REGULATION OF EXPRESSION OF RNA POLYMERASE II-TRANSCRIBED HUMAN snRNA GENES



Joana Guiro Carvalho da Rocha

Oriel College

University of Oxford

A thesis submitted for the degree of Doctor of Philosophy at the University of Oxford, Michaelmas Term 2019

Table of Contents

Abstract **5**

Acknowledgements **6**

Abbreviations **7**

List of Figures **11**

List of Tables **13**

Chapter 1 Introduction **14**

1.1 Transcription of human RNA polymerase II-dependent snRNA genes *14*

1.1.1 Transcription initiation 17

1.1.2 Transcription elongation 20

1.1.3 snRNA 3'end processing 21

1.1.4 Transcription termination 23

1.1.5 Coupling of initiation and 3'end formation 24

1.1.6 Variant snRNA genes 25

1.2 Transcription of protein-coding genes *26*

1.2.1 Transcription initiation 26

1.2.2 The role of Mediator in expression of protein-coding genes 27

1.2.3 Integrator in protein coding genes 30

1.2.4 Transcription elongation 30

1.2.5 Splicing 31

1.2.6 RNA 3' end formation and termination 32

1.3 CDK roles in transcription *34*

1.3.1 CTD phosphorylation during transcription of human snRNA genes 35

1.3.2 CTD phosphorylation on protein coding genes 38

1.3.3 CDK analog-sensitive cell lines 39

1.4 Unbiased methods to Pulldown proteins *41*

1.4.1 LNA oligos 41

1.4.2 enChIP 41

1.4.3 CLASP 42

1.4.4 CAPTURE 43

1.4.5 Genomic Locus Proteomics (GLoPro) 43

1.5 Project aims and description *44*

Chapter 2 Materials and Methods **46**

2.1 HEK293 Cell Culture ... *46*
2.2 dCas9 cloning and guide RNA design ... *46*
2.3 Transfection of HEK293 cells ... *47*
2.4 SDS-PAGE gels ... *48*
2.5 ChIP ... *50*
2.6 qPCR ... *51*
2.7 enChIP ... *52*
2.8 Chromatin Extraction and MNase digestion (Net-Seq) ... *54*
2.9 CAPTURE qPCR ... *55*
2.10 CAPTURE Proteomics ... *56*
2.11 RNA extraction ... *58*
2.12 RNase protection ... *58*
2.13 Immunoprecipitation ... *61*

Chapter 3 dCas9 Pulldown Approaches ... 64
3.1 Mutating Cas9 in pX462 and assessing the efficiency and specificity of guides ... *64*
3.1.1 Insertion of the H840A mutation into pX462 ... 64
3.1.2 Design and assessment of the guide (g)RNAs targeting the U2 snRNA gene ... 65
3.2 Stable cell lines expressing dCas9-FLAG + guide RNA ... *67*
3.3 enChIP on the PSE of the U2 snRNA gene ... *69*
3.4 U2 snRNA gene DSE dCas9 pulldown from the chromatin fraction ... *71*
3.5 CAPTURE of Telomeres and the U2 snRNA gene DSE and PSE ... *73*
3.6 Conclusions ... *77*

Chapter 4 U2 snRNA gene CAPTURE strategy ... 78
4.1 U2 snRNA gene PSE and DSE FLAG CAPTURE ... *78*
4.2 U2 snRNA gene PSE and DSE Streptavidin-Biotin CAPTURE ... *80*
4.3 Conclusions ... *84*

Chapter 5 Roles of transcriptional CDKs in snRNA gene expression ... 85
5.1 The effect of inhibiting analog-sensitive (as) CDKs on expression of the U2 snRNA gene. ... *85*
5.2 The effect of inhibiting analog-sensitive (as) cyclin-dependent kinases (CDKs) on 3' box recognition. ... *87*
5.3 Investigating CDK12 and CDK13 redundancy ... *93*
5.3.1 The association of RPAP2 and Integrator with the U2 snRNA gene ... 93
5.3.2 Pol II ChIPs on the protein coding gene KPNB1 ... 94
5.4 Conclusions ... *98*

Chapter 6 Discussion **99**

6.1 CAPTURE identifies potential new candidates that regulate snRNA gene expression. *99*

6.1.1. Future Directions 102

6.2 The roles of CDKs in snRNA gene expression *102*

6.2.1. Future directions 105

Supplementary Material **107**

References **130**

Abstract

Human small non-coding RNAs (snRNAs) are required for expression of the vast majority of our protein-coding genes due to their important role in pre-mRNA processing events. Transcription initiation of snRNA genes is mediated by transcription factors like Oct-1, which recognize the DSE element, and by PTF, which recognizes the PSE and nucleates a pre-initiation complex (PIC) comprising general transcription factors TFIIA, B, C, E, F and H, TBP and some TBP-associated factors, and Mediator. RNA polymerase II (pol II) is then recruited with the carboxyl-terminal domain (CTD) of the large subunit in a hypophosphorylated form. In humans, the CTD comprises 52 repeats of the consensus heptapeptide, $Y_1S_2P_3T_4S_5P_6S_7$. After initiation, the pol II CTD is phosphorylated on S2, S5 and S7 as the polymerase progresses into elongation, which is mediated by the little elongation complex (LEC) and finally 3' end formation and termination are mediated by recognition of the 3' box RNA processing element by the Integrator complex. It has been shown that there is compulsory coupling between the promoter element PSE and the 3' box; transcription by pol II must initiate from a PSE-containing promoter for the 3' box to be recognised. However, the mechanism underlying this coupling is still unclear. In order to better understand how snRNA genes are regulated, I performed an unbiased dCas9-mediated pulldown approach to characterize the transcription machinery present on the DSE and PSE of the U2 snRNA gene. This analysis has identified some novel potential players in snRNA gene expression.

After initiation of transcription of protein-coding genes, phosphorylation of S5 of the pol II CTD by the cyclin-dependent kinase (CDK) 7 subunit of TFIIH facilitates promoter escape and 5' capping of the emerging transcript. CTD S2 phosphorylation by the CDK9 subunit of the super elongation complex (SEC) and CDK12 facilitates processive elongation, splicing and recognition of the polyadenylation (poly(A)) signal in the nascent RNA. It has been shown that the activity of CDK9 is required for recognition of the snRNA gene 3' box. However, the roles of other CDKs was unknown. Characterisation of the *in vivo* roles of the CDKs has now been facilitated by using CRISPR/Cas9 genome engineering to produce cell lines with analog-sensitive (as) versions of many of these CDKs. In order to determine the roles of several CDKs in snRNA gene transcription, I performed pol II chromatin immunoprecipitation (ChIP) assays, RNase protection assays and immunoprecipitation/mass spectrometry analysis after specific inhibition of analog-sensitive CDKs. I have confirmed the role of CDK9 in expression of the human U2 snRNA gene and uncovered redundant roles for CDK12 and CDK13 in recognition of the 3' box.

Acknowledgements

First I would like to thank Prof. Shona Murphy for the supervision during the past 4 years. Her support, expertise, input and feedback were essential for the design of the project, planning of the experiments and for developing critical judgment of my own data as well as my academic growth in being a scientist.

I'm very grateful for having co-workers that became actual friends. Thank you Justyna Zaborowska, Michael Tellier, Firdaus Isa, Isabella Maudlin and Gilbert Ansa from Prof. Murphy lab and also Agnese Cristini (which I will really miss when leaving Oxford!), Clelia Accalai, Chiara Beghe, Flavia Constantinescu and Helene Borrmann from Dr. Gromak lab not only for the lab meetings, advices on experiments and support throughout my studies but also for all lunch and coffee breaks and for making the Dunn School basement a very social, nice and friendly place to be.

I am also very grateful for all support from my family which greatly contributed to my achievements, in special to my mother Jacira, my father Enivaldo, my aunt Doroti and my grandmother Deolinda, and for their encouragement for pursuing this DPhil even though it meant being at a distance. Last but not least, I am very grateful for Nici for supporting me so much that probably became the only mathematician that can explain CRISPR and for making Oxford a truly happy place for me.

Abbreviations

as – analog-sensitive

BAP – Biotin acceptor peptide

CAPTURE – CRISPR affinity purification in situ of regulatory elements

CASPEX – dCas-APEX2 fusion

CDK - cyclin-dependent kinases

CHD8 – chromodomain-helicase-DNA binding protein 8

ChIP – Chromatin immunoprecipitation

ChEP - Chromatin enrichment for proteomics

CLASP - Cas9 locus-associated proteome

CPSF – Cleavage and polyadenylation specificity factor

CstF – Cleavage stimulatory factor

CTD – carboxyl terminal domain

dCas9 - dead Cas9

DMEM – Dulbecco's Modified Eagle's Medium

DSE – Distal sequence element

DSIF – DRB sensitivity-inducing factor

EGF – Epidermal growth factor

enChIP- Engineered DNA-binding molecule-mediated chromatin immunoprecipitation

ePICh – End-targeting PICh

GLoPro – Genomic Locus Proteomics

gRNA - guide RNA

HEK – Human embryonic kidney

IEG – Immediate early gene

IRF – Interferon regulatory factor 1

LEC – Little elongation complex

LNA – locked nucleic acids

NELF – Negative elongation factor

PAP – Poly-A polymerase

PBP – PSE-binding protein

PICh – Proteomics of isolated chromatin segments

PHD – Plant homeodomain

PIC – pre-inititation complex

pol – Polymerase

PSE – Proximal sequence element

P-TEFb – Positive elongation factor b

PTF – PSE-binding transcription factor

rRNA – ribosomal RNA

SEC – Super elongation complex

SEC-L2 - Super elongation complex like 2

SEC-L3 - Super elongation complex like 3

Ser2- Serine 2

Ser2P – Phospho-Ser2

Ser5- Serine 5

Ser5P – Phospho-Ser5

Ser7- Serine 7

Ser7P – Phospho-Ser7

SNAPc – snRNA activating protein complex

snoRNAs – small nucleolar RNAs

snRNA - small nuclear RNA

snRNP – small nuclear ribonuncleoprotein

snTAFc – snRNA TAF complex

SOSS – sensor of single stranded DNA

TAFs – TBP-associated factors

TBP – TATA-binding protein

tRNA – transfer RNA

Tyr1 – Tyrosine 1

Tyr1P – Phospho-Tyr1

U2AF – U2 auxiliary factor

Wt – Wild type

List of Figures

Figure 1.1. Expression of human snRNA genes.

Figure 1.2. Model for transcription termination of the U2 snRNA gene.

Figure 1.3. Pol II promoter release in protein coding genes.

Figure 1.4. Pol II CTD phosphorylation events in snRNA gene transcription.

Figure 1.5. Scheme of engineered analog sensitive kinases.

Figure 2.1. Insertion of the H840A mutation on pX462.

Figure 3.1. Scheme of the U2 snRNA gene and the gRNAs and qPCR primers used.

Figure 3.2. dCas9 ChIPs on U2 snRNA gene upon qRNA transient transfection.

Figure 3.3. Expression of dCas9 in stable cell lines and U2 snRNA gene targeting.

Figure 3.4. Silver staining and western blot of enChIP on the U2 snRNA gene PSE.

Figure 3.5. Western blot, silver staining and mass spectrometry of the dCas9 Pulldown from chromatin fraction.

Figure 3.6. CAPTURE on telomeres.

Figure 3.7. CAPTURE on the DSE and PSE of the U2 snRNA gene.

Figure 3.8. Characterisation of a stable cell lines expressing dCas9 and BirA.

Figure 3.9. Western blot of CAPTURE output of the U2 snRNA gene PSE.

Figure 4.1. U2 snRNA gene PSE FLAG CAPTURE mass spectrometry.

Figure 4.2 U2 snRNA gene DSE and PSE FLAG CAPTURE mass spectrometry.

Figure 4.3 Scheme of the Streptavidin-Biotin CAPTURE.

Figure 4.4. U2 snRNA gene PSE Streptavidin-biotin CAPTURE mass spectrometry.

Figure 4.5. U2 snRNA gene DSE and PSE Streptavidin-biotin CAPTURE mass spectrometry.

Figure 5.1. The effect of inhibiting analog-sensitive (as) cyclin-dependent kinases (CDKs) on expression of the U2 snRNA gene.

Figure 5.2. The effect of inhibiting CDK 12/13as cells on expression of the U2 snRNA gene.

Figure 5.3. Scheme of the RNase protection.

Figure 5.4. The effect of inhibiting analog-sensitive CDKs on pre-snRNA 3' end formation.

Figure 5.5. RNase protections on CDK 7, 8, 9, 10, 12, 13 and 12/13 as.

Figure 5.6. RNase protections after treatment of HEK293 with TZH531.

Figure 5.7 Time course of CDK12/13as inhibition.

Figure 5.8. RPAP2, Int 9 and Int 11 ChIPs on the U2 snRNA in 12/13as after drug inhibition.

Figure 5.9. Pol II ChIP on KNPB1 in CDK12/13and CDK13 as cells after drug treatment.

Figure 6.1. List of the top hits of the Int11 pulldown.

Figure S2.1 Sequence of the dCas9 mutated on the pX462.

Figure S3.1. dCas9 occupancy on the u2 snRNA gene, U1 PSE and off targets in Chr 6 and 7.

Figure S3.2. MNase digestion of chromatin.

Figure S4.1. catRAPID predicted proteins to interact with the U2 snRNA.

Figure S.5.1. Ratio of the RT/pre-snRNA transcripts on the HEK293 RNase protections.

Figure S.5.2. Ratio of the RT/pre-snRNA transcripts on the 7, 8, 9, 10, 12 and 13 as RNase protections.

Figure. S.5.3 Ratio of the RT/pre-snRNA transcripts on the 12/13 as RNase protections.

List of Tables

Table 1.1. Factors regulating expression of snRNA and protein-coding genes.

Table 1.2. Transcriptional CDK and their pol II CTD targets.

Table 2.1. Sequence of primers for dCas9 cloning and sequencing, U2 guides and U2 snRNA and KPNB1 qPCR primers.

Table 2.2. List of antibodies used either for ChIP or Western Blot.

Table S4.1. U2 snRNA PSE FLAG CAPTURE.

Table S4.2. U2 snRNA PSE Streptavidin-Biotin CAPTURE.

Table S4.3. Optimized U2 snRNA PSE Streptavidin-Biotin CAPTURE.

Table S4.4. U2 snRNA DSE Streptavidin-Biotin CAPTURE

Table S.5.1. Int11 IP Mass spectrometry list in 12/13 as cells DMSO and NM conditions.

Table S.5.2. Int11 IP Mass spectrometry list in HEK293 cells DMSO and NM conditions.

Table S.5.3. List of proteins that changed the phosphorylation status under inhibition of CDK12/13 in Int11 IP mass spectrometry list in 12/13 as cells NM conditions.

Chapter 1

Introduction

1.1 Transcription of human RNA polymerase II-dependent snRNA genes

RNA polymerase II (pol II) is responsible for the transcription of all 25,000 or so protein-coding genes in the human genome and many long non-coding RNAs. Abundant non-coding RNAs, including ribosomal (r)RNA, transfer (t)RNA, the 7SK RNA that regulates transcription elongation by pol II, and the spliceosomal U6 and U6atac small nuclear (sn)RNAs are transcribed by pol I or pol III (1-5). However, other smaller non-coding RNAs such as microRNAs, small nucleolar (sno)RNAs and the remaining snRNAs are transcribed by pol II (3, 6, 7). snRNAs are short, mainly uridylic acid-rich RNAs (less than 350 nts) that associate with proteins to form small nuclear ribonucleoprotein particles (snRNPs) (7).

Due to their roles in pre-mRNA splicing, the pol II-transcribed U1, U2, U4, U4atac, U5, U11 and U12 snRNAs are required for expression of intron-containing protein-coding genes. The major spliceosome includes the U1, U2, U4, U5 and U6 snRNPs and recognizes the canonical GT/AG splice sites flanking introns to assemble in a stepwise manner with the pre-mRNA. Some introns are instead flanked by AT/AC splice sites, which recruit the minor spliceosome containing instead U11, U12, U4atac, U5 and U6atac snRNPs (8). U7

snRNA is required for 3′ end formation of replication-activated histone mRNA (9) and the U3 sn(o)RNA is required for processing of rRNA (10).

Although the functions of snRNAs are well understood, the regulation of their expression is still not fully characterized. The human snRNA genes have a much simpler promoter structure than most protein-coding genes; this comprises a distal sequence element (DSE) which acts as an enhancer and an essential snRNA gene-specific proximal sequence element (PSE) which is the core promoter (3, 11, 12) (Figure 1.1). The transcripts are intronless and non-polyadenylated, and a 3′ box directs formation of the 3′ end of the pre-snRNA, which is further processed to produce the mature snRNA (11).

In vertebrates, the genes for the major snRNAs (U1, U2, U4 and U5) are well conserved with recognizable DSE, PSE and 3′ box sequences (3), and Drosophila snRNA genes also have PSE-like sequences (13). The vertebrate snRNA genes are often present in multiple copies. For example, there are four copies of the U1 snRNA gene/human haploid genome (14) and 15 copies of the U2 snRNA gene/human haploid genome) (15). In contrast, the genes for the minor snRNAs (U11, U12 and U4atac) are usually single-copy genes and are missing in many invertebrates including Cnidaria and Annelida (16).

An important feature of snRNA genes is the obligatory coupling between the 3′ box RNA processing element and the snRNA gene-type promoter; replacement of an snRNA gene promoter with the promoter of a protein-coding gene results in failure of RNA 3′ end formation (17, 18).

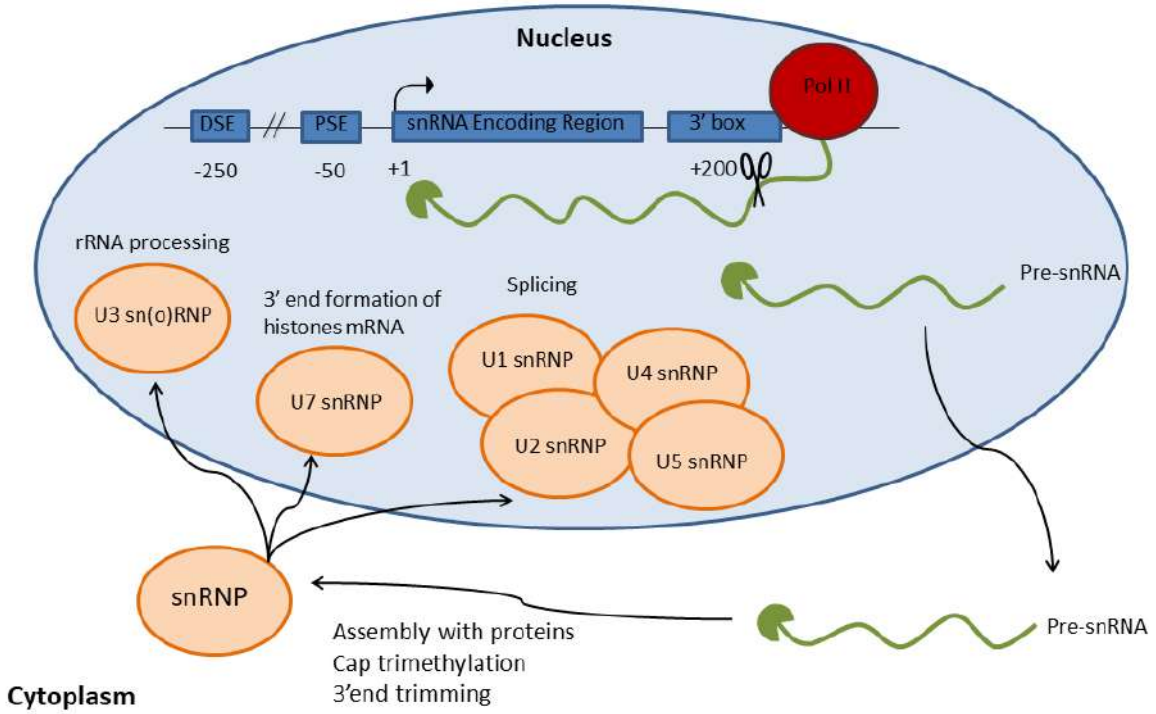


Figure 1.1. Expression of human pol II-transcribed snRNA genes.

Distal sequence element (DSE) and proximal sequence element (PSE) are the enhancer and promoter elements, respectively. The 3′ box is the snRNA gene-specific RNA processing element. An arrow represents the start site of transcription and the numbers below the line indicate the approximate position of the elements with respect to the transcription start site. DSE comprises the Oct1 binding site ATTTGCAT (19), three Sp1 binding sites (G/T)(G/A)GGCG(G/T)(G/A)(G/A)(G/T) (20) and a STAF binding site YY(A/T)CCC(A/G)N(A/C)AT(G/C)C(A/C)YYRCR (21). The PSE sequence consensus sequence is TCACCNTNA(G/C)NNNAA(A/T)(G/A)N [2]. The consensus sequence of the 3′ box is $GTTYN_{0-3}AARRYAGA$ (17). The snRNA transcript is represented in green with the cap in the 5′ end. The pre-snRNA is exported to the cytoplasm for maturation by 3′ end trimming, cap trimethylation and assembly with the snRNP proteins [6]. The functions of the various snRNPs after re-import into the nucleus are noted.

The pol III-transcribed genes for 7SK snRNA and U6 snRNA also have a DSE and PSE, but in addition they have a TATA box at −25 (3). Placing a TATA box downstream of the PSE in a pol II-transcribed snRNA gene converts transcription from pol II to pol III (3), emphasizing that the DSE and PSE can work as *cis*-acting elements for either polymerase.

1.1.1 Transcription initiation

Transcription factors Oct1, Sp1, NF1 and Staf bind to sequences in the DSE, which has the properties of a transcriptional enhancer (3, 22). Oct1 enhances snRNA gene expression by stabilizing binding of the PSE-binding protein/PSE-binding transcription factor/snRNA activating protein complex, PTF (also known as PBP and SNAPc) (19, 23, 24), to the PSE of both pol II- and pol III-transcribed snRNA genes through direct interaction (19, 25, 26). In turn, PTF helps to recruit the TATA-binding protein (TBP) and the TBP-associated factors (TAFs) that make up the snRNA TAF complex (snTAFc) to the pol II-transcribed snRNA genes (23, 27, 28). The snTAFc on the U2 snRNA gene, which comprises TAF5, TAF6, TAF8, TAF9, TAF11 and TAF13 (28), is a subset of the TAFs found in TFIID, the TBP/TAF-containing complex required for transcription of many protein-coding genes (29). However, the U1, U4, U5 and U11 snRNA genes appear to have a different TAF complex, which includes TAF7 (30). TAF7 may help to recruit the large multisubunit Mediator complex, which is required for efficient transcription of snRNA genes (30). TAF7 has also been implicated in regulation of promoter escape (see below). However, TAF7 appears to be absent from the U2 snRNA genes (28), suggesting that these genes employ a different mechanism of Mediator recruitment and promoter escape. The general protein-coding gene transcription factors TFIIA, TFIIB, TFIIE, TFIIF and TFIIH are also required for transcription of snRNA genes (31, 32). Table 1.1 indicates which factors are shared between snRNA genes and protein-coding genes and which are specific to each class of gene. As noted above, the DSE and PSE are interchangeable between the pol II- and pol III-transcribed snRNA genes (3). However, in the presence of a TATA box in the promoter, a pol III- and snRNA gene-specific TBP-containing complex is recruited (3).

There is evidence that a nucleosome between the DSE and PSE in the U1, U2, 7SK and U6 snRNA genes brings those two regions together, facilitating the interaction between Oct1 and PTF (33-36). In addition, PTF plays a role in depletion of histones from the transcription unit of the U2 gene (37). Thus, chromatin structure plays a role in the regulation of expression of human snRNA genes. Expression of the pol III-transcribed U6 snRNA gene is regulated by interaction of a protein involved in chromatin modification, the chromodomain-helicase-DNA binding protein 8 (CHD8), with Staf (38). As CHD8 has also been shown to interact with the elongating form of pol II (38), it may also regulate transcription of pol II-transcribed snRNA genes.

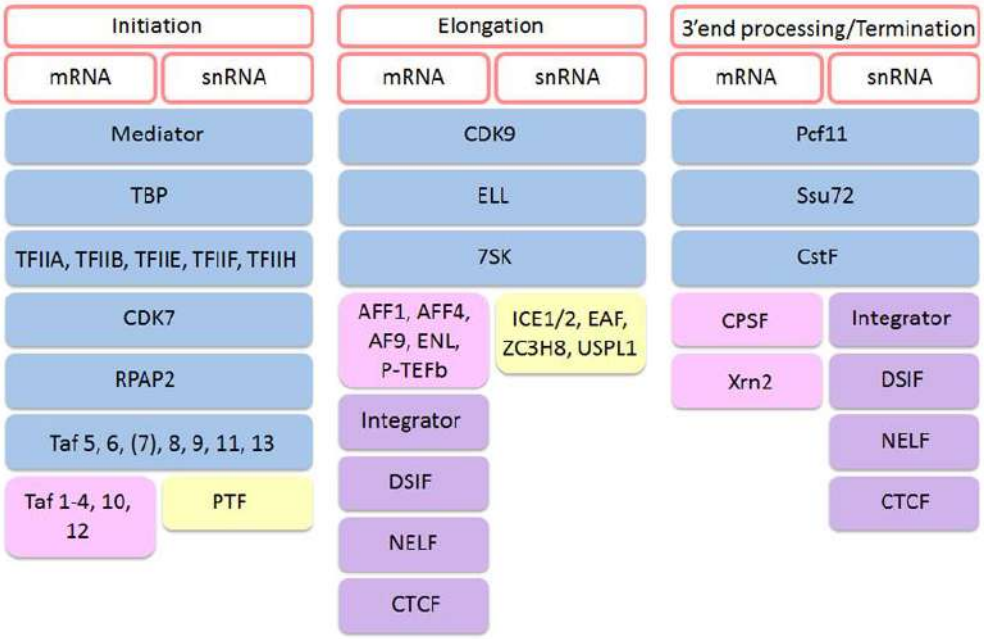
Association of the large 26-subunit Mediator complex with protein-coding genes has been well studied. This complex is recruited to the pre-initiation complex (PIC) and can bind to several transcription factors at the same time. Mediator has an important role as a binding platform for interaction between the transcription factors bound to sequences in protein-coding gene promoters and pol II (39).

The Mediator complex is made up of modules, with the head module comprising the subunits Med6, Med8, Med11, Med17, Med18, Med20 and Med22 (40). The head module is responsible for the interaction with the pol II CTD through Med17 (41). Med17 is a subunit on the 'nose' of the head module whereas Med18 and Med20 are located on the 'jaw'. A mutation on the alpha helix of Med18 is predicted to strengthen the interaction between the jaw and the rest of the module. The jaw is therefore thought to be the moveable element that regulates binding to the pol II CTD (40). This flexibility may also have a role in interactions between Mediator and pol II and also between Mediator and transcription factors such as TFIIH (42) and TBP (43).

The association of Mediator subunits with snRNA genes was only described for the first time recently. Med1, Med23 and Med26 are associated with a subset of snRNA genes (U1, U4 and U5) and in significantly higher levels than in the promoter region of some protein-coding genes (30). However, the chromatin immunoprecipitation (ChIP) experiments did not define the region of the snRNA genes associated with Mediator. Med26 is required for the expression of U1, U2, U4, U5 and U11 genes (30) and is essential for the recruitment of the little elongation complex (LEC) (30). Interestingly, Med26 also interacts with TAF7, which represses initiation and blocks LEC recruitment. In addition, TAF7 has protein sequence similarities to EAF, which is part of the LEC. Accordingly, it has been suggested that EAF could snatch Med26 from TAF7 to facilitate elongation (30). Med26 may therefore act as a molecular switch triggering elongation after promoter escape, at least for the snRNA genes where TAF7 is recruited.

Table 1.1. Factors regulating expression of snRNA and protein-coding genes.

In blue are the factors that have roles in the same event of transcription in both classes of genes. In pink and yellow are the factors unique to protein-coding genes and snRNA genes, respectively. Factors that regulate both genes but at different steps are shown in purple (23, 25, 27-32, 37, 39, 44-65).



1.1.2 Transcription elongation

The first identified elongation complex that regulates protein-coding genes is known as the super elongation complex (SEC) (52). Later, another elongation complex was identified that regulates snRNA genes, called the little elongation complex (LEC). LEC is an ELL-containing complex that specifically regulates elongation of transcription of pol II-transcribed snRNA genes and comprises ELL, ICE1, ICE2, EAF and ZC3H8 (51). It has been suggested that ICE1, ICE2, ELL and EAF are recruited to the promoter by the Mediator complex (30). ICE1 has been shown to be the scaffold for LEC formation and is essential for recruitment of the other components. ELL interacts with the ICE1 N-terminal domain, whereas ICE2 and ZC3H8 interact with the ICE1 C-terminal domain. In addition, ICE1 is required for pol II recruitment (51). ELL knockdown shifts the distribution of pol II on snRNA genes towards the 5' end, as expected of a factor that facilitates transcription elongation (66). It has been suggested that ELL is necessary to ensure a high level of transcription of snRNA genes (66). The function of subunits ICE2 and ZC3H8 is currently unknown.

