INVESTIGATION INTO FACTORS INVOLVED IN THE TRANSCRIPTIONAL REGULATION OF SNRNA GENES

Mathias Opeyemi Fagbemi

Lady Margaret Hall

University of Oxford

A thesis submitted for the degree of Doctor of Philosophy at the University of Oxford, Michaelmas 2024

Table of Contents

Acknowledgements

I want to begin by expressing my deepest gratitude to **Professor Shona Murphy.** You took a chance on me and gave me the opportunity to work on the project I was passionate about while giving me the freedom to explore and the unwavering support I needed. If I could go back in time, I would choose the same project under the same supervisor without hesitation.

I would also like to thank all current and past members of the Murphy Lab. **Dr Joana Guiro**, for igniting the fire with the project and providing essential training during the early stages. **Steph Barker,** your relentless support with my experiments and lab admin was invaluable! **Dr Isabella Maudlin,** I am very grateful for your technical and theoretical guidance and support throughout my journey. **Dr Cecilia Studniarek**, the lab would not be the same without you; thank you for all the fun memories. **Dr. Micheal Tellier,** I appreciate all your support with my project; thank you. **Dr. Callum Henfrey,** thank you for your support with bioinformatic analysis. **Dr. Gilbert Ansa**, your encouragement through my DPhil kept me going, especially through our late nights in the lab. **Aleksandra Pluta**, my PhD buddy, I am glad we went through this experience together. It would not have been the same without you.

I would also like to thank **Professor Chris Norbury**, **Professor Natasha Gromak, Professor Nick Proudfoot, Professor Andre Furger, Professor Ervin Fodor and the Gromak lab** for their insightful discussions during my assessments and presentations, which helped to refine my work.

To my family – **Dad, Mum, Dami and Mololuwa**, your unwavering support means the world to me. I could not have done this without you. To my late sister, **Deborah**, your encouragement inspires me to “give life my best shot”. Thinking of you made cycling to the lab on those freezing mornings much easier. Finally, I am immensely grateful to the **Ments**, particularly **Jessica**, for your love and support over the past year, especially throughout the writing process.

I dedicate this thesis to my late Grandpa, **Francis Ademola Fagbemi**, whose desire to become a scientist was only limited by circumstance.

Declaration

The work presented here was conducted between October 2020 and October 2024 at the Murphy Laboratory, Sir William Dunn School of Pathology, Oxford.

Except where stated otherwise, the data presented here is wholly mine. The data presented in section 5.2 is now published in *Guiro et al.,* 2022, where I am the co-first author. Subsection (2.9.1) was written in collaboration with the Genome Engineering Oxford (GEO) facility, and Subsection (2.5.7.3) was written in collaboration with Dr Henfrey. The HDR templates containing degradation, SMASH and 3X FLAG tags for respective proteins were designed by Gilbert Ansa and constructed by Prof. Murphy (section 2.9.2). Additionally, I collaborated with Mrs Barker for the screening of degron-tagged cell lines using western blot screening (section 5.6).

All collaborations are duly acknowledged within the body of the thesis.

Mathias O. Fagbemi

October 2024

Abbreviations

AS	Analog-sensitive
BRAT1	BRCA1-associated ATM activator 1
BrdU	5-Bromo-2'-Deoxyuridine
BRE	Transcription factor IIB recognition element
BRF2	TFIIB-related factor 2
CA	Calyculin-A
CAPTURE	CRISPR Affinity Purification *in situ* of Regulatory element
CBC	Cap-binding complex
CBCA	CBC-Arsenic resistance
CDK	Cyclin-dependent kinases
cDNA	complementary DNA
CPSF	Cleavage and polyadenylation specificity factor-73
CstF	Cleavage stimulatory factor
CTD	Carboxyl-terminal domain
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DRB	5, 6-dichloro-1-β-D-ribofuranosylbenzimidazole
DSB	Double-stranded break
DSE	Downstream sequence element
DSIF	DRB sensitivity-inducing factor
ETS	external transcribed spacer
FLASH	FLICE-associated huge protein
g	Grams
g	Relative centrifugal force
gDNA	genomic DNA
GRO-Seq	Global Run-On sequencing
GTF	General transcription factor
HEK	Human Embryonic Kidney
HP1	Heterochromatin protein-1

ICE	interactor of the little elongation complex
ICM	Integrator cleavage module
IntS	Integrator complex subunit
ITS	internal transcribed spacer
LEC	Little elongation complex
lncRNA	long noncoding RNA
m3G	$N^{2,2,7}$-trimethyl guanosine
m7G	N^{7}-methyl guanosine
miRNA	Micro RNA
mNET-seq	mammalian Native Elongation Transcript
mRNA	Messenger RNA
NELF	negative elongation factor
NM	1-NM-PP1
P-TEFb	positive transcription elongator factor b
PAS	Polyadenylation signal
PBP	PSE-binding protein
PCR	polymerase chain reaction
PHAX	Phosphorylated adaptor for RNA export
PIC	Preinitiation complex
Pol II	RNA polymerase II
PSE	Proximal Sequence Element
pSer	phosphorylated Serine
PTF	PSE-transcription factor
PTM	post-translational modification
pTyr	phosphorylated Tyrosine
qPCR	quantitative PCR
RNA	Ribonucleic acid
RNase	Ribonuclease
RPAP2	RNA Pol II-associated protein 2
RPM	Revolutions per minute
RT	reverse transcription; readthrough snRNA; room temperature
SEC	Super elongation complex

SLBP	Stem-loop binding protein
SNAPc	snRNA activating protein complex
snoRNP	small nucleolar RNP
snRNA	Small nuclear RNA
snRNP	Small nuclear ribonucleoproteins
SOSS	Sensor of the single-stranded DNA
SR	Splicing regulatory
SS	Splice site
SSU	small subunit
TAF	TATA-binding protein (TBP) associated factors
tCDK	transcriptional CDK
TFII	Transcription factor II
TOE	Target of EGR
tRNA	Transfer RNA
TSS	transcription start site
TT	Tautomycetin
U	Uridine-rich
μM	micromolar

Abstract

The human pol II-dependent small nuclear (sn)RNAs are non-coding RNAs which function as small nuclear ribonucleoproteins (snRNPs) and play key roles in the processing of other RNAs, including splicing of pre-mRNA. Unlike most mRNAs, these snRNAs are not spliced or polyadenylated, and 3' end formation is directed by a gene-specific 3' box rather than a poly(A) site. Cleavage by the Integrator complex upstream of the 3' box produces pre-snRNAs that are further processed in the cytoplasm. In addition, these snRNA genes have a specialised promoter comprising an enhancer-like DSE and an essential PSE that functions as the core promoter. The enzymatic activity of several transcriptional cyclin-dependent kinases (CDKs) regulates the transcription of protein-coding genes and the processing of the transcripts. CDK9 has been shown to regulate the processing of snRNA gene transcripts by ensuring the recruitment of the Integrator complex. However, the role of other CDKs is not yet clear. I have used analog-sensitive CDK cell lines, RT-qPCR, and chromatin immunoprecipitation to investigate the role of other transcriptional CDKs in the expression of human snRNA genes. I have found that the activity of either CDK12 or CDK13 is required for the 3' end processing of snRNA gene transcripts, and inhibition of both CDK12 and CDK13 causes failure to recognise the 3' box. In addition, CDK12 and CDK13, as well as the elongation factor, SPT6, help to recruit Integrator subunits to snRNA genes. Using ChIP-qPCR, I have demonstrated that inhibition of CDK12 or CDK13 affects phosphorylation of CTD residues other than Ser2 and Ser 5 during the transcription of U2 snRNA genes. This work provides new insights into factors that regulate the 3' end processing of snRNA genes and the potential mechanisms underlying their function.

Chapter 1 | Introduction

1.1 Eukaryotic gene expression

Transcription is the process by which the inherited genetic blueprint, DNA, is copied into RNA. It is one of three fundamental processes of the central dogma of molecular biology, which includes DNA replication, RNA synthesis (transcription) and protein synthesis (translation) [1]. In contrast to prokaryotes and bacteriophages, eukaryotic genetic information in chromosomes is tightly stored in the nucleus as nucleoprotein complexes called Chromatin [2] (Fig. 1.1). The fundamental repeating block of chromatin is the nucleosome, which consists of approximately 146-147 base pairs of DNA wrapped 1.65 times around an octamer of four highly conserved core histone proteins, H2A, H2B, H3 and H4 [3–5] (Fig. 1.1). There are two major chromatin regions on the chromosomes: Heterochromatin and Euchromatin[6,7]. Heterochromatin is a higher-order chromatin structure characterised by highly compacted nucleosome structures (~30 nm) due to the activity of a linker histone H1 protein and the interaction of neighbouring nucleosomes through flexible N-termini histone tails. As a result, heterochromatin is transcriptionally silent. Euchromatin is the transcriptionally active, extended chromatin structure with single nucleosome units connected by short linker DNA fragments (Fig. 1.1) and is described as an array of 'beads-on-string' structures (~10 nm) [8].

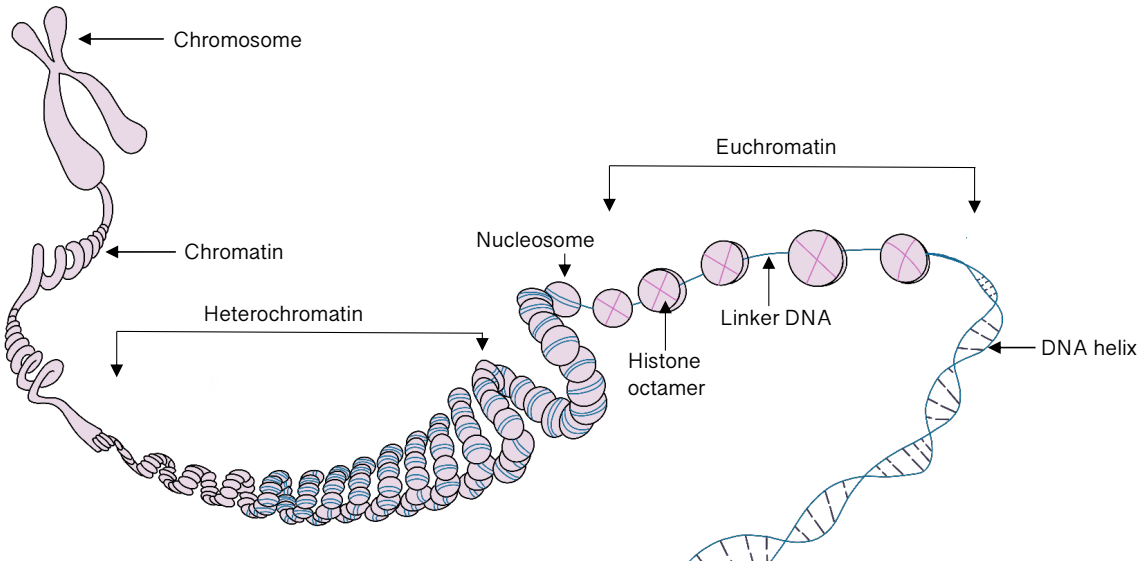


Figure 1.1 Eukaryotic gene expression

Schematic showing how chromosomes unravel to allow access to the genetic information in DNA. Chromatin consists of DNA tightly packed histone protein, forming compact structures known as heterochromatin. In contrast, euchromatin represents the more extended and transcriptionally active form, where DNA is less condensed, allowing access to regulatory proteins and the transcriptional machinery.

The initiation of eukaryotic gene expression is regulated by modifying local chromatin structure through histone modification and the activity of ATP-dependent chromatin remodelling enzymes [9,10]. Histone proteins are highly enriched in basic amino acid residues, particularly arginine and lysine. Post-translational modifications (PTMs) of histones alter their electrostatic charges, subsequently influencing their DNA-binding capacity and the overall chromatin structure. Histone acetylation and methylation are the most studied PTMs in transcription. However, other modifications, such as phosphorylation and ubiquitination also play a

significant role in chromatin organisation and gene expression [11,12]. Acetylated lysine residues, such as H3K9ac and H3K27ac, promote a relaxed and open chromatin structure and gene activation [13–15], Whereas lysine methylation, such as H3K9Me3, can facilitate the interaction of neighbouring nucleosomes to form the heterochromatin and recruit histone deacetylases and heterochromatin protein-1 (HP1), which promotes gene silencing [16–18]. However, methylation of these side chains does not always imply transcriptional repression, as H3K4Me, H3R2Me, and H4R3Me are also associated with gene activation [19–21]. These histone PTMs create an open euchromatin structure that exposes DNA sites recognised by specific effector protein domains to enhance gene activation [22]. For example, bromodomains recognise K-Ac sites, while chromodomains recognise KMe sites. Activator proteins with these domains bind to these sites to further remodel chromatin and open promoter regions to the binding of transcriptional factors and RNA polymerase [22–24].

1.2 RNA Polymerase II

RNA polymerases are central to the transcription of all genes. The activity of RNA polymerases was observed as early as 1959 in both prokaryotic and eukaryotic systems [25–29]. Roeder and Rutter *et al.,* along with Chambon *et al.,* subsequently showed that transcription of the eukaryotic genome is achieved by three specialised polymerases: I, II, and III [30–32]. RNA polymerase (Pol) I is a multiprotein complex comprising 14 subunits. It is located in the nucleolus of eukaryotic cells and is responsible for synthesising 47S pre-ribosomal RNA (rRNA), the precursor molecule

for 5.8S, 18S and 28S rRNA [31]. RNA polymerase (Pol) II and III consist of 12 and 17 subunits, respectively, and are located in the nucleoplasm [31]. Pol II synthesises a diverse range of RNAs, including messenger RNA (mRNA), most small nuclear RNA (snRNA), micro-RNA (miRNA), and long non-coding RNAs (lncRNAs), among others. Pol III synthesises transfer RNA (tRNA), 5S RNA, U6 snRNA, 7SK snRNA and 7 SL RNA [33,34]. The structures of all three polymerases are highly conserved, sharing several subunits. Notably, Rpb5, Rpb6, Rpb8, Rpb10 and Rpb12 are common to all three polymerase complexes [35].

The carboxyl-terminal domain (CTD) of the largest subunit of Pol II, Rpb1, consists of a tandemly repeated heptapeptide with the consensus sequence Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7 (Y1-S2-P3-T4-S5-P6-S7) [36,37]. This sequence is conserved in eukaryotes, and there is a correlation between the complexity of an organism and the number of repeats. For example, *Saccharomyces cerevisiae* contains 26 repeats, *Drosophila melanogaster* contains 44 repeats, and mammalian CTD contains 52 repeats [38]. The significance of these heptads varies across species. For instance, only 8 CTD heptads are required for cell viability in *S. cerevisiae*, while mouse cells require a minimum of 26 CTD heptads for survival [39,40]. Similarly, the importance of each residue is dependent on the species. For example, substituting Thr4 with valine in chicken DT40 or alanine in human cells results in cell lethality [41,42], whereas mutating Thr4 does not impact the viability of budding yeast [43]. On the other hand, mutation of Tyr1, Ser2 or Ser5 residues is lethal in budding yeast [44]. In mammalian CTD, 21 of the first 30 repeats at the N-terminal region strictly follow the consensus sequence, whereas the other 31 repeats, mainly at the C-terminal half, often present mutations at positions 2, 5 and 7 [45–48]. Tyr1 and Pro6 are present

in all repeats, while Ser7 is the least conserved residue and is present in only 26 repeats across the CTD. The C-terminus of the mammalian CTD also contains a conserved 10-residue sequence after the 52 repeats, ISPDDSDEEN, which stabilises the CTD [48,49].

1.3.1 Cyclin-dependent kinases and Pol II CTD phosphorylation

Cyclin-dependent kinases (CDKs) are a family of serine/threonine kinases initially discovered through their roles in regulating the cell cycle [50]. As the name suggests, CDKs are characterised by their dependence on a regulatory cyclin to perform their enzymatic functions. Since the discovery of the first CDK, 19 other CDKs have been identified with distinct roles in transcription and the cell cycle [50]. These CDKs can be widely categorised into cell cycle CDKs: CDKs 1-6 and CDKs 14-18, and transcriptional CDKs (tCDKs): CDKs 7-13, CDK19 and CDK20. Notably, some cyclins serve as regulatory subunits for multiple CDKs [51–53], underscoring the complexity of CDK activities (Table 1.1). Unlike cell cycle CDKs, the regulatory cyclins for tCDKs are constantly expressed and are not regulated by the cell cycle [52]. This study focuses on the activities of tCDKs.

Table 1.1 Human transcriptional cyclin-dependent kinases (tCDKs), their regulatory cyclins and their Pol II CTD target residue.
List of transcriptional CDKs, their alternative names, regulatory cyclin and enzymatic activities on Pol II CTD. The CTD target for CDK10 and CDK19 remains unknown. However, CDK19 is a paralog of CDK8. Table adapted from Zaborowska *et al*. [54].

CDKs	**Alternative name(s)**	**Regulatory Cyclin**	**Pol II CTD residue target**
CDK7	CAK, MO15, STK1	Cyclin H [55]	Ser5 and Ser7 [56–60]
CDK8	K35	Cyclin C [61]	Ser5 [62–64]
CDK9	CDC2L4, PITALRE,	Cyclin T [65]	Ser2, Thr4, Ser5 and Ser7 [57,63,66–69]
CDK10	PISSLRE	Cyclin M [70]	-
CDK11	CDC2L1, CDC2L2	Cyclin L [71]	Ser2 [72,73]
CDK12	CD2L7, CRKRS, CRK7,	Cyclin K [51]	Ser2, Ser5 [64,74–77]
CDK13	CHED, CDC2L5	Cyclin K [51]	Ser2, Ser5 [75,76,78–80]
CDK19	CDC2L6, CDKL8	Cyclin C [53]	-
CDK20	CCRK	Cyclin H [52]	-

CDKs regulate the expression of Pol II transcription through the reversible phosphorylation of the CTD, which subsequently helps to recruit factors required for different phases of the transcription cycle [54,81]. In mammals, cyclin-dependent kinases are the most prominent kinase family that regulate the phosphorylation of the CTD, and most tCDKs can phosphorylate the CTD [54]. Modification of the Pol II CTD is involved in regulating several steps during Pol II transcription, such as

chromatin modification [19,82–85], promoter escape [86,87], capping [88–90], transcript elongation [91–95], RNA 3' end processing [96–100], and termination [101,102]. All repeats except proline 6 (which can undergo isomerisation) are subject to phosphorylation and dephosphorylation modifications by several conserved kinases and phosphatases [54].

1.3.2 Employing analog-sensitive kinases to study CDK function

All CDKs contain key elements such as the G-loop, T-loop, and ATP-binding pocket (Figure 1.2). Despite the functional specialisation of CDKs over time, their ATP-binding sites are highly conserved. This conservation presents a challenge when investigating the roles of each kinase in cellular processes. For example, THZ1, a CDK7 inhibitor, has off-target effects on CDK12 and CDK13 when used at high concentrations [103]. Similarly, the available kinase inhibitors for CDK12 lack specificity due to the high degree of homology between the active sites of CDK12 and CDK13 [79].

Targeting CDKs indirectly, such as by inhibiting their regulatory cyclins, is also difficult to achieve, as some CDKs can function with more than one cyclin partner [65], and cyclins are not always exclusive to one CDK. For example, Cyclin K is the regulatory cyclin for both CDK12 and CDK13 [51]. The G-loop positions ATP for the transfer of the phosphate group [104,] and the T-loop (activation loop) regulates nucleotide and substrate binding pockets by adopting different conformations [105.] The C- helix stabilises the T-loop[106] (Figure 1.2). Most available commercial inhibitors are designed to target the ATP-binding site but often lack kinase selectivity

[52,107,108]. CDK degraders have been developed to target the G-loop of CDKs, but resistance to degradation has been observed in some kinases [109]. The T-loop contains a highly-conserved terminal Asp-Phe-Gly (DFG) peptide motif, making it challenging to design kinase-specific inhibitors [107]. To address this issue, I adopted the analog-sensitive kinase approach to specifically target kinases (See Chapter 3). This involves mutating the gatekeeper residue, which controls the capacity of the ATP binding pocket, to an amino acid with a smaller side chain, such as glycine or alanine. This enlarges the active site of the kinase, making it susceptible to competitive bulky ATP-analog inhibitors such as 1-NM-PP1 or 1-NA-PP1 (Chapter 3, Fig. 3.1A) [110,111]. Due to the specificity of the mutation, this method of inhibition is highly selective for the kinase of interest.

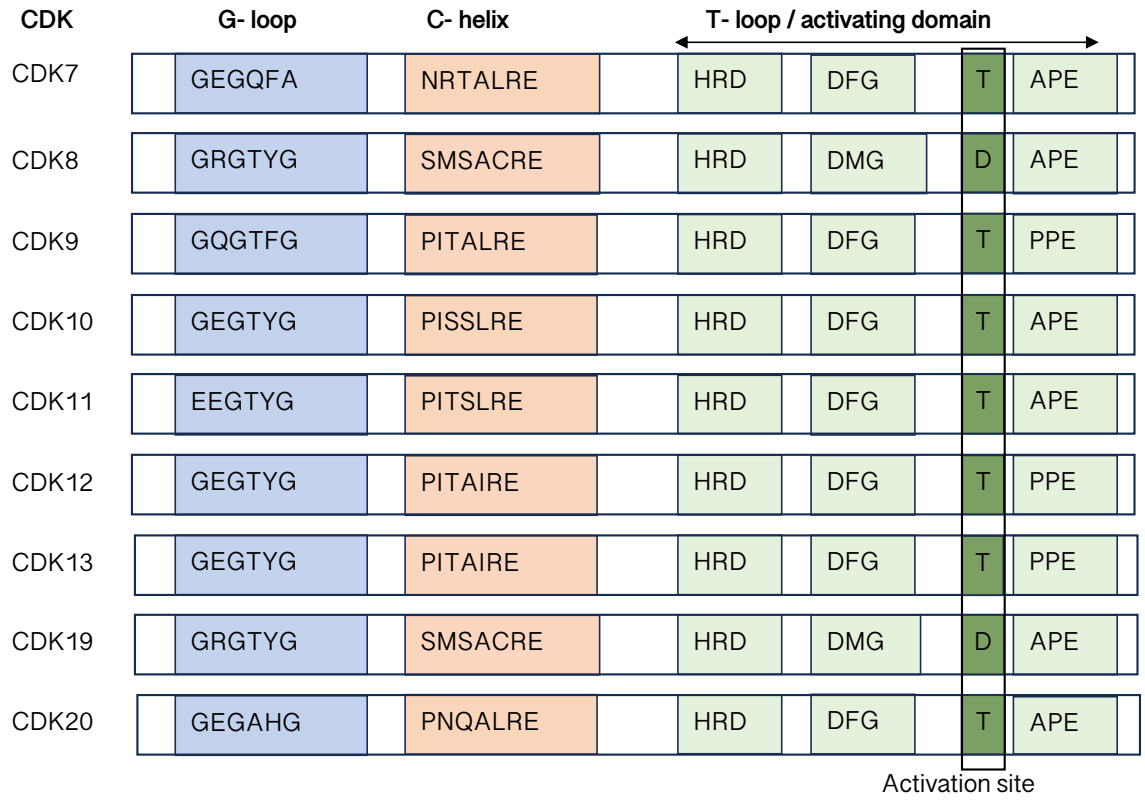


Figure 1.2 Schematic of the core Motifs of the human transcriptional cyclin-dependent kinases

Schematic of the primary structure of transcriptional CDKs showing the G-loop (ATP binding domain), C- helix (cyclin-binding domain) and the T-loop (activating domain). The T-loop is marked by the HRD motif at the start and the APE or PPE motif at the end. The DFG/ DMG motif is required for the structural regulation of the active site, and the activation residues are phosphorylation targets to activate respective kinases. Blue boxes represent the G-loop region; orange represents the C-helix motif; the T-loop motif is depicted in green, and the activation site is highlighted in darker green. Figure adapted from Lim *et al.*[81].

1.4 Expression of Protein-coding genes

1.4.1 Transcription initiation

The presence of a hypo-phosphorylated CTD tail marks the commencement of a new transcription cycle. This unphosphorylated Pol II requires the assembly of general transcription factors (GTFs) to bind to core promoters and form the pre-initiation complex. Core promoters have conserved consensus sequences and can be categorised into focused and dispersed promoters, which govern the pattern of initiation [112,113]. Focused promoters allow transcription initiation from a single site or a specific cluster of sites. Conversely, dispersed promoters initiate transcription from several sites spanning 50-100 nucleotides and are primarily observed in CpG islands. Indeed, approximately 60% of Pol II promoters are located in CpG islands [114,115]. Most focused core promoter motifs are recognised by Transcription factor II D (TFIID). The TFIIB recognition element (BRE) is recognised by TFIIB [116]. The TATA box and Initiator element (Inr) are the most common promoter motifs. The TATA box was the first promoter sequence to be discovered [117] and is similar to the prokaryotic Pribnow box. However, these sequences are not homologous [118]. It was named after

its consensus sequence, TATA(A/T)AA(G/A), although a limited number of human promoters contain this sequence [116,119,120].

TFIID can independently bind to the TATA box without the assistance of other TFs, as it contains the TATA-binding protein (TBP) and a core set of TBP-associated factors (TAFs). The binding of TFIID to the promoter induces a conformation in the local DNA, bending it by approximately 90° and causing it to unwind [121,122]. This alteration facilitates the recruitment of additional transcription factors (Table 1.2) that enable the formation of the pre-initiation complex. TFIIB binds to DNA to stabilise the TBP-promoter interaction and promotes the recruitment of the TFIIF to RNA polymerase to form a closed complex on the DNA. Following this, TFIIE is recruited, which is required for promoter melting and promotes the recruitment of TFIIH. TFIIH has helicase activities, which promote DNA melting and opening of the DNA duplex to form a transcriptional bubble called the open complex. Additionally, the CDK7 subunit of TFIIH phosphorylates the CTD during transcription initiation [118,123].

Table 1.2 RNA Polymerase II general transcription initiation factors and their functions

Factor	Function
TFIID	Binds the TATA box; initiates pre-initiation complex assembly, recruits TFIIB.[124,125]
TFIIB	Stabilises the TATA-TFIID complex, recruits Pol II-TFIIF complex [126].
TFIIF	Targets Pol II to the promoter; promotes elongation [127,128].
TFIIE	Facilitates ATP independent promoter melting; promoter clearance; recruits TFIIH [129].

TFIIH	Promoter melting; promoter clearance; helicase and kinase activities [129,130].

During transcription initiation, the Pol II CTD is largely unphosphorylated, which facilitates its interaction with the mediator complex [131–133], a 30-subunit complex comprising head, middle and tail subcomplexes, which plays a role in Pol II-mediated transcription initiation and elongation [58,134]. The head unit interacts with the CTD, the middle region controls the conformation of a mediator to ensure it maintains the correct orientation to bind DNA [135], and the tail interacts with regulatory transcriptional activators and repressors, acting as a bridge to relay information to the transcriptional complex [136–138]. The Mediator complex binds to Pol II and its associated transcription factors (Table 1.2) to complete the pre-initiation complex (PIC). Following PIC formation, the Mediator complex stimulates the activity of CDK7, which induces mediator release, facilitating promoter escape (see 1.4.2.1) and transcription initiation [86]. RNA polymerase produces burst synthesises of short ~15 nucleotides long RNA in a process known as abortive transcription [139,140]. The transcription bubble formed during this process is further unwound downstream to enable transcription. Following the synthesis of approximately seven nucleotides of the nascent RNA, the upstream end of the bubble is then rewound [141]. Structural studies have shown that this nascent RNA clashes with TFIIB, signalling the end of transcription initiation [142]. However, this interaction does not displace the TFIIB [143].

1.4.2 Transcription elongation

1.4.2.1 Promoter escape

The first step towards transcription elongation is an intermediary phase where RNA polymerase is released from its host of general transcription factors (GTFs) in a process known as promoter escape [144]. The rewinding of the bubble during abortive transcription leads to the dislodging of TFIID and TFIIB. However, TFIIE and TFIIH are still present. CDK7, the kinase component of TFIIH, phosphorylates CTD Ser5 to stimulate the loss of the mediator complex and facilitate promoter escape [58,60]. Phosphorylated Ser5 (pSer5) also promotes the recruitment of capping enzymes, which co-transcriptionally adds 7-methyl guanosine 'cap' to the 5' end of nascent RNA, subsequently protecting it from degradation [145,146]. The recruitment of elongation factors negative elongation factor (NELF) and DRB-sensitivity inducing factor (DSIF) as the nascent RNA is further synthesised promotes the eviction of TFIIE and TFIIH[147,148]. DSIF, a heterodimer of SPT4 and SPT5 subunits, associates with the Pol II complex through SPT5 [149] and NELF subunits (NELF A, B, C/D, and E) subsequently interact with Pol II by recognising the SPT5-Pol II interface [150]. The exact mechanism by which TFIIF is displaced is unknown. However, the loss of TFIIF coincides with the recruitment of other elongation factors [148].

1.4.2.2 Promoter-proximal pausing

The binding of NELF and DSIF facilitates Pol II promoter escape by causing a loss of GTFs, stabilises the transcriptional complex, and inhibits Pol II from further RNA synthesis [151. This forms] a regulatory early-elongation checkpoint involving promoter-

proximal pausing [152]. The paused Pol II complex, which occurs after synthesising approximately 25-50 RNA nucleotides [153–155], faces two possible outcomes: productive elongation or premature termination.

1.4.2.2.1 Productive elongation

The release from this checkpoint into productive elongation requires the activity of positive transcription elongator factor b (P-TEFb), which comprises CDK9 and cyclin T [150.] Elongation factors, such as PAF1 and SPT6, are also required for Pol II release into productive elongation [156,157]. CDK9 phosphorylates Ser2 of the CTD, the NELF-E subunit of NELF, the SPT5 subunit of DSIF and the RPB1 CTD linker [92,156,158,159]. Phosphorylation of NELF facilitates its dissociation from Pol II, which allows the recruitment of PAF1. The binding of PAF1 and NELF are mutually exclusive. Therefore, the recruitment of PAF may inhibit the rebinding of NELF [156]. The phosphorylation of SPT5 converts it from a negative to a positive elongation factor, and the phosphorylation of the CTD linker promotes the association of SPT6 with Pol II [92,156]. The phosphorylation of the CTD and association of PAF1 and SPT6 leads Pol II into productive elongation. SPT6 is a histone chaperone, facilitating pol II transcription through nucleosomes while maintaining chromatin structure [160]. The activity of CDK9 facilitates transcription but is also highly regulated. CDK9 primarily exists in an inactive form, sequestered within the 7SK snRNP complex, while the active form is found mainly in the super elongation complex (SEC)[94,161,162]. SEC is a multi-protein complex comprising AFF1, AFF4, AF9, ELL/EAF and CDK9 subunits [163]. SEC subunits promote transcription elongation and interact with multiple factors that facilitate the recruitment of CDK9 to the paused transcriptional complex [164],

including the Med26 subunit of the mediator complex. The mediator complex is proposed to function as a transcriptional regulator, which facilitates the loss of TFIID and provides a docking site for the SEC complex.

1.4.2.2.2 Premature termination

The failure to progress into productive elongation leads to premature termination of transcription. This process is mediated by a multisubunit complex named Integrator -Protein phosphatase 2A, INTAC) [165,166]. The Integrator complex (Integrator) is a conserved machinery comprising at least 15 subunits (IntS 1-15). The first 12 subunits were initially discovered through their role in coupling the pol II CTD to the 3' end processing of U1 and U2 snRNA gene transcripts [167,168]. IntS11, the endonuclease subunit, and its binding partner IntS9 are paralogs of the cleavage and polyadenylation specificity factor-73 (CPSF73) and CPSF100, respectively. CPSF73 and CPSF100 are the catalytic subunits of the CPSF cleavage module, which co-transcriptionally cleaves mRNA at the 3' end of the gene [167,169]. The activity of the Integrator was initially thought to be specific to snRNA genes. However, it has been demonstrated to play diverse roles in the transcription of other Pol II-transcribed (or protein-coding) genes [166,170,171]. Integrator interacts with Pol II, DSIF and NELF and recruits phosphatase, PP2A, to form the INTAC complex [165,172,173]. The presence of the INTAC complex prevents the association of PAF and SPT6 with Pol II [172]. Following this, the Integrator positions PP2A to dephosphorylate the Pol II CTD and DSIF; the nascents are then cleaved using the endonuclease activity of the Integrator and degraded by Xrn2 [174], consequently inhibiting gene transcription, productive elongation, and terminating the transcription cycle [166,170,175,176].

1.4.3 Transcription termination

At the end of the RNA synthesis cycle, pre-mRNA must be cleaved to release it from Pol II and the DNA template, facilitating downstream export and resetting Pol II for another transcription cycle. For protein-coding genes that encode polyadenylated mRNA, this termination process is initiated by the recognition of two sequences which flank the cleavage site – the upstream Polyadenylation signal (PAS), AAUAAA, recognised by the cleavage and polyadenylation specificity factor (CPSF), and the GU-rich downstream sequence element (DSE) recognised by the cleavage stimulatory factor (CstF) [177,178]. CPSF subunits, CPSF160, CPSF30, WDR33, and FIP1, mediate the PAS recognition, while CPSF100 and CPSF73, in association with symplexin, cleavage factor I (CFI) and cleavage factor II (CFII11), mediate mRNA cleavage [179,180]. Poly(A) polymerase is also recruited to facilitate cleavage, after which it catalyses the addition of poly (A) tails (a stretch of adenosine nucleotides) to the cleaved 3'-OH end [181]. The length of the poly (A) tail is crucial to its stability and essential for its export into the cytoplasm [182,183]. Following initial polyadenylation, poly(A) binding proteins further extend and regulate the length of the poly (A) tail. A notable outlier to this termination mechanism is the replication-dependent histone mRNAs, which are not polyadenylated [184]. Thus, a different 3' end machinery involving the stem-loop binding protein (SLBP) is required for CPSF73 to 3' process these mRNAs (see section 1.6.1.2)

mRNA processing is coupled to the transcription termination of protein-coding genes [185]. Following mRNA cleavage, transcription termination involves the torpedo model pathway. The torpedo model suggests that transcription past the PAS facilitates the phosphorylation of Ser2 by CDK9 [185,186]. This enables the recruitment

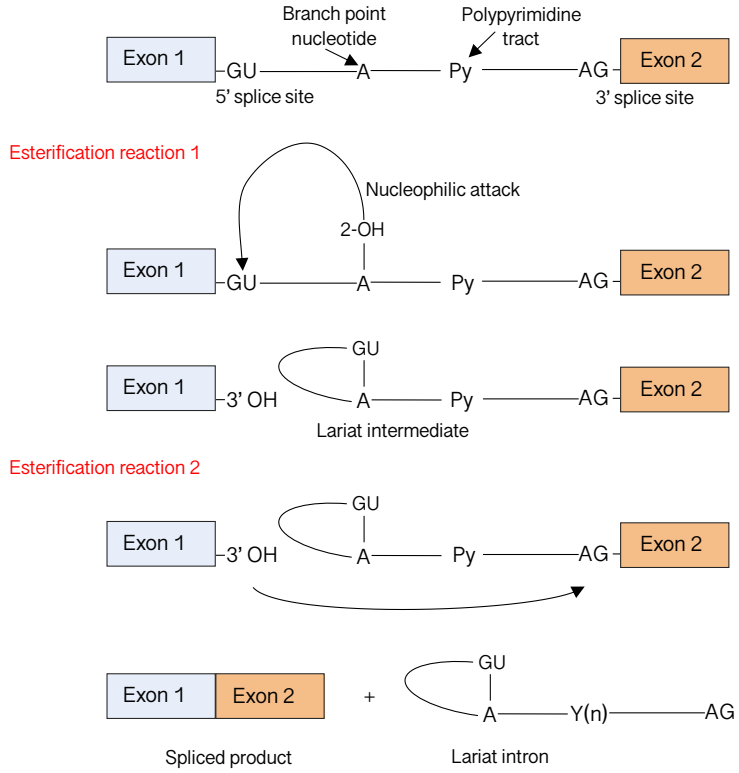
of termination factors, such as PCF11, which interacts with phosphorylated Ser2 (Ser2P) and disassembles the elongation complex [187–189]. The association of PCF11 consequently reduces Pol II processivity, allowing CPSF73 to cleave the nascent RNA at the poly(A)site [169,190]. The downstream RNA remains attached to the transcription machinery, is no longer protected by a cap, and is degraded by Xrn2, a 5' to 3' exoribonuclease, promoting termination[191–194]. This model also accounts for anti-termination factors, such as SR-related CTD-associated factor (SCAF) 4 and SCAF8, which bind to Ser2P and Ser5P and are redundant in preventing the use of early polyadenylation sites[195].

1.5 Pre-mRNA splicing

Splicing, the excision of introns from pre-mRNA, is a major co-transcriptional RNA processing event alongside capping and polyadenylation [196,197]. This process is not only required for the accurate translation of mRNA but can also increase transcript diversity from a single mRNA through alternative splicing [198]. The fundamental chemistry of splicing is two transesterification reactions, where the number of phosphodiester bonds remains unchanged (Fig. 1.3A). Thus, the primary mechanism does not require ATP hydrolysis or an energy source to occur. However, the process is catalysed by a multi-megadalton spliceosome complex, which requires ATP [199]. The spliceosome complex consists of five small nuclear RNAs (snRNAs), which are associated with other proteins in their complexes and form small nuclear ribonucleoproteins (snRNPs). snRNPs form the building blocks of the spliceosome and assemble with RNA splicing factors and ~ 200 proteins to form the

spliceosome complex [200]. The pre-mRNA precursor contains consensus sequences on the 5' splice site (SS) and exon-intron 3' end SS, which provide the signal and docking site for the recruitment of the spliceosome complex. The 5' end of the intron generally contains the GU/AU sequence and the AG/AC sequence at the 3' end. Additionally, introns contain an internal branch site located upstream of their pyrimidine-rich region (polypyrimidine tract) at their 3' end, which is required for splicing. Assembly of the spliceosome complex is dependent on the sequence present at the 5' and 3' SS. The GU-AG sequence promotes the assembly of the major spliceosome, while an AU-AC sequence facilitates the assembly of the minor spliceosome [201–204]. Assembly of the major spliceosome complex occurs in a stepwise process [205]. Spliceosome assembly is initiated by the association of factors at the 5'SS, branch point, and 3'SS. U1 snRNP and splicing regulatory (SR) proteins bind to the 5' splice site, the branch-point binding protein (BBP) binds to the branch point, and U2AF65-U2AF35 heterodimer binds to the polypyrimidine tract and 3' splice site, respectively. The SR proteins interacts with U2AF to bridge these complexes, which form the E complex, and the base-pair association of the U2 snRNP with the branchpoint forms the A complex. Following this, the U4/U6 and U5 tri-complex bind to form the mature spliceosome which undergoes conformational changes, called the B complex. Firstly, U1 and U4 snRNPs are released. This allows U6 snRNP to interact with the 5'SS and bind directly to the U2 snRNP to form the catalytically active spliceosome, and both transesterification reactions occur in the C complex (Fig. 1.3B).

A


Branch point nucleotide
Polypyrimidine tract
Exon 1
GU
A
Py
AG
Exon 2
5' splice site
3' splice site
Esterification reaction 1
Nucleophilic attack
2-OH
3' OH
Lariat intermediate
Esterification reaction 2
Spliced product
Lariat intron
Y(n)

B

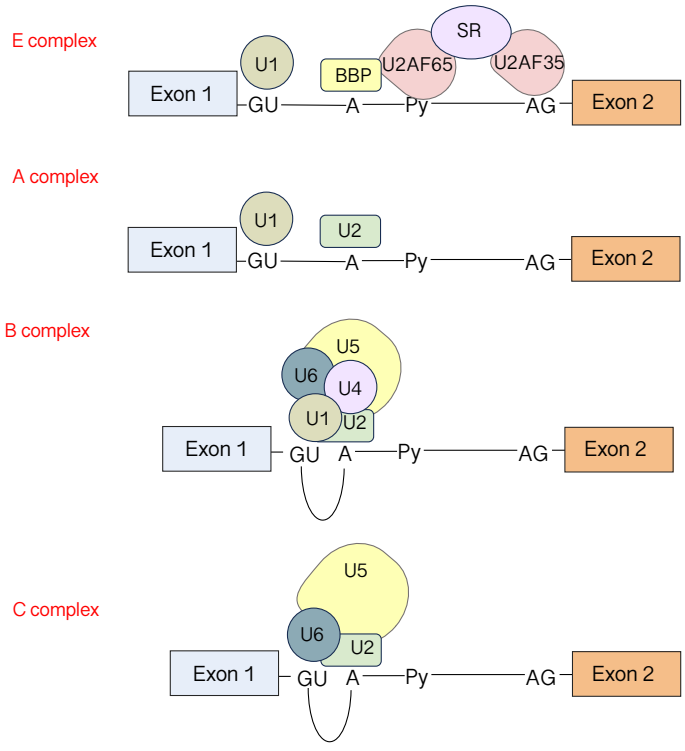

E complex
SR
U1
BBP
U2AF65
U2AF35
Exon 1
GU
A
Py
AG
Exon 2
A complex
U2
B complex
U5
U6
U4
C complex

Figure 1.3 Spliceosome assembly and mechanism
A. Splicing occurs by two esterification reactions. In the first step, A nucleophilic hydroxyl group of the branch point nucleotide initiates a nucleophilic attack on the 5' splice site, leading to the cleavage of the first exon and the formation of a lariat intermediate. In the second step, the free hydroxyl group of the released exon performs a nucleophilic attack on the 3' splice site, joining the two exons and excising the lariat intron. **B.** Assembly of the spliceosome complex on pre-mRNA occurs in a stepwise manner. The E complex is formed through the association of the U1 snRNP with the 5' splice site, U2AF proteins binding to the polypyrimidine site and 3' splice sites, and the interaction of BBP and SR proteins. U2 snRNP binds to the branchpoint to form the A complex. U4 and U6 snRNP subsequently bind to form the B complex, and U1 and U4 snRNPs are lost during activation, which forms the catalytic C complex.

snRNPs are not only essential for mRNA synthesis through their roles in the spliceosome, but splicing is also required for efficient mRNA 3' end processing [206]. This highlights the importance of snRNPs in the overall expression of mRNA genes. Furthermore, it emphasizes how the roles of other RNAs, such as snRNAs, facilitate the expression of protein-coding genes.

1.6 snRNA gene expression: from transcription to functional snRNPs.

The uridine-rich (U) snRNAs are a class of non-coding RNA transcribed in the nucleus by RNA polymerase II, with the exceptions of U6, U6atac and 7SK snRNAs, which are transcribed by RNA polymerase III. snRNAs are functional units of the spliceosome required to splice pre-mRNAs and are highly expressed within the cell [207]. For instance, the U1 snRNA genes are loosely clustered and repeated 4 times on human chromosome 1, and the U2 snRNA genes are tandemly repeated 10 - 20 times within an approximately 6.0 Kb repeat unit on human chromosome 17 [208–210]. Pol II- transcribed snRNAs undergo the co-transcriptional addition of an N^7-methyl guanosine cap (m7G) to the 5' end, which signals and is required for snRNA nuclear export upon cleavage at the 3' end [211–213]. The m7G cap binds to the nuclear cap-binding complex (CBC), facilitating the interaction of phosphorylated adaptor for RNA export (PHAX) [214]. Phosphorylated PHAX acts as an adaptor to bind to nuclear export factors CRM1, XPO1, and RanGTP to form the export complex [214]. snRNA is then exported into the cytoplasm, where PHAX is dephosphorylated, and RanGTP is hydrolysed to form RanGDP, subsequently causing the dissociation of the export complex. The SMN complex then facilitates the assembly of Sm proteins, which are assembled onto the snRNA in a stepwise manner to form the specific snRNP core domain [215–218]. The formation of the core domain initiates snRNP biogenesis and is required for the hypermethylation of the m7G cap, resulting in the formation of the $N^{2,2,7}$-trimethyl guanosine (m3G) cap [219,220]. The core domain and m3G cap then facilitate the snRNA 3' end trimming by Target of EGR1 (TOE1) exonuclease [221,222] and import of the snRNP back into the nucleus by binding to importin β [223]. The presence

of the m3G cap ensures the nuclear retention of newly imported snRNPs, after which coilin targets them to Cajal bodies, where they are processed and then distributed within the nucleus to play specific roles [224,225] (Fig. 1.4).

