCherry appearing prior to that from the 3' GFP. For simplicity, the illustration shows only one stem loop per fluorophore; however, the actual construct includes 24 repeats of each stem loop. (B) Exemplar fluorescence data on transcription elongation dynamics in live cell *Drosophila* embryos. When a tagged gene is active, fluorescent puncta will appear in a microscope's field of view in a sequential manner, red for the MS2 stem loop and green for the PP7. (C) Sample raw single-cell MS2 and PP7 fluorescence traces in light green and red. The raw data is then filtered (in medium

green and red) fitted to a piece-wise function (dark green and red) to determine the rising time for each channel (Indicated by vertical red and green lines for MS2 and PP7 channels). Created with BioRender.com.

In this study, we developed a computational method to accurately measure the transcription elongation rate based on fluorescence data. The fluorescence signals from each channel display four distinct phases. Initially, as RNAP begins transcription at the promoter and moves through the sequence preceding the MS2 region, the fluorescence signals maintain a flat, background level. Once the first RNAP reaches the MS2 or PP7 regions, the corresponding fluorescence signals start to rise at a constant rate. This increase continues until RNAP reaches the end of the gene and the nascent transcript is cleaved, leading to a plateau in the fluorescence as new RNAP molecules initiate transcription from the promoter. Eventually, after transcription is halted and no new RNAP molecules are initiated, the fluorescence signals begin to decline.

To calculate the transcription elongation rate, the analysis starts by pinpointing the exact frame where the fluorescence signals begin to rise, marking the moment RNAP reaches the MS2 or PP7 repeat region in the reporter. The transcriptional dynamics in the fluorescence data are modeled using a continuous piecewise function that includes both a flat and a linear rise segment to represent the background and increasing phases, respectively. To improve accuracy, the analysis focuses only on the initial rise, terminating once the trace achieves 75% of its maximum intensity to avoid distortion from the plateau and decline phases. The crucial step involves identifying the transition point where the fluorescence signal shifts from a flat background to a linear increase, indicating the first encounter of RNAP with the MS2 or PP7 stem loop sequences. Simultaneous analysis of both the green and red fluorescence channels allows for the precise measurement of the time delay between the increases in these channels, providing a quantitative estimate of the time required for RNAP to transcribe the intervening 4.3 kb segment, which includes the MS2 stem loops and spacer sequences (Figure 4.1C).

4.2.2 NPS do not slow down transcription elongation

In this study, we assessed the transcription elongation rates of two distinct reporter constructs. The first construct incorporates a portion of the *lacZ* gene as a spacer between the MS2 and PP7 stem loops (*lacZ* reporter). This construct was previously studied by Liu et al. (2021), who reported an average transcription elongation rate of 1.72 ± 0.05 kb/min (SEM, n = 355). Our measurements using the same construct showed a rate of 2.04 ± 0.06 kb/min (SEM, n = 310), aligning closely with Jonathan's findings and confirming earlier results (Garcia et al., 2013; Fukaya et al., 2017). On the other hand, a different reporter construct utilizing twelve copies of the Widom 601 NPS as the spacer (12x601 reporter) yielded a slightly higher elongation rate of 2.10 ± 0.03 kb/min (SEM, n = 1435). These results, illustrated in Figure 4.2, are consistent with Liu et al. (2021),

demonstrating the reliability of our analysis method despite variations in approach. However, the data distributions for both the *lacZ* and 12x601 constructs were broader and included extreme outliers, likely reflecting limitations in our analytical method. These issues will be further discussed in the subsequent section.

These observations suggest that the chromatin context provided by the Widom 601 NPS does not dramatically alter the transcription elongation rate compared to the *lacZ* sequence. One possible explanation for this could be that the Widom 601 NPS, although well-designed to dictate nucleosome positioning in vitro, may not effectively influence nucleosome behavior in vivo. We will further discuss the nucleosome positioning in the embryo in the next chapter. In the complex environment of a living cell, other factors such as chromatin remodelers, histone modifiers, and the natural chromatin landscape can diminish the impact of engineered NPS like Widom 601. This highlights the challenges of translating in vitro findings to in vivo contexts, where the cellular machinery and chromatin dynamics are more intricate and less predictable.

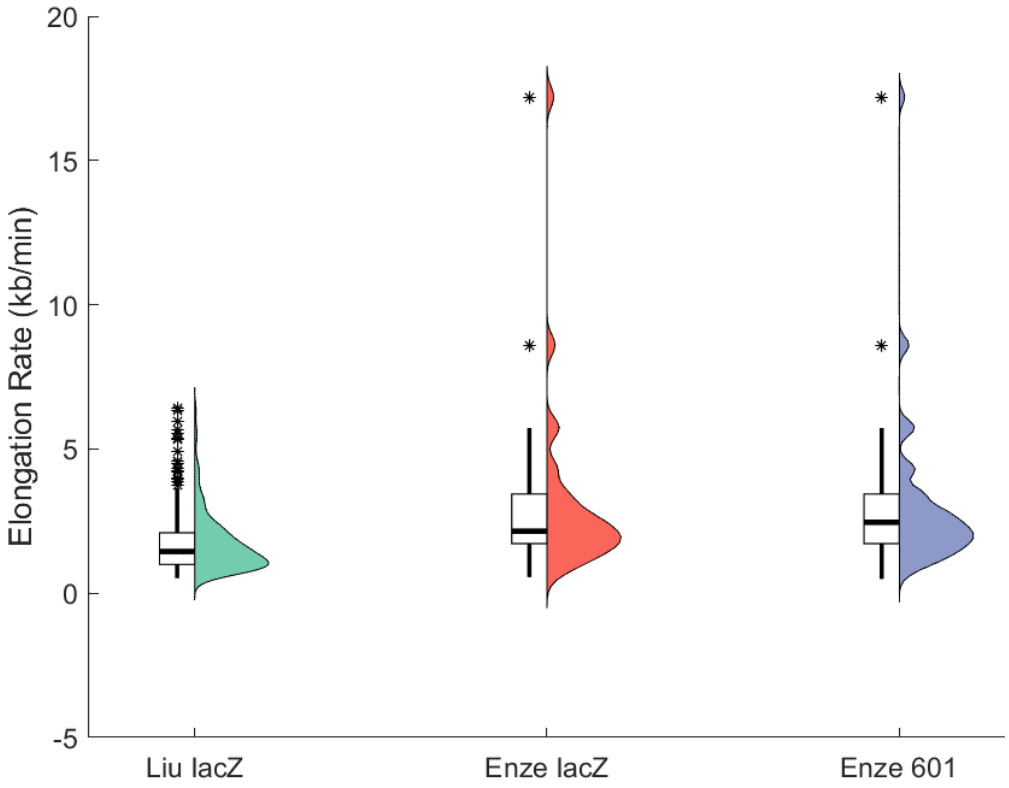

Figure 4.2 Comparison of transcription elongation rate measured with dual-color construct in Drosophila embryo. Both our *lacZ* and 601 constructs demonstrate mean elongation rates similar to those reported by Liu et al. While the distribution shapes around the means are comparable, ours show greater spread, indicating wider variability. Additionally, our results include more pronounced outliers, highlighting potential limitations in our analysis method.

However, our observed average transcription elongation rates with both *lacZ* and 12x601 are slightly faster than the earlier estimations derived from traditional methods. Specifically, northern blotting assays conducted on *Drosophila* larval tissues (Thummel et al., 1990) and nuclear run-on assay utilizing extracts from *Drosophila* S2 cell (O'Brien and Lis, 1993). A plausible explanation for the observed increase in the average transcription elongation rate is that the earlier estimations were primarily based on the analysis of heat shock genes. These studies required elevated temperatures for induction, which might have led to attenuated rates of transcription elongation. Heat stress can impact cellular mechanisms and potentially slow down enzymatic activities (Cardiello et al., 2018), hence potentially explaining the slower elongation rates observed in these earlier studies. In our experimental setup, measurements were conducted under optimal culturing conditions, specifically maintained at around 22°C. This temperature is considered ideal for *Drosophila* development and does not impose the thermal stress associated with heat shock protocols used in previous studies. Consequently, this could contribute to the higher transcription elongation rates observed in our experiments, as the transcription machinery operates more efficiently without the stress factors that potentially hinder its activity in elevated temperatures.

4.2.3 Cell cycle does not affect transcription elongation rate

We next aimed to explore whether variations in the cell cycle stage could influence the rate of transcription elongation. Our data acquisition spanned nuclear cycle (NC) 13, 14, and sometimes 12, allowing us to assess if different stages of the cell cycle impact the efficiency of the transcriptional machinery. As previously mentioned, we employed a computational approach to analyze the transcription data collected from both MS2 and PP7 fluorescence channels. We fitted a piecewise function to the data from each channel to determine the point at which the fluorescent signals began to rise significantly above the background level. By calculating the delay in frames between the rise in fluorescence in the MS2 and PP7 channels, we could quantify the time difference in transcription progression between the two sites. Combining this with the known length of the transcript between the sites (4.3 kb) and the duration of each frame (15 seconds), we were able to calculate transcription elongation rate from the fluorescent data.

Except that we only have data from 12x601 construct in NC12 (2.66 ± 0.21 kb/min, SEM, n = 45), the transcription elongation rates for *lacZ* (2.09 ± 0.12 kb/min, SEM, n = 103 in NC13 and 2.01 ± 0.08 kb/min, SEM, n = 207) and 12x601 constructs (2.00 ± 0.05 kb/min, SEM, n = 458 in NC13 and 2.13 ± 0.04 kb/min, SEM, n = 932) are very similar to each other, as illustrated in Figure 4.3. This consistency suggests that the transcription elongation rates for these constructs are not significantly impacted by the different spacers used (*lacZ* versus Widom 601 sequences), across these stages of the cell cycle. This could indicate that the transcriptional machinery in *Drosophila* embryos maintains a relatively stable elongation rate across these nuclear cycles, irrespective of the specific underline

sequences provided by different spacer elements.

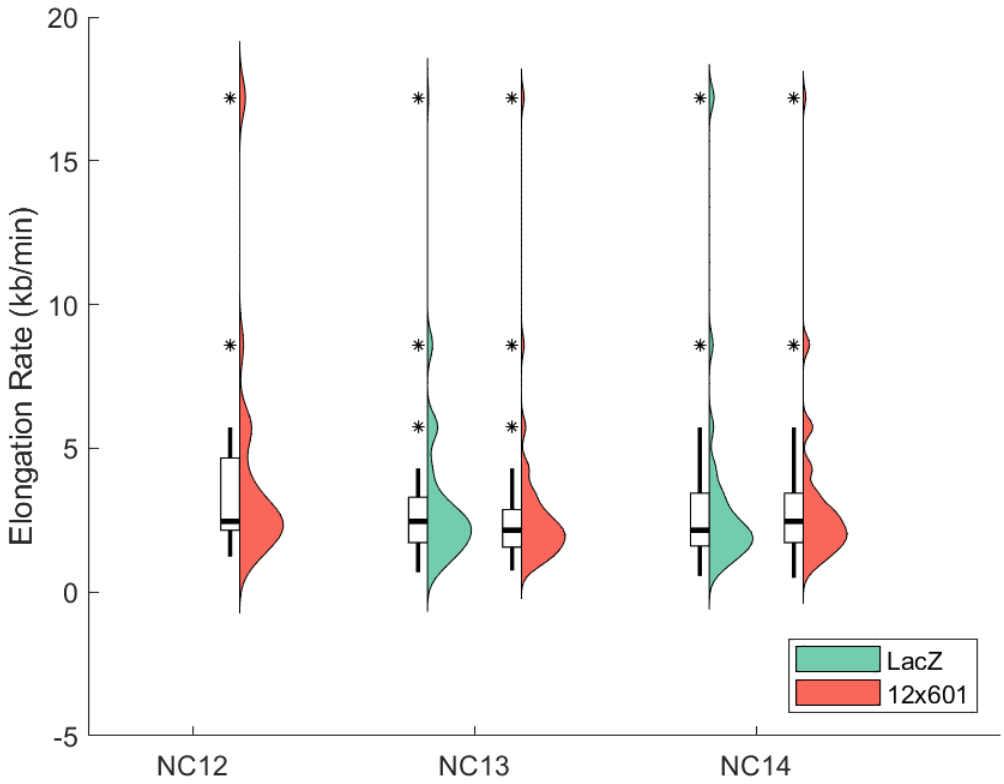

Figure 4.3 Transcription elongation rate of lacZ and 12x601 constructs at different NC. The transcription elongation rates of the *lacZ* and 12x601 constructs across different nuclear cycles (NC13 and NC14) are depicted through a combination of box plots and violin plots, which illustrate the distribution of the rates. Outliers in the data are marked with asterisks for clarity. The analysis shows no significant differences in transcription elongation rates within the same construct across different nuclear cycles. Additionally, the introduction of 12 copies of Widom 601 NPS into the constructs did not influence the transcription elongation rate at either NC13 or NC14.

4.2.4 Low SNR of MS2 channel results in poor fitting

The pivotal step in calculating transcription elongation rates from fluorescent traces involves identifying the rise time for each channel. As detailed in section 4.2.1, the fluorescent traces were modeled using a two-phase piece-wise function, as depicted in Figure 4.4A. The fitting shows that the PP7 signal begins to rise above the background 7 frames after the MS2 signal, yielding an elongation rate of 2.51 kb/min. However, the MS2 channel fitting was occasionally inaccurate due to poor SNR, as highlighted by two problematic examples in Figures 4.4B and C, where the PP7 signal rises either before or

immediately after the MS2 signal, resulting in biologically implausible elongation rates, such as 17.6 kb/min.

The frequent fitting failures can be attributed to the optical properties of the fluorophores; MCP, which binds to the MS2 stem loop, is conjugated to mCherry, while PCP, associated with the PP7 stem loop, is linked to GFP. The raw fluorescent signal from the MS2 channel, shown in Figure 4.4D, exhibits a lower peak and higher background compared to the GFP channel, presented in Figure 4.4G. Analysis across both the *lacZ* and 12x601 reporters reveals that the MS2 channel has an average background of 64.8 AU (Figure 4.4E) and a peak mean of 103 (Figure 4.4F), resulting in an SNR of 0.59 after background subtraction. Conversely, the PP7 channel maintains a lower average background of 47.6 (Figure 4.4H) and a higher peak mean of 198 (Figure 4.4I), achieving an SNR of 3.16—approximately five times higher than the MS2 channel. The elevated background, reduced peak, and low SNR in the MS2 channel significantly impair fitting accuracy, leading to outliers in our transcription elongation rate measurements. Addressing this issue might involve adopting a more effective labeling strategy or employing advanced modeling techniques for fluorescent signal analysis.

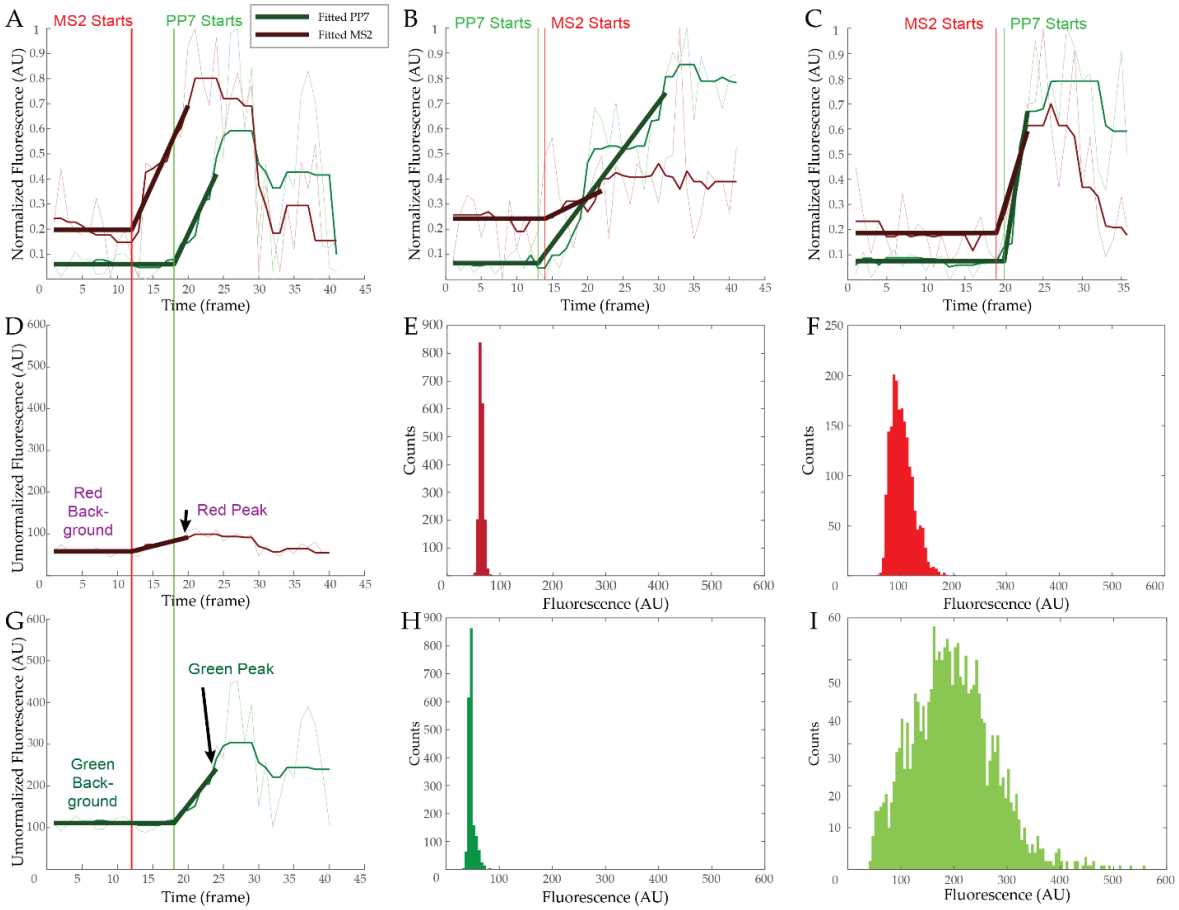

Figure 4.4 Low SNR of MS2 Channel Compromises Fitting Accuracy. (A) Accurate fittings of both MS2 and PP7 channels reveal a 7-frame delay in the PP7 signal rising above background, compared to the MS2 signal. (B) Incorrect fitting where the PP7 signal

rises before the MS2 signal, resulting in a negative elongation rate. (C) Incorrect fitting where the PP7 signal rises just 1 frame after the MS2 signal, leading to a biologically implausible elongation rate of 17.6 kb/min. (D) Raw MS2 fluorescent signal from (A); the background level is determined by the flat phase of the fitting, and the peak level corresponds to the highest value achieved during the fitting, marked at 75% of the maximum raw signal. (E) Distribution of MS2 background levels with an average of 64.8 AU. (F) Distribution of MS2 peak levels with an average of 103 AU; the resulting SNR for the MS2 channel is 0.59 after background subtraction. (G) Raw PP7 fluorescent signal from (A), with background and peak levels defined similarly to the MS2 channel. (H) Distribution of PP7 background levels, averaging 47.6 AU. (I) Distribution of PP7 peak levels, averaging 198 AU; the SNR for the PP7 channel is 3.16.