It has been shown that the tumour suppressor p53 is recruited to snRNA gene promoters and prevents the recruitment of LEC by interacting with ELL and disrupting the association between ELL/EAF and ICE1 and therefore arresting pol II in the promoter and preventing transcription (67).

The LEC has been shown to co-localize with coilin (51), which is found in Cajal bodies, where snRNP maturation takes place (7). ICE1 recruits another protein, USPL1, to Cajal bodies. USPL1 also interacts with U1 and U2 snRNPs and its knockdown causes a

reduction in snRNA levels (53). However, the molecular mechanism underlying the effect of USPL1 knockdown is unclear.

More recently, it has been shown that the 7SK snRNP, comprising 7SK snRNA, MePCE and Larp7, plays a role in the recruitment of the LEC to snRNA genes. Knockdown of 7SK snRNP components disrupts LEC integrity, affects pol II recruitment and reduces expression of snRNA genes (68). This was a surprising finding as the 7SK snRNP is a negative regulator of the CTD Ser2 kinase P-TEFb complex (69).

1.1.3 snRNA 3'end processing

The Integrator complex was discovered more recently than Mediator. It is present only in metazoans and is a large complex of 14 subunits that is required for 3' end processing of pol II-transcribed pre-snRNAs in humans and flies (44, 47, 54).

Integrator is able to bind the pol II CTD phosphorylated on both Ser2 and Ser7 (70). However, it is still unclear which integrator subunits are involved in Ser2P and Ser7P recognition. Int11 is a paralogue of CPSF73, which is responsible for RNA cleavage directed by poly(A) sites and the RNA processing signal in the replication-activated histone pre-mRNAs (71). Depletion of Int11 causes accumulation of the pre-U1 and pre-U2 snRNAs (37, 47), which is in agreement with the expected catalytic role of this subunit.

So far, Integrator subunits have not been found to interact directly with chromatin. However, subunit Int12 has a PHD (plant homeodomain) domain, which is a common chromatin-binding motif that recognizes methylated histones (72). In fact, Int12 has been found to bind the genome and to be enriched in transcriptionally active regions. Knock

down of Int12 shows little effect in snRNA processing and it has been shown to regulate translation by regulating the expression of genes involved in protein synthesis pathways (73). Subunits Int3 and Int6 are also part of another complex involved in detecting DNA double strand breaks, called sensor of single stranded DNA (SOSS) (74-78). As snRNA genes are highly expressed and at least the U2 snRNA gene is depleted of canonical histones (55), it has been suggested that these Integrator subunits could help to prevent DNA damage that is facilitated by a constitutively open chromatin state (79).

Int5 is found associated with both the snRNA gene promoter region and the 3' box, whereas Int11 is associated mainly with the 3' box (80). Integrator may therefore be a modular complex that assembles in a stepwise manner, with some subunits playing a role early in transcription, while others are active in RNA 3' end processing.

It has been demonstrated that Int4 interacts with NELF-A and Int6 interacts with the Spt5 subunit of DSIF (56). In expression of snRNA genes, DSIF has been proposed to bind pol II prior to transcription elongation (56). Integrator could therefore be recruited both by pol II CTD phosphorylation and by association with DSIF. Recruitment of NELF through interaction with RNA and Integrator could contribute to specific 3' box recognition by preventing recruitment of the cleavage stimulatory factor (CstF) polyadenylation factor (56). These events would then allow Integrator to properly process the nascent snRNA transcript (56).

1.1.4 Transcription termination

The 3’ box is a signal for processing nascent snRNA transcripts and pol II continues transcribing for a few hundred base pairs past the 3’ box before transcription is terminated (50). As already mentioned, Integrator binds DSIF, NELF and pol II, recognizes the 3’ box signal on the RNA and processes the nascent transcript. Pre-snRNA 3’ end formation and pol II transcription termination on snRNA genes is intrinsically linked, as demonstrated by a recent study in which knockdown of the catalytic subunits Int9 and Int11 caused termination defects (37). NELF is also involved in termination. It accumulates downstream of the 3’ box of the U2 snRNA gene and its knockdown also causes a termination defect (55). A CTCF binding site is located close to the region of NELF association, and delimits a nucleosome-depleted region spanning the transcription unit (55). CTCF is best known as a transcription factor for protein-coding genes through its action as an insulator and organizer of chromatin structure (81). CTCF has been shown to position nucleosomes and to impede pol II elongation (60, 82). By acting in this way, CTCF could control the chromatin structure and elicit termination of transcription. U2 snRNA genes may therefore have a ‘late’ elongation checkpoint where CTCF-positioned nucleosomes slow down pol II, allowing NELF to terminate transcription (55).

In agreement with these findings, a more recent study implicates CTCF in DSIF, NELF and P-TEFb recruitment to the U2 snRNA gene (61). The findings suggest a model linking transcription termination to recognition of the 3’ box. In this model, NELF is required for termination of transcription, since its knockdown results in termination defects without affecting CTCF recruitment or RNA 3’ end formation. Additionally, P-TEFb phosphorylates the pol II CTD on Ser2, leading to the recruitment of Integrator subunits 9/11, which can

elicit RNA processing driven by the 3’ box (61). This model would explain how transcription termination and RNA processing are tightly linked during expression of snRNA genes (figure 1.2).

Intriguingly, mRNA 3’ end processing factors Pcf11, Ssu72 and Cstf64 also associate with snRNAs genes. However, Pcf11 and Ssu72 were shown to affect termination of transcription of snRNA genes rather than 3’ end processing (37).

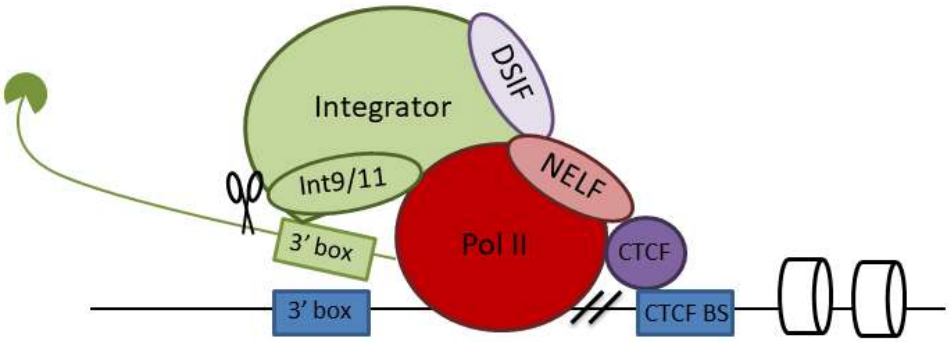


Figure 1.2. Model for transcription termination of the U2 snRNA gene.

The snRNA transcript is represented in green with the cap in the 5’ end and nucleosomes are represented by barrels. Pol II continues to transcribe after the 3’ box and Integrator processes the nascent RNA after 3’ box recognition. CTCF recognizes the CTCF binding site downstream of the 3’ box and controls nucleosome occupancy. Negative elongation factor (NELF) is recruited by DRB sensitivity-inducing factor (DSIF) and CTCF at the end of the transcription unit and causes transcription termination (61).

1.1.5 Coupling of initiation and 3’end formation

The 3’ box is only recognized if transcription is initiated from a pol II-dependent snRNA gene promoter (17, 18). If the PSE is substituted by a protein-coding gene promoter, the transcripts become longer as the 3’ box signal is ignored and pol II transcribes until a functional polyadenylation site is reached (83). The mechanism

underlying coupling of transcription initiation and RNA 3' end formation is still not understood but may involve the snRNA gene-specific promoter factor PTF or the snRNA genes-specific components of the LEC. Alternatively, recognition of the 3' box may only occur in the absence of SEC recruitment.

1.1.6 Variant snRNA genes

With the exception of teleosts and birds, all metazoans studied so far have pseudogenes derived from snRNA genes, with mammals possessing a large number (16). Some sequences in the human genome that were previously thought to be U1 snRNA pseudogenes have been shown to be true genes that have recognizable DSE, PSE and 3' box sequences and are actively transcribed (14). These genes have been termed variant (v)U1 genes (14, 84). The DSE, PSE and 3' box sequences of the vU1 snRNA genes are, however, sufficiently different from the U1 snRNA gene sequences to suggest that they are differentially regulated. Interestingly, the vU1 snRNAs appear to be most highly expressed in embryonic stem cells (14). U1b is a mouse U1 variant gene only expressed in early embryonic stages, although the U1 and U1b snRNAs differ by only seven base changes (85, 86). However, the promoter of U1b differs significantly from that of U1 (86). Although the function of U1b is still unclear, there are clear parallels between mouse U1b and the human vU1 snRNAs.

U5 snRNA variants in human and Drosophila have also been described (87, 88). The variants form different functional snRNP complexes from the canonical U5 snRNP and are thought to play roles in splicing (87). Interestingly, the U5 variants are differentially expressed during development and may promote tissue-specific splicing (88).

1.2 Transcription of protein-coding genes

1.2.1 Transcription initiation

As already mentioned, pol II is also responsible for the transcription of all protein-coding genes. Transcription starts with the formation of the PIC at the promoter and this comprises the general transcription factors TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH. Initially, the TBP subunit of TFIID binds the TATA box region of the core promoter region (although a TATA box is present in a minority of promoters) and unwinds the DNA for open complex formation (89, 90). TFIID also contains TBP-associated factors (TAFs) which are essential for TFIID assembly (91). TAFs also have a critical role in activator-responsive transcription (92-96)

Following TFIID binding, TFIIA binds to TBP and stabilizes the TBP-DNA interaction (97, 98). TFIIB then joins the PIC, to recruit pol II/TFIIF and help to select the transcription start site (99). In turn, TFIIF recruits TFIIE and TFIIH. TFIIE is required for the transition from initiation to elongation, whereas TFIIH is a complex with DNA-dependent ATPase activity, ATP-dependent helicase activity and CTD kinase activity (100). The TFIIH subunit CDK7 is responsible for phosphorylation of Ser5 of the pol II CTD is essential for transcription initiation (32).

After initiation of transcription of protein-coding genes, the capping machinery is recruited to the 5’ end of the nascent transcript and allosterically activated by Ser5P (101, 102).

1.2.2 The role of Mediator in expression of protein-coding genes

As mentioned above (Section 1.1.1), mediator is a large 26 subunit key transcription complex that acts a platform to bridge interaction between several transcription factors and pol II (39). Mediator subunit composition is variable and it has been shown that some mammalian Mediator complexes lack some of the subunits (103, 104), suggesting that different subunit combinations might be required depending of the differentiation of the cell or even for different classes of genes. The CDK8 module comprises CDK8, CCNC, Med 12 and Med13 and associates with the core Mediator in a reversible manner (105, 106). Additionally, CDK19, Med12L and Med13L, paralogues of CDK8, Med12 and Med13 respectively, can also associate with the core Mediator in a mutually exclusive way (107) and have been implicated in neuronal and developmental disorders (108, 109).

Mediator has been implicated in PIC assembly and stability as it can interact with several PIC factors simultaneously. In fact, Mediator recruits and regulates the activity of the PIC components TFIIA, TFIIB, TFIIE, TFIIH and TFIIF (42, 110-113) and has also been shown to interact cooperatively with TFIID (114-118).

During transcription initiation, pol II has to break the interactions with Mediator and the other PIC components for promoter release and the transition to productive elongation (Figure 1.3). It has been proposed that this release is facilitated by pol II CTD phosphorylation by CDK7 and that Mediator plays a role in this phosphorylation (119-122).

Additionally, Mediator has been shown to interact with and recruit the SEC. The same region of Med26 has the ability to bind either the SEC or TFIID, suggesting that

Med26 interaction with the SEC acts as a molecular switch allowing pol II to escape the promoter and enter productive elongation (117). Med26 is therefore required for productive elongation of both protein coding genes and snRNA genes (Section 1.1.1.).

A model has been proposed for Mediator-dependent regulation on pol II transcription (Figure 1.3) (39).

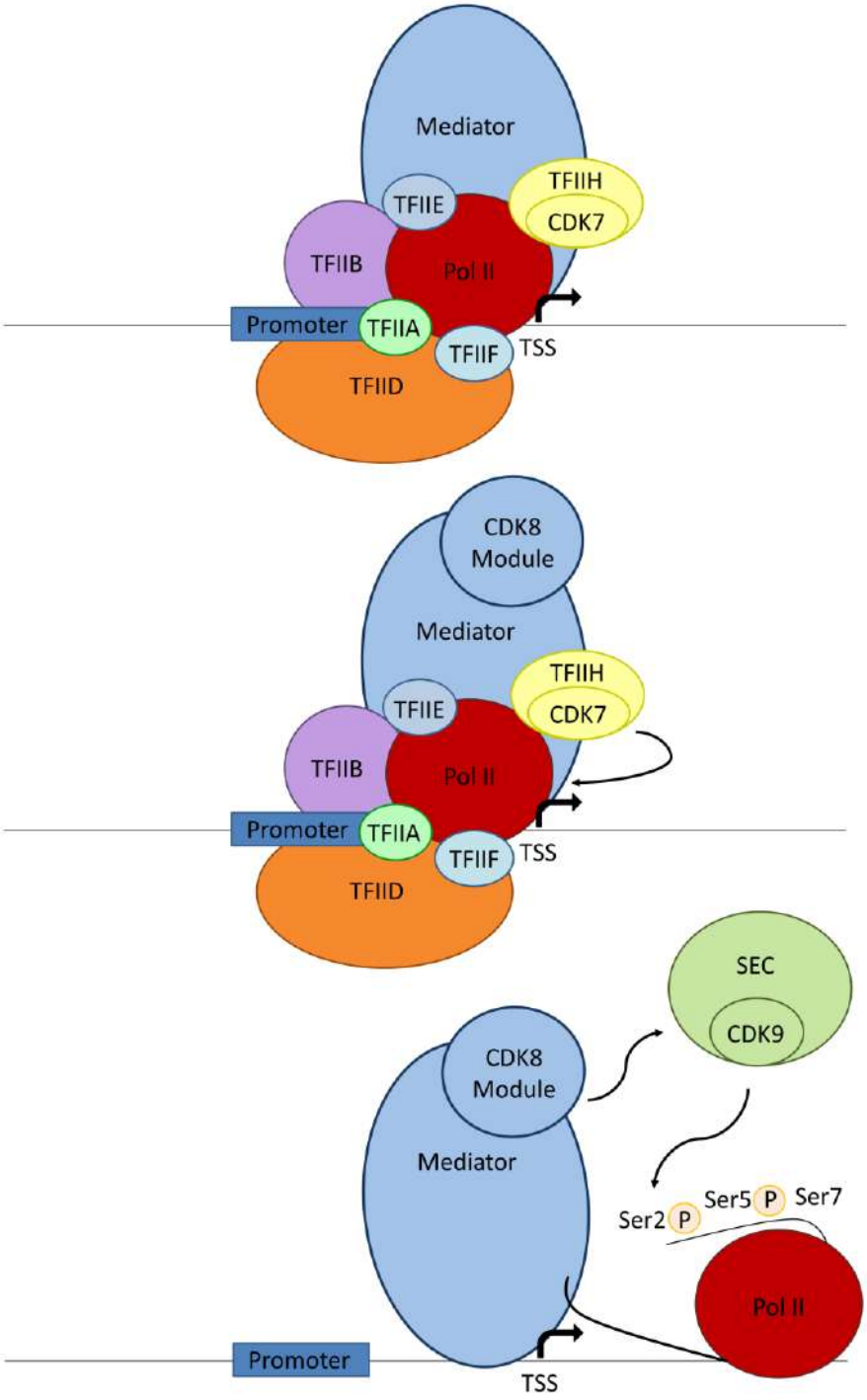


Figure 1.3. Pol II promoter release in protein coding genes.

The general transcription factors are recruited to the protein coding genes promoter forming the PIC together with Mediator. Pol II is then recruited and its Ser5 residue of the CTD is subsequently phosphorylated by TFIIH subunit CDK7 resulting in promoter release. SEC then is recruited by Med26 and the subunit CDK9 of SEC phosphorylates Ser2 of pol II CTD resulting in engagement in productive elongation (39).

1.2.3 Integrator in protein coding genes

Until recently, Integrator was thought to be an snRNA-specific processing complex, but it has now been shown to also regulate expression of protein-coding genes (57-59). Expression of the immediate early genes (IEGs) activated by epidermal growth factor (EGF) is regulated by pausing prior to elongation and this pause is overcome by EGF induction. EGF induction regulates the recruitment of Integrator, which in turn recruits the SEC (57). Integrator also plays a role in release of the NELF-mediated pause during transcription of human genes and the HIV genome (58, 59). Interestingly, the human genes targeted by Integrator possess 3' box-like sequences close to the end of the genes, and show a decrease in transcript levels upon Int11 knockdown, suggesting that Int11 is also involved in the production of mRNA (58).

Integrator has also been implicated in eRNA biogenesis by having a role in 3' end cleavage of the primary transcript. Depletion of Integrator has also shown loss of enhancer-promoter interaction and termination defects in eRNA (123). Integrator therefore appears to play major roles in expression of at least some protein-coding genes.

1.2.4 Transcription elongation

In order to engage in productive elongation, pol II has to overcome a pause at the early elongation checkpoint (EEC). The positive transcription elongation factor b (P-TEFb) which contains CDK9 plays a key role in regulating elongation as it phosphorylates pol II

CTD, the negative elongation factor (NELF) and the DRB sensitivity-inducing factor (DSIF). Upon CDK9 phosphorylation, NELF dissociates from pol II allowing progression into productive elongation (124, 125). The SEC, which comprises the elongation factor ELL, AFF1, AFF4, AF9, ENL and P-TEFb is involved in productive elongation of transcription of protein-coding genes (52).

Two similar elongation complexes were later identified; SEC-like complexes 2 and 3 (SEC-L2 and SEC-L3), which are similar in composition and activity for pol II CTD phosphorylation. SEC-L2 comprises P-TEFb, AF9, ENL and AFF2 whereas SEC-L3 comprises P-TEFb, AF9, ENL and AFF3. Although the complexes are very similar, they might be required for regulation of expression of different subsets of genes. It has been shown that HSP70 gene expression requires SEC, but not SECL2 or SEC-L3. In addition, the group of genes affected by each AFF knockdown is different, reinforcing the idea that there is some functional redundancy among these complexes (126).

1.2.5 Splicing

Protein coding-gene primary transcripts (pre-messenger RNA (pre-mRNA)) include both introns and exons and the introns are removed mainly co-transcriptionally to produce the mature messenger RNA. This intron removal process is called splicing and is performed by the spliceosome. The spliceosome is formed by the U snRNPs U1, U2, U4, U5 and U6. U1 snRNA initially interacts with the 5’ splice site at the boundary of an exon and an intron by base pair interaction. The splicing factor 1 (SF1) then binds to the branch

site and the dimeric splicing factor U2 auxiliary factor (U2AF) binds to the polypyrimidine tract and the 3’ splice site (127, 128). The U2 snRNA then displaces SF1 and base pairs to the pre-mRNA at the branch point (129). The spliceosome is then formed in a stepwise manner followed by the recruitment of U4, U5 and U6 snRNPs (130). Once the spliceosome is properly assembled the introns are then removed by two transesterification reactions. First, an OH from a nucleotide on the branch point attacks the first nucleotide of the intron at the 5’ splice site resulting in a lariat. Secondly, an OH from the released 5’ exon attacks the first nucleotide after the 3’ splice site of the intron resulting in the release of the lariat and therefore removal of the intron and joining of the exons (131).

Transcription and pre-mRNA splicing are coordinated with transcription by the pol II CTD (132). The pol II CTD recruits splicing regulatory factors and promotes spliceosome assembly (133). Additionally, it has been shown that the rate of transcription elongation influences splicing as pol II with a low elongation rate promotes the inclusion of alternative exons in alternative splicing; an event where exons can be included in or excluded from the mature RNA resulting in different RNA transcripts from the same encoding region (134).

1.2.6 RNA 3' end formation and termination

3’ end of mRNAs are also formed by co-transcriptional cleavage, as seen for snRNAs, but unlike snRNAs, mRNAs are polyadenylated. Pre-mRNA cleavage is directed by a conserved AAUAAA sequence, the polyA site, and a downstream GU rich sequence (135). The Cleavage and Polyadenylation Specificity Factor (CPSF) complex is responsible for the

recognition of the poly(A) site and cleavage of the transcript. CPSF1 is the subunit that acts as a scaffold for CPSF4 and WRD33 to assemble and recognize the polyA site (136, 137) and also interacts with TFIID and the pol II CTD *in vitro*, linking polyadenylation and transcription (138). CPSF3 is the subunit with catalytic activity and responsible for the pre-mRNA cleavage (139) CPSF2 and CPSF3 are homologues of Int9 and Int11, the catalytic subunits of Integrator and essential for 3'end processing of snRNA genes (140). CstF64, a member of the CstF complex, binds the downstream GU-rich element required for polyadenylation (141). Following the cleavage by CPSF3, the poly-A polymerase (PAP) enzyme adds around 200 A residues to the 3' end of the RNA newly cleaved (135).

There are two models that try to explain the link between termination and cleavage and polyadenylation: the torpedo model and the allosteric model. On the torpedo model there is a cleavage of the nascent RNA which promotes an entry site for the exonuclease Xrn2 to degrade the RNA downstream of the cleavage site while still attached to the pol II. Once Xrn2 reaches pol II, it destabilizes the elongation complex and promotes termination. This has been proposed as the torpedo model of termination (142). In yeast, Xrn2 depletion does not affect cleavage, however, increased levels of pol II downstream of the polyA site are observed by ChIP assays suggesting termination defects which supports the torpedo model (143). The allosteric model proposed that the transcription of the polyA site leads to conformational changes on the elongation complex machinery by either recruiting negative elongation factors or release of anti-termination factors that were travelling along with pol II (142).

1.3 CDK roles in transcription

The largest subunit of mammalian pol II has a carboxyl-terminal domain (CTD) comprising 52 repeats of the consensus heptapeptide $Tyr_1Ser_2Pro_3Thr_4Ser_5Pro_6Ser_7$. During transcription, the pol II CTD is reversibly modified, which helps to recruit transcription and RNA processing factors at the right point of the transcription cycle. Poll II CTD modifications such as phosphorylation are carried out by cyclin-dependant kinases (CDKs) (144). Table 1.2 shows the targets of the transcriptional CDKs, the majority of the kinases are promiscuous and phosphorylate more than one residue of the pol II CTD. CDK10 and CDK19 targets are yet to be discovered (Table 1.2). The RNA polymerase phosphorylation states change during transcription; unphosphorylated pol II can be found in initiation complexes while elongating pol II is heavily phosphorylated. The changes in the phosphorylation state of pol II allow recruitment of different factors at different stages of the transcription cycle. Phosphorylation is reversed by CTD phosphatases, including RPAP2, Ssu72 and Fcp1, allowing pol II to be recycled at the end of the transcription cycle (64).

Table 1.2. Transcriptional CDKs and their pol II CTD targets.

Transcription CDKs 7, 8, 9, 10, 11, 12, 13 and 19 and their targets on the pol II CTD. CDK 10 and CDK19 targets are still not identified. Adapted from Zaborowska *et al*., 2016 (144).

Kinases	Targets
CDK7	Ser5, Ser7
CDK8	Ser5
CDK9	Ser2, Thr4, Ser5, Ser7
CDK10	?
CDK11	Ser2
CDK12	Ser2, Ser5
CDK13	Ser2, Ser5
CDK19	?

1.3.1 CTD phosphorylation during transcription of human snRNA genes

Ser7 is phosphorylated during transcription of all human genes. However its phosphorylation appears to be essential only for expression of snRNA genes (44). Ser5 and Ser7 are both phosphorylated by the CDK7 subunit of TFIIH soon after initiation of transcription (45) and phospho-Ser7 (Ser7P) mediates recruitment of the phospho-Ser5 (Ser5P) phosphatase RPAP2 (80). RPAP2 not only dephosphorylates Ser5P soon after initiation of transcription but also helps to recruit subunits Int1, Int4, Int5, Int6 and Int7 of the large multisubunit Integrator complex (80). The complete Integrator complex comprises 14 subunits and is responsible for the recognition of the 3' box and RNA cleavage to produce the pre-snRNA (47). However, the catalytic subunit, Int11 is absent

when Integrator is associated with RPAP2 (80), and Int9 is also likely to be absent as it interacts strongly with Int11 (145). The CDK9 kinase subunit of the positive-transcription elongation factor b (P-TEFb), which is involved in the expression of both protein-coding genes and of snRNA genes, phosphorylates Ser2 (Ser2P) of the pol II CTD (144) and Ser7 and Ser2 phosphorylation together recruit the missing subunits Int9/11 to activate 3' box recognition and RNA 3' end processing (70). Accordingly, CDK9 inhibitors have a drastic effect on 3' end processing of snRNA gene transcripts (46). This sequence of events is shown in Figure 1.4.

Ser2 can also be phosphorylated by CDK12 (146, 147) and CDK12 knockdown affects snRNA gene expression (148). Phosphorylation of the DRB-sensitivity-inducing factor (DSIF) and the negative elongation factor (NELF) by P-TEFb are also required for productive elongation of protein-coding genes (48, 49, 69) (Section 1.2.4). However, CDK9 inhibitors have little effect on transcription of the U1 and U2 genes (50).

Like protein-coding gene transcripts, snRNA gene transcripts are co-transcriptionally capped (149) (Section 1.2.5). It is therefore likely that Ser5P also ensures capping of pre-snRNAs.

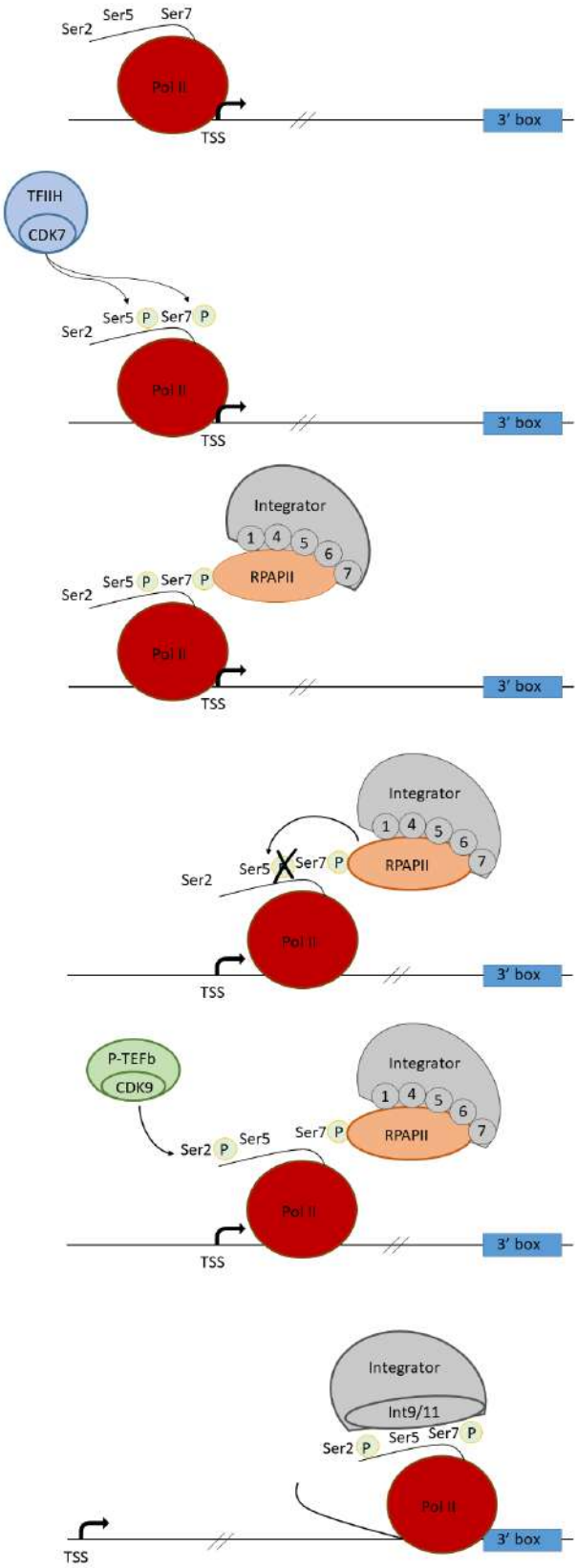


Figure 1.4. Pol II CTD phosphorylation events in snRNA gene transcription.

Initially, the cyclin-dependent kinase (CDK)7 subunit of TFIIH phosphorylates Ser5 and Ser7. RPAP2 interacts with Ser7P. RPAP2 in turn recruits the Integrator subunits Int1, Int4, Int5, Int6 and Int7. RPAP2 dephosphorylates Ser5P, and positive-transcription elongation factor b (P-TEFb) is recruited by a mechanism still unknown. The P-TEFb subunit CDK9 phosphorylates Ser2. The double phosphorylation on Ser2 and Ser7 recruits Int9/11, which allows RNA processing to occur (80).