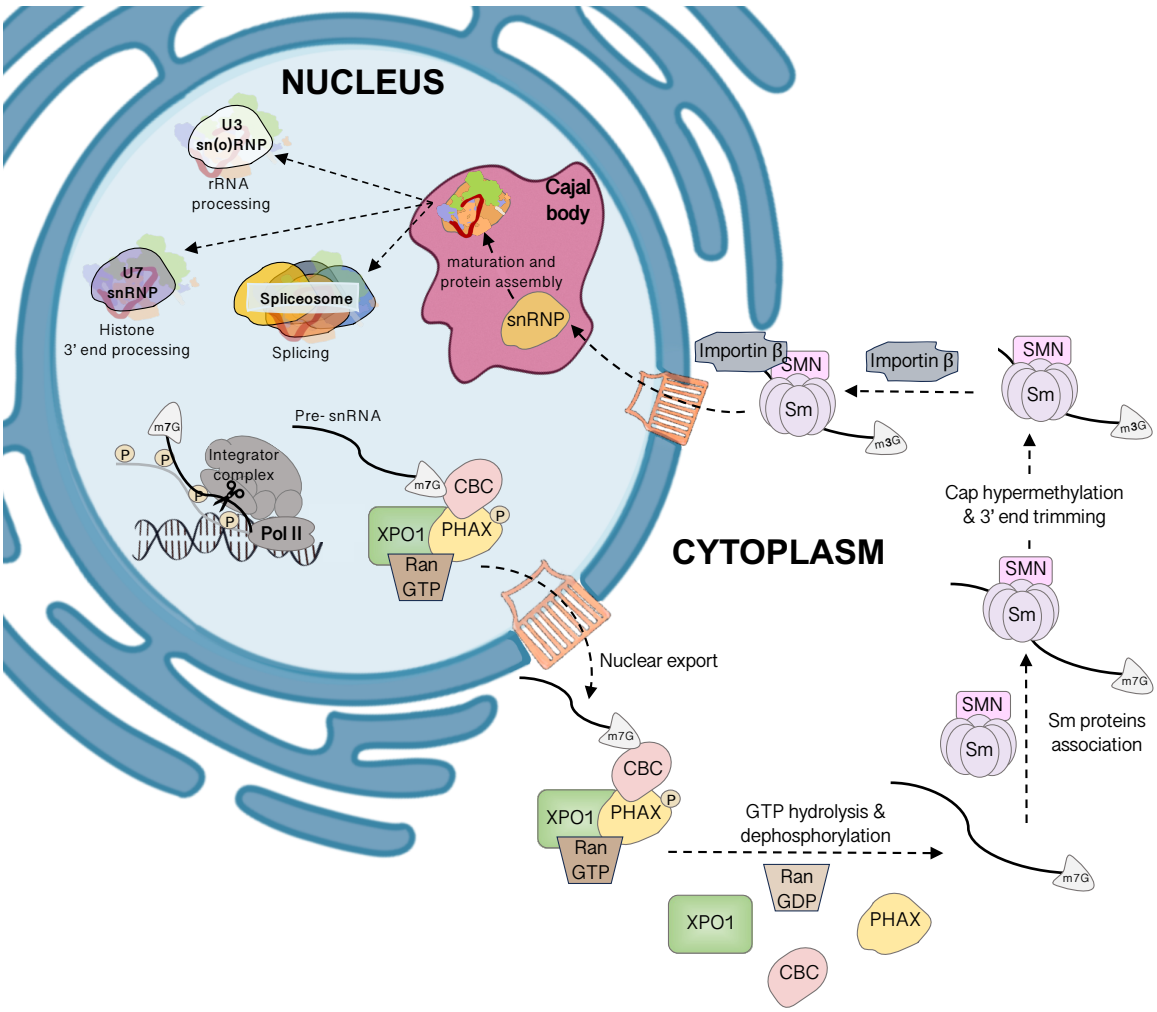


Figure 1.4 Schematic pathway of snRNA gene expression, snRNA maturation, and assembly into functional snRNPs.

The snRNA gene is transcribed by its RNA polymerase II, co-transcriptionally capped and cleaved by the Integrator complex to generate the pre-snRNA transcript. The m7G cap interacts with CBC, phosphorylated PHAX, XPO1 and RanGTP to form the export complex and is exported into the cytoplasm. PHAX is dephosphorylated in the cytoplasm, and RanGTP is hydrolysed to form RANGDP, facilitating the disassembly of the export complex. snRNAs then interact with Sm proteins to start snRNP complex formation. snRNAs then undergo cap hypermethylation and 3' end trimming before reimporting into the nucleus through association with Importin β. snRNPs are transported into Cajal bodies, where they undergo further maturation, interact with functional proteins and function in various cellular processes.

Abbreviations: CBC: cap-binding complex, PHAX: phosphorylated adaptor for RNA export, XPO1: Exportin1, SMN: survival motor neuron.

1.6.1 snRNPs mediate various cellular processes

While most snRNPs are components of the spliceosome complex, facilitating pre-snRNA splicing (see section 1.5), not all snRNPs function in this capacity. For example, U3 snRNPs are essential for rRNA processing and ribosome biogenesis [226], and U7 snRNPs are involved in the 3' end formation of histones mRNA [227–229].

1.6.1.1 U3 snRNPs are required for pre-rRNA processing

The rRNA precursors for the 60S and 40S eukaryotic ribosomes are synthesised in the nucleolus by RNA Pol I [33]. The synthesised 35S pre-rRNA transcript contains the 28S and 5.8S rRNA, two of the three rRNA components of the 60S ribosomal subunit, and the 18S rRNA for the 40S ribosomal subunit [230]. The third rRNA component of the 60S subunit, 5S rRNA, is synthesised by RNA Pol III [34]. The 35S primary rRNA transcript contains four transcribed spacer units which separate rRNA precursor transcripts (Fig. 1.5) [230]. The 5' external transcribed spacer (ETS) is located upstream of the 18S rRNA, two internal transcribed spacers, ITS 1 and ITS2, flank the 5.8S rRNA, and the 3' ETS following the 28S rRNA [231] (Fig 1.5A). The maturation of the 35S rRNA involves a multistep co-transcriptional cleavage process of the transcribed spacer units to yield 28S, 5.8S, and 18S rRNA [232]. This process begins with the association of various non-ribosomal proteins, endonucleases, exonucleases and small nucleolar RNPs (snoRNPs) [231,233–235]. The assembly of these factors occurs at the 5' ETS of the 35S rRNA to form a macromolecular complex called the small subunit (SSU) processome [233,235].

snoRNAs, the core components of snoRNPS, are a class of non-coding RNAs with two major functions: modifications of rRNA and pre-rRNA cleavage [236]. These modifications include 2'-O- ribose methylation, directed by C/D box snoRNAs [237,238], and pseudouridylation, which is directed by H/ACA box snoRNAs [239,240]. Together, these processes are critical for the maturation and biogenesis of ribosomes. The U3 snRNA contains the C/D box motif, which localises it to the nucleolus [241] where it functions as a snoRNP required for ribosome biogenesis [242]. During the formation of the SSU, processome complex, the U3 snoRNP base-pairs to the 5' ETS, and the 18S rRNA providing a guide for the cleavage of the 35S rRNA, and a platform for the interaction of other SSU processome proteins [243,244]. Following the SSU processome complex formation, the 35S RNA is cleaved at the A0, A1 and A2 positions, leading to the formation of the 20S rRNA [245]. The 20S RNA, as part of the ribosome, is subsequently exported to the cytoplasm, where it undergoes further maturation [245,246] and is further cleaved into 18S rRNA [247,248], the precursor to the 40S ribosome. The cleaved 5' ETS is degraded by the nuclear exosomes and Xrn2 [249,250].

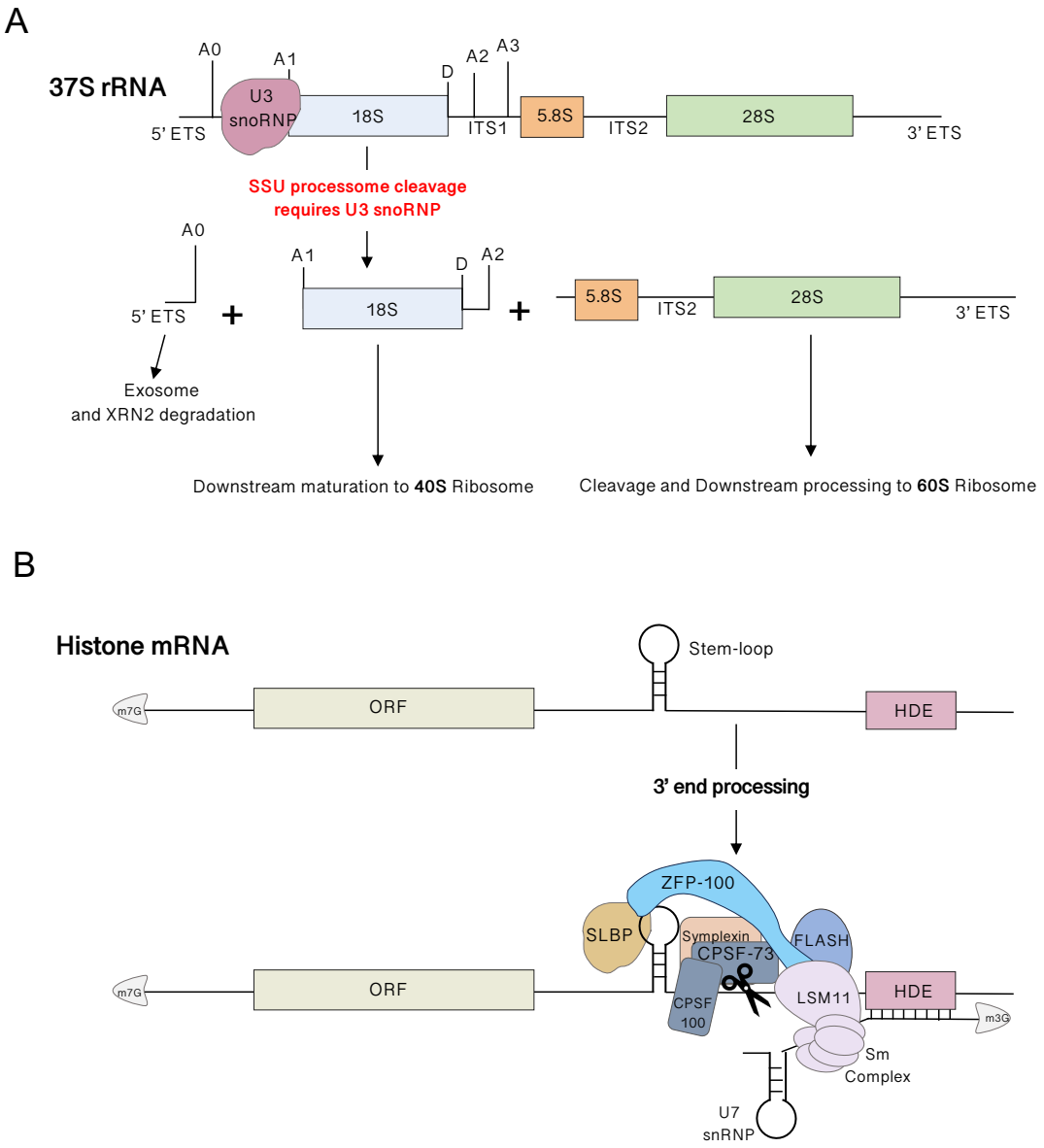


Figure 1.5. snRNPs mediate ribosome biogenesis and histone mRNA 3' end formation.

A. Schematic showing the 37S rRNA, the precursor to the 60S and 40S eukaryotic ribosome subunits. The rRNA encodes the 18S rRNA precursor to the 40S, and 5.8 and 28S rRNA, precursors to the 60S ribosome. These rRNA sequences are interspersed with external and internal transcribed spacers (ETS and ITS, respectively), that are cleaved out during maturation. Cleavage of the 5' ETS, which releases the 18S rRNA, requires the formation of a small subunit processome complex. U3 snoRNP is an integral part of the assembly and function this complex by base pairing with the 18S rRNA and 5' ETS, facilitating their cleavage **B.** Histone mRNA contains a stem-loop structure and HDE at the 3' end as opposed to a poly(A) tail. U7 snRNP binds to the HDE and is stabilised by FLASH and SLBP. The LSM11

SM protein binds to FLASH, while ZFP100 bridges the interaction between SLBP and U7 snRNP. Following the formation of the SLBP-FLASH-LSM100 complex, histone mRNA processing factors are recruited, and CPSF73 cleaves the mRNA. Abbreviations: HDE: histone downstream element, SLBP: Stem-loop binding protein, FLASH: FLICE-associated huge protein.

1.6.1.2 U7 snRNPs are required for the 3' end processing of histone mRNAs.

As earlier described, histones are a primary component of chromatin and play a key role in regulating gene expression (see section 1.1). Replication-dependent histone mRNAs encode for these histones and are highly expressed in the S-phase to package and store newly synthesised DNA [251,252]. Histone mRNAs are not normally polyadenylated. Instead, their 3' end contains a highly conserved stem-loop structure [253] (Fig. 1.5). Therefore, the 3' end formation of histone mRNAs requires an additional set of factors, including the stem-loop binding protein (SLBP), ZFP100 and U7 snRNPs [254]. SLBP contains an RNA-binding motif which binds to the stem-loop sequence [255,256], while U7 snRNP binds to the histone downstream element (HDE) [228,229]. ZFP100, a histone hairpin binding protein, interacts with both SLBP and U7 snRNP, bridging both complexes [257] (Fig. 1.5B) and so stabilising the u7 snRNP complex [258]. FLICE-associated huge protein (FLASH) also interacts with and stabilises the U7 snRNP [259]. The U7 snRNA contains a non-consensus SM-binding site, which ensures the recruitment of slightly different SM proteins. This different SM binding site recruits LSM10 and LSM11 SM proteins [259,260] as opposed to SMD1 and SMD2 SM proteins, which are directed to spliceosomal snRNAs [261]. The LSM11 has larger structure compared to other SM proteins and contains an extended N-terminal region, which is required for histone mRNA processing [260,262] and is

proposed to provide a platform for association of histone-mRNA-specific processing factors [260,262]. U7 snRNA also interacts with FLASH using this extended LSM11 N-terminus [263]. Following the interaction of SLBP with the U7 snRNP, the LSM11-FLASH-SLBP complex leads to the recruitment of CPSF73, CSF100 and symplexin, which cleaves histone mRNA (Fig. 1.5B).

1.7 snRNA gene expression

snRNA genes are generally regarded as a model system for transcription due to their small size and simplicity (Fig. 1.6). They have a more compact structure than most protein-coding genes and have a simple specialised promoter consisting of two elements: the distal sequence element (DSE), which acts as an enhancer and the proximal sequence element (PSE), which functions as the core promoter [264,265]. In addition, snRNA transcripts are intronless, non-polyadenylated, and 3' end formation is directed by a 3' box rather than a poly-A site (Fig. 1.6) [266]. Furthermore, the 3' box is only recognised if transcription is initiated by an snRNA-specific promoter, indicating an obligate coupling of transcription initiation and formation of the 3’ end of pre-snRNA [267,268].

1.7.1 snRNA gene transcription initiation

The transcription of snRNA genes follows the conserved sequential order of initiation, elongation and termination. For the U2 gene, transcription is initiated when Oct1, Sp1 and STAF bind to sequences in the DSE. Oct-1 stabilises the binding of basal transcription factor, PSE transcription factor (PTF), to the PSE. PTF, also

known as PSE-binding protein (PBP) [269] or snRNA activating protein complex (SNAPc) [270], is a multiprotein complex consisting of PTFα (SNAP190), PTFβ (SNAP50), PTFδ (SNAP45), PTFγ (SNAP43) and SNAP19 [264]. The PTF complex is essential in forming the preinitiation complex and regulating transcription through its interaction with various activators and repressors. For example, PTFα and PTFγ interact with the p53 tumour repressor protein[265]. Also, the interaction of PTFα with the activator protein, Oct-1, promotes stable recruitment of PTF to the PSE and formation of the pre-initiation complex, which contains TFIIA, TFIIB, TFIIE, TFIIF and TFIIH, a specialised TBP/TAF complex termed snTAFc and, finally, Pol II [265,271–273].

Unlike protein-coding genes, TAFs 1-4, TAF10 and TAF12 are not required to initiate U2 snRNA gene expression. Instead, TAF5, TAF6, TAF8, TAF9, TAF11 and TAF13 interact with the TBP to form a distinct snRNA TAF complex (snTAFc) on the U2 snRNA gene [273]. Interestingly, the TFIID subunit TAF7, which plays a role in promoter escape and recruitment of the mediator complex, is not found on the U2 snRNA gene. TAF7 is, however, present on the U1, U4, U5, and U11 snRNA genes, indicating that the U2 genes may use an alternative mechanism for mediator recruitment [274,275].

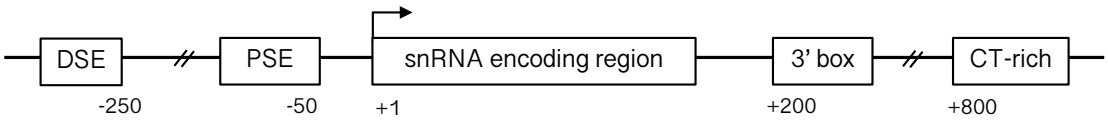


Figure 1.6 Schematic structure of the human Pol II transcribed snRNA gene showing cis-elements which regulate snRNA expression.
Structure of the human Pol II- transcribed snRNA gene showing the cis-acting elements required for the expression of a functional snRNA transcript and their location relative to the transcription start site (TSS). The distal sequence element (DSE) functions as a transcription enhancer; the Proximal Sequence Element (PSE) is the core promoter regulatory element which recognises the snRNA activator protein complex; the 3' box regulates the 3' end processing of the snRNA transcript

and the CT-rich region which provides a binding platform for the CTCF protein. The TATA box, which directs pol III recruitment, is absent in pol II transcribed snRNA genes. The black arrow represents the TSS.

1.7.1.2 Polymerase recruitment

Human pol III-transcribed snRNA genes, such as U6 snRNA and 7SK snRNA, contain an additional cis-element, the TATA box, which directs the association of a different set of factors to recruit pol III [276,277]. In contrast, the TATA box is absent in most human snRNA genes, which paradoxically still allows RNA pol II recruitment. However, the presence of a TATA box is not the key determinant of polymerase specificity in all species. *A. thaliana* U6 and U2 snRNA genes both contain the TATA box yet recruit different polymerases (Fig. 1.7). Indeed, the polymerase specificity in *A. thaliana* is determined by the distance between the TATA box and the upstream sequence element [278]. Similarly, in *D. melanogaster, B. mori, A. gambiae, and A. mellifera snRNA* genes, a variation in the PSEA sequence determines the specificity of polymerase recruitment [279] (Fig. 1.7).

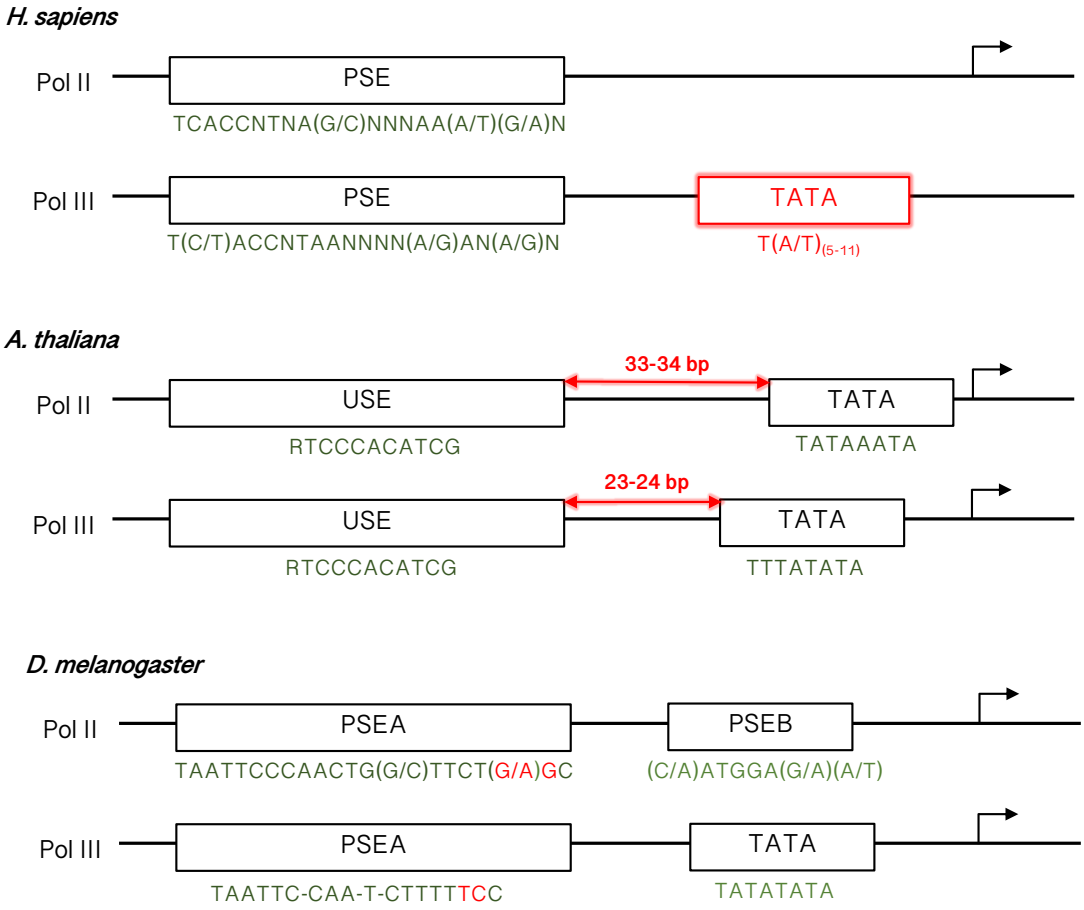


Figure 1.7 Different factors govern polymerase recruitment to snRNA genes
Organization of the genomic elements that determine RNA polymerase recruitment in *Homo sapiens, Arabidopsis thaliana* and *Drosophila melanogaster* snRNA genes. PSE(A) and USE are cis-promoter elements required for transcription. The determinants of RNA polymerase II/III recruitment are highlighted in red, and the consensus sequence of each element is in green, below its respective element. Numbers and letters under boxes represent the distance from TSS or consensus sequence. Abbreviations: TSS; transcription start site, DSE: Distal sequence element PSE; Proximal sequence element, USE; Upstream sequence element.

TBP is recruited to human snRNA genes regardless of the presence of a TATA box and is part of the PIC for both pol II and pol II-transcribed snRNA genes. TBP then recruits TFIIB or BRF2 (TFIIB-related factor 2) and its GTFs to pol II or pol III genes, respectively [272,280,281]. TFIIB and BRF2 are structural homologs with similar core

domains and a zinc domain for interaction with their respective RNA polymerases [282]. TFIIB and BRF2 also bind to the PSE-PTF complex via protein-protein interactions (with PTF), and both have similar binding pockets. TFIIA is then recruited to pol II-dependent genes to stabilise the TBP-TFIIB complex on pol II snRNA genes and promote TFIIB-specific recruitment by acting as a competitive inhibitor to BRF2. TFIIA can also be recruited to pol III-transcribed snRNA genes [283]. However, the binding of TFIIA to TBP structurally restricts the binding of BRF2 to TBP [284]. Likewise, BDP1, a pol III-specific transcription factor, stabilises the BRF2-PTF-TBP complex to ensure pol III recruitment. TBP, TFIIA, TFIIB, TFIIF, TFIIE, TFIIH and Pol II form the pre-initiation complex on Pol II-dependent snRNA genes. In contrast, TBP, BDP1, BRF2 and the TATA are sufficient to form the pre-initiation complex on pol III-dependent snRNA genes [285].

1.7.2 snRNA transcription elongation

1.7.2.1 The Little elongation complex

The Little elongation complex (LEC) comprises ELL, EAF, ICE1 and ICE2 core subunits, which are associated with ZC3H8 and USPL1 proteins [286]. The ELL and EAF subunits are shared with the SEC of protein-coding genes. The ICE1 and ICE2 subunits are snRNA-gene specific for transcription. However, the ICE1 subunit has been recently shown to interact with the exon junction complex on pre-mRNA [287]. The discovery of the SEC and LEC subunits stemmed from the identification of ELL as an elongation factor on protein-coding genes [288,289]. EAF was subsequently discovered as the ELL binding partner [290], and two separate elongation complexes,

SEC [163] and LEC [286], were later characterised for their roles on protein-coding [163,291,292] and snRNA genes [286,293], respectively. ICE1 is critical for the formation of the LEC complex and the recruitment and stabilisation of Pol II to snRNA genes, and the knockdown of ICE1 results in the loss of Pol II, ELL, and ICE2 from the U11 and U12 snRNA gene promoters [293]. Similarly, ELL is a key factor for Pol II elongation on snRNA genes. While knockdown of ELL does not impair Pol II recruitment, it prevents Pol II from elongating, resulting in its accumulation at the 5' end of the gene[293]. ICE1 has also been shown to interact with snRNA 3' end processing and termination factors such as Integrator complex subunits and the CBC-Arsenic resistance protein 2 (CBCA) complex [294]. This suggests that the role of LEC may extend beyond elongation, potentially influencing 3' box-dependent RNA processing and termination of transcription of snRNA genes. The role of ICE2 in the expression of snRNA genes remains unknown. However, it is related to protein families with RNA 5' end exonuclease activity [295]. Although ICE2 lacks the motif that confers exonuclease activity, it was proposed that the molecular function of ICE2 may involve binding to the 5' end of RNA[295].

1.7.2.2 The initiation to elongation checkpoint on snRNA genes

The 30-subunit mediator complex functions as a transcriptional hub due to its ability to interact with various initiation and elongation transcription factors [136,137,274,296–299]. Indeed, the highly conserved NTD of Med26 shares a common ARM repeat with protein-coding elongation factors SPT6, IWS1, and Elongin A [299,300]. Med26 recruits the LEC complex to the snRNA genes by binding to the EAF subunit of LEC with its NTD domain [274,300]. Interestingly, the sequence through which EAF

interacts with Med26 is highly conserved and shares similarities with the sequence used by TAF7 to interact with Med26. However, these factors do not interact with Med26 in a mutually exclusive manner [300]. TAF7 has been shown to inhibit the interaction of EAF with the mediator complex *in vitro and* repress the recruitment of the LEC subunits to the snRNA genes [300]. Notably, the knockdown of TAF7 causes an increase in snRNA gene expression [300]. This led to a model where Med26 acts as a bridge between transcription initiation and elongation [300]. Essentially, TAF7 binds to Med26 to prevent premature transition into elongation, after which TAF7 releases Med26, allowing the interaction of EAF and LEC and productive elongation[300]. The depletion of Med26 causes a loss of ICE1, the LEC scaffold, and the loss of transcription of snRNA genes [300]. The EAF subunit is common to both LEC and SEC complexes. ENL, a SEC subunit, binds to the enhancers and promoter of protein-coding genes and is required for SEC recruitment [301]. Therefore, it is thought that ENL, in addition to Med26, directs the recruitment of SEC [294]. Similarly, it was proposed that a subunit of the LEC complex or an snRNA transcription factor is required to direct the specific recruitment of LEC in association with Med26 [294].

1.7.3 Regulation of snRNA gene expression by cyclin-dependent kinases.

During the expression of snRNA genes, CDK7, as part of the TFIIH complex, phosphorylates Ser5 and Ser7 CTD residues at transcription initiation [56,97,100]. Phosphorylation of Ser5 likely facilitates the recruitment of capping enzymes to cap

the nascent snRNA, similar to the process observed during the transcription of protein-coding genes [88,146,302].

1.7.3.1 snRNA 3' end processing

Upon CDK7 activity, RNA Pol II-associated protein 2 (RPAP2), a Ser5 phosphatase, binds to Ser7P and dephosphorylates Ser5P [99]. RPAP2 also co-recruits specific Integrator subunits, including IntS1, IntS4, IntS5, IntS6, and IntS7, via protein-protein interactions (Fig. 1.8). Notably, RPAP2 does not interact with nor co-recruit the endonuclease subunit IntS11 or its binding partner IntS9. However, these non-catalytic subunits are essential for IntS11 recruitment, as RPAP2 knockdown results in the loss of IntS11 [99]. At the 3' end of the gene, CDK9 phosphorylates Ser2 [54,303] and the presence of a phosphorylated Ser7 and an adjacent phosphorylated Ser2 facilitates the recruitment of IntS11 and IntS9, which are required for cleavage of nascent RNA upstream from the 3' box [96] (Fig. 1.8). Therefore, in addition to the requirement of an snRNA-specific promoter [267,268], CDK9 activity is required for the 3' end processing of nascent snRNA[303,304]. This model of snRNA 3' end processing indicates that the Integrator complex is recruited in a stepwise manner. However, beyond the subunits discussed here, it remains unclear how the other noncatalytic Integrator subunits are recruited to the snRNA gene or whether the complete Integrator complex is required for snRNA cleavage. Notably, specific Integrator subunits can function in sub-complexes independently of the complete Integrator complex. For example, IntS3 and IntS6 are subunits of the sensor of the single-stranded DNA (SOSS)1 and SOSS2 complexes, which are essential for double-stranded DNA break (DSB) repair process [305–309]. Similarly, IntS13 is recruited to

enhancers and plays a role in enhancer activation and myeloid differentiation [310]. IntS10, IntS13, and IntS14 form an independent functional module with an affinity for nucleic acids, preferentially RNA hairpins [311]. Integrator subunits also play unique pathophysiological roles in metazoans, ranging from hematopoietic development to carcinomatosis [168,312,313]. Taken together, it is plausible that specific Integrator subunits are recruited to snRNA genes by different transcription factors, for example, CDK9 [96,304], with the whole complex assembling when the 3' end processing signal is reached. For instance, SPT6, a histone chaperone and a key elongation factor for transcription of human protein-coding genes [156,314–319] is present on snRNA genes and is required for IntS3 recruitment to snRNA genes [320].

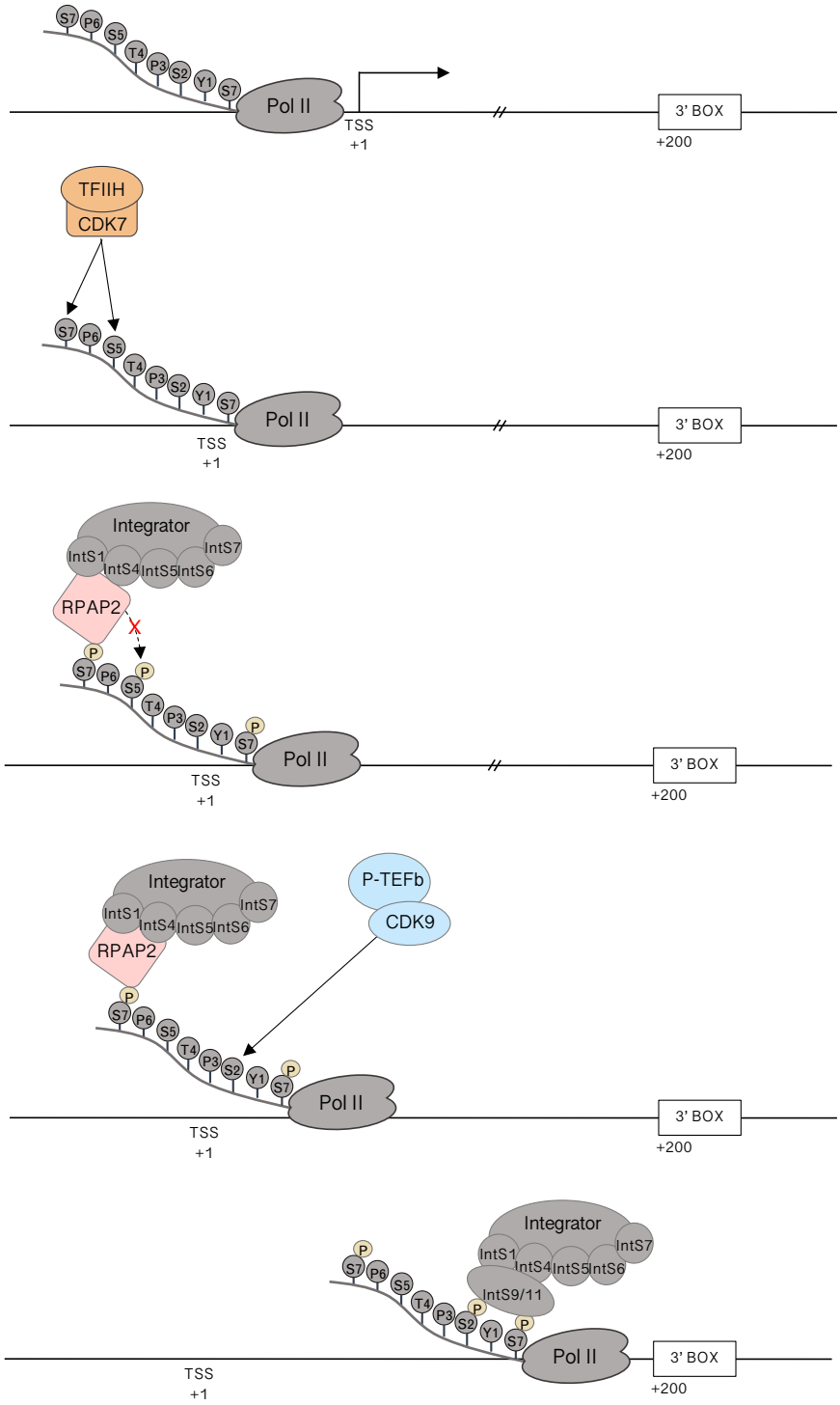


Figure 1.8 CDKs mediate the expression of snRNA genes through CTD phosphorylation

CDK7 phosphorylates Ser5 and Ser7 of the CTD after transcription initiation. RPAP2 binds to pSer7 and dephosphorylates pSer5. Additionally, RPAP2 associates with non-catalytic subunits of the Integrator complex. CDK9 phosphorylates CTD Ser2. Upon the recognition of the 3' box, the presence of pSer7 and an adjacent pSer2 facilitates the recruitment of IntS9/11, the endonuclease module.

1.7.3.2 Potential Roles of tCDK in snRNA Gene Expression

Transcriptional CDKs (tCDKs) may also play a role in recruiting Integrator subunits and other transcriptional factors to facilitate snRNA expression. For example, CDK12 was recently shown to be present on all snRNA genes [320], and its regulatory cyclin, Cyclin K, was significantly enriched at the 3' box of U2 snRNA genes. Additionally, PTFα (alternatively SNAP190 or SNAPC4) was found to have significantly decreased phosphorylation upon inhibition of CDK12 and CDK13 together [75]. CDK12 or CDK13 knockdown also affects snRNA gene expression [80].

While the roles of many tCDKs in snRNA gene expression remain unclear, these kinases may play a role through CTD phosphorylation, similar to CDK7 and CDK9. CDK8 is known to phosphorylate Ser5 [321], while CDK10 and CDK11 can phosphorylate Ser2 [72,322]. Both CDK12 and CDK13 are capable of phosphorylating Ser2 and Ser5 [77,80,323,324]. Additionally, the roles of CDK9 on snRNA gene expression may extend beyond Ser2 phosphorylation as CDK9 is known to target Thr4 [41], Ser5 [67], and Ser7 [54] residues of the CTD.

1.7.4 snRNA transcription termination

Similar to protein-coding genes, the processing of snRNA gene transcripts is generally believed to be coupled with transcription termination [325]., Termination is highly dependent on the recruitment and activity of the Integrator complex, as evidenced by the termination defects caused upon the loss of IntS11 and IntS9 [325]. Interestingly, recent studies suggest that snRNA termination can occur independently of the RNA cleavage by the Integrator complex, indicating the

existence of an alternative or complementary mechanism [326]. In human U1 and U2 snRNA genes, Pol II travels approximately 600 nucleotides past the 3' box, where it supposedly encounters nucleosomes and, in response, terminates transcription [325,327]. Nucleosome repositioning and open-state chromatin are required for Pol II accessibility during transcription [328]. Indeed, it has been demonstrated that the PTF complex maintains a constitutionally open chromatin state across the U2 snRNA gene during interphase to facilitate gene expression [325,329]. NELF is recruited and accumulates at the end of the nucleosome-free region, approximately 800bp downstream of the transcription start site, where it plays a role in termination [330]. Importantly, this position is very close to the CTCF binding site [330]. CTCF is a versatile protein which plays a role in enhancer-blocking insulation to prevent unwanted promoter-enhancer interactions [331]. CTCF binding also plays a role in nucleosome repositioning [332] and can cause Pol II pausing [333]. However, CTCF does not function as an enhancer-blocking insulator on snRNA genes as it is recruited at the 3' end of the gene [330]. Instead, CTCF recruits SPT5, the large subunit of DSIF [334], CDK9, which is required for 3' end processing [303,304], and NELF, an essential termination factor[330], thereby coupling 3' end processing to termination [335]. The current model for human snRNA transcription proposes a "late" elongation checkpoint, where CTCF binds to its binding site downstream of the 3' box, repositions nucleosomes to slow down Pol II, then may recruit CDK9 for 3' processing, and NELF and DSIF for transcription termination.

Despite the orthologous relationship between yeast and human snRNA genes and the similarities in termination mechanisms of protein-coding genes in both species, the termination of transcription of yeast snRNA genes follows a different pathway,

which requires the recruitment of Nrd1, Nab3, and Sen1 proteins [336]. Nrd1 is a CTD interaction and RNA-binding protein, which binds to Ser5P at the 5' end of the gene [337–339]. Nrd1 also interacts with Nab3, an RNA-binding protein and Sen1, a helicase proposed to disrupt the elongation complex. This termination complex is also associated with the exosome complex through interactions of Nrd1, facilitating 3' end processing [340]. Termination of transcription of human snRNA is more similar to the process in human protein-coding genes than in yeast snRNA genes. As RNA processing is coupled to termination, protein-coding 3' end processing factors such as PCF11, SSU72 and CstF are also found on human snRNA genes. However, they function as termination factors [325]. SCAF4, the human ortholog of Nrd1, is an antitermination factor on protein-coding genes [195]. There is no known human ortholog of Nab3, and Senataxin, the human ortholog of Sen1, plays a role in R-loops resolution and protein-coding termination [341,342], but its knockdown does not affect snRNA gene transcription[325].

1.7.5 Overlap of regulatory factors in protein-coding and snRNA genes: distinct functional roles

Several transcription factors are shared between protein-coding genes and Pol II-transcribed snRNA genes, although some exhibit distinct functions depending on the gene type. Advances in identifying novel and shared factors have enhanced our understanding of transcriptional regulation across both gene types. Nevertheless, many questions persist. For instance, transcription factors whose activity is at the 5' end or gene body of protein-coding genes, such as CTCF, are often involved in

processes at the 3' end of snRNA genes [332,335]. Could this be because snRNA genes are short and intronless? Moreover, do snRNA gene-specific factors, such as the PTF complex and ICE1 and ICE2 of the LEC complex, contribute to the obligate coupling between the snRNA promoter and the 3' box? Beyond RPAP2 recruitment, the mechanisms by which other non-catalytic Integrator subunits are recruited remain unknown. Furthermore, could other tCDKs, in addition to CDK7 and CDK9, play roles in snRNA expression?

1.7.5.1 How does obligate coupling of an snRNA promoter to 3' box function work?

Transcription and RNA 3' end processing are tightly coupled during the expression of protein-coding and snRNA genes [96,343]. The RNA cleavage of the transcript of both genes involves the activity of the respective homologous endonucleases, CPSF73 for protein-coding genes and IntS11 for snRNA gene transcripts [344]. However, different mechanisms are required for 3' end processing and formation of the cleavage machinery. In both cases, this process requires Pol II. As earlier described (1.4.3), on protein-coding genes, the CPSF complex binds to the Poly (A) site, and the interaction of CstF with the DSE stimulates the endonuclease activity of CPSF73 to cleave nascent mRNA just downstream of the poly (A) signal [345,346] (Fig. 1.9). The activity of Pol II is required for this process to occur as these complexes interact with the CTD, and the truncation of Pol II CTD leads to an mRNA processing defect [343,347]. Also, the CPSF complex is recruited earlier in transcription by TFIID, and it docks on phosphorylated Pol II at the start of elongation [348], essentially travelling with Pol II until the requirement for nascent mRNA cleavage. Similarly, Pol II acts as a 3' end

processing co-factor during the expression of snRNA genes by promoting the recruitment of the Integrator cleavage module through the interaction with its phosphorylated Ser 2 and an adjacent Ser7 [96] (Fig. 1.9).

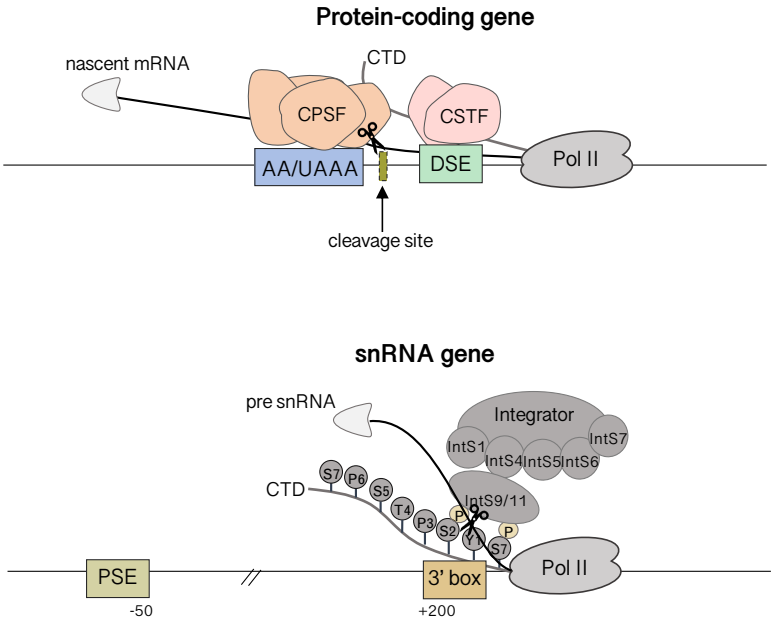


Figure 1.9 Pol II couples transcription and 3' end processing in protein-coding and snRNA genes
Schematic illustration of the involvement of Pol II in coupling transcription to 3' end processing during expression of protein-coding and snRNA genes. CPSF and CstF interact with Pol II, and upon their association with their respective cis-elements and recruitment of other cleavage factors, pre-mRNA cleavage occurs. Similarly, Pol II CTD is phosphorylated at Ser2 and Ser7 to recruit the cleavage module of the Integrator complex to snRNA genes.

Similar to how CPSF travels with Pol II along the protein-coding gene and how non-catalytic Integrator subunits are recruited early in the expression of snRNA genes [99,348], the factor which controls snRNA promoter and 3' box coupling likely interacts with and/or regulates the Pol II CTD. This coupling conundrum can be likened to a navigational system, where the PSE initiates transcription and tells Pol II when to

terminate transcription while also informing the transcriptional complex of regulatory checkpoints such as the 3' box. The 3' box serves as a critical checkpoint to avoid transcriptional readthrough, analogous to a traffic light signalling a car to stop. However, other factors (i.e. inhibition of CDK9 [96,304]) can also contribute to similar transcriptional outcomes. To better understand why the recognition of this checkpoint fails, it is essential to study the factors that regulate Pol II, particularly at this 'traffic light' point of misregulation – the 3' end of the gene.

1.8 Project aims

This project aims to identify the factors involved in the transcription of snRNA genes and the processing of the transcripts to identify the specific mechanisms regulating the expression of these genes. This work uses the RNU2 gene in the Human Embryonic Kidney (HEK) 293 cell line as the model gene for Pol II-transcribed snRNA gene expression, as the RNU2 genes are tandemly repeated approximately 15 times in the haploid human genome [210,349], and the U2 locus is likely triploid in HEK293 cells [350].

Chapter 2 lists the materials and methods employed in this study.

Chapter 3 focuses on elucidating the roles of transcriptional CDKs in the recognition of the 3' box and snRNA 3' end processing.

Chapter 4 investigates the involvement of tCDKs in Pol II CTD phosphorylation.

Chapter 5 explores the mechanisms by which redundant CDKs may regulate 3' box recognition by identifying novel factors potentially involved in snRNA gene expression.

Chapter 6 summarises the findings from the previous chapters, highlights models by which potential new factors are involved in snRNA 3' end processing and discusses future work needed to test these models.

Chapter 2 | Materials and Methods

2.1 Cell culture

Human Embryonic Kidney (HEK) 293 were cultured at 37°C and 5% CO_2 in Gibco Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% heat-inactivated Foetal Bovine Serum (FBS) (Sigma Aldrich, F9665), 2mM L-Glutamine (Gibco, 25030024) and 100U/ mL Penicillin streptomycin (Gibco, 15140122). HEK293 is the parental cell line for all analog-sensitive (AS) and degron-tagged (dTAG) cell lines that were created. Thus, AS and dTAG cell lines were cultured accordingly. Cells were passaged every three days at approximately 80% confluency using PBS containing 1 mM EDTA. All cell culture was performed in a Class II laminar flow hood.

2.1.1 Cell line storage

Cells were grown to 80% confluency in a T-75 flask and resuspended in 1 ml 100% FBS; 900 µl of this was added to 100 µl DMSO (Merck, D8418) in a cryogenic vial (Starlab, E3090-6222) and stored at -70°C.

2.2 Transfection into HEK293 or AS cells

Transfection was performed using jetPRIME reagent (Polyplus) according to the manufacturer's instructions. Cells were seeded in a 6-well plate and grown overnight to approximately 70% confluency. Subsequently, 2 µg of plasmid DNA was

diluted in 200 µl jetPRIME buffer, jetPRIME reagent was added, and the mix was incubated at room temperature (RT) for 10 minutes and added to cells.

2.3 alamarBlue HS Cell Viability Assay

Cells were grown to approximately 80% confluency and seeded into 95 µl of 250 cells per well in a clear-bottom 96-well plate (Greiner Bio-one, 655090). The top row of the 96-well plate contained media only to assess the background signal of the assay. Cell proliferation was read every 12h or 24h by adding 5 µl alamarBlue HS cell viability reagent (ThermoFisher, A50101). Fluorescence was measured with a 560 nm excitation and 590 nm emission wavelength, respectively, and reading was taken after 1 hour using a fluorimeter, according to the manufacturer's instructions.