4.2.5 Anterior-Posterior (AP) position does affect transcription elongation rate

Finally, we turned our attention to exploring the potential correlation between transcription elongation rate and gene expression levels. This investigation is particularly interesting in the context of *Drosophila* embryogenesis, where the maternal Bicoid activator establishes a gradient that is essential for regulating AP patterning genes like *hunchback* (*hb*) in a concentration-dependent manner (Nien et al., 2011). Given that our reporter constructs are all driven by the *hb* promoter, by analyzing how AP patterning influences transcription elongation rates, we can gain deeper insights into how varying levels of gene expression may modulate the transcription elongation rate during early developmental processes.

To analyze transcription elongation rates in relation to gene expression levels, we segmented data from both the *lacZ* and 12x601 construct based on its anterior-posterior (AP) position within *Drosophila* embryos, as shown in Figure 4.5A. This segmentation allowed for a detailed examination of how transcription elongation rates vary across different spatial regions of the embryo. By comparing these rates at various AP positions in both NC13 and NC14, we confirmed that the dynamics of transcription elongation remain consistent across these developmental stages, supporting conclusions previously drawn in section 4.2.3. In the anterior 40% of the embryo, the transcription elongation rate consistently remains around 2 kb/min in both NC13 and NC14, as illustrated in Figure 4.5A and C. Beyond this anterior region, there is a noticeable decline in transcription elongation rate, coinciding with the decreasing gradient of Bicoid protein concentration. As Bicoid concentration decreases along the AP axis, it leads to a corresponding reduction in hb promoter activity. This correlation suggests a complex interaction between transcription initiation and chromatin structure, indicating that the level of transcription initiation—affected by transcription factor concentrations—may influence chromatin configuration, which in turn impacts transcription elongation rates. However, for the *lacZ* reporter, insufficient data were collected in NC13 (Figure 4.5C) to establish a clear trend. In NC14, the elongation rate begins to drop after the anterior 50%

of the embryo, exhibiting a trend similar to that observed with the 12x601 reporter. This pattern suggests that the introduction of an array of NPS into the genome may not efficiently position nucleosomes. We will further explore this issue in the next chapter.

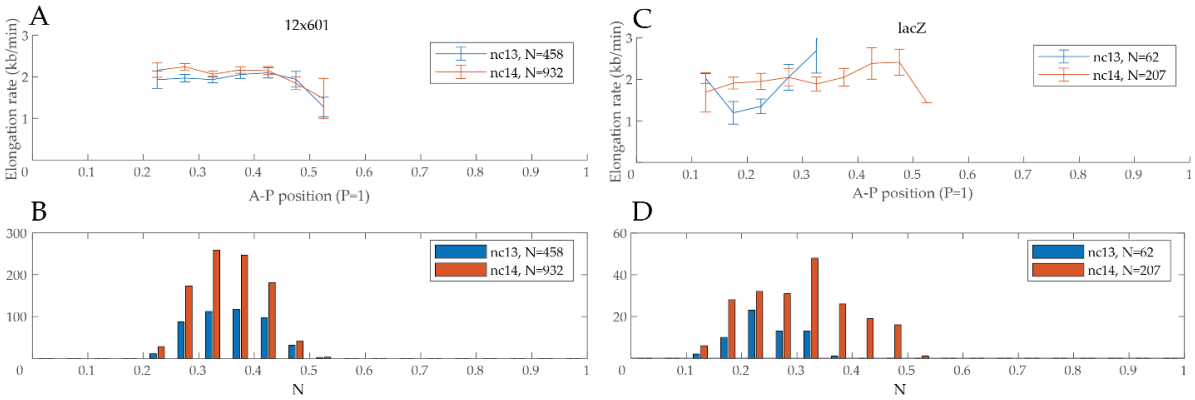

Figure 4.5 Variation in transcription elongation rates across the anterior-posterior (A-P) axis, independent of sequence context. Transcription elongation rates from multiple embryos using the 12x601 and *lacZ* reporters were analyzed and categorized based on their positions along the A-P axis. (A) Shows a consistent pattern in transcription elongation rates for both NC13 and NC14 of the 12x601 reporter, with higher rates observed in nuclei closer to the anterior. (B) Displays the number of nuclei analyzed in each positional bin for the 12x601 reporter. (C) Illustrates that in the *lacZ* reporter, the transcription elongation rate begins to decline after approximately the anterior 45% of the embryo in NC14. Data beyond the anterior 30% in NC13 are lacking, preventing a conclusive analysis. (D) Shows the number of nuclei analyzed in each positional bin for the *lacZ* reporter.

4.3 Conclusions

In this chapter, we have expanded our investigations from in vitro reconstituted systems to explore the dynamics of transcription elongation within the context of *Drosophila* embryogenesis. Employing both the MS2 and PP7 stem-loop systems along with confocal microscopy, we monitored real-time transcription, providing a direct view of transcriptional processes in live embryos. Our studies were specifically focused on assessing the effects of NPS and exploring how variations in nuclear cycles and anterior-posterior positional effects influence transcription elongation rates. This comprehensive approach has allowed us to gain a nuanced understanding of gene expression regulation during critical developmental phases, highlighting the complex interplay between chromatin structure and transcriptional machinery.

Our findings confirmed consistent transcription elongation rates across nuclear cycles 13 and 14, regardless of the spacer sequence used. This consistency underscores the robustness of the transcription machinery across different developmental stages and suggests that introducing array of NPS in vivo is not sufficient to alter transcription

elongation. Furthermore, our analysis revealed that the transcription elongation rate remains stable at approximately 2 kb/min in the anterior 40% of the embryo, demonstrating that the rates are not significantly impacted by the AP positional variance within this range. However, beyond this anterior segment, we noted a decline in the transcription elongation rate, correlating with a decrease in Bicoid protein concentration and *hb* promoter activity. This pattern suggests a potential linkage between transcription factor concentration gradients, chromatin structure alterations, and transcription elongation dynamics.

The approach and findings discussed in this chapter offer valuable insights into the complex interplay between chromatin environment and transcriptional control mechanisms during early developmental processes in *Drosophila*. These insights are crucial for further studies on gene regulation and embryonic development, providing a foundation for understanding how genetic and epigenetic factors influence transcription across biological systems.

4.4 Acknowledgements

We extend our gratitude to Prof. Hernan Garcia for generously providing the fly lines and the microscope, and for his insightful comments on this thesis. We are also thank Dr. Jonathan Liu and Dr. Jiaxi Zhao for their expert training in fly manipulation, imaging acquisition, and image processing. Additionally, we would like to thank Dr. Alex Tong for his pivotal role in developing the data analysis pipeline.

4.5 Material and Methods

4.5.1 DNA constructs

The plasmid containing the reporter construct P2P-MS2-lacZ-PP7 (referenced in Liu et al., 2021) was kindly provided by the Garcia Lab. Subsequent reporter construct, the P2P-MS2-12x601-PP7 plasmid was cloned using custom services from Genescript, involving the introduction of twelve copies of the Widom 601 NPS into the *lacZ* region. These plasmids were integrated into the *Drosophila* genome at the 38F1 landing site utilizing PhiC31-mediated Recombinase Mediated Cassette Exchange (RMCE), with integration services provided by BestGene (Chino Hills, Carlifornia).

4.5.2 Fly strains

Transcription of the reporters were measured by imaging embryos resulting from

crossing yw;MCP-NoNLS-mCherry,Histone-iRFP;MCP-NoNLS-mCherry, PCP-NoNLS-GFP (referenced in Liu er al., 2021) female virgins with yw;P2P-MS2-lacZ-PP7 or yw;P2P-MS2-12x601-PP7 males.

4.5.3 Sample preparation and data collection

Sample preparation for imaging was conducted according to the protocols detailed in Garcia and Gregor (2018) and Liu et al. (2021). Briefly, embryos were collected two hours after replacing the cage lid, dechorinated using bleach, and then mounted between a semipermeable membrane (Lumox film, Starstedt, Germany) and a coverslip. The embryos were embedded in Halocarbon 27 oil (Sigma) and excess oil was removed from the sides using absorbent paper to slightly flatten the embryos. Imaging was performed on a Leica SP8 scanning confocal microscope (Leica Microsystems, Wetzlar, Germany). Fluorescent proteins MCP-mCherry, PCP-eGFP, and Histone-iRFP were excited using a White Light Laser at wavelengths of 488 nm, 587 nm, and 670 nm, respectively. Average laser powers measured at the output of a 10X objective were 35 μW for eGFP and 20 μW for mCherry. Three Hybrid Detectors (HyD) captured the fluorescent signals within spectral windows of 496-546 nm for eGFP, 600-660 nm for mCherry, and 700-800 nm for iRFP. The confocal imaging stack included 15 slices, spaced 0.5 μm apart, covering a total z-height of 7 μm. Images were acquired every 15 seconds at a resolution of 512 × 128 pixels with a pixel size of 202 nm and a pixel dwell time of 1.2 μs. Signal acquisition involved accumulating data over three repetitions for each frame. Data were collected from 310 cells in 2 embryos with the *lacZ* reporter and 1435 cells in 8 embryos with the 12x601 reporter, with imaging conducted up to 25 minutes post-NC14.

4.5.4 Image processing and data analysis

Images were analyzed using custom-written MATLAB software, adhering to protocols from Garcia et al. (2013) and Lammers et al. (2020), accessible at a public Github repository (https://github.com/GarciaLab/mRNADynamics). This software automates the analysis of microscope images by segmenting individual nuclei using the Histone-iRFP signal as a nuclear mask, segmenting transcription spots based on fluorescence, and quantifying the intensity of each MCP-mCherry and PCP-eGFP spot within a nucleus over time.

For precise localization of fluorescence puncta, we used the Trainable Weka Segmentation plugin for FIJI (Arganda-Carreras et al., 2017), which employs the FastRandomForest algorithm. This generates a probability map for each 2D image (Z-slice at a specific time) that predicts the likelihood of each pixel belonging to a punctum. To process fluorescence and probability data, which consist of stacks of 2D images at different Z-levels, we flattened the images into a single 2D image by selecting the maximum pixel value along the Z-axis. Each punctum is defined at each time point as

the area where the probability map exceeds half of the image's maximum probability value—for instance, if the highest value is 80%, the threshold is set at 40%.

Puncta associations across frames and colors were determined by checking for overlaps: a punctum at time i is linked to one at time i-1 if they overlap. Likewise, a red punctum at time i is associated with a green punctum at the same time if they overlap. The centroids of these spots provide coordinates over time for each color (x, y, time, color), marking the center of each punctum in the images. Fluorescence values are then summed within a 4-pixel radius around each centroid, creating a dataset of red and green fluorescence for each spot.

The rise times are calculated by cropping the data to focus solely on the initial increase in fluorescence, discontinuing the analysis once the trace reaches 75% of its maximum. This selected portion is fitted to a flat - linear piecewise continuous function, representing both the stable background and the transcription-driven rise. The time at which this fit begins to rise is considered the moment when RNAP reaches the start of the corresponding MS2/PP7 repeat section in the transcript. By analyzing both the green and red fluorescence channels in this manner, we can calculate the time delay between their increases, providing an estimate of the time required to transcribe the intervening 4.3 kb segment containing the MS2 loops and insert sections. The code is written in MATLAB and is available at https://github.com/abmtong/BLabOTMatlab in the folder \Misc\Drosophila.

Chapter 5
Evaluate nucleosome occupancy using ATAC-seq

5.1 Introduction

In Chapter 4, the transcription elongation rate is assessed in a *Drosophila* embryo using artificially constructed reporter genes embedded into the genome by confocal microscopy. These reporters include segments of the *LacZ* gene or multiple copies of the Widom 601 sequences, but the actual nucleosome occupancy of these reporter constructs within the chromatin context of the embryo remains unmeasured. Understanding nucleosome occupancy is crucial because it can significantly influence transcriptional dynamics by affecting how RNAP interacts with the DNA. Nucleosomes can act as barriers to transcription elongation.

To address this gap, further NGS experiments could be conducted. This method would provide detailed insights into the chromatin structure around these reporter constructs, clarifying how nucleosome arrangement might be influencing transcriptional activity and elongation rates observed in confocal microscopy experiments. Such data are essential for a comprehensive understanding of the interplay between chromatin architecture and gene regulation within the developmental context of *Drosophila* embryos.

Various NGS techniques have been developed to provide deep insights into chromatin biology (Yan et al., 2020), each serving unique purposes in the study of the epigenetic landscape, chromatin accessibility, and TF binding. These techniques include:

1) ATAC-seq (Assay for Transposase-Accessible Chromatin with sequencing): Utilizes a hyperactive Tn5 transposase to probe open chromatin areas, making it excellent for studying chromatin accessibility and indirectly inferring nucleosome locations (Buenrostro et al., 2015; Buenrostro et al., 2015a; Cusanovich et al., 2018).
2) DNase-seq (DNase I hypersensitive sites sequencing): Identifies regions of DNA that are sensitive to cleavage by the enzyme DNase I, which preferentially targets accessible chromatin regions, providing insights into regulatory elements that are active in particular cells or under specific conditions (Boyle et al., 2008; Song and Crawford, 2010; Meuleman et al., 2020).
3) FAIRE-seq (Formaldehyde-Assisted Isolation of Regulatory Elements sequencing): Enriches for nucleosome-depleted regions of DNA by exploiting the differential solubility of nucleosome-bound and free DNA following formaldehyde cross-linking, which is useful for identifying regions like enhancers and promoters that are free of nucleosomes (Giresi et al., 2007; Simon et al., 2013).

4) ChIP-seq (Chromatin Immunoprecipitation sequencing): Targets protein-DNA interactions, using antibodies specific to either histone modifications or transcription factors. This method allows researchers to determine the specific locations where proteins bind to DNA, revealing regulatory sites and the impact of histone modifications on gene expression (Park, 2009; Raha et al., 2010).
5) MNase-seq (Micrococcal Nuclease sequencing): Uses micrococcal nuclease, an enzyme that digests linker DNA between nucleosomes, thereby identifying positions of nucleosomes along the genome. This technique provides precise mapping of nucleosome locations, essential for understanding the role of nucleosome positioning in regulating access to genetic information (Schones et al., 2008; Cui and Zhao, 2011).

Each of these methods offers a distinctive perspective on the complex mechanisms regulating gene expression, providing critical data that help delineate the functional elements of the genome. By understanding how these elements are influenced by chromatin structure and dynamics, researchers can gain insights into the intricate processes that control gene activity across different biological contexts and conditions.

In this chapter, we employed ATAC-seq, a technique that is becoming increasingly popular for evaluating nucleosome occupancy across the genome. One of the significant advantages of ATAC-seq is its efficiency with minimal cell input, 500-50000 cells or nuclei, allowing for the analysis of chromatin accessibility even in samples where cell numbers are limited. While the sensitivity and specificity of ATAC-seq are comparable to DNase-seq but superior to FAIRE-seq where both methods require millions of cells as input materials (Buenrostro et al., 2013). Additionally, ATAC-seq does not require prior knowledge of the epigenetic marks or transcription factors involved in regulating the chromatin environment (Grandi et al., 2022). This feature is particularly advantageous as it permits a broad application across various biological systems without the need for customized reagents or probes specific to particular DNA-binding proteins or histone modifications.

ATAC-seq utilizes a hyperactive Tn5 transposase that performs two critical functions: it cuts DNA at open chromatin sites and simultaneously inserts sequencing adapters. This dual functionality streamlines the library preparation process for high-throughput sequencing, effectively highlighting genomic regions that are not tightly bound by nucleosomes or other regulatory proteins (Buenrostro et al., 2015). As a result, ATAC-seq provides a comprehensive view of chromatin accessibility across the genome. The data generated offers a detailed map of accessible DNA sites, revealing insights into the regulatory landscapes that influence gene expression (Grandi et al., 2022). This method is highly valued for its efficiency and the depth of information it provides, making it a crucial tool in the study of genomic regulation and function.

In the following, we have conducted ATAC-seq to assess chromatin accessibility and nucleosome positioning during nuclear cycle 14, aligning this analysis with the collection of our confocal microscopy data. This approach provides a detailed examination of chromatin dynamics at this particular phase of *Drosophila* embryonic development. By obtaining a comprehensive view of the chromatin landscape, we aim to understand how

it influences transcriptional elongation and the regulatory mechanisms active during these pivotal developmental stages. Integrating ATAC-seq data with confocal microscopy observations enables us to explore the reasons why introducing multiple NPS into the *Drosophila* genome does not alter transcription elongation rates as expected. This analysis helps to uncover the complexities of chromatin architecture and its impact on gene expression, potentially revealing why certain theoretical modifications do not produce the anticipated practical effects in the dynamics of transcriptional regulation.

5.2 Results

5.2.1 ATAC-seq library preparation and data collection

The ATAC-seq libraries were created using multiple embryos at nuclear cycle 14 from specific reporter fly lines that were previously utilized in confocal microscopy studies. These libraries were made from fly lines with variations in spacer sequences: one library utilized the *LacZ* line which incorporates a portion of *lacZ* gene as a spacer (*lacZ* construct), while another 12x601 line which features twelve copies of the Widom 601 sequence (12x601 construct) to assess the nucleosome positioning. Additional libraries were also generated from lines containing a single copy of the Widom 601 sequence (1x601 construct) and a single 5sRNA sequence (1x5sRNA construct). This setup enables a detailed comparative analysis of how various DNA sequences influence nucleosome occupancy and chromatin accessibility within the cellular context.