1.3.2 CTD phosphorylation on protein coding genes

Protein coding genes transcription regulation by CDKs starts with the phosphorylation of Ser5 by CDK7. Ser5P recruits the 5'end capping machinery and helps pol II to escape the promoter to engage in productive elongation. Ser5P has been shown to correlate with splicing on protein coding genes in mammals (150). In fact, Ser5P has also been shown to recruit the spliceosome and regulate splicing in yeast (151). Those studies suggest that Ser5P couples splicing to transcription.

Ser7 phosphorylation is also high at the transcription start site (152). However, the relevance of this phosphomark for expression of protein coding genes is unclear as mutation of Ser7 to alanine has little effect on transcription of these genes (153). Tyr1P is also increased at the transcription start site and has been associated with antisense promoter transcription and active enhancers (154).

Ser2 and Thr4 phosphorylation is highest towards the polyadenylation and transcription termination sites. In yeast, Thr4P recruits the termination factor Rtt103 and regulates pol II pausing after the polyadenylation site (155). Thr4P has also been shown to be involved in elongation (156), splicing (155) and in 3'end formation of histones transcripts (157).

CDK12 and 13 have been shown to phosphorylate Ser2 *in vitro* and have been suggested to contribute to the majority of the Ser2P during transcription elongation (146, 158). Ser2P is recognized by the elongation factors Spt6 which recruits the nuclear RNA export factor REF1/Aly in mammalian cells (159). The connection between transcription and splicing has also been shown by the recognition of Ser2P by the splicing factor Prp40

in yeast (133). Further studies have shown that both Ser2P and Ser5P are recognized by the splicing factor U2AF65 which leads to the recruitment of the Prp19 complex (160), which adds to the increasing evidence of the coupling between splicing and transcription. In summary, the increase of Ser2P and Thr4P towards the 3'end of protein-coding genes facilitates the recruitment of factors and complexes involved in transcription elongation, termination and mRNA splicing and 3'end processing. For instance, Ser2P is recognized the Cleavage and Polyadenylation Factor Subunit pcf11 (161). Also, Thr4P has been shown in yeast to recruit the termination factor Rtt103 and regulates pol II pausing downstream of polyA sites and facilitates transcription (155).

1.3.3 CDK analog-sensitive cell lines

CDKs are highly conserved kinases and very similar in their active sites. This high similarity makes it difficult to produce specific inhibitors for a single CDK. DRB is the most well-known drug that has been used for inhibition of CDK9 (162), but not CDK12 (163). However, it is still unknown whether DRB inhibits other kinases. THZ531 has been used to inhibit CDK12 and CDK13 (164) and Cortistatin A inhibits both CDK8 and CDK19 (165). Therefore, discerning the role of each kinase in transcription is not an easy task.

To circumvent this problem, Bishop and Shokat pioneered the development of so-called "analog-sensitive" kinases where the bulky "gatekeeper" residue of the active site is mutated to a smaller amino acid. This mutation selectively allows the entry of a bulky ATP analogue inhibitor. Thus, other wild type kinases are not affected by the analogue, allowing selective inhibition of the kinase being studied by chemical genetics (Figure 1.5) (166). Using CRISPR/Cas9 genome engineering, it now possible to make endogenous

analog-sensitive kinases. Analog-sensitive kinases have recently allowed the transcription field to advance and identify roles for previously unknown CDKs (167, 168). For example, a recombinant analog sensitive (as) CDK12 was engineered and its activity shown to be unaffected in vitro kinase assays. However, the change of the gatekeeper residue to glycine causes the CDK12as to be effectively inhibited by the bulky adenine analog 1-NM-PP1 (167). The same group then engineered a CDK12as cell line by using CRISPR/Cas9 genome engineering. Cas9 was associated with a guide RNA (gRNA) targeting the gatekeeper and a repair template was provided carrying the mutation for the exchange of the phenylalanine residue 813 to glycine. The repair template was inserted into the genome by homologous repair generating a CDK12as cell line (168). Inhibiting CDK12as causes a change in the phosphorylation status of the pol II CTD, in little as 15 minutes inhibition with 1-NM-PP1. A 50% reduction of Ser2P was observed within 1h. A reduction of Ser5P was also observed, which is in line with the finding that CDK12 can also phosphorylate Ser5 in vitro (168). Recent work on CDK12as shows that CDK12 is required for the transcription of core DNA replication genes and therefore controls G1/S progression linking transcriptional CDKs to cell cycle (169).

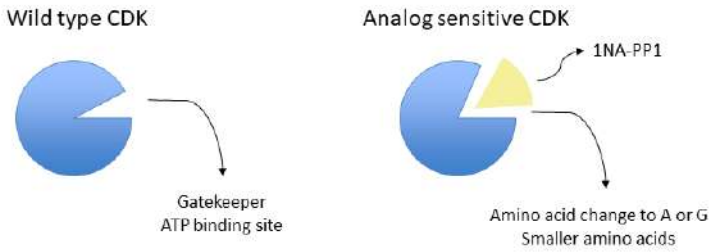


Figure 1.5. Scheme of engineered analog sensitive kinases.

The wild type CDK on left showing the gatekeeper and on the right the analog sensitive CDK with the larger gatekeeper due to the amino acid substitution (166).

1.4 Unbiased methods to Pulldown proteins

1.4.1 LNA oligos

Proteomics of isolated chromatin segments (PICh) uses DNA oligo as baits to pulldown proteins associated with a specific genomic locus and further identify them by mass spectrometry. The DNA oligos have an altered backbone made of locked nucleic acids (LNA), which has an increased melting temperature. PICh has been used for repetitive regions such as telomeres (170) and centromeres (171). PICh also been further developed into end-targeting PICh (ePICh) and used for the identification of proteins bound to the promoter of ribosomal RNA genes.

1.4.2 enChIP

Engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) was one of the first methods using dCas9 for pulldowns (172). This consists of a dCas9-FLAG that targets the region of interest and can be pulled down by using a FLAG antibody. Fujita *et. al* targeted the Interferon Regulatory Factor 1 (IRF1) promoter region in order to identify new proteins that might be involved in transcription regulation of that gene (172). Although claimed to successfully identify a list of proteins, the results are contradictory. The authors did not perform western blot analysis to analyse whether dCas9 had been pulled down. In addition, their mass spectrometry list was full of potential contaminants such as actin and tubulin and abundant factors involved in elongation of transcription and translation. Furthermore, the lack of negative controls

such as a pulldown with dCas9 without a specific guide made it impossible to discriminate between proteins enriched on their region of interest and non-specific background.

1.4.3 CLASP

Cas9 locus-associated proteome (CLASP) is another method for unbiased chromatin-associated protein isolation, but it differs from the other methods due to *in vitro* targeting of chromatin (173). A dCas9 assembled with the guide is purified and combined with sonicated chromatin and the eluted proteins are submitted to mass spectrometry. Tsui *et al.* established the method and used dCas9 to target the histone cluster in *Drosophila melanogaster* in order to identify new factors involved in histone gene expression. They have used a set of guides targeting the H2A/H2B promoter and narrowed their list of candidates by normalizing the protein abundance to the pulldown using a non-targeting gRNA. To validate their candidates, they performed dsRNA-mediated RNAi in S2 cells and showed that two of their candidates, Vig and Vig2 bind to the 3' UTR of the H2A mRNA and can downregulate its levels (173). Intriguingly, all factors identified by CLASP were RNA binding proteins instead of DNA binding transcription factors as expected. Therefore, although CLASP has the major advantage of not having to generate stable cell lines expressing dCas9, this method does not appear an ideal way to identify transcription factors.

1.4.4 CAPTURE

CRISPR affinity purification *in situ* of regulatory elements (CAPTURE) stands out from the other dCas9 Pulldown methods by having *in vivo* dCas9 biotinylation (174). It consists of a system with three plasmids: one encoding dCas9 with a biotin acceptor peptide (BAP); a second plasmid encoding BirA which biotinylates dCas9 *in vivo*; and a third plasmid containing the guides. Stable cell lines are generated for the dCas9 and BirA plasmids, whereas the guides are transfected transiently and the transfected cells sorted by FACS using mCherry. The cells are crosslinked before analysis to stabilize interactions between proteins and nucleic acids. Another attractive feature of the CAPTURE method is that the urea washes enrich for the crosslinked chromatin-bound factors (175). CAPTURE has been used not only to identify proteins bound to a specific region of interest by dCas9 pulldown followed by mass spectrometry but also for analysing chromatin long range interactions identified by 3C, HiC and 5C (174).

1.4.5 Genomic Locus Proteomics (GLoPro)

The Genomic Locus Proteomics (GLoPro) is another dCas9 pulldown method to identify proteins in an unbiased way. In this case, dCas9 is fused to an engineered peroxidase APEX2 (CASPEX). In the presence of hydrogen peroxide, the enzyme oxidises biotin, turning it into radicals that react with tyrosine residues, thereby labelling the proteins within a 400bp each side of the targeted site with biotin. The proteins are pulled down by using streptavidin magnetic beads followed by trypsin digestion on beads and further mass spectrometry (176).

To circumvent the issue of dCas9 preventing proteins from binding due to physically blocking their binding site, GLoPro uses sgRNAs tiling the region of interest, with non-targeting guides as controls. They have targeted CASPEX to the hTERT locus, incubated the whole non-crosslinked cell lysate with streptavidin beads, followed by stringent washing. The bound peptides were digested on beads and analysed by LC-MS/MS. Two proteins already described to be in that region were identified by GLoPro, CTBP1 and MAZ (177-179). CASPEX was also targeted to the MYC promoter and was also successful in identifying proteins bound to that region (176).

1.5 Project aims and description

The overall aim of my DPhil studies was to better understand the regulation of expression of the human snRNA genes, with a focus on regulation at the levels of transcription and RNA processing. The two main approaches I have used are complementary; the first is an unbiased approach to identify proteins involved in snRNA gene transcription and the second is a candidate approach where I have assessed the roles of transcriptional CDKs that could potentially regulate snRNA gene transcription and processing of the transcripts.

The first project aims to identify new factors that might be involved in snRNA gene transcription that could help explain the coupling of the PSE and the 3'box in snRNA genes. To address this question in an unbiased way, dCas9 was targeted to different parts of the U2 snRNA gene by specific guide RNAs to the region of interest. dCas9 was then pulled down and the protein complex submitted to mass spectrometry, in order to identify components of the transcription apparatus, followed by validation.

The second project aimed to characterise the roles of the transcriptional CDKs in transcription of snRNA genes and processing of the transcripts. Cell lines with analog-sensitive versions of CDK7, CDK8, CDK9, CDK10, CDK12, CDK13 and double CDK12/CDK13 have been used to assessing the effect of inhibition on pol II levels. In addition, the effect of inhibition on RNA 3'end processing was assessed by RNA protection assays.

Together, these two complementary approaches have identified new factors involved in expression of snRNA genes and suggested a new model for the coupling of initiation from snRNA gene promoters and recognition of the 3' box.

Chapter 2

Materials and Methods

2.1 HEK293 Cell Culture

Human embryonic kidney (HEK) 293 were grown in Dulbecco's Modified Eagle's Medium (DMEM) with 10% fetal bovine serum, 2 mM L-glutamine, 50 U/mL penicillin and 50 µg/mL streptomycin and incubated at 37˚C in 5% CO2.

2.2 dCas9 cloning and guide RNA design

I have introduced the mutation H840A on the nickase Cas9 in order to produce a dCas9. Using pX462 as a template, PCR was used to amplify 500bp from an EcoRV site until the mutation to be inserted and another 500bp with the mutation (H-> A) on the left primer and the right primer across an FseI site. The two 500bp fragments were then ligated and amplified with the EcoRV and FseI primers and inserted into the pX462 cut with EcoRV and FseI in order to replace the fragment and insert the mutation (Fig. 2.1). dCas9 was sequenced to guarantee that unwanted mutations were not acquired and that no original DNA sequence was present. A short fragment of the dCas9 sequence containing the mutation inserted on the gatekeeper is shown in Fig.S2.1 (See page 107).

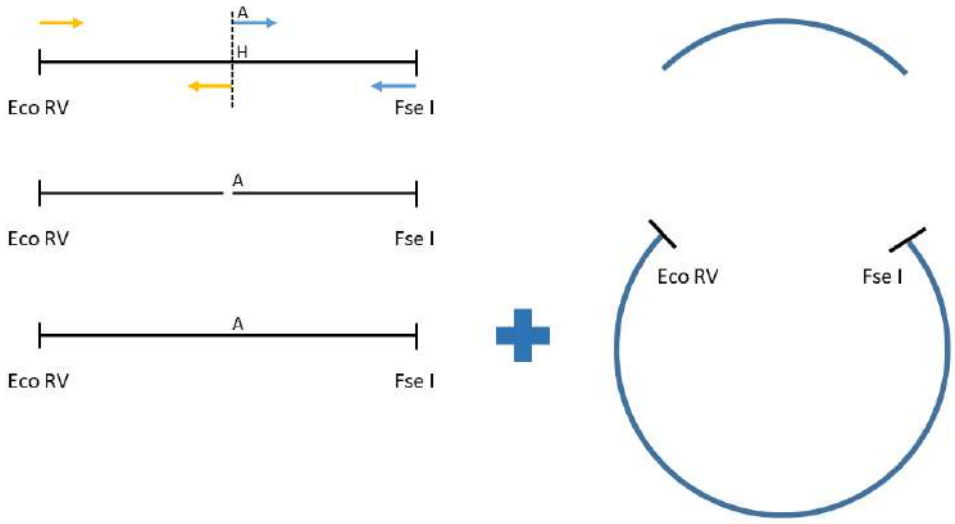


Figure 2.1. Insertion of the H840A mutation on pX462.

Scheme showing the design of the primers to insert the H840A mutation into pX462. The codon for the alanine to be inserted is present in the left primer (on the second half of the 1kb product generated). The primers illustrated are: in yellow, dCas9 EcoRV F and R; and in blue, dCas9 FseI F and R. The pX462 vector was cut with EcoRV and FseI and ligated with the newly *in vitro* transcribed template.

The guide RNAs were designed using the online Guide Design Resources tool developed by the Zhang Lab from MIT. A range of single and paired guides were selected based on the software output for the DSE, PSE and 3'box regions.

2.3 Transfection of HEK293 cells

Transfections were performed using jetPRIME (Polypus transfection) according to the manufacturer's instructions. The amount of DNA transfected was according to the datasheet specific for the size dish used, following the ratio suggested of 1 µg of DNA per

2 µl of jetPRIME. The DNA was diluted into the jetPRIME buffer, vortexed and the jetPRIME reagent was then added. The DNA and jetPRIME mix was vortexed again and incubated for 10 minutes at room temperature. The mix was then pipetted onto the dish with the growing cells and incubated for 24 hours at 37˚C in 5% CO2.

2.4 SDS-PAGE gels

2.4.1 Western Blot

Proteins were first separated by SDS-PAGE with different concentrations of acrylamide depending of the size of the protein-of-interest. Resolving gels ranged from 8% to 15% of acrylamide, while the stacking gels were 4% of acrylamide mix. Proteins were denatured in Laemmli buffer by heating at 95˚C for 5 minutes. Gels were run in 1x running buffer, first at 100V until the proteins ran past the stacking gel, and then at 140V until the proteins reached the bottom of the gel. The proteins were then transferred using a semi-dry transfer system (Trans-blot SD Semi-Dry transfer cell (Bio-Rad) to a nitrocellulose membrane in a transfer buffer containing methanol for 1 hour at 10V. The membrane was then blocked in PBS containing 5% milk and 0.1% Tween-20 for 1 hour at room temperature. After blocking, the membrane was incubated for 1 hour at room temperature or overnight at 4˚C with the primary antibody against the protein of interest in PBS containing 1% milk and 0.1% Tween-20 (See Table 2.2 for references for the antibodies used). The membranes were then washed three times for 10 minutes with PBS containing 1% milk and 0.1% Tween-20 and further incubated for 1 h at room temperature with the appropriated HRP-conjugated secondary antibody diluted in PBS

1% milk and 0.1% Tween-20. The membrane was washed another three times for 10 min and developed by using ECL according to manufacturer's instructions (Visualizer WB detection kit) and X-ray films in a compact X4 X-ray film processor (Xograph imaging system, UK).

2.4.2 Silver staining

Proteins were first separated by SDS-PAGE on a gradient gel ranging from 5% to 20% acrylamide from top to bottom. As for the western blot (Section 2.4.1), proteins were denatured in Laemmli buffer by heating at 95˚C for 5 minutes and the electrophoresis was run in 1x running buffer, first at 100V until the proteins moved to separating gel, and then at 140V until the proteins reached bottom of the gels. The gel was then fixed in fixative solution (50% ethanol, 10% acetic acid, 40% H_2O, 1.85% formaldehyde) shaking for 1 hour. The gel was then washed several times with dH_2O and kept shaking in dH_2O for 30 minutes. The gel was then incubated with DTT solution (5µg/ml DTT in dH_2O) shaking for 1 hour. After rinsing the DTT solution, the gel was incubated with the silver solution (0.1% $AgNO_3$) shaking for 1 hour and protected from light. The gel was rinsed with dH_2O for 1 min and developed with the freshly made developer solution (3% Na_2CO_3, 1.85% formaldehyde).

2.5 ChIP

ChIP were performed as described in (Laitem et al., 2015). Approximately 1 x 10^7 HEK293 cells were crosslinked with 1% formaldehyde at room temperature for 10 min. The reaction was quenched with 0.125 M glycine for 5 minutes at room temperature. Cells were washed twice with pre-chilled PBS, and then collected by scraping. All subsequent steps were carried out on ice. Cells were pelleted by centrifugation at 1500 RPM for 10 minutes. The cells were resuspended in 0.5 ml ChIP Lysis Buffer (10mM Tris-HCl pH 8.0, 0.25% Triton, 10mM EDTA + cOmplete EDTA-free Protease Inhibitor Cocktail (Roche) + PhosSTOP (Sigma)) and incubated for 10 min on ice. Nuclei were spun down at 1500 x *g* for 5 minutes at 4˚C and the supernatant was discarded. The pelleted nuclei were resuspended in 0.5 ml of ChIP wash buffer (10mM Tris-HCl pH 8.0, 200mM NaCl, 1mM EDTA + cOmplete EDTA-free Protease Inhibitor Cocktail (Roche) + PhosSTOP (Sigma)) and spun down at 1500 x *g* for 5 minutes at 4˚C. Nuclei were re-suspended in 300 µl ChIP sonication buffer (10mM Tris-HCl pH 8.0, 100mM NaCl, 1mM EDTA, 0.1% SDS + cOmplete EDTA-free Protease Inhibitor Cocktail (Roche) + PhosSTOP (Sigma)) and sonicated for 1 hour (30s on, 30s off) on the Q Sonica Thermocube Q800R. After sonication, the samples were spun down and the supernatant containing the chromatin was transferred to a new 1.5mL eppendorf tube. To pre-clear the sheared DNA, the chromatin was incubated with 10 µl Dynabeads Protein G for 30 minutes on a rotating wheel at 4˚C. Chromatin was then quantified by NanoDrop One. 85 µg of chromatin were incubated with 1 µg of the antibody against the protein of interest rotating at 4˚C overnight (See Table 2.2 for references for the antibodies used). For each IP samples, 15 µl of Dynabeads Protein G were blocked with RIPA containing 4 mg/ml BSA rotating at 4˚C

overnight. The blocking solution was removed from Dynabeads and the sonicated samples pre-incubated with antibody were added to the beads and left rotating for 1 hour at 4˚C. The unbound fraction of the IgG control was kept as input. Beads were then separated on a magnetic rack DynaMag-2 Magnet (Thermo Fisher) and washed three times with 300 µl RIPA, three times with 300 µl high salt buffer, two times with 300 µl LiCl buffer and two times with 300 µl TE buffer. The immunoprecipitated DNA-protein complexes bound to the beads were then eluted in two times 15 min steps with 50 µl of elution buffer containing 10 mM DTT on a horizontal shaker at 1300 RPM room temperature. The input samples were diluted 1:10 in elution buffer. Samples were incubated with 10 µg RNaseA for 30 min at 37˚C to degrade the RNA. The samples were then incubated for 5 h at 65˚C with a final concentration of 200 mM NaCl for reverse crosslinking. The DNA was then precipitated with 300 µl ethanol at -20˚C overnight. The DNA was centrifuged at 20000 x *g* for 20 minutes and each DNA pellet was re-suspended in 100 µl TE buffer with 20 µg of proteinase K in 25 µl of 5x proteinase K buffer, and incubated for 1h30 to 2h at 42˚C. The DNA was then purified using the QIAGEN QIAquick PCR purification kit and eluted in 50 µl EB for qPCR analysis.

2.6 qPCR

ChIP assays were analysed by qPCR on a Rotor-Gene 3000 cycler (Corbett Research), equipped with Rotor-Gene 3000 Software analysis program. Each reaction contained: 5 µl of 2x QuantiTect SYBR Green PCR, 3 µl RNase/DNase-free water, 10 pmol/µl of each oligonucleotides and 1 µl of purified ChIP DNA template. A standard curve was produced for each primer set using serial dilutions of the input DNA sample (1:5, 1:25, 1:125 and

1:625 dilutions) The immunoprecipitated and control DNA samples were assessed technical triplicate. Oligonucleotides used are listed in Table 2.1 at the end of this chapter. The thermal cycles were: 95˚C for 15 minutes; 40 cycles of 94˚C for 15 seconds, 57˚C for 15 seconds, 72˚C for 15 seconds. The last cycle ended with a 10 minutes melt curve from 50˚C to 99˚C, with 1˚C increase every 5 seconds.

The Ct value (copies/µl) for each samples were quantified relative to the standard and subtracted by the IgG control Ct value. The resulting values are the % of input being pulled down as shown in the ChIP graphs. Experiments were replicated at least three times and each ChIP sample was measured in triplicate by qPCR. Data presented represent the mean ± SEM (standard error of the mean) of three independent experiments. P-values were calculated by two-tailed unpaired t-test. $p<0.05$ was considered to be statistically significant. The qPCR primers used can be found in Table 2.1 at the end of this chapter.

2.7 enChIP

enChIP was performed as in (Fujita, 2013). Approximately 2 x 10^6 HEK293 cells were transfected with 2 µg of gRNA cloning vector with Lipofectamine 2000 (Invitrogen) on the first day. On the following day, cells were re-plated onto 10-cm tissue culture plates. On the third day, cells were fixed with 1% formaldehyde at 37 °C for 5 minutes. Cells were washed twice with ice-cold PBS and pelleted for 10 min at 1500 rpm at 4°C. Cell pellet was resuspended in 0.5mL cell lysis buffer (10 mM Tris pH 8.1, 1 mM EDTA, 0.5% NP-40, 50 mM NaF + cOmplete EDTA-free Protease Inhibitor Cocktail (Roche) + PhosSTOP

(Sigma)) and the nuclei pellet was collected by centrifugation for 5min at 500g and 4°C. Nuclei was resuspended in 0.5 ml nuclear lysis buffer (10 mM Tris pH 8.1, 1 mM EDTA, 0.5 M NaCl, 1% Triton, 0.5% deoxycholate, 0.5% Sarcosyl, 50 mM NaF + cOmplete EDTA-free Protease Inhibitor Cocktail (Roche) + PhosSTOP (Sigma)) spun down for 5minutes, 800 x *g* at 4°C. Chromatin was re-suspended in 0.8 ml of Sonication Buffer (10 mM Tris pH 8.0, 150 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% sodium deoxycholate, 0.1% SDS, 1% triton X-100 + cOmplete EDTA-free Protease Inhibitor Cocktail (Roche) + PhosSTOP (Sigma)). The chromatin fraction was extracted and fragmented by sonication to 200bp on the Q Sonica Thermocube Q800R (35% amplitude, 30 seconds on and 30 seconds off for 1 hour). The sonicated chromatin in Sonication Buffer with 1% Triton X-100 was pre-cleared with 15 µg of normal mouse IgG (Santa Cruz Biotechnology) conjugated to 150 µl of Dynabeads-Protein G (Invitrogen) and further incubated with 15 µg of anti-FLAG M2 Ab (Sigma–Aldrich) conjugated to 150 µl of Dynabeads-Protein G at 4 °C overnight. The Dynabeads were washed twice each with 1 ml of Low Salt Wash Buffer (20 mM Tris, pH 8.0, 150 mM NaCl, 2 mM EDTA, 1% TritonX-100, 0.1% SDS), High Salt Wash Buffer (20 mM Tris, pH 8.0, 500 mM NaCl, 2 mM EDTA, 1% TritonX-100, 0.1% SDS), LiCl Wash Buffer (10 mM Tris, pH 8.0, 250 mM LiCl, 1 mM EDTA, 0.5% IGEPAL-CA630, 0.5% sodium deoxycholate), and TE Buffer (10 mM Tris, pH 8.0, 1 mM EDTA. Elution was done by incubating 35 µl of 3xFLAG peptide (Sigma–Aldrich) (500 µg/ml) in TBS with 0.1% IGEPAL CA-630 at 37 °C for 20 min. The eluted samples were precipitated by isopropanol. The precipitants were suspended in 40 µl of 2× sample buffer, incubated at 100 °C for 30 min for reverse-crosslinking and denaturation of proteins, and subjected to SDS–PAGE followed by either silver staining or immunoblotting. When sent to mass spectrometry the precipitants from isopropanol were resuspended in 1M TEAB buffer. The samples

were sent to the Advanced Proteomics Facility, Department of Biochemistry Oxford. The samples were digested by the facility with trypsin FASP and run in a Ultimate 3000 RSLCnano HPLC (Dionex, Camberley, UK) coupled to a QExactive Orbitrap mass spectrometer (Thermo Electron, Hemel Hempstead, UK). The database search algorithm used was Mascot v2.2.04, MS/MS Ion Search. The analysis was performed by the facility.

2.8 Chromatin Extraction and MNase digestion (Net-Seq)

1×10^7 cells were washed twice with ice-cold PBS and collected in a 15 ml tube. Cells were centrifuged at 420g for 5 min at 4 °C. The cell pellet was resuspended with 4ml of ice-cold HLB+N (10mM Tris-HCl (pH 7.5), 10 mM NaCl, 2.5 mM $MgCl_2$ and 0.5% NP-40, freshly added cOmplete EDTA-free Protease Inhibitor Cocktail (Roche) and PhosSTOP (Sigma)) and incubated on ice for 5 minutes. The cell pellet was underlaid with 1ml of ice-cold HLB+NS (10 mM Tris-HCl (pH 7.5), 10 mM NaCl, 2.5 mM MgCl2, 0.5% NP-40 and 10% sucrose, freshly added cOmplete EDTA-free Protease Inhibitor Cocktail (Roche) and PhosSTOP (Sigma)) and centrifuged at 420 x *g* for 5 minutes at 4°C. The nuclear pellet was re-suspended in 125 µl of ice-cold NUN1 buffer (20 mM Tris-HCl (pH 7.9), 75 mM NaCl, 0.5 mM EDTA and 50% glycerol, freshly added cOmplete EDTA-free Protease Inhibitor Cocktail (Roche) and PhosSTOP (Sigma)) and transferred to a 1.5 ml tube. 1.2 ml of ice-cold NUN2 (20 mM HEPES-KOH (pH 7.6), 300 mM NaCl, 0.2 mM EDTA, 7.5 mM MgCl2, 1% NP-40 and 1 M urea, freshly added cOmplete EDTA-free Protease Inhibitor Cocktail (Roche) and PhosSTOP (Sigma)) buffer was added and the sample vortexed at maximum speed to mix. Samples were incubated on ice for 15 minutes and vortexed every 3–4 minutes. The samples were centrifuged at 16,000 x *g* for 10 minutes at 4 °C and the

supernatant was discarded to collect the chromatin pellets. The chromatin was then digested by MNase: first, 100 µl of MNase buffer was added to the chromatin pellets and centrifuged at 16,000g for 10 s at 4 °C and the supernatant was discarded to collect the chromatin pellets; then, 100 µl of MNase reaction mixture (1x MNase buffer, 1x BSA, 40 gel unit/µl) was added to the pellet. The samples were incubated at 37 °C for 90 seconds at 1,400 rpm. on a thermomixer. The MNase was inactivated by adding 10 µl of 250 mM EGTA to the samples. The samples were then centrifuged at 16,000 x *g* for 5 minutes at 4 °C. The supernatant was then used as input for the pulldown and the enChIP protocol was followed from the incubation with the conjugated Dynabeads. (Section 2.7).