2.4 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

2.4.1 Western blot

SDS-PAGE gels were prepared according to the molecular weight of the protein of interest. Table 2.1 details the composition of one 5% gel cast in a 1.5mm plate. The formulation for different % gels can be adjusted accordingly. Cells were harvested in media and centrifuged at 13,000 x g for 2 min at RT. The pellet was lysed in Laemmli buffer containing 100 mM Dithiothreitol (DTT). Samples were then sonicated using QSonica Thermocube Q800R (QSONICA) at 100% Amplitude, with a cycle of 2 min on and 2 min off. After sonication, samples were heated at 95°C for 5 minutes to denature proteins. Gels were run in 1x SDS-running buffer (250 mM Tris, 1.9 M Glycine, 34 mM SDS) at 100V at 4°C.

Table 2.1 SDS protein gels components for 5% gel

Reagent	Resolving gel	Stacking gel
30% Acrylamide (Avantor, J61505.K2)	1.6 ml	0.9 ml
Distilled H_2O	5.7 ml	3.6 ml
Resolving or stacking buffer	2.7 ml resolving buffer	1.5 ml stacking buffer
25% Ammonium persulphate (APS) (Sigma, A3678)	40 µl	20 µl
TEMED (Merck, T9281)	8 µl	15 µl

Proteins were transferred to the nitrocellulose membrane (ThermoFisher, 88018) using the wet transfer method. The gel was transferred with a wet transfer buffer containing ethanol. The membranes were then blocked using 2.0% Bovine Serum Albumin (BSA) in Tris Buffered Saline containing 0.1% Tween-20 (TBS-T) for 30 minutes and incubated with primary antibodies diluted in 2.5% BSA in TBS-T overnight at 4°C. Membranes were washed three times, 5 minutes each, with TBS-T at RT and incubated with HRP conjugated secondary antibody diluted in 1% BSA in TBS-T at RT for an hour. Following antibody incubations, membranes were washed twice, for 5 minutes each, with TBS-T and once for 10 minutes with TBS. The protein bands were then detected using ECL chemiluminescent reagent (ThermoFisher), according to manufacturers' instructions and visualised using the iBright FL1000 imaging system (ThermoFisher).

2.5 Chromatin Immunoprecipitation (ChIP)

2.5.1 Crosslinking

Cells were seeded on 15 cm dishes and grown to 80% confluency overnight. Cells were then treated with the respective drug/vehicle and subsequently crosslinked with 1% formaldehyde (F8775, Sigma) for 10 minutes at RT. Glycine was added to a final concentration of 125 mM and incubated for 5 minutes at RT to quench formaldehyde activity.

2.5.2 Cell lysis and sonication

Cells were washed twice with 10 ml of cold PBS, centrifuged at 500 x g and resuspended in 2 ml Lysis buffer (10 mM Tris-HCl pH 8.0, 0.25% Triton X-100, 10 mM EDTA, and 1% SDS) prepared with protease and phosphatase inhibitors (cOmplete, EDTA-free Protease Inhibitor Cocktail, Sigma and PhosSTOP, Roche, respectively), which were added freshly to all buffers whenever required throughout the ChIP protocol. For ChIP-sequencing (seq), 10% crosslinked SNH mouse cells were added for spike-in. Cells were incubated with lysis buffer on ice for 10 minutes, split into 2 x ml samples and centrifuged at 1500 x g, 4°C for 5 minutes, and the 2 ml was split into 2 x 1 ml samples. Cells were then resuspended in 1ml wash buffer (0.2 M NaCl, 10 mM Tris-HCl, and 1 mM EDTA) each and centrifuged at 1500 x g, 4°C for 5 minutes. Cell pellet was resuspended in 300 µl sonication buffer (100 mM NaCl, 10mM Tris-HCl pH 8.0, 1mM EDTA, 0.1% SDS, and prepared with fresh protease and phosphatase inhibitors), transferred into sonication tubes (Diagenode, C30010016)

and chromatin was sheared to approximately 250bp using Bioruptor Pico sonication device (Diagenode, B01080010) or QSonica Thermocube Q800R.

2.5.3 Immunoprecipitation (IP)

The sheared chromatin was centrifuged at 20 000 x g for 20 minutes, and the supernatant of the 2 x 300 µl samples was combined and transferred into DNA LoBind tubes (Fisher Scientific, 10051232). Samples were subsequently quantified, and 100 µg per IP was precleared using 20 µl of Dynabeads G (Life Technologies, 10004D) washed in 200 µl of RIPA buffer (10mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 0.1% SDS, 1% Triton X-100, 0.1% Sodium deoxycholate, and prepared with fresh protease and phosphatase inhibitors) and rotated at 16 rpm, 4°C for 30 minutes. Precleared samples were incubated with antibodies/ IgG and rotated at 16 rpm, 4°C overnight. For ChIP-seq, 7 Ips were performed per condition. Another 20 µl of Dynabeads G per IP was washed in 200 µl ChIP buffer, saturated with 4 mg/ml BSA in 100 µl of RIPA buffer, and rotated at 16 rpm, 4°C overnight. The blocking solution was subsequently removed from beads and washed twice in lysis buffer (prepared with protease and phosphatase inhibitors) using a DynaMag magnetic rack (ThermoFisher, 12321D). Sonicated extracts were then incubated with blocked beads and rotated at 16 rpm, 4°C for an hour. Following this, IgG supernatant was saved as total input and stored on ice, and both IgG and antibody-bound beads were washed three times with 300 µl of pre-chilled RIPA buffer, three times with pre-chilled high-salt wash buffer (10mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM EDTA, 0.1% SDS, 1% Triton X-100, 0.1% sodium deoxycholate), two times with 200 µl of

prechilled LiCl buffer (10mM Tris-HCl pH 8.0, 250 mM LiCl, 1mM EDTA, 1% NP-40, 1% Sodium deoxycholate) and two times with 600µl of pre-chilled TE buffer (10mM Tris-HCl pH 7.5, 1mM EDTA). IP samples were then resuspended in 50 µl of elution buffer (100 mM $NaHCO_3$, 1% SDS, and prepared with freshly added 10 mM DTT) and mixed at 1400 rpm, RT for 15 minutes on Eppendorf Thermomixer (Merck, EP5387000030). The elution process was repeated once, and eluates were combined. The total input was also diluted 1:10 in the elution buffer.

2.5.4 Decrosslinking and DNA Purification

All samples were treated with 1µl of 10 mg/ml RNase A (ThermoFisher, EN0531), 200mM NaCl and were incubated for 5 hours before chromatin was precipitated using Ethanol. Samples were centrifuged for 20 min at 20 000g, 4°C, and the pellet was dissolved in 100 µl of TE buffer. Proteins were degraded by treating the sample with 1.5 µl Proteinase K (ThermoFisher, 53155) in 25µl of 5X Proteinase K buffer (50 mM Tris-HCl pH 7.5, 25 mM EDTA, 1.25% SDS) for 2 hours. DNA was purified using QIAQuick PCR purification Kit-250 (Qiagen, 28106) according to the manufacturer's instructions and eluted in 50 µl. For ChIP-Seq, a single column was repeatedly used during the elution stage.

2.5.5 DNA shearing assessment

Following DNA purification, 2 µl of the purified sample was diluted in 10X orange DNA loading dye, and the size of the DNA fragment was visualised on a 2% agarose gel.

2.5.6 ChIP-qPCR

A serial dilution of the 1/10 input was performed to generate the following concentrations: 1:5, 1:25, 1:125 and 1:625. A master mix comprising of 1µl ChIP sample or input dilution,1µl primer mix (5% left primer, 5% right primer and 95% buffer), 3 µl nuclease-free water (Life Technologies, AM9937) and 5µl SYBR Green (Qiagen, 204145) was run on in triplicates on the Rotor-Gene Q Thermocycler (Qiagen, 9001862) with a thermal cycling protocol of 15 minutes at 95°C, 40 cycles of 15 seconds at 94°C, 15 seconds at 57°C, 15 seconds at 72°C and 10 minutes melt curve from 50°C to 99°C. A standard curve was generated using serial dilutions quantitation on the Qiagen Rotor-Gene Q software (Qiagen, version 2.3.1.49). The quantitation data was analysed relative to input dilutions, and IgG levels were subtracted during analysis.

2.5.7 ChIP-Seq

2.5.7.1 Library preparation

DNA was converted into libraries for next-generation sequencing using the NEBNext Ultra™ II DNA Library Prep Kit (NEB, E7645S/E7103S) and was performed according to the manufacturer's instructions. Notably, 500 pg of fragment DNA was used as starting material for end repair and less than 5ng of input was used for adaptor ligation. SPRIselect beads (Fisher Scientific, 15676104) were used for the cleanup of adaptor ligation, and the thermal cycling protocol for adaptor amplification was as follows: 1 cycle at 98°C for 30 seconds, 11 cycles of 98°C for 10 seconds and 65°C for 75 seconds and 1 cycle of 65°C for 5 minutes. PCR reaction cleanup was

performed twice, and the library fragment (approximately 400bp) was visualised using the TapeStation system (Agilent, G2991BA). Sample DNA concentrations were quantified using Qubit HS DNA assay, pooled, and sent for partial lane sequencing with 200G raw data per sample at Novogene.

Table 2.2 Primer sequence for ChIP-qPCR

Region	Forward sequence	Reverse sequence
Negative	GGAGCGGAGCGTTCTCTGTCTCCCC	AGAGTGTGAGCCCTCATTCACGCCC
DSE	TGGCTCGATACGAACAAGGAAG	GTTCCCGGGCTTTCATTTCG
upPSE	GGGAACGCCGAAGAAGCACGGG	CCCCAGCCTCGCTCCTTGCCC
PSE	ATGAGAGTGGGACGGTGA	CACTTGATCTTAGCCAAAAGG
3' Box	ACGAGTCCTGTGACGCGCCGGCTTG	CTCCGGGTGGGTCCCATTCCTTTAA
CT-rich	CCTCCCCGCCTCTCCCTCGCTC	GGACAAATAGCCAACGCATGCGG

2.5.7.3 Bioinformatic analysis

This subsection (2.5.7.3) was written in collaboration with Dr Henfrey, whose work on this project is acknowledged with thanks

Dr. Callum Henfrey, a postdoc and bioinformatician in the lab, performed bioinformatic analysis. The datasets were downloaded in fastq.gz format from Novogene, and read quality metrics were assessed using FASTQC (version 0.12.1) [351]. Adapters were trimmed with CUTADAPT (version 1.18) in paired-end mode with the following options: -minimum-length 10 -q 15,10 -j 16 -A GATCGTCGGACTGTAGAACTCTGAAC -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC. Alignment of trimmed reads to the human genome (hg38), using the Gencode V44 annotation, was performed with STAR version 2.7.8a with the parameters runThreadN 16 --readFilesCommand gunzip -c -k --limitBAMsortRAM 20000000000 --outSAMtype BAM SortedByCoordinate. SAMTOOLS was used to retain the properly paired and

mapped reads (-f 3) and to remove PCR duplicates for paired-end sequencing. Reads mapping to the DAC Exclusion List Regions (accession: ENCSR636HFF) were removed with BEDTOOLS version 2.29.2 [352]. Read counts were scaled to the SNH spike-in by normalising all files using the number of reads mapping to the *Mus musculus* genome. FPKM-normalized bigwig files were created with DEEPTOOLS2 bamCoverage tool with the parameters -bs 10 -p max –e –normalize using RPKM. Bigwig files were used to generate read matrices using the DEEPTOOLS2 compute matrix function, set to scale-regions mode, splitting each gene body into 400 bins with 2.5kb upstream of the TSS and 2.5kb downstream of the TES. Reads were separately mapped to the *RNU2* gene and normalised for library size to determine U2 metagene profiles.

2.6 RNA extraction

Cells were seeded on 15 cm dishes and grown to 70% confluency overnight. Cells were then treated with the respective drug or vehicle, harvested and centrifuged at 20 000 x g for 5 minutes. The cell pellet was resuspended in 100 µl RNA lysis buffer (140 mM NaCl, 1.5 mM $MgCl_2$, 10mM Tris-HCl pH 7.5, 0.5% v/v NP-40) and 100 µl of 2X HSB buffer (280 mM NaCl, 10mM KCl, 1.5 mM Na_2HPO4, 12 mM Glucose, 50mM HEPES pH 7.05) was added. Samples were treated with 5 units of DNase I (NEB, M0303S) and incubated at 30°C for 5 minutes, after which Proteinase K mix (0.4 mg/ml Proteinase K, 0.6% SDS) was added, and samples were incubated at 37°C for 30 minutes. Following DNase and Proteinase K treatment, 30 µl of 4M ammonium acetate and an equal volume of Phenol: chloroform pH 4.5 (ThermoFisher, AM9722)

were added, and samples were vortexed and centrifuged at 20 000 x g for 10 minutes. The RNA in the aqueous phase (top layer) was precipitated using ethanol, and samples were centrifuged at 20,000 x g for 10 minutes. The pellet was resuspended in 100 µl DNase buffer (10mM Tris-HCl pH 7.5, 10mM $MgCl_2$) and treated with 10 units of DNase I at 37°C for an hour. Following this, 10µl 3M NaOAc and an equal volume of Phenol: chloroform pH 4.5 were added to the samples, and RNA was precipitated in ethanol and stored at -20°C.

2.7 First-strand complementary DNA (cDNA) synthesis

Ethanol-precipitated RNA was centrifuged at 20,000 x g at 4°C for 20 minutes, and the pellet was resuspended in 50 µl nuclease-free water. Following this, 1µg of RNA per condition was diluted in 14.75 of water, and 2µl Primer mix (1 µl 10mM dNTPs, 1 µM 100 µM reverse gene-specific primer) was added to the sample. Samples were subsequently incubated at 65°C on a PCR thermocycler, and 8.25 µl of superscript mix (5µl 5 X FS buffer, 1.25 µl 0.1M DTT, 1µl Superscript III, 1 µl RNase out) was added. Samples were vortexed, pulse-centrifuged, and run in a thermocycler with the protocol: 55°C for an hour, 70°C for 55 minutes and 4°C for 5 minutes. cDNA samples were subsequently stored at -20°C.

2.8 Reverse transcription qPCR (RT-qPCR)

cDNA samples were diluted in equal volume of nuclease-free water. A master mix comprising of 1µl cDNA sample,1µl primer mix, 3 µl nuclease-free water (and 5µl

SYBR Green was run in triplicates on the Rotor-Gene Q Thermocycler (Qiagen, 9001862) with the thermal cycling protocol of 15 minutes at 95°C, 40 cycles of 15 seconds at 94°C, 15 seconds at 57°C, 15 seconds at 72°C and 10 minutes melt curve from 50°C to 99°C. Average CT values from housekeeping samples are then deducted from experimental sample CT values before CT values in the gene-specific primer are normalised to their respective control. The fold change in expression is then derived by the formula $2^{-\text{CT value}}$, where CT value has been normalised to control and housekeeping gene.

Table 2.3 Primer sequence for RT-qPCR

Name	Forward sequence	Reverse sequence
Pre snRNA	ATCTGATACGTCCTCTATCCGA	GAAACACGTTGTATCCCCGG
Readthrough snRNA	AGAGGGAGGTAAGAGACGGT	CCTTCCCAAAGCGCAGTAG
7SK snRNA	CTGATCTGGCTGGCTAGGCGGG	GAAGACCGGTCCTCCTCTATCGG

2.9 Cell lines creation

2.9.1 Analog-sensitive cells (AS)

This subsection (2.9.1) was written in collaboration with the Genome Engineering Oxford (GEO) facility, whose work on this project is acknowledged with thanks

Professor Murphy created the CDK12AS cells, as described in Tellier *et al*. [77]. GEO and Professor Murphy collaborated on the CDK13AS and CDK12/13AS cell lines. The CDK12 (accession: NM_016507) mutation was performed using CRISPR-Cas12, while the CDK13 (accession: NM_003718.5) mutation was performed using the CRISPR-Cas9 system. The guide and HDR template sequences can be found in Table 2.4.

2.9.1.1 Guide RNA complex

Cas9 guide RNA complexes were formed by mixing the crRNA and tracrRNA (IDT) at a final concentration of 100 µM in IDT Duplex Buffer (30 mM HEPES, pH 7.5, 100 mM potassium acetate), heating the mixture to 95 °C, and slowly cooling it to room temperature.

2.9.1.2 RNP formation

For Cas9 RNP formation, 5 µg Cas9 HiFi protein (IDT) was mixed with 0.5 µl guide RNA complexes from the previous step in a total volume of 2 µl IDT Duplex Buffer and incubated at 37ºC for 5 minutes. For Cas12a RNP formation, 2 µg Cas12a ULTRA protein (IDT) was mixed with 15 pmol of each crRNA in a total volume of 2 µl IDTE Buffer (IDT, 11-05-01-04), resuspended 5 times and incubated at 37ºC for 5 minutes. Resulting RNP mixes were added directly to cells before nucleofection.

2.9.1.3 RNP nucleofection

Cells were washed twice with PBS, resuspended in 8 µl buffer R and mixed with 2 µl RNP solution from the previous step, 2 µl of a 2.16 µM Alt-R Cas12a electroporation enhancer (IDT, 1076300), diluted in IDTE buffer, and 125 pmol ssODN homology-directed repair (HDR) template. This mixture was resuspended five times in a 10 µl Neon nucleofector tip and nucleofected with the Neon Transfection System using the following settings: 1700 V, 20 ms, 1 pulse (program “J”) or 1325 V, 10 ms, 3 pulses (program “J2”). Upon nucleofection, cells were transferred directly into a single well

of a 24-well plate containing 500 µl pre-warmed complete DMEM and recovered overnight in a humidified incubator at 37°C, 5% CO_2. Media was replaced the following morning, and cells were cultured to 70-80% confluency and genotyped using PCR and Sanger sequencing of the target amplicon. The gDNA and cDNA were prepared as described in Tellier *et al*. [77]. CDK12 fragment was amplified using GGTGGCAGAGTGAGAGTCTG forward primer and CCTTCCATTAGCTGTTTCATGAA backward primer. CDK13 fragment was amplified using GAAGGTACTTACGGACAAGTTTA forward primer and CCAAGGATACAGCCACAGCTC reverse primer.

Table 2.4 Guide and HDR template sequences used for AS cell creation

Name	Sequence (5' – 3')
CDK12 guide 1	AGTATATGGACCATGACTT
CDK12 guide 2	CCTTGTATTTGAGTATATA
CDK13 guide 1	TTATCTGGTGTTTGAATATA
CDK13 guide 2	CCAGTAGTCCCATCAGATCA
CDK12 HDR template	T*T*AATTTTTTTTCGTCTTTATGTAGGTGCCTTTTACCTTGTAGGCGAGT ATATGGACCATGACTTAATGGGACTGCTAGAATCTGGTT*T*G
CDK13 HDR template	A*A*TGTATTAATTATTCTCATTCTTAGGTGCATTTTATCTGGTGGGCGA ATATATGGACCATGATCTGATGGGACTACTGGAATCAGGC*T*T

2.9.2 Degron-tagged cells

STP6 (accession: NM_001320755.2), ICE2 (accession: NM_024611.6) and IntS3 (accession: NM_001324475.2) HDR template containing degradation, SMASH and 3X FLAG tags were designed by Gilbert Ansa (a DPhil student in the Murphy lab) and constructed by Prof. Murphy (Table 2.5, Figure 2.1) and cloned into the mammalian

expression vector pCDNA6/TR, which carries a blasticidin resistance gene. The SPT6 HDR template was transfected into HEK293 cells along with pX462, which carries a puromycin resistance gene, expressing dCas9 and gRNAs (Table 2.5). Transfected cells were seeded on 96-well plates with DMEM supplemented with 20% FBS, 2mM L-Glutamine and 100U/ mL Penicillin and selected using Blasticidin S and Puromycin antibiotics. Cells were subsequently grown until visible colonies were formed. Mrs Barker and I screened and validated the engineered cells using western blotting to identify homozygous clones. A list of HDR templates and guide RNA sequences used can be found in Table 2.5

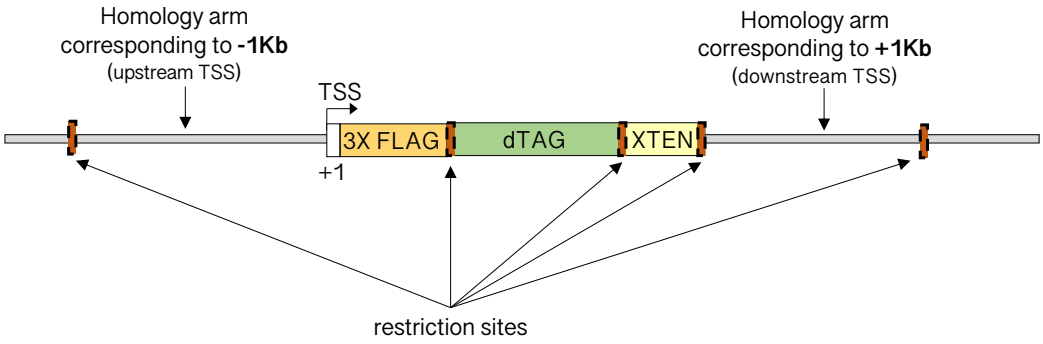


Figure 2.1 Schematic representation of the dTAG, 3x FLAG and XTEN linker relative to the transcription start site (TSS) of the construct in pCDNA6/TR that was transfected. The figure illustrates the transfected construct in pCDAN6/TR and highlights the positions of the tags, XTEN linker, restriction sites on the HDR template, and homology arms corresponding to the sequences (1kb) upstream or downstream from the transcription start site of the gene of interest.

Table 2.5 HDR templates and guide RNA sequences for degron cell lines. Abbreviations: UPF: upstream forward; UPR: Upstream reverse; DNF: downstream forward; DNR: downstream reverse

HDR template		
Name	Forward	Reverse
SPT6 upstream	GTACGGTACCTGTAATCCCAGCAC CTTGGGAG	TAGTCCATGGCTGCAGCTTCTCACTC TGCC

SPT6 downstream	GATCGGATCCTCTGATTTTGTGGAAAGCGAGGC	GATCGCGGCCGCGAAGTCATTTCGTGGTTCTACTC
IntS3 upstream	GTACGGTACC	TAGTCCATGGCTGCAGCCAGGAGC
IntS3 downstream	GATCGGATCCGAGTTGCAGAAGGGAAAAGGGG	GATCGCGGCCGCGTTGGCCAGGATGGTCTCGATCTC
ICE2 upstream	GTACGGTACCGAAAGGCCTAAGATTCCTGATATC	TAGTCCATGGTCCTAGATTTCTGCTTCACTCTAGC
ICE2 downstream	GATCGGATCCAGCTCCAAGATGGTCATAAGTG	GATCGCGGCCGCCCTTGGCTGGCTTAGTAAATGTG
Guides		
SPT6 upstream	Guide 1 CACCGATACAATGATGAAGGCGAGG Guide 2 AAACCCTCGCCTTCATCATTGTATC	
SPT6 downstream	Guide 1 CACCGACAAATTTCTTGGTGACTCG Guide 2 AAACCGAGTCACCAAGAAATTTGTC	
IntS3 upstream	Guide 1 CACCGCCGGCTCCTGGCTGCAGCCA Guide 2 AAACTGGCTGCAGCCAGGAGCCGGC	
IntS3 downstream	Guide 3 CACCGCCATGGCTGCAGCCAGGAGC Guide 4 AAACGCTCCTGGCTGCAGCCATGGC	
ICE2 upstream	Guide 1 CACCGCTAGAGTGAAGCAGAAATCT Guide 2 AAACAGATTTCTGCTTCACTCTAGC	
ICE2 downstream	Guide 3 CACCGTAGGAAGATGAGCTCCAAGA Guide 4 AAACTCTTGGAGCTCATCTTCCTAC	

Table 2.6 List of antibodies used for ChIP and Western blot (WB) analysis

Antibodies	Company	Application
Mouse anti-Flag M2	Merck, F3165	WB
Rabbit anti-BRAT1	Abcam, 11657-1-AP	ChIP
Rabbit anti-CDK9 (D-7)	Santa Cruz	ChIP
Rabbit anti-Cyclin K	Bethyl, A301-939A	ChIP
Rabbit anti-ICE2 (NARG2)	Life Technologies, PA598485	WB
Rabbit anti-ICE2 (NARG2)	Life Technologies, BS-5968R	WB
Rabbit anti-IntS11	Novus, NB100-60638	ChIP, WB
Rabbit anti-IntS3	ProteinTech, 16620-1-AP	ChIP, WB
Rabbit anti-IntS9	ProteinTech, 11657-1-AP	ChIP, WB
Rabbit anti-Nucleolin	Abcam, ab305947	WB
Rabbit anti-Oct1	Bethyl, A301-717A	WB
Rabbit anti-RNA Pol II RPB1 NTD	Cell signalling, 14948	ChIP, WB
Rabbit anti-RNA Pol II Phospho Ser2	Abcam, 5095	ChIP
Rabbit anti-RNA Pol II Phospho Ser5	Abcam, 5131	ChIP
Rabbit anti-RNA Pol II Phospho Tyr1	Active Motif, 61383	ChIP
Rabbit anti-SPT6	Novus, NB100-2582	ChIP, WB
Rat anti-RNA Pol II Phospho Ser 7	Merck, 4E12	ChIP, WB
Rat anti-RNA Pol II Phospho Ser2	Active Motif, 61083	WB
Rat anti-RNA Pol II Phospho Ser5	Active Motif, 61085	WB

Table 2.7 List of inhibitors used

Inhibitor	Company
1-NA-PP1	Cayman Chemical, 10954,
1-NM-PP1	Cayman Chemical, 13330,
5,6-Dichlorobenzimidazole 1-β-D-ribofuranoside (DRB)	Sigma Aldrich, D1916,
Calyculin A	Cayman Chemical,19246
dTAG-^{V}1	Tocris, 6914
Tautomycetin	Tocris Bioscience, 2305
THZ531	MedChemExpress, HY-103618

Chapter 3 | CDK7, CDK9, CDK12 and CDK13 all play roles in the expression of human U2 snRNA genes.

3.1 Background

Cyclin-dependent kinases (CDKs) have a highly conserved structure, which presents a significant challenge to studying their function as it is difficult to specifically inhibit an individual kinase [52,107,108]. To counteract specificity issues, I adopted a chemical biology approach, genetically modifying endogenous CDKs to make them susceptible to rapid and specific inhibition by ATP analogs [110]. The gatekeeper residue, a conserved amino acid that controls the capacity of the ATP binding pocket typically due to its bulky side chains, is mutated to an amino acid with a smaller side chain, such as glycine or alanine. This enlarges the active site of the kinase, making it susceptible to competitive bulky ATP-analog inhibitors such as 1-NM-PP1 or 1-NA-PP1 (Fig. 3.1A) [110,111]. Due to the specificity of the mutation, this method of inhibition is highly selective for the kinase of interest [353,354]. In addition, these bulky ATP analogs do not bind covalently to the active site, making the inhibition reversible. Importantly, the inhibition of the kinase occurs in minutes and allows the inhibition of the engineered kinases without affecting the protein level of the kinase in cells.

The first step in developing an analog-sensitive (AS) kinase is identifying the gatekeeper residue within the kinase domain. The canonical amino acid sequence of the human CDK of interest can be aligned with kinases with known gatekeeper

residue sequences, such as F103 of human CDK9 [95,355]. In most protein kinases, the gatekeeper residue is preceded by two non-polar, aliphatic amino acids (typically isoleucine/leucine and valine) and followed by a polar, acidic amino acid, typically glutamate, and two additional non-polar amino acids [355]. The identified residue is then mutated using CRISPR-mediated homology-directed repair and replaced with glycine or alanine to enlarge the binding pocket [353].

This chapter describes the generation and characterisation of analog-sensitive kinase CDK12AS, CDK13AS and CDK12AS/13AS cell lines using genome engineering, comparison of their growth relative to the wild-type cell lines, and establishment of the concentration of the analog, 1-NM-PP1 (NM), required to inhibit the kinases selectively. This is followed by the determination of the effect of inhibiting these kinases and the previously engineered CDK7AS [60] and CDK9AS [356] on the recognition of the U2 snRNA gene 3' box using a newly developed RT-qPCR assay.

3.2 Cell line generation

It was previously shown that the inhibition of CDK9 with 5, 6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB), a small molecule inhibitor, causes a loss of recognition of the 3' box during U2 snRNA gene expression [304]. I sought to understand the roles of other transcriptional kinases on the 3' end processing of U2 snRNA transcripts. The Murphy lab has previously developed HEK293 analog-sensitive cell lines for many transcriptional CDKs, including CDK7AS [60], CDK9AS [356], and CDK12 [77]. CDK13, a paralog of CDK12 [357], has also recently been identified as

an important transcriptional kinase. Therefore, I created analog-sensitive CDK13 and CDK12AS/13AS HEK293 cell lines.

In addition to being paralogs, CDK12 and CDK13 share the same regulatory protein, cyclin K [51]. CDK12 and CDK13 also often share the same targets—for example, Ser2 and Ser5 of RNA Polymerase II (Pol II) CTD [75–80]. Therefore, it was important to create cell lines where treatment with the drug inhibits both kinases. CDK12/13AS indicates a cell line where the active site of CDK12 was mutated before that of CDK13. CDK13/12AS indicates a cell line where the active site of CDK13 was mutated first. I tried both methods of creating a double mutant cell line to ensure that the order of making the double mutant did not affect the growth of the cell line or the response to inhibition.

CDK12AS, CDK13AS, CDK12AS/13AS, and CDK13/12AS cell lines were generated in collaboration with Genome Engineering Oxford (GEO) and Professor Murphy (Chapter 2.9.1). The CDK12AS cell line was generated before I arrived in the lab. For the CDK13AS, CDK12AS/CDK13AS, and CDK13/12AS cell lines, the colonies of the transfected cells were then sorted into single wells, and PCR was used to create sequencing templates for gDNA, and RT-PCR was used to create sequencing templates for cDNA to identify mutated cell lines. The PCR products were run on an agarose gel to check the size, purified, and sequenced to verify the HDR template insertion at the genomic (gDNA) and mRNA level (cDNA) (Fig. 3.1B).

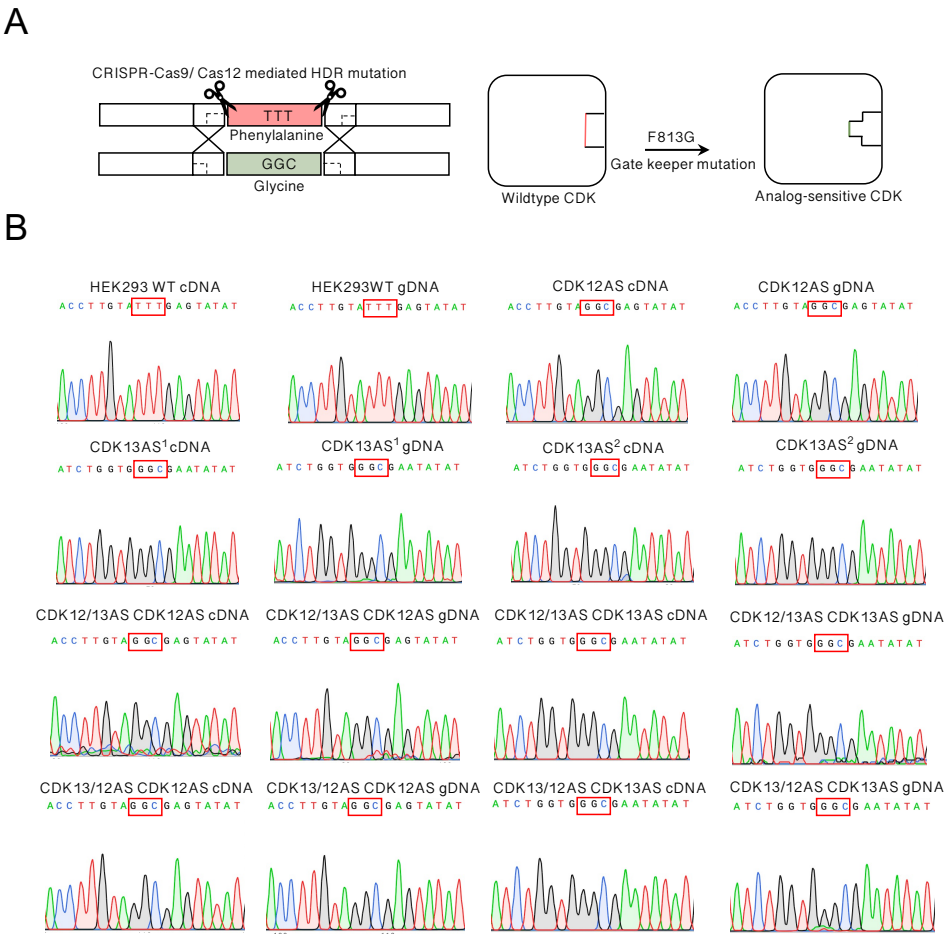


Fig.3.1 Creation of analog-sensitive cell lines

A. Schematic showing the CRISPR-Cas9 directed HDR mutation of the gatekeeper residue in the active site of cyclin-dependent kinases. The gatekeeper residue, phenylalanine (TTT), is mutated to a smaller amino acid, glycine (GGC), to enlarge the ATP binding pocket and make it sensitive to bulky non-hydrolysable ATP analogs. **B**. Sanger sequencing chromatogram showing the cDNA and gDNA of the CDK12, CDK13, CDK12/13, and CDK13/12 analog sensitive (AS) cells.

3.3. Cell Viability Assessment

Having confirmed that the selected cells carry the AS mutation, I compared the cell proliferation and viability of the mutant cell lines to the wild-type parental HEK293 cell lines. Cells were grown to approximately 90% confluency, and 250 cells were counted and seeded into each well of four 96-well clear bottom plates. The

proliferation of the cells was measured every 24 hours over four days; each plate represents a time point (0h, 24h, 48h and 72h).

The number of viable cells was then detected using the alamarBlue HS cell viability assay Chapter 3, and the fluorescence of viable cells was measured at an excitation of 530 nm and emission of 590 nm. To account for background, the fluorescence read of the media alone and cells untreated with alamarBlue reagent was subtracted from the final value of the cell count in each well. The average fluorescence unit of analog-sensitive cells was then compared to HEK293WT cells (Fig.3.2A). All AS cells initially grew slightly better than or at the same rate as the wild-type cells, and at 72 hours, the number of viable cells in the AS cell line and parental cell line was similar, which indicates that the gatekeeper mutation to create analog-sensitive cells did not negatively affect cell growth and proliferation.

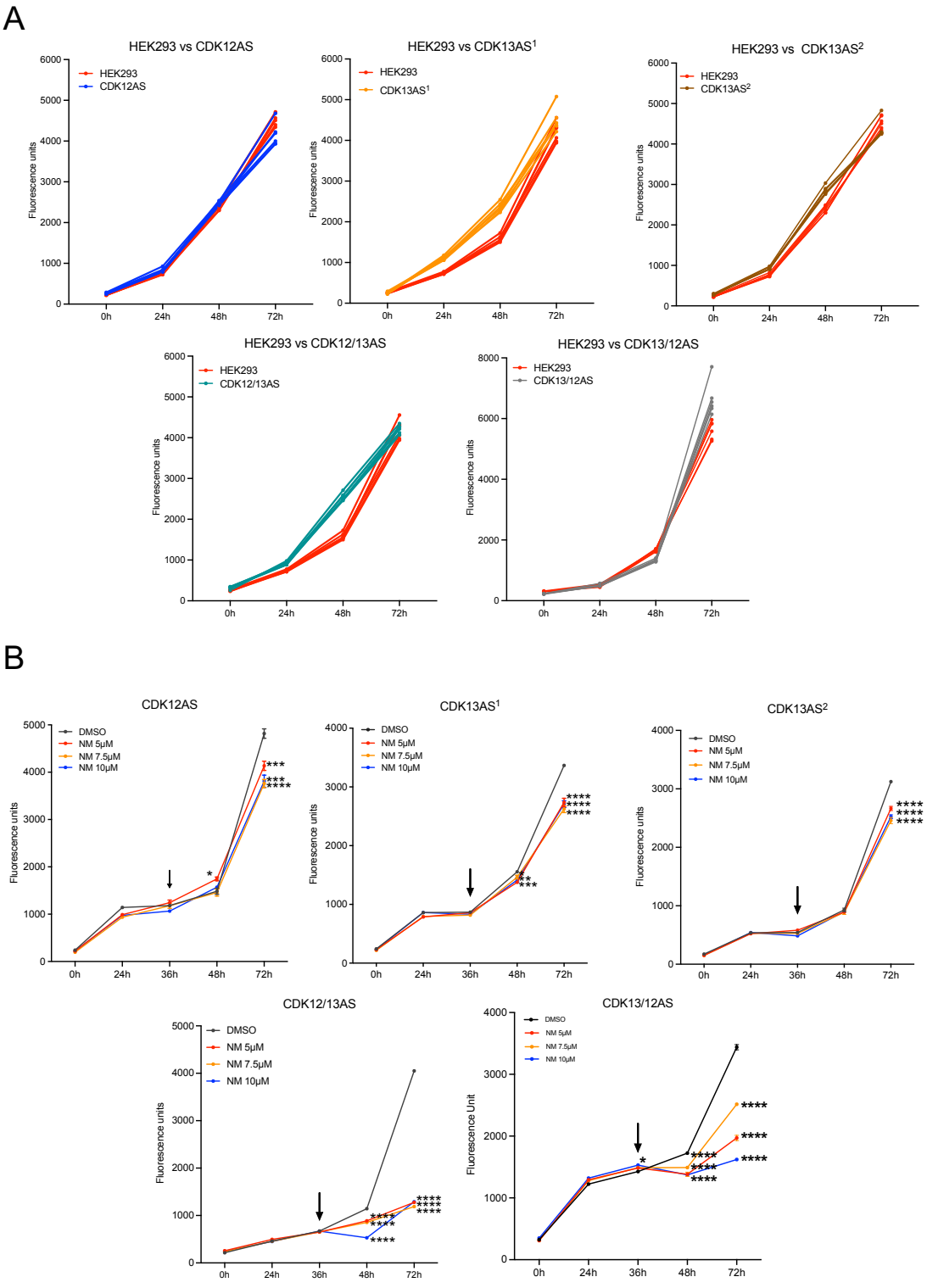


Fig. 3.2 Analog-sensitive mutation does not affect cell growth.

A. Time course of the growth of AS cells, compared to the wild-type HEK293 parental cell line. Cell growth was measured using the alamarBlue HS Cell viability assay. n=6 biological replicates.

B. Time course of the growth of AS cells compared to the wild-type HEK293 parental cell line. Respective concentrations of 1-NM-PP1 or DMSO were added at 36 h, as indicated by a black arrow. Cell growth was measured using the alamarBlue HS Cell viability assay. n=6 biological replicates. Statistical test = two-tailed unpaired t-test, * = p <0.05, ** = p<0.01, *** = p <0.001, **** =p <0.0001.

3.3.2 Cell drug treatment

I also sought to determine the optimal concentration of 1-NM-PP1 (NM) for treating the AS cells using the alamarBlueHS cell viability assay. Similar to the cell viability assay, 250 cells were seeded into five 96-well clear bottom plates, and cells were initially measured at 0 and 24 hours. At 36 hours, the remaining three plates were treated with different concentrations of NM (5 μM, 7.5 μM and 10 μM) or DMSO (vehicle). The viability of parental and AS cells was then measured immediately after drug (or vehicle) addition (36 h), 12 hours after drug addition (48h total time), and 36 hours after drug addition (72 h total time). All concentrations affected the growth of the CDK12AS, CDK13AS, and CDK12/13AS cells. 5.0 μM and 7.5 μM have no significant effect on the HEK293 parental cell line, in agreement with the same analysis carried out previously on CDK12AS HEK293 made using ectopic expression of Cas 9 and guides from plasmids (Tellier, NAR 2020). In line with Tellier *et al*. [77], 7.5μM was chosen as the concentration for specific inhibition.

3.4 Establishing a quantitative reverse transcription PCR (RT-qPCR) assay for U2 pre-snRNA 3' end formation

An RNase protection assay was previously used to measure the efficiency of 3' end formation of U2 pre-snRNA. This technique uses a complementary radiolabelled RNA probe to hybridise and protect the pre-snRNA and complement from RNase/T1 digestion. The amount of the protected radioactive product of the correct size allows quantitation of 3' end formation [303,304,335].

To eliminate the use of radioactivity, Dr Joanna Guiro, a previous DPhil student in the lab, and I optimised an RT-qPCR technique to measure pre-snRNA and snRNA readthrough quantity (Fig. 3.3A). Cells were treated with the respective drug/vehicle, after which they were harvested and centrifuged. The cell pellet was then lysed and treated with DNase and proteinase K, and total RNA was extracted using phenol-chloroform. We initially synthesised the complementary (cDNA) using random primers. However, despite trying different conditions, the positive controls failed (data not shown). We therefore designed gene-specific primers to quantitatively amplify different regions of the U2 snRNA gene transcript. The primers were designed according to the types of transcripts produced, i.e. pre-snRNA and readthrough snRNA. qPCR was subsequently used to amplify and analyse the level of pre-snRNA 3' end processing. (Fig. 3.3A).

To validate the technique, I inhibited CDK9 by treating CDK9AS cells [356] with 15 μM 1-NA-PP1 for 1 hour, as the CDK9 inhibitor DRB has been shown to affect pre-U2 snRNA 3' end formation in 1 hour [356]. 1-NA-PP1 was used instead of 1-NM-PP1 as the active site of CDK9AS contains the larger amino acid, alanine. 1-NA-PP1 contains an extra methyl group, which maximises its binding in an active site pocket

containing alanine [355], and 1-NA-PP1 is effective in inhibiting CDK9AS [356]. Cells were also treated with DRB as a positive control. The RT-qPCR showed a significant increase in readthrough snRNA produced when CDK9AS is inhibited or cells are treated with DRB (Fig. 3.3B). This agrees with previously published data demonstrating that CDK9 inhibition causes failure of snRNA 3' box recognition, resulting in a lack of nascent snRNA gene transcript cleavage to produce pre-snRNA. The effect of CDK9 inhibition on pre-snRNA 3'end processing provides a positive control for future experiments.

3.5 CDK12 and CDK13 are redundant for pre-U2 snRNA 3' end formation.

As mentioned previously, inhibition of CDK9 has been shown to affect the 3' end processing of snRNA genes [304]. In addition, Dr Guiro had previously tested the roles of analog-sensitive transcriptional CDKS in U2 pre-snRNA 3' end processing and highlighted the possible roles of CDK12 and CDK13 [358]. However, these cell lines were found to be puromycin-resistant. Therefore, I wanted to test the effect of drug treatment on the newly-created analog-sensitive CDK12 and CDK13 cell lines to check the roles of these kinases in snRNA gene expression.

CDK12AS, CDK13AS, CDK12/13AS, and CDK13/12AS cells were treated with 7.5 μM 1-NM-PP1 for an hour to assess their effect on U2 pre-snRNA 3' end processing using RT-qPCR (Fig. 3.3B). Inhibition of CDK12 or CDK13 independently does not affect the production of pre-snRNAs and does not affect the recognition of the 3' box. However, inhibiting CDK12 and CDK13 together results in a 9-fold increase in

the readthrough snRNAs produced, indicating pre-snRNA 3' end misprocessing. Drug treatment of either the CDK12/13AS or CDK13/12AS cell line increases the level of snRNA readthrough, suggesting that the order of kinase mutation to create the double cell line did not affect the reaction to inhibition (Fig. 3.3B). DRB, a positive control, causes an approximately 8-fold increase in readthrough (Fig. 3.3B).

Inhibition of both CDK12 and CDK13 causes an increase of pre-snRNA on both double AS cell lines, although this was not statistically significant for the CDK12/13 cell line. I determined the ratio of readthrough RNA to pre-snRNA produced to confirm that this increase in readthrough snRNA was not simply a result of pre-snRNA accumulation. The ratio of readthrough snRNA to pre-snRNA also indicates a significant increase in readthrough snRNA, consistent with previous results (Fig. 3.3B). This suggests that the observed readthrough is due to snRNA 3' end misprocessing and not solely due to increased transcription.