For each ATAC-seq library, ten embryos at the desired developmental stage were collected. The collection process involved several steps to ensure the integrity and cleanliness of the samples. First, the embryos were dechorionated to remove the outer layer, which facilitates access to the cells. After dechorionation, the embryos were thoroughly washed to eliminate any residues or contaminants that could interfere with the subsequent steps. Finally, the embryos were homogenized to break the cell membranes, a crucial step that ensures the release of nuclei needed for the ATAC-seq protocol. Then the nuclei were mixed with hyperactive Tn5 transposase to fragmentize the genomic DNA. After the collection and preparation of nuclei from the embryos, they are mixed with a hyperactive Tn5 transposase from Illumina. This enzyme plays a critical role in the ATAC-seq protocol by accessing exposed DNA in open chromatin regions. It performs a cut-and-paste reaction, simultaneously fragmenting the DNA and inserting single-strand DNA adapters at both the 5' and 3' ends of each DNA fragment. These adapters are crucial as they provide the necessary sequences for the subsequent amplification steps. The DNA fragments, now with adapters attached, were purified to remove enzymes and other impurities that could interfere with subsequent processes. Once the DNA is purified, a quantitative PCR (qPCR) reaction was conducted to determine the optimal number of PCR cycles for library amplification. This step is crucial to avoid over-amplification of the library, which can skew the representation of different

DNA fragments.

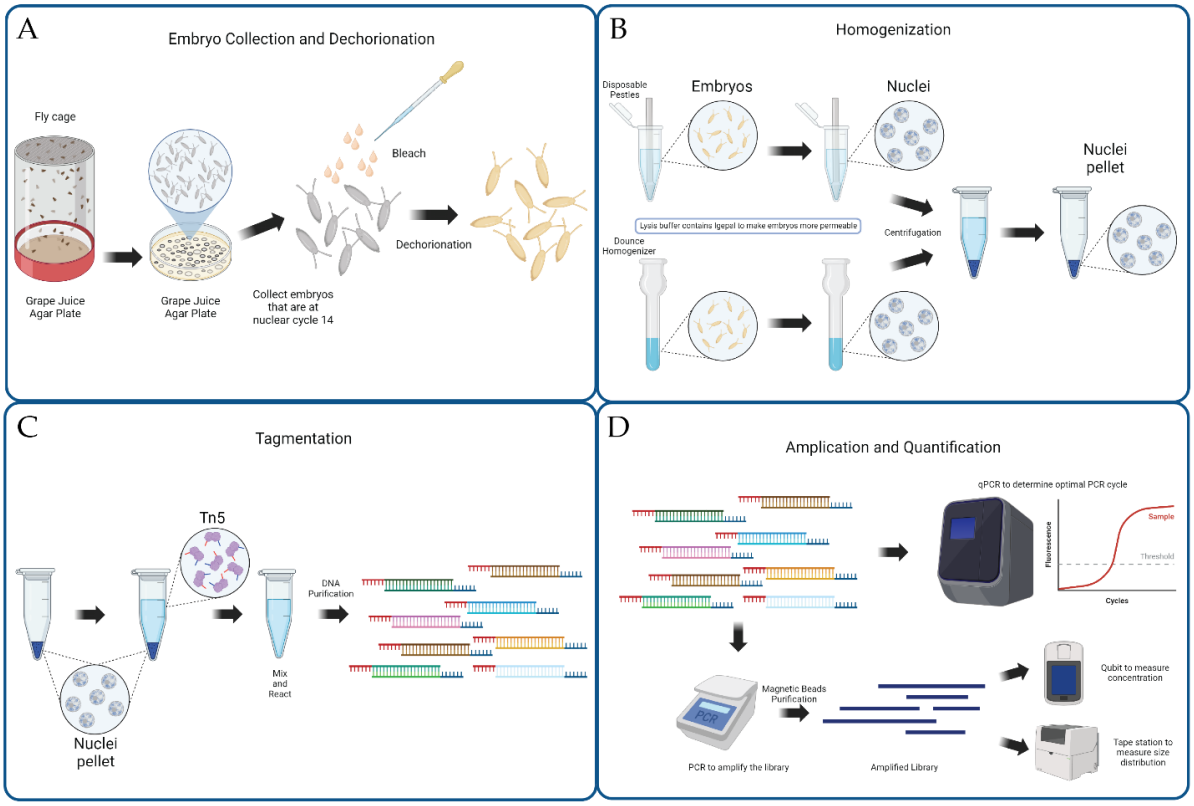

Figure 5.1 Overview of ATAC-seq library preparation. (A) Embryos are collected and treated with bleach to remove the chorion. (B) Dechorionated embryos are homogenized to release nuclei, followed by centrifugation to separate the cytoplasmic contents. (C) Hyperactive Tn5 transposase is introduced to fragment the genomic DNA. (D) The fragmented DNA is then PCR amplified and purified in preparation for sequencing. Created with BioRender.com.

Over-amplification can lead to disproportionate enrichment of certain sequences, reducing the overall quality and utility of the sequencing data by masking less abundant but potentially important genomic regions. The use of qPCR allows for precise calibration of the PCR process, ensuring that amplification is stopped at a point that preserves the diversity and relative abundance of the DNA fragments in the sample, the main PCR amplification was then performed accordingly. The PCR amplified DNA was then purified with magnetic beads. This purification step is essential to remove salts, primers, primer-dimers, and unincorporated dNTPs that could interfere with subsequent sequencing. The concentration of the ATAC-seq library was measured using a Qubit fluorometer to ensure there is sufficient DNA for sequencing. At the same time, the size distribution of the library was accessed using a Tapestation system. This analysis confirms that the DNA was neither under- nor over-fragmented. Finally, the libraries were sequenced using an Illumina sequencer, employing pair-end sequencing technology which reads both ends of each DNA fragment. This approach enhances the

accuracy and completeness of the sequencing data which is invaluable for accurately aligning the reads to the reference genome. Detailed experimental setup of each step will be discussed in the Material and Method section.

5.2.2 Check the developmental stage of embryo for each library

During the preparation of ATAC-seq libraries, the developmental stages of embryos were initially estimated using a stereomicroscope, with ten embryos pooled per genotype to prepare each library. To refine staging accuracy and confirm the developmental phase of each library, chromatin accessibility of specific developmental marker genes was analyzed. The genes *even-skipped* (*eve*) and *engrailed* (*en*), known for their distinct expression patterns during *Drosophila* embryogenesis, were utilized as markers. eve is typically expressed in a striped pattern during early segmentation, making it a reliable indicator for specific early to mid-embryonic stages, particularly within 20 to 30 minutes after the onset of NC14. engrailed expression, marking posterior compartment cells from mid-embryogenesis onwards, provides insights into slightly later stages.

ATAC-seq data from the *lacZ*, 1x5s, 1x601, and 12x601 genotypes, as illustrated in Figure 5.2, highlighting chromatin accessibility for these constructs. For comparative analysis, ATAC-seq libraries from embryos collected at 12 minutes (NC14+12) and 72 minutes (NC14+72) post-NC14, provided by Professor Shelby Blythe (Soluri et al., 2020), are also included. These additional datasets serve as critical references for accurately estimating the developmental stages of our ATAC-seq libraries. Notably, all four libraries were prepared approximately 40 minutes after the commencement of NC14, coinciding with the timing of our live cell imaging experiments with *Drosophila* embryos. While the *lacZ* sample may slightly precede the other three samples in developmental timing, and although all samples likely contain a mix of younger and older embryos, they are predominantly of younger stages on average. The slight variations observed are minimal. Examination of these data reveals consistent and satisfactory sample quality, aligning well with previous ATAC-seq datasets.

5.2.3 Quality check of the data over the whole genome and over the transgene

We conducted a deep sequencing analysis to ensure the quality of our samples. Each sample achieved an average of 32 million reads, indicating a robust depth that enhances the reliability of chromatin accessibility insights derived from the data. The fragment size distributions, which are crucial for assessing the efficiency of the transposition reaction and the overall quality of the library preparation, were analyzed and are displayed in

Figure 5.3A.

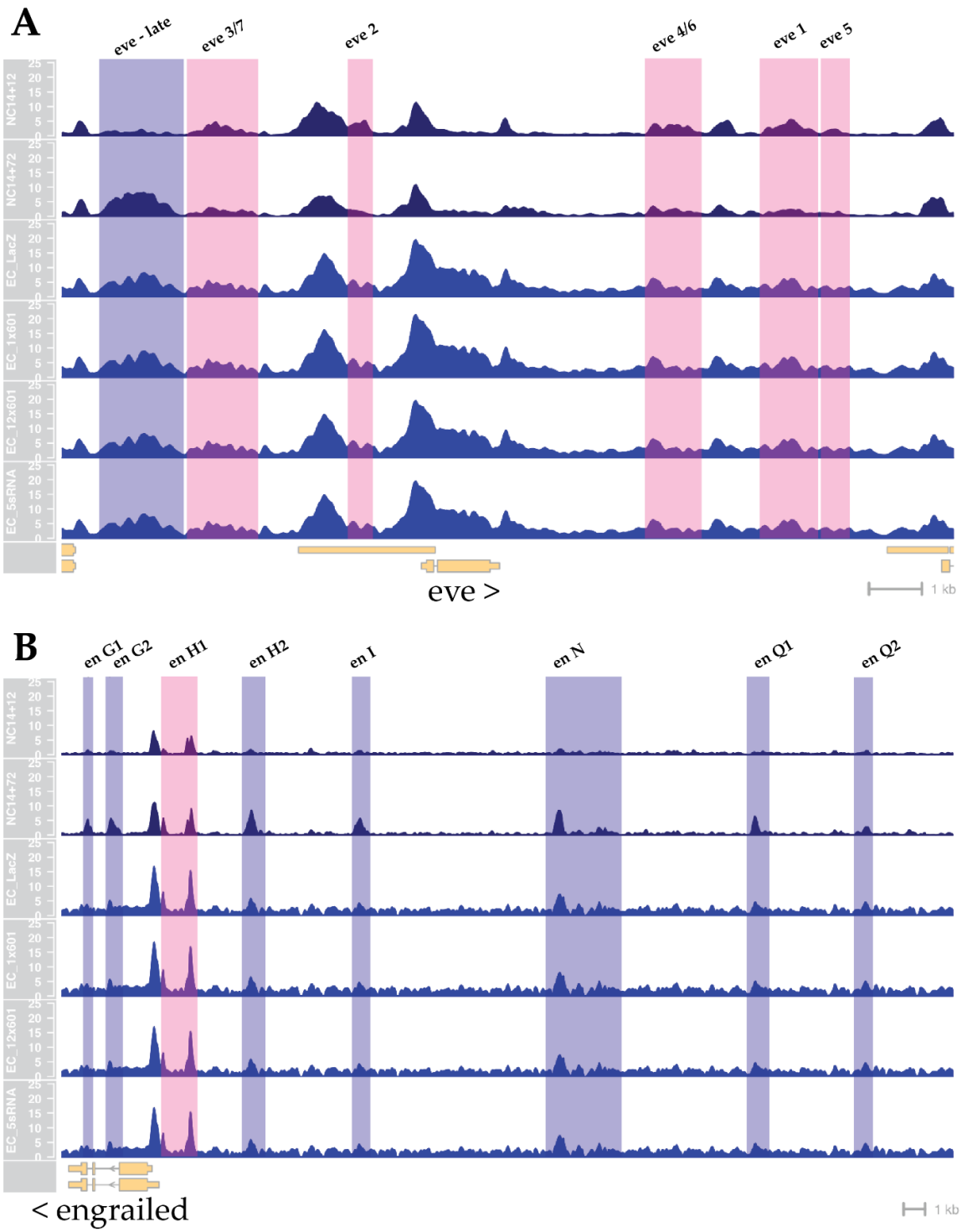

Figure 5.2 Assessing Developmental Stages Using Chromatin Accessibility of *eve* and *engrailed* Genes. (A-B) Chromatin accessibility at the representative pair-rule gene, eve, and the segment polarity gene, engrailed, across various developmental stages and

reporter constructs. The panels display early NC14 (NC14+12, first row), late NC14 (NC14+72, second row), *lacZ* (third row), 1x601 (fourth row), 12x601 (fifth row), and 5sRNA (sixth row). Enhancers that are significantly more accessible in early NC14 are highlighted in red, while those that become accessible later, around the onset of gastrulation, are in blue. ATAC accessibility is consistently shown on the same y-axis scale (0-25 counts per million, CPM) across all plots for direct comparison.

The read distribution pattern in the ATAC-seq data reveals a predominance of very short fragments, predominantly indicative of "open chromatin" areas, predominantly under 100 base pairs in length. This observation contrasts with the minimal presence of reads that align with the +1 nucleosome size of around 150 base pairs and an even smaller presence in the +2 nucleosome range around 300 base pairs. In contrast, during library preparation, each sample underwent analysis with a Tape station, where the distribution of DNA fragments appeared more uniform, as detailed in the Materials and Methods section. The observation of a higher prevalence of shorter DNA fragments in the ATAC-seq data from this experiment, which employed patterned flow cells in a NovaSeq system, suggests a potential technical factor influencing the results. Patterned flow cells are known to be more precise in cluster generation due to their structured layout, which can enhance sequencing efficiency and output. However, they may also exhibit a stronger size exclusion effect, leading to an enrichment of shorter fragments. This characteristic could particularly emphasize open chromatin reads, as these are typically shorter due to the nature of accessible DNA being more freely available for transposition.

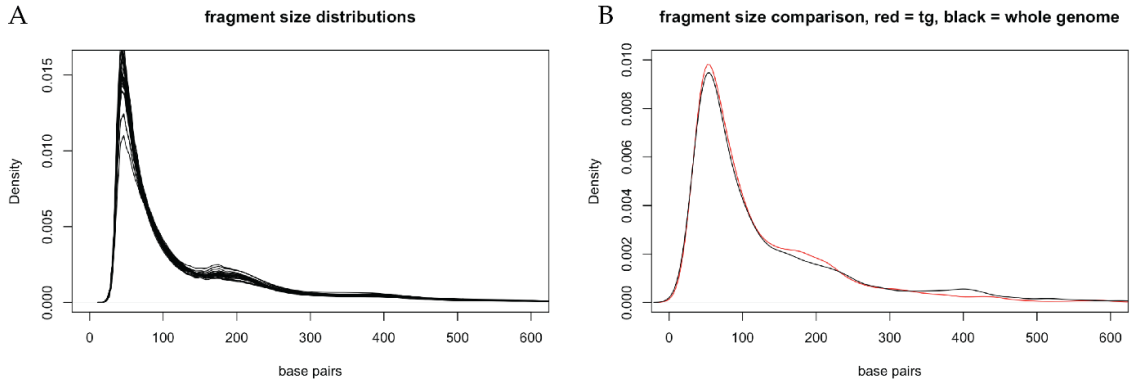

Figure 5.3 Fragment size distributions of reads from all libraries and reads mapped to the transgenes. (A) Displays the fragment size distributions for each individual library. (B) Shows the composite fragment size distribution of reads from all libraries, marked in black, alongside the fragment size distribution of reads specifically mapped to the transgenes, highlighted in red.

Mapping the ATAC-seq data to the individual transgenes by creating a bowtie index from the sequences of the reporter constructs has revealed important insights, as illustrated in Figure 5.3B. The analysis revealed that the size distribution of reads mapped to the transgenes closely aligns with that of the overall genomic data. This

similarity indicates that, despite a general bias towards shorter DNA fragments observed across the sequencing data, there is no discernible skew when comparing reads mapped to the transgenes versus those from the entire genome. Such uniformity in read distribution confirms that our experimental and analytical approaches are robust and reliable across various genomic contexts. This outcome is reassuring, verifying that data quality is not compromised by the unique features of the transgenes and that the observed bias towards shorter fragments is a characteristic of the sequencing method itself, not a targeted anomaly. The consistency in data quality across all four genotypes reinforces the dependability of our library preparation techniques and sets a solid foundation for further detailed analysis of nucleosome positioning across different reporters.

5.2.4 Single copy of NPS does not alter nucleosome positioning

To assess the nucleosome positioning within each transgene accurately, we use a strategy that quantifies ATAC-seq read coverage in distinct genomic regions characterized by their DNA fragment sizes. Specifically, we calculate the coverage of reads in the open chromatin regions, which are identified by fragment lengths between 50 and 100 base-pairs. These regions typically lack nucleosomes, suggesting higher accessibility and potential regulatory activity. Additionally, we evaluate the coverage in the mononucleosome regions, where fragment lengths range from 140 to 250 base-pairs. These lengths correspond to DNA wrapped around a single nucleosome, which can provide insights into standard nucleosome positioning along the DNA.

In figure 5.4A, we illustrate the coverage distribution of reads across open chromatin and nucleosomal regions for the *lacZ* construct, which spans 7962 base-pairs. In this visualization, open chromatin (open) coverage is marked in red, while coverage over nucleosomal (nuc) regions is depicted in black. Both of the coverages are counts per million (CPM) normalized. The P2 promoter fragment is distinctly highlighted with a thick gold rug, indicating its position within the transgenes. The specific sites where single NPS, 5sRNA or Widom 601 sequences, have been inserted in the 1x5sRNA or 1x601 constructs are marked with thick grey rugs and ticked lines for clear differentiation. The schematic is the same for the 1x5sRNA and 1x601 constructs. In the *lacZ* construct, both MS2 (around 2000 base-pairs) and PP7 (around 6000 base-pairs) stem loop regions are highly repetitive, leading to them being largely unmapped. Nevertheless, the reads mapping to the *lacZ* region were successfully aligned, and nucleosome positioning could be inferred where the black distribution, representing nucleosomal regions, visibly protrudes through the red distribution. For improved visualization of nucleosome positioning within the construct, figure 5.4B utilizes a more analytical approach by plotting the log ratio of nucleosomal to open chromatin reads. This method enhances visualization by providing a more quantitative comparison, effectively highlighting areas where nucleosomes are predominantly positioned compared to regions of open chromatin. By focusing on the log ratio, it becomes easier

to discern subtle variations in nucleosome density across the construct. Particularly within the P2 promoter region (gold rug), there are notable spots indicative of enriched nucleosome presence. While, in the *lacZ* region, there is also a reasonable degree of phasing of nucleosome sized reads relative to the open chromatin reads. The position where, in other transgenes, the single nucleosome has been added, does seem to be where under normal circumstances there would be a slight enrichment for nucleosome sized-reads.