2.9 CAPTURE qPCR

CAPTURE qPCR was performed as described in Liu et al, 2017 (174). 1×10^7 FB-dCas9/BirA-expressing HEK293 stable cells transduced with sequence-specific gRNAs crosslinked with 1% formaldehyde for 10 minutes, and then quenched with 0.125 M glycine for 5 minutes. Cell were washed twice wish PBS, scraped and lysed in 1 ml RIPA buffer (10 mM Tris-HCl, 1 mM EDTA, 0.1% sodium deoxycholate, 0.1% SDS, 1% Triton X-100, pH 8.0 + cOmplete EDTA-free Protease Inhibitor Cocktail (Roche)), and rotated for 15 minutes at 4°C. Cell lysates were centrifuged at 2,300 x *g* for 5 minutes at 4°C to isolate the nuclei. Nuclei were suspended in 500 µl of 0.5% SDS lysis buffer (0.5% SDS, 10 mM EDTA, 50 mM Tris-HCl, pH 8.0 + cOmplete EDTA-free Protease Inhibitor Cocktail (Roche)) and sonicated to 200bp on the Q Sonica Thermocube Q800R (35% amplitude, 30 seconds on and 30 seconds off for 1 hour). Sonicated chromatin was centrifuged at 16,100 x *g* for

10 minutes at 4°C. 450 µl of supernatant was transferred to a new Eppendorf tube and NaCl to a final concentration of 300 mM was added. The supernatant was then incubated with 10 µl of MyOne Streptavidin T1 Dynabeads (Thermo-Fisher Scientific) at 4°C overnight. After overnight incubation, Dynabeads were washed twice with 1 ml of 2% SDS, twice with 1 mL of RIPA buffer with 0.5 M NaCl, twice with 1 mL of LiCl buffer (250 mM LiCl, 0.5% NP-40, 0.5% sodium deoxycholate, 1 mM EDTA and 10 mM Tris-HCl, pH 8.0), and twice with 1 ml of TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). The chromatin was eluted in SDS elution buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl, pH 8.0) followed by reverse cross-linking at 65°C overnight. The ChIP DNA was treated with RNase A (5 µg/ml) and proteinase K (0.2 mg/ml) at 37°C for 30 min, and purified using QIAquick Spin columns (QIAGEN) eluted in 50 µl of EB for qPCR analysis.

2.10 CAPTURE Proteomics

4 x 10^6 to 1 × 10^7 FB-dCas9/BirA HEK293 stable cells transfected with sequence-specific sgRNAs were cross-linked with 2% formaldehyde for 10 minutes and then quenched with 0.25 M glycine for 5 minutes. Cells were washed twice with PBS, scraped and lysed with 10 ml of cell lysis buffer (25 mM Tris-HCl, 85 mM KCl, 0.1% Triton X-100, pH 7.4, freshly added 1 mM DTT and cOmplete EDTA-free Protease Inhibitor Cocktail (Roche) ++ PhosSTOP (Sigma)) and rotated for 15 minutes at 4°C. Cell lysates were centrifuged at 2,300 x *g* for 5 minutes at 4°C to isolate the nuclei. The nuclei were resuspended in 5 ml nuclear lysis buffer (50 mM Tris-HCl, 10 mM EDTA, 4% SDS, pH 7.4, freshly added 1 mM DTT and cOmplete EDTA-free Protease Inhibitor Cocktail (Roche) + PhosSTOP (Sigma)) and incubated for 10 minutes at room temperature. The nuclei

suspension was then mixed with 15 ml of 8 M urea buffer and centrifuged at 16,100 x *g* for 25 minutes at room temperature. Nuclei pellets were then re-suspended in 5 ml nuclear lysis buffer and mixed with 15 ml of 8 M urea buffer (10mM Tris, 1mM EDTA, 8M Urea, pH 7.4 + cOmplete EDTA-free Protease Inhibitor Cocktail (Roche) + PhosSTOP (Sigma)), and centrifuged at 16,100 x g for 25 min at room temperature. The samples were washed twice more in 5 ml nuclear lysis buffer and mixed with 15 ml of 8 M urea buffer, followed by centrifugation at 16,100 x *g* for 5 minutes at room temperature. Pelleted chromatin was then washed twice with 5 ml cell lysis buffer. The chromatin pellet was resuspended in 5 ml of IP binding buffer without NaCl (20 mM Tris-HCl, 1 mM EDTA, 0.1% NP-40, pH 7.5, freshly added cOmplete EDTA-free Protease Inhibitor Cocktail (Roche) and PhosSTOP (Sigma)) and aliquoted into Eppendorf tubes. Chromatin was then subjected to sonication to 200 bp on the Q Sonica Thermocube Q800R (35% amplitude, 30 seconds on and 30 seconds off for 1 hour). Fragmented chromatin was centrifuged at 16,100 x g for 25 min at 4°C. Supernatant was combined and final concentration 150 mM NaCl was added to the sheared chromatin. To prepare the streptavidin beads for affinity purification, 120 µL of streptavidin agarose slurry (Life Technologies) was washed 3 times in 1 ml of IP binding buffer and added to soluble chromatin. After overnight incubation at 4°C, streptavidin beads were collected by centrifugation at 800 x *g* for 3 minutes at 4°C. The beads were then washed 5 times with 1 ml of IP binding buffer (20 mM Tris-HCl, 1 mM EDTA, 0.1% NP-40, 150-300 mM NaCl, pH 7.5, freshly added proteinase inhibitor) and proteins were then de-crosslinked by increasing the NaCl concentration to 3M and boiling it at 95°C for 1 hour. The samples were sent to the Advanced Proteomics Facility, Department of Biochemistry Oxford. The analysis was performed by the facility, the samples were digested with trypsin FASP and run in an Ultimate 3000 RSLCnano HPLC (Dionex, Camberley, UK)

coupled to a QExactive Orbitrap mass spectrometer (Thermo Electron, Hemel Hempstead, UK). The database search algorithm used was Mascot v2.2.04, MS/MS Ion Search.

2.11 RNA extraction

Approximately 3 x 10^6 cells were scraped into a 1.5mL eppendorf and spun at 13000 RPM at room temperature for 1 minute. The cell pellet was resuspended in 100 µl RNA Lysis Buffer (140mM NaCl, 1.5mM $MgCl_2$, 10mM Tris-HCl pH 7.5, 0.5% v/v NP-40). 100µl of 2xHSB (280mM NaCl, 10mM KCl, 1.5mM Na_2HPO_4, 12mM Glucose, 50mM HEPES, pH 7.05) and 5 units of DNAse were added to the sample and incubated for 5 minutes at 30°C. 100 µl of PK solution (0.4mg/ml PK, 0.6% SDS) was added to the sample and incubated for 30 min at 37°C. The RNA was purified by phenol/chloroform extraction and ethanol precipitated. The RNA was resuspended in 100 µl DNase buffer (10mM Tris-HCl pH 7.5, 10mM $MgCl_2$) and 10 units of DNase were added. The sample was incubated for 1 hour at 37°C. The RNA was then purified a second time: 10 µl of 3M NaOAc was added and the RNA was phenol/chloroform extracted and kept in ethanol at -20^{o}C overnight.

2.12 RNase protection

DNA marker

The pBR322 DNA-MspI Digest DNA marker was 5' end labelled using adenosine triphosphate labelled on the gamma phosphate group with ^{32}P. The labelling was done by adding 1 µl of the DNA marker, 1µl of 10x PNK buffer, 1µl T4

PNK, 6µl autoclaved H_2O and 0.1M DTT and incubating at 37°C for 1hour. The DNA was then precipitated by adding 80 µl of autoclaved dH_2O, 10 µl of 4M AmoAc, 250 µl 100% ethanol and 10 µg of tRNA and resuspended in formamide blue juice (99% v/v formamide, 0.1% w/v xylene cyanol, 0.2% w/v bromophenol blue, 5mM EDTA) and kept at -20°C.

Riboprobe

The DNA template used was the U2G construct from Shona Murphy (180). 500 ng of the linearized DNA template was incubated with 1x Transcription buffer, 10mM DTT, 400µM ATP, CTP, GTP, 9.6µM UTP, 20uCi UTP, 10 units RNAsin and 20 units of T7 RNA Polymerase. The reaction was kept at 37^{o}C for 1 to 2 hours then 10 µl of formamide blue juice (99% v/v formamide, 0.1% w/v xylene cyanol, 0.2% w/v bromophenol blue, 5mM EDTA) was added and the sample was denaturated at 85°C for 5 minutes. The sample was then loaded onto a 6% polyacrylamide-urea gel and purified. The gel slice was placed in 400 µl of RNA elution buffer (0.5M NaOAc pH 5.2, 0.5% w/v SDS) for 1 hour at 37°C. The riboprobe was then phenol/chloroform extracted and ethanol precipitated with tRNA. The pellet was resuspended in autoclaved water so 1µl equals to 500cps. The riboprobe was kept at -20°C.

RNase Protection

In order to hybridize the riboprobe to RNA, 20 µg of the RNA extracted was re-suspended in 30 µl of R-loop buffer (80% v/v formamide, 400mM NaCl, 40mM pipes pH 6.4, 1mM EDTA) and 1 µl of the riboprobe. The reaction was incubated for 10 minutes at 80°C and hybridised at 56°C overnight. An RNase protection solution was made by adding 1 ml of RNase protection buffer (300mM NaCl, 10mM Tris-HCl pH 7.5, 5mM EDTA), 4µl RNase A (10mg/ml) and 2µl RNase T1 (500.000 U/ml). 300 µl of the RNase protection buffer was added to each sample and mixed. Samples were kept at 18°C for 1.5 to 2 hours. 30 µl of proteinase K mix (3.3 mg/ml proteinase K, 7.7% SDS) was added to each sample and incubated at 37°C for 30 min. The samples were then purified by phenol/chloroform extraction and re-suspended in 10 µl of formamide blue juice (99% v/v formamide, 0.1% w/v xylene cyanol, 0.2% w/v bromophenol blue, 5mM EDTA) and loaded onto 8% polyacrylamide-urea gels.

Polyacrylamide-urea gels

8% polyacrylamide-urea gel recipe: 20ml of 40% Acrylamide/Bisacrylamide; 50 g of Urea, 10 ml 10 x TBE (1M Tris, 1M Boric Acid, 0.02M EDTA (disodium salt)) and fill up to 100 ml with MilliQ H_2O. For the 6%

polyacrylamide-urea gel 15ml of 40% Acrylamide/Bisacrylamide was used. 100 µl of TEMED and 200 µl of 25% APS were added to the 100 ml gel solution and the gels were poured to a thickness of 0.4mm. 6% polyacrylamide-urea gels were run at 20 V and 8% polyacrylamide-urea gels were run at 40V.

2.13 Immunoprecipitation

40 µl of Dynabeads protein G per condition were pre-blocked overnight with 1 ml 0.5% BSA in PBS. The next day, the beads were washed 3 x with IP Buffer (25mM Tris-HCl pH8.0, 150mM NaCl, 0.5% NP-40, 10% Glycerol, 2.5mM $MgCl_2$) and 2 µg of the antibody against the protein of interest was added to 600 µl IP Buffer + pI and kept rotating at 4°C for at least 2 hours. 1×10^7 cells were treated either with DMSO or NM and washed twice and scrapped in 8 ml cold PBS + cOmplete EDTA-free Protease Inhibitor Cocktail (Roche). The cells were centrifuged at 500g for 5 min at 4°C. The pellet was resuspended in 800µl Lysis Buffer (25mM Tris-HCl pH8.0, 150mM NaCl, 0.5% NP-40, 10% Glycerol, 2.5mM $MgCl_2$ + cOmplete EDTA-free Protease Inhibitor Cocktail (Roche) + PhosSTOP (Sigma)) with SIGMA Benzonase (1 µl/5 ml buffer). The sample was incubated rotating at 4°C for 30 minutes. The cell extract was then centrifuged for 15 minutes at 13000 x *g* at 4°C. 800µl of Dilution Buffer (150mM NaCl, 10% Glycerol, 2.5mM MgCl2 + cOmplete EDTA-free Protease Inhibitor Cocktail (Roche) + PhosSTOP (Sigma)) were added to the supernatant and the protein concentration on the samples were determined by Bradford. The beads conjugated with the antibody of interest were washed 3 times with IP Buffer + cOmplete EDTA-free Protease Inhibitor Cocktail (Roche) and 1 mg of the sample was added. The samples were incubated at 16 rpm on rotating wheel in cold room for 2 hours. The beads were then washed 3 times with IP Buffer + cOmplete EDTA-free

Protease Inhibitor Cocktail (Roche) and another 3 times with IP Buffer + cOmplete EDTA-free Protease Inhibitor Cocktail (Roche) without NP-40. After last wash the beads were quickly spun down to completely remove any remaining IP Buffer. The elution was done twice in 60µl of Elution Buffer (0.4% SDS, 0.1% Tween 20, 50mM Tris pH 8) shaking at 1000 x *g* for 15 minutes at 25°C.

Table 2.1. Sequence of primers for dCas9 cloning and sequencing, U2 guides and U2 snRNA and KPNB1 qPCR primers.

Name	Sequence of Forward Primer	Sequence of Reverse Primer
dCas9 EcoVR	GAATGATATCGTGCTGACCCTGACAC	GTCCACATCGTAGTCGGACAG
dCas9 FseI	GCCATCGTGCCTCAGAGCTTTC	GAACGGCCGGCCTTTTTCGTGG
dCas9 ExF Seq	GGGCACATACCACGATCTGCTG	
dCas9 Mut Seq	GGCAAGAGCGACAACGTGCCCTC	
dCas9 ExR Seq	GGATTCTCCTCGACGTCACCGC	
DSE Guide	CACCAAATGAAAGCCCGGGAACGC	AAACGCGTTCCCGGGCTTTCATTT
PSE Guide	CACCAAGGCGAGCGCATCGCTTCT	AAACAGAAGCGATGCGCTCGCCTT
U2 N	GGAGCGGAGCGTTCTCTGTCTCCCC	AGAGTGTGAGCCCTCATTCACGCCC
U2 DSE	TGGCTCGATACGAACAAGGAAG	GTTCCCGGGCTTTCATTTCG
U2 UpPSE	GGGAACGCCGAAGAAGCACGGG	CCCCAGCCTCGCTCCTTGCCC
U2 PSE	ATGAGAGTGGGACGGTGA	CACTTGATCTTAGCCAAAAGG
U2 3'Box	ACGAGTCCTGTGACGCGCCGGCTTG	CTCCGGGTGGGTCCCATTCCTTTAA
U2 CTCF	CCTCCCCGCCTCTCCCTCGCTC	GGACAAATAGCCAACGCATGCGG
KPNB1 TSS	AGGGAAGGGAAGTTAGGTTGCG	GAGCTTCCACATGGCCCCTA
KPNB1 1b	GCAAAGCAGGATACTAAGTGATCCGA	GAAGCCACAATTGACCTAGAGC
KPNB1 C	TGCAAGAGCCAGTGGGAACACTT	CCTCTACTCAGCAATGATACTTC
KPNB1 Int	CTGAGGAAACTGAAGAACCAAG	GAAGGCAGTGCTTGCCAGAAT
KPNB1 K	GAGGAGTGTGCACGGATGCTGAA	CCAAGATGGCCGATGTTATGG
KPNB1 G	TAGTTACCGTCTGCTTGGGAAGATG	CCTCTGACAGCAAGTCCAACATT
KPNB1 E	GACTCATCACACCAAGGTCAC	GATAGTGCTGGGAAGGAAATGG
KPNB1 L	GTACATCTCAGCTTTGGCATATG	GCCCAGAACATAGCAGGCATTGC
KPNB1 M	GTTTCACCGTGTTAGCCAGGATGG	CCACAGCCATGTTCATTTCTGC

Table 2.2. List of antibodies used either for ChIP or Western Blot.

Antibody	Reference
Monoclonal ANTI-FLAG® M2	Sigma-Aldrich F1804
RNA Polymerase II/POLR2A	Novus Biologicals NBP2-32080
AffiniPure Donkey Anti-Rabbit IgG (H+L)	Jackson ImmunoResearch 711-005-152
PTFy	Courtesy of Yoon JB and Roeder RG (181)
Actin Antibody (C-11)	Santa Cruz Biotechnology sc-1615
Anti-TRF1 antibody [TRF-78]	Abcam ab10579
Sp1 (1C6)	Santa Cruz Biotechnology sc-420
TFIID Antibody (58C9)	Santa Cruz Biotechnology sc-421
Streptavidin (HRP)	Abcam ab7403
Anti-CRISPR-Cas9	Abcam ab191468
Anti-Biotin	Abcam ab53494
RPAP2	Proteintech 17401-1-AP
IntS9	Proteintech 11657-1-AP
IntS11	Novus Biologicals NB100-60638

Chapter 3

dCas9 Pulldown Approaches

3.1 Mutating Cas9 in pX462 and assessing the efficiency and specificity of guides

At the time the project started there were no plasmids commercially available for inserting the guides into a vector encoding dead (d)Cas9. Instead of using one plasmid to express dCas9 and another plasmid to express the guides, I inserted a second mutation into the Cas9 in the pX462 plasmid which also has site for inserting the guides, the mutated pX462 could then express both dCas9 and a guide.

3.1.1 Insertion of the H840A mutation into pX462

The nickase Cas9 already has the D10A mutation and in order to completely loose Cas9 activity, the H840A mutation was also inserted. Two sets of primers were used to amplify the region containing the mutation to be inserted (shown in yellow and blue in Fig. 3.1) and cloning was carried out as detailed in Section 2.2.

3.1.2 Design and assessment of the guide (g)RNAs targeting the U2 snRNA gene

Initially, 13 singled and paired guides were designed and tested. Paired guides should result in more dCas9 on the U2 snRNA gene, which could make it easier to pulldown dCas9. However, recruiting more dCas9 could interfere with factors binding to the DNA. Single guides were therefore also tested.

Guides targeting DSE, PSE and the 3'box were cloned into pX462 with the H840A mutation. The plasmids were transiently transfected into HEK293 cells, and to assess if dCas9 was being targeted to the region of interest, I performed ChIP using an antibody against FLAG on the U2 snRNA gene once the dCas9 used is FLAG tagged.

As is noted in the scheme below (Figure 3.1): G3, G4 and G5 target the DSE; G1, G2 and G8 target PSE; G10, G11, G12 and G13 target the 3'box. G4 most efficiently targets dCas9 to the DSE region. G8 most efficiently targets dCas9 to the PSE. G13 gives the highest level of dCas9 on the 3' box. Interestingly, G12 targets dCas9 to both the 3' box and the DSE, suggesting that there is a link between these two regions (Fig. 3.2).

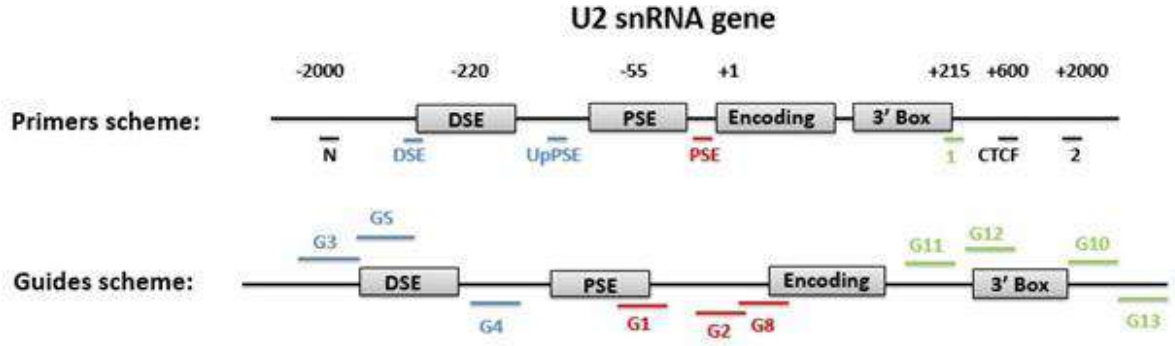


Figure 3.1. Scheme of the U2 snRNA gene and the gRNAs and qPCR primers used.

On the scheme above it is shown where the qPCR primers used to amplify the ChIP products are mapped to the U2 snRNA gene. On the scheme below it is noted where the gRNA designed target the U2 snRNA gene. The qRNA and the qPCR primers are colour coded to the regions they target.

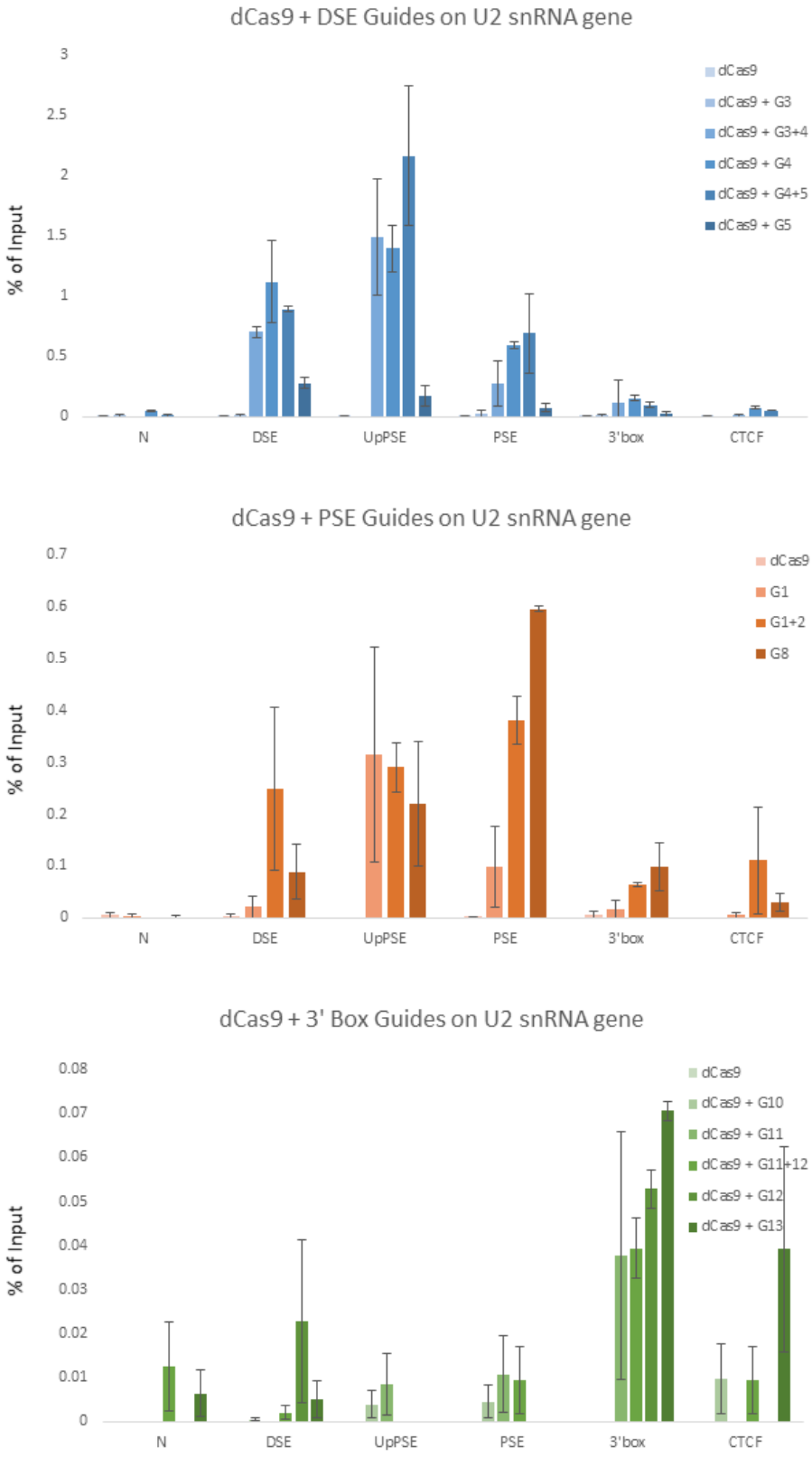


Figure 3.2. dCas9 ChIPs on U2 snRNA gene upon qRNA transient transfection.

The three graphs show dCas9 occupancy on the U2 snRNA gene upon transient transfection of the gRNAs. The top graph shows the DSE guides, middle one shows the PSE guides and the bottom one shows the 3'box guides. When pol II signal was below IgG signal it was considered as background noise and therefore no bars are shown.

ChIP-qPCR across the U2 snRNA gene shows clearly that dCas9 is not targeted to the different genomic regions with the same efficiency. This may be due to the transcriptional complex getting bigger towards the 3'end of the U2 snRNA gene and blocking dCas9 access. Based on these data, guides G4 and G8 were chosen for further experiments targeting the DSE and PSE regions of the u2 snRNA gene.

In order to assess the specificity of the guides, the U1 PSE region was also analysed by qPCR when targeting the U2 PSE region. Although the PSE region is very conserved between snRNA genes, G8 is specific to the U2 PSE snRNA gene region as no dCas9 can be detected on the U1 PSE. Additionally, G8 was blasted against the whole genome and the top two potential off-target regions were analysed by qPCR. The identity between the guide and the off-target regions were 15 and 16 on the chromosomes 6 and 7, respectively, and the qPCR results show that dCas9 is not present on that region, indicating that the guides are specific to the regions designed (Fig. S3.1; See page 107). Guide 4 off targets identities were lower than for Guide 8, so therefore guide 4 was considered to be specific as well.

3.2 Stable cell lines expressing dCas9-FLAG + guide RNA

Once the guides were chosen, stable cell (SC) lines were generated for dCas9 only, dCas9 + Guide 4 (dCas9 DSE stable Cell Line). The vector was linearized and transfected into HEK293 cells and the cells were selected and kept in culture with puromycin treatment. dCas9 expression was confirmed by western blot (Fig, 3.3A) and whether dCas9 was targeted to the U2 snRNA gene was checked by ChIP followed by qPCR (Fig. 3.3B). I also took into account the fact that pol II could be affected by dCas9 recruitment,

and therefore performed ChIP experiments using an antibody against pol II to analyse whether transcription was affected (Fig. 3.3C). ChIP-qPCR analysis showed that dCas9 targets the U2 snRNA gene in the stable cell lines, and the pol II profile is the same for the dCas9 stable cell lines leading to the conclusion that transcription is not affected by the presence of dCas9.

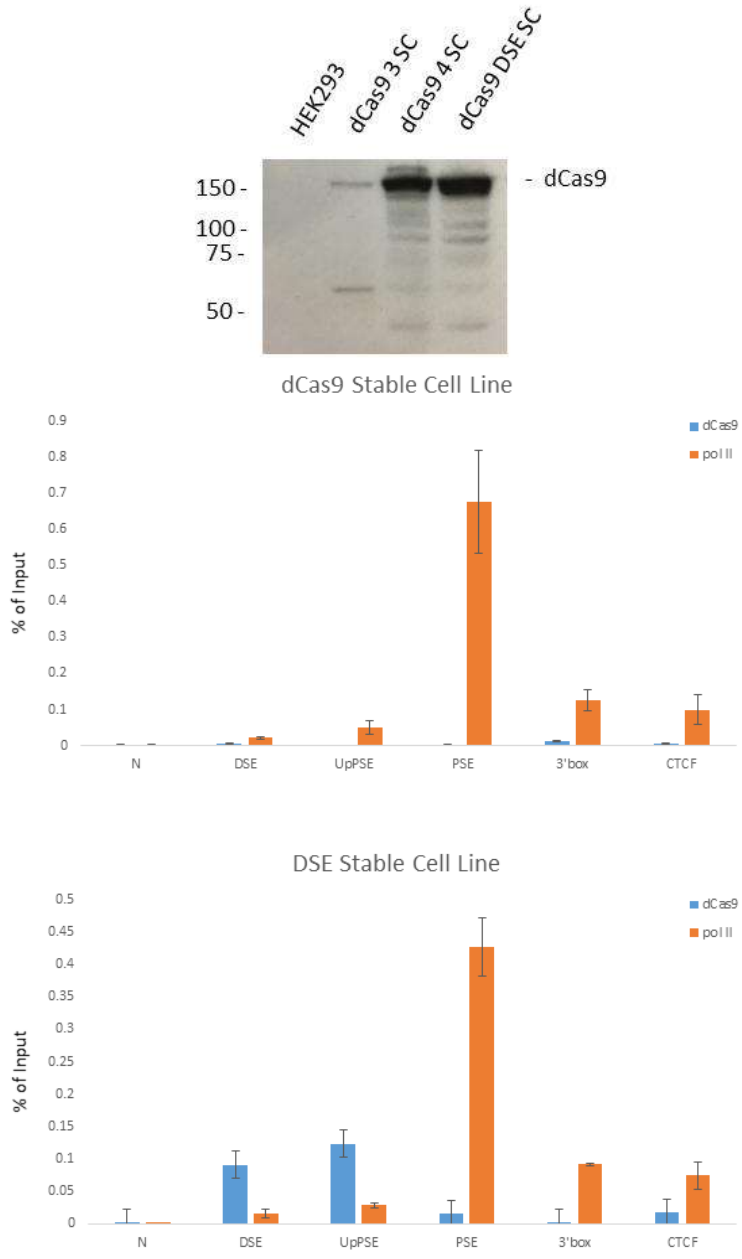


Figure 3.3. Expression of dCas9 in stable cell lines and U2 snRNA gene targeting.

Above, screening of dCas9 stable cell lines. dCas9 4 SC and dCas9 DSE SC express dCas9 and were chosen for further experiments. The graphs show dCas9 and pol II occupancy on the U2 snRNA gene for both stable cell lines.