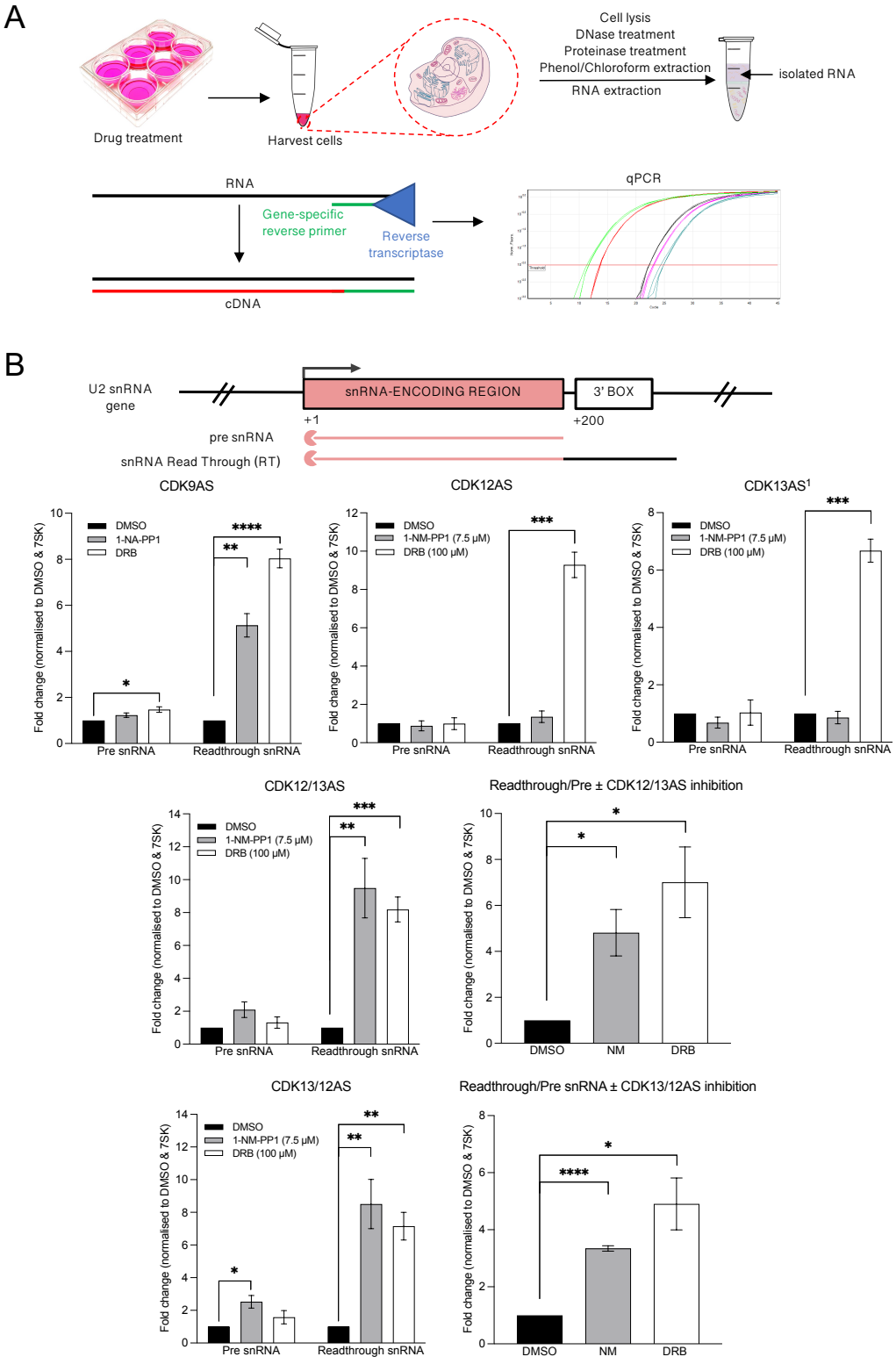


Figure 3.3. CDK12 and CDK13 are redundant for snRNA 3' end processing

A. Scheme of the reverse transcription qPCR. Cells are treated for 1 hour and harvested in media. The cell pellet is then lysed and treated with DNase and Proteinase K, and total RNA is isolated using phenol-chloroform extraction. Complementary DNA is synthesised using gene-specific reverse primers, and pre-

and readthrough snRNA levels are analysed using qPCR. **B.** Quantification of the pre-snRNA and readthrough snRNA transcripts produced after treating analog sensitive CDKs with drug or drug vehicle (DMSO) for an hour. DRB, which inhibits CDK9, is a positive control for loss of 3' box function. Transcripts were normalised to DMSO and the ratio to the transcribed 7SK transcript was determined. n≥3 biological repeats. Statistical test = two-tailed unpaired t-test, * = p <0.05, ** = p<0.01, *** = p <0.001, **** =p <0.0001.

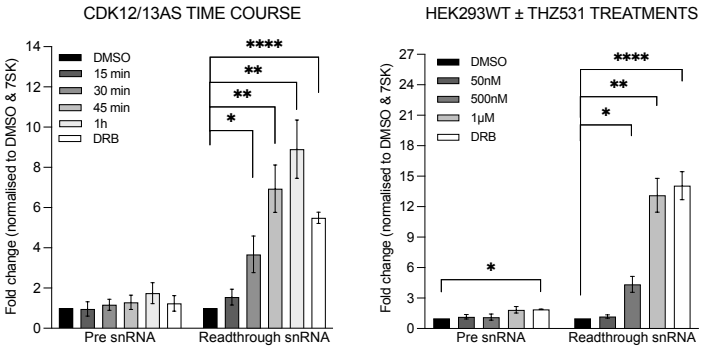
The time course of CDK12AS/13AS inhibition indicates that 30 minutes is sufficient to affect 3' box recognition and produce more readthrough snRNA (Fig. 3.3C). I also inhibited the wild-type parental cell line, Human Embryonic Kidney (HEK)-293 cells, with a range of concentrations of the well-known CDK12 and CDK13 inhibitor, THZ531, for 1 hour (Fig. 3.3C). Treating cells with 500nM and 1μM of THZ531 causes a 4-fold and 13-fold increase in readthrough snRNA transcripts, respectively. This further confirms the finding that although inhibiting CDK12 and CDK13 independently does not affect 3' box recognition, inhibiting both kinases leads to a defect in pre-U2 snRNA 3' end processing. Thus, either activity is sufficient for efficient 3' end formation of pre-U2 snRNA, indicating that CDK12 and CDK13 are functionally redundant in this process.

3.6 Effect of CDK7 inhibition on U2 snRNA 3'end formation

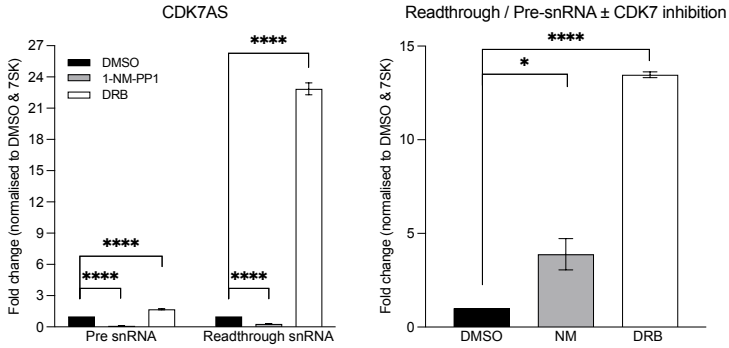
CDK7 phosphorylates Ser7 of the Pol II CTD, and CTD Ser7P plays a role in snRNA expression as it may help to recruit the integrator cleavage module (IntS9/11), which consequently causes cleavage of the nascent snRNA just upstream from the 3' box [96,100]. Prior to my arrival at the lab, Prof. Murphy and Dr Ivan Ferrer-Vicens generated

and characterised a CDK7 analog-sensitive cell line [60]. I inhibited CDK7AS cells with 7.5 μM NM to investigate the role of CDK7 in U2 pre-snRNA 3' end processing (Fig. 3.3D). Initial analysis of the RT-qPCR indicates that inhibition of CDK7 does not cause an increase in readthrough snRNA (Fig. 3.3D). However, a significant decrease in the levels of pre-snRNA is observed. A reduction of pre-snRNA levels is likely to mean lower transcription levels, and this is also observed on protein-coding genes when CDK7 is inhibited [60]. When I deteremined the ratio of readthrough snRNA to pre-snRNA, there is an approximately 4-fold increase in the level of readthrough snRNA relative to pre-snRNA produced, which supports the results of previous studies implicating CDK7 in snRNA 3' end processing [96,97]. These results suggest that CDK7 inhibition affects the 3' end processing of snRNA gene transcripts and impairs transcription. Little effect was seen when wild-type parental cells were treated with various concentrations of 1-NM-PP1 for 1 hour, indicating that the effects observed with all AS cells is due to inhibiting specific kinases (Fig.3.3E).

C



D



E

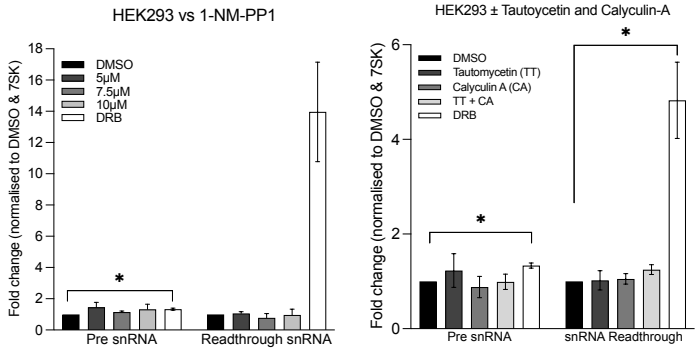


Fig 3.3 **C.** Time course of inhibition of CDK12 and CDK13 together with 7.5 μM of 1-NM-PP1 or DMSO up to 1 hour and quantification of transcripts produced after treating the HEK293 parental cell line with THZ531, a CDK12 and CDK13 inhibitor, for an hour. **D**. RT-qPCR analysis of pre-snRNA, readthrough (RT), and RT/Pre-snRNA transcripts after inhibition of CDK7AS with 7.5 μM of 1-NM-PP1 or DMSO for an hour. **E.** Treatment of HEK293 parental wild-type cells with different concentrations of 1-NM-PP1 for an hour and quantification of transcripts produced after treating the HEK293 parental cell line with PP1 and PP2A inhibitors. DRB acts as a positive control for loss of 3' box function. Transcripts were normalised to DMSO and the ratio to the transcribed 7SK transcript was determiend. n≥3 biological replicates,

error bars = SEM, Statistical test = two-tailed unpaired t-test, * =p <0.05, ** = p<0.01, *** = p <0.001, **** =p <0.0001.

3.7 Inhibition of Protein phosphatase (PP) 1 or 2A does not affect pre-U2 snRNA 3' end formation.

PP1 and PP2 are well-known serine/threonine phosphatases that play a role in various cellular processes, including regulating protein-coding gene expression [165,175,176,356,359,360]. PP1 and PP2A are also known to dephosphorylate some of the targets of CDK9 [175,361,362]. It was previously shown that the inhibition of PP1 with Tautomycetin (TT) causes a termination defect on protein-coding genes, and inhibition of PP2A with Calyculin-A (CA) causes an RNA processing defect on protein-coding genes [356,360]. To determine whether PP1 and PP2 play any role in the expression of snRNA genes, I inhibited wild-type HEK293 with 25nM TT and 2.5nM CA for 30 minutes. I also inhibited both phosphatases together to check for possible redundancies in regulating the 3' end processing of pre-snRNA. RT-qPCR analysis indicates that the inhibition of PP1 and PP2A independently or together does not affect pre-U2 3' end formation (Fig3.3E).

3.8 Summary & Discussion

In addition to regulating protein-coding genes, CDKs play a significant role in the expression of snRNA genes. A major limitation in studying the role of these kinases in human gene expression is the availability of selective inhibitors. Here, I have used the analog-sensitive kinase system developed by Shokat and colleagues [110] to target

transcriptional kinases CDK12 and CDK13 independently and together, using the bulky ATP analog 1-NM-PP1 and to inhibit CDK7 and CDK9.

Previous studies have shown that multiple cis-acting sequences and trans-acting factors mediate the recognition of the 3' box [267,268,303]. Egloff *et al*. also demonstrated that Pol II CTD Ser7 phosphorylation and an adjacent Ser2 are required to recruit the integrator cleavage module (ICM), which cleaves nascent snRNA [96]. CDK7 is recognised as the primary kinase for Ser7 phosphorylation, which would explain the effect on 3' box recognition when CDK7 is inhibited. Inhibition of CDK7 also causes a dramatic loss of pre-snRNA, indicating a reduction of transcription. This agrees with previous studies where the mutation of every Ser7 in the Pol II CTD to alanine caused a decrease in U2 snRNA gene transcription[96,97,100].

CDK9 is the major CTD Ser2 kinase [46,95,98,302,363–365]. However, Ser2 can also be phosphorylated by CDK12 and CDK13 [54,75,77,79,80]. I have shown here that inhibiting CDK12 and CDK13 together but not independently leads to failure to recognise the 3' box. Still, it is unclear whether CDK12 and CDK13 affect the snRNA 3' end processing through phosphorylation of CTD Ser2, other CTD residues, or other substrates, particularly since CDK9 should still be functional.

It has been shown that the 3' box is only recognised if transcription is initiated from an snRNA-specific promoter, indicating an obligate coupling of the promoter to the 3' box [267,268]. Additionally, Fan *et al*. [75] showed that phosphorylation of PSE-transcription factor (PTF)-α, the large subunit of the PSE-binding complex, is decreased upon CDK12 and CDK13 inhibition [75]. This suggests that CDK12 and CDK13 may modulate 3' box recognition by phosphorylating promoter-binding

proteins at the 5' end of snRNA genes. Immunoprecipitation of PTFα followed by mass spectrometry could provide a better understanding of how the inhibition of these kinases may regulate the interaction of key factors involved in pre-snRNA 3' end formation. The significant increase in pre-snRNA levels upon CDK13/12AS cell inhibition may reflect an effect on transcription. This increase can also be seen when the CDK12/13AS cells are inhibited, although this is not statistically significant. The DRB-treated positive controls in the treatment of CDK9AS cells, HEK293 with THZ531 and HEK293 with PP1/2A inhibition, also showed increased levels of products using the pre-snRNA primer. As this pre-snRNA increase is inconsistent across all positive controls, it may reflect that CDK9 inhibition using DRB has variable effects on transcription.

The inhibition of phosphatases has been shown to affect the termination of transcription of protein-coding genes [165,175,176,356,359,360,366,367]. However, my results indicate that the loss of these phosphatases independently or together does not affect the 3' end processing of U2 snRNA genes. Interestingly, inhibition of PP2A reverses the effect of CDK9 inhibition on the termination of transcription of protein-coding genes [356]. A future direction for this project will be to investigate if the lack of 3' box recognition caused by CDK9 or CDK12 and CDK13 inhibition can also be reversed by PP1 and/or PP2A inhibition.

3.9 Limitations of these experiments

Upon generation of the cell lines, their growth and proliferation of the cell were monitored using an alamarBlue HS viability assay. This assay recognises

metabolically active cells by testing the reducing power of the mitochondria. Resazurin, the active component of the assay, is blue and non-fluorescent. The reducing activity of metabolically active mitochondria reduces resazurin to resorufin, a pink and highly fluorescent compound subsequently detected with a fluorimeter. While this assay is highly reliable, the metabolic activity of a cell does not always directly correlate with cell viability, which could be a limitation of this method. This could be improved by incorporating an alternative technique with this assay. For example, the 5-Bromo-2'-Deoxyuridine (BrdU) proliferation assay directly measures a viable cell's DNA synthesis by incorporating a thymidine analogue, BrdU, into newly synthesised DNA. Cells containing BrdU are subsequently detected using anti-BrdU antibodies through flow cytometry [368]. In addition, a negative control could be incorporated into this experiment to provide a baseline for the assay. Cells could be treated with hydrogen peroxide, which causes oxidative stress or detergents such as Triton-X to disrupt cell membranes and reduce cell viability [369,370]. However, the alamarBlue HS assay did not detect any growth defects in the cell lines created.

Chapter 4 | Inhibition of transcriptional CDKs affects Pol II CTD phosphorylation on U2 snRNA genes.

4.1 Background

The heptapeptide repeats at the C-terminal of Pol II CTD undergo post-translational modifications, which generate a 'code' for the recruitment of transcriptional and co-transcriptional factors that direct the expression of Pol II-transcribed genes [54,98]. Phosphorylation is the most studied post-translation modification, with many potential combinations of phosphorylated tyrosine, serine, and threonine [54,98]. Despite the similarities between protein-coding and snRNA genes, the roles of many transcriptional (t)CDKs in Pol II CTD phosphorylation in the expression of snRNA genes remain largely unexplored. CDK12 and CDK13 are redundant for U2 snRNA 3' end processing (Chapter 3). This chapter details experiments investigating the roles of CDK12, CDK13 and other tCDKs in Pol II CTD phosphorylation using ChIP.

4.2 Inhibition of CDK7AS, CDK9AS and CDK12AS/13AS affects the profile of Pol II RPB1 across U2 snRNA genes.

To further understand the roles of CDK12, CDK13, CDK7 and CDK9 on Pol II actively transcribing U2 snRNA genes, the respective analog-sensitive kinases were inhibited with 7.5 μM of 1-NM-PP1 (CDK7AS, CDK12AS, and CDK13AS) or 15 μM of 1-NA-PP1 (CDK9AS) for 30 minutes and the Pol II (RPB1 NTD) profile was analysed using ChIP-qPCR (Fig. 4.1). ChIP-qPCR data shows that the inhibition of CDK7, the main Ser7 kinase [56–60,371,372], causes an approximate 50% loss of Pol II across the promoter, 3' box, and CT regions of the gene, indicating a loss of transcription. This result agrees with previously published studies where Ser7AS mutation inhibition causes a loss in transcription [60]. Upon inhibition of CDK9AS cells, the ChIP-qPCR profile showed no significant changes in the Pol II profile across the promoter and the 3' box regions. However, there is an increase in the levels of Pol II across the CT-rich region of the genes. The separate inhibition of CDK12AS or CDK13AS for 30 minutes does not affect the profile of actively transcribing Pol II on U2 snRNA genes. However, an increase in Pol II is observed at the promoter region, 3' box and CT-rich, upon the inhibition of CDK12AS and CDK13AS together, further exhibiting the redundancy of these kinases in the expression of U2 snRNA genes.

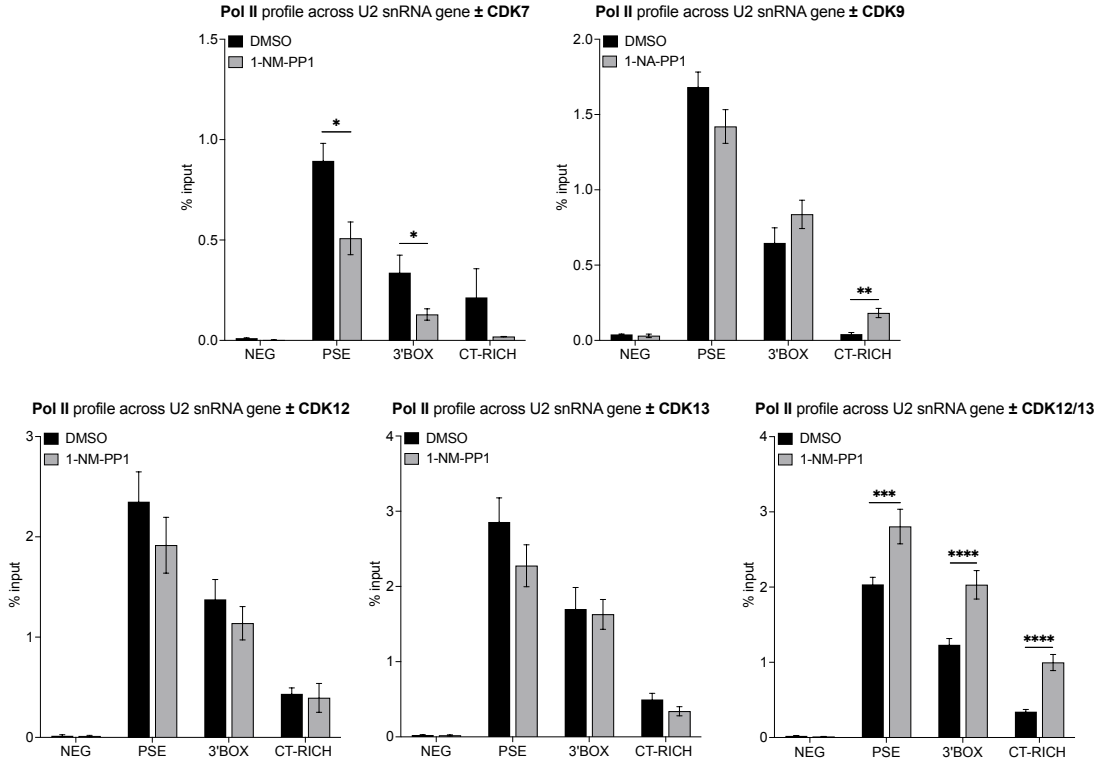


Figure 4.1. Inhibition of CDK7AS, CDK9AS and CDK12AS/13AS affects the profile of Pol II RPB1across U2 snRNA genes. ChIP-qPCR of Pol II across U2 snRNA gene upon inhibition of analog-sensitive transcriptional kinases with 7.5μM 1-NM-PP1, 15μM 1-NA-PP1 or DMSO (negative control) for 30 minutes. n≥3 biological repeats. Statistical test = two-tailed unpaired t-test, * = p <0.05, ** = p<0.01, *** = p <0.001, **** =p <0.0001.

4.3 Phosphorylation of the Pol II CTD upon inhibition of analog-sensitive kinases

To better understand the roles of tCDKs on the expression of snRNA genes, the AS cell lines were inhibited for 30 minutes, and the phosphorylation profiles of the key CTD marks – Tyr1, Ser2, Ser5, and Ser7 were assessed by ChIP-qPCR. The profile of these marks has also been analysed in terms of their ratio to Pol II (CTD phospho-residue/ Pol II) to account for the possible redistribution of Pol II across the genes upon kinase inhibition.

4.4 The pTyr1 ChIP-qPCR profile changes upon CDK12AS and CDK13AS inhibition

Although the role of pTyr1 in the transcription of protein-coding genes is unclear, this mark is highest at the 5' end of mammalian protein genes [373–375]. On the human U2 snRNA genes, pTyr1 peaks over the 3' box (Fig 4.2), suggesting that it may play a role in RNA 3' processing. The inhibition of CDK12AS causes a loss of pTyr1 relative to Pol II, decreasing pTyr1/Pol II levels by at least 50% across all assessed regions of the U2 snRNA gene (Fig 4.2). Inhibition of CDK13AS instead causes a slight increase in pTyr1/Pol II level across the PSE and 3' box region, with a significant loss of pTyr1 at the CT-rich region. The inhibition of CDK12AS and CDK13AS together causes a further loss of pTyr1/Pol II across all assessed regions of the gene compared to when inhibiting CDK12AS alone. This suggests that CDK13 cannot compensate for the absence of CDK12 in phosphorylating Tyr1 of the CTD at the promoter and 3' box region. Still, both kinases are required for Tyr1 phosphorylation at the CT-rich region. Interestingly, the drastic loss of pTyr1 at the CT-rich region when both kinases are inhibited is only observed when the pTyr1 level is expressed with ratio to Pol II, as Pol II accumulates at this region when both kinases are inhibited (Figure 4.1)

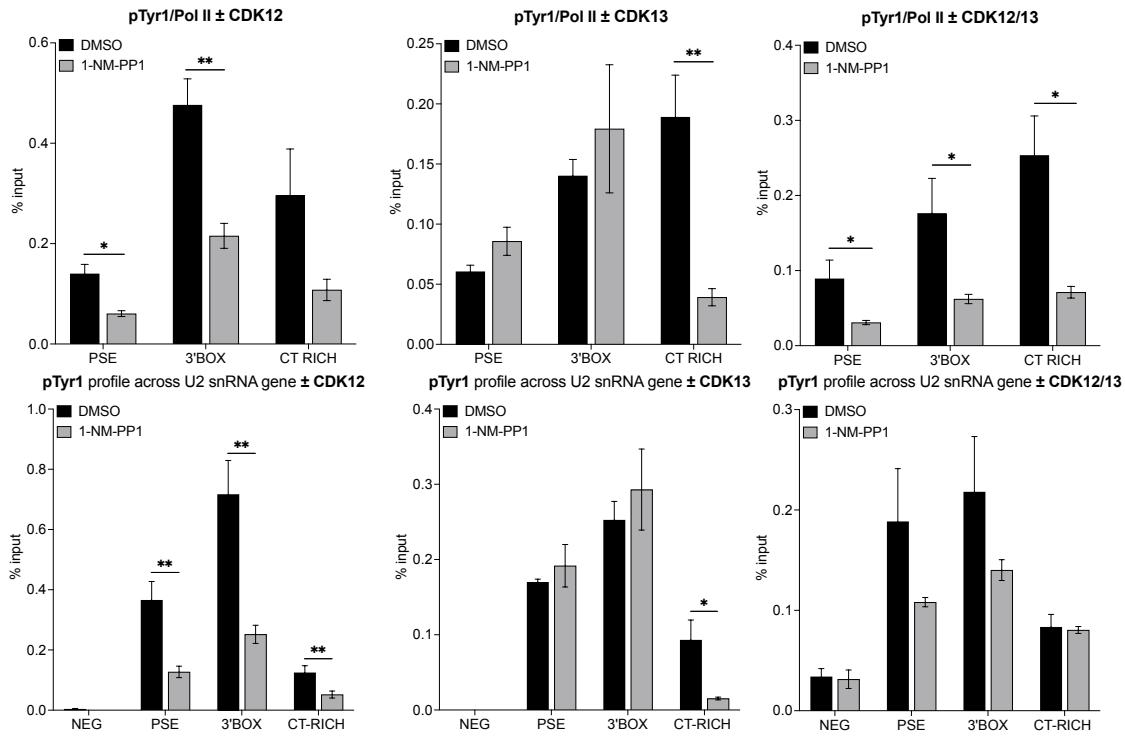


Figure 4.2. The pTyr1 ChIP-qPCR profile changes upon CDK12AS and CDK13AS inhibition. (Top) ChIP-qPCR of the ratio of pTyr1 to Pol II across the U2 snRNA gene after treatment of CDK12 and CDK13 AS cells with 7.5μM 1-NM-PP1 or DMSO for 30 minutes. (Bottom) ChIP-qPCR of pTyr1 across U2 snRNA gene upon inhibition of respective AS cells. n≥3 biological repeats. Statistical test = two-tailed unpaired t-test, * = p <0.05, ** = p<0.01.

4.5 CDK9, CDK12 and CDK13 phosphorylate Ser2 of the Pol II CTD on U2 snRNA genes

pSer2 has a major role in cleavage and polyadenylation of transcripts from protein-coding genes [77,376–378] and has been shown to help recruit the Integrator complex to cleave transcripts from human snRNA genes at the 3' end [96,330] (Chapter 1). Inhibition of CDK12AS and CDK13AS independently causes some loss of pSer2/Pol II across assessed regions of the gene (Fig. 4.3A). However, there is a further decrease in pSer2/Pol II across the 3' box and CT-rich region of the U2 snRNA gene

when both kinases are inhibited, which suggests that both kinases are required for pSer2 at the 3' end of the gene. Surprisingly, there is an increase in pSer2/Pol II at the promoter region when both kinases are inhibited together.

CDK7 is primarily known for phosphorylating Ser5 and Ser7 on Pol II CTD [56,57], but it has also been implicated in the phosphorylation of pSer2 at the 3' end of protein-coding genes such as c-Myc and GAPDH [147]. Therefore, CDK7AS cells were inhibited to assess whether CDK7 may influence snRNA 3' end-processing through Ser2 modification (Fig 4.3B). ChIP-qPCR analysis shows a moderate Ser2/Pol II loss at the PSE and 3' box regions. However, there is an increase in pSer2/Pol II at the CT-rich region. This may indicate a role of CDK7 in regulating Ser2 kinases or recruiting phospho-Ser2 phosphatases to the 3' end of the snRNA gene.

CDK9 is recognised as the main Ser2 kinase [91,379] and is implicated in the 3' end processing of transcripts from U2 snRNA genes as small molecule CDK9 inhibitors, including DRB, cause a major defect in 3' box recognition [96,303,304]. Therefore, the pSer2 profile was also analysed following the inhibition of CDK9AS (Fig. 4.3B). CDK9AS inhibition significantly reduces pSer2/Pol II levels by over 70% at the 3' box and CT-rich region of the U2 snRNA gene. This marked decrease at the 3' end of the gene confirms the critical role of CDK9 in phosphorylating Ser2 to activate U2 pre-snRNA 3' end processing [96,275,303,304].

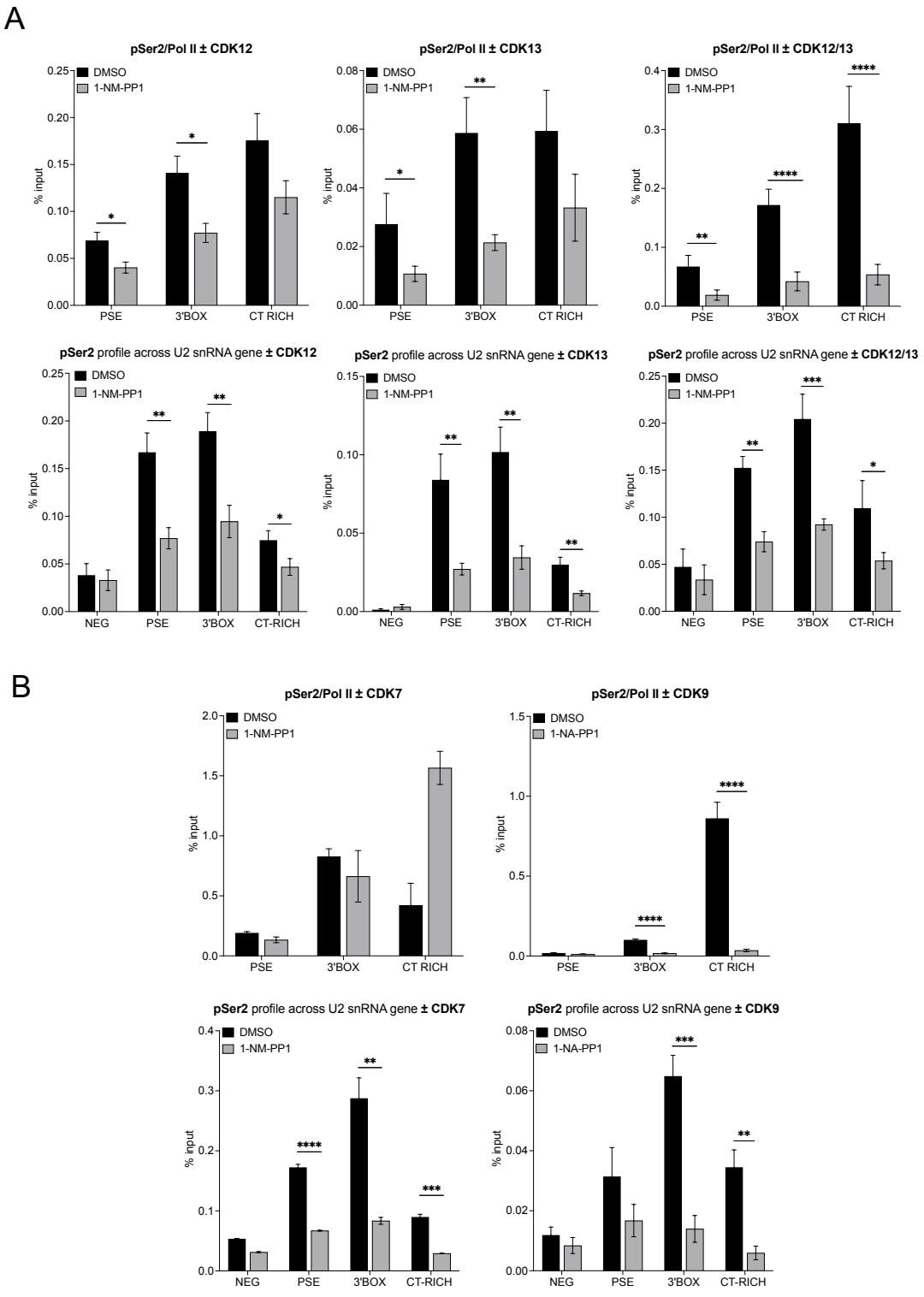


Figure 4.3 CDK9, CDK12, and CDK13 phosphorylate Ser2 of the Pol II CTD on U2 snRNA genes. **A.** (Top) ChIP-qPCR of the pSer2 to Pol II ratio across the U2 snRNA gene after treatment of CDK12AS and CDK13AS with 7.5μM 1-NM-PP1 or DMSO for 30 minutes. (Bottom) ChIP-qPCR of pSer2 across U2 snRNA gene upon inhibition of respective analog-sensitive cells **B.** (Top) ChIP-qPCR of the ratio of pSer2 to Pol II

across the U2 snRNA gene after treatment of CDK7 and CDK9AS with 7.5μM 1-NM-PP1 or 15μM 1-NA-PP1, or DMSO for 30 minutes. (Bottom) ChIP-qPCR of pSer2 across U2 snRNA gene upon inhibition of respective AS cells. n≥3 biological repeats. Statistical test = two-tailed unpaired t-test, * = p <0.05, ** = p<0.01, *** = p <0.001, **** =p <0.0001.

4.6 Inhibition of CDK7, CDK12 and CDK13 affects the pSer5 ChIP-qPCR profile on U2 snRNA genes

pSer5 plays a major role in the capping of Pol II transcripts, which can enhance downstream RNA processing events, including snRNA gene 3' box recognition [88,146,380] (see section 1.3.1 of the Introduction). In addition to CDK7, the major mammalian Ser5 kinase [56–60,371,372], CDK12 and CDK13 have been identified as potential Ser5 kinases on protein-coding genes [54,75]. Therefore, these kinases were inhibited to analyse their possible roles in snRNA gene expression. ChIP-qPCR analysis indicates a loss of pSer5/Pol II at the CT-rich region upon CDK12AS inhibition (Fig. 4.4A), while the inhibition of CDK13AS causes a more evident loss of pSer5 across the gene, with a significant loss of pSer5/Pol II at the 3' box region (Fig. 4.4A). The inhibition of both kinases causes a reduction in pSer5/Pol II levels across U2 snRNA genes. The accumulation of Pol II across the gene upon CDK12AS/13AS double inhibition makes it challenging to detect the loss of pSer5 at the PSE and 3' box regions when it is not normalised to Pol II levels (Fig. 4.4A). However, the pSer5/Pol II levels at the 3' box region upon CDK12AS/13AS inhibition do not reflect the loss of pSer5/ Pol II caused by inhibiting CDK13AS alone. It is possible that other Ser5 kinases (e.g. CDK7 and/or CDK9 [67]) may compensate when CDK12 and CDK13 are inhibited together.

During snRNA gene expression, pSer5 is dephosphorylated after initiation to allow the recruitment of integrator factors and downstream processing of 3' end processing [99]. Therefore, the pSer5 profile on U2 snRNA genes upon CDK7AS inhibition was also analysed (Fig. 4.4B). ChIP-qPCR indicates an approximate 60% loss in Ser5 levels at the promoter region and at least 80% loss at the 3' box and CT-rich regions when CDK7AS is inhibited. However, these losses are largely due to changes in Pol II levels as the pSer5/Pol II profile shows a minimal loss of pSer5 at the promoter, an approximate 50% reduction at the 3' box region and an increase at the CT-rich region.

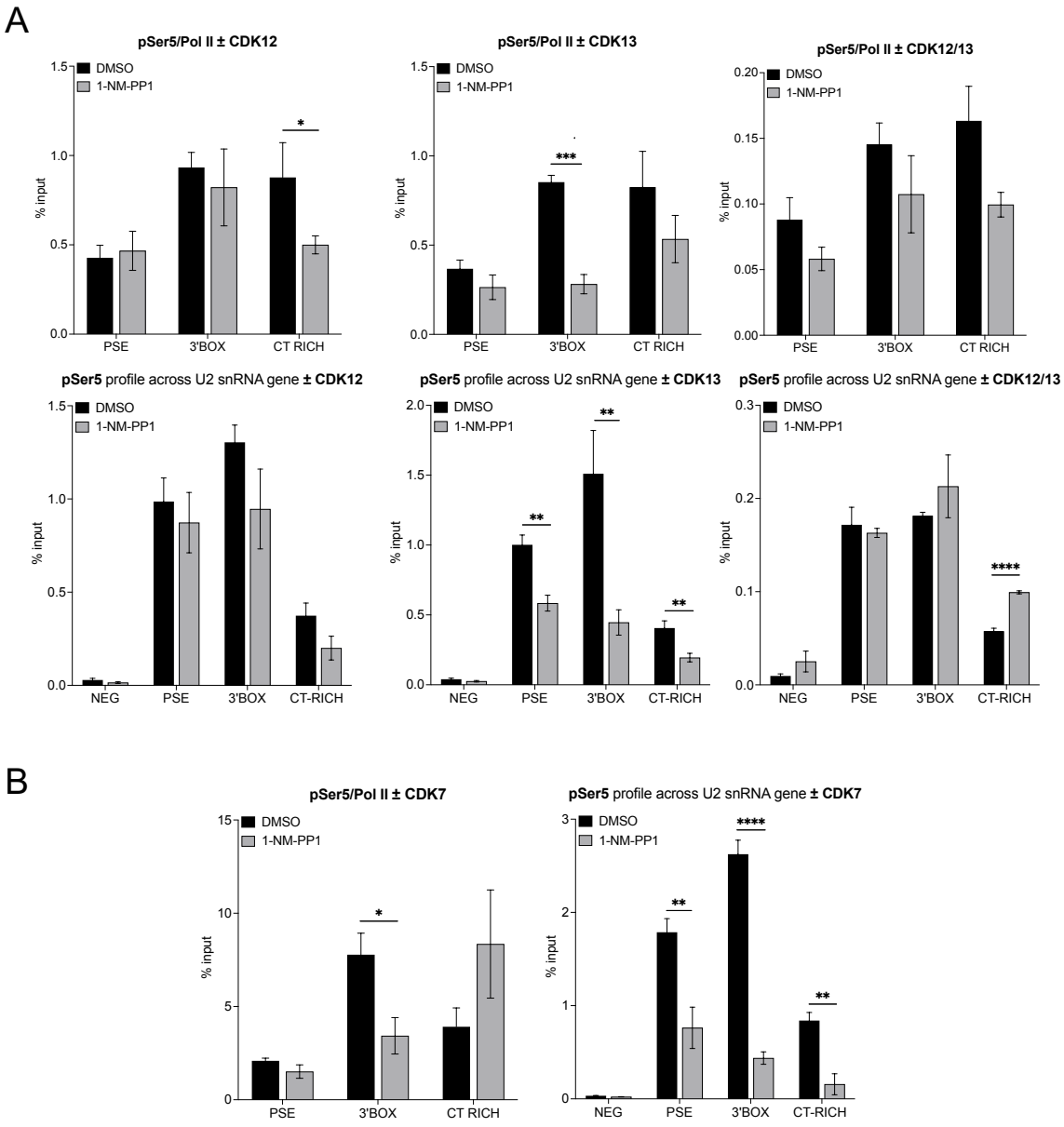


Figure 4.4 Inhibition of CDK7, CDK12 and CDK13 affects the pSer5 ChIP-qPCR profile on U2 snRNA genes. A. (Top) ChIP-qPCR of the pSer5 to Pol II ratio across the U2 snRNA gene after treatment of CDK12AS and CDK13 AS with 7.5μM 1-NM-PP1 or DMSO for 30 minutes. (Bottom) ChIP-qPCR of pSer5 across U2 snRNA gene upon inhibition of respective AS cells **B.** (Top) ChIP-qPCR of the ratio of pSer2 to Pol II across the U2 snRNA gene after treatment of CDK7 and CDK9AS with 7.5μM 1-NM-PP1 or 15μM 1-NA-PP1, or DMSO for 30 minutes. (Bottom) ChIP-qPCR of pSer2 across U2 snRNA gene upon inhibition of respective AS cells. n≥3 biological repeats. Statistical test = two-tailed unpaired t-test, * = p <0.05, ** = p<0.01, **** =p <0.0001.

4.7 Inhibition of CDK7AS, CDK9AS, and CDK12AS/13 AS affects pSer7 profiles on U2 snRNA genes

Ser7 phosphorylation of the Pol II CTD is required for snRNA gene expression and transcript 3' end processing and has been shown to help recruit the Integrator complex [96,97,100]. Therefore, CDK7AS, CDK9AS, and CDK12AS/13 AS cells were inhibited to understand how these kinases may regulate Ser7 phosphorylation across the U2 snRNA gene (Fig. 4.5). CDK12AS and CDK13AS inhibition together causes a moderate decrease in pSer7/Pol II levels at the promoter region, with an approximate 50% decrease in pSer7 at the 3' box, and a more significant loss at the CT-rich regions. The inhibition of CDK9AS indicates no changes in pSer7/Pol II levels observed at the promoter, but the pSer7/Pol level is reduced by approximately 50% at the 3' box region and is significantly lost at the CT-rich regions, comparable to CDK12AS/13AS inhibition (Fig.4.5). The analysis of CDK7AS inhibition indicates a near-complete loss of pSer7 at the promoter region and an approximate 70% decrease at the 3' box. However, there is an increase in pSer7/Pol II at the CT-rich region. This may be explained by the near-total loss of Pol II at the CT-rich region upon CDK7AS inhibition. Therefore, the presence of a small amount of residual pSer7 will show an increase when normalised to Pol II. This is also observed for Ser2/Pol II (Fig. 4.3B) and Ser5/Pol II (Fig. 4.4B) levels upon CDK7AS inhibition.

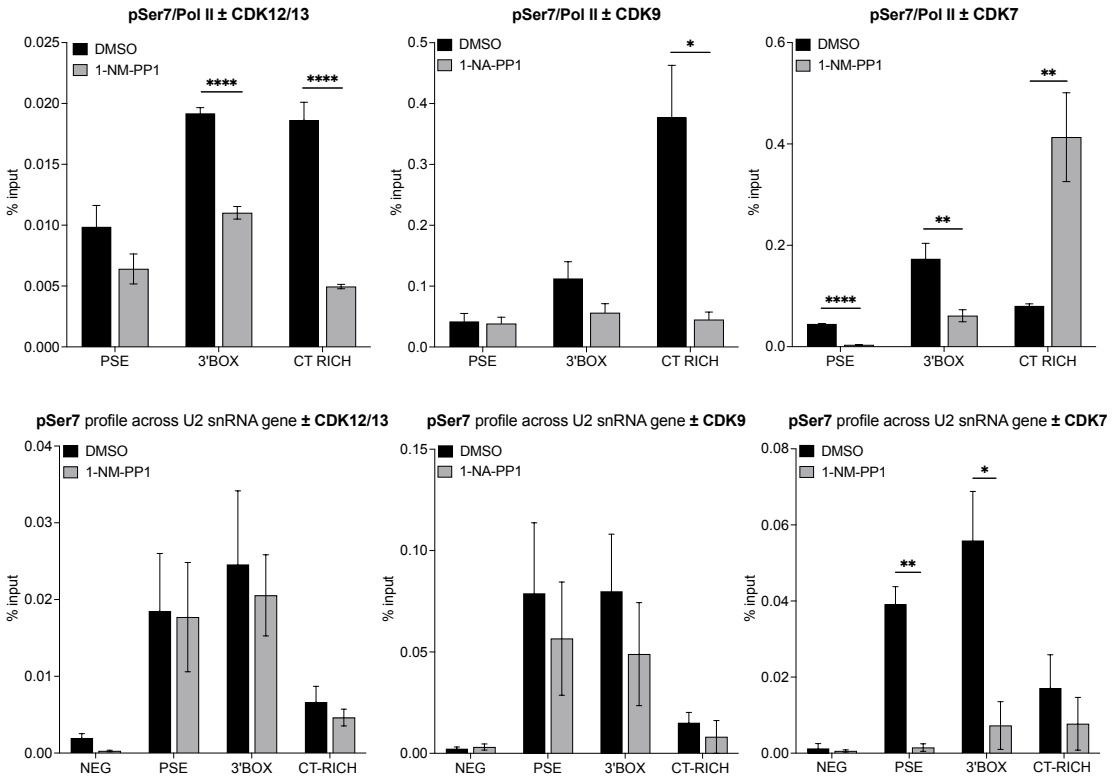


Figure 4.5. Inhibition of CDK7AS, CDK9AS, and CDK12AS/13 AS affects pSer7 profiles on U2 snRNA genes. (Top) ChIP-qPCR of the ratio of pSer7 to Pol II across the U2 snRNA gene after treatment of respective AS cells with 7.5μM 1-NM-PP1 or 15μM 1-NA-PP1, or 1-DMSO for 30 minutes. (Bottom) ChIP-qPCR of pTyr1 across U2 snRNA gene upon inhibition of respective AS cells. n≥3 biological repeats. * = p <0.05, ** = p<0.01.

4.8 Summary and Discussion

The aim of the experiments described in this chapter was to better understand the roles of CTD kinases in the phosphorylation of Pol II CTD residues during the expression of U2 snRNA genes. AS kinase cells were inhibited for a mere 30 minutes to avoid unwanted secondary effects and because it is the minimum time of inhibition of CDK12AS and CDK13 AS cells to observe U2 snRNA gene transcript misprocessing (Chapter 3, Fig. 3.3C). ChIP-qPCR of Pol II upon CDK12AS/13AS inhibition showed an accumulation of Pol II at the promoter, 3' box and termination regions of the gene. This suggests potential stalling of Pol II across these regions, indicating possible defects in elongation, processing and termination. However, further studies employing additional techniques, such as Global Run-On sequencing (GRO-Seq [381]) and mammalian Native Elongation Transcript sequencing (mNET-seq [382]), which track nascent RNA synthesis, would be required to confirm this. While CDK12AS/13AS inhibition mirrors the effect of CDK9 inhibition in the 3' end misprocessing of snRNA gene transcripts (Chapter 3, Fig. 3.3A), the Pol II profile upon CDK9 inhibition shows no Pol II accumulation across the gene (Fig 4.1), which confirms previous studies demonstrating that CDK9 functions as an RNA processing factor in expression of snRNA genes [303]. Additionally, this result suggests that CDK12/13 may have additional and unique roles in the RNA processing of snRNA gene transcripts.