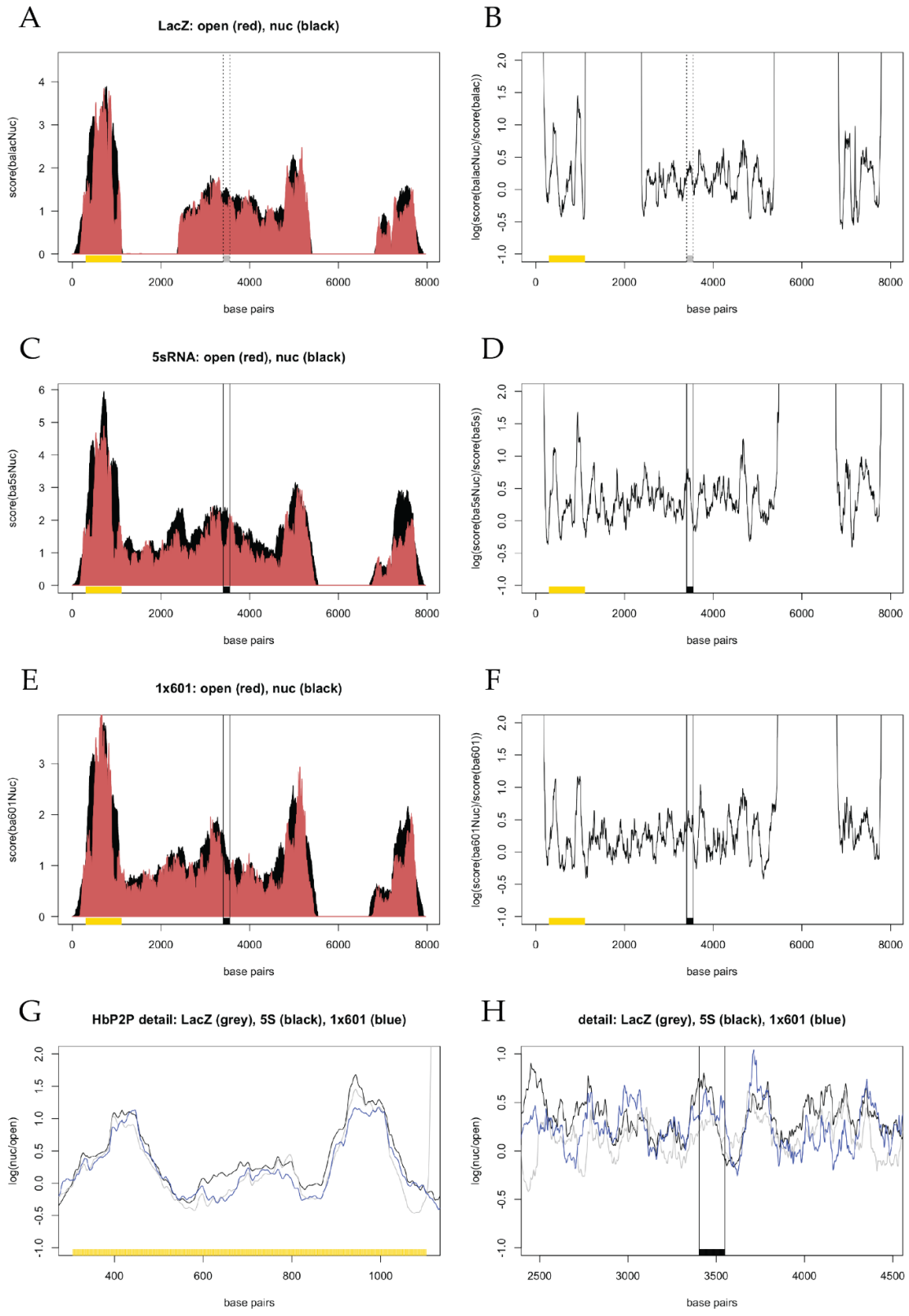

Figure caption in the following page.

Figure 5.4 Chromatin accessibility across the lacZ, 1x5sRNA, and 1x601 constructs. Gold rugs indicate the P2 promoter region consistent across all plots, while grey lines mark the locations where NPS was inserted. Black lines represent the positions of a 5sRNA or a Widom 601 NPS. (A) Distribution of reads for *lacZ* construct; black represents reads ranging from 140 to 250 base-pairs, corresponding to nucleosomal regions, and red represents reds under 100 base-pairs, corresponding to open chromatin area. Regions that could not be mapped due to repetitive sequences are left blank. (B) Log ratio of nucleosomal to open chromatin reads for *lacZ*. (C) Distribution of nucleosomal and open chromatin reads of 1x5sRNA construct. (D) Log ratio of nucleosomal to open chromatin reads for 1x5sRNA. (E) Distribution of nucleosomal and open chromatin reads of 1x601 construct. (F) Log ratio of nucleosomal to open chromatin reads for 1x601. (G, H) Log ratio within the P2 promoter and for region between MS2 and PP7 for *lacZ*, the 1x5sRNA, and the 1x601 construct. In these panels, the *lacZ* is highlighted in grey, 1x5sRNA in black, and 1x601 in blue.

In Figures 5.4C, 5.4D, 5.4E, and 5.4F, the individual coverage data and log ratios for the 1x5sRNA and 1x601 constructs are displayed in the same manner as with the *lacZ* construct, providing a cohesive analysis across different reporter constructs. The P2 promoter region, continues to be highlighted by a gold rug in these figures, ensuring consistent reference points across the different constructs. The specific locations where the 5sRNA or Widom 601 NPS have been integrated are marked with black rugs. Notably, these figures show a discernible enrichment for nucleosome-sized reads at these positions, suggesting that the insertion of these sequences does favor nucleosome positioning. Interestingly, similar enrichments are also observed in the *lacZ* construct, implying that these regions might naturally predispose towards nucleosome assembly, independent of the specific underline sequences. Furthermore, the use of non-repetitive versions of MS2 stem-loop repeats in the 1x5sRNA and 1x601 constructs has facilitated successful mapping of reads from these constructs to the genome. Such precise mapping is crucial for evaluating the nucleosome positioning near promoter regions.

In the analysis of the 1x5sRNA and 1x601 constructs, the observed periodicity of nucleosome enrichment aligning with the positions of sequence insertions provides intriguing insights into nucleosome positioning. As depicted in figure 5.4G, which offers a zoomed-in view of the spacer region for all three constructs, there is a noticeable alignment of nucleosome peaks with the inserted sequences. However, the significance of this alignment must be interpreted cautiously, as it could potentially occur by random chance. The intensity of nucleosome enrichment at these specific insertion sites is not markedly higher than that in neighboring regions, which suggests that while there is an observable pattern, it is not exceptionally pronounced. It's important to note that these positions do fall on the 'enriched' side of the plots for all three transgenes examined. When compared to the *LacZ* control, there is a visible, albeit not overwhelming, increase in nucleosome-sized fragment enrichment over these regions.

To validate the findings regarding the enrichment of nucleosome-sized reads in the 1x5sRNA, 1x601, and *lacZ* constructs, performing a comparative analysis over the P2

promoter region is a strategic approach. Since all three constructs share the same underlying sequence in this region, any consistent patterns observed can strengthen the argument that the nucleosome enrichment is not merely coincidental or due to random noise. In figure 5.4H, the log ratio of nucleosomal to open chromatin reads from all three constructs is overlaid with adjusted y-axis range to accommodate increased enrichment levels. The analysis successfully delineates the nucleosomal arrangement within the promoter region, illustrating the typical chromatin structure involved in gene regulation. The first peak at around 400 base-pairs is indicative of the -1 nucleosome. This nucleosome is positioned upstream of the transcription start site (TSS) and is often involved in regulating access to the core promoter elements by transcription machinery. The second peak at around 900 base-pairs represents the +1 nucleosome, located just downstream of the TSS. This positioning is strategic for controlling the transition from transcription initiation to elongation, effectively influencing the expression of the gene (Ramachandran et al., 2015; Wang et al., 2022). The valley between 600 and 800 base-pairs corresponds to the nucleosome depleted region (NDR) (Bai et al., 2011), a critical area typically free from nucleosomes that allows regulatory proteins and transcription factors to bind to the DNA more easily, facilitating the initiation of transcription (Hartley and Madhani, 2009; Jiang and Pugh, 2009; Kubik et al., 2015). This demonstrate that all three constructs exhibit a similar trend in this well-conserved region not only provides compelling evidence that the observed patterns in nucleosome positioning but also are likely reflective of real, biologically relevant phenomena rather than random variations. This consistency across different constructs, particularly in a shared sequence context like the P2 promoter, adds significant credibility to the conclusion that the observed nucleosome enrichments in the spacer regions are genuine.

5.5.1 Adding twelve Widom 601 NPS is still not able to alter the nucleosome positioning

Finally, we examined the 12x601 construct. This construct is created by integrating twelve Widom 601 NPS into the spacer region while maintaining the original linker sequence from the *LacZ* construct. The interesting periodicity observed in figure 5.5A highlights a critical issue in sequence alignment within the highly repetitive regions. The use of unique linker sequences helps to reduce this mappability issue to some extent by providing unique sequences that can assist in aligning reads to specific locations in the genome. However, as observed, this does not completely eliminate the problem, especially for reads that are closely matched to the Widom 601 sequence itself. If a read is exactly 146 base-pairs long and matches the Widom 601 sequence perfectly, it cannot be mapped to a specific instance of the sequence due to the repetitive nature. This also applies to reads shorter than 146 base-pairs that contain significant portions of the Widom 601 sequence. These sequences may overlap with multiple potential mapping

sites across the repeated Widom 601 sequences, leading to ambiguities in mapping.

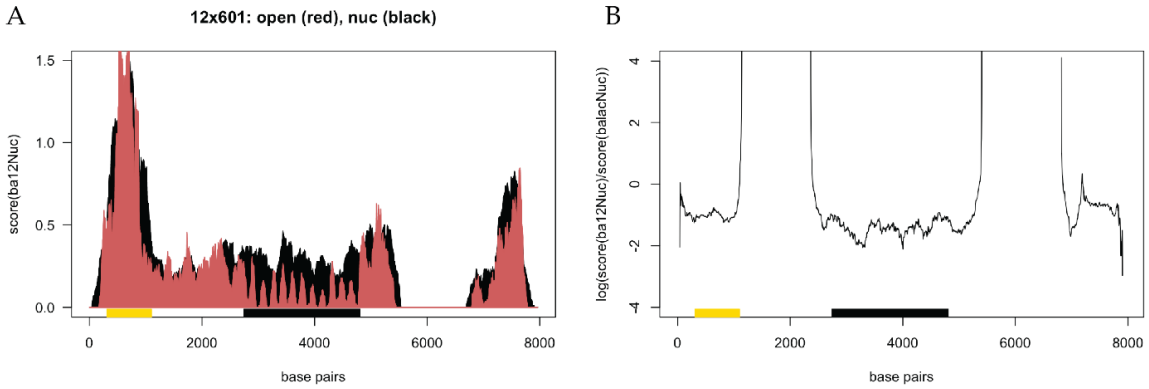

Figure 5.5 12x601 construct does not alter nucleosomal positioning. (A) Distribution of nucleosomal and open chromatin reads of 12x601 construct. (B) Log ratio of nucleosomal reads from 12x601 construct over that from *lacZ* construct.

To address the mappability issues in the 12x601 construct due to its repetitive nature, we employed a comparative approach by plotting the log ratio of nucleosome-sized reads from the *LacZ* construct against those from the 12x601 construct, as illustrated in Figure 5.5B. As before, the P2 promoter is marked by a gold rug and the spacer region is marked by a black rug. It is obvious that nucleosome-sized reads were not enriched in the 12x601 construct. In fact, it seems like there is a general under-representation of nucleosome-sized reads for this construct compared with the *LacZ* construct.

5.5.2 Transcription activity may alter the nucleosome positioning

Expanding the scope of ATAC-seq analysis to include genome-wide chromatin accessibility provides a comprehensive view of how transcription activities influence nucleosome distribution. In early NC14, the activity at the *hb* P2 promoter, used in all four reporter constructs, is particularly relevant because it is a phase of active transcription which can disrupt or reorganize nucleosomes. By comparing this with a gene where RNAP is predominantly paused at NC14, we can assess how transcriptional activity impacts nucleosome positioning across different genomic contexts. There are several examples of paused loci in early embryos. *Hedgehog* (*hh*) is one that is predominantly paused during NC14 before becoming selectively activated at gastrulation. The *hh* locus is large, so for a direct comparison with the *hb* locus, we will get an 8000 base-pairs region that includes 2000 base-pairs upstream of the *hh* TSS and 6000 base-pairs downstream.

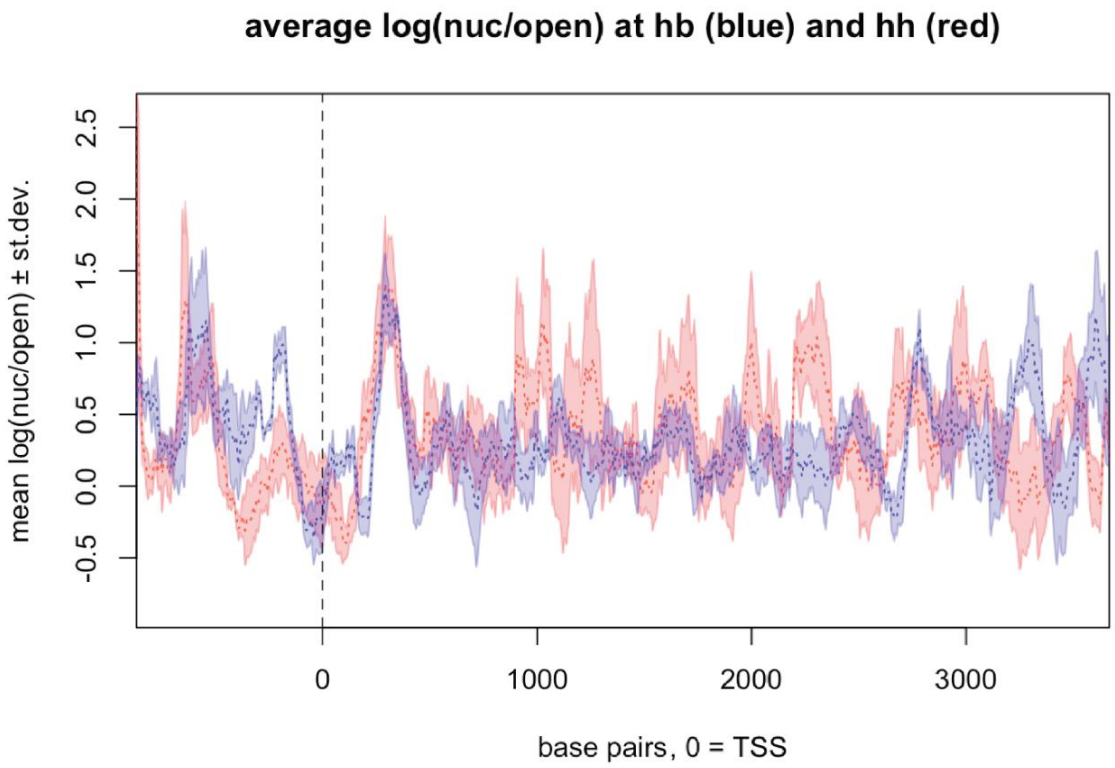

Figure 5.6 Enhanced nucleosome positioning is observed in silenced genes. Log ratio of nucleosomal to open chromatin reads for *hb* is highlighted in blue and that of *hh* is highlighted in red.

Applying the same analysis to the *hh* gene and plotting the log ratio of nucleosomal to open chromatin reads alongside the *hb* gene, provides a comparative view of nucleosome dynamics across these two genes in Figure 5.6. This analysis demonstrates how active transcription influences nucleosome positioning within these gene bodies. Although the *hh* gene does not show perfectly phased nucleosomes, it demonstrates clearer phasing compared to *hb*, suggesting that transcription activities impact nucleosome organization differently across genes. Both genes exhibit a consistent positioning of the +1 nucleosome just downstream of the transcription start site, highlighting its key role in transcription regulation. Additionally, there is a somewhat consistent placement of an upstream nucleosome around -500 bp from the TSS in both genes. However, the *hb* locus displays a more complex pattern with an additional nucleosome between this upstream position and the +1 nucleosome, indicating a potentially more intricate regulatory environment. Overall, these findings emphasize the interaction between transcription dynamics and chromatin architecture, providing insights into the regulatory mechanisms that govern gene expression.

5.3 Conclusion

In this chapter, we investigated the nucleosome occupancy of all four reporter constructs used in confocal microscopy experiment using ATAC-seq to provide a detailed view of chromatin dynamics during critical developmental stages. Our experiments revealed that the incorporation of single, 5sRNA or Widom 601 sequence, or multiple, twelve copies of Widom 601, NPS did not significantly alter nucleosome positioning, highlighting the robustness of the in vivo chromatin environment against structural reorganization. This resilience suggests that chromatin's inherent structure in a living organism is highly stable and resistant to artificial modifications aimed at altering nucleosome positioning.

Moreover, our comparative analysis of the *hb* and *hh* genes demonstrated that active transcription influences nucleosome dynamics, with the *hh* gene showing clearer nucleosome phasing than *hb*. This difference underscores the impact of transcriptional activity on chromatin structure, suggesting that gene-specific transcriptional mechanisms may dictate nucleosome organization.

The integration of ATAC-seq with confocal microscopy has provided comprehensive insights into the chromatin landscape, revealing the complex interplay between chromatin architecture and transcriptional regulation. These observations underscore the intricate nature of chromatin dynamics and pave the way for further research into the regulatory mechanisms that control gene expression during development. This work not only enriches our understanding of genomic regulation but also highlights the challenges and considerations necessary when attempting to manipulate chromatin architecture for experimental or therapeutic purposes.

5.4 Acknowledgements

We are deeply grateful to Prof. Shelby Blythe for his expert training on ATAC-seq library preparation, and for generously providing all the necessary reagents. We also extend our thanks for his invaluable contributions to the analysis of the data.