3.3 enChIP on the PSE of the U2 snRNA gene

The first attempt to pulldown dCas9 used enChIP (172). The original paper targets dCas9 to the promoter region of a protein-coding gene. However, the paper does not show that dCas9 is being pulled down or if there is an any enrichment of the proteins in the transcriptional region dCas9 is being targeted to as the experiment does not have a negative control such as dCas9 without a guide (172). In fact, the mass spectrometry list is full of likely contaminants, including actin and tubulin, and no transcription factors were pulled down, suggesting that the technique does not work not specifically and requires improvement.

Nonetheless, I tested enChIP targeting dCas9 to the PSE region of the U2 snRNA gene. The plasmid with the guide targeting the PSE, G8, was transiently transfected into HEK293 cells. dCas9 alone was used as a negative control for the dCas9 and enrichment of expected proteins, and pcDNA6 was used as a negative control for the dCas9 transfection. Western blot analysis indicates that dCas9 was pulled down (Fig. 3.4). However, there was no difference in the protein profile observed by silver staining between dCas9 with and without a guide (Fig 3.4). enChIP also failed to pull down factors known to bind to the PSE region, as PTFy was not detected (data not shown).

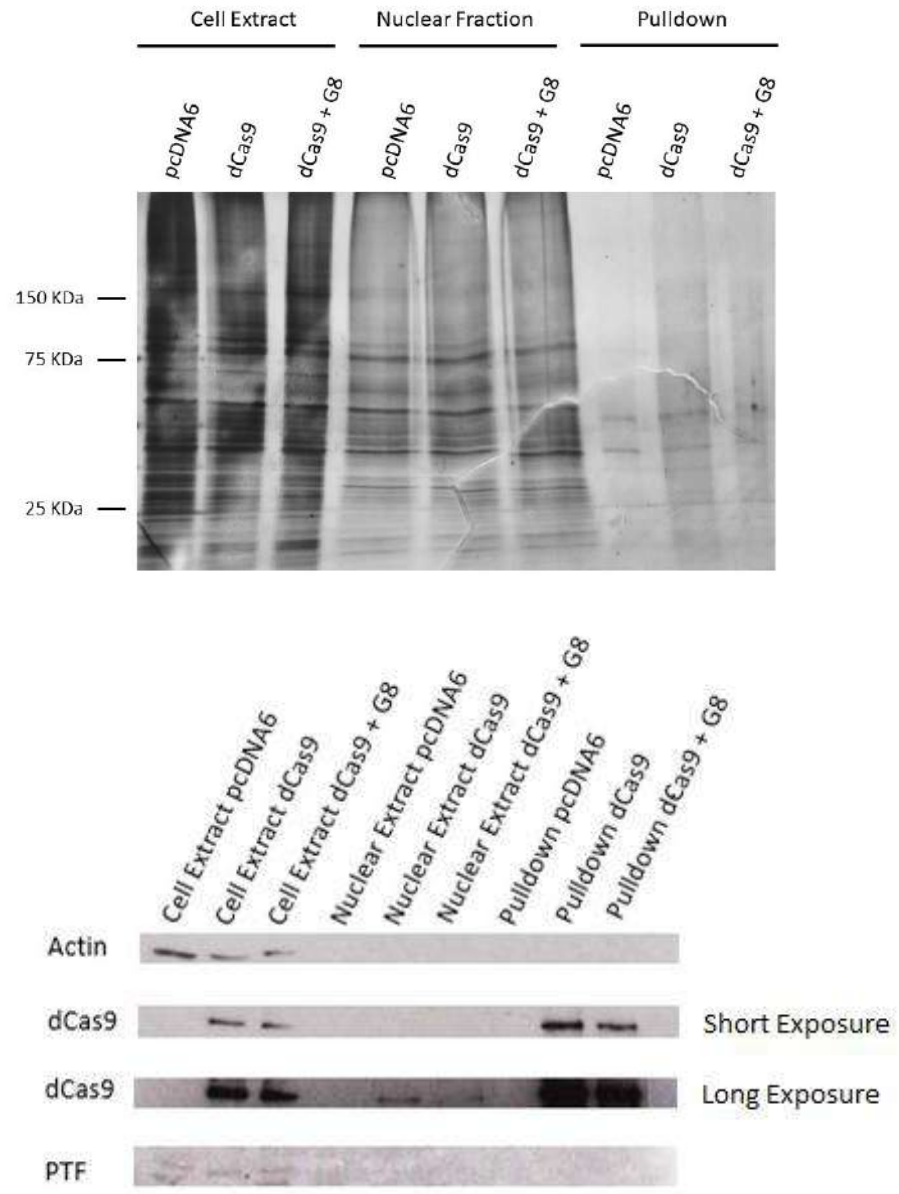


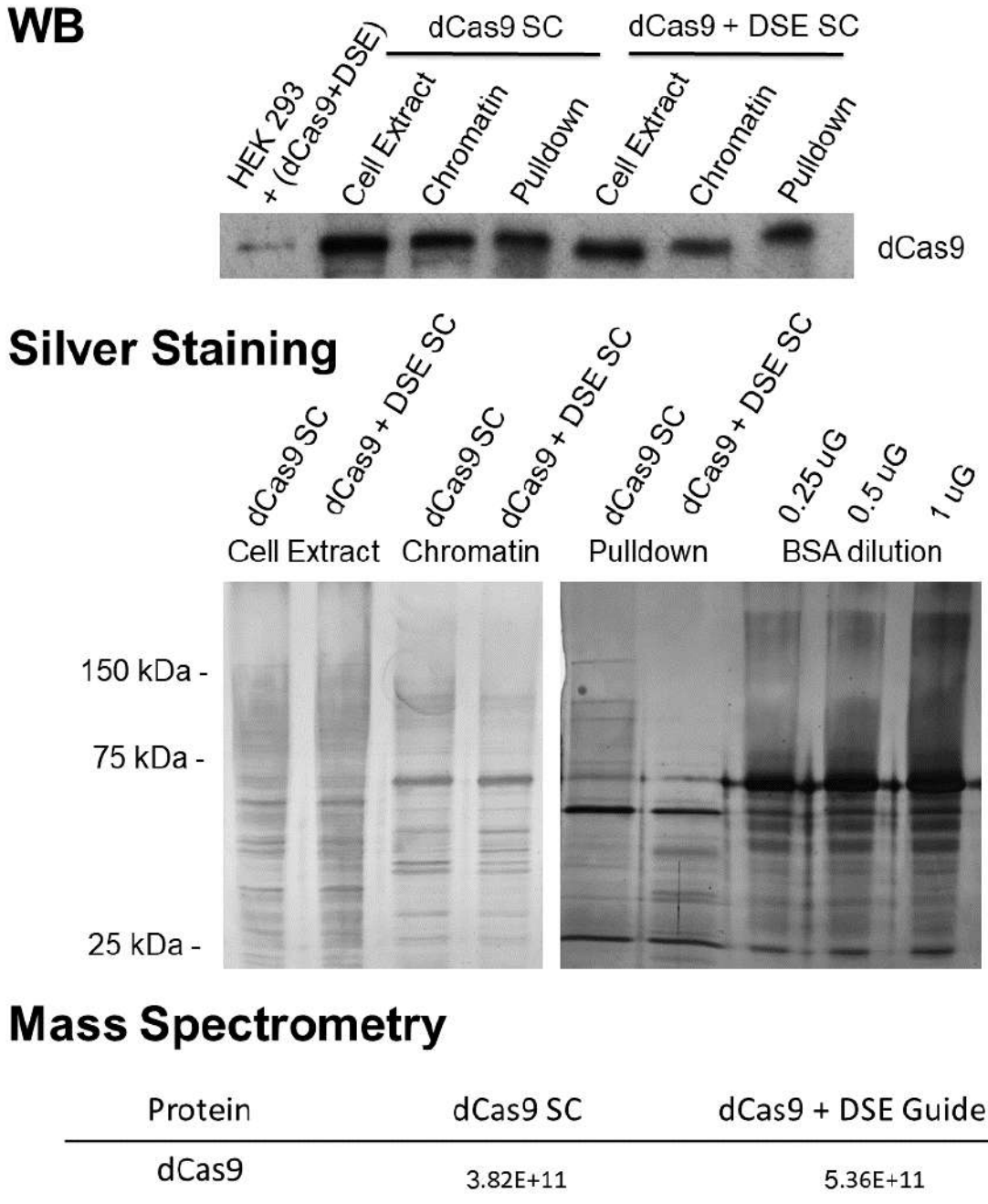
Figure 3.4. Silver staining and western blot of enChIP on the U2 snRNA gene PSE.

Top, silver staining of the cell extract, nuclear fraction and enChIP pulldown. Bottom, western blot of actin as a negative control, dCas9 and PTF as positive controls.

3.4 U2 snRNA gene DSE dCas9 pulldown from the chromatin fraction

In an attempt to improve the dCas9 pulldown, the chromatin fraction was used instead of total cell extract that was used in enChIP. Chromatin was prepared according to the mNetseq protocol, which uses a sucrose gradient to separate nuclei followed by fragmentation of the chromatin by MNase digestion (182). As formaldehyde crosslinking affects the production of the chromatin fraction (data not shown), chromatin preparation was carried out in native conditions without crosslinking. The MNase digestion was checked by agarose gel electrophoresis (Fig. S3.2; See page 108). The dCas9 pulldown was then performed using the chromatin fraction and the eluates from the beads were checked both by western blot and silver staining (Fig. 3.5).

The protein profile for dCas9 with or without guide was very similar, although some subtle changes are observed after silver staining (Fig. 3.5). These samples were submitted for mass spectrometry analysis. However, the only protein identified in the samples was dCas9. The failure to identify any other protein may be due to lack of crosslinking or that the MNase digestion produced DNA fragments that are too short to bind dCas9 and other proteins.



Mass Spectrometry

Protein	dCas9 SC	dCas9 + DSE Guide SC
dCas9	3.82E+11	5.36E+11

Figure 3.5. Western blot, silver staining and mass spectrometry of the dCas9 Pulldown from chromatin fraction.

Top, western blot of the cell extract, chromatin and pulldown of the dCas9 and dCas9 targeting DSE stable cell lines. HEK293 + dCas9 has been used as positive control. Middle, silver staining of the cell extract, chromatin, pulldown and BSA dilution to quantify the protein bands. Bottom, the mass spectrometry results for the pulldown.

3.5 CAPTURE of Telomeres and the U2 snRNA gene DSE and PSE

A new CAPTURE (CRISPR affinity purification in situ of regulatory elements) technique to identify proteins associated with genomic regions was published in late 2017 (174). CAPTURE targets an *in vivo* biotinylated dCas9 to specific regions on a given gene and its pulldown also brings together protein present in the region allowing identification of the associated protein profile on the region of interest in an unbiased way. I therefore attempted carry out a similar pulldown using the dCas9 stable cell lines already generated and using FLAG antibodies conjugated to dynabeads. CAPTURE was initially performed on telomeres as a proof of concept. Initially, I assessed dCas9 targeting by CAPTURE-qPCR (Fig. 3.6) and a further pulldown showed that dCas9 was being pulled down and also TRF1, a telomeric protein known to bind to telomeres (183).

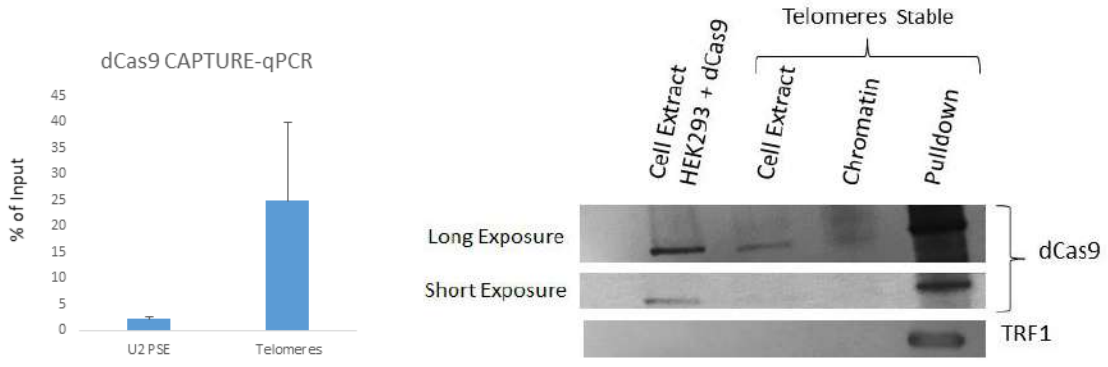


Figure 3.6. CAPTURE on telomeres.

Left, CAPTURE-qPCR of dCas9 targeting telomeres. The U2 PSE has been used as a negative control. Right, western blot of CAPTURE performed on telomeres against dCas9 and TRF1. HEK293 has been used a negative control.

As the CAPTURE technique seemed to be working to target telomeres, it was then performed on the U2 snRNA gene. dCas9 was targeted to the DSE and PSE. The western blots performed on the eluates of the pulldown and the mass spectrometry list indicate that the dCas9 pulldown has worked and has been successfully pulled down expected factors for the regions of interest; Sp1 for the DSE CAPTURE and TBP for the PSE CAPTURE. The unbound fraction of the beads was also loaded on the western of the PSE CAPTURE to check if there was any dCas9 in chromatin fraction that was not being pulled down by the beads (Fig. 3.7). The U2 snRNA PSE pulldown using FLAG was submitted to mass spectrometry analysis and the table of proteins identified on this region can be found in Section 4.1.

Considering that streptavidin and biotin interaction is stronger than antibody and protein interaction, stable cell lines expressing dCas9 with the biotin acceptor peptide (BAP) and the BirA were also generated. Western blot confirmed that dCas9 is expressed and biotinylated *in vivo* and ChIPs have confirmed that dCas9 is present on the U2 snRNA genes upon transient transfection of guides to the region of interest and that dCas9 is also biotinylated when associated with the region of interest. The presence of dCas9 on the U2 PSE does not impair pol II recruitment or levels showing a normal pol II profile for the U2 snRNA gene as seen by the pol II ChIPs on untransfected and transfected cells (Fig. 3.8).

CAPTURE on the DSE of the U2 snRNA gene shows that dCas9 is pulled down and also Sp1, as expected (Figure 3.7). When dCas9 is targeted to the PSE of the U2 snRNA gene, dCas9 is also pulled down and both dCas9 and PTFy are enriched 1.5 and 1.6x,

respectively, over the negative control dCas9 without the guide (Fig. 3.9). No enrichment could be observed for pol II or TBP.

Mass spectrometry analysis of the CAPTURE pulldown showed that expected factors are enriched. The CAPTURE mass spectrometry list for the DSE and PSE region of the U2 snRNA gene using streptavidin and biotin can be found in Chapter 4 (Section 4.2).

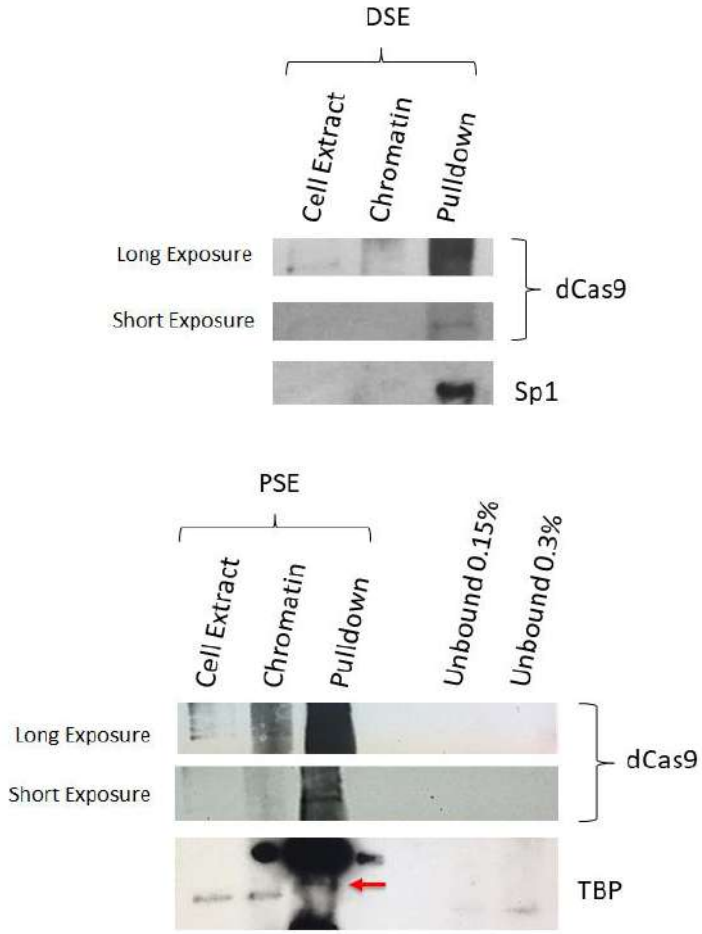


Figure 3.7. CAPTURE on the DSE and PSE of the U2 snRNA gene.

Top, western blot of the CAPTURE for the DSE region of the U2 snRNA gene. Cell Extract, chromatin fraction and pulldown were probed with dCas9 and Sp1 antibodies. Bottom, western blot of the CAPTURE for the PSE region of the U2 snRNA gene. Cell extract, chromatin, pulldown and unbound fractions were probed with dCas9 and TBP antibodies. The elution was done by boiling the beads which released the light and heavy chain of the antibody conjugated to it and therefor seen in the TBP membrane.

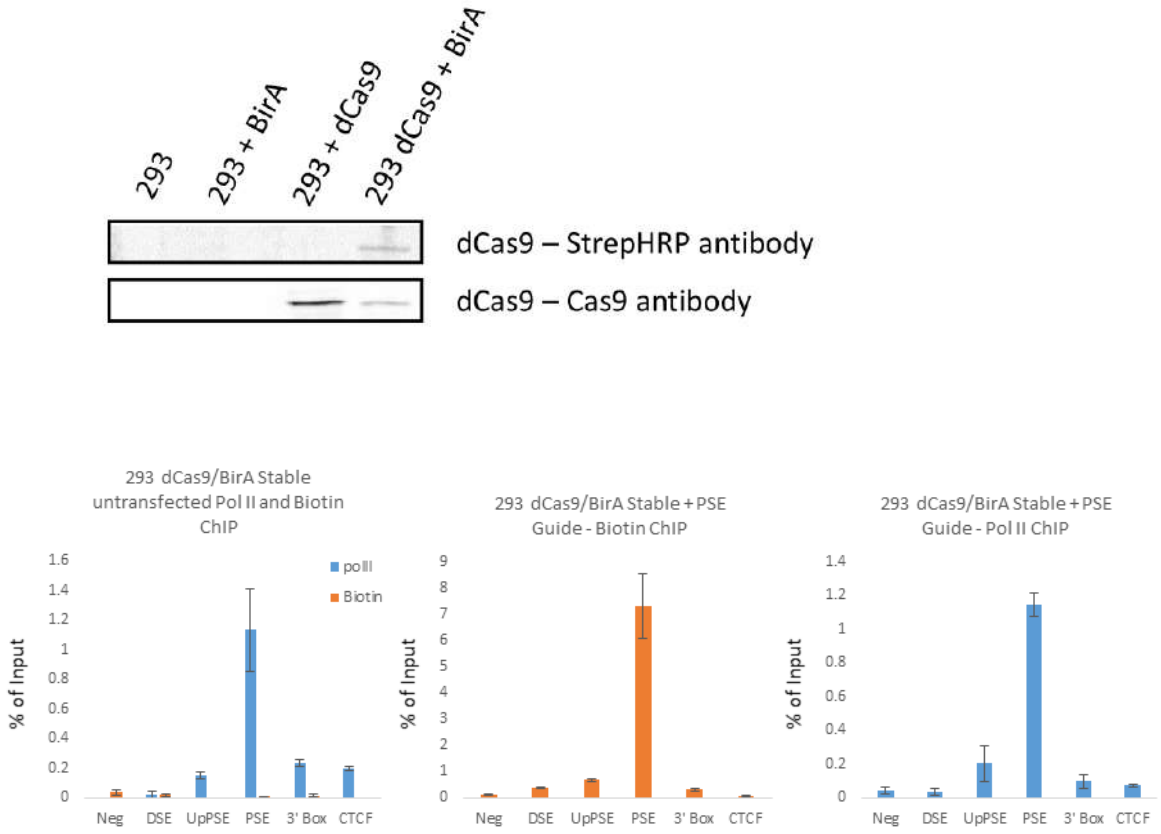


Figure 3.8. Characterisation of a stable cell line expressing dCas9 and BirA.

Top, western blot of HEK293, HEK293+BirA, HEK293+dCas9BAP and the stable HEK293 cell line expressing BirA and dCas9BAP. dCas9BAP is expressed and biotinylated in the stables cell lines generated. Bottom, ChIP using antibodies against biotin and pol II followed by qPCR on the U2 snRNA gene in the dCas9BAP stable cell lines untransfected or upon PSE guide transfection. ChIP-qPCR from untransfected cells are shown on the same graph, the transfected ones are shown in different graphs for better visualization.

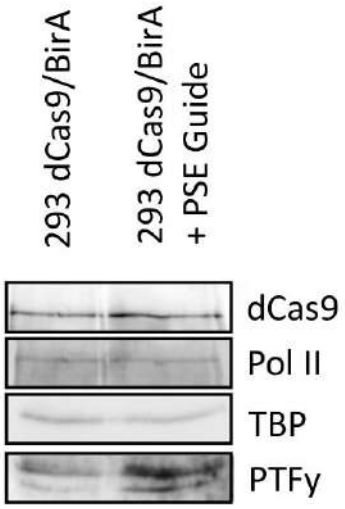


Figure 3.9. Western blot of CAPTURE output of the U2 snRNA gene PSE.

U2 snRNA gene PSE CAPTURE pulldown eluates were probed with dCas9, pol II, TBP and PTFy antibodies. HEK293 dCas9BAP/BirA stable cell lines were transiently transfected with the PSE guide and untransfected cells were used as negative control.

3.6 Conclusions

Guide RNAs were designed to specifically target dCas9 to the U2 snRNA gene. Additionally, dCas9 did not affect pol II profile on the region targeted. After optimization, CAPTURE has been shown to pulldown other proteins also present in the region of interest: telomeres, DSE or PSE. CAPTURE seemed as a good technique to identify proteins in an unbiased way and it was therefore performed in order to identify proteins that regulate the expression of the U2 snRNA gene in human cells.

Chapter 4

U2 snRNA gene CAPTURE strategy

4.1 U2 snRNA gene PSE and DSE FLAG CAPTURE

Mass spectrometry from the U2 snRNA gene PSE FLAG CAPTURE initially identified 35 proteins: 19 proteins were specific for the U2 snRNA gene and 16 proteins were enriched by at least 2.5 times the dCas9 only negative control (Table S4.1; See page 109). A summary of the factors that might be relevant for gene expression can be seen in Fig. 4.1. The mass spectrometry list does not contain any transcription factors that would be expected to be associated with a promoter region; not even TBP, that was detected by western blot (Section 3.5). The only factor involved in gene expression that was identified was the RNA processing factor Cstf3, by 2.5-fold. The zinc finger protein ZNF512B was a unique hit in the PSE region and it could be a transcription factor for the U2 snRNA gene. Additionally, the RNA binding protein Lin28B was in the mass spectrometry list, which might suggest that not only factors that interact with DNA are being pulled down, but also factors that interact with the U2 snRNA transcript. Lin28B has not yet been shown to interact with snRNA genes, but bioinformatic predictions by catRAPID shows that Lin28B is predicted to interact with the U2 snRNA as it comes as first on the list of predicted

Lin28B interactors (Fig.S4.1; See page 108). Overall, the FLAG CAPTURE failed to identify any transcription factors that are known to be involved in snRNA gene regulation.

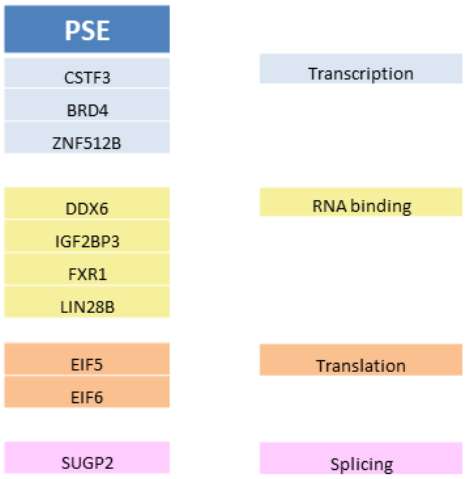


Figure 4.1. U2 snRNA gene PSE FLAG CAPTURE mass spectrometry.

List of proteins involved in gene expression and pulled down by CAPTURE using FLAG on the PSE of the U2 snRNA gene. Cut off: 2.5-fold enrichment. See Table S4.1 in page 109 for the full list of proteins.

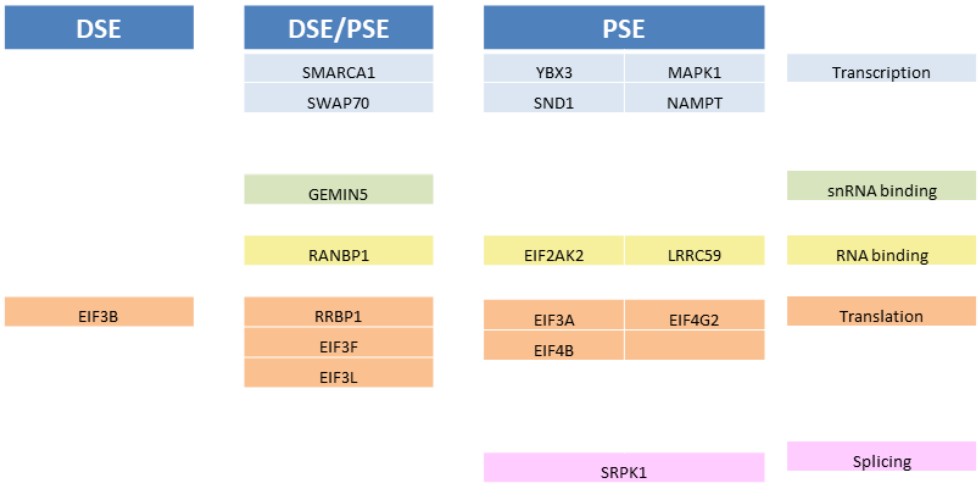


Figure 4.2 U2 snRNA gene DSE and PSE FLAG CAPTURE mass spectrometry.

List of proteins involved in gene expression and pulled down by CAPTURE using FLAG on the DSE and PSE of the U2 snRNA gene. On the left, factors present only on the DSE CAPTURE, on the right, factors present only on the PSE CAPTURE, on the middle column factors present both in DSE and PSE. Cut off: 2.5-fold enrichment.

I repeated FLAG CAPTURE on the PSE and also DSE regions and, although more factors were identified, it also failed to identify factors known to be involved in snRNA gene transcription (Fig. 4.2).

4.2 U2 snRNA gene PSE and DSE Streptavidin-Biotin CAPTURE

After extensive optimisation of the CAPTURE pulldown protocol, I found that using streptavidin-biotin (as detailed in the original CAPTURE paper (174)) to pulldown dCas9 was much more effective than FLAG. The final protocol utilized is shown in the scheme in Fig. 4.3. The gRNAs were transiently transfected in the cells stably expressing dCas9 and BirA for 24h. The cells were then crosslinked with 2% formaldehyde, harvested, pelleted and the chromatin was extracted and subjected to several urea washes in order to enrich for chromatin bound factors. The chromatin was then sonicated and the sheared DNA was incubated with streptavidin beads overnight. The following day the beads were washed with 150mM NaCl several times to remove non-specific proteins. The proteins were then de-crosslinked on beads and submitted to the mass spectrometry facility (Advanced Proteomics Facility, Department of Biochemistry, Oxford). The samples were digested with trypsin and run on an Ultimate 3000 RSLCnano HPLC coupled to a QExactive Orbitrap mass spectrometer.

The first mass spectrometry list obtained with streptavidin-biotin CAPTURE targeting the PSE region has 83 proteins, 44 of which are unique to the PSE region and 39 proteins enriched on the PSE region over the dCas9 negative control without a guide (Table S4.2). The cut off used for the enrichment was 4-fold the ratio of dCas9 + PSE guide over dCas9 only. The mass spectrometry failed to identify several expected factors such

as pol II, general transcription factors, TAFs, CDKs or Integrator. However, it did identify Med20, one of the small subunits of Mediator. Some transcription factors that have not yet been described to be associated with the U2 snRNA genes such as BTF3, NR2F2 and the associated transcription factor BCLAF1 were also identified. Interestingly, the cleavage and polydenylation factor CPSF6 was also found to be associated with the U2 snRNA gene. Among the proteins also found to be there were the serine threonine kinases PRPF4B and SLK.