Previous studies showed increased Pol II levels across the body of protein-coding genes when CDK12AS was inhibited with 7.5 μM NM for 30 minutes in HEK293-

CDK12AS cells [77]. In contrast, there are no changes in Pol II levels across the U2 snRNA gene when CDK12AS or CDK13AS are inhibited independently. This demonstrates how these kinases may exert different functions on different gene types, similar to CDK9, a transcription elongation and RNA processing factor for mammalian protein-coding genes [67,95,176,303,383–385].

CDK7AS inhibition causes at least a 50% reduction in Pol II levels across the U2 snRNA gene (Fig. 4.1), which is consistent with the observed result in Chapter 3 (Fig. 3.3C), where CDK7AS inhibition resulted in a significant loss of pre–snRNA synthesis, indicating a loss of transcription. However, CDK7 was inhibited for an hour when investigating the pre-snRNA levels, whereas the CDK7AS cells were inhibited for 30 minutes when checking Pol II occupancy. Thus, it is not clear what level of Pol II loss from the gene is required to cause a near-total loss of pre-snRNA synthesis.

CDK12 and CDK13 are Ser2 plus Ser5 kinases and have been shown to regulate transcriptional processes such as elongation [75,78,323]. However, the role of these kinases in phosphorylating other CTD marks on snRNA genes is understudied. ChIP-qPCR shows that CDK12AS inhibition affects pTyr1 across the U2 snRNA genes, as at least a 50% drop in pTyr1/Pol II levels is observed upon its inhibition. On the other hand, CDK13AS inhibition causes a slight increase in pTyr1/Pol II levels across the PSE and 3’ box and a major decrease in pTyr1 at the CT-rich region of the gene. The inhibition of both kinases together reflects a higher loss in the Tyr1/Pol II profile compared to inhibiting CDK12AS alone, suggesting that although CDK12 may compensate for CDK13 in Tyr1 phosphorylation at the PSE and 3’ box regions, both kinases are required for Tyr1 phosphorylation. Previous studies suggest that

the Integrator complex may be recruited to protein-coding genes through pTyr1 residues and that this phosphorylation helps mediate integrator-dependent termination near promoter regions [386,387]. The present work indicates that inhibiting CDK12AS alone does not cause a significant 3' end processing defect (Chapter 3). However, the inhibition of CDK12/13AS causes a further loss in pTyr1/Pol II levels across the gene. It is possible that the function of CDK12/13 in regulating 3' end processing could also be mediated by Tyr1 phosphorylation on U2 snRNA genes.

ChIP-qPCR analysis also confirmed the role of CDK12 and CDK13 in Ser2 phosphorylation. Inhibiting each kinase independently causes a reduction in Ser2 across all regions, with and without normalisation to Pol II (Fig 4.3). Upon inhibition together, there is a further loss of pSer2/Pol II at the 3' box and CT-rich regions, demonstrating partial redundancy as the activity of either kinase may partially compensate for the loss of the other. The increase of the pSer2 relative to Pol II (at the promoter) when CDK12AS and 13AS are both inhibited may reflect the regulation of other Ser2 kinases (for instance, CDK9 [91,379]) or the loss of a pSer2 phosphatase (e.g. Fcp1 [388,389]) as both kinases cause loss of pSer2/Pol II when inhibited independently. Further ChIP studies will be needed to investigate the pSer2/Pol II profile upon the inhibition of CDK9, CDK12 and CDK13 together and pSer2 phosphatase profiles when CDK12 and CDK13 are inhibited together.

Unlike CDK12AS and CDK13AS, the inhibition of CDK9AS causes a near-total loss of pSer2/Pol II at the 3' box and CT-rich region. Active CDK12 and CDK13 are presumably still present when CDK9AS is inhibited but cannot compensate for the loss of CDK9 activity (Chapter 3, Fig. 3.3A). This begs the question of which kinase

is upstream. As inhibiting CDK9AS gives the most significant loss of pSer2, CDK12 and CDK13 may control Ser2 phosphorylation by regulating CDK9 activity. However, it remains possible that CDK9 regulates the Ser2 phosphorylation activity of CDK12 and CDK13. Alternatively, CDK12 and CDK13 may control 3' end formation through other pathway/s in addition to phosphorylating Ser2. RT-qPCR of snRNA readthrough levels upon inhibition of all three kinases together will be required to confirm this.

CDK7 inhibition causes a decrease of pSer5/Pol II only at the 3' box region. The lack of pSer5 drop at the PSE may be due to the simultaneous loss of pSer7 at the promoter region as CDK7 is the major mammalian Ser7 kinase, [56–60,371,372] and pSer7 recruits RPAP2, a Ser5 phosphatase [99]. Hence, the loss of pSer7 at the PSE may cause less dephosphorylation of Ser5 by RPAP2. Alternatively, CDK9, CDK12 or CDK13 may contribute to the Ser5 phosphorylation [67,75–77,80] in the absence of CDK7, or the presence of CDK7, CDK9, CDK12 or CDK13 may be sufficient for Ser5 phosphorylation at the promoter region.

ChIP-qPCR also showed the loss of pSer7/Pol II across the PSE and 3' box when CDK7 is inhibited, confirming the role of CDK7 in Ser7 phosphorylation of the CTD on the U2 snRNA genes. There is also a total loss of pSer7/Pol II at the PSE, which is not observed when the other kinases are inhibited. This agrees with the current view of snRNA gene expression, where CDK7 phosphorylates Ser7 after initiation [97,99,100,275]. Interestingly, the inhibition of CDK12AS/13AS also decreases pSer7/Pol II levels across the gene, which suggests that their regulation of 3' end processing may involve the recruitment of the integrator cleavage module (ICM), as inhibiting CDK12AS and CDK13AS together causes a loss of both pSer2 and pSer7 at the 3'

box region. However, ChIP-qPCR analysis of pSer7 and IntS9/11 when both kinases are inhibited independently will be needed to confirm this.

CDK9 inhibition causes a loss of Ser7/Pol II at the 3' box and CT-rich regions. CDK9 has been previously shown to phosphorylate Ser7 *in vitro* [57,372] and is required for snRNA 3' end processing [96,275,303,304]. The activity of CDK9 on Ser7 phosphorylation at the 3' end suggests that it is possible CDK9 phosphorylates both Ser2 and Ser7 to recruit the ICM at the 3' end, while CDK7 primarily phosphorylates Ser7 at the promoter region. However, further studies, including ChIP of integrator subunits upon inhibition of these kinases, will be needed to test this.

While these ChIP-qPCR results provide further insight into the roles of CDK12 and CDK13 in CTD phosphorylation during the expression of snRNA genes, additional studies will be required to establish how these kinases control CTD modifications to regulate 3' end formation of snRNA genes.

4.9 Limitations

A limitation of the experiments described in this chapter is the specificity and sensitivity of the various antibodies used. Also, ChIP-qPCR monitors actively transcribing Pol II, and masking the epitope recognised by the antibody by different protein-protein interactions will introduce bias to the profile generated. Conformational changes in phosphorylated residues or neighbouring residues can also affect the effective binding of the antibodies. To mitigate antibody specificity issues, multiple repeats were run, only antibodies that gave a sufficient percentage

input over negative regions were used, and control experiments were carried out where possible.

Chapter 5 | Cdk12 and Cdk13 are redundant for the recruitment of key proteins to human U2 snRNA genes

5.1 Background

Transcriptional kinases such as CDK7 and CDK9 mediate the 3' end processing of snRNA genes through the phosphorylation of the Pol II CTD Ser7 and Ser2, respectively [97,99,100,275,303,304] (see chapter 1.7.3.1 of the Introduction). Modifications of CTD residues lead to the recruitment of factors that control gene expression. For example, phosphorylation of Ser7 by CDK7 and an adjacent Ser2 by CDK9 leads to the recruitment of IntS9 and IntS11 cleavage modules[96]. RT-qPCR analysis after inhibiting analog-sensitive cells has shown that CDK12 and CDK13 are redundant for 3' box-directed processing of transcripts from U2 snRNA genes (Chapter 3). The experiments described in this chapter aimed to investigate the mechanism underlying CDK12 and CDK13 control 3' box recognition by identifying new factors which may play a role in the expression of snRNA genes.

5.2 SPT6 and Cyclin K are present on the U2 snRNA genes.

CRISPR Affinity Purification *in situ* of Regulatory element (CAPTURE) is a pulldown technique which uses guide RNAs to direct biotinylated catalytically dead cas9 to specific regions of a gene, enabling specific pulldown of associated proteins at these regions followed by identification by mass spectrometry [390]. Prior to my

arrival at the lab, Dr Joana Guiro used CAPTURE to pull down proteins associated with the promoter region of the human U2 snRNA gene [320]. Mass spectrometry analysis identified several proteins which were enriched over the DSE and PSE regions of the gene. Several transcription factors known to be associated with snRNA genes were pulled down, such as NELFB, MED23, and IntS3. Other factors that had not been previously shown to be associated with snRNA genes were also identified, such as SPT6 and Cyclin K [320,358]. SPT6 is a histone chaperone which recruits SETD2 to deposit the K3K36me3 mark, [317,319] and a known elongation factor required for the expression of protein-coding genes in yeast [391,392] and humans [77,316,318,393]. Cyclin K is the regulatory cyclin and binding partner for CDK12 and CDK13 [51]. Using ChIP-qPCR, I confirmed the presence of SPT6 and Cyclin K on the U2 snRNA gene in HEK293 cells and showed they have similar gene distributions (Fig. 5.1). Previously published studies indicate that siRNA-mediated knockdown of SPT6 causes a loss of IntS3 from protein-coding genes, promoter enhancers and long noncoding RNA (lncRNA) genes [319.] Bioinformatic reanalysis of this data revealed that SPT6 knockdown also results in a loss of IntS3 across U2 and other Pol II-transcribed snRNA genes [320]. Additionally, a reanalysis of previously published chromatin RNA-seq data indicated that chromatin-associated RNA from snRNA genes is reduced when SPT6 is knocked down, suggesting that SPT6 plays a role in the expression of these genes [320.]

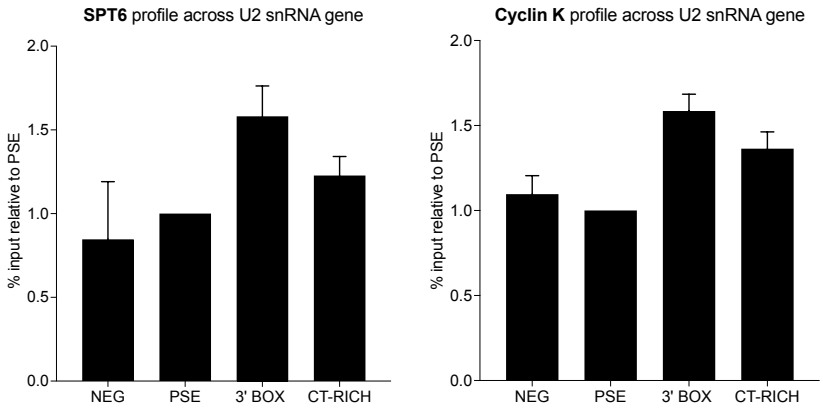


Figure 5.1. SPT6 and Cyclin K are present on the U2 snRNA gene.
ChIP-qPCR analysis of SPT6 and Cyclin K showing the enrichment over the negative region at the 3' box of the U2 snRNA gene in HEK293 cells. % input is normalised to PSE. n=3 biological repeats.

5.3 CDK12 and CDK13 are both involved in SPT6 recruitment to the U2 snRNA gene.

The identification of SPT6 as an snRNA gene-associated protein through CAPTURE, its role in the recruitment of IntS3 to snRNA genes [320], and in the termination of transcription of lncRNA genes[319] suggests a role for SPT6 at the 3' end of snRNA genes. This hypothesis is further supported by the SPT6 ChIP profile, which shows enrichment at the 3' box region. To explore the possibility that CDK12 and CDK13 affect SPT6 recruitment, these kinases were inhibited, and SPT6 association was measured by ChIP-qPCR (Fig. 5.2). CDK12AS inhibition causes a marked decrease in SPT6 recruitment across the gene. However, most of this reduction is due to the loss of Pol II, as SPT6 levels relative to Pol II (SPT6/Pol II) show no change at the PSE and a more modest but not significant decrease at the 3' end of the gene. In contrast, CDK13AS inhibition leads to an increase in SPT6 association across the U2 snRNA gene, both with and without normalisation to Pol II.

Interestingly, inhibiting CDK12 and CDK13 AS cells together causes a marked loss of SPT6 across the gene, particularly at the 3' box region (Fig. 5.2). This loss is also apparent when SPT6 is expressed as a ratio to Pol II and is stronger than the loss observed when CDK12 alone is inhibited. This result may suggest that CDK12 activity compensates for the loss of CDK13 while CDK13 activity only partially compensates for the loss of CDK12.

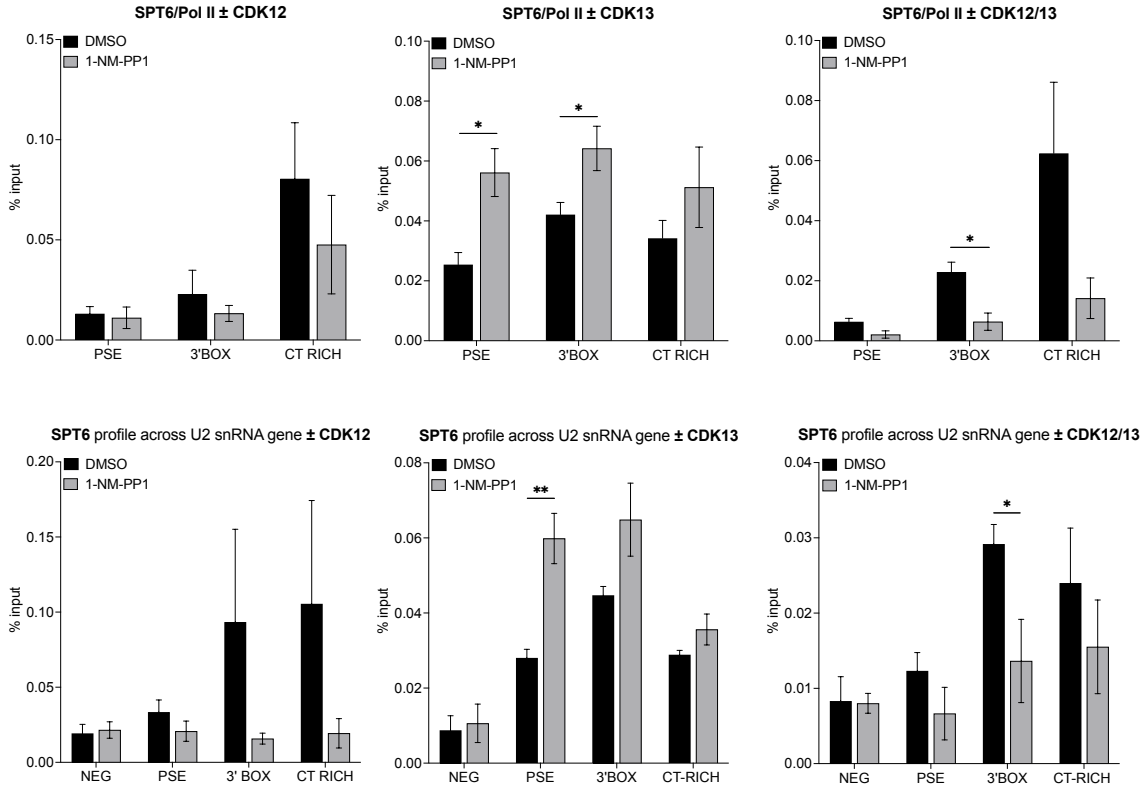


Figure 5.2. CDK12 and CDK13 are both involved in SPT6 recruitment to the U2 snRNA gene.

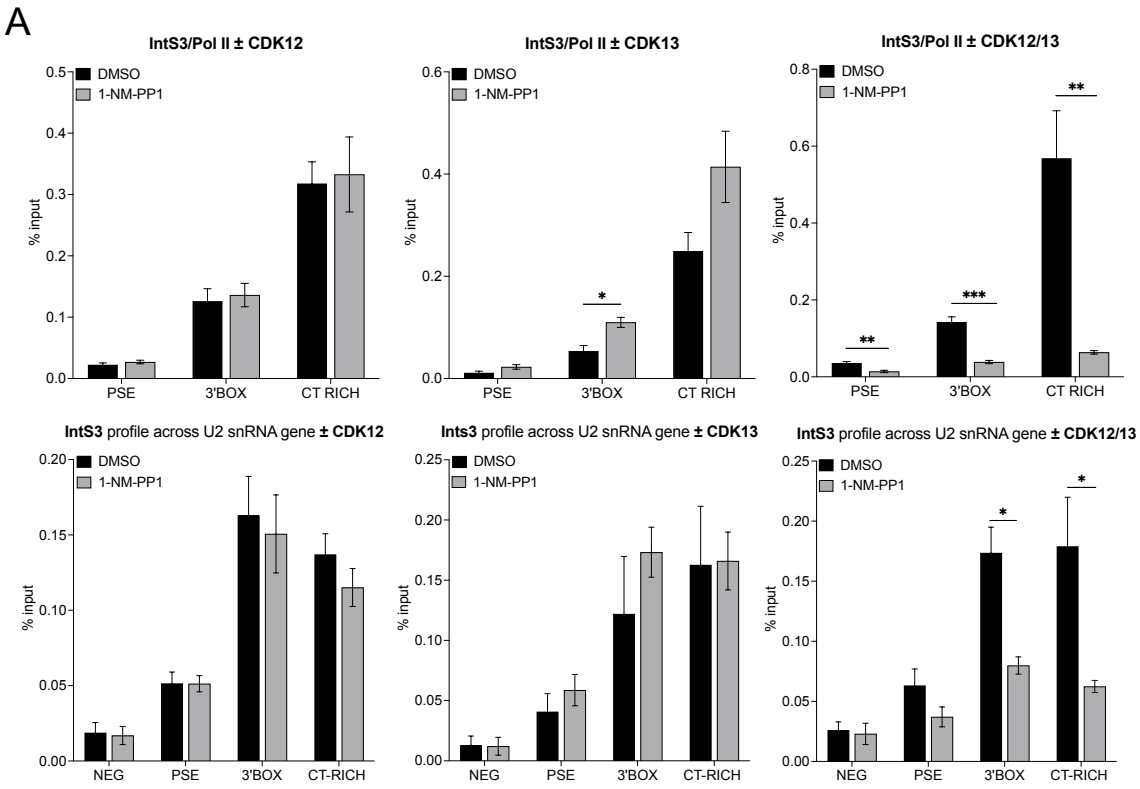
(Top) ChIP-qPCR of SPT6 relative to Pol II across the U2 snRNA gene after treatment of CDK12 and CDK13 AS cells with 7.5 μM 1-NM-PP1 or DMSO for 30 minutes. (Bottom) ChIP-qPCR of SPT6 across U2 snRNA gene upon inhibition of respective AS cells. n≥3 biological repeats. Statistical test = two-tailed unpaired t-test, * = $p < 0.05$, ** = $p < 0.01$.

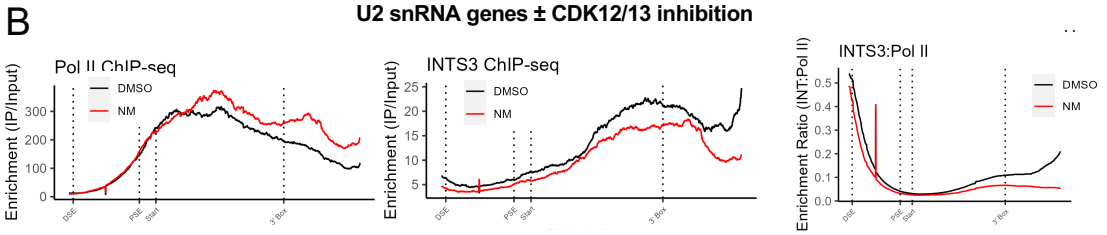
5.4 CDK12 and CDK13 are redundant for the recruitment of Integrator complex subunits.

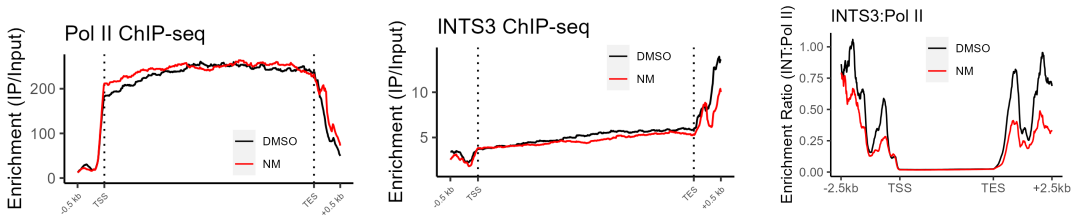
The regulation of IntS3 recruitment by SPT6 to snRNA genes [320] and the loss of SPT6 upon CDK12 and CDK13 AS cells inhibition (Fig. 5.2) suggests that CDK12 and CDK13 may regulate 3' end processing through the recruitment of Integrator subunits by SPT6. Therefore, the profile of IntS3 was assessed upon CDK12AS and CDK13AS inhibition (Fig 5.3A-D). ChIP-qPCR analysis revealed that inhibition of CDK12 does not affect IntS3 recruitment, CDK13AS inhibition causes an increase in IntS3/Pol II levels across the gene, and the levels of IntS3 are decreased by at least 50% by inhibition of both kinases. A more marked loss is observed when IntS3 is expressed with ratio to Pol II (Fig. 5.3A). This result suggests that CDK12 and CDK13 exhibit functional redundancy in the recruitment of IntS3 to the U2 snRNA gene. The loss of SPT6 and IntS3 when CDK12 and CDK13 AS cells are inhibited together suggests that CDK12 and CDK13 may regulate IntS3 directly or indirectly through SPT6 recruitment. ChIP-sequencing (ChIP-seq) of IntS3 upon inhibition of CDK12 and CDK13 AS cells together was also performed (Fig. 5.3 B-D). The sequencing results were analysed and plotted by Dr. Callum Henfrey, a postdoc and bioinformatician in the lab. ChIP-seq analysis on the U2 snRNA gene reveals that the inhibition of CDK12/13 inhibition causes a loss of IntS3, particularly at the 3' end. When normalised to Pol II, the loss of IntS3 is also most evident at the 3' box of the U2 snRNA and other snRNA genes (Fig. 5.3B, C). On protein-coding genes, the inhibition of CDK12 and CDK13 AS cells causes a loss of IntS3 across the gene (Fig. 5.3D). Normalisation to Pol II shows a slight decrease after the transcription start

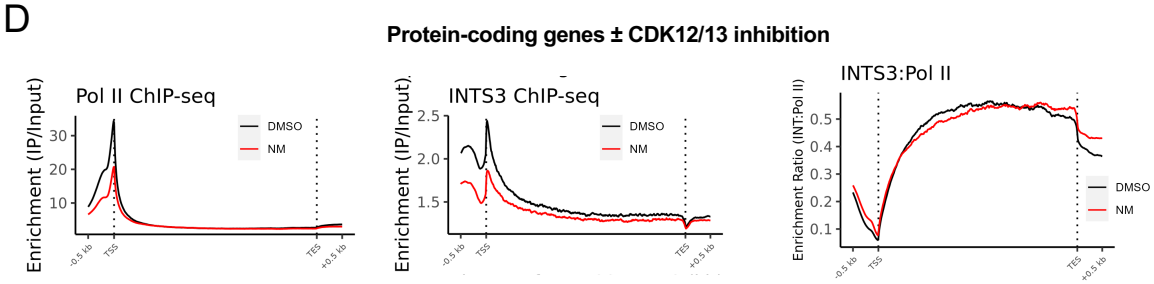
site (TSS) but an increase in IntS3/Pol II levels at the 3' end of the gene, which may be due to a processing or termination defect [77].

CDK7 and CDK9 regulate 3' end processing by facilitating IntS9/11 recruitment [96,99,100]. To determine whether the activity of these kinases is specific to the cleavage module or if they also mediate the recruitment of other Integrator subunits, CDK7AS and CDK9AS were inhibited, and the IntS3 profile on the U2 snRNA gene was analysed by ChIP-qPCR (Fig 5.3. E). CDK9 inhibition causes an approximate 50% loss of IntS3 across the gene, mirroring the IntS3 profile when CDK12/13 are inhibited together. A more significant loss is observed at the 3' box and CT-rich region when IntS3 is normalised to Pol II after CDK9 inhibition, similar to that seen after CDK12/13 inhibition. This suggests that CDK9 activity is either necessary to recruit other Integrator subunits besides the cleavage module or that the cleavage module is required to recruit IntS3 to the gene or stabilise its interaction. CDK7 inhibition also causes an approximate 50% loss of IntS3 levels across the U2 snRNA gene (Fig. 5.3E). However, the normalisation of IntS3 to Pol II shows a significant increase of IntS3 across the gene. This increase suggests that CDK7 is a negative regulator of IntS3 recruitment or that other kinases (e.g. CDK9, CDK12 and CDK13) can compensate for CDK7 inhibition.


A
IntS3/Pol II ± CDK12
IntS3/Pol II ± CDK13
IntS3/Pol II ± CDK12/13
DMSO
1-NM-PP1
% input
PSE
3'BOX
CT RICH
IntS3 profile across U2 snRNA gene ± CDK12
Ints3 profile across U2 snRNA gene ± CDK13
IntS3 profile across U2 snRNA gene ± CDK12/13
NEG
PSE
3'BOX
CT-RICH


B
U2 snRNA genes ± CDK12/13 inhibition
Pol II ChIP-seq
INTS3 ChIP-seq
INTS3:Pol II
Enrichment (IP/Input)
Enrichment Ratio (INT:Pol II)
DMSO
NM


C
Other snRNA genes ± CDK12/13 inhibition
Pol II ChIP-seq
INTS3 ChIP-seq
INTS3:Pol II
Enrichment (IP/Input)
Enrichment Ratio (INT:Pol II)
DMSO
NM
-2.5kb
TSS
TES
+2.5kb


D
Protein-coding genes ± CDK12/13 inhibition
Pol II ChIP-seq
INTS3 ChIP-seq
INTS3:Pol II
Enrichment (IP/Input)
Enrichment Ratio (INT:Pol II)
DMSO
NM
TSS
TES

E

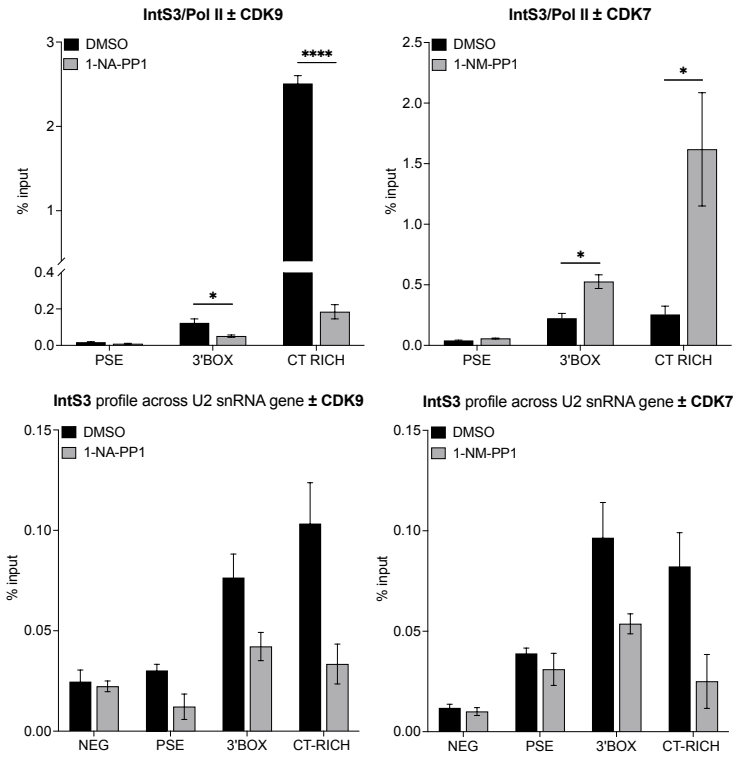


Figure 5.3 (A-E) CDK12 and CDK13 are redundant for recruitment of IntS3 to the U2 snRNA gene A. (Top) ChIP-qPCR of IntS3 relative to Pol II across the U2 snRNA gene after treatment of CDK12 and CDK13 AS cells with 7.5 μM 1-NM-PP1 or DMSO for 30 minutes. (Bottom) ChIP-qPCR of IntS3 across U2 snRNA gene upon inhibition of respective AS cells. B. ChIP-seq analysis of Pol II, IntS3 and IntS3 relative to Pol II on U2 snRNA genes upon inhibition of CDK12/13 AS cells with 7.5 μM 1-NM-PP1 or DMSO for 30 minutes. C. ChIP-seq analysis of Pol II, IntS3 and IntS3 relative to Pol II on other snRNA genes upon inhibition of CDK12/13 AS cells with 7.5 μM 1-NM-PP1 or DMSO for 30 minutes. D. ChIP-seq analysis of Pol II, IntS3 and IntS3 relative to Pol II on protein-coding genes upon inhibition of CDK12/13 AS cells with 7.5 μM 1-NM-PP1 or DMSO for 30 minutes. E. (Top) ChIP-qPCR of IntS3 relative to Pol II across the U2 snRNA gene after treatment of CDK9AS or CDK7AS with 15 μM 1-NA-PP1, 7.5 μM 1-NM-PP1 or DMSO for 30 minutes. (Bottom) ChIP-qPCR of IntS3 across U2 snRNA gene upon inhibition of respective AS cells. Statistical test = two-tailed unpaired t-test, * = p <0.05, ** = p<0.01, *** = p <0.001, **** = p <0.0001.

Given the role of CDK12 and CDK13 in IntS3 recruitment (Fig. 5.3A), I have also investigated the recruitment of the Integrator complex cleavage module and its binding partner, IntS9, when CDK12 and CDK13 are inhibited. I carried out several ChIP studies using a commercial IntS11 antibody. However, the results were

inconclusive as the negative control region did not give a significantly lower IP pulldown than the U2 snRNA gene (Supp Fig. 1). This is likely due to the antibody lacking specificity for IntS11. Therefore, the effect on the cleavage module was assessed using an anti-IntS9 antibody (Fig. 5.3 F, G).

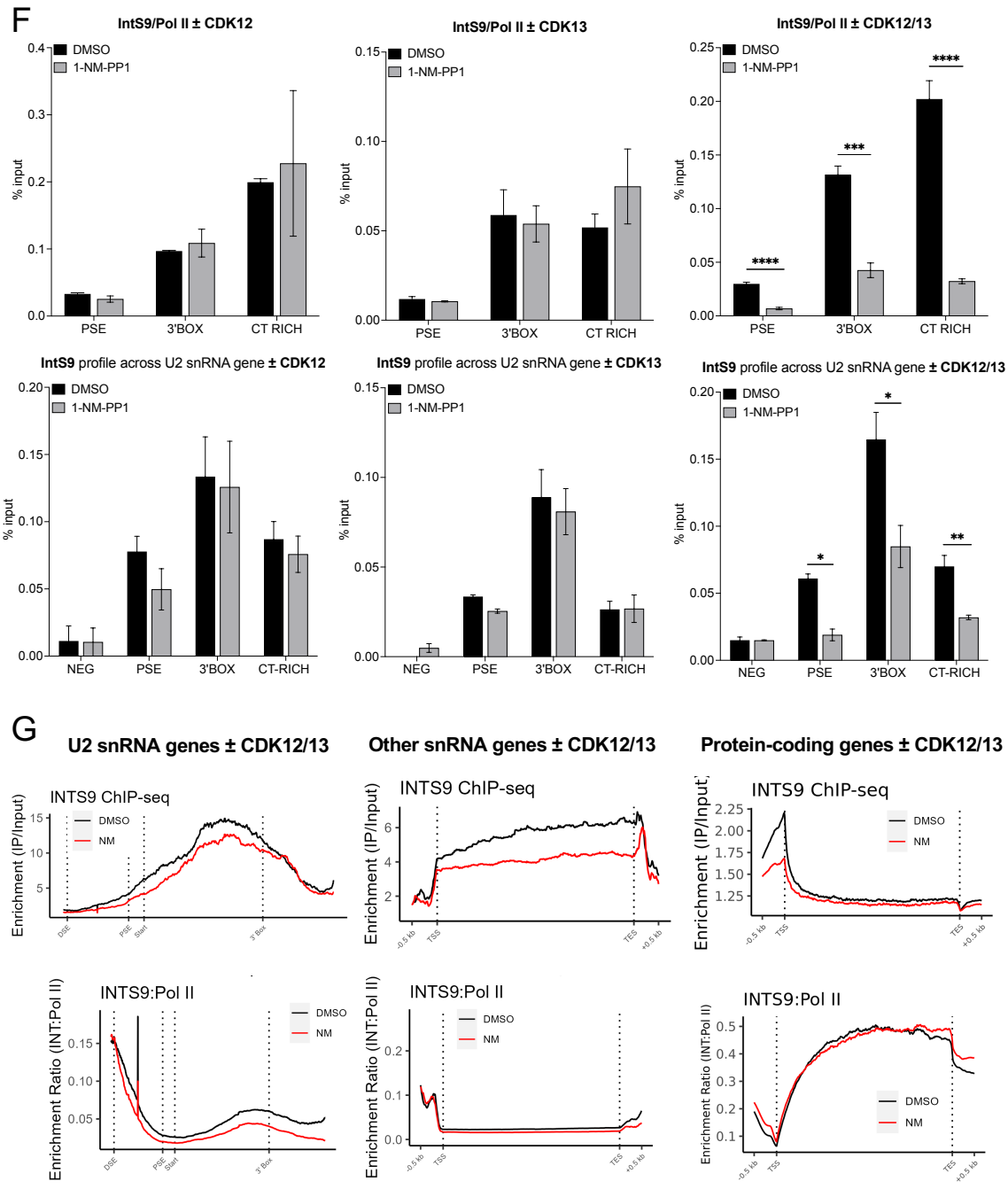


Figure 5.3 (F, G) CDK12 and CDK13 are redundant for the recruitment of IntS9/11 to the U2 snRNA gene F. (Top) ChIP-qPCR of IntS9 relative to Pol II across the U2 snRNA gene after treatment of CDK12 and CDK13 AS cells with 7.5 μM 1-NM-PP1 or DMSO for 30 minutes. (Bottom). ChIP-qPCR of IntS9 across U2 snRNA gene upon inhibition of respective AS cells. G.(Top) ChIP-seq analysis of IntS9 on U2 snRNA, other snRNA genes, and protein-coding genes upon inhibition of CDK12/13 AS cells with 7.5 μM 1-NM-PP1 or DMSO for 30 minutes. (Bottom) ChIP-seq analysis IntS9 relative to Pol II across respective genes upon inhibition of CDK12/13 AS cells. n≥3 biological repeats. Statistical test = two-tailed unpaired t-test, * = p <0.05, ** = p<0.01, *** = p <0.001, **** = p <0.0001.

Inhibition of CDK12 or CDK13 independently does not affect the recruitment of IntS9 across the snRNA gene. However, inhibition of both kinases together reduces the levels of IntS9 across the gene by approximately 50%. However, when normalised to Pol II, the loss of IntS9 observed across all assessed gene regions is greater than 50% (Fig. 5.3F). This result indicates that CDK12 and CDK13 are redundant for IntS9 recruitment to the U2 snRNA gene and may regulate the 3' end processing by affecting Integrator recruitment. ChIP-seq analysis of IntS9 upon CDK12/13 inhibition also shows the loss of IntS9 across the U2 snRNA and other snRNA genes. This loss is also observed when normalisation to Pol II is carried out. Conversely, the observed loss of IntS9 on protein-coding genes reflects a loss of Pol II as there is a slight loss of IntS9/Pol II after the TSS but an increase in IntS9/Pol II levels at the 3' end of the gene, similar to IntS3/Pol II (Fig. 5.3. D)

5.5 CDK12 are CDK13 are required for BRAT1 recruitment to the U2 snRNA gene.

BRCA1-associated ATM activator 1 (BRAT1), a protein involved in the DNA damage repair and neuronal differentiation pathways, was recently shown to interact with

IntS9 and IntS11 [394–396]. It was proposed that BRAT1 binds to IntS11 to preserve its structural and functional integrity, and BRAT1 mutation causes loss of IntS11 activity, which results in snRNA misprocessing [394,395]. Given the functional relevance of BRAT1 in snRNA gene expression and the limitations faced with the anti-IntS11antibody for ChIP-qPCR, the anti-BRAT1 antibody was used as a proxy to assess the effects of CDK12AS and CDK13AS inhibition on IntS11 recruitment (Fig. 5.4). Inhibition of CDK12AS and CDK13 AS cells independently causes a loss of BRAT1 across all regions of the gene (Fig. 5.4). However, a further and more significant reduction of BRAT1/Pol II levels is observed when both kinases are inhibited together, suggesting that CDK12 or CDK13 partially compensates for BRAT1 recruitment in the absence of the other.

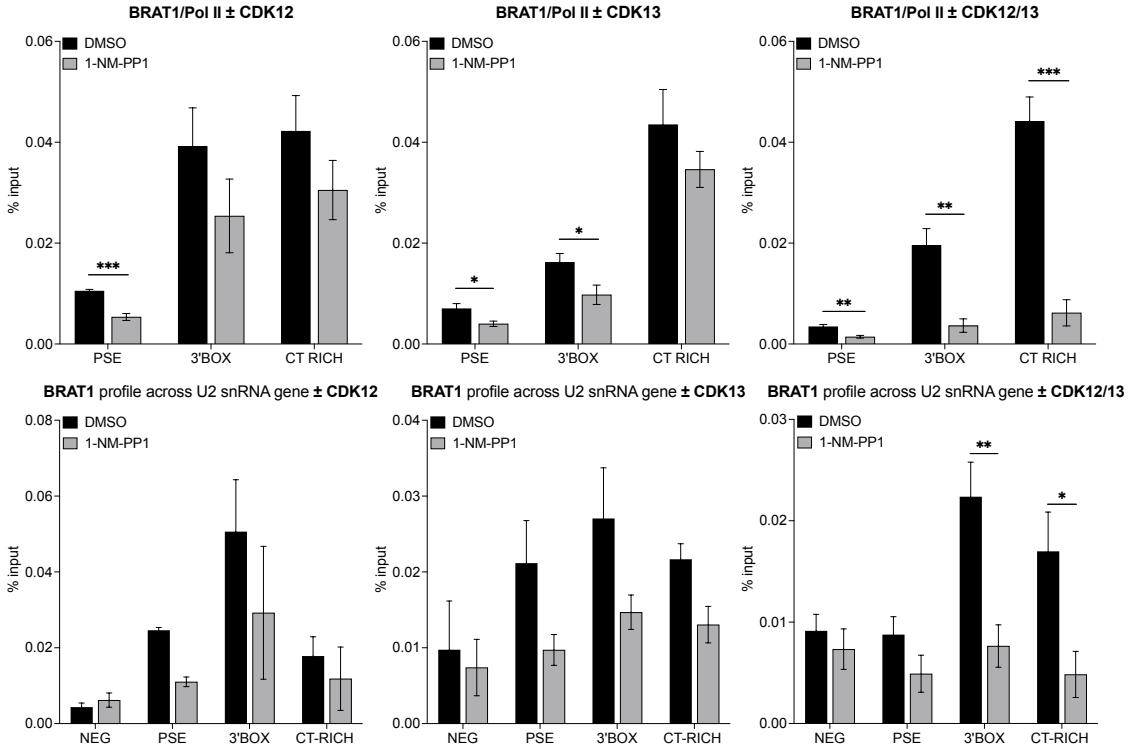


Figure. 5.4. CDK12 and CDK13 activity is required to recruit BRAT1 to the U2 snRNA gene.

(Top) ChIP-qPCR of BRAT1 relative to Pol II across the U2 snRNA gene after treatment of CDK12 and CDK13 AS cells with 7.5μM 1-NM-PP1 or DMSO for 30 minutes. (Bottom) ChIP-qPCR of BRAT1 across U2 snRNA gene upon inhibition of respective AS cells. n≥3 biological repeats. Statistical test = two-tailed unpaired t-test, * = p <0.05, ** = p<0.01, *** = p <0.001.

5.6 Can SPT6, IntS3 or ICE2 mediate the 3' end processing of snRNA gene transcripts?

Three highly conserved transcription factors have functions exclusive to snRNA gene expression: PTF complex, the interactor of the little elongation complex ELL subunit 1 (ICE1), and ICE2 subunits of the little elongation complex (LEC) [397–399]. The snRNA gene-specific promoter is required for the recognition of the 3' box and subsequent 3' end processing [267,268]. Therefore, LEC subunits involved in transcription elongation may bridge the gap between transcription initiation and RNA processing by potentially relaying signals between the promoter and the 3' box. In addition, the observed redundancy of CDK12 and CDK13 in IntS3, IntS9 and SPT6 recruitment highlights the possible involvement of these factors in snRNA 3' end processing. Therefore, Prof. Murphy, Mrs Barker, and I aimed to create dTAG degron cell lines for SPT6, IntS3 and ICE2 proteins using CRISPR-Cas9 genome editing to investigate the primary effects of rapid degradation of these proteins. The dTAG system utilizes the ubiquitination system to target proteins of interest for degradation upon induction by a small molecule, d-TAGV-1 [400,401]. Prof Murphy designed the constructs, which contain a dTAG degron tag, a 3X FLAG tag to circumvent the limitations of commercial antibodies during immunoprecipitation experiments, and an XTEN linker to provide flexibility and avoid interference of the degron or degradation machinery with the structure of the protein of interest (see

Chapter 2). Mrs Barker and Prof. Murphy performed the cloning, aiming to N-terminally tag the proteins of interest in the HEK293 parental cell line. I transfected the parental cell line and screened colonies for positive dTAG cell lines. After drug selection post-transfection, cells were sorted into single cells in single wells using flow cytometry. PCR was used to identify colonies with the dTAG that were correctly integrated. However, the colonies had a very low survival rate following cell sorting, and PCR was not consistently successful. Therefore, I implemented an alternative technique where cells were plated sparsely on 15cm dishes after drug selection to allow the growth of single colonies. Subsequently, these colonies were individually harvested and screened by western blotting (Fig. 5.5). Unfortunately, generating a homozygous cell line for SPT6 and IntS3 dTAG cell lines proved unsuccessful. However, several heterozygous colonies were found, and one clone was selected (marked by an asterisk). (Fig. 5.5 A, B). Two ICE2 cell lines were generated as indicated by ICE2 western screening (Fig. 5.5 C). The selected clones were treated with 1 μM of $dTAG^{V}$-1 over a 6-hour time course. Western blot analysis shows that degron-tagged SPT6, IntS3 and ICE2 are degraded within two hours of treatment (Figure 5.5. A-C). U2 pre and readthrough snRNA transcript levels were assessed using an RT-qPCR assay when these proteins were conditionally degraded. RT-qPCR analysis shows that degradation of the tagged SPT6, IntS3, and ICE2 does not affect the 3' end processing or levels of pre-snRNA transcripts produced (Fig 5.5 A-C). This suggests that the activity of the untagged SPT6 and IntS3 is sufficient for 3' end processing, and ICE2 may not play a role in the 3'end processing of snRNA gene transcripts. Alternatively, ICE2 present on the U2 snRNA gene may not be degraded after a 6-hour treatment. It is also possible that SPT6 and IntS3 are not removed from

the gene in the time of degron activation. ChIP-qPCR of SPT6, IntS3 and ICE2 upon degradation will be needed to confirm this.