5.5 Material and Method

5.5.3 DNA constructs

Additional plasmids containing the reporter construct P2P-MS2-5sRNA-PP7 and P2P-MS2-1x601-PP7, were developed in-house by substituting a single copy of 5sRNA or the Widom 601 NPS into the lacZ region of the P2P-MS2-lacZ-PP7 plasmid used in

chapter 4. These plasmids were integrated into the *Drosophila* genome as detailed in section 4.4.1.

5.5.4 Library preparation

ATAC-seq was performed according to the procedures detailed in Soluri et al. (2020). Briefly, embryos at specific developmental stages, observed under a stereomicroscope, were selected and maintained at a constant room temperature to minimize staging variability. Ten embryos were dechorionated with bleach on a paper towel, washed, then floated off in PBS + 0.1% Tween-20. Sinking embryos were collected, transferred to a LoBind tube (Eppendorf, Hamburg, Germany), and homogenized in Lysis Buffer (Buenrostro et al., 2013) (10 mM Tris-HCl pH 7.4; 10 mM NaCl; 3 mM MgCl2; 0.1% Igepal CA-630) using the Monarch Pestle Set (NEB, Ipswich, Massachusetts). Nuclei were pelleted by gentle centrifugation at 4°C at 500 RCF for 10 minutes, and the supernatant was discarded before freezing the pellet on dry ice. Samples were briefly thawed on ice before undergoing tagmentation in a 50 µL reaction mixture containing 25 µL of 2x TD Buffer (Illumina, San Diego, California), 20 µL ultrapure water, and 5 µL TDE1 enzyme (Illumina, San Diego, California). The mixture was incubated for 30 minutes at 37°C with 800 RPM agitation in an Eppendorf Thermomixer HotShakey device. Post-tagmentation, samples were purified using the Qiagen MiniElute Reaction Clean Up kit, eluting in 10 µL EB Buffer (10 mM Tris-HCl, pH 8.5).

ATAC libraries were then amplified using a set of Dual-Unique Index PCR Primers (Soluri et al., 2020) and purified with Ampure XP beads (Beckman Coulter, Brea, California), with the final product eluted in 15 µL EB Buffer. The quality and fragment size distribution of the amplified libraries were verified on an Agilent Bioanalyzer (Agilent Biotechnologies, Santa Clara, California) to ensure the presence of distinct open and nucleosome-associated fragment sizes, and concentrations were quantified using the Qubit dsDNA HS Assay Kit (ThermoFisher, Waltham, Massachusetts). To minimize preparation bias, three replicates of each reporter construct were prepared on different days.

5.5.5 ATAC-seq analysis

The analysis was carried out in accordance with the methods described in Soluri et al. (2020). Demultiplexed reads were trimmed off adapters using TrimGalore! and subsequently mapped to the dm6 assembly of the *Drosophila* genome using Bowtie2 with the option -X 2000. Suspected optical and PCR duplicates were identified and marked using Picard MarkDuplicates. Mapped, trimmed, and duplicate-marked reads were then processed in R, utilizing the GenomicAlignments (Lawrence et al., 2013) and Rsamtools libraries (http://bioconductor.org/packages/release/bioc/html/Rsamtools.html). Only properly paired, non-secondary, and mapped reads with mapping quality scores of 10

or higher were retained. Reads with mapped lengths of 100 base-pairs or less were categorized as originating from 'open' chromatin, while those between 140 and 250 base-pairs were considered to have come from nucleosome-protected regions. Coverage of reads on different reporters was visualized by binning the data in 5 base-pair intervals.

Chapter 6

Conclusions

In this dissertation, we conducted a comprehensive investigation into transcription over nucleosomes, combining *in vitro* and *in vivo* methods to explore the behavior of transcriptional machinery as it interacts with chromatin.

We illuminated the dynamic interactions between RNAP and nucleosomes, shedding light on transcriptional mechanics in various contexts. Initial studies using purified components established the nucleosome as a significant barrier in vitro, a stark contrast to the more efficient RNAP navigation observed in vivo. This discrepancy was explored through cryo-ET, which captured intermediate states of RNAP transcribing through nucleosomes, providing a structural basis for understanding how RNAP contends with this chromatin barrier.

Further *in vivo* investigations using *Drosophila* embryos demonstrated that transcription elongation rates remain consistent regardless of the presence of NPS that hinder RNAP *in vitro*. This suggests additional cellular factors *in vivo* that facilitate RNAP's passage, a phenomenon further explored through ATAC-seq analysis, which indicated that underlying DNA sequences play a less critical role in nucleosome positioning within a cellular context than previously thought.

Overall, these findings challenge existing *in vitro* models of transcription, presenting a more nuanced view of the nucleosome as a barrier and emphasizing the complexity of transcription regulation in a cellular environment. Future research may benefit from integrating these diverse approaches to further dissect the interplay between transcription machinery and chromatin, potentially leading to enhanced models that more accurately reflect the *in vivo* state.

Bibliography

[1] Ahn, S. H., M. Kim, and S. Buratowski. "Phosphorylation of Serine 2 within the Rna Polymerase Ii C-Terminal Domain Couples Transcription and 3' End Processing." Mol Cell 13, no. 1 (Jan 16 2004): 67-76.

[2] Akai, Y., N. Adachi, Y. Hayashi, M. Eitoku, N. Sano, R. Natsume, N. Kudo, et al. "Structure of the Histone Chaperone Cia/Asf1-Double Bromodomain Complex Linking Histone Modifications and Site-Specific Histone Eviction." Proc Natl Acad Sci U S A 107, no. 18 (May 4 2010): 8153-8.

[3] Akolpoglu, M. B., Y. Alapan, N. O. Dogan, S. F. Baltaci, O. Yasa, G. Aybar Tural, and M. Sitti. "Magnetically Steerable Bacterial Microrobots Moving in 3d Biological Matrices for Stimuli-Responsive Cargo Delivery." Sci Adv 8, no. 28 (Jul 15 2022): eabo6163.

[4] Alam, K. K., K. D. Tawiah, M. F. Lichte, D. Porciani, and D. H. Burke. "A Fluorescent Split Aptamer for Visualizing Rna-Rna Assembly in Vivo." ACS Synth Biol 6, no. 9 (Sep 15 2017): 1710-21.

[5] Ardehali, M. B., J. Yao, K. Adelman, N. J. Fuda, S. J. Petesch, W. W. Webb, and J. T. Lis. "Spt6 Enhances the Elongation Rate of Rna Polymerase Ii in Vivo." EMBO J 28, no. 8 (Apr 22 2009): 1067-77.

[6] Arganda-Carreras, I., V. Kaynig, C. Rueden, K. W. Eliceiri, J. Schindelin, A. Cardona, and H. Sebastian Seung. "Trainable Weka Segmentation: A Machine Learning Tool for Microscopy Pixel Classification." Bioinformatics 33, no. 15 (Aug 1 2017): 2424-26.

[7] Artsimovitch, I., and T. M. Henkin. "In Vitro Approaches to Analysis of Transcription Termination." Methods 47, no. 1 (Jan 2009): 37-43.

[8] Awrey, D. E., N. Shimasaki, C. Koth, R. Weilbaecher, V. Olmsted, S. Kazanis, X. Shan, et al. "Yeast Transcript Elongation Factor (Tfiis), Structure and Function. Ii: Rna Polymerase Binding, Transcript Cleavage, and Read-Through." J Biol Chem 273, no. 35 (Aug 28 1998): 22595-605.

[9] Baek, S., and I. Lee. "Single-Cell Atac Sequencing Analysis: From Data Preprocessing to Hypothesis Generation." Comput Struct Biotechnol J 18 (2020): 1429-39.

[10] Bai, L., G. Charvin, E. D. Siggia, and F. R. Cross. "Nucleosome-Depleted Regions in Cell-Cycle-Regulated Promoters Ensure Reliable Gene Expression in Every Cell Cycle." Dev Cell 18, no. 4 (Apr 20 2010): 544-55.
[11] Barski, A., S. Cuddapah, K. Cui, T. Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev, and K. Zhao. "High-Resolution Profiling of Histone Methylations in the Human Genome." Cell 129, no. 4 (May 18 2007): 823-37.
[12] Becker, P. B., and J. L. Workman. "Nucleosome Remodeling and Epigenetics." Cold Spring Harb Perspect Biol 5, no. 9 (Sep 1 2013).
[13] Bednar, J., V. M. Studitsky, S. A. Grigoryev, G. Felsenfeld, and C. L. Woodcock. "The Nature of the Nucleosomal Barrier to Transcription: Direct Observation of Paused Intermediates by Electron Cryomicroscopy." Mol Cell 4, no. 3 (Sep 1999): 377-86.
[14] Behjati, S., and P. S. Tarpey. "What Is Next Generation Sequencing?". Arch Dis Child Educ Pract Ed 98, no. 6 (Dec 2013): 236-8.
[15] Bertrand, E., P. Chartrand, M. Schaefer, S. M. Shenoy, R. H. Singer, and R. M. Long. "Localization of Ash1 Mrna Particles in Living Yeast." Mol Cell 2, no. 4 (Oct 1998): 437-45.
[16] Bilokapic, S., M. Strauss, and M. Halic. "Histone Octamer Rearranges to Adapt to DNA Unwrapping." Nat Struct Mol Biol 25, no. 1 (Jan 2018): 101-08.
[17] Bondarenko, V. A., L. M. Steele, A. Ujvari, D. A. Gaykalova, O. I. Kulaeva, Y. S. Polikanov, D. S. Luse, and V. M. Studitsky. "Nucleosomes Can Form a Polar Barrier to Transcript Elongation by Rna Polymerase Ii." Mol Cell 24, no. 3 (Nov 3 2006): 469-79.
[18] Bowman, E. A., and W. G. Kelly. "Rna Polymerase Ii Transcription Elongation and Pol Ii Ctd Ser2 Phosphorylation: A Tail of Two Kinases." Nucleus 5, no. 3 (May-Jun 2014): 224-36.
[19] Bowman, G. D., and M. G. Poirier. "Post-Translational Modifications of Histones That Influence Nucleosome Dynamics." Chem Rev 115, no. 6 (Mar 25 2015): 2274-95.
[20] Boyle, A. P., S. Davis, H. P. Shulha, P. Meltzer, E. H. Margulies, Z. Weng, T. S. Furey, and G. E. Crawford. "High-Resolution Mapping and Characterization of Open Chromatin across the Genome." Cell 132, no. 2 (Jan 25 2008): 311-22.
[21] Branco, M. R., and A. Pombo. "Intermingling of Chromosome Territories in

Interphase Suggests Role in Translocations and Transcription-Dependent Associations." PLoS Biol 4, no. 5 (May 2006): e138.
[22] Brannan, K., H. Kim, B. Erickson, K. Glover-Cutter, S. Kim, N. Fong, L. Kiemele, et al. "Mrna Decapping Factors and the Exonuclease Xrn2 Function in Widespread Premature Termination of Rna Polymerase Ii Transcription." Mol Cell 46, no. 3 (May 11 2012): 311-24.
[23] Braselmann, E., C. Rathbun, E. M. Richards, and A. E. Palmer. "Illuminating Rna Biology: Tools for Imaging Rna in Live Mammalian Cells." Cell Chem Biol 27, no. 8 (Aug 20 2020): 891-903.
[24] Briley, W. E., M. H. Bondy, P. S. Randeria, T. J. Dupper, and C. A. Mirkin. "Quantification and Real-Time Tracking of Rna in Live Cells Using Sticky-Flares." Proc Natl Acad Sci U S A 112, no. 31 (Aug 4 2015): 9591-5.
[25] Buenrostro, J. D., P. G. Giresi, L. C. Zaba, H. Y. Chang, and W. J. Greenleaf. "Transposition of Native Chromatin for Fast and Sensitive Epigenomic Profiling of Open Chromatin, DNA-Binding Proteins and Nucleosome Position." Nat Methods 10, no. 12 (Dec 2013): 1213-8.
[26] Buenrostro, J. D., B. Wu, H. Y. Chang, and W. J. Greenleaf. "Atac-Seq: A Method for Assaying Chromatin Accessibility Genome-Wide." Curr Protoc Mol Biol 109 (Jan 5 2015): 21 29 1-21 29 9.
[27] Buenrostro, J. D., B. Wu, U. M. Litzenburger, D. Ruff, M. L. Gonzales, M. P. Snyder, H. Y. Chang, and W. J. Greenleaf. "Single-Cell Chromatin Accessibility Reveals Principles of Regulatory Variation." Nature 523, no. 7561 (Jul 23 2015): 486-90.
[28] Buratowski, S. "Connections between Mrna 3' End Processing and Transcription Termination." Curr Opin Cell Biol 17, no. 3 (Jun 2005): 257-61.
[29] Buratowski, S., S. Hahn, L. Guarente, and P. A. Sharp. "Five Intermediate Complexes in Transcription Initiation by Rna Polymerase Ii." Cell 56, no. 4 (Feb 24 1989): 549-61.
[30] Cao, J., and Q. Yan. "Histone Ubiquitination and Deubiquitination in Transcription, DNA Damage Response, and Cancer." Front Oncol 2 (2012): 26.
[31] Cardiello, J. F., J. A. Goodrich, and J. F. Kugel. "Heat Shock Causes a Reversible Increase in Rna Polymerase Ii Occupancy Downstream of Mrna Genes, Consistent with a Global Loss in Transcriptional Termination." Mol Cell Biol 38, no.

18 (Sep 15 2018).
[32] Chalamcharla, V. R., H. D. Folco, J. Dhakshnamoorthy, and S. I. Grewal. "Conserved Factor Dhp1/Rat1/Xrn2 Triggers Premature Transcription Termination and Nucleates Heterochromatin to Promote Gene Silencing." Proc Natl Acad Sci U S A 112, no. 51 (Dec 22 2015): 15548-55.
[33] Chandrasekharan, M. B., F. Huang, and Z. W. Sun. "Ubiquitination of Histone H2b Regulates Chromatin Dynamics by Enhancing Nucleosome Stability." Proc Natl Acad Sci U S A 106, no. 39 (Sep 29 2009): 16686-91.
[34] Chao, J. A., Y. Patskovsky, S. C. Almo, and R. H. Singer. "Structural Basis for the Coevolution of a Viral Rna-Protein Complex." Nat Struct Mol Biol 15, no. 1 (Jan 2008): 103-5.
[35] Chen, P., W. Li, and G. Li. "Structures and Functions of Chromatin Fibers." Annu Rev Biophys 50 (May 6 2021): 95-116.
[36] Chen, Z., R. Gabizon, A. I. Brown, A. Lee, A. Song, C. Diaz-Celis, C. D. Kaplan, et al. "High-Resolution and High-Accuracy Topographic and Transcriptional Maps of the Nucleosome Barrier." Elife 8 (Jul 31 2019).
[37] Chubb, J. R., T. Trcek, S. M. Shenoy, and R. H. Singer. "Transcriptional Pulsing of a Developmental Gene." Curr Biol 16, no. 10 (May 23 2006): 1018-25.
[38] Connelly, S., and J. L. Manley. "A Functional Mrna Polyadenylation Signal Is Required for Transcription Termination by Rna Polymerase Ii." Genes Dev 2, no. 4 (Apr 1988): 440-52.
[39] Coons, A. H. "The Beginnings of Immunofluorescence." J Immunol 87 (Nov 1961): 499-503.
[40] Cramer, P., D. A. Bushnell, J. Fu, A. L. Gnatt, B. Maier-Davis, N. E. Thompson, R. R. Burgess, et al. "Architecture of Rna Polymerase Ii and Implications for the Transcription Mechanism." Science 288, no. 5466 (Apr 28 2000): 640-9.
[41] Cramer, P., D. A. Bushnell, and R. D. Kornberg. "Structural Basis of Transcription: Rna Polymerase Ii at 2.8 Angstrom Resolution." Science 292, no. 5523 (Jun 8 2001): 1863-76.
[42] Cui, K., and K. Zhao. "Genome-Wide Approaches to Determining Nucleosome Occupancy in Metazoans Using Mnase-Seq." Methods Mol Biol 833 (2012): 413-9.
[43] Cusanovich, D. A., A. J. Hill, D. Aghamirzaie, R. M. Daza, H. A. Pliner, J. B.