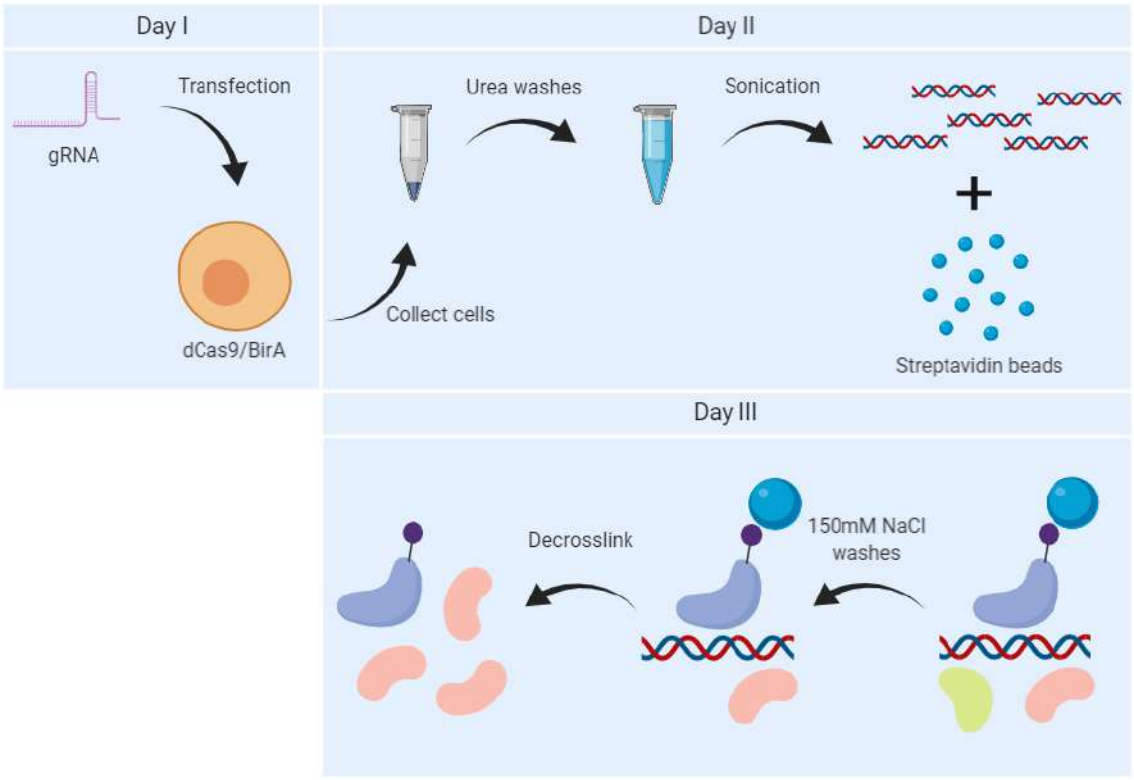


Figure 4.3 Scheme of the Streptavidin-Biotin CAPTURE.

Day I: The gRNA is transiently transfected into cells stably expressing dCas9/BirA Day II: The cells are crosslinked with 2% formaldehyde, harvested, chromatin is extracted and subjected to several 8M urea washes. The chromatin is sonicated and incubated with streptavidin beads. Day III: The beads are washed with 150mM NaCl and proteins are de-crosslinked. The samples are then subjected to mass spectrometry.

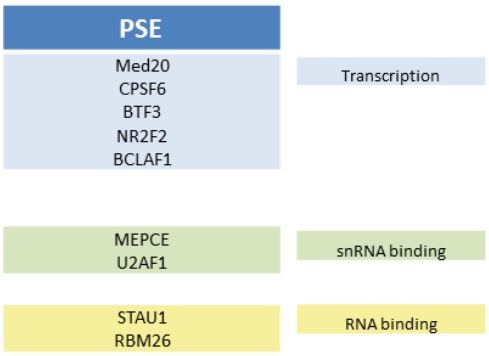


Figure 4.4. U2 snRNA gene PSE Streptavidin-biotin CAPTURE mass spectrometry.

List of proteins involved in gene expression and pulled down by CAPTURE using streptavidin-biotin on the PSE of the U2 snRNA gene. Cut off: 4-fold enrichment. See Table S4.2 in page 109 for the full list of proteins.

After further optimization of the ratio between the chromatin input and the amount of beads used for pulldown, the second mass spectrometry list obtained for the PSE region using streptavidin-biotin CAPTURE has 404 proteins when using the same cut off of 4-fold enrichment over dCas9 only, 301 of these are only found when the guide to the PSE region is co-expressed and 103 proteins enriched are enriched more than 4-fold (Table S4.3). In this experiment, expected factors such as Integrator subunits Int1, Int2, Int3, Int4, Int7, Int11, Med16, Med23, pol II subunit RPB2, Spt5, CDK7, NELFA, NELFB, CDK9 were pulled-down. Cyclin K, the binding partner of CDK12 and CDK13, was also found associated with the U2 snRNA genes, suggesting that those kinases might be involved in regulating transcription of this gene. Additionally, the nuclear capping binding protein NCBP1 was also found to be on the U2 snRNA genes, which is expected as capping occurs co-transcriptionally (101, 102).

Interestingly, the protein-coding gene factors involved in termination and RNA 3' end formation Cstf1, Cstf2, Cstf3, CPSF1 and XRN2 were also found to be present on the U2 snRNA gene. In addition, Spt6 was found associated with the U2 snRNA gene for the first

time, Spt6 is involved in elongation of transcription of protein coding genes (184) and may be responsible for the recruitment of Integrator to long non-coding genes, which allows proper transcription termination to occur (185).

The splicing factors SF1, SF3A1, SF3B2, SF3B4, and SF3B6 were also surprisingly associated the U2 snRNA gene. As U snRNA genes are intronless, these factors might have a splicing-independent role in the regulation of expression of the U2 snRNA gene.

One of the most interesting findings from the final mass spectrometry list is the enrichment of SMARC factors. SMARC factors are involved in chromatin remodelling and have been described as putative transcriptional factors (186). SMARCA2/4, SMARCA5, SMARCC1 and SMARCE1 were found associated with the U2 snRNA gene and could be involved in initiation of transcription. Another transcription factor that was found to be present on U2 snRNA genes was TCEA1/2 (or TFIIS). TCEA1/2 has been shown to facilitate pol II transcription pausing and is proposed to induce pol II to cleave the 3'end of the nascent transcript (187-189).

I also performed a DSE streptavidin-biotin CAPTURE in parallel with the PSE streptavidin-biotin CAPTURE. The DSE mass spectrometry list includes 255 proteins that were enriched when the DSE guide was co-expressed (Table S4.4). Among these are Int2, Int3, Int11, Med23, pol II subunit RPB2, which would be expected. A novel factor that was specifically enriched for the DSE region was the TFIIH XPD helicase subunit.

The SMARC proteins SMARCA2/4, SMARCA5 and SMARCC1 are also present in the DSE region, reinforcing the idea that this complex might be involved in the regulation of transcription of snRNA genes.

The streptavidin-biotin CAPTURE therefore appears to be much more efficient and reliable in identifying proteins associated with the U2 snRNA gene as can be seen in the summary table in Fig. 4.5. The table merges proteins identified as associated with the PSE and DSE region that are involved in gene expression, including expected factors.

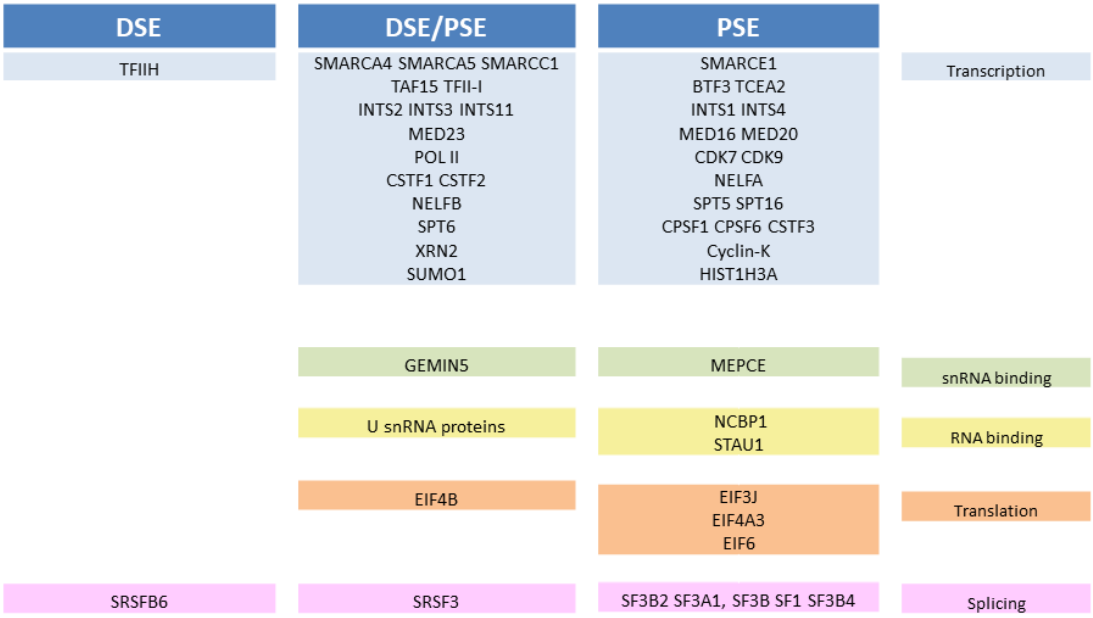


Figure 4.5. U2 snRNA gene DSE and PSE Streptavidin-biotin CAPTURE mass spectrometry.

List of proteins involved in gene expression and pulled down by CAPTURE using streptavidin-biotin on the DSE and PSE of the U2 snRNA gene. On the left, factors present only on the DSE CAPTURE, on the right, factors present only on the PSE CAPTURE, in the middle column, factors present both in the DSE and PSE. Cut off: 4-fold enrichment. See Table S4.3 in pages 111 to 115 and Table S4.4 in pages 116-118 for the full list of proteins.

4.3 Conclusions

Streptavidin-biotin CAPTURE was shown to be a powerful tool in identifying factors involved in gene expression of the U2 snRNA gene. Several proteins known to regulate snRNA gene expression were identified and CAPTURE also revealed new candidates that need to be further confirmed and assessed.

Chapter 5

Roles of transcriptional CDKs in snRNA gene expression

5.1 The effect of inhibiting analog-sensitive (as) CDKs on expression of the U2 snRNA gene.

Several cells lines where the endogenous CDKs are mutated to become analog-sensitive have been made in the Murphy lab. These include analog-sensitive versions of CDK7, 8, 9, 10, 12, 13 and double 12/13. To assess whether inhibition of any individual CDK could affect pol II levels on the U2 snRNA gene, I performed ChIP experiments with the addition of control DMSO or the bulky kinase inhibitors NM-PP1 or NA-PP1 (Fig. 5.1). ChIP-qPCR showed that inhibition of any CDK individually does not affect pol II levels on the snRNA gene (Fig 5.1). In order to assess redundancy of the CDKs, I also analysed the effect of inhibiting CDK12 and CDK13 at the same time (CDK12as/13as). When CDK12 and 13 were co-inhibited, an increase of pol II on the PSE was observed by ChIP-qPCR (Fig. 5.2).



Figure 5.1. The effect of inhibiting analog-sensitive (as) cyclin-dependent kinases (CDKs) on expression of the U2 snRNA gene.

ChIP using an antibody against pol II followed by qPCR on the U2 snRNA gene without (blue bars) and with drug inhibition (orange bars) in the CDK 7, 8, 9, 10, 12 and 13 cell lines.

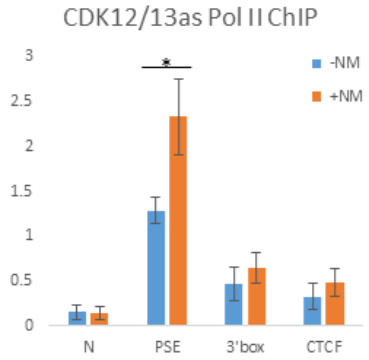


Figure 5.2. The effect of inhibiting CDK 12/13as cells on expression of the U2 snRNA gene. ChIP using an antibody against pol II followed by on the U2 snRNA gene in the CDK 12/13as without (blue bars) and with drug inhibition (orange bars). $P \leq 0.05$.

5.2 The effect of inhibiting analog-sensitive (as) cyclin-dependent kinases (CDKs) on 3' box recognition.

Inhibition of CDK9 by DRB has already been shown to affect pre-snRNA 3'end formation (180). In order to investigate whether inhibition of any of the analog-sensitive CDKs can affect pre-snRNA 3' end formation, RNase protections were performed on RNA from each analog-sensitive cell upon drug inhibition. For RNase protection, the RNA extracted from the cell is hybridised with a labelled RNA probe complementary to the region of interest and any unprotected RNA is digested by nucleases. If there is no 3’ end processing defect the expected pre-snRNA would be produced. However if there is 3’ end processing defect, the longer read through (RT) RNA is produced and the protected RNA size changes accordingly (Fig. 5.3).

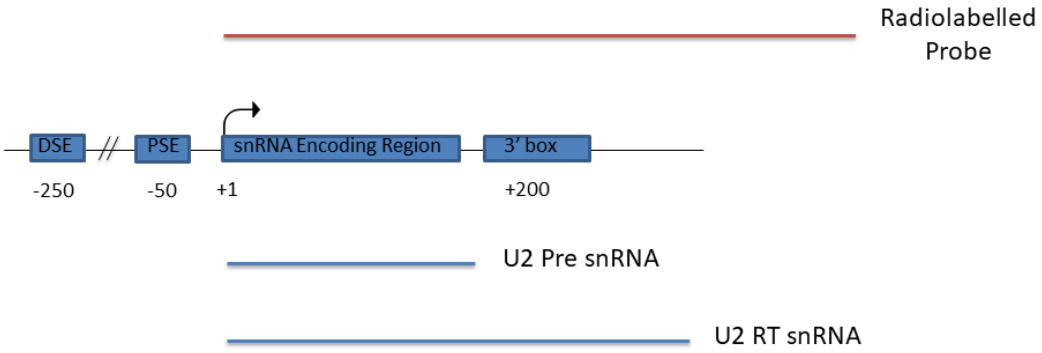


Figure 5.3. Scheme of the RNase protection.

The radiolabelled probe is hybridised to the RNA extracted from the cells. The shorter RNA fragment (U2 Pre snRNA) indicates that 3' processing is not affected. However, if there is 3' end processing defect a longer fragment is generated (U2 RT snRNA).

For every analog-sensitive cell line three treatment conditions were used: DMSO (as a negative control); 1-NM-PP1 or 1-NA-PP1, which are the ATP bulky analog drugs; and DRB which is the positive control as it is known to inhibit CDK9 and affect 3' box recognition (46, 162).

As controls for the RNase protections, HEK293 cells were treated with a range of concentrations of NA (10-20mM) and NM (7.5-12.5 mM) and no 3' end processing defect was observed at any of the concentrations tested (Fig, 5.4). CDK9as was treated with 1-NA-PP1 at 15mM, all other cell lines were treated with 7.5mM 1-NM-PP1. As expected, DRB and inhibition of CDK9as causes a 3'end processing defect (Fig. 5.5). Inhibition of CDK9 is in fact the only CDK that causes a 3'end processing defect on its own as inhibition of CDK 7, 8, 10, 12 and 13 do not affect 3' end processing (Fig. 5.5). However, when the paralogues CDK 12 and 13 were inhibited at the same time, there is 3'end processing defect. These results indicate that CDK9, CDK12 and CDK13 play roles in 3' box

recognition and that CDK12 and CDK13 are redundant for their role in this processing (Fig. 5.5).

THZ531 is an inhibitor of both CDK12 and CDK13. It has been shown by ChIP to inhibit transcriptional elongation in Jurkat cells and to decrease S2P levels (164). HEK293 cells were therefore treated with THZ531 to determine whether it also affects pre-snRNA 3'end formation. THZ531 was used at concentrations of 50nM, 500nM and 1uM. All conditions cause a 3'end processing defect, reinforcing my finding that CDK12 and CDK13 are involved in recognition of the snRNA gene 3' box (Fig. 5.6).

In addition, I carried out a time course of CDK12/13as inhibition (15 minutes, 30 minutes and 1 hour; Figure 5.7). A 3' end processing defect is apparent within 15 minutes, with a further increase in the read-through transcript within 1h. As would be expected, there is also a decrease in the pre-snRNA produced (Fig. 5.7).

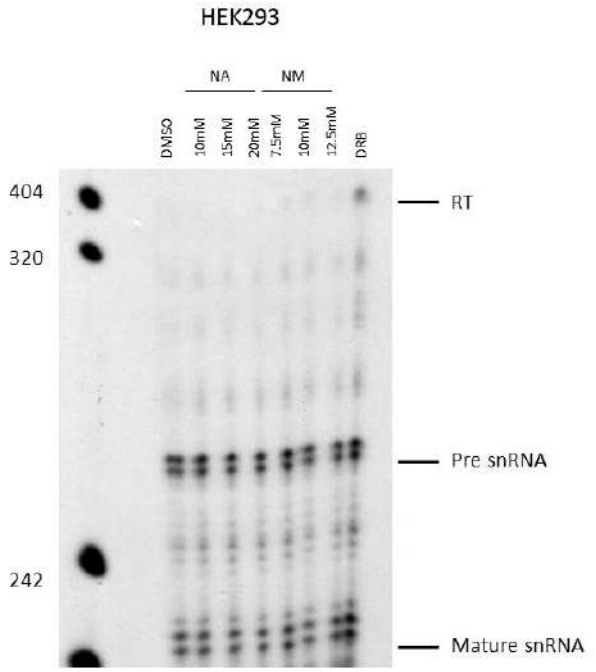


Figure 5.4. The effect of inhibiting analog-sensitive CDKs on pre-snRNA 3' end formation

HEK293 cells were treated with DMSO, NA, NM and DRB at the concentrations stated. The ratio of the RT and the pre-snRNA is given in Fig. S5.1 under each lane (See page 119).

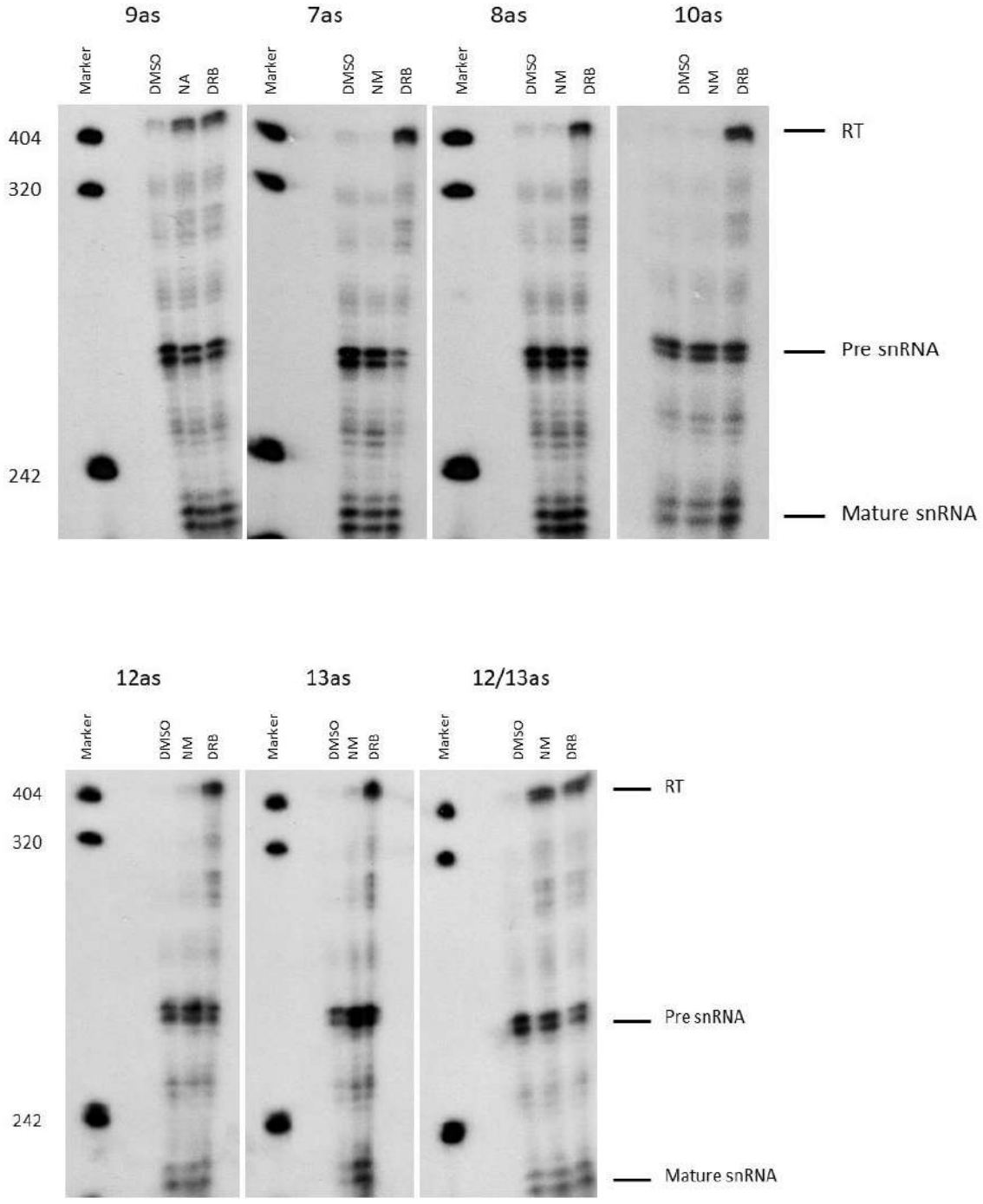


Figure 5.5. RNase protections on CDK 7, 8, 9, 10, 12, 13 and 12/13 as.

RNase protections of RNA from CDK 7, 8, 9, 10, 12, 13 and 12/13as cells after DMSO, NA, NM or DRB treatment. Inhibition of CDK 9 by DRB was used as a postive control for the 3 end processing defect. Inhibition of CDK 7, 8, 9, 10, 12 and 13as on their own do not show 3' end processing defect as there is no RT. Inhibition of CDK12 and 13 at the same time causes RT accumulation. The ratio of the RT and the pre-snRNA can be found in Fig. S5.2 and S5.3 under each lane (See pages 120 and 121 respectively).

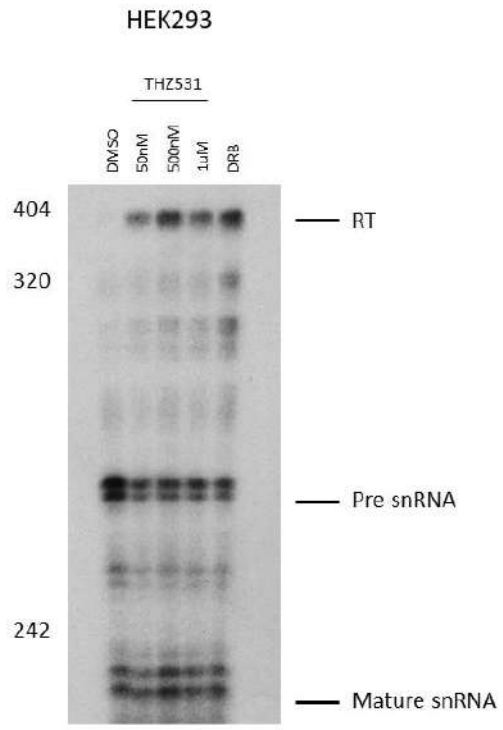


Figure 5.6. RNase protections after treatment of HEK293 with TZH531.

Wild type HEK293 cells were treated with 50nM, 500nM and 1uM THZ532 as indicated. All conditions cause a 3'end processing defect. DMSO was used as a negative control and DRB was used as a positive control. The ratio of the RT and the pre-snRNA is given in Fig. S5.1 under each lane (See page 119).

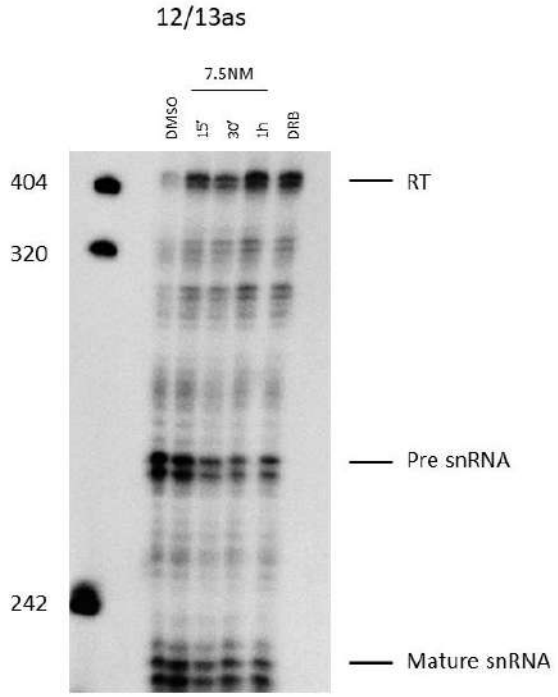


Figure 5.7 Time course of CDK12/13as inhibition

Time course of inhibition of CDK12/13as with 7.5mM NM after 15 minutes, 30 minutes and 1 hour. There is an RT accumulation within 15min and a further increase at 1h. The pre-snRNA levels also decrease over time.

5.3 Investigating CDK12 and CDK13 redundancy

5.3.1 The association of RPAP2 and Integrator with the U2 snRNA gene

As inhibiting CDK12 and CDK13 has an impact on 3' end processing, these kinases could be involved in the recruitment of RPAP2, as RPAP2 recruits Integrator, which is the complex that cleaves the pre-snRNA. ChIP experiments using an antibody against RPAP2 were performed after inhibition of CDK12/CDK13as. However, inhibition causes no reduction of RPAP2 levels on the U2 snRNA gene (Fig. 5.8). ChIP using antibodies against Integrator subunits 9 and 11, which are the catalytic subunits of Integrator, also show that Integrator recruitment is unaffected by CDK12/CDK13as inhibition, and those subunits are still present on the U2 snRNA gene at normal levels (Fig. 5.8). These results indicate that Integrator is present but may require phosphorylation by CDK12/13 to be active and cleave the pre-snRNA. Alternatively, the function of other proteins involved in 3' end processing are affected by CDK12 and CDK13 inhibition.

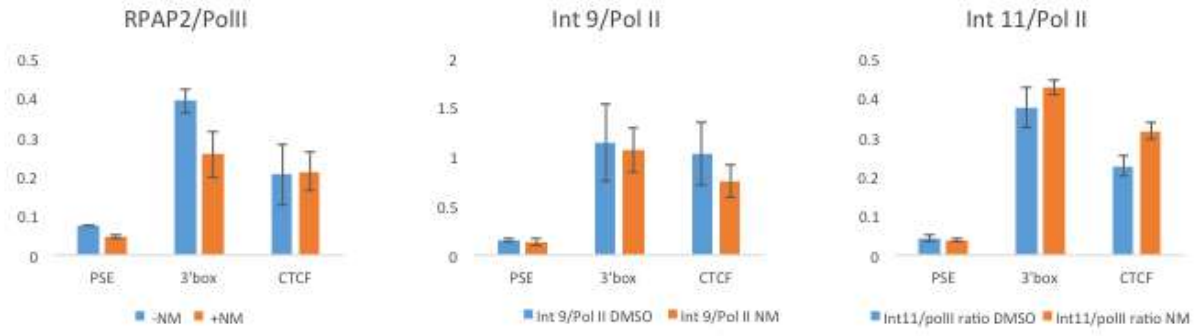


Figure. 5.8. RPAP2, Int 9 and Int 11 ChIP on the U2 snRNA in 12/13as after drug inhibition.

Left, RPAP2 ChIP on the U2 snRNA genes in normal conditions (blue bars) and after drug treatment (orange bars). Middle, Int9 ChIP. Right, Int11 ChIP. All ChIP results have been ratioed to pol II as these factors interact with it.

5.3.2 Pol II ChIPs on the protein coding gene KPNB1

Considering CDK12 and CDK13 are redundant for 3'end processing of pre-snRNAs, they may also have redundant roles in protein coding genes. For this reason, ChIP using an antibody against pol II on the KPNB1 gene were also performed.

The pol II ChIP on KPNB1 gene after CDK12/13as inhibition indicate that those CDKs also have an effect on protein coding genes (Fig 5.9). Pol II increases at the TSS, as it does for snRNA genes, and additionally, pol II decreases around the 3' end of the gene. In order to show that this effect is only seen in the inhibition of both CDKs, ChIP using an antibody against pol II after inhibition of CDK13as only were also performed. Upon CDK13 inhibition in the CDK13as cells, the pol II levels are not affected along the KPNB1 gene. CDK12 inhibition on CDK12 as cells were previously performed by Tellier et al (2019, unpublished) and its inhibition shows an increase of pol II in the gene body as a result of polymerase slowing down during elongation. The effect of CDK12/13 inhibition on elongation and termination in protein coding genes are recapitulated by CDK12 on its own (Tellier, 2019. Unpublished). However, the single kinases effects on pol II at the TSS contrast with the CDK12/13 effect, which might suggest that these kinases are redundant in the same regulatory pathway.