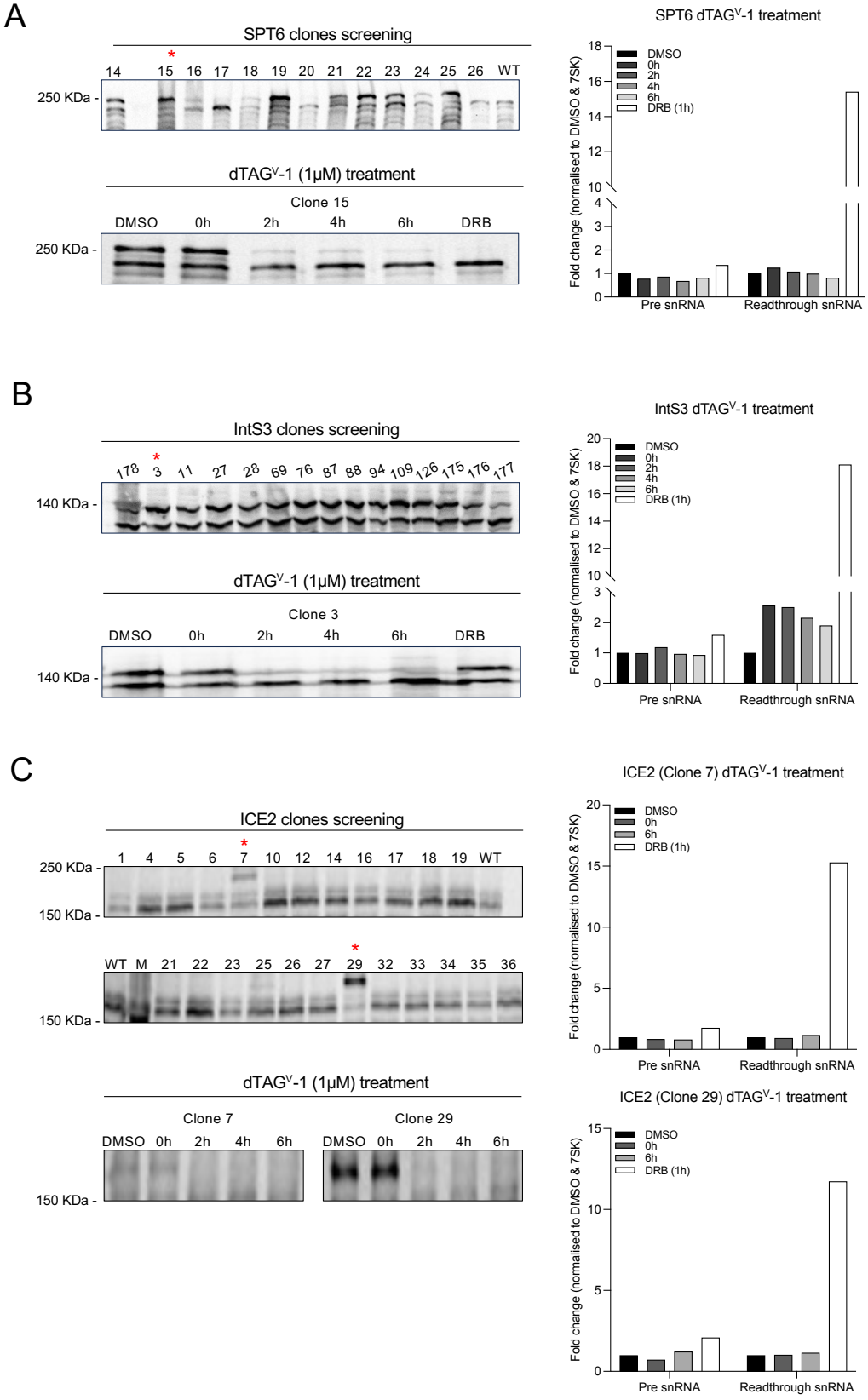

A
SPT6 clones screening
14 15 16 17 18 19 20 21 22 23 24 25 26 WT
250 KDa -
dTAGV-1 (1μM) treatment
Clone 15
DMSO 0h 2h 4h 6h DRB
250 KDa -
SPT6 dTAGV-1 treatment
DMSO
0h
2h
4h
6h
DRB (1h)
Fold change (normalised to DMSO & 7SK)
Pre snRNA
Readthrough snRNA
B
IntS3 clones screening
178 3 11 27 28 69 76 87 88 94 109 126 175 176 177
140 KDa -
dTAGV-1 (1μM) treatment
Clone 3
DMSO 0h 2h 4h 6h DRB
140 KDa -
IntS3 dTAGV-1 treatment
DMSO
0h
2h
4h
6h
DRB (1h)
Fold change (normalised to DMSO & 7SK)
Pre snRNA
Readthrough snRNA
C
ICE2 clones screening
1 4 5 6 7 10 12 14 16 17 18 19 WT
250 KDa -
150 KDa -
WT M 21 22 23 25 26 27 29 32 33 34 35 36
150 KDa -
dTAGV-1 (1μM) treatment
Clone 7
DMSO 0h 2h 4h 6h
Clone 29
DMSO 0h 2h 4h 6h
150 KDa -
ICE2 (Clone 7) dTAGV-1 treatment
DMSO
0h
6h
DRB (1h)
Fold change (normalised to DMSO & 7SK)
Pre snRNA
Readthrough snRNA
ICE2 (Clone 29) dTAGV-1 treatment
DMSO
0h
6h
DRB (1h)
Fold change (normalised to DMSO & 7SK)
Pre snRNA
Readthrough snRNA

Figure 5.5. Can SPT6, IntS3 or ICE2 mediate the 3' end processing of snRNA gene transcripts?

(Left) Western blot screening for SPT6, IntS3 or ICE2 dTAG mutant with respective antibodies. Western blot analysis of selected clone (red asterisks) following time course treatment with 1μM dTAG^{V}-1, DMSO or DRB, using respective antibodies. (Right) RT-qPCR quantification of the U2 pre-snRNA and readthrough snRNA transcripts produced after treating the selected clone with 1μM dTAG^{V}-1, DMSO (negative control), or DRB (positive control) for an hour. Red asterisks indicate the selected clone.

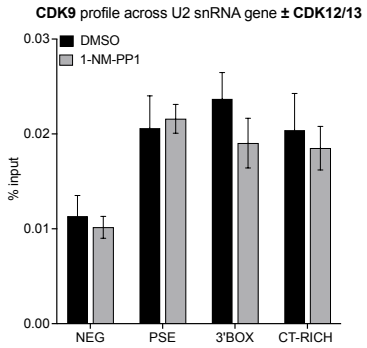
5.7 CDK9 or CDK12 and CDK13 are required for the 3' end processing of snRNA gene transcripts.

The similarity between the roles of CDK9 and the joint role of CDK12 and CDK13 in Integrator subunit recruitment and regulating snRNA 3' end processing suggests that CDK9 functions downstream or upstream of CDK12 and CDK13. ChIP-qPCR analysis indicates that the inhibition of CDK12AS and CDK13AS together does not affect the recruitment of CDK9 to the snRNA gene (Fig.5.6A) but could regulate CDK9 activity by phosphorylating CDK9 at the Thr186 residue to produce the active form of this kinase on snRNA genes [94,402]. Alternatively, CDK9 may function upstream of CDK12 and CDK13. This would require an analysis of the effect of CDK9 inhibition on CDK12/13 and/or Cyclin K recruitment and activation on snRNA genes.

RT-qPCR was also used to investigate whether CDK12 and CDK13 control the 3' box recognition through the same pathway as CDK9 (Fig 5.6B) CDK12/13 AS cells were inhibited with 7.5 μM NM or 100μM DRB separately, or NM and DRB together for 30 minutes and the levels of pre- and readthrough snRNA transcripts analysed. RT-qPCR analysis shows that the inhibition of CDK9 independently by DRB, CDK12AS and CDK13AS, or all three kinases together, causes an snRNA 3' end processing

defect. Importantly, the ratio of readthrough to pre-snRNA indicates that the inhibition of the three kinases together causes a higher increase in snRNA 3' end processing than CDK9 inhibition alone or CDK12 and CDK13 together (Fig. 5.6B). This result indicates that CDK12 and CDK13 may mediate snRNA 3' end processing, at least partly, through a different mechanism to CDK9. To confirm that the observed effect was a combined effect of inhibiting all three kinases and not a result of inhibiting CDK9 and CDK12 or CDK9 and CDK13. CDK12AS or CDK13AS cell lines were inhibited with NM and DRB (Fig.5.3C). The inhibition of CDK9 and CDK12AS or CDK9 and CDK13AS together does not affect the level of readthrough snRNA or pre-snRNA transcripts produced. This confirms that the previous increase in snRNA readthrough resulted from inhibiting all three kinases together.

A



B

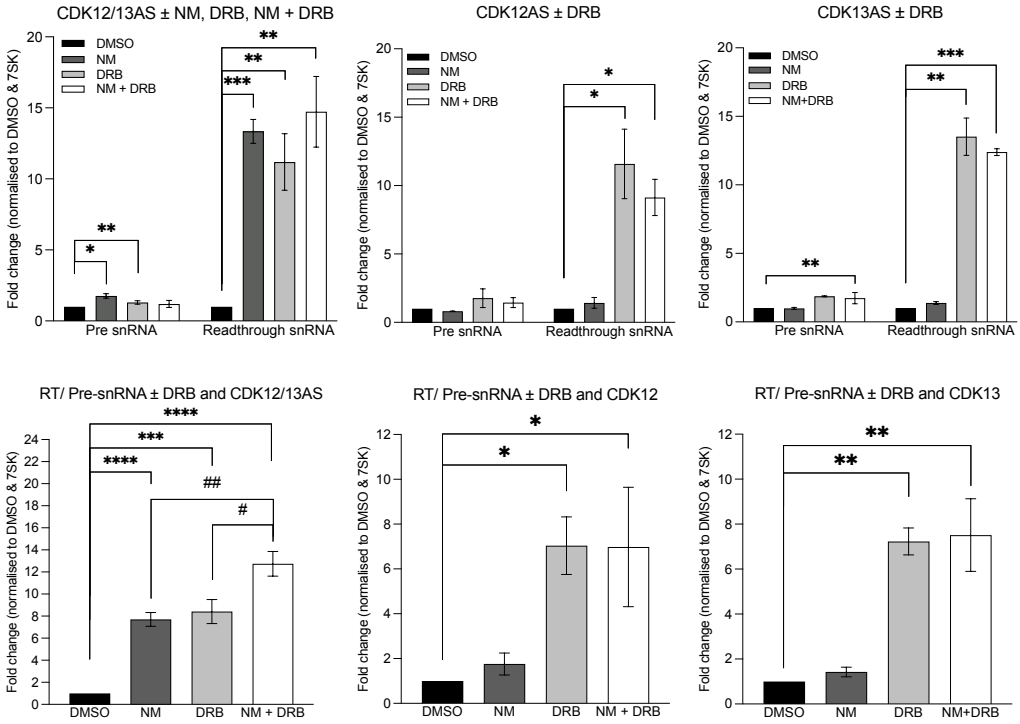


Figure 5.6 CDK12 and CDK13 do not facilitate the recruitment of CDK9 to the U2 snRNA gene and may regulate 3' end processing through distinct pathways.

A. ChIP-qPCR of CDK9 profile across U2 snRNA gene upon inhibition of CDK12/13 AS cells with 7.5μM 1-NM-PP1 or DMSO for 30 minutes. B. (Top) Quantification of transcripts produced after treating CDK12AS, CDK13AS and CDK12/13 AS cells with NM, DRB, or DMSO. (Bottom) The ratio of pre-snRNA to readthrough snRNA transcripts after inhibiting CDK12, CDK13 and CDK12/13 AS cells with NM, DRB, or DMSO. n≥3 biological repeats. Statistical test = two-tailed unpaired t-test, * or #= p <0.05, ** or ## = p < 0.01*** = p <0.001, **** =p <0.0001. * denotes t-test between DMSO and drug, # denotes t-test between NM+DRB and other drug treatments.

5.8 Discussion

The experiments described in this chapter aimed to elucidate how CDK12 and CDK13 mediate snRNA 3'end processing by identifying factors involved in this mechanism. We have previously shown the presence of SPT6 and Cyclin K on human snRNA genes using CAPTURE pulldown [320]. A re-analysis of previously published data confirms that SPT6 is present on snRNA genes and shows that its knockdown also causes a loss of IntS3 on snRNA genes [319,320]. However, the role of SPT6 in snRNA gene expression remains unknown. Factors associated with elongation on protein-coding genes are often associated with 3' end processing and termination on snRNA genes. For example, NELF and DSIF accumulate at the transcription start site of protein-coding genes and bind to Pol II to form the paused elongation complex [150,154]. Instead, NELF and DSIF accumulate on snRNA genes at the 3' end [173,335].

Similarly, PTEF-b, an elongation factor on protein-coding genes, functions at the 3' end of snRNA genes [303]. This gene-specific regulation, where elongation factors on protein-coding genes are implicated in the 3' end processing of snRNA gene transcripts, led to the hypothesis that SPT6, a histone chaperone and elongation factor on protein-coding genes, may also exhibit functional duality and be required to regulate the 3' end RNA processing or termination of snRNA gene transcripts. Indeed, SPT6 is implicated in the termination of transcription of other non-coding RNAs through the recruitment of IntS3 [319]. Furthermore, ChIP-qPCR analysis indicates that SPT6 recruitment is highest at the 3' box of the U2 gene, like other snRNA processing factors, e.g. Integrator subunits (Fig. 5.3). ChIP-qPCR analysis

showed that CDK12 and CDK13 are redundant in recruiting Integrator subunits IntS9 and IntS3, indicating that these kinases may mediate the 3' end processing in an Integrator-dependent manner. ChIP-seq analysis also shows a loss of Ints3/Pol II and IntS9/Pol II, particularly at the 3' end of U2 and other snRNA genes when CDK12 and CDK13 AS cells are inhibited. Indeed, CDK12/13 may mediate IntS3 recruitment through SPT6. Therefore, the observed increase in IntS3 recruitment following CDK13AS inhibition (Fig. 5.3A) may reflect the increase in SPT6 recruitment (Fig 5.2). On the other hand, CDK12AS and CDK13AS inhibition causes a slight decrease in the levels of IntS3/Pol II and IntS9/Pol II after the TSS of protein-coding genes. However, there is an accumulation of IntS3/Pol II and IntS9/Pol II around the TES, which indicates that CDK12 and CDK13 regulate the recruitment of IntS3 on a gene-dependent basis.

The loss of BRAT1, which has been shown to interact with and be required for the activity of IntS11 [394,396], observed upon inhibition of CDK12AS and CDK13AS cells independently suggests that CDK12 and CDK13 may cause moderate loss of the IntS11 subunit. However, this loss may not be sufficient to cause snRNA gene readthrough.

We also sought to create degron-tagged cell lines to investigate the possible role of SPT6, IntS3 and ICE2 in the recognition of the 3' box. Similarly, SPT6, CDK9 and CDK12/13 affect the recruitment of IntS3 to the snRNA gene. Given the roles of these kinases in 3' end processing, it was necessary to investigate the possible roles of SPT6 and IntS3 in this process. In addition, ICE2 is a subunit of one of two snRNA-specific complexes, the little elongation complex [398], which is required for snRNA gene expression [398]. Hence, ICE2 may play a role in mediating the obligate coupling

of the promoter to 3' end processing [267,268]. Of all three factors, ICE2 is the only factor where homozygous degron cell lines were created (Fig.5.3 C). However, the conditional degradation of ICE2 in two independent cell lines did not affect recognition of the 3' box. It is possible that a 6-hour dTAG-mediated degradation of ICE2 is not sufficient to fully degrade the protein in complexes on the gene, and the absence of an increase in snRNA readthrough transcripts by RT-qPCR analysis could reflect the presence of residual ICE2 on the snRNA gene. Similarly, when SPT6 and IntS3 are conditionally degraded, these factors on the gene are not easily accessible for degradation, allowing snRNA expression to proceed normally. The lack of homozygous IntS3 and SPT6 degron-tagged cell lines identified in the screening may also highlight the importance of these proteins for gene expression in HEK 293 cells. It is also possible that the N-terminal tagging of these highly conserved proteins changes their structure and, subsequently, their function, which is lethal to the cell. Therefore, only heterozygotes cell lines survive as the other allele compensates.

5.9 Limitations

A limitation of these experiments is the absence of SPT6 ChIP profiles after CDK7 and CDK9 inhibition, which would help to elucidate a possible mechanism for their recruitment. ChIP-qPCR of other Integrator subunits upon inhibition of CDK7, CDK9, CDK12 and CDK13 AS cells would also provide further insights into how these kinases regulate the recruitment of the Integrator complex. Additionally, the possibility that CDK9 functions upstream of CDK12 and CDK13 requires ChIP-qPCR

analysis of CDK12, CDK13 and Cyclin K after CDK9 inhibition. Also, ChIP-qPCR of pThr186-CDK9, the active form of CDK9, upon CDK12/13 inhibition will be required to determine whether CDK12/13 control CDK9 activity on the snRNA genes. Another limitation of this experiment is the need for more validation that ICE2 is absent from snRNA genes upon dTAG degradation. ChIP-qPCR analysis of the ICE2 profile upon dTAG degradation would be needed to investigate this.

Chapter 6 | Summary and Outlook

The two major processes known to mediate the recognition of the 3' box during the transcription of human snRNA genes are transcription initiation by an snRNA-specific promoter and the enzymatic activity of transcriptional CDKs through the phosphorylation of Ser2 by CDK9 and phosphorylation of Ser7 by CDK7 [96,99,100,267,268,303]. The current study investigated the roles of other transcriptional CDKs in recognition of the 3' box and uncovered a joint role of CDK12 and CDK13 in RNA 3' end processing (Chapter 3). However, the mechanism behind this role remains unclear. The presence of CDK9 and the presumed presence of CDK7 in the absence of CDK12 and CDK13 activity suggest that Ser2 and Ser7 would still be phosphorylated. However, a significant loss of pSer2, IntS9 and IntS3 (Chapters 4 and 5) suggests that CDK12 and CDK13 control RNA 3' end processing in an Integrator-dependent manner. Additionally, the recruitment of CDK9 to snRNA genes is not dependent on CDK12 and CDK13, and the inhibition of all three kinases together causes a significantly higher increase in U2 snRNA readthrough than inhibiting CDK9 alone or CDK12 and CDK13 together, indicating that these kinases may work through different pathways (Chapter 5). Furthermore, SPT6 is associated with the U2 snRNA gene, and the activity of CDK12 and CDK13 together is required for its recruitment. SPT6 is also required for IntS3 recruitment to snRNA genes [320], as are either CDK12 or CDK13 (Chapter 5).

6.1 The roles of the implicated factors

6.1.1 The Integrator Complex

The Integrator complex comprises 15 canonical subunits (IntS1-IntS15) [167,403,404], which are grouped into several subcomplexes that play various roles in the expression of protein-coding genes [175,311,386,403,405,406]. Notably, IntS11/IntS9 heterodimer and IntS4 make up the cleavage module [167,405]. IntS11 and IntS9 are homologs of CPSF73 and CPSF100, respectively [407]. IntS11 functions as the endonuclease of the cleavage module, IntS9 acts as the heterodimeric binding partner of IntS11, and IntS4 as a scaffold of the complex, similar to the CPSF73-CPSF100-Symplexin mRNA cleavage factor [405,408]. Integrator is recruited to protein-coding genes early in the transcription cycle to cause early termination of transcription by mediating the dephosphorylation of the paused Pol II CTD through protein phosphatase 2A (PP2A) [165,386,409] and cleaving nascent transcripts, which may allow the degradation of transcripts by exonucleases and Pol II destabilisation [174,410–412]. In the expression of human snRNA genes, Ints11, as part of the Integrator cleavage module, cleaves nascent snRNA gene transcripts just upstream from the 3' box [167]. RPAP2 has been shown to help to recruit IntS1, IntS4, IntS5, IntS6, and IntS7, whereas Ints9 and Ints11 require phosphorylation of Ser2 and Ser7 of the Pol II CTD for stable recruitment [96,97,99,275]. It is unclear what other IntS subunits are recruited to human snRNA genes and if the whole Integrator complex may be necessary. IntS3 may function independently of the Integrator complex and is a subunit of the sensor of the single-stranded DNA (SOSS)1 and SOSS2 complexes required for the double-stranded DNA break (DSB) repair process [305]. IntS3 interacts

with IntS6 through its C-terminal domain (CTD) and dimerises to facilitate the formation of a larger SOSS1 complex, which recognises longer single-stranded (ss) DNA and plays a role in R-loops resolution [309,413,414]. Interestingly, IntS3 has been proposed to recruit the Integrator-PP2A (INTAC)-SOSS1 complex to the paused Pol II complex due to its ssDNA recognition functions [415]. Until recently, IntS3 has been excluded from most Integrator complex structural models due to its flexibility [165,386,416]. Furthermore, recent structural studies of IntS3 CTD have demonstrated that this subunit has a stronger affinity for ssRNA over ssDNA despite its role in DNA damage repair, indicating that IntS3 may have a primary role in RNA processing [414]. My results suggest that IntS3 plays a key role in recruiting or stabilising other Integrator subunits, including the cleavage module in 3' box-dependent RNA processing.

6.1.2 SPT6

SPT6 is a histone chaperone which recruits SETD2 to deposit the K3K36me3 epigenetic mark, [317,319] and a known elongation factor required for the expression of protein-coding genes. The knockdown of this factor impairs transcription of coding genes and causes loss of epigenetic marks such as H3K36me3 [156,314–319]. As a histone chaperone, SPT6 is thought to help Pol II negotiate transcription through nucleosomes [392]. Its role in the expression of human snRNA genes is surprising as these genes are largely nucleosome-depleted [325]. However, previously published data [319,320] and the data presented here suggest that a key role is the recruitment or stabilization of IntS3.

6.1.3 CDK9

CDK9 is a master transcriptional kinase for protein-coding genes and plays roles at the beginning and end of these genes. For instance, at the 5' end of the gene, CDK9 phosphorylates NELF, DSIF and CTD Ser2 to allow progression into productive elongation and coupling elongation to pre-mRNA processing [92,95,159,417–419]. At the 3' end of the gene, CDK9 phosphorylates Xrn2 and PP1 and helps to recruit polyadenylation factors through Pol II CTD Ser2 phosphorylation to regulate termination [356,383,385]. During snRNA gene expression, CDK9 primarily functions at the 3' end of the gene [303]. Here, the phosphorylation of Ser2 by CDK9 is required for the recruitment of IntS11/9 subunits, which facilitates the subsequent RNA 3' end processing of the nascent transcript [96,304]. My data suggests that CDK9 may also regulate the snRNA 3' end processing through the recruitment of other Integrator subunits (Chapter 5). The inhibition of CDK9 causes a near total loss of IntS3/Pol II at the 3' box and CT-rich regions of the U2 snRNA gene (Fig. 5.3E). CDK9 is required for the 3' end processing of snRNA genes by phosphorylating Ser2, which is required to recruit the Integrator cleavage module [96,303,304]. Previous studies suggest that other non-cleavage subunits of the Integrator complex are recruited earlier during transcription through their Integrator subunit association with RPAP2 [99]. The data presented here (Fig. 5.6E) indicates that not only is IntS3 enriched at the 3' end of the snRNA genes, but CDK9 is also required to recruit it to U2 snRNA gene.

6.1.4 CDK12 and CDK13

CDK12 and CDK13 are paralogous CTD kinases which share the same regulatory cyclin, cyclin K [51], and are involved in regulating phosphorylation of Ser2 and Ser5 CTD residues [54,76,420].

CDK12 and CDK13 are involved in the transcription elongation of protein-coding genes and pre-mRNA processing [77,323,378,421–423]. However, not much was known about the roles of CDK12 and CDK13 in the expression of snRNA genes. Previously published data and the presented data (Chapter 5) indicate that CDK12 and cyclin K are associated with U2 snRNA genes [320]. Additionally, my data (Chapter 5) indicates that CDK12 and CDK13 together play roles in the phosphorylation of Tyr1, Ser2, Ser5 and Ser7 CTD residues (Chapter 4) and can mediate the recruitment of SPT6 (Fig. 5.1A). In addition, CDK12 and CDK13 are redundant for the recruitment of integrator subunits, IntS3 and IntS9, and in the 3' end processing of human U2 snRNA (Chapter 5). However, the exact mechanism by which CDK12 and CDK13 mediate the snRNA 3' end processing remains unclear.

6.2 Potential mechanisms of actions of CDK12 and CDK13 in snRNA 3' end processing.

Several models could explain how CDK12/13 regulate the 3' end processing of snRNA genes.

6.2.1. CDK12 and CDK13 may recruit SPT6 to function at the 3' end of snRNA genes

SPT6 is absent from the recent structural studies of the Integrator pre-termination complex at the promoter-proximal pausing of protein-coding genes [406]. This is likely because the presence of SPT6 will facilitate the progression of Pol II into productive elongation [156,314,316], and the SPT6 structure sterically clashes with IntS11, the endonuclease, and IntS6, which has been previously shown to bind to the CTD of IntS3 [172,413]. In the context of snRNA gene transcription, SPT6 may be recruited and stabilised on the snRNA genes by CDK12 and CDK13. CDK7 phosphorylates Ser7 to recruit RPAP2 and some Integrator subunits early in the transcription cycle of snRNA genes [99]. At the 3' end, CDK9 recruits IntS9/11 through Ser2 phosphorylation [96], and CDK12/13 may recruit IntS3 directly or indirectly through SPT6. IntS3 may then stabilise the Integrator complex and facilitate the cleavage of nascent snRNA (Fig. 6.1). In the absence of IntS3, the formation of a stable Integrator complex is lost. Also, SPT6 may still be recruited, potentially causing steric clashes with other Integrator subunits, e.g. IntS6, and sterically hindering the recruitment of IntS11.

6.2.1.1 Future directions

Analysis of 3' box-dependent RNA processing after degradation of SPT6 and/or IntS3 could determine their involvement in the RNA 3' end processing of snRNA gene transcripts. An IP-Mass spectrometry pulldown of IntS3 upon CDK12 and CDK13 inhibition will also provide further insights into factors that interact with IntS3 and whether CDK12 and CDK13 mediate these interactions. Additionally, IntS11 ChIP

profile upon IntS3 degradation will determine whether this subunit is required to recruit the IntS9/11 complex.

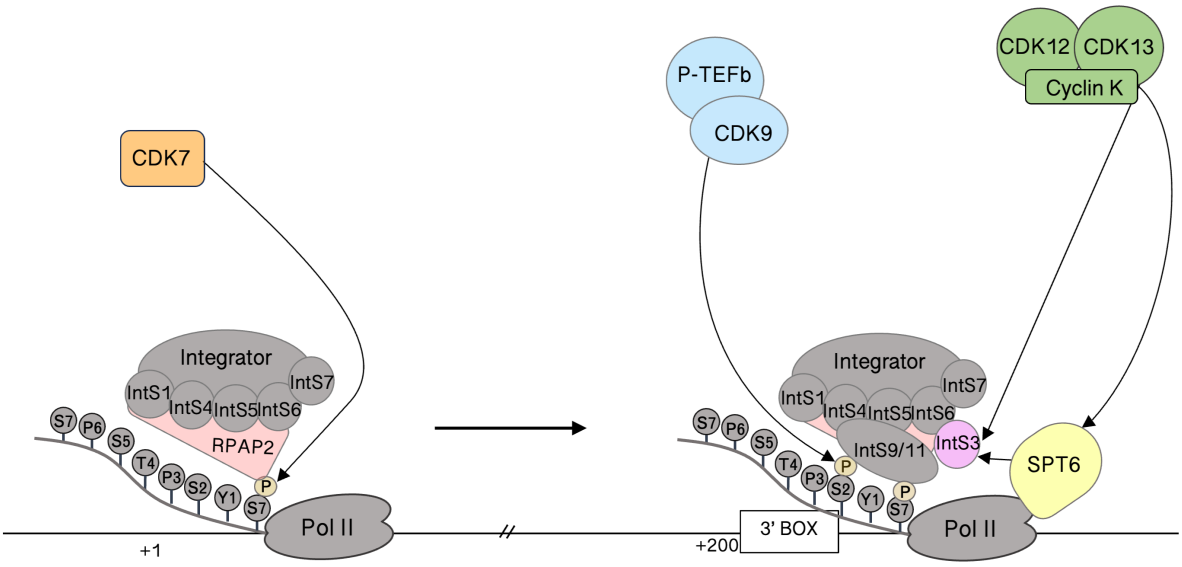


Figure 6.1 SPT6 may function at the 3' end of the snRNA gene to aid the recruitment of IntS3. Schematic model showing CDK7 recruits some Integrator subunits at the beginning of transcription through phosphorylated Ser7. At the 3' end, CDK12/13 may mediate the recruitment of IntS3 directly or through SPT6, which completes and stabilises the Integrator complex, and the snRNA transcript is subsequently cleaved. P denotes phosphorylation.

6.2.2 Interactions between CDK12/CDK13 and CDK9

6.2.2.1 CDK12 and CDK13 may function upstream of CDK9

CDK12/13 may function upstream of active CDK9 and mediate the RNA 3' end processing through the known pathway of Ser2 phosphorylation to recruit the Integrator cleavage module [96,99,100] (Fig. 6.2 A). CDK12/13 does not affect the recruitment of CDK9 to snRNA genes (Fig. 5.6 A). However, this experiment does not measure the active form phospho-CDK9 (pThr186-CDK9). CDK12 and CDK13 may

activate CDK9 at the 3' end by phosphorylation of its Thr186 residue, thereby regulating snRNA 3' end processing through pCDK9 activity (Fig. 6.2A). Previous studies have shown an association of CDK9 with cyclin K, the regulatory cyclin for CDK12 and CDK13, both in vitro [424]and in vivo [93]and CDK9 activity is regulated through its interaction with cyclin K [425]. Additionally, CDK9 primarily functions at the 3' end of snRNA genes as an RNA processing kinase [303,304]. ChIP-qPCR data presented here (Fig. 4.1) demonstrate increased Pol II levels upon inhibition of CDK12/13 across all regions of the U2 snRNA gene. CDK12 and CDK13 therefore regulate Pol II transcription throughout the gene but may control snRNA 3' end processing through activating CDK9.

6.2.2.2 CDK12 and CDK13 may function downstream of CDK9

The data presented in Chapter 4 indicate that inhibition of CDK12 and CDK13 causes a significant loss of pSer2/Pol II at the 3' box and CT-rich regions of the U2 snRNA gene (Fig. 4.3A). However, a near total loss of pSer2/Pol II is observed across the 3' end of the gene when CDK9 is inhibited.

It is possible that CDK9 mediates the phosphorylation of Ser2 partly through CDK12 and CDK13 activity but also through direct phosphorylation of Ser2 at the 3' end of snRNA genes. This may explain why the inhibition of CDK12 and CDK13 also causes a loss of IntS9/Pol II. CDK9 may also help to recruit IntS3 to stabilise the Integrator complex through the activity of CDK12 and CDK13 (Fig. 6.2B).

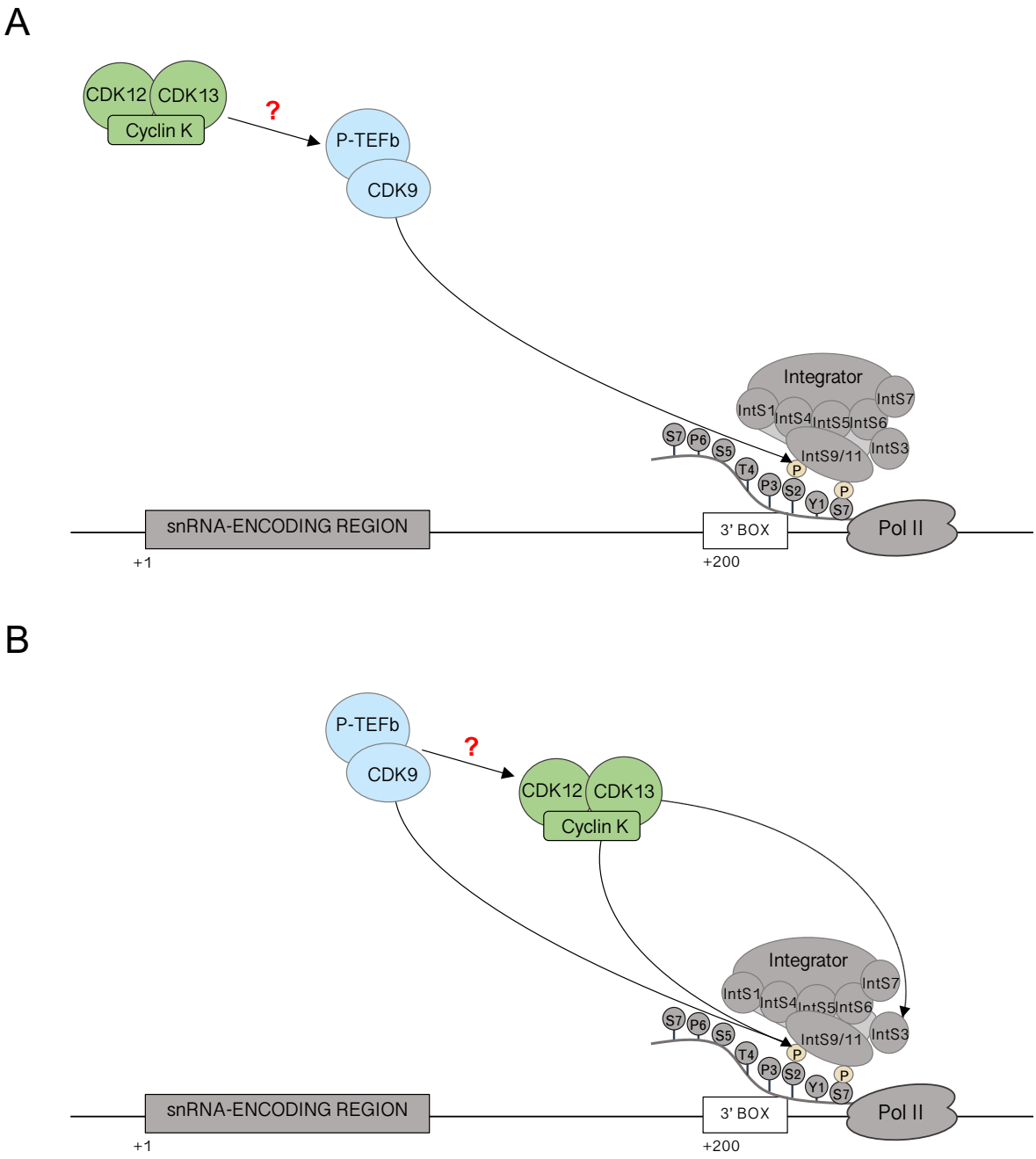


Figure 6.2 CDK12 and CDK13 may function upstream or downstream of CDK9
Schematic model showing CDK12 and CDK13 may function A. Upstream of CDK9, subsequently phosphorylating Ser2 to recruit IntS9/11 and cleave nascent snRNA. B. Downstream of CDK9 where CDK12 and CDK13 phosphorylate Ser2 in addition to the activity of CDK9. CDK9 may also mediate the recruitment of IntS3 through CDK12 and CDK13. P denotes phosphorylation.

6.2.2.3 Future directions

pCDK9-Thr186 ChIP upon CDK12 and CDK13 inhibition will help to determine whether CDK12 and CDK13 activate CDK9 on snRNA genes. Likewise, cyclin K, active CDK12 and CDK13 ChIP upon CDK9 inhibition will clarify if CDK9 works upstream and helps to recruit/activate CDK12 and CDK13. It would also be necessary to investigate the effect of CDK9 inhibition on the association of SPT6 with snRNA genes.

6.2.3 CDK12 and CDK13 may recruit initiation and elongation factors to control snRNA 3' end processing.

CDK12 and CDK13 may help to recruit initiation factors, which are later required for snRNA 3' end processing. The PTF complex and ICE1 and ICE2 of the LEC complex are the only snRNA-specific factors on snRNA genes. Therefore, these proteins may play a role in the obligate coupling of the promoter and the 3' box [267,268]. Previous studies have demonstrated that the knockdown of ELL and ICE1 subunits of the LEC leads to a loss of snRNA transcription [293,398]. Additionally, CDK12 and CDK13 were recently shown to be required for the phosphorylation of PTFα [75], which binds to the PSE. Furthermore, the work presented in Chapter 4 indicates that CDK12 and CDK13 inhibition causes an accumulation of Pol II at the PSE region of U2 snRNA gene (Figure 4.1). CDK12 and CDK13 may phosphorylate initiation factors, such as PTFα, which subsequently recruits other elongation and RNA 3' end processing factors, e.g., SPT6, the LEC complex, and IntS3, which mediate snRNA 3' end processing. On protein-coding genes, the CPSF complex associates with active

TFIID at transcription initiation before dissociating and associating with polymerase upon the start of transcription elongation [348]. CDK12 and CDK13 may complete the obligate coupling of the snRNA-specific promoter and 3' box through a similar mechanism. Using their activity at both ends of the gene, they may phosphorylate an initiation factor at the 5' end, which is subsequently expressed through the gene and required for the snRNA 3' end processing (Fig. 6.3).

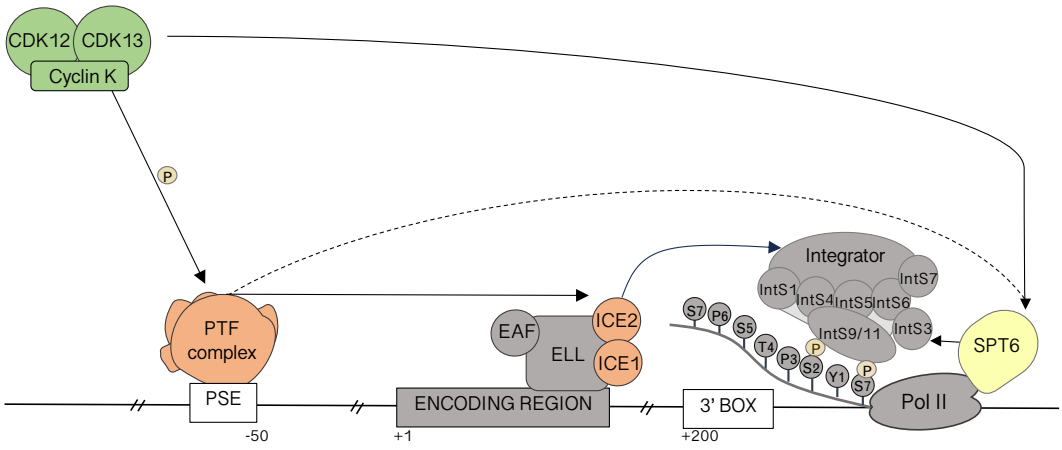


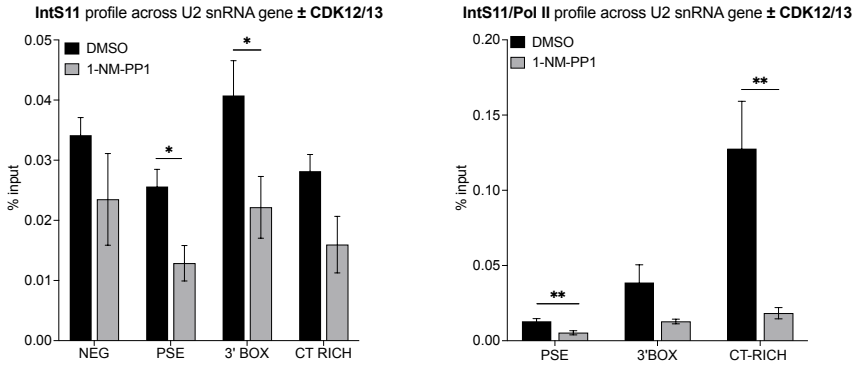
Figure 6.3 CDK12 and CDK13 may recruit initiation and elongation factors to control snRNA 3' end processing.
Schematic model of how CDK12 and CDK13 may regulate the 3'end processing through initiation factors. CDK12/13 may phosphorylate the PTF complex at the beginning of transcription, which may subsequently phosphorylate elongation factors (e.g. ICE2), which mediate the snRNA 3' end processing. Alternatively (dashed line), the phosphorylated initiation factors recruit other factors at the 3' end, e.g. (SPT6), which recruit Integrator subunits and may control snRNA 3' end processing. P denotes phosphorylation and snRNA-specific factors are displayed in orange.

6.2.3.1 Future directions

ICE2 ChIP upon degradation of ICE2 in the ICE2-dTAG cell lines will be required to ensure this factor is lost from the snRNA gene before quantifying the levels of pre and read-through snRNA transcripts by RT-qPCR (Chapter 5). Additionally, it will be useful to determine if the inhibition of CDK9 or CDK12 and CDK13 causes a loss of ICE2 from the snRNA genes. Furthermore, IP-Mass spectrometry of an snRNA-specific initiation factor, such as PTF-α upon inhibition of CDK12 and CDK13 together, may also provide possible candidates which may link 3' end processing of snRNA gene transcripts to the promoter. Taken together, this work provides further insights into the roles of 3' end processing factors (e.g. CDK9) involved in the expression of human snRNA genes while also identifying new factors which play a role in this mechanism. Additionally, I have identified

novel factors which may be implicated in snRNA 3' end regulation, including SPT6, IntS3, and ICE2. However, the mechanism by which these new factors regulate the 3' end processing remains unclear and requires further investigation. The follow-up of this study may reveal a novel mechanism that furthers our understanding of the obligate coupling of the snRNA promoter to the 3' box.

Supplementary



Supplementary figure 1. IntS11 profile upon inhibition of CDK12 and CDK13 AS cells.

(Left) ChIP-qPCR of IntS11 across U2 snRNA gene upon inhibition of CDK12 and CDK13 AS cells. (Right) ChIP-qPCR analysis of IntS11 relative to Pol II across respective genes upon inhibition of CDK12/13 AS cells. n≥3 biological repeats. Statistical test = two-tailed unpaired t-test, * = $p < 0.05$, ** = $p < 0.01$.

References

1. Crick F. Central Dogma of Molecular Biology. *Nature*. 1970;227:561-563.
2. Kornberg RD. Chromatin Structure- A Repeating Unit of Histones and DNA. *Science (1979)*. 1974;184:868-871.
3. Luger K, Mä Der AW, Richmond RK, Sargent DF, Richmond TJ. *Crystal Structure of the Nucleosome Core Particle at 2.8 A ˚ Resolution*. Vol 389.; 1997.
4. Woodcock CLF, Safer JP, Stanchfield JE. *STRUCTURAL REPEATING UNITS IN CHROMATIN I. Evidence for Their General Occurrence*. Vol 97.; 1976.
5. Davey CA, Sargent DF, Luger K, Maeder AW, Richmond TJ. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J Mol Biol*. 2002;319(5):1097-1113. doi:10.1016/S0022-2836(02)00386-8
6. Tremethick DJ. Higher-Order Structures of Chromatin: The Elusive 30 nm Fiber. *Cell*. 2007;128(4):651-654. doi:10.1016/j.cell.2007.02.008
7. Tamaru H. Confining euchromatin/heterochromatin territory: Jumonji crosses the line. *Genes Dev*. 2010;24(14):1465-1478. doi:10.1101/gad.1941010
8. Olins AL, Olins DE. Spheroid Chromatin Units (v Bodies). *Science (1979)*. 1974;183(4122):330-332.
9. Vignali M, Hassan AH, Neely KE, Workman JL. ATP-Dependent Chromatin-Remodeling Complexes. *Mol Cell Biol*. 2000;20(6):1899.

10. Fry CJ, Peterson CL. Chromatin remodeling enzymes: who's on first? *Current Biology*. 2001;11:185-197.

11. Nadal S, Raj R, Mohammed S, Davis BG. Synthetic post-translational modification of histones. *Curr Opin Chem Biol*. 2018;45:35-47. doi:10.1016/j.cbpa.2018.02.004

12. Zhang D, Tang Z, Huang H, et al. Metabolic regulation of gene expression by histone lactylation. *Nature*. 2019;574(7779):575-580. doi:10.1038/s41586-019-1678-1

13. Noma K ichi, David Allis C, S Grewal SI. Transitions in Distinct Histone H3 Methylation Patterns at the Heterochromatin Domain Boundaries. *Science (1979)*. 2001;293:1150-1154. doi:10.1126/science

14. Litt MD, Simpson M, Gaszner M, Allis CD, Felsenfeld G. Correlation between histone lysine methylation and developmental changes at the chicken β-globin locus. *Science (1979)*. 2001;293(5539):2453-2455. doi:10.1126/science.1064413

15. Grunstein M. Yeast Heterochromatin: Minireview Regulation of Its Assembly and Inheritance by Histones. *Cell*. 1998;93:325-328.

16. Rea S, Eisenhaber F, Â nal DO, et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature*. 2000;406:593-599. www.nature.com593

17. Nakayama J, Rice JC, Strahl BD, Allis CD, Grewal SIS. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science (1979)*. 2001;292(5514):110-113. doi:10.1126/science.1060118

18. Lachner M, O'Carroll D, Rea S, Mechtler KT. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature*. 2001;410:116-119.

19. Hui Ng H, ois Robert F, Young RA, Struhl K. Targeted Recruitment of Set1 Histone Methylase by Elongating Pol II Provides a Localized Mark and Memory of Recent Transcriptional Activity. *Orphanides and Reinberg*. 2003;11:709-719.

20. Kirmizis A, Santos-Rosa H, Penkett CJ, et al. Arginine methylation at histone H3R2 controls deposition of H3K4 trimethylation. *Nature*. 2007;449(7164):928-932. doi:10.1038/nature06160

21. Pesavento JJ, Bullock CR, LeDuc RD, Mizzen CA, Kelleher NL. Combinatorial modification of human histone H4 quantitated by two-dimensional liquid chromatography coupled with top down mass spectrometry. *Journal of Biological Chemistry*. 2008;283(22):14927-14937. doi:10.1074/jbc.M709796200

22. Cosgrove MS, Boeke JD, Wolberger C. Regulated nucleosome mobility and the histone code. *Nat Struct Mol Biol*. 2004;11(11):1037-1043. doi:10.1038/nsmb851

23. Struhl K. *Fundamentally Different Logic Minireview of Gene Regulation in Eukaryotes and Prokaryotes*. Vol 98.; 1999.

24. Whitehouse I, Flaus A, Cairns B, White MF, Workman J, Owen-Hughes T. Nucleosome mobilization catalysed by the yeast SWI:SNF complex. *Nature*. 1999;400:784-787.

25. Weiss S, Gladstone L. A mammalian system for the incorporation of cytidine triphosphate into ribonucleic acid. *J Am Chem Soc*. 1959;81(51):4118-4119. https://pubs.acs.org/sharingguidelines

26. Stevens A. Incorporation of the Adenine Ribonucleotide into RNA by cell fraction from E. Coli B. *Biochem Biophys Res Commun*. 1960;3(1):92-96.