Berletch, G. N. Filippova, et al. "A Single-Cell Atlas of in Vivo Mammalian Chromatin Accessibility." Cell 174, no. 5 (Aug 23 2018): 1309-24 e18.
[44] Darzacq, X., Y. Shav-Tal, V. de Turris, Y. Brody, S. M. Shenoy, R. D. Phair, and R. H. Singer. "In Vivo Dynamics of Rna Polymerase Ii Transcription." Nat Struct Mol Biol 14, no. 9 (Sep 2007): 796-806.
[45] Di Cerbo, V., F. Mohn, D. P. Ryan, E. Montellier, S. Kacem, P. Tropberger, E. Kallis, et al. "Acetylation of Histone H3 at Lysine 64 Regulates Nucleosome Dynamics and Facilitates Transcription." Elife 3 (Mar 25 2014): e01632.
[46] Doma, M. K., and R. Parker. "Endonucleolytic Cleavage of Eukaryotic Mrnas with Stalls in Translation Elongation." Nature 440, no. 7083 (Mar 23 2006): 561-4.
[47] Dvir, A., R. C. Conaway, and J. W. Conaway. "A Role for Tfiih in Controlling the Activity of Early Rna Polymerase Ii Elongation Complexes." Proc Natl Acad Sci U S A 94, no. 17 (Aug 19 1997): 9006-10.
[48] Dvir, A., K. P. Garrett, C. Chalut, J. M. Egly, J. W. Conaway, and R. C. Conaway. "A Role for Atp and Tfiih in Activation of the Rna Polymerase Ii Preinitiation Complex Prior to Transcription Initiation." J Biol Chem 271, no. 13 (Mar 29 1996): 7245-8.
[49] Dyba, M., S. Jakobs, and S. W. Hell. "Immunofluorescence Stimulated Emission Depletion Microscopy." Nat Biotechnol 21, no. 11 (Nov 2003): 1303-4.
[50] Dye, M. J., and N. J. Proudfoot. "Multiple Transcript Cleavage Precedes Polymerase Release in Termination by Rna Polymerase Ii." Cell 105, no. 5 (Jun 1 2001): 669-81.
[51] Eaton, J. D., L. Davidson, D. L. V. Bauer, T. Natsume, M. T. Kanemaki, and S. West. "Xrn2 Accelerates Termination by Rna Polymerase Ii, Which Is Underpinned by Cpsf73 Activity." Genes Dev 32, no. 2 (Jan 15 2018): 127-39.
[52] Eberharter, A., and P. B. Becker. "Histone Acetylation: A Switch between Repressive and Permissive Chromatin. Second in Review Series on Chromatin Dynamics." EMBO Rep 3, no. 3 (Mar 2002): 224-9.
[53] Ehara, H., T. Kujirai, M. Shirouzu, H. Kurumizaka, and S. I. Sekine. "Structural Basis of Nucleosome Disassembly and Reassembly by Rnapii Elongation Complex with Fact." Science 377, no. 6611 (Sep 9 2022): eabp9466.
[54] Endoh, M., W. Zhu, J. Hasegawa, H. Watanabe, D. K. Kim, M. Aida, N. Inukai, et al. "Human Spt6 Stimulates Transcription Elongation by Rna Polymerase

Ii in Vitro." Mol Cell Biol 24, no. 8 (Apr 2004): 3324-36.
[55] Engel, C., S. Neyer, and P. Cramer. "Distinct Mechanisms of Transcription Initiation by Rna Polymerases I and Ii." Annu Rev Biophys 47 (May 20 2018): 425-46.
[56] Fong, N., H. Kim, Y. Zhou, X. Ji, J. Qiu, T. Saldi, K. Diener, et al. "Pre-Mrna Splicing Is Facilitated by an Optimal Rna Polymerase Ii Elongation Rate." Genes Dev 28, no. 23 (Dec 1 2014): 2663-76.
[57] Foster, E. R., and J. A. Downs. "Histone H2a Phosphorylation in DNA Double-Strand Break Repair." FEBS J 272, no. 13 (Jul 2005): 3231-40.
[58] Fuchs, G., Y. Voichek, S. Benjamin, S. Gilad, I. Amit, and M. Oren. "4sudrb-Seq: Measuring Genomewide Transcriptional Elongation Rates and Initiation Frequencies within Cells." Genome Biol 15, no. 5 (May 9 2014): R69.
[59] Fukaya, T., B. Lim, and M. Levine. "Rapid Rates of Pol Ii Elongation in the Drosophila Embryo." Curr Biol 27, no. 9 (May 8 2017): 1387-91.
[60] Garcia, H. G., M. Tikhonov, A. Lin, and T. Gregor. "Quantitative Imaging of Transcription in Living Drosophila Embryos Links Polymerase Activity to Patterning." Curr Biol 23, no. 21 (Nov 4 2013): 2140-5.
[61] Garcia, Hernan G, and Thomas Gregor. "Live Imaging of Mrna Synthesis in Drosophila." RNA Detection: Methods and Protocols (2018): 349-57.
[62] Gauto, D. F., L. F. Estrozi, C. D. Schwieters, G. Effantin, P. Macek, R. Sounier, A. C. Sivertsen, et al. "Integrated Nmr and Cryo-Em Atomic-Resolution Structure Determination of a Half-Megadalton Enzyme Complex." Nat Commun 10, no. 1 (Jun 19 2019): 2697.
[63] Gaykalova, D. A., O. I. Kulaeva, O. Volokh, A. K. Shaytan, F. K. Hsieh, M. P. Kirpichnikov, O. S. Sokolova, and V. M. Studitsky. "Structural Analysis of Nucleosomal Barrier to Transcription." Proc Natl Acad Sci U S A 112, no. 43 (Oct 27 2015): E5787-95.
[64] Giresi, P. G., J. Kim, R. M. McDaniell, V. R. Iyer, and J. D. Lieb. "Faire (Formaldehyde-Assisted Isolation of Regulatory Elements) Isolates Active Regulatory Elements from Human Chromatin." Genome Res 17, no. 6 (Jun 2007): 877-85.
[65] Golding, I., J. Paulsson, S. M. Zawilski, and E. C. Cox. "Real-Time Kinetics of Gene Activity in Individual Bacteria." Cell 123, no. 6 (Dec 16 2005): 1025-36.

[66] Gonzalez-Jimenez, A., A. Campos, F. Navarro, A. Clemente-Blanco, and O. Calvo. "Regulation of Eukaryotic Rnaps Activities by Phosphorylation." Front Mol Biosci 8 (2021): 681865.
[67] Goodwin, S., J. D. McPherson, and W. R. McCombie. "Coming of Age: Ten Years of Next-Generation Sequencing Technologies." Nat Rev Genet 17, no. 6 (May 17 2016): 333-51.
[68] Grandi, F. C., H. Modi, L. Kampman, and M. R. Corces. "Chromatin Accessibility Profiling by Atac-Seq." Nat Protoc 17, no. 6 (Jun 2022): 1518-52.
[69] Greer, E. L., and Y. Shi. "Histone Methylation: A Dynamic Mark in Health, Disease and Inheritance." Nat Rev Genet 13, no. 5 (Apr 3 2012): 343-57.
[70] Greger, I. H., A. Aranda, and N. Proudfoot. "Balancing Transcriptional Interference and Initiation on the Gal7 Promoter of Saccharomyces Cerevisiae." Proc Natl Acad Sci U S A 97, no. 15 (Jul 18 2000): 8415-20.
[71] Greiss, F., M. Deligiannaki, C. Jung, U. Gaul, and D. Braun. "Single-Molecule Imaging in Living Drosophila Embryos with Reflected Light-Sheet Microscopy." Biophys J 110, no. 4 (Feb 23 2016): 939-46.
[72] Gui, L., W. J. O'Shaughnessy, K. Cai, E. Reetz, M. L. Reese, and D. Nicastro. "Cryo-Tomography Reveals Rigid-Body Motion and Organization of Apicomplexan Invasion Machinery." Nat Commun 14, no. 1 (Mar 30 2023): 1775.
[73] Hall, M. A., A. Shundrovsky, L. Bai, R. M. Fulbright, J. T. Lis, and M. D. Wang. "High-Resolution Dynamic Mapping of Histone-DNA Interactions in a Nucleosome." Nat Struct Mol Biol 16, no. 2 (Feb 2009): 124-9.
[74] Han, Y., S. Wang, Z. Zhang, X. Ma, W. Li, X. Zhang, J. Deng, et al. "In Vivo Imaging of Protein-Protein and Rna-Protein Interactions Using Novel Far-Red Fluorescence Complementation Systems." Nucleic Acids Res 42, no. 13 (Jul 2014): e103.
[75] Hartley, P. D., and H. D. Madhani. "Mechanisms That Specify Promoter Nucleosome Location and Identity." Cell 137, no. 3 (May 1 2009): 445-58.
[76] Hattne, J., D. Shi, C. Glynn, C. T. Zee, M. Gallagher-Jones, M. W. Martynowycz, J. A. Rodriguez, and T. Gonen. "Analysis of Global and Site-Specific Radiation Damage in Cryo-Em." Structure 26, no. 5 (May 1 2018): 759-66 e4.
[77] Ho, C. K., and S. Shuman. "Distinct Roles for Ctd Ser-2 and Ser-5 Phosphorylation in the Recruitment and Allosteric Activation of Mammalian Mrna

Capping Enzyme." Mol Cell 3, no. 3 (Mar 1999): 405-11.
[78] Holstege, F. C., P. C. van der Vliet, and H. T. Timmers. "Opening of an Rna Polymerase Ii Promoter Occurs in Two Distinct Steps and Requires the Basal Transcription Factors Iie and Iih." EMBO J 15, no. 7 (Apr 1 1996): 1666-77.
[79] Hong, Jong Chan. "General Aspects of Plant Transcription Factor Families." In Plant Transcription Factors, 35-56: Elsevier, 2016.
[80] Hoppe, C., and H. L. Ashe. "Live Imaging and Quantitation of Nascent Transcription Using the Ms2/Mcp System in the Drosophila Embryo." STAR Protoc 2, no. 1 (Mar 19 2021): 100379.
[81] Hsieh, F. K., O. I. Kulaeva, S. S. Patel, P. N. Dyer, K. Luger, D. Reinberg, and V. M. Studitsky. "Histone Chaperone Fact Action During Transcription through Chromatin by Rna Polymerase Ii." Proc Natl Acad Sci U S A 110, no. 19 (May 7 2013): 7654-9.
[82] Hsin, J. P., and J. L. Manley. "The Rna Polymerase Ii Ctd Coordinates Transcription and Rna Processing." Genes Dev 26, no. 19 (Oct 1 2012): 2119-37.
[83] Hu, Y., J. Xu, E. Gao, X. Fan, J. Wei, B. Ye, S. Xu, and W. Ma. "Enhanced Single Rna Imaging Reveals Dynamic Gene Expression in Live Animals." Elife 12 (Mar 3 2023).
[84] Huranova, M., J. A. Jablonski, A. Benda, M. Hof, D. Stanek, and M. Caputi. "In Vivo Detection of Rna-Binding Protein Interactions with Cognate Rna Sequences by Fluorescence Resonance Energy Transfer." RNA 15, no. 11 (Nov 2009): 2063-71.
[85] Hussmann, J. A., J. Ling, P. Ravisankar, J. Yan, A. Cirincione, A. Xu, D. Simpson, et al. "Mapping the Genetic Landscape of DNA Double-Strand Break Repair." Cell 184, no. 22 (Oct 28 2021): 5653-69 e25.
[86] Hyun, K., J. Jeon, K. Park, and J. Kim. "Writing, Erasing and Reading Histone Lysine Methylations." Exp Mol Med 49, no. 4 (Apr 28 2017): e324.
[87] Imbalzano, A. N., K. S. Zaret, and R. E. Kingston. "Transcription Factor (Tf) Iib and Tfiia Can Independently Increase the Affinity of the Tata-Binding Protein for DNA." J Biol Chem 269, no. 11 (Mar 18 1994): 8280-6.
[88] Ishibashi, T., M. Dangkulwanich, Y. Coello, T. A. Lionberger, L. Lubkowska, A. S. Ponticelli, M. Kashlev, and C. Bustamante. "Transcription Factors Iis and Iif Enhance Transcription Efficiency by Differentially Modifying Rna Polymerase

Pausing Dynamics." Proc Natl Acad Sci U S A 111, no. 9 (Mar 4 2014): 3419-24.
[89] Izban, M. G., and D. S. Luse. "Factor-Stimulated Rna Polymerase Ii Transcribes at Physiological Elongation Rates on Naked DNA but Very Poorly on Chromatin Templates." J Biol Chem 267, no. 19 (Jul 5 1992): 13647-55.
[90] Jennings, Barbara H. "Drosophila–a Versatile Model in Biology & Medicine." Materials today 14, no. 5 (2011): 190-95.
[91] Jeon, C., H. Yoon, and K. Agarwal. "The Transcription Factor Tfiis Zinc Ribbon Dipeptide Asp-Glu Is Critical for Stimulation of Elongation and Rna Cleavage by Rna Polymerase Ii." Proc Natl Acad Sci U S A 91, no. 19 (Sep 13 1994): 9106-10.
[92] Jeronimo, C., A. Angel, V. Q. Nguyen, J. M. Kim, C. Poitras, E. Lambert, P. Collin, et al. "Fact Is Recruited to the +1 Nucleosome of Transcribed Genes and Spreads in a Chd1-Dependent Manner." Mol Cell 81, no. 17 (Sep 2 2021): 3542-59 e11.
[93] Jiang, C., and B. F. Pugh. "Nucleosome Positioning and Gene Regulation: Advances through Genomics." Nat Rev Genet 10, no. 3 (Mar 2009): 161-72.
[94] Jiang, Y., M. Yan, and J. D. Gralla. "A Three-Step Pathway of Transcription Initiation Leading to Promoter Clearance at an Activation Rna Polymerase Ii Promoter." Mol Cell Biol 16, no. 4 (Apr 1996): 1614-21.
[95] Johansson, Hans E, Lars Liljas, and Olke C Uhlenbeck. "Rna Recognition by the Ms2 Phage Coat Protein." Paper presented at the Seminars in VIROLOGY, 1997.
[96] Kalhor, R., H. Tjong, N. Jayathilaka, F. Alber, and L. Chen. "Genome Architectures Revealed by Tethered Chromosome Conformation Capture and Population-Based Modeling." Nat Biotechnol 30, no. 1 (Dec 25 2011): 90-8.
[97] Kaplan, C. D., K. M. Larsson, and R. D. Kornberg. "The Rna Polymerase Ii Trigger Loop Functions in Substrate Selection and Is Directly Targeted by Alpha-Amanitin." Mol Cell 30, no. 5 (Jun 6 2008): 547-56.
[98] Kaufmann, I., G. Martin, A. Friedlein, H. Langen, and W. Keller. "Human Fip1 Is a Subunit of Cpsf That Binds to U-Rich Rna Elements and Stimulates Poly(a) Polymerase." EMBO J 23, no. 3 (Feb 11 2004): 616-26.
[99] Keniry, A., L. J. Gearing, N. Jansz, J. Liu, A. Z. Holik, P. F. Hickey, S. A. Kinkel, et al. "Setdb1-Mediated H3k9 Methylation Is Enriched on the Inactive X and Plays a Role in Its Epigenetic Silencing." Epigenetics Chromatin 9 (2016): 16.

[100] Khatter, H., M. K. Vorlander, and C. W. Muller. "Rna Polymerase I and Iii: Similar yet Unique." Curr Opin Struct Biol 47 (Dec 2017): 88-94.
[101] Kim, M., N. J. Krogan, L. Vasiljeva, O. J. Rando, E. Nedea, J. F. Greenblatt, and S. Buratowski. "The Yeast Rat1 Exonuclease Promotes Transcription Termination by Rna Polymerase Ii." Nature 432, no. 7016 (Nov 25 2004): 517-22.
[102] Kireeva, M. L., B. Hancock, G. H. Cremona, W. Walter, V. M. Studitsky, and M. Kashlev. "Nature of the Nucleosomal Barrier to Rna Polymerase Ii." Mol Cell 18, no. 1 (Apr 1 2005): 97-108.
[103] Kireeva, M. L., W. Walter, V. Tchernajenko, V. Bondarenko, M. Kashlev, and V. M. Studitsky. "Nucleosome Remodeling Induced by Rna Polymerase Ii: Loss of the H2a/H2b Dimer During Transcription." Mol Cell 9, no. 3 (Mar 2002): 541-52.
[104] Kirov, N., I. Tsaneva, E. Einbinder, and R. Tsanev. "In Vitro Transcription through Nucleosomes by T7 Rna Polymerase." EMBO J 11, no. 5 (May 1992): 1941-7.
[105] Knezetic, J. A., and D. S. Luse. "The Presence of Nucleosomes on a DNA Template Prevents Initiation by Rna Polymerase Ii in Vitro." Cell 45, no. 1 (Apr 11 1986): 95-104.
[106] Komarnitsky, P., E. J. Cho, and S. Buratowski. "Different Phosphorylated Forms of Rna Polymerase Ii and Associated Mrna Processing Factors During Transcription." Genes Dev 14, no. 19 (Oct 1 2000): 2452-60.
[107] Kornberg, R. D. "Chromatin Structure: A Repeating Unit of Histones and DNA." Science 184, no. 4139 (May 24 1974): 868-71.
[108] Kouzarides, T. "Chromatin Modifications and Their Function." Cell 128, no. 4 (Feb 23 2007): 693-705.
[109] Kristjuhan, A., and J. Q. Svejstrup. "Evidence for Distinct Mechanisms Facilitating Transcript Elongation through Chromatin in Vivo." EMBO J 23, no. 21 (Oct 27 2004): 4243-52.
[110] Kubik, S., M. J. Bruzzone, P. Jacquet, J. L. Falcone, J. Rougemont, and D. Shore. "Nucleosome Stability Distinguishes Two Different Promoter Types at All Protein-Coding Genes in Yeast." Mol Cell 60, no. 3 (Nov 5 2015): 422-34.
[111] Kuehner, J. N., E. L. Pearson, and C. Moore. "Unravelling the Means to an End: Rna Polymerase Ii Transcription Termination." Nat Rev Mol Cell Biol 12, no. 5 (May 2011): 283-94.
[112] Kujirai, T., H. Ehara, Y. Fujino, M. Shirouzu, S. I. Sekine, and H.