The change in pol II profile at the TSS of protein coding genes after CDK12/13as inhibition could also be an effect of Integrator being affected as Integrator is involved in promoter proximal pausing (57, 59, 79).

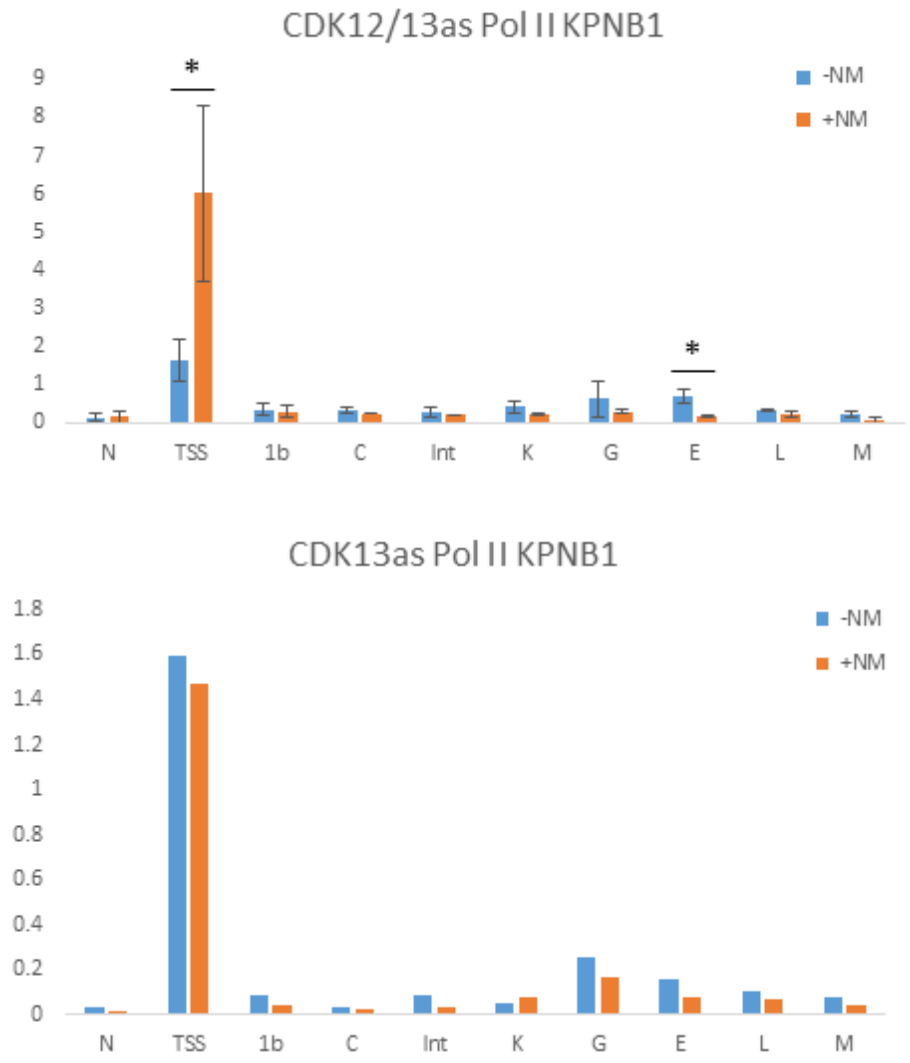


Figure 5.9. Pol II ChIP on KNPB1 in CDK12/13and CDK13 as cells after drug treatment.

Top: Pol II levels on the KPNB1 gene in CDK12/13 as on normal conditions (blue bars) and after drug treatment (orange bars). Bottom: Pol II levels on the KPNB1 gene in CDK13 as normal conditions (blue bars) and after drug treatment (orange bars).

5.3.3 Int11 Immunoprecipitation in CDK12/13 as cells

In order to further explore the effect of CDK12/13 inhibition, I performed an immunoprecipitation (IP) of Int11 in HEK293 and CDK12/13 under normal conditions and

after drug treatment and analysed by mass spectrometry interaction partners identified (See Table S5.1 from page 122 to 125 and Table S5.2 from page 126 to 128 for the full list of proteins). Int1, 2, 3, 4, 5, 6, 7, 8, 9 and 12 are all co-immunoprecipitated with Int11 in normal conditions and after drug treatment at the same levels. Additionally, the Int6 paralog DDX26b or Int6L (76) also co-immunoprecipitates with Int11. Those results indicate that CDK12/13as inhibition has no impact on Integrator recruitment or assembly.

Interestingly, upon CDK12 and 13 inhibition of CDK12/13 as, cells there is an increase in the amount of pol II pulled down: from 1 peptide in normal conditions to 10 peptides upon drug inhibition; 0.14 to 1.11 ratio to Int11. This might indicate pol II arrest, which would agree with the ChIP results where there is an increase of pol II in the PSE for the U2 snRNA gene or TSS for KPNB1 (Section 5.1 and Section 5.3.2, respectively). The only loss that has been observed was CPSF2, which changed from 2.57 in normal conditions to 1.78 upon drug inhibition when ratioed to Int11.

The pol II subunits RPB1, 2, 3, 7, and 11 and RPABC1 and 3 were also found to co-immunoprecipitate with Int11. Protein coding gene factors, such as the polyA binding proteins PABC1 and PABC4 were co-immunoprecipitated with Int11. Integrator may also be involved in splicing as splicing factors such as SRSF7, SRSF10 and PUF60 were also found associated with Int11. snRNPs proteins such as U1-70, U2A, U2B and the Sm proteins B, D2 and D3 were also found to be associated with Int11, suggesting that Integrator might interact with the spliceosome. Previous studies have shown crosstalk between 3'end complex factors and splicing, possibly coupling the terminal intron removal with polyadenylation (190)

Int11 recruitment is not impaired by CDK12/13as inhibition as seen by ChIP, and the IP results indicate that the assembly of Integrator is also not affected. However, Integrator is still not able to process when CDK12/13 are inactivated, as indicated by RNase protection analysis. Thus, Integrator is being recruited but not active. This could be due to a lack of phosphorylation of its subunits. For this reason, phosphorylation changes upon CDK12/13as inhibition were also analysed.

The top 10 hits on the mass spectrometry list include the CPSF complex (CPSF1, 2, 3 and 4), Fip1 and WDR33. This could suggest that integrator has a role in cleavage and polyadenylation or that poly(A) factors play a role in snRNA gene 3' box recognition. A phopsho(S/T)Y search led to a short list of two proteins as putative targets of CDK12 and CDK13 that also interact with Integrator. The phosphorylation changes only upon CDK12 and CDK13 inhibition suggest that CPSF4 and WDR33 are targets of those kinases and also interact with Integrator (Table S.5.3; See page 129). This could indicate that CPSF4 is a target of CDK12 and CDK13 and also have a role in 3'end processing in addition to Integrator.

Cryo-EM structure has shown the association of CPSF1, CPSF4 and WDR33 with the polyadenylation signal AAUAAA hexamer. CPSF1 acts as a scaffold for CPSF4 and WDR33 to interact and recognize the polyadenylation signal (137, 191).

My mass spectrometry results indicate that CDK12 and CDK13 might have redundant roles in 3' end processing of protein coding genes in addition to 3' box recognition. The CPSF complex is recruited and assembled but it seems to require CDK12 and CDK13 activity to phosphorylate CPSF4 and WDR33 in order to recognise the PAS and affect 3' end processing. CDK12 and 13 by phosphorylating polyA factors would then be

involved in similar mechanisms both in snRNA genes and protein coding genes regulating 3’ end processing of both classes of genes and therefore indicating that polyA factors might have a role in 3’end processing of snRNA genes.

5.4 Conclusions

Pol II is not affected by a single CDK inhibition, but inhibition of CDK12 and 13 at the same time affects pol II occupancy on both the U2 snRNA gene and the protein coding gene KPNB1 indicating an increase of pol II levels in the PSE and the TSS, respectively. Since RNase protection experiments has shown that CDK12 and 13 are redundant for 3’end processing of the U2 snRNA gene I also tried to identify the mechanism by which it would be regulating its processing. Surprisingly, ChIP experiments show that upon inhibition CDK12 and 13 there is no change in RPAP2 or Integrator levels. Pulldown of Int11 and further phospho mass spectrometry showed that CPSF4 and WDR33 are targets of CDK12 and 13.

Chapter 6

Discussion

6.1 CAPTURE identifies potential new candidates that regulate snRNA gene expression.

Aiming to identify proteins involved in snRNA gene expression in an unbiased way, I initially tried using the enChIP method (172), which failed to enrich for proteins over the control pulldown of dCas9 without a guide. In fact, a critical analysis of the original paper describing enChIP led me to the conclusion that this technique is not successful in identifying proteins associated with the region of interest. The experiment lacks negative controls and the mass spectrometry list shows that no transcription factors are being brought down from the promoter region targeted (172). I have confirmed that enChIP is not a useful technique by including the dCas9 alone control and showing that no protein profile changes or enrichment for proteins involved in transcription is achieved (Section 3.3). In addition, I concluded that chromatin extraction is required to reduce background and non-specific protein binding, as there are several cytoplasmatic proteins on the enChIP mass spectrometry list (172). CAPTURE is instead a much more efficient method than enChIP as it not only has a chromatin extraction step but also subjects the chromatin to urea washes, which enrich for chromatin-bound proteins (175). An additional

advantage of CAPTURE over enChIP is that it does not rely on a tag-antibody interaction such as FLAG and instead uses biotin-streptavidin interaction, which is much stronger as seen by comparison of CAPTURE with FLAG and biotin-streptavidin (174). I have shown using CAPTURE that the biotin-streptavidin interaction is crucial for the identification of proteins in the region of interest (Section 4.2); FLAG-CAPTURE experiments failed to identify known factors involved in snRNA gene expression, whereas biotin-streptavidin CAPTURE identified several factors involved in different stages of expression such as Integrator subunits Int1, Int2, Int3, Int4, Int7, Int11, Mediator subunits Med16, Med23, pol II subunit RPB2, Spt5, CDK7, NELFA, NELFB and CDK9.

Biotin-streptavidin CAPTURE also identified new proteins potentially required for regulation of snRNA gene expression. The SMARC factors are the most interesting candidates identified; SMARC factors are part of the SWI/SNF complex involved in chromatin remodelling (186). In yeast, the SWI/SNF complex binds proximal to the promoter and allows the binding of transcription factors, such as the nuclear factor (NF) 1, which activates transcription (192). SWI/SNF is recruited by the CREB-binding protein and acetylated histone tails. It promotes DNA twisting and removal of histones thereby allowing TFIID to bind and displace nucleosomes at the TSS, thereby promoting transcription initiation (193). The SWI/SNF complex is also able to remodel chromatin and relocate nucleosomes to facilitate transcription. SMARC2 and SMARC4 factors are ATPase and helicase catalytic subunits, the other SMARC factors functions are still unknown (194). The SWI/SNF complex has also been implicated in tumour suppression by acting as an epigenetic regulator and has been used as a biomarker for several types of cancers as truncation mutations on SMARCB1 were found in malignant rhabdoid tumours (194). SWI/SNF could be involved in regulating transcription of snRNA genes by changing the

epigenetic environment of the PSE region, allowing transcription factors to bind and promote transcription. PTF binding to the PSE is required to maintain an open chromatin structure as the nucleosome density across the U1 and U2 snRNA genes increases after PTFy knockdown (37). However, the underlying mechanism is unclear. One possibility is that SMARC factors could be recruited by or interact with PTFy to remove nucleosomes from the transcribed region.

Several protein-coding gene factors involved in RNA 3'end formation and termination such as Cstf1, Cstf2, Cstf3, CPSF1 and XRN2 were also found associated with the U2 snRNA gene. In fact, ChIP-seq experiments have already shown that Xrn2 is present in the U1, U4 and U5 snRNA genes and that Xrn2 inactivation can cause termination delay as seen by a pol II shift towards the 3' end of the genes (195). Although this result contrasts with results from a different group where ChIP of pol II upon Xrn2 siRNA knockdown on U1 and U2 snRNA genes shows no effect on pol II profile (37). However, the role of Xrn2 in termination of snRNA genes it is still unclear. Surprisingly, the splicing factors SF1, SF3A1, SF3B2, 4, and 6 were also found associated the U2 snRNA gene. The protein-coding genes termination factors Ssu72 and Pcf11 have also been detected on the U1 and U2 snRNA gene by ChIP assays and their knockdown has an effect on pol II as seen by an increase in pol II levels at the end of the U1 and U2 snRNA genes (37). RNase protections have shown that knockdown of Pcf11 and Ssu72 has little effect on 3'box recognition, implicating these factors in pol II termination on snRNA genes rather than in 3' end processing (37). Cstf2 has also been shown to be present on the U1 and U2 snRNA genes but its role has not been further assessed (37). Many protein-coding gene factors are now implicated in snRNA gene expression and CPSF2 and WDR33 could just as well be protein-coding genes factors involved in regulation of expression of the U2 snRNA gene.

6.1.1. Future Directions

The candidates revealed by CAPTURE still need to be validated to be on the U2 snRNA gene by ChIP and in order to further assess their occupancy on other snRNA genes, ideally, ChIP-seq should be performed. Knockdown of candidates such as the SMARC factors and pol II and histone ChIPs would determine if the SMARC factors have an impact on pol II and nucleosome levels across the U2 snRNA gene. Additionally, further investigation of the roles of protein-coding gene factors in snRNA gene expression could reveal a new layer of regulation of expression. ChIP assays would confirm their occupancy on the U2 snRNA genes and map the region bound. Furthermore, pol II ChIPs and RNase protection after knockdown of these factors would determine how those factors affect pol II and if they are involved in 3' end processing.

It would also be interesting to carry out CAPTURE after inhibiting the analog-sensitive CDKs. This would help to determine how these kinases affect transcription/RNA processing as it would allow the transcription/RNA processing machinery with and without inhibition to be characterised.

6.2 The roles of CDKs in snRNA gene expression

Another aspect of my project was to investigate the roles of the transcriptional CDKs in snRNA gene expression. By using analog-sensitive cells for CDK7, 8, 9, 10, 12, 13 and the double 12/13, I have assessed the pol II profile upon inhibition of each of those kinases. Inhibition of an individual kinase does not show any effect on pol II levels across

the U2 snRNA gene. However, co-inhibition of CDK12 and 13 at the same time causes an increase in pol II over the PSE, as seen by ChIP (Section 5.1.2). RNase protections confirm that CDK9 is required for 3' end processing of the U2 snRNA gene and shows for the first time that CDK12 and CDK13 are redundant for 3' end processing of transcripts from the U2 snRNA gene. Protein-coding gene expression after inhibition of CDK12 and 13 has also been assessed. Inhibition of CDK12 results in an increase of pol II in the gene body and a loss of pol II after the polyA site (Tellier, 2019; under revision) whereas CDK13 inhibition does not alter the pol II profile across KPNB1 (Section 5.1.2). However, inhibition of CDK12 and 13 together causes increased pol II levels on the TSS, but also a decrease of pol II in the termination site suggesting that the inhibition of those kinases might cause premature termination. A recent paper has suggested that the premature termination observed upon inhibition with THZ531 would be effect of CDK12 only rather than CDK12 and 13 (196).

In order to further investigate the roles of CDK12 and CDK13 in snRNA gene expression, I checked if co-inhibition of CDK12 and 13 impairs recruitment of Integrator, which would lead to a 3' end processing defect. I also checked S7P phosphorylation levels by ChIP as S7P is necessary for recruitment of Integrator (80). S7P is not affected by inhibition of these kinases (Section 5.1.1). Recruitment of the catalytic subunits Int9 and Int11 is also not affected as seen by ChIP experiments (Section 5.1.1). Int9 and Int11 both have sites that could be phosphorylated by CDK12 and CDK13 and these phospho-acceptor sites are in the beta-lactamase domain that is responsible for the catalytic activity. I therefore performed an Int11 pulldown and submitted the resulting proteins to phospho mass spectrometry to investigate whether the phosphorylation status of those subunits changes upon inhibition of CDK12 and 13.

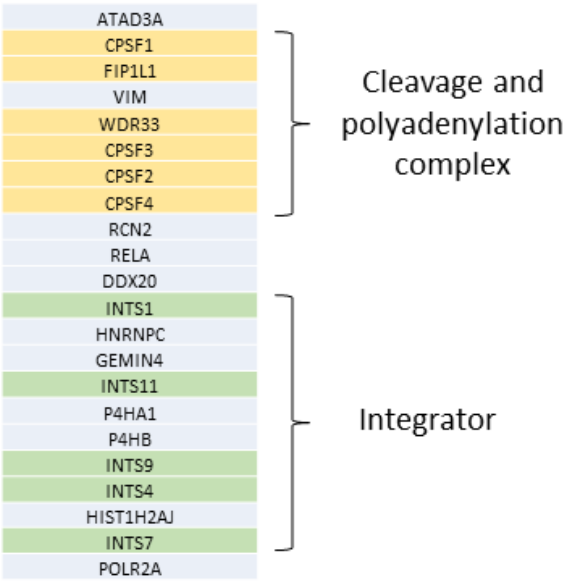


Figure 6.1. List of the top hits of the Int11 pulldown.

List of the top proteins from the Int11 pulldown. The cleavage and polyadenylation complex is indicated in yellow and the Integrator complex is indicated in green. Only subunits Int1, 4, 7, 9 and 11 are shown in the figure.

Interestingly, the Int11 pulldown also brought down its paralog CPSF3. Analysing the protein list, it is clear that the top complex to be pulled down is the protein-coding gene cleavage and polyadenylation complex, comprising the CPSF complex (CPSF1, 2, 3 and 4), Fip1 and WDR33 (Fig. 6.1). The second complex pulled down was the snRNA gene 3' end processing complex Integrator and all subunits were co-immunoprecipitated with the exception of Int10 (Fig. 6.1). Regarding the phosphorylation changes, there were no changes for any of the Integrator subunits. However, the phosphorylation status of the two subunits CPSF4 and WDR33, which are involved in recognition of the polyA signal, was affected. These results indicate that CPSF4 and WDR33 are targets of CDK12 and CDK13 and that phosphorylation of these factors might play a role in 3' end processing of both protein-coding and snRNA gene transcripts.

CDK12 and 13 may therefore play a central role in 3’ end processing of both snRNA and protein coding genes. However, the mechanism by which CDK12 and 13 activates 3’end processing of pre-snRNA is still unclear. My results suggest that Integrator may not be the only complex responsible for 3’end processing of snRNA genes as I see a 3' end processing defect even though Integrator recruitment and its phosphorylation are unchanged. CDK12 and 13 affect the phosphorylation of the polyadenylation factors CPSF4 and WDR33, suggesting that there might be new layers of regulating 3’ end processing. As it is possible that the polyadenylation factors, CPSF4 and WDR33 are being pulled down because the Int11 antibody is cross-reacting with CPSF3, further experiments are needed to determine the role of these polyadenylation factors in 3' box-directed 3' end formation. As CPSF1 is found to be associated with the U2 snRNA gene by CAPTURE, a role for polyadenylation factors in U2 gene expression cannot be ruled out. However, a role for the phosphorylation of factors that do not interact directly with Integrator is also possible.

6.2.1. Future directions

To better understand how CDK12 and CDK13 activate 3’ box-dependent RNA processing, I have submitted protein samples from CDK12as, CDK13as and CDK12as/CDK13as cells untreated or treated for 30 minutes or 1 hour with inhibitor for label-free phospho-proteomics analysis at the Target Discovery Institute (Raphael Heilig and Roman Fischer) Tandem Mass Tag (TMT) to identify the targets of CDK12 and CDK13. I am also performing ChIP-seq of pol II and phospho-pol II upon inhibition of CDK12as/13as as well as on HEK293 cells treated with THZ531, which inhibits both

enzymes. I am also producing cell lines expressing FLAG-tagged Int11 and Int9 to repeat the pulldown experiments to test whether CPSF4 and WDR33 are associated with Integrator. If they are, I will carry out ChIP to determine whether these factors are associated with snRNA genes.

CONCLUSIONS

My work has identified new candidate proteins to regulate snRNA gene expression, and also has revealed several protein-coding genes factors that might play major roles in snRNA gene expression. Additionally, the redundancy of CDK12 and CDK13 for the 3' end processing of snRNA gene transcripts, their roles in protein-coding gene expression and their putative targets CPSF4 and WDR33 are novel and significant in explaining the roles of these two CDKs in expression of human genes.

Parts of the introduction were based on my published literature review (Guiro & Murphy, 2017) (197).

Supplementary Material

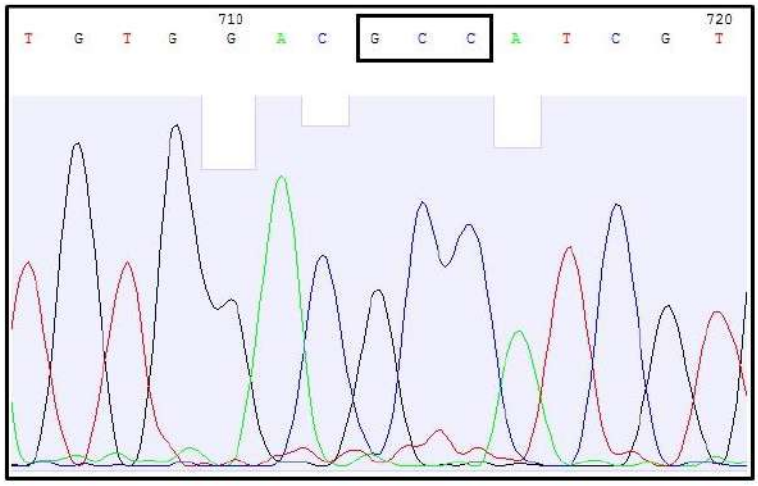


Figure S2.1 Sequence of the dCas9 mutated on the pX462.

Sequencing trace for dCas9 with the mutated gatekeeper highlighted with a box.

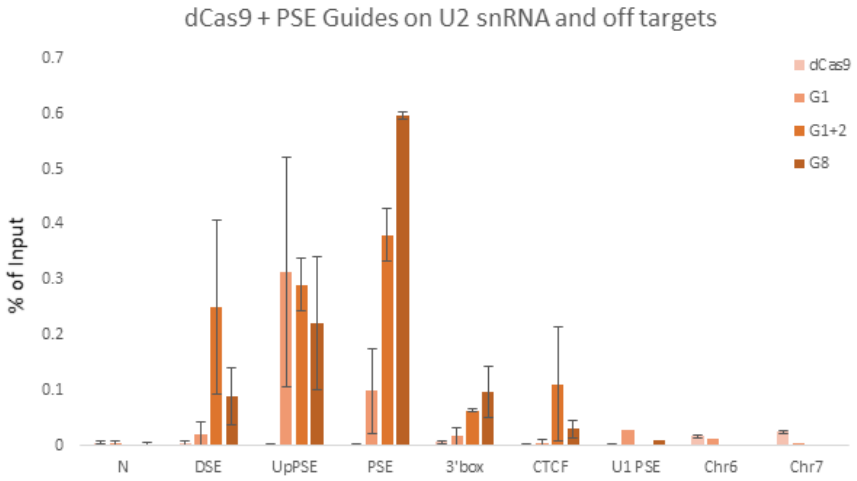


Figure S3.1. dCas9 occupancy on the U2 snRNA gene, U1 PSE and off targets in Chr 6 and 7.

Primers N, DSE, UpPSE, PSE, 3'box and CTCF shows that dCas9 is being targeted to the U2 snRNA with a peak on the PSE seen by the G8. dCas9 is not present on the U1PSE gene even though the PSE region is very conversed between snRNA genes. Off targets on Chr6 and Chr7 are also nor targeted by dCas9.

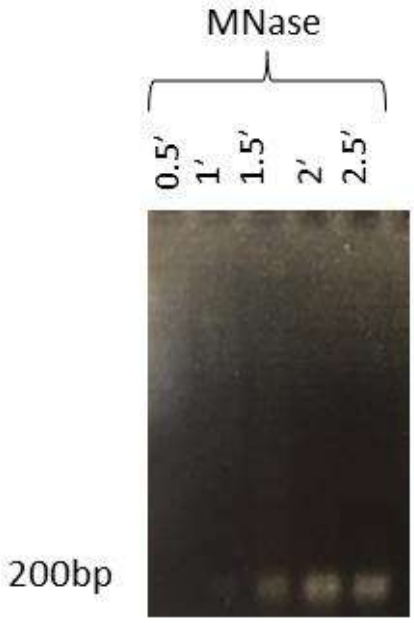


Figure S3.2. MNase digestion of chromatin.

The chromatin was digested and run on an agarose gel, the fragments had around 200bp length.

Complete list of all the interactions

#	Protein ID	RNA ID	Z-score?	Discriminative Power (%)?	Interaction Strength (%)?	Domain?	Motif?	Ranking?
1	LN28B_HUMAN	input_rna_1	-0.25	37	91	yes	yes	
2	HNRPC_HUMAN	input_rna_1	-0.32	33	65	yes	yes	
3	SRSF2_HUMAN	input_rna_1	-0.58	17	35	yes	yes	
4	RSMB_HUMAN	input_rna_1	-0.57	17	43	yes	yes	
5	MBNL1_HUMAN	input_rna_1	-0.71	14	15	yes	yes	
6	SRSF9_HUMAN	input_rna_1	-0.73	14	9	yes	yes	

Figure S4.1. catRAPID predicted proteins to interact with the U2 snRNA.

Top hits on the catRAPID for the mature U2 snRNA sequence against the proteome database.

Table S4.1. U2 snRNA PSE FLAG CAPTURE.

List of proteins specific to the U2 snRNA PSE condition (shown as guide only) or enriched in the U2 snRNA PSE (the values are fold difference in relation to the dCas9 negative control; dCas9 + PSE guide/dCas9 without guide).

Gene names	G8/dCas9 ratio
ZNF512B	guide only
STAU2	guide only
DHFR	guide only
MOV10	guide only
GNL1	guide only
USP9X;USP9Y	guide only
CSK	guide only
ACTN4;ACTN1	guide only
SNRNP40	guide only
UBAP2L	guide only
ITGAX	guide only
COPA	guide only
LSM2	guide only
IPO5	guide only
KTN1	guide only
GCLM	guide only
SRP72	guide only
LIN28B	guide only
IRF2BP2	guide only
PRH1	5.830176768
CALML5	5.375462353
CENPV	4.072910676
CCDC124	3.996928747
SUGP2	3.784700244
FXR1	3.291967825
BZW2	3.179676172
HMGN2	3.037988593
DSG1	3.003321367
KPNA2	2.930960716
EIF6	2.928728472
NPM3	2.914909526
DSC1	2.839677892
TMA16	2.67163189
CARD8	2.642821425
CSTF3	2.57622693

Table S4.2. U2 snRNA PSE Streptavidin-Biotin CAPTURE.

Gene names	G8/dCas9	Gene names	G8/dCas9
HSPA6;HSPA7	Guide only	UTRN	20.37017937
TTN	Guide only	TRA2A	12.23150327
SHPRH	Guide only	SMN1	10.77023769
DBN1	Guide only	IRF3	8.693293046
BTF3	Guide only	HEATR6	7.880493033
MMS22L	Guide only	XYLB	7.841714957
ADAR	Guide only	PLEC;EPPK1	7.43374947
TUBGCP4	Guide only	TMX1	7.220688041
BSG	Guide only	MED20	6.729851461
PRPF4B	Guide only	DNMT3B	6.670113137
MAK16	Guide only	NOP9	6.533030122
CPSF6	Guide only	MEPCE	5.645752599
MARH5	Guide only	NKRF	5.594394673
SRBD1	Guide only	CBX1	5.512822518
C12orf10	Guide only	SAR1B	5.441459183
RPA3	Guide only	SDHA	5.417557702
DBR1	Guide only	TCAF1	5.215915688
METTL3	Guide only	UBE2L3	5.211488585
CRKL	Guide only	ZNF326	5.109811511
LRRIQ1	Guide only	CDC42	5.053189109
DYNLT1	Guide only	STMN2;STMN1	5.02429311
RNF126	Guide only	UMPS	4.724221802
PSPC1	Guide only	U2AF1	4.678885272
KIF23	Guide only	GPD1L	4.623669294
POGK	Guide only	H1FX	4.6180443
STAU1	Guide only	CDC37	4.570570361
SAR1A	Guide only	IQGAP2	4.533869017
YTHDC2	Guide only	ZC3H11A	4.432471787
PSMB2	Guide only	RBM22	4.425352831
EXOSC8	Guide only	MBNL1;MBNL2	4.330021627
HIRIP3	Guide only	PPIL4	4.166109051
TAT	Guide only	GNB1	4.139098199
CXorf38	Guide only	BCLAF1	4.113679089
NR2F2	Guide only	IRF2BP2	4.068834832
NOM1	Guide only	YWHAG	4.050186866
CHMP2B	Guide only	MCCC1	4.042720008
HIST1H3A;HIST3H3	Guide only	CRNKL1	4.033875355
PSME2	Guide only		
DDX31	Guide only		
UBE2K	Guide only		
GLS	Guide only		
RBM26	Guide only		
SESN2	Guide only		
NUDCD2	Guide only		
SLK;STK10	33.44456551		
ABRACL	25.16736784		

Table S4.3. Optimized U2 snRNA PSE Streptavidin-Biotin CAPTURE.