27. Huang RCC, Maheshwari N, Bonner J. Enzymatic synthesis of RNA. *Biochem Biophys Res Commun*. 1960;3(6):689-694.

28. WEISS SB, NAKAMOTO T. Net synthesis of ribonucleic acid with a microbial enzyme requiring deoxyribonucleic acid and four ribonucleoside triphosphates. *J Biol Chem*. 1961;236:PC18-PC20. doi:10.1016/s0021-9258(18)64345-5

29. Hurwitz R, Bresler AE. The Incorporation of Ribonucleotides into Ribonucleic Acid*. *Journal of Biological Chemistry* . 1961;236(2):542-548.

30. Roeder R, Rutter W. Multiple Forms of DNA-dependent in Eukaryotic Organisms. *Nature*. 1969;224:234-237.

31. Roedert RG, Rutter WJ. Specific Nucleolar and Nucleoplasmic RNA Polymerases*. *Proceedings of the National Academy of Sciences*. 1970;65(3):675-682. https://www.pnas.org

32. Redinger C, Mandel JL, Gissinger F, Chambon P. α-amanitin: A specific inhibitor of one of two DNA-dependent RNA polymerase activities from calf thymus. *Biochem Biophys Res Commun*. 1969;38(1):165-171.

33. Sentenac A. Eukaryotic RNA polymerase. *Crit Rev Biochem Mol Biol*. 1985;18(1):31-90. doi:10.3109/10409238509082539

34. Dieci G, Fiorino G, Castelnuovo M, Teichmann M, Pagano A. The expanding RNA polymerase III transcriptome. *Trends in Genetics*. 2007;23(12):614-622. doi:10.1016/j.tig.2007.09.001

35. Griesenbeck J, Tschochner H, Grohmann D. Structure and function of RNA polymerases and the transcription machineries. *Subcell Biochem*. 2017;83:225-270. doi:10.1007/978-3-319-46503-6_9

36. Allison LA, Moyle M, Shales M, James Lngles Banting C. Extensive Homology among the Largest Subunits of Eukaryotic and Prokaryotic RNA Polymerases. *Cell*. 1985;42:599-610.

37. Corden JL, Cadenat DL, Ahearn JM, Dahmust ME. A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II (transcription/gene expression/protein phosphorylation). *PNAS*. 1985;82:7934-7938. https://www.pnas.org

38. Rosonina E, Blencowe BJ. Analysis of the requirement for RNA polymerase II CTD heptapeptide repeats in pre-mRNA splicing and 3'-end cleavage. *RNA*. 2004;10(4):581-589. doi:10.1261/rna.5207204

39. Allison LA, K-c Wong J, Danial Fitzpatrick V, Moyle M, James Ingles C. *The C-Terminal Domain of the Largest Subunit of RNA Polymerase II of Saccharomyces Cerevisiae, Drosophila Melanogaster, and Mammals: A Conserved Structure with an Essential Function*. Vol 8.; 1988.

40. Bartolomei MS, Halden NF, Ruta Cullen C, Corden JL. *Genetic Analysis of the Repetitive Carboxyl-Terminal Domain of the Largest Subunit of Mouse RNA Polymerase II*. Vol 8.; 1988.

41. Hsin JP, Sheth A, Manley J. RNAP II CTD Phosphorylated on Threonine-4 Is Required for Histone mRNA 3' End Processing. *Science (1979)*. 2011;334:683-686. doi:10.1126/science.1207056

42. Hintermair C, Heidemann M, Koch F, et al. Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. *EMBO Journal*. 2012;31(12):2784-2797. doi:10.1038/emboj.2012.123

43. Stiller JW, Mcconaughy BL, Hall BD. Evolutionary complementation for polymerase II CTD function. *Yeast*. 2000;16(1):57-64. doi:10.1002/(SICI)1097-0061(20000115)16:1<57::AID-YEA509>3.0.CO;2-E

44. West ML, Corden JL. Construction and Analysis of Yeast RNA Polymerase II CTD Deletion and Substitution Mutations. *Genetics*. 1995;140:1223-1233. https://academic.oup.com/genetics/article/140/4/1223/6013373

45. Corden JL, Cadenat DL, Ahearn JM, Dahmust ME. A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II (transcription/gene expression/protein phosphorylation). *PNAS*. 1985;82:7934-7938. https://www.pnas.org

46. Chapman RD, Heidemann M, Hintermair C, Eick D. Molecular evolution of the RNA polymerase II CTD. *Trends in Genetics*. 2008;24(6):289-296. doi:10.1016/j.tig.2008.03.010

47. Liu P, Kenney JM, Stiller JW, Greenleaf AL. Genetic organization, length conservation, and evolution of RNA polymerase II carboxyl-terminal domain. *Mol Biol Evol*. 2010;27(11):2628-2641. doi:10.1093/molbev/msq151

48. Hsin JP, Manley JL. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev*. 2012;26(19):2119-2137. doi:10.1101/gad.200303.112

49. Chapman RD, Palancade B, Lang A, Bensaude O, Eick D. The last CTD repeat of the mammalian RNA polymerase II large subunit is important for its stability. *Nucleic Acids Res*. 2004;32(1):35-44. doi:10.1093/nar/gkh172

50. Russell P, Nurse P. Schizosacchammyces pombe and Saccharomyces cerevisiae: A Look at Yeasts Divided. *Cell*. 1966;45:761-762.

51. Kohoutek J, Blazek D. Cyclin K goes with Cdk12 and Cdk13. *Cell Div*. 2012;7. doi:10.1186/1747-1028-7-12

52. Malumbres M. Cyclin-dependent kinases. *Genome Biol*. 2014;15(6). doi:10.1186/gb4184

53. Li N, Fassl A, Chick J, et al. Cyclin C is a haploinsufficient tumour suppressor. *Nat Cell Biol*. 2014;16(11):1080-1091. doi:10.1038/ncb3046

54. Zaborowska J, Egloff S, Murphy S. The pol II CTD: new twists in the tail. *Nat Struct Mol Biol*. 2016;23:771-777.

55. Fisher RP, Morgan D 0. *A Novel Cyclin Associates with MOWCDK7 to Form the CDK-Activating Kinase*. Vol 78.; 1994.

56. Akhtar MS, Heidemann M, Tietjen JR, et al. TFIIH Kinase Places Bivalent Marks on the Carboxy-Terminal Domain of RNA Polymerase II. *Mol Cell*. 2009;34(3):387-393. doi:10.1016/j.molcel.2009.04.016

57. Glover-Cutter K, Larochelle S, Erickson B, et al. TFIIH-Associated Cdk7 Kinase Functions in Phosphorylation of C-Terminal Domain Ser7 Residues,

Promoter-Proximal Pausing, and Termination by RNA Polymerase II. *Mol Cell Biol*. 2009;29(20):5455-5464. doi:10.1128/mcb.00637-09

58. Wong KH, Jin Y, Struhl K. TFIIH Phosphorylation of the Pol II CTD Stimulates Mediator Dissociation from the Preinitiation Complex and Promoter Escape. *Mol Cell*. 2014;54(4):601-612. doi:10.1016/j.molcel.2014.03.024

59. Fisher RP. Cdk7: a kinase at the core of transcription and in the crosshairs of cancer drug discovery. *Transcription*. 2019;10(2):47-56. doi:10.1080/21541264.2018.1553483

60. Velychko T, Mohammad E, Ferrer-Vicens I, et al. CDK7 kinase activity promotes RNA polymerase II promoter escape by facilitating initiation factor release. *Mol Cell*. 2024;84(12):2287-2303.e10. doi:10.1016/j.molcel.2024.05.007

61. Tassan JP, Jaquenoud M, Leopoldt P, Schultz SJ, Nigg EA. *Identification of Human Cyclin-Dependent Kinase 8, a Putative Protein Kinase Partner for Cyclin C*. Vol 92.; 1995.

62. Rickert P, Corden JL, Lees E. Cyclin C/CDK8 and cyclin H/CDK7/p36 are biochemically distinct CTD kinases. *Oncogene*. 1999;18:1093-1102. http://www.stockton-press.co.uk/onc

63. Ramanathan Y, Rajpara SM, Reza SM, et al. Three RNA Polymerase II Carboxyl-terminal Domain Kinases Display Distinct Substrate Preferences. *Journal of Biological Chemistry*. 2001;276(14):10913-10920. doi:10.1074/jbc.M010975200

64. Bartkowiak B, Greenleaf AL. Expression, purification, and identification of associated proteins of the full-length hCDK12/CyclinK complex. *Journal of*

Biological Chemistry. 2015;290(3):1786-1795. doi:10.1074/jbc.M114.612226

65. Peng J, Zhu Y, Milton JT, Price DH. Identification of multiple cyclin subunits of human P-TEFb. *Genes Dev*. 1998;12:755-762. www.genesdev.org

66. Zhou M, Halanski MA, Radonovich MF, et al. *Tat Modifies the Activity of CDK9 To Phosphorylate Serine 5 of the RNA Polymerase II Carboxyl-Terminal Domain during Human Immunodeficiency Virus Type 1 Transcription*. Vol 20.; 2000.

67. Kim YK, Bourgeois CF, Isel C, Churcher MJ, Karn J. Phosphorylation of the RNA Polymerase II Carboxyl-Terminal Domain by CDK9 Is Directly Responsible for Human Immunodeficiency Virus Type 1 Tat-Activated Transcriptional Elongation. *Mol Cell Biol*. 2002;22(13):4622-4637. doi:10.1128/mcb.22.13.4622-4637.2002

68. Pinhero R, Liaw P, Bertens K, Yankulov K. Three cyclin-dependent kinases preferentially phosphorylate different parts of the C-terminal domain of the large subunit of RNA polymerase II. *Eur J Biochem*. 2004;271(5):1004-1014. doi:10.1111/j.1432-1033.2004.04002.x

69. Czudnochowski N, Bösken CA, Geyer M. Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. *Nat Commun*. 2012;3. doi:10.1038/ncomms1846

70. Guen VJ, Gamble C, Flajolet M, et al. CDK10/cyclin M is a protein kinase that controls ETS2 degradation and is deficient in STAR syndrome. *Proc Natl Acad Sci U S A*. 2013;110(48):19525-19530. doi:10.1073/pnas.1306814110

71. Berke JD, Ronique Sgambato V, Zhu PP, et al. Dopamine and Glutamate Induce Distinct Striatal Splice Forms of Ania-6, an RNA Polymerase II-Associated Cyclin. *Neuron*. 2001;32:277-287.

72. Gajdušková P, Ruiz de los Mozos I, Rájecký M, Hluchý M, Ule J, Blazek D. CDK11 is required for transcription of replication-dependent histone genes. *Nat Struct Mol Biol*. 2020;27(5):500-510. doi:10.1038/s41594-020-0406-8

73. Pak V, Eifler TT, Jäger S, Krogan NJ, Fujinaga K, Peterlin BM. CDK11 in TREX/THOC regulates HIV mRNA 3' end processing. *Cell Host Microbe*. 2015;18(5):560-570. doi:10.1016/j.chom.2015.10.012

74. Bösken CA, Farnung L, Hintermair C, et al. The structure and substrate specificity of human Cdk12/Cyclin K. *Nat Commun*. 2014;5. doi:10.1038/ncomms4505

75. Fan Z, Devlin JR, Hogg SJ, et al. CDK13 cooperates with CDK12 to control global RNA polymerase II processivity. *Sci Adv*. 2020;6.

76. Greenleaf AL. Human CDK12 and CDK13, multi-tasking CTD kinases for the new millenium. *Transcription*. 2019;10(2):91-110. doi:10.1080/21541264.2018.1535211

77. Tellier M, Zaborowska J, Caizzi L, et al. CDK12 globally stimulates RNA polymerase II transcription elongation and carboxyl-terminal domain phosphorylation. *Nucleic Acids Res*. 2020;48(14):7712-7727. doi:10.1093/nar/gkaa514

78. Greifenberg AK, Hönig D, Pilarova K, et al. Structural and Functional Analysis of the Cdk13/Cyclin K Complex. *Cell Rep*. 2016;14(2):320-331. doi:10.1016/j.celrep.2015.12.025

79. Zhang T, Kwiatkowski N, Olson CM, et al. Covalent targeting of remote cysteine residues to develop CDK12 and CDK13 inhibitors. *Nat Chem Biol*. 2016;12(10):876-884. doi:10.1038/nchembio.2166

80. Liang K, Gao X, Gilmore JM, et al. Characterization of Human Cyclin-Dependent Kinase 12 (CDK12) and CDK13 Complexes in C-Terminal Domain Phosphorylation, Gene Transcription, and RNA Processing. *Mol Cell Biol*. 2015;35(6):928-938. doi:10.1128/mcb.01426-14

81. Lim S, Kaldis P. Cdks, cyclins and CKIs: Roles beyond cell cycle regulation. *Development*. 2013;140(15):3079-3093. doi:10.1242/dev.091744

82. Kajitani T, Kato H, Chikashige Y, et al. Ser7 of RNAPII-CTD facilitates heterochromatin formation by linking ncRNA to RNAi. *Proc Natl Acad Sci U S A*. 2017;114(52):E11208-E11217. doi:10.1073/pnas.1714579115

83. Yague-Sanz C, Migeot V, Larochelle M, et al. Chromatin remodeling by Pol II primes efficient Pol III transcription. *Nat Commun*. 2023;14(1). doi:10.1038/s41467-023-39387-4

84. Singh N, Asalam M, Ansari MO, Gerasimova NS, Studitsky VM, Akhtar MS. Transcription by RNA polymerase II and the CTD-chromatin crosstalk. *Biochem Biophys Res Commun*. 2022;599:81-86. doi:10.1016/j.bbrc.2022.02.039

85. Hampsey M, Reinberg D. Tails of Intrigue: Phosphorylation of RNA Polymerase II Mediates Histone Methylation. *Cell*. 2003;113:429-432.

86. Søgaard TMM, Svejstrup JQ. Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with

mediator. *Journal of Biological Chemistry*. 2007;282(19):14113-14120. doi:10.1074/jbc.M701345200

87. Wong KH, Jin Y, Struhl K. TFIIH Phosphorylation of the Pol II CTD Stimulates Mediator Dissociation from the Preinitiation Complex and Promoter Escape. *Mol Cell*. 2014;54(4):601-612. doi:10.1016/j.molcel.2014.03.024

88. Fabrega C, Shen V, Shuman S, Lima CD. Structure of an mRNA Capping Enzyme Bound to the Phosphorylated Carboxy-Terminal Domain of RNA Polymerase II. *Mol Cell*. 2003;11:1549-1561.

89. Komarnitsky P, Cho EJ, Buratowski S. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev*. 2000;14(19):2452-2460. doi:10.1101/gad.824700

90. Schroeder SC, Schwer B, Shuman S, Bentley D. Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes Dev*. 2000;14(19):2435-2440. doi:10.1101/gad.836300

91. Price DH. P-TEFb, a Cyclin-Dependent Kinase Controlling Elongation by RNA Polymerase II. *Mol Cell Biol*. 2000;20(8):2629-2634.

92. Yamada T, Yamaguchi Y, Inukai N, Okamoto S, Mura T, Handa H. P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Mol Cell*. 2006;21(2):227-237. doi:10.1016/j.molcel.2005.11.024

93. Lin X, Taube R, Fujinaga K, Matija Peterlin B. P-TEFb containing cyclin K and Cdk9 can activate transcription via RNA. *Journal of Biological Chemistry*. 2002;277(19):16873-16878. doi:10.1074/jbc.M200117200

94. Li Q, Price JP, Byers SA, Cheng D, Peng J, Price DH. Analysis of the large inactive P-TEFb complex indicates that it contains one 7SK molecule, a dimer of HEXIM1 or HEXIM2, and two P-TEFb molecules containing Cdk9 phosphorylated at threonine 186. *Journal of Biological Chemistry*. 2005;280(31):28819-28826. doi:10.1074/jbc.M502712200

95. Laitem C, Zaborowska J, Isa NF, Kufs J, Dienstbier M, Murphy S. CDK9 inhibitors define elongation checkpoints at both ends of RNA polymerase II-transcribed genes. *Nat Struct Mol Biol*. 2015;22(5):396-403. doi:10.1038/nsmb.3000

96. Egloff S, Szczepaniak SA, Dienstbier M, Taylor A, Knight S, Murphy S. The integrator complex recognizes a new double mark on the RNA polymerase II carboxyl-terminal domain. *Journal of Biological Chemistry*. 2010;285(27):20564-20569. doi:10.1074/jbc.M110.132530

97. Egloff S. Role of Ser7 phosphorylation of the CTD during transcription of snRNA genes. *RNA Biol*. 2012;9(8):1033-1038. doi:10.4161/rna.21166

98. Egloff S, Murphy S. Cracking the RNA polymerase II CTD code. *Trends in Genetics*. 2008;24(6):280-288. doi:10.1016/j.tig.2008.03.008

99. Egloff S, Zaborowska J, Laitem C, Kiss T, Murphy S. Ser7 phosphorylation of the CTD recruits the RPAP2 ser5 phosphatase to snRNA genes. *Mol Cell*. 2012;45(1):111-122. doi:10.1016/j.molcel.2011.11.006

100. Egloff S, O'Reilly D, Chapman R, et al. Serine-7 of the RNA Polymerase II CTD Is Specifically Required for snRNA Gene Expression. *Science (1979)*. 2007;318(5857):1775-1777. doi:10.1126/science.1149374

101. Zhang DW, Rodríguez-Molina JB, Tietjen JR, Nemec CM, Ansari AZ. Emerging Views on the CTD Code. *Genet Res Int*. 2012;2012:1-19. doi:10.1155/2012/347214

102. Richard P, Manley JL. Transcription termination by nuclear RNA polymerases. *Genes Dev*. 2009;23(11):1247-1269. doi:10.1101/gad.1792809

103. Kwiatkowski N, Zhang T, Rahl PB, et al. Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. *Nature*. 2014;511(7511):616-620. doi:10.1038/nature13393

104. Steinberg SF. Post-translational modifications at the ATP-positioning G-loop that regulate protein kinase activity. *Pharmacol Res*. 2018;135:181-187. doi:10.1016/j.phrs.2018.07.009

105. Reinhardt R, Leonard TA. A critical evaluation of protein kinase regulation by activation loop autophosphorylation. *Elife*. 2023;12. doi:10.7554/eLife.88210

106. Li T, Tang HC, Tsai KL. Unveiling the noncanonical activation mechanism of CDKs: insights from recent structural studies. *Front Mol Biosci*. 2023;10:1-9. doi:10.3389/fmolb.2023.1290631

107. Łukasik P, Baranowska-bosiacka I, Kulczycka K, Gutowska I. Inhibitors of cyclin-dependent kinases: Types and their mechanism of action. *Int J Mol Sci*. 2021;22(6):1-23. doi:10.3390/ijms22062806

108. Echalier A, Endicott JA, Noble MEM. Recent developments in cyclin-dependent kinase biochemical and structural studies. *Biochim Biophys Acta Proteins Proteom*. 2010;1804(3):511-519. doi:10.1016/j.bbapap.2009.10.002

109. Jiang B, Gao Y, Che J, et al. Discovery and resistance mechanism of a selective CDK12 degrader. *Nat Chem Biol*. 2021;17(6):675-683. doi:10.1038/s41589-021-00765-y

110. Bishop A, Ubersax J, Petsch D, et al. A chemical switch for inhibitor-sensitive alleles of any protein kinase. *Nature*. 2000;407. www.nature.com

111. Liu Y, Shah K, Yang F, Witucki L, Shokat KM. A Molecular Gate which Controls Unnatural ATP Analogue Recognition by the Tyrosine Kinase V-Src. *Bioorg Med Chem*. 6:1219-1226.

112. Kadonaga JT. Perspectives on the RNA polymerase II core promoter. *Wiley Interdiscip Rev Dev Biol*. 2012;1(1):40-51. doi:10.1002/wdev.21

113. Juven-Gershon T, Kadonaga JT. Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev Biol*. 2010;339(2):225-229. doi:10.1016/j.ydbio.2009.08.009

114. Juven-Gershon T, Hsu JY, Theisen JW, Kadonaga JT. The RNA polymerase II core promoter - the gateway to transcription. *Curr Opin Cell Biol*. 2008;20(3):253-259. doi:10.1016/j.ceb.2008.03.003

115. Antequera F. Structure, function and evolution of CpG island promoters. *Cellular and Molecular Life Sciences*. 2003;60(8):1647-1658. doi:10.1007/s00018-003-3088-6

116. Vo L, Wang YL, Kassavetis GA, Kadonaga JT. The punctilious RNA polymerase II core promoter. *Genes Dev*. 2017;31:1289-1301. doi:10.1101/gad.303149

117. Goldberg ML. *Sequence Analysis of Drosophila Histone Genes*. Stanford University; 1979.

118. Butler JEF, Kadonaga JT. The RNA polymerase II core promoter: A key component in the regulation of gene expression. *Genes Dev*. 2002;16(20):2583-2592. doi:10.1101/gad.1026202

119. Ngoc LV, Cassidy CJ, Huang CY, Duttke SHC, Kadonaga JT. The human initiator is a distinct and abundant element that is precisely positioned in focused core promoters. *Genes Dev*. 2017;31(1):6-11. doi:10.1101/gad.293837.116

120. Suzuki Y, Tsunoda T, Sese J, et al. Identification and Characterization of the Potential Promoter Regions of 1031 Kinds of Human Genes. *Genome Res*. 2001;11(5):677-684. doi:10.1101/gr.164001

121. Kim Y, Geiger JH, Hahn S, Sigler PB. Crystal structure of a yeast TBP/TATA-box complex Orientation and structure of the TATA box. *Nature*. 1993;365:512-520.

122. Kim J, Nikolov D, Burley S. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature*. 1993;365:520-527.

123. Kuhlman TC, Cho H, Reinberg D, Hernandez AN. The General Transcription Factors IIA, IIB, IIF, and IIE Are Required for RNA Polymerase II Transcription from the Human U1 Small Nuclear RNA Promoter. *Mol Cell Biol*. 1999;19(3):2130-2141.

124. Hahn S, Buratowski S, Sharp PA, Guarentes L. *Isolation of the Gene Encoding the Yeast TATA Binding Protein TFIID: A Gene Identical to the SW5 Suppressor of Ty Element Insertions*. Vol 58.; 1999.

125. Eisenmann DM, Doilard C, Winston F. SPT15, the Gene Encoding the Yeast TATA Binding Factor TFIID, Is Required for Normal Transcription Initiation In Vivo. *Cell*. 1989;58:1183-1191.

126. Neigeborn A N D L, Carlson M. GENES AFFECTING THE REGULATION OF SUC2 GENE EXPRESSION BY GLUCOSE REPRESSION IN SACCHAROMYCES CEREVISIAE. *Genetics*. 1984;108:845-858. https://academic.oup.com/genetics/article/108/4/845/5996403

127. Henry NL, Campbell AM, Feaver WI, et al. TFIIF-TAF-RNA polymerase II connection. *Genes Dev*. 1994;8:2868-2878.

128. Sun Z wen, Hampsey M. *Identification of the Gene (SSU71/TFGJ) Encoding the Largest Subunit of Transcription Factor TFIIF as a Suppressor of a TFIIB Mutation in Saccharomyces Cerevisiae (SSU71/Transcription Start-Site Selection) Abbreviations: Pol II, RNA Polymerase II; TBP, TATA-Binding Protein; Csm-, Cold-Sensitive; Tsm-, Heat-Sensitive*. Vol 92.; 1995.

129. Feaver WJ, Henry NL, Bushnell DA, et al. Yeast TFIIE: Cloning, expression, and homology to vertebrate proteins. *Journal of Biological Chemistry*. 1994;269(44):27549-27553. doi:10.1016/s0021-9258(18)47019-6

130. Guzder SN, Qiu H, Sommers CH, Sung P, Prakashttll L, Prakash S. DNA repair gene RAD3 of s. cerevisiae is essential for transcription by RNA polymerase II. *Nature*. 1994;367:91-94.

131. Myers LC, Gustafsson CM, Bushnell DA, et al. The Med proteins of yeast and their function through the RNA polymerase II carboxy-terminal domain. *Genes Dev*. 1998;12:45-54. www.genesdev.org

132. Asturias FJ, Jiang YW, Myers LC, Gustafsson CM, Kornberg RD. Conserved Structures of Mediator and RNA Polymerase II Holoenzyme. *Science (1979)*. 1999;283:985-987.

133. Davis JA, Takagi Y, Kornberg RD, Asturias FJ. Structure of the Yeast RNA Polymerase II Holoenzyme: Mediator Conformation and Polymerase Interaction. *Mol Cell*. 2002;10:409-415.

134. Boube M, Joulia L, Cribbs DL, Bourbon HM. *Review Evidence for a Mediator of RNA Polymerase II Transcriptional Regulation Conserved from Yeast to Man Lee and Young, 2000; Woychik and Hampsey, 2002). However, in Vitro Transcription Assays Strongly Sug-Gested the Requirement for Additional Components Link-Ing Gene-Specific Transcriptional Regulators to Pol II and GTFs. The First Evidence Came from in Vivo Studies In*. Vol 110.; 2002. http://www.cell.com/cgi/content/

135. Sato S, Tomomori-Sato C, Tsai KL, et al. Role for the MED21-MED7 Hinge in Assembly of the Mediator-RNA Polymerase II Holoenzyme. *Journal of Biological Chemistry*. 2016;291(52):26886-26898. doi:10.1074/jbc.M116.756098

136. Malik S, Roeder RG. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nat Rev Genet*. 2010;11(11):761-772. doi:10.1038/nrg2901

137. Allen BL, Taatjes DJ. The Mediator complex: A central integrator of transcription. *Nat Rev Mol Cell Biol*. 2015;16(3):155-166. doi:10.1038/nrm3951

138. Tsai KL, Yu X, Gopalan S, et al. Mediator structure and rearrangements required for holoenzyme formation. *Nature*. 2017;544(7649):196-201. doi:10.1038/nature21393

139. Luse DS, Jacob GA. Abortive initiation by RNA polymerase II in vitro at the adenovirus 2 major late promoter. *J Biol Chem*. 1987;262(31):14990-14997. doi:10.1016/s0021-9258(18)48127-6

140. CPHolstege F, Fiedler U, Timmers Ht. *Three Transitions in the RNA Polymerase II Transcription Complex during Initiation*. Vol 16.; 1997. https://www.embopress.org

141. Pal M, Ponticelli AS, Luse DS. The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. *Mol Cell*. 2005;19(1):101-110. doi:10.1016/j.molcel.2005.05.024

142. Kostrewa D, Zeller ME, Armache KJ, et al. RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature*. 2009;462(7271):323-330. doi:10.1038/nature08548

143. Čabart P, Újvári A, Pal M, Luse DS. Transcription factor TFIIF is not required for initiation by RNA polymerase II, but it is essential to stabilize transcription factor TFIIB in early elongation complexes. *PNAS*. 2011;108:15786-15791. doi:10.1073/pnas.1104591108/-/DCSupplemental

144. Dvir A. Promoter escape by RNA polymerase II. *Biochim Biophys Acta*. 2002;1577:208-223. www.bba-direct.com

145. Rodriguez CR, Cho EJ, Keogh MC, Moore CL, Greenleaf AL, Buratowski S. *Kin28, the TFIIH-Associated Carboxy-Terminal Domain Kinase, Facilitates*

the Recruitment of MRNA Processing Machinery to RNA Polymerase II. Vol 20.; 2000.

146. Ho CK, Shuman S. Distinct Roles for CTD Ser-2 and Ser-5 Phosphorylation in the Recruitment and Allosteric Activation of Mammalian mRNA Capping Enzyme. *Mol Cell*. 1999;3:405-411.
147. Larochelle S, Amat R, Glover-Cutter K, et al. Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II. *Nat Struct Mol Biol*. 2012;19(11):1108-1115. doi:10.1038/nsmb.2399
148. Zhan Y, Grabbe F, Oberbeckmann E, Dienemann C, Cramer P. Three-step mechanism of promoter escape by RNA polymerase II. *Mol Cell*. 2024;84(9):1699-1710.e6. doi:10.1016/j.molcel.2024.03.016
149. Missra A, Gilmour DS. Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the Drosophila RNA polymerase II transcription elongation complex. *PNAS*. 2010;107:11301-11306. doi:10.1073/pnas.1000681107/-/DCSupplemental
150. Core L, Adelman K. Promoter-proximal pausing of RNA polymerase II: A nexus of gene regulation. *Genes Dev*. 2019;33(15-16):960-982. doi:10.1101/gad.325142.119
151. Yamaguchi Y, Takagi T, Wada T, et al. *NELF, a Multisubunit Complex Containing RD, Cooperates with DSIF to Repress RNA Polymerase II Elongation*. Vol 97.; 1999.
152. Vos SM, Farnung L, Urlaub H, Cramer P. Structure of paused transcription complex Pol II–DSIF–NELF. *Nature*. 2018;560(7720):601-606. doi:10.1038/s41586-018-0442-2

153. Rougviet AE, Lis JT. Postinitiation Transcriptional Control in Drosophila melanogaster. *Molecular and CellularBiology*. 1990;10(11).

154. Krumm A, Hickey LB, Groudine M. Promoter-proximal pausing of RNA Eolymerase II defines a general rate-limiting step after transcription initiation. *Genes Dev*. 1995;9:559-572.

155. Rasmussen EB, Lis JT. *Short Transcripts of the Ternary Complex Provide Insight into RNA Polymerase II Elongational Pausing*. Vol 252.; 1995.

156. Vos SM, Farnung L, Boehning M, et al. Structure of activated transcription complex Pol II–DSIF–PAF–SPT6. *Nature*. 2018;560(7720):607-612. doi:10.1038/s41586-018-0440-4

157. Yu M, Yang W, Ni T, et al. RNA polymerase II–associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. *Science (1979)*. 2015;350(6266):1379-1383. doi:10.1126/science.aab0892

158. Chun Y, Joo YJ, Suh H, et al. Selective Kinase Inhibition Shows That Bur1 (Cdk9) Phosphorylates the Rpb1 Linker In Vivo . *Mol Cell Biol*. 2019;39(15). doi:10.1128/mcb.00602-18

159. Fujinaga K, Irwin D, Huang Y, Taube R, Kurosu T, Peterlin BM. Dynamics of Human Immunodeficiency Virus Transcription: P-TEFb Phosphorylates RD and Dissociates Negative Effectors from the Transactivation Response Element. *Mol Cell Biol*. 2004;24(2):787-795. doi:10.1128/mcb.24.2.787-795.2004

160. Žumer K, Maier KC, Farnung L, et al. Two distinct mechanisms of RNA polymerase II elongation stimulation in vivo. *Mol Cell*. 2021;81(15):3096-3109.e8. doi:10.1016/j.molcel.2021.05.028

161. Yik JHN, Chen R, Nishimura R, Jennings JL, Link AJ, Zhou Q. Inhibition of P-TEFb (CDK9/Cyclin T) Kinase and RNA Polymerase II Transcription by the Coordinated Actions of HEXIM1 and 7SK snRNA. *Mol Cell*. 2003;12:971-982.

162. Luo Z, Lin C, Guest E, et al. The Super Elongation Complex Family of RNA Polymerase II Elongation Factors: Gene Target Specificity and Transcriptional Output. *Mol Cell Biol*. 2012;32(13):2608-2617. doi:10.1128/mcb.00182-12

163. Lin C, Smith ER, Takahashi H, et al. AFF4, a Component of the ELL/P-TEFb Elongation Complex and a Shared Subunit of MLL Chimeras, Can Link Transcription Elongation to Leukemia. *Mol Cell*. 2010;37(3):429-437. doi:10.1016/j.molcel.2010.01.026

164. Shilatifard A;, Lane WS, Jackson KW, Conaway RC, Conaway JW. An RNA polymerase II elongation factor encoded by the human ELL gene. *Science (1979)*. 1999;271(5257).

165. Zheng H, Qi Y, Hu S, et al. Identification of Integrator-PP2A complex (INTAC), an RNA polymerase II phosphatase. *Science (1979)*. 2020;370(6520). doi:10.1126/science.abb5872

166. Elrod ND, Henriques T, Huang KL, et al. The Integrator Complex Attenuates Promoter-Proximal Transcription at Protein-Coding Genes. *Mol Cell*. 2019;76(5):738-752.e7. doi:10.1016/j.molcel.2019.10.034

167. Baillat D, Hakimi MA, Näär AM, Shilatifard A, Cooch N, Shiekhattar R. Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell*. 2005;123(2):265-276. doi:10.1016/j.cell.2005.08.019

168. Azuma N, Yokoi T, Tanaka T, et al. Integrator complex subunit 15 controls mRNA splicing and is critical for eye development. *Hum Mol Genet*. 2023;32(12):2032-2045. doi:10.1093/hmg/ddad034

169. Ryan K, Calvo O, Manley JL. Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *RNA*. 2004;10(4):565-573. doi:10.1261/rna.5214404

170. Tatomer DC, Elrod ND, Liang D, et al. The Integrator complex cleaves nascent mRNAs to attenuate transcription. *Genes Dev*. 2019;33(21-22):1525-1538. doi:10.1101/gad.330167.119

171. Lai F, Gardini A, Zhang A, Shiekhattar R. Integrator mediates the biogenesis of enhancer RNAs. *Nature*. 2015;525(7569):399-403. doi:10.1038/nature14906

172. Fianu I, Chen Y, Dienemann C, et al. Structural basis of Integrator-mediated transcription regulation. *Science (1979)*. 2021;374:883-887. https://www.science.org

173. Yamamoto J, Hagiwara Y, Chiba K, et al. DSIF and NELF interact with Integrator to specify the correct post-transcriptional fate of snRNA genes. *Nat Commun*. 2014;5. doi:10.1038/ncomms5263

174. West S, Gromak N, Proudfoot NJ. Human 5' → 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature*. 2004;432(7016):522-525. doi:10.1038/nature03035

175. Huang KL, Jee D, Stein CB, et al. Integrator Recruits Protein Phosphatase 2A to Prevent Pause Release and Facilitate Transcription Termination. *Mol Cell*. 2020;80(2):345-358.e9. doi:10.1016/j.molcel.2020.08.016

176. Vervoort SJ, Welsh SA, Devlin JR, et al. The PP2A-Integrator-CDK9 axis fine-tunes transcription and can be targeted therapeutically in cancer. *Cell*. 2021;184(12):3143-3162.e32. doi:10.1016/j.cell.2021.04.022

177. Hu J, Lutz CS, Wilusz J, Tian B. Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation. *RNA*. 2005;11(10):1485-1493. doi:10.1261/rna.2107305

178. Proudfoot NJ. Ending the message: Poly(A) signals then and now. *Genes Dev*. 2011;25(17):1770-1782. doi:10.1101/gad.17268411

179. Chan S, Choi EA, Shi Y. Pre-mRNA 3'-end processing complex assembly and function. *Wiley Interdiscip Rev RNA*. 2011;2(3):321-335. doi:10.1002/wrna.54

180. Shi Y, Di Giammartino DC, Taylor D, et al. Molecular Architecture of the Human Pre-mRNA 3' Processing Complex. *Mol Cell*. 2009;33(3):365-376. doi:10.1016/j.molcel.2008.12.028

181. Wahle E, Rüegsegger U. 3'-End processing of pre-mRNA in eukaryotes. *FEMS Microbiol Rev*. 1999;23(3):277-295. doi:10.1111/j.1574-6976.1999.tb00400.x

182. Schönemann L, Kühn U, Martin G, et al. Reconstitution of CPSF active in polyadenylation: Recognition of the polyadenylation signal by WDR33. *Genes Dev*. 2014;28(21):2381-2393. doi:10.1101/gad.250985.114

183. Tudek A, Lloret-Llinares M, Heick Jensen T. The multitasking polyA tail: Nuclear RNA maturation, degradation and export. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2018;373(1762). doi:10.1098/rstb.2018.0169

184. Romeo V, Schümperli D. Cycling in the nucleus: Regulation of RNA 3' processing and nuclear organization of replication-dependent histone genes. *Curr Opin Cell Biol*. 2016;40:23-31. doi:10.1016/j.ceb.2016.01.015

185. Hoon Ahn S, Kim M, Buratowski S. Phosphorylation of Serine 2 within the RNA Polymerase II C-Terminal Domain Couples Transcription and 3' End Processing. *Mol Cell*. 2004;13:67-76. http://www.molecule.org/cgi/content/full/13/1/67/

186. Logan J, Falck-Pedersent E, Darnell JE, Shenk T. A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse BmaJ-globin gene. *PNAS*. 1987;84:8306-8310. https://www.pnas.org

187. Lunde BM, Reichow SL, Kim M, et al. Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat Struct Mol Biol*. 2010;17(10):1195-1201. doi:10.1038/nsmb.1893

188. Zhang Z, Gilmour DS. Pcf11 is a termination factor in Drosophila that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. *Mol Cell*. 2006;21(1):65-74. doi:10.1016/j.molcel.2005.11.002

189. Zhang Z, Fu J, Gilmour DS. CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11. *Genes Dev*. 2005;19(13):1572-1580. doi:10.1101/gad.1296305

190. Mandel CR, Kaneko S, Zhang H, et al. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature*. 2006;444(7121):953-956. doi:10.1038/nature05363

191. Connelly S, Manley JL. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev*. 1988;2:440-452.

192. Proudfoot N. How RNA polymerase II terminates transcription in higher eukaryotes. *Trends Biochem Sci*. 1989;41:105-110.

193. Eaton JD, Davidson L, Bauer DLV, Natsume T, Kanemaki MT, West S. Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity. *Genes Dev*. 2018;32(2):127-139. doi:10.1101/gad.308528.117

194. Kim M, Krogan NJ, Vasiljeva L, et al. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature*. 2004;433:517-661. www.nature.com/nature

195. Gregersen LH, Mitter R, Ugalde AP, et al. SCAF4 and SCAF8, mRNA Anti-Terminator Proteins. *Cell*. 2019;177(7):1797-1813.e18. doi:10.1016/j.cell.2019.04.038

196. Proudfoot NJ, Furger A, Dye MJ. Integrating mRNA Processing with Transcription. *Cell*. 2002;108:501-512.

197. Neugebauer KM. On the importance of being co-transcriptional. *J Cell Sci*. 2002;115(20):3865-3871. doi:10.1242/jcs.00073

198. Marasco LE, Kornblihtt AR. The physiology of alternative splicing. *Nat Rev Mol Cell Biol*. 2023;24(4):242-254. doi:10.1038/s41580-022-00545-z

199. Matera AG, Wang Z. A day in the life of the spliceosome. *Nat Rev Mol Cell Biol*. 2014;15(2):108-121. doi:10.1038/nrm3742

200. Black CS, Whelan TA, Garside EL, Macmillan AM, Fast NM, Rader SD. Spliceosome assembly and regulation: insights from analysis of highly reduced spliceosomes. *RNA*. 2023;(29):531-550. doi:10.1261/rna

201. Wu Q, Krainer AR. *MINIREVIEW AT-AC Pre-MRNA Splicing Mechanisms and Conservation of Minor Introns in Voltage-Gated Ion Channel Genes*. Vol 19.; 1999.

202. Breathnach R, Benoist C, O'hare K, Gannon F, Chambon P. Ovalbumin gene: Evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries*. *PNAS*. 1978;75(10):4853-4857. https://www.pnas.org

203. Jackson IJ. A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res*. 1991;19(14):3795-3798.

204. Aebi M, Hornig H. 5' Cleavage Site in Eukaryotic Pre-mRNA Splicing Is Determined by the Overall 5' Splice Region, Not by the Conserved 5' GU. *Cell*. 1987;50:237-246.

205. Görnemann J, Kotovic KM, Hujer K, Neugebauer KM. Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. *Mol Cell*. 2005;19(1):53-63. doi:10.1016/j.molcel.2005.05.007

206. Nesic D, Cheng J, Maquat LE. Sequences within the Last Intron Function in RNA 3'-End Formation in Cultured Cells. *Mol Cell Biol*. 1993;13(6):3359-3369.

207. Lerner M, Boyle J, Mount S, Wolin S, Steitz J. Are snRNPs involved in splicing? *Nature*. 1980;283:220-224.

208. Lund E, Bostock C, Robertson M, Christie S, Mitchen JL, Dahlberg' JE. Ul Small Nuclear RNA Genes Are Located on Human Chromosome 1 and Are Expressed in Mouse-Human Hybrid Cells. *Mol Cell Biol*. 1983;3(12):2211-2220.

209. Lindgren V, Ares Jr M, Weiner A, Francke U. Human genes for U2 small nuclear RNA map to a major adenovirus 12 modification site on chromosome 17. *Nature*. 1985;314.

210. Van Arsdell SW, Weiner AM. Human Genes for U2 Small Nuclear RNA Are Tandemly Repeated. *Mol Cell Biol*. 1984;4(3):492-499.

211. Saiditt-Georgieff M, Harpold M, Chen-Kiang S, Darnell JE. *The Addition of 5' Cap Structures Occurs Early in HnRNA Synthesis and Prematurely Terminated Molecules Are Capped*. Vol 19.; 1980.

212. Hamm J, Mattaj W. *Monomethylated Cap Structures Facilitate RNA Export from the Nucleus*. Vol 63.; 1990.

213. Izaurralde E, Lewis J, Gamberl C, Jarmolowski A, McGuigan C, Mattaj I. A cap-binding protein complex mediating U snRNA export. *Nature*. 1995;376:709-712.

214. Ohno M, Segref A, Bachi A, Wilm M, Mattaj IW. PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. *Cell*. 2000;101(2):187-198. doi:10.1016/S0092-8674(00)80829-6

215. Raker VA, Plessel G, Luhrmann1 R. *The SnRNP Core Assembly Pathway: Identification of Stable Core Protein Heteromeric Complexes and an SnRNP Subcore Particle in Vitro*. Vol 15.; 1996. https://www.embopress.org

216. Rundle HD, Nagel L, Boughman JW, et al. *Essential Role for the SMN Complex in the Specificity of SnRNP Assembly*. Vol 287. Columbia Univ. Press; 2000. https://www.science.org

217. Meister G, Eggert C, Fisher U. SMN-mediated assembly of RNPs- a complex story. *Trends Cell Biol*. 2002;12(10):472-478.

218. Fischer U, Liu Q, Dreyfuss G. *The SMN-SIP1 Complex Has an Essential Role in Spliceosomal SnRNP Biogenesis*. Vol 90.; 1997.

219. Wlattaj RW. Cap Trimethylation of U snRNA Is Cytoplasmi and Dependent on U snRNP Protein Binding. *Cell*. 1986;46:905-911.

220. Cory S, Adams J. Modified 5'- Termini in small nuclear RNAs of mouse myeloma cells. *Mol Biol Rep*. 1975;2:287-294.

221. Lardelli RM, Schaffer AE, Eggens VRC, et al. Biallelic mutations in the 3' exonuclease TOE1 cause pontocerebellar hypoplasia and uncover a role in snRNA processing. *Nat Genet*. 2017;49(3):457-464. doi:10.1038/ng.3762

222. Ma T, Xiong ES, Lardelli RM, Lykke-Andersen J. Sm complex assembly and 5' cap trimethylation promote selective processing of snRNAs by the 3' exonuclease TOE1. *Proc Natl Acad Sci U S A*. 2024;121(3). doi:10.1073/pnas.2315259121

223. Palacios I, Hetzer M, Adam SA, Mattaj IW. Nuclear import of U snRNPs requires importin β. *EMBO J*. 1997;16(22):6783-6792. https://www.embopress.org

224. Shaw DJ, Eggleton P, Young PJ. Joining the dots: Production, processing and targeting of U snRNP to nuclear bodies. *Biochim Biophys Acta Mol Cell Res*. 2008;1783(11):2137-2144. doi:10.1016/j.bbamcr.2008.07.025

225. Broome HJ, Hebert MD. In vitro RNase and nucleic acid binding activities implicate coilin in U snRNA processing. *PLoS One*. 2012;7(4). doi:10.1371/journal.pone.0036300

226. Kass S, Kaximierz :, Steitz JA, Sollner-Webb' B. The U3 Small Nucleolar Ribonucleuprotein Functions in the First Step of Preribosomal RNA Processing. *Cell*. 1990;60:897-908.