Kurumizaka. "Structural Basis of the Nucleosome Transition During Rna Polymerase Ii Passage." Science 362, no. 6414 (Nov 2 2018): 595-98.
[113] Kujirai, T., and H. Kurumizaka. "Transcription through the Nucleosome." Curr Opin Struct Biol 61 (Apr 2020): 42-49.
[114] Kulaeva, O. I., F. K. Hsieh, H. W. Chang, D. S. Luse, and V. M. Studitsky. "Mechanism of Transcription through a Nucleosome by Rna Polymerase Ii." Biochim Biophys Acta 1829, no. 1 (Jan 2013): 76-83.
[115] Kulish, D., and K. Struhl. "Tfiis Enhances Transcriptional Elongation through an Artificial Arrest Site in Vivo." Mol Cell Biol 21, no. 13 (Jul 2001): 4162-8.
[116] Kuras, Laurent, and Kevin Struhl. "Binding of Tbp to Promoters in Vivo Is Stimulated by Activators and Requires Pol Ii Holoenzyme." Nature 399, no. 6736 (1999): 609-13.
[117] Lackford, B., C. Yao, G. M. Charles, L. Weng, X. Zheng, E. A. Choi, X. Xie, et al. "Fip1 Regulates Mrna Alternative Polyadenylation to Promote Stem Cell Self-Renewal." EMBO J 33, no. 8 (Apr 16 2014): 878-89.
[118] Lai, W. K. M., and B. F. Pugh. "Understanding Nucleosome Dynamics and Their Links to Gene Expression and DNA Replication." Nat Rev Mol Cell Biol 18, no. 9 (Sep 2017): 548-62.
[119] Lambert, S. A., A. Jolma, L. F. Campitelli, P. K. Das, Y. Yin, M. Albu, X. Chen, et al. "The Human Transcription Factors." Cell 175, no. 2 (Oct 4 2018): 598-99.
[120] Lammers, N. C., V. Galstyan, A. Reimer, S. A. Medin, C. H. Wiggins, and H. G. Garcia. "Multimodal Transcriptional Control of Pattern Formation in Embryonic Development." Proc Natl Acad Sci U S A 117, no. 2 (Jan 14 2020): 836-47.
[121] Larson, D. R., D. Zenklusen, B. Wu, J. A. Chao, and R. H. Singer. "Real-Time Observation of Transcription Initiation and Elongation on an Endogenous Yeast Gene." Science 332, no. 6028 (Apr 22 2011): 475-8.
[122] Lauberth, S. M., T. Nakayama, X. Wu, A. L. Ferris, Z. Tang, S. H. Hughes, and R. G. Roeder. "H3k4me3 Interactions with Taf3 Regulate Preinitiation Complex Assembly and Selective Gene Activation." Cell 152, no. 5 (Feb 28 2013): 1021-36.
[123] Lawrence, M., W. Huber, H. Pages, P. Aboyoun, M. Carlson, R. Gentleman, M. T. Morgan, and V. J. Carey. "Software for Computing and Annotating Genomic Ranges." PLoS Comput Biol 9, no. 8 (2013): e1003118.
[124] Lee, C., H. Shin, and J. Kimble. "Dynamics of Notch-Dependent

Transcriptional Bursting in Its Native Context." Dev Cell 50, no. 4 (Aug 19 2019): 426-35 e4.
[125] Lee, D. K., J. DeJong, S. Hashimoto, M. Horikoshi, and R. G. Roeder. "Tfiia Induces Conformational Changes in Tfiid Via Interactions with the Basic Repeat." Mol Cell Biol 12, no. 11 (Nov 1992): 5189-96.
[126] Leuther, K. K., D. A. Bushnell, and R. D. Kornberg. "Two-Dimensional Crystallography of Tfiib- and Iie-Rna Polymerase Ii Complexes: Implications for Start Site Selection and Initiation Complex Formation." Cell 85, no. 5 (May 31 1996): 773-9.
[127] Li, B., M. Carey, and J. L. Workman. "The Role of Chromatin During Transcription." Cell 128, no. 4 (Feb 23 2007): 707-19.
[128] Li, M., A. Hada, P. Sen, L. Olufemi, M. A. Hall, B. Y. Smith, S. Forth, et al. "Dynamic Regulation of Transcription Factors by Nucleosome Remodeling." Elife 4 (Jun 5 2015).
[129] Li, W., B. You, M. Hoque, D. Zheng, W. Luo, Z. Ji, J. Y. Park, et al. "Systematic Profiling of Poly(a)+ Transcripts Modulated by Core 3' End Processing and Splicing Factors Reveals Regulatory Rules of Alternative Cleavage and Polyadenylation." PLoS Genet 11, no. 4 (Apr 2015): e1005166.
[130] Li, X. Y., A. Virbasius, X. Zhu, and M. R. Green. "Enhancement of Tbp Binding by Activators and General Transcription Factors." Nature 399, no. 6736 (Jun 10 1999): 605-9.
[131] Lin, Y. S., and M. R. Green. "Mechanism of Action of an Acidic Transcriptional Activator in Vitro." Cell 64, no. 5 (Mar 8 1991): 971-81.
[132] Lindstrom, D. L., S. L. Squazzo, N. Muster, T. A. Burckin, K. C. Wachter, C. A. Emigh, J. A. McCleery, J. R. Yates, 3rd, and G. A. Hartzog. "Dual Roles for Spt5 in Pre-Mrna Processing and Transcription Elongation Revealed by Identification of Spt5-Associated Proteins." Mol Cell Biol 23, no. 4 (Feb 2003): 1368-78.
[133] Liu, J., D. Hansen, E. Eck, Y. J. Kim, M. Turner, S. Alamos, and H. G. Garcia. "Real-Time Single-Cell Characterization of the Eukaryotic Transcription Cycle Reveals Correlations between Rna Initiation, Elongation, and Cleavage." PLoS Comput Biol 17, no. 5 (May 2021): e1008999.
[134] Liu, L., Y. Li, S. Li, N. Hu, Y. He, R. Pong, D. Lin, L. Lu, and M. Law. "Comparison of Next-Generation Sequencing Systems." J Biomed Biotechnol 2012

(2012): 251364.
[135] Lo, W. S., E. R. Gamache, K. W. Henry, D. Yang, L. Pillus, and S. L. Berger. "Histone H3 Phosphorylation Can Promote Tbp Recruitment through Distinct Promoter-Specific Mechanisms." EMBO J 24, no. 5 (Mar 9 2005): 997-1008.
[136] Logan, J., E. Falck-Pedersen, J. E. Darnell, Jr., and T. Shenk. "A Poly(a) Addition Site and a Downstream Termination Region Are Required for Efficient Cessation of Transcription by Rna Polymerase Ii in the Mouse Beta Maj-Globin Gene." Proc Natl Acad Sci U S A 84, no. 23 (Dec 1987): 8306-10.
[137] Lorch, Y., J. W. LaPointe, and R. D. Kornberg. "Nucleosomes Inhibit the Initiation of Transcription but Allow Chain Elongation with the Displacement of Histones." Cell 49, no. 2 (Apr 24 1987): 203-10.
[138] Lowary, P. T., and J. Widom. "New DNA Sequence Rules for High Affinity Binding to Histone Octamer and Sequence-Directed Nucleosome Positioning." J Mol Biol 276, no. 1 (Feb 13 1998): 19-42.
[139] Lucic, V., A. Leis, and W. Baumeister. "Cryo-Electron Tomography of Cells: Connecting Structure and Function." Histochem Cell Biol 130, no. 2 (Aug 2008): 185-96.
[140] Lue, N. F., and R. D. Kornberg. "Accurate Initiation at Rna Polymerase Ii Promoters in Extracts from Saccharomyces Cerevisiae." Proc Natl Acad Sci U S A 84, no. 24 (Dec 1987): 8839-43.
[141] Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent, and T. J. Richmond. "Crystal Structure of the Nucleosome Core Particle at 2.8 a Resolution." Nature 389, no. 6648 (Sep 18 1997): 251-60.
[142] Luk, E., A. Ranjan, P. C. Fitzgerald, G. Mizuguchi, Y. Huang, D. Wei, and C. Wu. "Stepwise Histone Replacement by Swr1 Requires Dual Activation with Histone H2a.Z and Canonical Nucleosome." Cell 143, no. 5 (Nov 24 2010): 725-36.
[143] Luo, L., M. Gribskov, and S. Wang. "Bibliometric Review of Atac-Seq and Its Application in Gene Expression." Brief Bioinform 23, no. 3 (May 13 2022).
[144] Luo, W., A. W. Johnson, and D. L. Bentley. "The Role of Rat1 in Coupling Mrna 3'-End Processing to Transcription Termination: Implications for a Unified Allosteric-Torpedo Model." Genes Dev 20, no. 8 (Apr 15 2006): 954-65.
[145] Mandal, S. S., C. Chu, T. Wada, H. Handa, A. J. Shatkin, and D. Reinberg. "Functional Interactions of Rna-Capping Enzyme with Factors That Positively and

Negatively Regulate Promoter Escape by Rna Polymerase Ii." Proc Natl Acad Sci U S A 101, no. 20 (May 18 2004): 7572-7.
[146] Maxon, M. E., J. A. Goodrich, and R. Tjian. "Transcription Factor Iie Binds Preferentially to Rna Polymerase Iia and Recruits Tfiih: A Model for Promoter Clearance." Genes Dev 8, no. 5 (Mar 1 1994): 515-24.
[147] McCracken, S., N. Fong, E. Rosonina, K. Yankulov, G. Brothers, D. Siderovski, A. Hessel, et al. "5'-Capping Enzymes Are Targeted to Pre-Mrna by Binding to the Phosphorylated Carboxy-Terminal Domain of Rna Polymerase Ii." Genes Dev 11, no. 24 (Dec 15 1997): 3306-18.
[148] Meaburn, E., and R. Schulz. "Next Generation Sequencing in Epigenetics: Insights and Challenges." Semin Cell Dev Biol 23, no. 2 (Apr 2012): 192-9.
[149] Meuleman, W., A. Muratov, E. Rynes, J. Halow, K. Lee, D. Bates, M. Diegel, et al. "Index and Biological Spectrum of Human Dnase I Hypersensitive Sites." Nature 584, no. 7820 (Aug 2020): 244-51.
[150] Meyer, C. A., and X. S. Liu. "Identifying and Mitigating Bias in Next-Generation Sequencing Methods for Chromatin Biology." Nat Rev Genet 15, no. 11 (Nov 2014): 709-21.
[151] Millan-Zambrano, G., A. Burton, A. J. Bannister, and R. Schneider. "Histone Post-Translational Modifications - Cause and Consequence of Genome Function." Nat Rev Genet 23, no. 9 (Sep 2022): 563-80.
[152] Miller, J. L., and P. A. Grant. "The Role of DNA Methylation and Histone Modifications in Transcriptional Regulation in Humans." Subcell Biochem 61 (2013): 289-317.
[153] Moebel, E., and C. Kervrann. "A Monte Carlo Framework for Missing Wedge Restoration and Noise Removal in Cryo-Electron Tomography." J Struct Biol X 4 (2020): 100013.
[154] Morrison, O., and J. Thakur. "Molecular Complexes at Euchromatin, Heterochromatin and Centromeric Chromatin." Int J Mol Sci 22, no. 13 (Jun 28 2021).
[155] Morse, R. H. "Nucleosomes Inhibit Both Transcriptional Initiation and Elongation by Rna Polymerase Iii in Vitro." EMBO J 8, no. 8 (Aug 1989): 2343-51.
[156] Moter, A., and U. B. Gobel. "Fluorescence in Situ Hybridization (Fish) for Direct Visualization of Microorganisms." J Microbiol Methods 41, no. 2 (Jul 2000):

85-112.
[157] Muller, J., C. M. Hart, N. J. Francis, M. L. Vargas, A. Sengupta, B. Wild, E. L. Miller, et al. "Histone Methyltransferase Activity of a Drosophila Polycomb Group Repressor Complex." Cell 111, no. 2 (Oct 18 2002): 197-208.
[158] Murata, K., and M. Wolf. "Cryo-Electron Microscopy for Structural Analysis of Dynamic Biological Macromolecules." Biochim Biophys Acta Gen Subj 1862, no. 2 (Feb 2018): 324-34.
[159] Mutskov, V., D. Gerber, D. Angelov, J. Ausio, J. Workman, and S. Dimitrov. "Persistent Interactions of Core Histone Tails with Nucleosomal DNA Following Acetylation and Transcription Factor Binding." Mol Cell Biol 18, no. 11 (Nov 1998): 6293-304.
[160] Mutz, K. O., A. Heilkenbrinker, M. Lonne, J. G. Walter, and F. Stahl. "Transcriptome Analysis Using Next-Generation Sequencing." Curr Opin Biotechnol 24, no. 1 (Feb 2013): 22-30.
[161] Nabavi, S., and R. N. Nazar. "Nonpolyadenylated Rna Polymerase Ii Termination Is Induced by Transcript Cleavage." J Biol Chem 283, no. 20 (May 16 2008): 13601-10.
[162] Nakajima, N., M. Horikoshi, and R. G. Roeder. "Factors Involved in Specific Transcription by Mammalian Rna Polymerase Ii: Purification, Genetic Specificity, and Tata Box-Promoter Interactions of Tfiid." Mol Cell Biol 8, no. 10 (Oct 1988): 4028-40.
[163] Nathan, D., D. E. Sterner, and S. L. Berger. "Histone Modifications: Now Summoning Sumoylation." Proc Natl Acad Sci U S A 100, no. 23 (Nov 11 2003): 13118-20.
[164] Ng, H. H., F. Robert, R. A. Young, and K. Struhl. "Targeted Recruitment of Set1 Histone Methylase by Elongating Pol Ii Provides a Localized Mark and Memory of Recent Transcriptional Activity." Mol Cell 11, no. 3 (Mar 2003): 709-19.
[165] Ng, P. C., and E. F. Kirkness. "Whole Genome Sequencing." Methods Mol Biol 628 (2010): 215-26.
[166] Nguyen, V. T., T. Kiss, A. A. Michels, and O. Bensaude. "7sk Small Nuclear Rna Binds to and Inhibits the Activity of Cdk9/Cyclin T Complexes." Nature 414, no. 6861 (Nov 15 2001): 322-5.
[167] Niedringhaus, T. P., D. Milanova, M. B. Kerby, M. P. Snyder, and A. E.

Barron. "Landscape of Next-Generation Sequencing Technologies." Anal Chem 83, no. 12 (Jun 15 2011): 4327-41.
[168] Nien, C. Y., H. L. Liang, S. Butcher, Y. Sun, S. Fu, T. Gocha, N. Kirov, J. R. Manak, and C. Rushlow. "Temporal Coordination of Gene Networks by Zelda in the Early Drosophila Embryo." PLoS Genet 7, no. 10 (Oct 2011): e1002339.
[169] Nogales, E., R. K. Louder, and Y. He. "Structural Insights into the Eukaryotic Transcription Initiation Machinery." Annu Rev Biophys 46 (May 22 2017): 59-83.
[170] Nuebler, J., G. Fudenberg, M. Imakaev, N. Abdennur, and L. A. Mirny. "Chromatin Organization by an Interplay of Loop Extrusion and Compartmental Segregation." Proc Natl Acad Sci U S A 115, no. 29 (Jul 17 2018): E6697-E706.
[171] O'Brien, T., and J. T. Lis. "Rapid Changes in Drosophila Transcription after an Instantaneous Heat Shock." Mol Cell Biol 13, no. 6 (Jun 1993): 3456-63.
[172] Ohkuma, Y., S. Hashimoto, C. K. Wang, M. Horikoshi, and R. G. Roeder. "Analysis of the Role of Tfiie in Basal Transcription and Tfiih-Mediated Carboxy-Terminal Domain Phosphorylation through Structure-Function Studies of Tfiie-Alpha." Mol Cell Biol 15, no. 9 (Sep 1995): 4856-66.
[173] Olins, A. L., and D. E. Olins. "Spheroid Chromatin Units (V Bodies)." Science 183, no. 4122 (Jan 25 1974): 330-2.
[174] Orphanides, G., G. LeRoy, C. H. Chang, D. S. Luse, and D. Reinberg. "Fact, a Factor That Facilitates Transcript Elongation through Nucleosomes." Cell 92, no. 1 (Jan 9 1998): 105-16.
[175] Orth, J. D., R. H. Kohler, F. Foijer, P. K. Sorger, R. Weissleder, and T. J. Mitchison. "Analysis of Mitosis and Antimitotic Drug Responses in Tumors by in Vivo Microscopy and Single-Cell Pharmacodynamics." Cancer Res 71, no. 13 (Jul 1 2011): 4608-16.
[176] Palangat, M., M. H. Larson, X. Hu, A. Gnatt, S. M. Block, and R. Landick. "Efficient Reconstitution of Transcription Elongation Complexes for Single-Molecule Studies of Eukaryotic Rna Polymerase Ii." Transcription 3, no. 3 (May-Jun 2012): 146-53.
[177] Park, P. J. "Chip-Seq: Advantages and Challenges of a Maturing Technology." Nat Rev Genet 10, no. 10 (Oct 2009): 669-80.
[178] Park, S. Y., H. C. Moon, and H. Y. Park. "Live-Cell Imaging of Single Mrna Dynamics Using Split Superfolder Green Fluorescent Proteins with Minimal

Background." RNA 26, no. 1 (Jan 2020): 101-09.
[179] Penagos-Puig, A., and M. Furlan-Magaril. "Heterochromatin as an Important Driver of Genome Organization." Front Cell Dev Biol 8 (2020): 579137.
[180] Perez Canadillas, J. M., and G. Varani. "Recognition of Gu-Rich Polyadenylation Regulatory Elements by Human Cstf-64 Protein." EMBO J 22, no. 11 (Jun 2 2003): 2821-30.
[181] Peterlin, B. M., and D. H. Price. "Controlling the Elongation Phase of Transcription with P-Tefb." Mol Cell 23, no. 3 (Aug 4 2006): 297-305.
[182] Proudfoot, N. J., A. Furger, and M. J. Dye. "Integrating Mrna Processing with Transcription." Cell 108, no. 4 (Feb 22 2002): 501-12.
[183] Rackham, O., and C. M. Brown. "Visualization of Rna-Protein Interactions in Living Cells: Fmrp and Imp1 Interact on Mrnas." EMBO J 23, no. 16 (Aug 18 2004): 3346-55.
[184] Raha, D., M. Hong, and M. Snyder. "Chip-Seq: A Method for Global Identification of Regulatory Elements in the Genome." Curr Protoc Mol Biol Chapter 21 (Jul 2010): Unit 21 19 1-14.
[185] Ramachandran, S., G. E. Zentner, and S. Henikoff. "Asymmetric Nucleosomes Flank Promoters in the Budding Yeast Genome." Genome Res 25, no. 3 (Mar 2015): 381-90.
[186] Rando, O. J., and F. Winston. "Chromatin and Transcription in Yeast." Genetics 190, no. 2 (Feb 2012): 351-87.
[187] Ranzoni, A. M., A. Tangherloni, I. Berest, S. G. Riva, B. Myers, P. M. Strzelecka, J. Xu, et al. "Integrative Single-Cell Rna-Seq and Atac-Seq Analysis of Human Developmental Hematopoiesis." Cell Stem Cell 28, no. 3 (Mar 4 2021): 472-87 e7.
[188] Rasmussen, E. B., and J. T. Lis. "In Vivo Transcriptional Pausing and Cap Formation on Three Drosophila Heat Shock Genes." Proc Natl Acad Sci U S A 90, no. 17 (Sep 1 1993): 7923-7.
[189] Ray-Soni, A., M. J. Bellecourt, and R. Landick. "Mechanisms of Bacterial Transcription Termination: All Good Things Must End." Annu Rev Biochem 85 (Jun 2 2016): 319-47.
[190] Richard, P., and J. L. Manley. "Transcription Termination by Nuclear Rna Polymerases." Genes Dev 23, no. 11 (Jun 1 2009): 1247-69.

[191] Richards, E. J., and S. C. Elgin. "Epigenetic Codes for Heterochromatin Formation and Silencing: Rounding up the Usual Suspects." Cell 108, no. 4 (Feb 22 2002): 489-500.
[192] Roberts, J. W. "Mechanisms of Bacterial Transcription Termination." J Mol Biol 431, no. 20 (Sep 20 2019): 4030-39.
[193] Roeder, R. G. "Transcriptional Regulation and the Role of Diverse Coactivators in Animal Cells." FEBS Lett 579, no. 4 (Feb 7 2005): 909-15.
[194] Roeder, R. G., and W. J. Rutter. "Multiple Forms of DNA-Dependent Rna Polymerase in Eukaryotic Organisms." Nature 224, no. 5216 (Oct 18 1969): 234-7.
[195] Rosonina, E., S. Kaneko, and J. L. Manley. "Terminating the Transcript: Breaking up Is Hard to Do." Genes Dev 20, no. 9 (May 1 2006): 1050-6.
[196] Rossetto, D., N. Avvakumov, and J. Cote. "Histone Phosphorylation: A Chromatin Modification Involved in Diverse Nuclear Events." Epigenetics 7, no. 10 (Oct 2012): 1098-108.
[197] Roth, S. Y., J. M. Denu, and C. D. Allis. "Histone Acetyltransferases." Annu Rev Biochem 70 (2001): 81-120.
[198] Rougvie, A. E., and J. T. Lis. "Postinitiation Transcriptional Control in Drosophila Melanogaster." Mol Cell Biol 10, no. 11 (Nov 1990): 6041-5.
[199] ———. "The Rna Polymerase Ii Molecule at the 5' End of the Uninduced Hsp70 Gene of D. Melanogaster Is Transcriptionally Engaged." Cell 54, no. 6 (Sep 9 1988): 795-804.
[200] Rust, M. J., M. Bates, and X. Zhuang. "Sub-Diffraction-Limit Imaging by Stochastic Optical Reconstruction Microscopy (Storm)." Nat Methods 3, no. 10 (Oct 2006): 793-5.
[201] Ryu, H. Y., and M. Hochstrasser. "Histone Sumoylation and Chromatin Dynamics." Nucleic Acids Res 49, no. 11 (Jun 21 2021): 6043-52.
[202] Saksouk, N., E. Simboeck, and J. Dejardin. "Constitutive Heterochromatin Formation and Transcription in Mammals." Epigenetics Chromatin 8 (2015): 3.
[203] Saldi, T., K. Riemondy, B. Erickson, and D. L. Bentley. "Alternative Rna Structures Formed During Transcription Depend on Elongation Rate and Modify Rna Processing." Mol Cell 81, no. 8 (Apr 15 2021): 1789-801 e5.
[204] Schneider, R., and R. Grosschedl. "Dynamics and Interplay of Nuclear Architecture, Genome Organization, and Gene Expression." Genes Dev 21, no. 23

(Dec 1 2007): 3027-43.
[205] Schones, D. E., K. Cui, S. Cuddapah, T. Y. Roh, A. Barski, Z. Wang, G. Wei, and K. Zhao. "Dynamic Regulation of Nucleosome Positioning in the Human Genome." Cell 132, no. 5 (Mar 7 2008): 887-98.
[206] Schroeder, S. C., B. Schwer, S. Shuman, and D. Bentley. "Dynamic Association of Capping Enzymes with Transcribing Rna Polymerase Ii." Genes Dev 14, no. 19 (Oct 1 2000): 2435-40.
[207] Schwabish, M. A., and K. Struhl. "Asf1 Mediates Histone Eviction and Deposition During Elongation by Rna Polymerase Ii." Mol Cell 22, no. 3 (May 5 2006): 415-22.
[208] Schweikhard, V., C. Meng, K. Murakami, C. D. Kaplan, R. D. Kornberg, and S. M. Block. "Transcription Factors Tfiif and Tfiis Promote Transcript Elongation by Rna Polymerase Ii by Synergistic and Independent Mechanisms." Proc Natl Acad Sci U S A 111, no. 18 (May 6 2014): 6642-7.
[209] Shaw, P. A., C. G. Sahasrabuddhe, H. G. Hodo, 3rd, and G. F. Saunders. "Transcription of Nucleosomes from Human Chromatin." Nucleic Acids Res 5, no. 8 (Aug 1978): 2999-3012.
[210] Shiio, Y., and R. N. Eisenman. "Histone Sumoylation Is Associated with Transcriptional Repression." Proc Natl Acad Sci U S A 100, no. 23 (Nov 11 2003): 13225-30.
[211] Shmueli, M. D., D. Sheban, A. Eisenberg-Lerner, and Y. Merbl. "Histone Degradation by the Proteasome Regulates Chromatin and Cellular Plasticity." FEBS J 289, no. 12 (Jun 2022): 3304-16.
[212] Simon, J. M., P. G. Giresi, I. J. Davis, and J. D. Lieb. "Using Formaldehyde-Assisted Isolation of Regulatory Elements (Faire) to Isolate Active Regulatory DNA." Nat Protoc 7, no. 2 (Jan 19 2012): 256-67.
[213] Simpson, D. A., A. J. Thompson, M. Kowarsky, N. F. Zeeshan, M. S. Barson, L. T. Hall, Y. Yan, et al. "In Vivo Imaging and Tracking of Individual Nanodiamonds in Drosophila Melanogaster Embryos." Biomed Opt Express 5, no. 4 (Apr 1 2014): 1250-61.
[214] Sims, R. J., 3rd, R. Belotserkovskaya, and D. Reinberg. "Elongation by Rna Polymerase Ii: The Short and Long of It." Genes Dev 18, no. 20 (Oct 15 2004): 2437-68.

[215] Singh, A. K., and F. Mueller-Planitz. "Nucleosome Positioning and Spacing: From Mechanism to Function." J Mol Biol 433, no. 6 (Mar 19 2021): 166847.
[216] Singh, A. K., T. Schauer, L. Pfaller, T. Straub, and F. Mueller-Planitz. "The Biogenesis and Function of Nucleosome Arrays." Nat Commun 12, no. 1 (Dec 1 2021): 7011.
[217] Singh, J., and R. A. Padgett. "Rates of in Situ Transcription and Splicing in Large Human Genes." Nat Struct Mol Biol 16, no. 11 (Nov 2009): 1128-33.
[218] Sklar, V. E., L. B. Schwartz, and R. G. Roeder. "Distinct Molecular Structures of Nuclear Class I, Ii, and Iii DNA-Dependent Rna Polymerases." Proc Natl Acad Sci U S A 72, no. 1 (Jan 1975): 348-52.
[219] Slatko, B. E., A. F. Gardner, and F. M. Ausubel. "Overview of Next-Generation Sequencing Technologies." Curr Protoc Mol Biol 122, no. 1 (Apr 2018): e59.
[220] Soluri, I. V., L. M. Zumerling, O. A. Payan Parra, E. G. Clark, and S. A. Blythe. "Zygotic Pioneer Factor Activity of Odd-Paired/Zic Is Necessary for Late Function of the Drosophila Segmentation Network." Elife 9 (Apr 29 2020).
[221] Song, L., and G. E. Crawford. "Dnase-Seq: A High-Resolution Technique for Mapping Active Gene Regulatory Elements across the Genome from Mammalian Cells." Cold Spring Harb Protoc 2010, no. 2 (Feb 2010): pdb prot5384.
[222] Stephens, D. J., and V. J. Allan. "Light Microscopy Techniques for Live Cell Imaging." Science 300, no. 5616 (Apr 4 2003): 82-6.
[223] Struhl, K. "Histone Acetylation and Transcriptional Regulatory Mechanisms." Genes Dev 12, no. 5 (Mar 1 1998): 599-606.
[224] Studitsky, V. M., D. J. Clark, and G. Felsenfeld. "Overcoming a Nucleosomal Barrier to Transcription." Cell 83, no. 1 (Oct 6 1995): 19-27.
[225] Studitsky, V. M., E. V. Nizovtseva, A. K. Shaytan, and D. S. Luse. "Nucleosomal Barrier to Transcription: Structural Determinants and Changes in Chromatin Structure." Biochem Mol Biol J 2, no. 2 (2016).
[226] Szabo, Q., F. Bantignies, and G. Cavalli. "Principles of Genome Folding into Topologically Associating Domains." Sci Adv 5, no. 4 (Apr 2019): eaaw1668.
[227] Takagaki, Y., and J. L. Manley. "Rna Recognition by the Human Polyadenylation Factor Cstf." Mol Cell Biol 17, no. 7 (Jul 1997): 3907-14.
[228] Takagaki, Y., J. L. Manley, C. C. MacDonald, J. Wilusz, and T. Shenk. "A

Multisubunit Factor, Cstf, Is Required for Polyadenylation of Mammalian Pre-Mrnas." Genes Dev 4, no. 12A (Dec 1990): 2112-20.
[229] Tan, S., T. Aso, R. C. Conaway, and J. W. Conaway. "Roles for Both the Rap30 and Rap74 Subunits of Transcription Factor Iif in Transcription Initiation and Elongation by Rna Polymerase Ii." J Biol Chem 269, no. 41 (Oct 14 1994): 25684-91.
[230] Teixeira, A., A. Tahiri-Alaoui, S. West, B. Thomas, A. Ramadass, I. Martianov, M. Dye, et al. "Autocatalytic Rna Cleavage in the Human Beta-Globin Pre-Mrna Promotes Transcription Termination." Nature 432, no. 7016 (Nov 25 2004): 526-30.
[231] Thastrom, A., P. T. Lowary, H. R. Widlund, H. Cao, M. Kubista, and J. Widom. "Sequence Motifs and Free Energies of Selected Natural and Non-Natural Nucleosome Positioning DNA Sequences." J Mol Biol 288, no. 2 (Apr 30 1999): 213-29.
[232] Thummel, C. S., K. C. Burtis, and D. S. Hogness. "Spatial and Temporal Patterns of E74 Transcription During Drosophila Development." Cell 61, no. 1 (Apr 6 1990): 101-11.
[233] Tian, B., and J. H. Graber. "Signals for Pre-Mrna Cleavage and Polyadenylation." Wiley Interdiscip Rev RNA 3, no. 3 (May-Jun 2012): 385-96.
[234] Tirosh, I., and N. Barkai. "Two Strategies for Gene Regulation by Promoter Nucleosomes." Genome Res 18, no. 7 (Jul 2008): 1084-91.
[235] Tutucci, E., M. Vera, J. Biswas, J. Garcia, R. Parker, and R. H. Singer. "An Improved Ms2 System for Accurate Reporting of the Mrna Life Cycle." Nat Methods 15, no. 1 (Jan 2018): 81-89.
[236] Veloso, A., K. S. Kirkconnell, B. Magnuson, B. Biewen, M. T. Paulsen, T. E. Wilson, and M. Ljungman. "Rate of Elongation by Rna Polymerase Ii Is Associated with Specific Gene Features and Epigenetic Modifications." Genome Res 24, no. 6 (Jun 2014): 896-905.
[237] Venkatesh, S., and J. L. Workman. "Histone Exchange, Chromatin Structure and the Regulation of Transcription." Nat Rev Mol Cell Biol 16, no. 3 (Mar 2015): 178-89.
[238] Vera, M., E. Tutucci, and R. H. Singer. "Imaging Single Mrna Molecules in Mammalian Cells Using an Optimized Ms2-Mcp System." Methods Mol Biol 2038 (2019): 3-20.

[239] Vermeulen, M., K. W. Mulder, S. Denissov, W. W. Pijnappel, F. M. van Schaik, R. A. Varier, M. P. Baltissen, et al. "Selective Anchoring of Tfiid to Nucleosomes by Trimethylation of Histone H3 Lysine 4." Cell 131, no. 1 (Oct 5 2007): 58-69.
[240] Visser, A. E., and J. A. Aten. "Chromosomes as Well as Chromosomal Subdomains Constitute Distinct Units in Interphase Nuclei." J Cell Sci 112 (Pt 19) (Oct 1999): 3353-60.
[241] Vogelstein, B., N. Papadopoulos, V. E. Velculescu, S. Zhou, L. A. Diaz, Jr., and K. W. Kinzler. "Cancer Genome Landscapes." Science 339, no. 6127 (Mar 29 2013): 1546-58.
[242] Vos, S. M., L. Farnung, H. Urlaub, and P. Cramer. "Structure of Paused Transcription Complex Pol Ii-Dsif-Nelf." Nature 560, no. 7720 (Aug 2018): 601-06.
[243] Wang, H., S. Schilbach, M. Ninov, H. Urlaub, and P. Cramer. "Structures of Transcription Preinitiation Complex Engaged with the +1 Nucleosome." Nat Struct Mol Biol 30, no. 2 (Feb 2023): 226-32.
[244] Wang, H. W., and J. W. Wang. "How Cryo-Electron Microscopy and X-Ray Crystallography Complement Each Other." Protein Sci 26, no. 1 (Jan 2017): 32-39.
[245] Wang, J., C. Zibetti, P. Shang, S. R. Sripathi, P. Zhang, M. Cano, T. Hoang, et al. "Atac-Seq Analysis Reveals a Widespread Decrease of Chromatin Accessibility in Age-Related Macular Degeneration." Nat Commun 9, no. 1 (Apr 10 2018): 1364.
[246] Wansink, D. G., W. Schul, I. van der Kraan, B. van Steensel, R. van Driel, and L. de Jong. "Fluorescent Labeling of Nascent Rna Reveals Transcription by Rna Polymerase Ii in Domains Scattered Throughout the Nucleus." J Cell Biol 122, no. 2 (Jul 1993): 283-93.
[247] Weber, M. S., M. Wojtynek, and O. Medalia. "Cellular and Structural Studies of Eukaryotic Cells by Cryo-Electron Tomography." Cells 8, no. 1 (Jan 16 2019).
[248] Wei, Y., L. Yu, J. Bowen, M. A. Gorovsky, and C. D. Allis. "Phosphorylation of Histone H3 Is Required for Proper Chromosome Condensation and Segregation." Cell 97, no. 1 (Apr 2 1999): 99-109.
[249] Wells, A. L., J. S. Condeelis, R. H. Singer, and D. Zenklusen. "Imaging Real-Time Gene Expression in Living Systems with Single-Transcript Resolution: Construct Design and Imaging System Setup." CSH Protoc 2007 (Nov 1 2007): pdb top28.

[250] Wen, Y., and A. J. Shatkin. "Transcription Elongation Factor Hspt5 Stimulates Mrna Capping." Genes Dev 13, no. 14 (Jul 15 1999): 1774-9.
[251] West, S., N. Gromak, and N. J. Proudfoot. "Human 5' --> 3' Exonuclease Xrn2 Promotes Transcription Termination at Co-Transcriptional Cleavage Sites." Nature 432, no. 7016 (Nov 25 2004): 522-5.
[252] Whitelaw, E., and N. Proudfoot. "Alpha-Thalassaemia Caused by a Poly(a) Site Mutation Reveals That Transcriptional Termination Is Linked to 3' End Processing in the Human Alpha 2 Globin Gene." EMBO J 5, no. 11 (Nov 1986): 2915-22.
[253] Williams, R. R. "Transcription and the Territory: The Ins and Outs of Gene Positioning." Trends Genet 19, no. 6 (Jun 2003): 298-302.
[254] Workman, J. L., and R. E. Kingston. "Alteration of Nucleosome Structure as a Mechanism of Transcriptional Regulation." Annu Rev Biochem 67 (1998): 545-79.
[255] Wutz, G., C. Varnai, K. Nagasaka, D. A. Cisneros, R. R. Stocsits, W. Tang, S. Schoenfelder, et al. "Topologically Associating Domains and Chromatin Loops Depend on Cohesin and Are Regulated by Ctcf, Wapl, and Pds5 Proteins." EMBO J 36, no. 24 (Dec 15 2017): 3573-99.
[256] Yamane, K., K. Tateishi, R. J. Klose, J. Fang, L. A. Fabrizio, H. Erdjument-Bromage, J. Taylor-Papadimitriou, P. Tempst, and Y. Zhang. "Plu-1 Is an H3k4 Demethylase Involved in Transcriptional Repression and Breast Cancer Cell Proliferation." Mol Cell 25, no. 6 (Mar 23 2007): 801-12.
[257] Yan, F., D. R. Powell, D. J. Curtis, and N. C. Wong. "From Reads to Insight: A Hitchhiker's Guide to Atac-Seq Data Analysis." Genome Biol 21, no. 1 (Feb 3 2020): 22.
[258] Yang, Z., Q. Zhu, K. Luo, and Q. Zhou. "The 7sk Small Nuclear Rna Inhibits the Cdk9/Cyclin T1 Kinase to Control Transcription." Nature 414, no. 6861 (Nov 15 2001): 317-22.
[259] Young, R. A. "Rna Polymerase Ii." Annu Rev Biochem 60 (1991): 689-715.
[260] Zaugg, J. B., and N. M. Luscombe. "A Genomic Model of Condition-Specific Nucleosome Behavior Explains Transcriptional Activity in Yeast." Genome Res 22, no. 1 (Jan 2012): 84-94.
[261] Zhang, M., C. Diaz-Celis, B. Onoa, C. Canari-Chumpitaz, K. I. Requejo, J. Liu, M. Vien, et al. "Molecular Organization of the Early Stages of Nucleosome

Phase Separation Visualized by Cryo-Electron Tomography." Mol Cell 82, no. 16 (Aug 18 2022): 3000-14 e9.
[262] Zhang, S., S. Aibara, S. M. Vos, D. E. Agafonov, R. Luhrmann, and P. Cramer. "Structure of a Transcribing Rna Polymerase Ii-U1 Snrnp Complex." Science 371, no. 6526 (Jan 15 2021): 305-09.
[263] Zhang, T., S. Cooper, and N. Brockdorff. "The Interplay of Histone Modifications - Writers That Read." EMBO Rep 16, no. 11 (Nov 2015): 1467-81.
[264] Zhang, Y. "Transcriptional Regulation by Histone Ubiquitination and Deubiquitination." Genes Dev 17, no. 22 (Nov 15 2003): 2733-40.
[265] Zhang, Y., and D. Reinberg. "Transcription Regulation by Histone Methylation: Interplay between Different Covalent Modifications of the Core Histone Tails." Genes Dev 15, no. 18 (Sep 15 2001): 2343-60.
[266] Zuleger, N., M. I. Robson, and E. C. Schirmer. "The Nuclear Envelope as a Chromatin Organizer." Nucleus 2, no. 5 (Sep-Oct 2011): 339-49.