227. Dominski Z, Marzluff WF. Formation of the 3' end of histone mRNA. *Gene*. 1999;239:1-14. www.elsevier.com/locate/gene

228. Mowry KL, Steitz JA. Identification of the Human U7 snRNP as One of Several Factors Involved in the 3' End Maturation of Histone Premessenger RNA's. *Science (1979)*. 1987;238:1682-1687. https://www.science.org

229. Strub K, Birnstiel ML. Genetic complementation in the Xenopus oocyte: co-expression of sea urchin histone and U7 RNAs restores 3' processing of H3 pre-mRNA in the oocyte. *EMBO J*. 1986;5(7):1675-1682. doi:10.1002/j.1460-2075.1986.tb04411.x

230. Woolford JL, Baserga SJ. Ribosome biogenesis in the yeast Saccharomyces cerevisiae. *Genetics*. 2013;195(3):643-681. doi:10.1534/genetics.113.153197

231. Barandun J, Chaker-Margot M, Hunziker M, Molloy KR, Chait BT, Klinge S. The complete structure of the small-subunit processome. *Nat Struct Mol Biol*. 2017;24(11):944-953. doi:10.1038/nsmb.3472

232. Pérez-Fernández J, Román Á, De Las Rivas J, Bustelo XR, Dosil M. The 90S Preribosome Is a Multimodular Structure That Is Assembled through a

Hierarchical Mechanism. *Mol Cell Biol*. 2007;27(15):5414-5429. doi:10.1128/mcb.00380-07

233. Osheim YN, French SL, Keck KM, et al. Pre-18S ribosomal RNA is structurally compacted into the SSU processome prior to being cleaved from nascent transcripts in Saccharomyces cerevisiae. *Mol Cell*. 2004;16(6):943-954. doi:10.1016/j.molcel.2004.11.031
234. Fatica A, Tollervey D, Dlakić M. PIN domain of Nob1p is required for D-site cleavage in 20S pre-rRNA. *RNA*. 2004;10(11):1698-1701. doi:10.1261/rna.7123504
235. Mougey EB, O'reilly M, Osheim Y, Miller OL, Beyer A, Sollner-Webb B. The terminal balls characteristic of eukaryotic rRNA transcription units in chromatin spreads are rRNA processing complexes. *Genes Dev*. 1993;7:1609-1619.
236. Culver GM. Sno-Capped: 5' Ends of Preribosomal RNAs Are Decorated with a U3 SnoRNP. *Chem Biol*. 2002;9:777-780.
237. Kiss-Lá Z, Henry Y, Kiss TS. Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *EMBO J*. 1998;17(3):797-807. https://www.embopress.org
238. Kiss-La ´szlo ´ Z, Henry Y, Bachellerie JP, Caizergues-Ferrer M, Kiss T. Site-Specific Ribose Methylation of Preribosomal RNA: A Novel Function for Small Nucleolar RNAs. *Cell*. 1996;85:1077-1088.
239. Ganot P, Caizergues-Ferrer M, Kiss^ T. The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and

ubiquitous sequence elements essential for RNA accumulation. *Genes Dev*. 1997;11:941-956.

240. Balakin AG, Smith L. The RNA World of the Nucleolus: Two Major Families of Small RNAs Defined by Different Box Elements with Related Functions. *Cell*. 1996;86:823-834.

241. Samarsky DA, Fournier MJ, Singer RH, Bertrand E. The snoRNA box C/D motif directs nucleolar targeting and also couples snoRNA synthesis and localization. *EMBO J*. 1998;17(13):3747-3757. https://www.embopress.org

242. Ois Dragon F, Gallagher JEG, Compagnone-Post PA, et al. A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. *Nature*. 2002;417:967-970. www.nature.com/nature

243. Dutca LM, Gallagher JEG, Baserga SJ. The initial U3 snoRNA:pre-rRNA base pairing interaction required for pre-18S rRNA folding revealed by in vivo chemical probing. *Nucleic Acids Res*. 2011;39(12):5164-5180. doi:10.1093/nar/gkr044

244. Marmier-Gourrier N, Cléry A, Schlotter F, Senty-Ségault V, Branlant C. A second base pair interaction between U3 small nucleolar RNA and the 5'-ETS region is required for early cleavage of the yeast pre-ribosomal RNA. *Nucleic Acids Res*. 2011;39(22):9731-9745. doi:10.1093/nar/gkr675

245. Turowski TW, Tollervey D. Cotranscriptional events in eukaryotic ribosome synthesis. *WIREs RNA*. 2015;6:129-139. doi:10.1002/wrna.1263

246. Kornprobst M, Turk M, Kellner N, et al. Architecture of the 90S Pre-ribosome: A Structural View on the Birth of the Eukaryotic Ribosome. *Cell*. 2016;166(2):380-393. doi:10.1016/j.cell.2016.06.014

247. Schafer T, StrauB D, Petfalski E, Tollervey D, Hurt E. The path from nucleolar 90S to cytoplasmic 40S pre-ribosomes. *EMBO*. 2003;22(6):1370-1380. https://www.embopress.org

248. Lamanna AC, Karbstein K. Nob1 binds the single-stranded cleavage site D at the 3-end of 18S rRNA with its PIN domain. *PNAS*. 2009;106(34). www.pnas.orgcgidoi10.1073pnas.0905403106

249. Petfalski E, Dandekar T, Henry Y, Tollervey D. Processing of the Precursors to Small Nucleolar RNAs and rRNAs Requires Common Components. *Mol Cell Biol*. 1998;18(3):1181-1189.

250. Allmang C, Mitchell P, Petfalski E, Tollervey D. Degradation of ribosomal RNA precursors by the exosome. *Nucleic Acids Res*. 2000;28(8):1684-1691.

251. De Koning L, Corpet A, Haber JE, Almouzni G. Histone chaperones: An escort network regulating histone traffic. *Nat Struct Mol Biol*. 2007;14(11):997-1007. doi:10.1038/nsmb1318

252. Groth A, Rocha W, Verreault A, Almouzni G. Chromatin Challenges during DNA Replication and Repair. *Cell*. 2007;128(4):721-733. doi:10.1016/j.cell.2007.01.030

253. Marzluff WF. Metazoan replication-dependent histone mRNAs: A distinct set of RNA polymerase II transcripts. *Curr Opin Cell Biol*. 2005;17(3):274-280. doi:10.1016/j.ceb.2005.04.010

254. Dominski Z, Marzluff WF. Formation of the 3' end of histone mRNA: Getting closer to the end. *Gene*. 2007;396(2):373-390. doi:10.1016/j.gene.2007.04.021

255. Wang ZF, Whitfield ML, Iii TCI, Dominski Z, Marzluff WF. The protein that binds the 3' end of histone mRNA: a novel RNA-binding protein required for histone pre-mRNA processing. *Genes Dev*. 1996;10:3028-3040.

256. Melin L, Soldati D, Mital R, Streit A, Schümperli D. Biochemical demonstration of complex formation of histone pre-mRNA with U7 small nuclear ribonucleoprotein and hairpin binding factors. *EMBO J*. 1992;11(2):691-697. doi:10.1002/j.1460-2075.1992.tb05101.x

257. Dominski Z, Erkmann JA, Yang X, Sànchez R, Marzluff WF. A novel zinc finger protein is associated with U7 snRNP and interacts with the stem-loop binding protein in the histone pre-mRNP to stimulate 3'-end processing. *Genes Dev*. 2002;16(1):58-71. doi:10.1101/gad.932302

258. Dominski Z, Zheng LX, Sanchez R, Marzluff WF. *Stem-Loop Binding Protein Facilitates 3-End Formation by Stabilizing U7 SnRNP Binding to Histone Pre-MRNA*. Vol 19.; 1999.

259. Skrajna A, Yang XC, Bucholc K, et al. U7 snRNP is recruited to histone pre-mRNA in a FLASH-dependent manner by two separate regions of the stem-loop binding protein. *RNA*. 2017;23:938-951. doi:10.1261/rna

260. Pillai RS, Grimmler M, Meister G, et al. Unique Sm core structure of U7 snRNPs: Assembly by a specialized SMN complex and the role of a new component, Lsm11, in histone RNA processing. *Genes Dev*. 2003;17(18):2321-2333. doi:10.1101/gad.274403

261. Kamback C, Walke S, Young R, et al. Crystal Structures of Two Sm Protein Complexes and Their Implications for the Assembly of the Spliceosomal snRNPs. *Cell*. 1999;96:375-387.

262. Azzouz TN, Gruber A, Schümperli D. U7 snRNP-specific Lsm11 protein: Dual binding contacts with the 100 kDa zinc finger processing factor (ZFP100) and a ZFP100-independent function in histone RNA 3' end processing. *Nucleic Acids Res*. 2005;33(7):2106-2117. doi:10.1093/nar/gki516

263. Yang X cui, Burch BD, Yan Y, Marzluff WF, Dominski Z. FLASH, a Proapoptotic Protein Involved in Activation of Caspase-8, Is Essential for 3' End Processing of Histone Pre-mRNAs. *Mol Cell*. 2009;36(2):267-278. doi:10.1016/j.molcel.2009.08.016

264. Hernandez N. Small Nuclear RNA Genes: A Model System to Study Fundamental Mechanisms of Transcription. *Journal of Biological Chemistry*. 2001;276(29):26733-26736. doi:10.1074/jbc.R100032200

265. Gridasova AA, Henry RW. The p53 Tumor Suppressor Protein Represses Human snRNA Gene Transcription by RNA Polymerases II and III Independently of Sequence-Specific DNA Binding. *Mol Cell Biol*. 2005;25(8):3247-3260. doi:10.1128/mcb.25.8.3247-3260.2005

266. Egloff S, O'Reilly D, Murphy S. Expression of human snRNA genes from beginning to end. *Biochem Soc Trans*. 2008;36(4):590-594. doi:10.1042/BST0360590

267. Neuman De Vegvar HE, Lund E, Dahlberg JE. 3' End Formation of U1 snRNA Precursors Is Coupled to Transcription from snRNA Promoters. *Cell*. 1986;47:259-266.

268. Hernandez N, Weiner AM. Formation of the 3' End of U1 snRNA Requires Compatible snRNA Promoter Elements. *Cell*. 1986;47:249-258.

269. Yoon JB, Murphy S, Bai L, Wang Z, Roeder RG. Proximal Sequence Element-Binding Transcription Factor (PTF) Is a Multisubunit Complex Required for Transcription of both RNA Polymerase II- and RNA Polymerase III-Dependent Small Nuclear RNA Genes. *Mol Cell Biol*. 1995;15(4):2019-2027.

270. Sadowski CL, Henry RW, Lobo SM, Hernandez N. Targeting TBP to a non-TATA box cis-re ulatory element: a TBP-containing complex activates transcription trom snRNA promoters through the PSE. *Genes Dev*. 1993;7:1535-1548.

271. Hirsch HA, Gu L, Henry ARW. The Retinoblastoma Tumor Suppressor Protein Targets Distinct General Transcription Factors To Regulate RNA Polymerase III Gene Expression. *Mol Cell Biol*. 2000;20(24):9182-9191.

272. Kuhlman TC, Cho H, Reinberg D, Hernandez AN. *The General Transcription Factors IIA, IIB, IIF, and IIE Are Required for RNA Polymerase II Transcription from the Human U1 Small Nuclear RNA Promoter*. Vol 19.; 1999.

273. Zaborowska J, Taylor A, Roeder RG, Murphy S. A novel TBP-TAF complex on RNA polymerase II-transcribed snRNA genes. *Transcription*. 2012;3(2):92-104. doi:10.4161/trns.19783

274. Takahashi H, Parmely TJ, Sato S, et al. Human mediator subunit MED26 functions as a docking site for transcription elongation factors. *Cell*. 2011;146(1):92-104. doi:10.1016/j.cell.2011.06.005

275. Guiro J, Murphy S. Regulation of expression of human RNA polymerase II-Transcribed snRNA genes. *Open Biol*. 2017;7(6). doi:10.1098/rsob.170073

276. Margottin F, Dujardin G, Gerard M, Egly JM, Huet J, Sentenac A. Participation of the TATA Factor in Transcription of the Yeast U6 Gene by RNA Polymerase C. *Science (1979)*. 1991;251:424-426. https://www.science.org

277. Lobo SM, Hernandez N. *A 7 Bp Mutation Converts a Human RNA Polymerase II SnRNA Promoter into an RNA Polymerase Ill Promoter*. Vol 58.; 1989.

278. Waibel F, Filipowicz W. RNA-polymerase specificity of transcription of Arabidopsis U snRNA genes determined by promoter element spacing. *Nature*. 1990;346:199-202.

279. Hernandez G, Valafar F, Stumph WE. Insect small nuclear RNA gene promoters evolve rapidly yet retain conserved features involved in determining promoter activity and RNA polymerase specificity. *Nucleic Acids Res*. 2007;35(1):21-34. doi:10.1093/nar/gkl982

280. Čabart P, Murphy S. BRFU, a TFIIB-like Factor, Is Directly Recruited to the TATA-box of Polymerase III Small Nuclear RNA Gene Promoters through Its Interaction with TATA-binding Protein. *Journal of Biological Chemistry*. 2001;276(46):43056-43064. doi:10.1074/jbc.M108515200

281. Teichmann M, Wang Z, Roeder RG. A stable complex of a novel transcription factor IIB-related factor, human TFIIIB50, and associated proteins mediate selective transcription by RNA polymerase III of genes with upstream promoter elements. *PNAS*. 2000;97:14200-14205. www.pnas.org

282. Buratowski S, Zhou H, Fink GR. *Functional Domains of Transcription Factor TFIIB (RNA Polymerase Fl/Transcription Inltiatlon/TATA-Binding Protein)*. Vol 90.; 1993.

283. Waldschmidt R, Seifart KH. TFIIA is required for in vitro transcription of mammalian U6 genes by RNA polymerase III. *Journal of Biological Chemistry*. 1992;267(23):16359-16364. doi:10.1016/s0021-9258(18)42010-8

284. Gouge J, Satia K, Guthertz N, et al. Redox Signaling by the RNA Polymerase III TFIIB-Related Factor Brf2. *Cell*. 2015;163(6):1375-1387. doi:10.1016/j.cell.2015.11.005

285. Gouge J, Guthertz N, Kramm K, et al. Molecular mechanisms of Bdp1 in TFIIIB assembly and RNA polymerase III transcription initiation. *Nat Commun*. 2017;8(1). doi:10.1038/s41467-017-00126-1

286. Smith ER, Lin C, Garrett AS, et al. The Little Elongation Complex Regulates Small Nuclear RNA Transcription. *Mol Cell*. 2011;44(6):954-965. doi:10.1016/j.molcel.2011.12.008

287. Baird TD, Chih-Chien Cheng K, Chen YC, et al. ICE1 promotes the link between splicing and nonsense-mediated mRNA decay. *Elife*. 2018;7:e33178. doi:10.7554/eLife.33178.001

288. Shilatifard A, Lane W, Jackson KW, Conaway RC, Conaway JW. An RNA Polymerase II Elongation Factor Encoded by the Human ELL Gene. *Science (1979)*. 1996;271:1873-1876.

289. Gerber M, Ma J, Dean K, Eissenberg J, Shilatifard A. Drosophila ELL is associated with actively elongating RNA polymerase II on transcriptionally active sites in vivo. *EMBO*. 2001;20(21):6104-6114. https://www.embopress.org

290. Kong SE, S Banks CA, Shilatifard A, Weliky Conaway J, Conaway RC. ELL-associated factors 1 and 2 are positive regulators of RNA polymerase II elongation factor ELL. *PNAS*. 2005;102(29):10094-10098. www.pnas.orgcgidoi10.1073pnas.0503017102

291. Sobhian B, Laguette N, Yatim A, et al. HIV-1 Tat Assembles a Multifunctional Transcription Elongation Complex and Stably Associates with the 7SK snRNP. *Mol Cell*. 2010;38(3):439-451. doi:10.1016/j.molcel.2010.04.012

292. Che Z, Liu X, Dai Q, et al. Distinct roles of two SEC scaffold proteins, AFF1 and AFF4, in regulating RNA polymerase II transcription elongation. *J Mol Cell Biol*. 2023;15(8). doi:10.1093/jmcb/mjad049

293. Hu D, Smith ER, Garruss AS, et al. The little elongation complex functions at initiation and elongation phases of snRNA gene transcription. *Mol Cell*. 2013;51(4):493-505. doi:10.1016/j.molcel.2013.07.003

294. Takahashi H, Ranjan A, Chen S, et al. The role of Mediator and Little Elongation Complex in transcription termination. *Nat Commun*. 2020;11(1). doi:10.1038/s41467-020-14849-1

295. Margelevičius M, Laganeckas M, Venclovas Č. COMA server for protein distant homology search. *Bioinformatics*. 2010;26(15):1905-1906. doi:10.1093/bioinformatics/btq306

296. Esnault C, Ghavi-Helm Y, Brun S, et al. Mediator-Dependent Recruitment of TFIIH Modules in Preinitiation Complex. *Mol Cell*. 2008;31(3):337-346. doi:10.1016/j.molcel.2008.06.021

297. Takagi Y, Calero G, Komori H, et al. Head Module Control of Mediator Interactions. *Mol Cell*. 2006;23(3):355-364. doi:10.1016/j.molcel.2006.06.007

298. Cai G, Imasaki T, Yamada K, Cardelli F, Takagi Y, Asturias FJ. Mediator Head module structure and functional interactions. *Nat Struct Mol Biol*. 2010;17(3):273-279. doi:10.1038/nsmb.1757

299. Diebold ML, Koch M, Loeliger E, et al. The structure of an Iws1/Spt6 complex reveals an interaction domain conserved in TFIIS, Elongin A and Med26. *EMBO Journal*. 2010;29(23):3979-3991. doi:10.1038/emboj.2010.272

300. Takahashi H, Takigawa I, Watanabe M, et al. MED26 regulates the transcription of snRNA genes through the recruitment of little elongation complex. *Nat Commun*. 2015;6. doi:10.1038/ncomms6941

301. Erb MA, Scott TG, Li BE, et al. Transcription control by the ENL YEATS domain in acute leukaemia. *Nature*. 2017;543(7644):270-274. doi:10.1038/nature21688

302. Phatnani HP, Greenleaf AL. Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev*. 2006;20(21):2922-2936. doi:10.1101/gad.1477006

303. Medlin J, Scurry A, Taylor A, Zhang F, Peterlin BM, Murphy S. P-TEFb is not an essential elongation factor for the intronless human U2 snRNA and histone H2b genes. *EMBO Journal*. 2005;24(23):4154-4165. doi:10.1038/sj.emboj.7600876

304. Medlin J, Uguen P, Taylor A, Bentley D, Murphy S. The C-terminal domain of pol II and a DRB-sensitive kinase are required for 3' processing of U2 snRNA. *EMBO*. 2003;22(4):925-934.

305. Skaar JR, Richard DJ, Saraf A, et al. INTS3 controls the hSSB1-mediated DNA damage response. *Journal of Cell Biology*. 2009;187(1):25-32. doi:10.1083/jcb.200907026

306. Li Y, Bolderson E, Kumar R, et al. hSSB1 and hSSB2 form similar multiprotein complexes that participate in DNA damage response. *Journal of Biological Chemistry*. 2009;284(35):23525-23531. doi:10.1074/jbc.C109.039586

307. Huang J, Gong Z, Ghosal G, Chen J. SOSS Complexes Participate in the Maintenance of Genomic Stability. *Mol Cell*. 2009;35(3):384-393. doi:10.1016/j.molcel.2009.06.011

308. Zhang F, Ma T, Yu X. A core hSSB1-INTS complex participates in the DNA damage response. *J Cell Sci*. 2013;126(21):4850-4855. doi:10.1242/jcs.132514

309. Xu C, Li C, Chen J, et al. R-loop-dependent promoter-proximal termination ensures genome stability. *Nature*. 2023;621(7979):610-619. doi:10.1038/s41586-023-06515-5

310. Barbieri E, Trizzino M, Welsh SA, et al. Targeted Enhancer Activation by a Subunit of the Integrator Complex. *Mol Cell*. 2018;71(1):103-116.e7. doi:10.1016/j.molcel.2018.05.031

311. Sabath K, Stäubli ML, Marti S, Leitner A, Moes M, Jonas S. INTS10–INTS13–INTS14 form a functional module of Integrator that binds nucleic acids and the cleavage module. *Nat Commun*. 2020;11(1). doi:10.1038/s41467-020-17232-2

312. Mendoza-Figueroa MS, Tatomer DC, Wilusz JE. The Integrator Complex in Transcription and Development. *Trends Biochem Sci*. 2020;45(11):923-934. doi:10.1016/j.tibs.2020.07.004

313. Rienzo M, Casamassimi A. Integrator complex and Transcription: Recent Findings And Pathophysiology . *Biochim Biophys Acta*. 2016;1859(10):1259-1280.

314. Ardehali MB, Yao J, Adelman K, et al. Spt6 enhances the elongation rate of RNA polymerase II in vivo. *EMBO Journal*. 2009;28(8):1067-1077. doi:10.1038/emboj.2009.56

315. Kaplan CD, Laprade L, Winston F. *Transcription Elongation Factors Repress Transcription Initiation from Cryptic Sites*. Vol 5.; 1991. www.sciencemag.org/cgi/content/full/301/5636/1094/DC1VOL301SCIENCE www.sciencemag.org

316. Dronamraju R, Strahl BD. A feed forward circuit comprising Spt6, Ctk1 and PAF regulates Pol II CTD phosphorylation and transcription elongation. *Nucleic Acids Res*. 2014;42(2):870-881. doi:10.1093/nar/gkt1003

317. McCullough L, Connell Z, Petersen C, Formosa T. The abundant histone chaperones Spt6 and FACT collaborate to assemble, inspect, and maintain chromatin structure in saccharomyces cerevisiae. *Genetics*. 2015;201(3):1030-1045. doi:10.1534/genetics.115.180794

318. Endoh M, Zhu W, Hasegawa J, et al. Human Spt6 Stimulates Transcription Elongation by RNA Polymerase II In Vitro. *Mol Cell Biol*. 2004;24(8):3324-3336. doi:10.1128/mcb.24.8.3324-3336.2004

319. Nojima T, Tellier M, Foxwell J, et al. Deregulated Expression of Mammalian lncRNA through Loss of SPT6 Induces R-Loop Formation, Replication Stress, and Cellular Senescence. *Mol Cell*. 2018;72(6):970-984.e7. doi:10.1016/j.molcel.2018.10.011

320. Guiro J, Fagbemi M, Tellier M, et al. CAPTURE of the Human U2 snRNA Genes Expands the Repertoire of Associated Factors. *Biomolecules*. 2022;12(5). doi:10.3390/biom12050704

321. Donner AJ, Ebmeier CC, Taatjes DJ, Espinosa JM. CDK8 is a positive regulator of transcriptional elongation within the serum response network. *Nat Struct Mol Biol*. 2010;17(2):194-201. doi:10.1038/nsmb.1752

322. Düster R, Ji Y, Pan KT, Urlaub H, Geyer M. Functional characterization of the human Cdk10/Cyclin Q complex. *Open Biol*. 2022;12(3). doi:10.1098/rsob.210381

323. Bartkowiak B, Liu P, Phatnani HP, et al. CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes Dev*. 2010;24(20):2303-2316. doi:10.1101/gad.1968210

324. Blazek D, Kohoutek J, Bartholomeeusen K, et al. The cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. *Genes Dev*. 2011;25(20):2158-2172. doi:10.1101/gad.16962311

325. O'Reilly D, Kuznetsova O V., Laitem C, Zaborowska J, Dienstbier M, Murphy S. Human snRNA genes use polyadenylation factors to promote efficient transcription termination. *Nucleic Acids Res*. 2014;42(1):264-275. doi:10.1093/nar/gkt892

326. Davidson L, Rouvière JO, Sousa-Luís R, et al. DNA-directed termination of mammalian RNA polymerase II. *Genes Dev*. 2024;38(21-24):998-1019. doi:10.1101/gad.351978.124

327. Cuello P, Boyd DC, Dye MJ, Proudfoot NJ, Murphy S. *Transcription of the Human U2 SnRNA Genes Continues beyond the 3 Box in Vivo*. Vol 18.; 1999. https://www.embopress.org

328. Nock A, Ascano JM, Barrero MJ, Malik S. Mediator-Regulated Transcription through the +1 Nucleosome. *Mol Cell*. 2012;48(6):837-848. doi:10.1016/j.molcel.2012.10.009

329. Pavelitz T, Bailey AD, Elco CP, Weiner AM. Human U2 snRNA Genes Exhibit a Persistently Open Transcriptional State and Promoter Disassembly at Metaphase. *Mol Cell Biol*. 2008;28(11):3573-3588. doi:10.1128/mcb.00087-08

330. Egloff S, Al-Rawaf H, O'Reilly D, Murphy S. Chromatin Structure Is Implicated in "Late" Elongation Checkpoints on the U2 snRNA and β-Actin Genes. *Mol Cell Biol*. 2009;29(14):4002-4013. doi:10.1128/mcb.00189-09

331. Wallace JA, Felsenfeld G. We gather together: insulators and genome organization. *Curr Opin Genet Dev*. 2007;17(5):400-407. doi:10.1016/j.gde.2007.08.005

332. Fu Y, Sinha M, Peterson CL, Weng Z. The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet*. 2008;4(7). doi:10.1371/journal.pgen.1000138

333. Shukla S, Kavak E, Gregory M, et al. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature*. 2011;479(7371):74-79. doi:10.1038/nature10442

334. Wada T, Takagi T, Yamaguchi Y, et al. DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev*. 1998;12:343-356. www.genesdev.org

335. Laitem C, Zaborowska J, Tellier M, et al. CTCF regulates NELF, DSIF and P-TEFb recruitment during transcription. *Transcription*. 2015;6(5):79-90. doi:10.1080/21541264.2015.1095269

336. Kim M, Vasiljeva L, Rando OJ, Zhelkovsky A, Moore C, Buratowski S. Distinct Pathways for snoRNA and mRNA Termination. *Mol Cell*. 2006;24(5):723-734. doi:10.1016/j.molcel.2006.11.011

337. Arigo JT, Eyler DE, Carroll KL, Corden JL. Termination of Cryptic Unstable Transcripts Is Directed by Yeast RNA-Binding Proteins Nrd1 and Nab3. *Mol Cell*. 2006;23(6):841-851. doi:10.1016/j.molcel.2006.07.024

338. Vasiljeva L, Kim M, Mutschler H, Buratowski S, Meinhart A. The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nat Struct Mol Biol*. 2008;15(8):795-804. doi:10.1038/nsmb.1468

339. Schulz D, Schwalb B, Kiesel A, et al. Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell*. 2013;155(5):1075-1087. doi:10.1016/j.cell.2013.10.024

340. Vasiljeva L, Buratowski S. Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Mol Cell*. 2006;21(2):239-248. doi:10.1016/j.molcel.2005.11.028

341. Groh M, Albulescu LO, Cristini A, Gromak N. Senataxin: Genome Guardian at the Interface of Transcription and Neurodegeneration. *J Mol Biol*. 2017;429(21):3181-3195. doi:10.1016/j.jmb.2016.10.021

342. Skourti-Stathaki K, Proudfoot NJ, Gromak N. Human Senataxin Resolves RNA/DNA Hybrids Formed at Transcriptional Pause Sites to Promote Xrn2-Dependent Termination. *Mol Cell*. 2011;42(6):794-805. doi:10.1016/j.molcel.2011.04.026

343. McCracken S, Fong N, Yankulov K, et al. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature*. 1997;385(9):807-809.

344. Pettinati I, Brem J, Lee SY, McHugh PJ, Schofield CJ. The Chemical Biology of Human Metallo-β-Lactamase Fold Proteins. *Trends Biochem Sci*. 2016;41(4):338-355. doi:10.1016/j.tibs.2015.12.007

345. Zhao J, Hyman L, Moore C. *Formation of MRNA 3 Ends in Eukaryotes: Mechanism, Regulation, and Interrelationships with Other Steps in MRNA Synthesis*. Vol 63.; 1999. https://journals.asm.org/journal/mmbr

346. Colgan DF, Manley JL. Mechanism and regulation of mRNA polyadenylation. *Genes Dev*. 1997;11:2755-2766.

347. Manley J, Hirose Y. RNA polymerase II is an essential mRNA polyadenylation factor. *Nature*. 1998;395.

348. Dantonel JC, Murthy KGK, Manley JL, Tora L. Transcription factor TFIID recruits factor CPSF for formation of 3′ end of mRNA. *Nature*. 1997;389.

349. Westin G, Zabielski J, Hammarstrom K, Monstein HJ, Bark C, Pettersson U. Clustered genes for human U2 RNA (small nuclear RNA/tandem repeat/heterotriplex). *PNAS*. 1984;81:3811-3815. https://www.pnas.org

350. Bylund L, Kytölä S, Lui WO, Larsson C, Weber G. Analysis of the cytogenetic stability of the human embryonal kidney cell line 293 by cytogenetic and STR profiling approaches. *Cytogenet Genome Res*. 2004;106(1):28-32. doi:10.1159/000078556

351. Andrews S. A Quality Control Tool for High Throughput Sequence Data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

352. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841-842. doi:10.1093/bioinformatics/btq033

353. Bartkowiak B, Yan C, Greenleaf AL. Engineering an analog-sensitive CDK12 cell line using CRISPR/Cas. *Biochim Biophys Acta Gene Regul Mech*. 2015;1849(9):1179-1187. doi:10.1016/j.bbagrm.2015.07.010

354. Horiuchi D, Huskey NE, Kusdra L, et al. Chemical-genetic analysis of cyclin dependent kinase 2 function reveals an important role in cellular transformation by multiple oncogenic pathways. *Proc Natl Acad Sci U S A*. 2012;109(17). doi:10.1073/pnas.1111317109

355. Lopez MS, Kliegman JI, Shokat KM. The logic and design of analog-sensitive kinases and their small molecule inhibitors. In: *Methods in Enzymology*. Vol 548. Academic Press Inc.; 2014:189-213. doi:10.1016/B978-0-12-397918-6.00008-2

356. Tellier M, Zaborowska J, Neve J, et al. CDK9 and PP2A regulate RNA polymerase II transcription termination and coupled RNA maturation . *EMBO Rep*. 2022;23(10). doi:10.15252/embr.202154520

357. Chou J, Quigley DA, Robinson TM, Feng FY, Ashworth A. Transcription-associated cyclin-dependent kinases as targets and biomarkers for cancer therapy. *Cancer Discov*. 2020;10(3):351-370. doi:10.1158/2159-8290.CD-19-0528

358. Carvalho Da Rocha JG. *Regulation of Expression of RNA Polymerase II-Transcribed Human SnRNA Genes*. University of Oxford; 2019.

359. Eaton JD, Francis L, Davidson L, West S. A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes. *Genes Dev*. 2020;34(1-2):132-145. doi:10.1101/gad.332833.119

360. Cortazar MA, Sheridan RM, Erickson B, et al. Control of RNA Pol II Speed by PNUTS-PP1 and Spt5 Dephosphorylation Facilitates Termination by a "Sitting Duck Torpedo" Mechanism. *Mol Cell*. 2019;76(6):896-908.e4. doi:10.1016/j.molcel.2019.09.031

361. Washington K, Ammosova T, Beullens M, et al. Protein phosphatase-1 dephosphorylates the C-terminal domain of RNa polymerase-II. *Journal of Biological Chemistry*. 2002;277(43):40442-40448. doi:10.1074/jbc.M205687200

362. Cossa G, Parua PK, Eilers M, Fisher RP. Protein phosphatases in the RNAPII transcription cycle: Erasers, sculptors, gatekeepers, and potential drug targets. *Genes Dev*. 2021;35(9-10):658-676. doi:10.1101/GAD.348315.121

363. Buratowski S. Progression through the RNA Polymerase II CTD Cycle. *Mol Cell*. 2009;36(4):541-546. doi:10.1016/j.molcel.2009.10.019

364. Saunders A, Core LJ, Lis JT. Breaking barriers to transcription elongation. *Nat Rev Mol Cell Biol*. 2006;7(8):557-567. doi:10.1038/nrm1981

365. Meinhart A, Kamenski T, Hoeppner S, Baumli S, Cramer P. A structural perspective of CTD function. *Genes Dev*. 2005;19(12):1401-1415. doi:10.1101/gad.1318105

366. Mosley AL, Pattenden SG, Carey M, et al. Rtr1 Is a CTD Phosphatase that Regulates RNA Polymerase II during the Transition from Serine 5 to Serine 2 Phosphorylation. *Mol Cell*. 2009;34(2):168-178. doi:10.1016/j.molcel.2009.02.025

367. Torchet C, Jacq C, Quevillon-Cheruel S, Labesse G, Facca C, Faye G. Ssu72 is a phosphatase essential for transcription termination of snoRNAs and speci®c mRNAs in yeast. *EMBO*. Published online 2003. https://www.embopress.org

368. Rothaeusler K, Baumgarth N. Assessment of cell proliferation by 5-bromodeoxyuridine (BrdU) labeling for multicolor flow cytometry. *Current protocols in cytometry / editorial board, J Paul Robinson, managing editor . [et al]*. 2007;Chapter 7. doi:10.1002/0471142956.cy0731s40

369. Oraki Kohshour M, Najafi L, Heidari M, Ghaffari Sharaf M. Antiproliferative effect of H2O2 against human acute myelogenous leukemia KG1 cell line. *JAMS Journal of Acupuncture and Meridian Studies*. 2013;6(3):134-141. doi:10.1016/j.jams.2012.08.004

370. Picache ER, Hassanieh L, Broek D, Schönthal AH. Inhibition of tumor cell growth by triton X-100 through specific effects on cell-cycle-regulatory components. *J Biomed Sci*. 2004;11(1):95-103. doi:10.1159/000075292

371. Mayer A, Lidschreiber M, Siebert M, Leike K, Söding J, Cramer P. Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol*. 2010;17(10):1272-1278. doi:10.1038/nsmb.1903

372. Kim M, Suh H, Cho EJ, Buratowski S. Phosphorylation of the yeast Rpb1 C-terminal domain at serines 2,5, and 7. *Journal of Biological Chemistry*. 2009;284(39):26421-26426. doi:10.1074/jbc.M109.028993

373. Descostes N, Heidemann M, Spinelli L, et al. Tyrosine phosphorylation of RNA Polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. *Elife*. 2014;3:e02105. doi:10.7554/eLife.02105

374. Hsin JP, Li W, Hoque M, Tian B, Manley JL. RNAP II CTD tyrosine 1 performs diverse functions in vertebrate cells. *Elife*. 2014;3:e02112. doi:10.7554/eLife.02112

375. Schlackow M, Nojima T, Gomes T, Dhir A, Carmo-Fonseca M, Proudfoot NJ. Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. *Mol Cell*. 2017;65(1):25-38. doi:10.1016/j.molcel.2016.11.029

376. Davidson L, Muniz L, West S. 3' end formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II CTD are reciprocally coupled in human cells. *Genes Dev*. 2014;28(4):342-356. doi:10.1101/gad.231274.113

377. Gu B, Eick D, Bensaude O. CTD serine-2 plays a critical role in splicing and termination factor recruitment to RNA polymerase II in vivo. *Nucleic Acids Res*. 2013;41(3):1591-1603. doi:10.1093/nar/gks1327

378. Eifler TT, Shao W, Bartholomeeusen K, et al. Cyclin-Dependent Kinase 12 Increases 3' End Processing of Growth Factor-Induced c-FOS Transcripts. *Mol Cell Biol*. 2015;35(2):468-478. doi:10.1128/mcb.01157-14

379. Shim EY, Walker AK, Shi Y, Blackwell TK. CDK-9/cyclin T (P-TEFb) is required in two postinitiation pathways for transcription in the C. elegans embryo. *Genes Dev*. 2002;16(16):2135-2146. doi:10.1101/gad.999002

380. Uguen P, Murphy S. The 3’ ends of human pre-snRNAs are produced by RNA polymerase II CTD-dependent RNA processing. *EMBO*. 2003;22(17):4544-4554.

381. Core LJ, Waterfall JJ, Lis JT. Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. *Science (1979)*. 2008;322:1845-1848. doi:10.1261/rna.1246808

382. Nojima T, Gomes T, Grosso ARF, et al. Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell*. 2015;161(3):526-540. doi:10.1016/j.cell.2015.03.027

383. Parua PK, Booth GT, Sansó M, et al. A Cdk9-PP1 switch regulates the elongation-termination transition of RNA polymerase II. *Nature*. 2018;558(7710):460-464. doi:10.1038/s41586-018-0214-z

384. Anshabo AT, Milne R, Wang S, Albrecht H. CDK9: A Comprehensive Review of Its Biology, and Its Role as a Potential Target for Anti-Cancer Agents. *Front Oncol*. 2021;11. doi:10.3389/fonc.2021.678559

385. Sansó M, Levin RS, Lipp JJ, et al. P-TEFb regulation of transcription termination factor Xrn2 revealed by a chemical genetic screen for Cdk9 substrates. *Genes Dev*. 2016;30:117-131. doi:10.1101/gad.269589

386. Fianu I, Chen Y, Dienemann C, et al. Structural basis of Integrator-mediated transcription regulation. *Science (1979)*. 2021;374:883-887. https://www.science.org

387. Shah N, Maqbool MA, Yahia Y, et al. Tyrosine-1 of RNA Polymerase II CTD Controls Global Termination of Gene Transcription in Mammals. *Mol Cell*. 2018;69(1):48-61.e6. doi:10.1016/j.molcel.2017.12.009

388. Cho H, Kim TK, Mancebo H, Lane WS, Flores O, Reinberg D. A protein phosphatase functions to recycle RNA polymerase II. *Genes Dev*. 1999;13:1540-1552. www.genesdev.org

389. Archambault J, Chambers RS, Kobor MS, et al. An essential component of a C-terminal domain phosphatase that interacts with transcription factor IIF in Saccharomyces cerevisiae. *PNAS*. 1997;94:14300-14305. www.pnas.org.

390. Liu X, Zhang Y, Chen Y, et al. In Situ Capture of Chromatin Interactions by Biotinylated dCas9. *Cell*. 2017;170(5):1028-1043.e19. doi:10.1016/j.cell.2017.08.003

391. Kaplan CD, Holland MJ, Winston F. Interaction between transcription elongation factors and mRNA 3'-end formation at the Saccharomyces cerevisiae GAL10-GAL7 locus. *Journal of Biological Chemistry*. 2005;280(2):913-922. doi:10.1074/jbc.M411108200

392. DeGennaro CM, Alver BH, Marguerat S, et al. Spt6 Regulates Intragenic and Antisense Transcription, Nucleosome Positioning, and Histone

Modifications Genome-Wide in Fission Yeast. *Mol Cell Biol*. 2013;33(24):4779-4792. doi:10.1128/mcb.01068-13

393. Narain A, Bhandare P, Adhikari B, et al. Targeted protein degradation reveals a direct role of SPT6 in RNAPII elongation and termination. *Mol Cell*. 2021;81(15):3110-3127.e14. doi:10.1016/j.molcel.2021.06.016

394. Cihlarova Z, Kubovciak J, Sobol M, et al. BRAT1 links Integrator and defective RNA processing with neurodegeneration. *Nat Commun*. 2022;13(1). doi:10.1038/s41467-022-32763-6

395. Lin MH, Jensen MK, Elrod ND, et al. Cytoplasmic binding partners of the Integrator endonuclease INTS11 and its paralog CPSF73 are required for their nuclear function. *Mol Cell*. Published online July 2024. doi:10.1016/j.molcel.2024.06.017

396. Dokaneheifard S, Dos Santos HG, Valencia MG, Arigela H, Edupuganti RR, Shiekhattar R. Neuronal differentiation requires BRAT1 complex to remove REST from chromatin. *Proc Natl Acad Sci U S A*. 2024;121(23). doi:10.1073/pnas.2318740121

397. Yoon JB, Murphy S, Bai L, Wang Z, Roeder RG. Proximal Sequence Element-Binding Transcription Factor (PTF) Is a Multisubunit Complex Required for Transcription of both RNA Polymerase II- and RNA Polymerase III-Dependent Small Nuclear RNA Genes. *Mol Cell Biol*. 1995;15(4):2019-2027.

398. Smith ER, Lin C, Garrett AS, et al. The Little Elongation Complex Regulates Small Nuclear RNA Transcription. *Mol Cell*. 2011;44(6):954-965. doi:10.1016/j.molcel.2011.12.008

399. Hu D, Smith ER, Garruss AS, et al. The little elongation complex functions at initiation and elongation phases of snRNA gene transcription. *Mol Cell*. 2013;51(4):493-505. doi:10.1016/j.molcel.2013.07.003

400. Nabet B, Roberts JM, Buckley DL, et al. The dTAG system for immediate and target-specific protein degradation. *Nat Chem Biol*. 2018;14(5):431-441. doi:10.1038/s41589-018-0021-8

401. Nabet B, Ferguson FM, Seong BKA, et al. Rapid and direct control of target protein levels with VHL-recruiting dTAG molecules. *Nat Commun*. 2020;11(1). doi:10.1038/s41467-020-18377-w

402. Chen R, Yang Z, Zhou Q. Phosphorylated Positive Transcription Elongation Factor b (P-TEFb) Is Tagged for Inhibition through Association with 7SK snRNA. *Journal of Biological Chemistry*. 2004;279(6):4153-4160. doi:10.1074/jbc.M310044200

403. Chen J, Ezzeddine N, Waltenspiel B, et al. An RNAi screen identifies additional members of the Drosophila Integrator complex and a requirement for cyclin C/Cdk8 in snRNA 3'-end formation. *RNA*. 2012;18(12):2148-2156. doi:10.1261/rna.035725.112

404. Offley SR, Pfleiderer MM, Zucco A, et al. A combinatorial approach to uncover an additional Integrator subunit. *Cell Rep*. 2023;42(3). doi:10.1016/j.celrep.2023.112244

405. Albrecht TR, Shevtsov SP, Wu Y, et al. Integrator subunit 4 is a 'Symplekin-like' scaffold that associates with INTS9/11 to form the Integrator cleavage module. *Nucleic Acids Res*. 2018;46(8):4241-4255. doi:10.1093/nar/gky100

406. Fianu I, Ochmann M, Walshe JL, et al. Structural basis of Integrator-dependent RNA polymerase II termination. *Nature*. 2024;629(8010):219-227. doi:10.1038/s41586-024-07269-4

407. Dominski Z, Yang X cui, Purdy M, Wagner EJ, Marzluff WF. A CPSF-73 Homologue Is Required for Cell Cycle Progression but Not Cell Growth and Interacts with a Protein Having Features of CPSF-100. *Mol Cell Biol*. 2005;25(4):1489-1500. doi:10.1128/mcb.25.4.1489-1500.2005

408. Pfleiderer MM, Galej WP. Structure of the catalytic core of the Integrator complex. *Mol Cell*. 2021;81(6):1246-1259.e8. doi:10.1016/j.molcel.2021.01.005

409. Huang KL, Jee D, Stein CB, et al. Integrator Recruits Protein Phosphatase 2A to Prevent Pause Release and Facilitate Transcription Termination. *Mol Cell*. 2020;80(2):345-358.e9. doi:10.1016/j.molcel.2020.08.016

410. Gardini A, Baillat D, Cesaroni M, et al. Integrator regulates transcriptional initiation and pause release following activation. *Mol Cell*. 2014;56(1):128-139. doi:10.1016/j.molcel.2014.08.004

411. Elrod ND, Henriques T, Huang KL, et al. The Integrator Complex Attenuates Promoter-Proximal Transcription at Protein-Coding Genes. *Mol Cell*. 2019;76(5):738-752.e7. doi:10.1016/j.molcel.2019.10.034

412. Stadelmayer B, Micas G, Gamot A, et al. Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. *Nat Commun*. 2014;5. doi:10.1038/ncomms6531

413. Jia Y, Cheng Z, Bharath SR, et al. Crystal structure of the INTS3/INTS6 complex reveals the functional importance of INTS3 dimerization in DSB repair. *Cell Discov*. 2021;7(1). doi:10.1038/s41421-021-00283-0

414. Li J, Ma X, Banerjee S, et al. Structural basis for multifunctional roles of human Ints3 C-Terminal domain. *Journal of Biological Chemistry*. 2021;296. doi:10.1074/jbc.RA120.016393

415. Xu C, Li C, Chen J, et al. R-loop-dependent promoter-proximal termination ensures genome stability. *Nature*. 2023;621(7979):610-619. doi:10.1038/s41586-023-06515-5

416. Wagner EJ, Tong L, Adelman K. Integrator is a global promoter-proximal termination complex. *Mol Cell*. 2023;83(3):416-427. doi:10.1016/j.molcel.2022.11.012

417. McCracken S, Fong N, Yankulov K, et al. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature*. 1997;385:357-361.

418. Harlen KM, Churchman LS. The code and beyond: Transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat Rev Mol Cell Biol*. 2017;18(4):263-273. doi:10.1038/nrm.2017.10

419. Kim JB, Sharp PA. Positive Transcription Elongation Factor b Phosphorylates hSPT5 and RNA Polymerase II Carboxyl-terminal Domain Independently of Cyclin-dependent Kinase-activating Kinase. *Journal of Biological Chemistry*. 2001;276(15):12317-12323. doi:10.1074/jbc.M010908200

420. Bösken CA, Farnung L, Hintermair C, et al. The structure and substrate specificity of human Cdk12/Cyclin K. *Nat Commun*. 2014;5. doi:10.1038/ncomms4505

421. Liang K, Gao X, Gilmore JM, et al. Characterization of Human Cyclin-Dependent Kinase 12 (CDK12) and CDK13 Complexes in C-Terminal Domain Phosphorylation, Gene Transcription, and RNA Processing. *Mol Cell Biol*. 2015;35(6):928-938. doi:10.1128/mcb.01426-14

422. Yu M, Yang W, Ni T, et al. RNA polymerase II–associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. *Science (1979)*. 2015;350(6266):1383-1386. doi:10.1126/science.aab0892

423. Quereda V, Bayle S, Vena F, et al. Therapeutic Targeting of CDK12/CDK13 in Triple-Negative Breast Cancer. *Cancer Cell*. 2019;36(5):545-558.e7. doi:10.1016/j.ccell.2019.09.004

424. Fu TJ, Peng J, Lee G, Price DH, Flores O. *Cyclin K Functions as a CDK9 Regulatory Subunit and Participates in RNA Polymerase II Transcription**. Vol 274.; 1999. http://www.jbc.org

425. Yu DS, Cortez D. A role for cdk9-cyclin k in maintaining genome integrity. *Cell Cycle*. 2011;10(1):28-32. doi:10.4161/cc.10.1.14364