Gene names	G8/dCas9	Gene names	G8/dCas9
SNRPD1	Guide only	RAE1	Guide only
FBL	Guide only	MND1	Guide only
SRSF3	Guide only	FUBP3	Guide only
PDS5B	Guide only	CCAR2	Guide only
HNRNPH3	Guide only	MSH6	Guide only
DFFA	Guide only	SMARCA4;SMARCA2	Guide only
NANS	Guide only	BRIX1	Guide only
GLUL	Guide only	INTS3	Guide only
FANCI	Guide only	TCOF1	Guide only
TRRAP	Guide only	ANAPC7	Guide only
NOL9	Guide only	NUMA1	Guide only
DIS3	Guide only	THOC6	Guide only
FUS;TAF15	Guide only	PPP1CC	Guide only
WDR75	Guide only	CLNS1A	Guide only
YKT6	Guide only	IRS4	Guide only
WDR5	Guide only	NACA	Guide only
SNX2	Guide only	HSPB1	Guide only
NAT2	Guide only	INTS1	Guide only
HCFC1	Guide only	ALYREF	Guide only
NAT10	Guide only	SUMO1	Guide only
ETF1	Guide only	SF1	Guide only
EIF4B	Guide only	UTP20	Guide only
KIF22	Guide only	MED23	Guide only
HDGF	Guide only	STK26;STK25;STK24	Guide only
MTAP	Guide only	TOR1AIP1	Guide only
MDC1	Guide only	SCRIB	Guide only
MAT2B	Guide only	AASDHPPT	Guide only
SBDS	Guide only	UBXN1	Guide only
HIST1H2AC	Guide only	NXN	Guide only
MCM4	Guide only	ADD1	Guide only
SF3B6	Guide only	UCHL5	Guide only
RPLP0	Guide only	RBM14	Guide only
NUP214	Guide only	POLR2B	Guide only
SMC2	Guide only	DEK	Guide only
NSUN2	Guide only	HNRNPF	Guide only
DDX46	Guide only	SUPT5H	Guide only
ANP32E	Guide only	PPP2R5C;PPP2R5D	Guide only
UMPS	Guide only	RAB3GAP2	Guide only
GEMIN5	Guide only	HNRNPA3	Guide only
PABPN1	Guide only	UTP15	Guide only
EIF3J	Guide only	PDLIM1	Guide only
PPP1R8	Guide only	OSBP	Guide only
RPP30	Guide only	HMGB2	Guide only
RECQL	Guide only	POP1	Guide only
SART3	Guide only	DRG2	Guide only
SSR4	Guide only	DPM1	Guide only

Gene names	G8/dCas9	Gene names	G8/dCas9
PFN2	Guide only	UFD1L	Guide only
HP1BP3	Guide only	CHD1L	Guide only
YWHAG	Guide only	RDX	Guide only
USP39	Guide only	SMAD3;SMAD9	Guide only
WDR36	Guide only	YTHDF1;YTHDF3	Guide only
BOP1	Guide only	CCAR1	Guide only
RBM4;RBM4B	Guide only	VPS26B	Guide only
CDC16	Guide only	PIN4	Guide only
ATXN2L	Guide only	POLR3C	Guide only
VBP1	Guide only	RTN4	Guide only
GRWD1	Guide only	OGT	Guide only
PHF5A	Guide only	GTF3C2	Guide only
PLRG1	Guide only	PES1	Guide only
SERPINH1	Guide only	AK2	Guide only
TARDBP	Guide only	ADRM1	Guide only
SUN2	Guide only	RAB5B	Guide only
GNL3	Guide only	SMEK1;SMEK2	Guide only
SMC3	Guide only	CPSF1	Guide only
NOL6	Guide only	RIF1	Guide only
INTS2	Guide only	ZFR	Guide only
LSM14A	Guide only	MTA2	Guide only
PWP2	Guide only	NUP50	Guide only
CSNK2A2	Guide only	CDK7	Guide only
THOC5	Guide only	LMAN1	Guide only
HDLBP	Guide only	THOC2	Guide only
GLYR1	Guide only	NUP153	Guide only
FAM50A	Guide only	AK1	Guide only
PRPF4	Guide only	NELFB	Guide only
AP3S1	Guide only	PRPF3	Guide only
GLE1	Guide only	CDC5L	Guide only
NEK9	Guide only	HEATR3	Guide only
EXOSC9	Guide only	CPSF3L	Guide only
PLAA	Guide only	PPIH	Guide only
SMU1	Guide only	DNAJC2	Guide only
LIMS1	Guide only	DDX47	Guide only
LUC7L2	Guide only	POLR3A	Guide only
CBX5	Guide only	DIMT1	Guide only
SLC35E1	Guide only	KIF2A	Guide only
COG6	Guide only	CNP	Guide only
VPS4A;VPS4B	Guide only	SRRM2	Guide only
CSTF2;CSTF2T	Guide only	PNO1	Guide only
HSPA6;HSPA7	Guide only	SEC16A	Guide only
DDX52	Guide only	BYSL	Guide only
EXOSC2	Guide only	COPS2	Guide only
POLD1	Guide only	ELAC2	Guide only
SMARCA5	Guide only	OXR1;NCOA7	Guide only

Gene names	G8/dCas9	Gene names	G8/dCas9
UTP6	Guide only	ARAF;RAF1	Guide only
H2AFY2	Guide only	TRMT1	Guide only
RRP9	Guide only	PRIM2	Guide only
RBM39	Guide only	CDKN2A	Guide only
TRMT1L	Guide only	HIST1H1D	Guide only
ZNF207	Guide only	NOP9	Guide only
LAS1L	Guide only	GINS1	Guide only
NOC2L	Guide only	CCNK	Guide only
CENPV	Guide only	AKAP11	Guide only
ILKAP	Guide only	NOC3L	Guide only
SMARCC1	Guide only	DNAJB1	Guide only
NOC4L	Guide only	CTNNBL1	Guide only
U2AF1	Guide only	KDM1A	Guide only
ORC3	Guide only	URB1	Guide only
RBM12	Guide only	RCL1	Guide only
TLE4;TLE1;TLE2;TLE3	Guide only	NPLOC4	Guide only
CIRH1A	Guide only	MTA1;MTA3	Guide only
KPNA6;KPNA5	Guide only	NCAPD3	Guide only
CHD4;CHD5;CHD3	Guide only	CHAF1B	Guide only
PRIM1	Guide only	LIG3	Guide only
URB2	Guide only	TCEA2;TCEA1	Guide only
BCCIP	Guide only	EMG1	Guide only
AP3M1	Guide only	THOC1	Guide only
RPL7L1	Guide only	POLR3B	Guide only
SRSF9	Guide only	CTNND1	Guide only
RALY	Guide only	UBE4B	Guide only
SRSF7	Guide only	NCAPG2	Guide only
CHERP	Guide only	GNL2	Guide only
TOMM34	Guide only	POLD2	Guide only
PSPC1	Guide only	PPP2R2A	Guide only
NUPL1	Guide only	SEC24A	Guide only
CRKL	Guide only	EBNA1BP2	Guide only
EXOSC4	Guide only	ARHGEF2	Guide only
SNRNP40	Guide only	TROVE2	Guide only
SNRPC	Guide only	MED16	Guide only
WDR12	Guide only	NKRF	Guide only
SEC23IP	Guide only	RPRD1B	Guide only
SF3B4	Guide only	SAFB;SAFB2	Guide only
UROD	Guide only	KIDINS220	Guide only
SLC30A6	Guide only	SCAMP2	Guide only
RAD50	Guide only	RPIA	Guide only
PPAN	Guide only	DDX56	Guide only
MYO5A	Guide only	RBBP5	Guide only
XAB2	Guide only	NELFA	Guide only
SRSF1	Guide only	SCD	Guide only
INTS7	Guide only	CEBPZ	Guide only

Gene names	G8/dCas9	Gene names	G8/dCas9
CSTF1	Guide only	GTPBP4	10.20440959
Sep-09	Guide only	XRN2	9.656695095
DDX23	Guide only	SEC31A	9.264885629
IRF2BP1	Guide only	FAM98B	9.135737803
THEM6	Guide only	MCM5	9.112434041
POLDIP3	Guide only	HNRNPM	8.833645774
CWF19L1	Guide only	ADAR	8.101772778
CDC45	Guide only	UTP18	7.73052229
VPS16	Guide only	MSH2	7.481373899
RAD21L1;RAD21	Guide only	TOP2A	7.470047832
HNRNPLL	Guide only	NUP62	7.427542098
LIMA1	Guide only	TEX10	7.164342676
INTS4	Guide only	RPA1	7.101548048
SUPT6H	Guide only	RCC1	6.918578602
SRRT	Guide only	EIF4A3	6.601025088
POM121C;POM121	Guide only	SF3A1	6.59672643
SCYL1	Guide only	API5	6.593354514
PPIL4	Guide only	KHDRBS1	6.432472261
SUGP2	Guide only	MCM3	6.347324888
KIAA0196	Guide only	SMC4	6.134982522
SMARCE1	Guide only	DDX5	6.119793232
PLEC;EPPK1	Guide only	TUBB2B;TUBB2A	6.09950591
EIF6	Guide only	HNRNPDL	6.044040785
DDX18	Guide only	SFPQ	6.023839398
PDCD11	Guide only	GALK1	6.022164368
NOP2	28.74377421	PICALM	5.857791311
GTF2I	28.61616233	RBMX;RBMXL1	5.821141562
TBL3	26.29212437	MAGED2	5.646068491
RSL1D1	25.02387723	NHP2L1	5.61207965
MYBBP1A	22.98348553	DHX15	5.580981091
DDX21	19.43688083	U2AF2	5.44942222
SSRP1	18.75820388	CPNE1	5.357618608
HNRNPA0	18.69292151	DAZAP1	5.280661494
NONO	17.76871013	NUP88	5.210607176
VIM	17.66524727	CSTF3	5.186732628
HNRNPUL2	16.1599665	CBX3	5.134968167
DHX16	15.89984351	HNRNPA2B1	5.106926623
RRP12	15.41803142	MCM2	5.065344922
NUP107	15.06867489	STMN1	4.982156227
TOP1	14.09018472	NCBP1	4.931220188
NME1	13.58739509	NOP56	4.887407194
SNRNP70	12.31604957	SNRNP200	4.88018719
MCM6	12.29261775	MCM7	4.816335631
MATR3	11.85565167	CSK	4.73335875
FUBP1	10.75570989	IDH1	4.687954849
HNRNPUL1	10.47743587	SKIV2L2	4.687914066

Gene names	G8/dCas9
PPM1G	4.674638844
COG5	4.642966498
SYNE1	4.602067805
SNRPF	4.600307365
TMEM33	4.549707829
MCMBP	4.535527236
SF3B2	4.498225234
NCL	4.491234018
NUP133	4.486941677
SUPT16H	4.483754846
WDR43	4.477919494
SARNP	4.444444444
HNRNPU	4.429816632
H1F0	4.417641229
SNRPA	4.412082692
NUP54	4.409714706
PTBP1	4.395638072
CHTOP	4.361070077
ELAVL1	4.358953722
TRIM28	4.355186722
NOP58	4.332957181
NUP43	4.329447623
SMC1A	4.32912125
GBE1	4.315923633
CDK9	4.292590127
PRPF8	4.263583598
PRPF31	4.234469219
DDX17	4.230105937
SPTBN1	4.229911023
EWSR1	4.208072551
RAB8A	4.194886661
XRCC5	4.118283448
CEPT1	4.099490401
KHSRP	4.085076046
NOL11	4.082059477
HNRNPR	4.046500936

Table S4.4. U2 snRNA DSE Streptavidin-Biotin CAPTURE.

Gene names	G4/dCas9	Gene names	G4/dCas9
ACTG2;ACTA1;ACTC1;ACTA2	Guide only	RECQL	Guide only
SNRPD1	Guide only	SART3	Guide only
FBL	Guide only	RAE1	Guide only
SRSF3	Guide only	FUBP3	Guide only
PDS5B	Guide only	CCAR2	Guide only
HNRNPH3	Guide only	MSH6	Guide only
DFFA	Guide only	MAP4	Guide only
SSB	Guide only	SMARCA4;SMARCA2	Guide only
GLUL	Guide only	BRIX1	Guide only
FANCI	Guide only	INTS3	Guide only
TRRAP	Guide only	TCOF1	Guide only
DIS3	Guide only	NUMA1	Guide only
FUS;TAF15	Guide only	THOC6	Guide only
WDR75	Guide only	HSPA14	Guide only
DHRS7	Guide only	UNC45A	Guide only
YKT6	Guide only	IRS4	Guide only
WDR5	Guide only	RAP1B;RAP1A	Guide only
PUM1;PUM2	Guide only	HSPB1	Guide only
SNX2	Guide only	ALYREF	Guide only
NAT2	Guide only	SUMO1	Guide only
HCFC1	Guide only	RPL31	Guide only
NAT10	Guide only	UTP20	Guide only
ABCF1	Guide only	MED23	Guide only
ETF1	Guide only	TUBGCP2	Guide only
FBLL1	Guide only	UBXN1	Guide only
EIF4B	Guide only	NXN	Guide only
KIF22	Guide only	ADD1	Guide only
HDGF	Guide only	UCHL5	Guide only
MTAP	Guide only	RBM14	Guide only
MAT2B	Guide only	POLR2B	Guide only
SBDS	Guide only	DEK	Guide only
HIST1H2AC;HIST3H2A;HIST1H2AE	Guide only	HNRNPF	Guide only
MCM4	Guide only	ELP2	Guide only
SF3B6	Guide only	RAB3GAP2	Guide only
RPLP0	Guide only	KPNA4	Guide only
EIF5	Guide only	UTP15	Guide only
NUP214	Guide only	DUT	Guide only
SMC2	Guide only	HMGB2	Guide only
DDX46	Guide only	POP1	Guide only
ANP32E	Guide only	ERCC2	Guide only
UMPS	Guide only	PFN2	Guide only
GEMIN5	Guide only	HP1BP3	Guide only
PABPN1	Guide only	YWHAG	Guide only
PPP1R8	Guide only	USP39	Guide only
RAB21	Guide only	WDR36	Guide only
RPP30	Guide only	BOP1	Guide only

Gene names	G4/dCas9	Gene names	G4/dCas9
RBM4;RBM4B	Guide only	RTN4	Guide only
ATXN2L	Guide only	OGT	Guide only
VBP1	Guide only	PES1	Guide only
HDGFRP2	Guide only	ADRM1	Guide only
PHF5A	Guide only	RAB5B	Guide only
PLRG1	Guide only	RIF1	Guide only
SERPINH1	Guide only	MTA2	Guide only
TARDBP	Guide only	MTHFD1L	Guide only
SUN2	Guide only	WDR44	Guide only
PCK2	Guide only	THOC2	Guide only
GNL3	Guide only	PRKAG1	Guide only
SMC3	Guide only	NELFB	Guide only
NOL6	Guide only	CDC5L	Guide only
INTS2	Guide only	IGF2BP2	Guide only
LSM14A	Guide only	HEATR3	Guide only
PWP2	Guide only	CPSF3L	Guide only
CSNK2A2	Guide only	NHLRC2	Guide only
SH3GL1;SH3GL2	Guide only	PPIH	Guide only
HDLBP	Guide only	DNAJC2	Guide only
GLYR1	Guide only	LPCAT1	Guide only
LRBA	Guide only	CRNKL1	Guide only
KIF5C;KIF5B;KIF5A	Guide only	DDX47	Guide only
PRPF4	Guide only	POLR3A	Guide only
AP3S1	Guide only	KIF2A	Guide only
CNN2;CNN3	Guide only	CNP	Guide only
GLE1	Guide only	GPN1	Guide only
NEK9	Guide only	PNO1	Guide only
EXOSC9	Guide only	BYSL	Guide only
PLAA	Guide only	COPS2	Guide only
SMU1	Guide only	ELAC2	Guide only
LUC7L2	Guide only	OXR1;NCOA7	Guide only
CBX5	Guide only	CSTF1	Guide only
ELP3	Guide only	DDX18	Guide only
BAG6	Guide only	H2AFY2	Guide only
VPS4A;VPS4B;FIGNL1	Guide only	RBM39	Guide only
CSTF2;CSTF2T	Guide only	TRMT1L	Guide only
HSPA6;HSPA7	Guide only	LAS1L	Guide only
DDX52	Guide only	NOC2L	Guide only
EXOSC2	Guide only	SMARCC1	Guide only
SMARCA5	Guide only	U2AF1	Guide only
HTATSF1	Guide only	RBM12	Guide only
RABL3	Guide only	KPNA6;KPNA5	Guide only
CCAR1	Guide only	BCCIP	Guide only
VPS26B	Guide only	AP3M1	Guide only
PIN4	Guide only	TOMM34	Guide only
PSMF1	Guide only	EXOSC4	Guide only

Gene names	G4/dCas9	Gene names	G4/dCas9
SNRNP40	Guide only	SSRP1	12.9538694
SNRPC	Guide only	DDX21	12.05231447
WDR12	Guide only	GTF2I	9.062697465
SEC23IP	Guide only	NME1	8.963847644
UROD	Guide only	NOP2	8.819867183
SLC30A6	Guide only	PCBP2	7.307767838
RAD50	Guide only	NONO	7.303993005
SRSF1	Guide only	VIM	7.117729604
TRMT1	Guide only	HNRNPA0	6.80117722
HIST1H1D	Guide only	RSL1D1	6.550947041
AKAP11	Guide only	PRMT3	6.455090286
DNAJB1	Guide only	GTPBP4	6.416404191
KDM1A	Guide only	CPNE1	6.339276565
URB1	Guide only	MYBBP1A	6.076549446
MTA1;MTA3	Guide only	XRN2	5.558682194
POLD2	Guide only	SNRNP70	5.44811145
MBNL1	Guide only	RRP12	5.34502859
SAFB;SAFB2	Guide only	DDX20	5.18077667
SCAMP2	Guide only	TOP1	5.012676566
DDX56	Guide only	TSN	4.990251524
SCD	Guide only	MCM5	4.906417715
CEBPZ	Guide only	NHP2L1	4.898372304
DDX23	Guide only	H1F0	4.476610505
THEM6	Guide only	TMEM33	4.207450836
POLDIP3	Guide only	TOPBP1	4.147804155
LIMA1	Guide only		
SUPT6H	Guide only		
SRRT	Guide only		
ACAD8	Guide only		
CDC37	Guide only		
HSPA1L	Guide only		
RPRD1A	Guide only		
HMGB3	Guide only		
UBE2N;UBE2NL	Guide only		
AMK2D;CAMK2B;CAMK2A;CAMK	Guide only		
PRKACB;PRKACA	Guide only		
HNRNPK	Guide only		
KIAA1524	Guide only		
UBE3C	Guide only		
RRP1B	Guide only		
NT5C2	Guide only		
RRP1	Guide only		
RPS6KA3	Guide only		
DCAF13	Guide only		
DDX50	Guide only		
TAB1	Guide only		

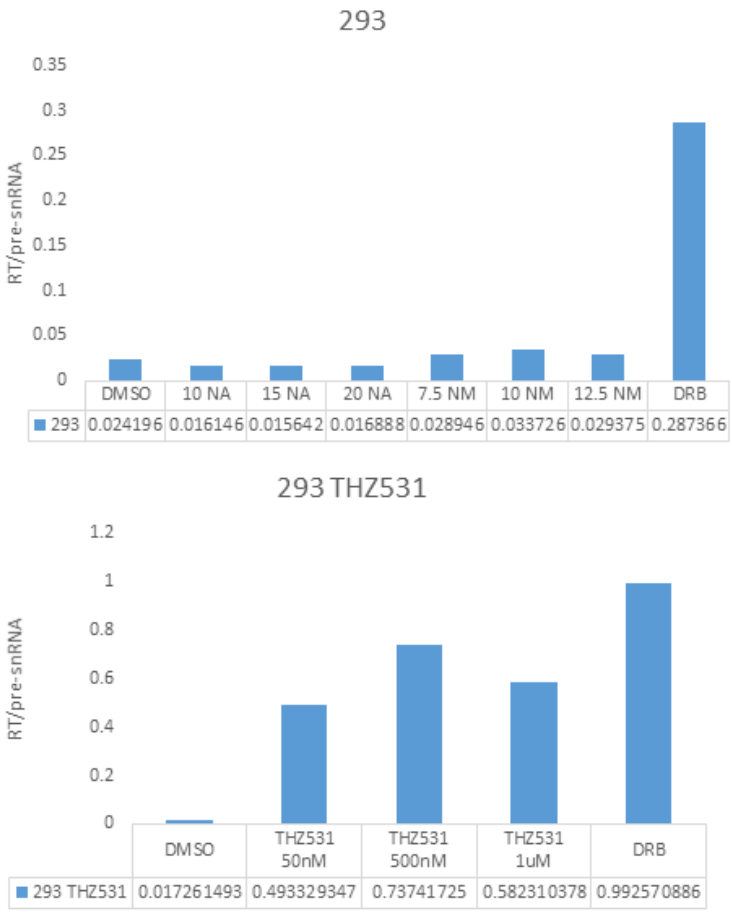


Figure S.5.1. Ratio of the RT/pre-snRNA transcripts on the HEK293 RNase protections.

On the top, RT/pre-snRNA ratios for the NA and NM tritation. On the bottom, Rt/pre-snRNA ratio for the HEK293 + THZ531 RNase protections.

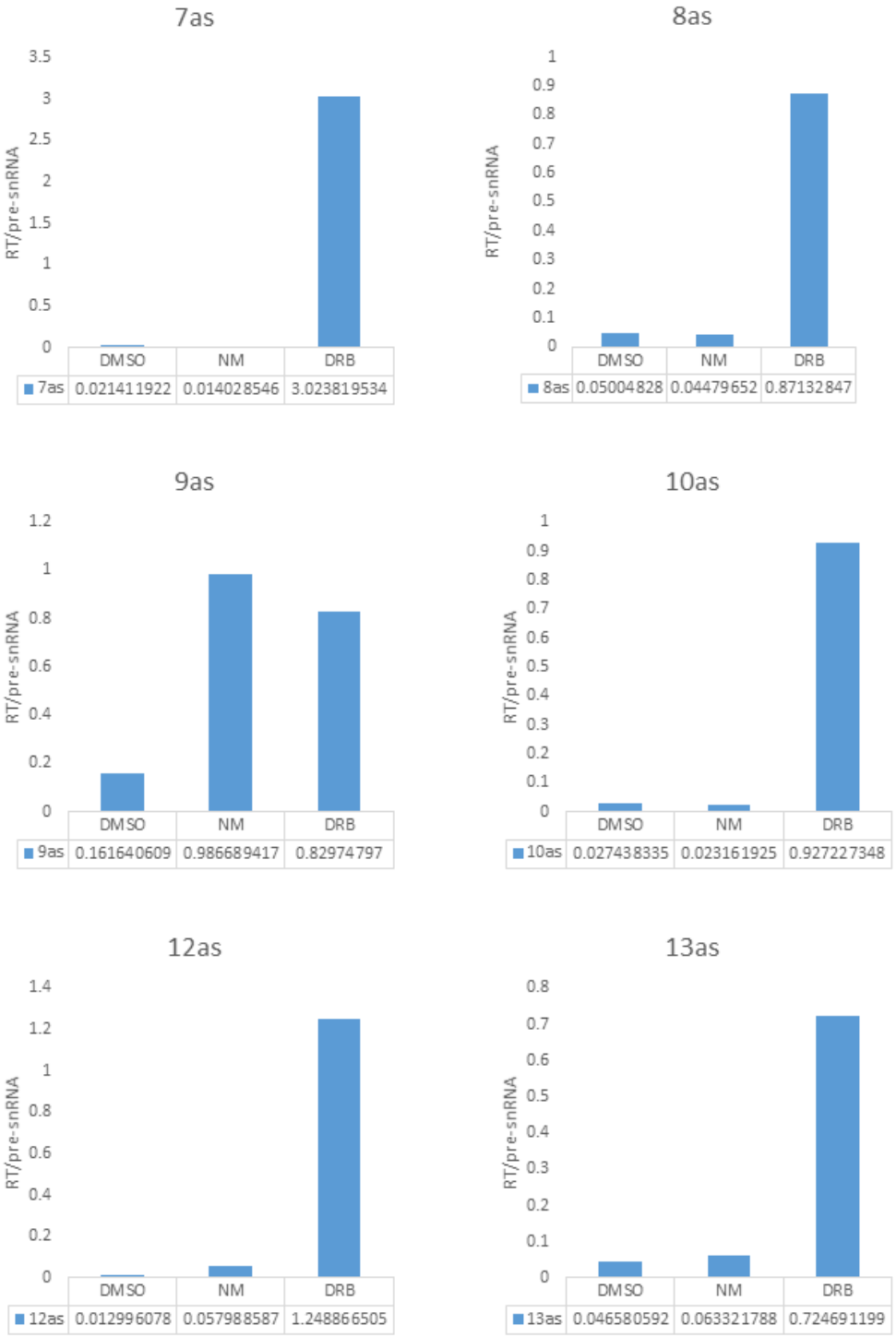


Figure S.5.2. Ratio of the RT/pre-snRNA transcripts on the 7, 8, 9, 10, 12 and 13 as RNase protections.

The RT/pre-snRNA ratios are stated under the bars of each graphs corresponding to 7, 8, 9, 10, 12 and 13 as.

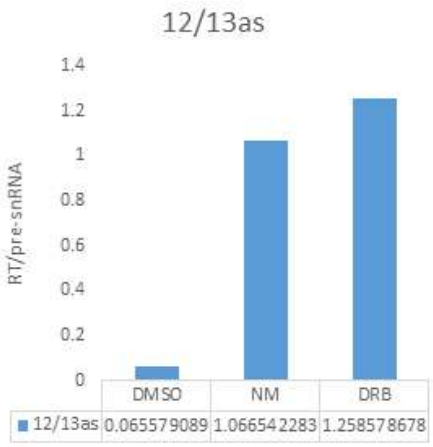


Figure. S.5.3 Ratio of the RT/pre-snRNA transcripts on the 12/13 as RNase protections.

The RT/pre-snRNA ratios are stated under the bars of the graph.

Table S.5.1. Int11 IP Mass spectrometry list in 12/13 as cells DMSO and NM conditions.
The ratio of the proteins by Int11 is shown for comparison. The values of CPSF2 and pol II are highlighted in red.

Gene names	12/12 DMSO IgG	12/13 DMSO Int11	12/13 DMSO Int11 ratio	12/13 NM IgG	12/13 NM Int11	12/13 NM Int11 ratio
ATAD3A	0	24	3.43	0	24	2.67
CPSF1	0	21	3.00	0	28	3.11
FIP1L1	0	9	1.29	0	9	1.00
VIM	1	23	3.29	0	19	2.11
WDR33	0	13	1.86	0	17	1.89
CPSF3	0	10	1.43	0	11	1.22
CPSF2	0	18	2.57	0	16	1.78
CPSF4	0	5	0.71	0	6	0.67
RCN2	0	5	0.71	0	4	0.44
RELA	0	6	0.86	0	7	0.78
DDX20	0	8	1.14	0	9	1.00
INTS1	0	12	1.71	0	14	1.56
HNRNPC	0	4	0.57	0	5	0.56
GEMIN4	0	9	1.29	0	10	1.11
CPSF3L	0	7	1.00	0	9	1.00
P4HA1	0	6	0.86	0	10	1.11
P4HB	0	6	0.86	0	7	0.78
INTS9	0	8	1.14	0	10	1.11
INTS4	0	10	1.43	0	10	1.11
HIST1H2AH	1	2	0.29	1	2	0.22
INTS7	0	6	0.86	0	8	0.89
POLR2A	0	1	0.14	0	10	1.11
HIST1H2BL	1	2	0.29	1	3	0.33
YBX3	0	2	0.29	0	3	0.33
PABPC1;PABPC3	0	4	0.57	0	5	0.56
UFL1	0	3	0.43	0	8	0.89
HNRNPK	0	2	0.29	0	3	0.33
RPSA	0	2	0.29	1	3	0.33
SNRNP70	0	6	0.86	0	6	0.67
SYMPK	0	2	0.29	0	5	0.56
INTS3	0	4	0.57	0	4	0.44
RPS3	0	5	0.71	1	5	0.56
IKZF3	1	0	0.00	0	1	0.11
NAV1	0	3	0.43	0	7	0.78
SRRM2	0	3	0.43	0	5	0.56
INTS2	0	3	0.43	0	4	0.44
HNRNPM	0	2	0.29	0	4	0.44
CXorf56	0	3	0.43	0	2	0.22
RPLP2	0	2	0.29	1	2	0.22
NEFM	0	5	0.71	0	4	0.44
RPS20	0	0	0.00	0	3	0.33
C2CD2L	0	0	0.00	0	1	0.11
GNB2L1	0	2	0.29	0	3	0.33
INTS12	0	2	0.29	0	2	0.22
RPS8	0	3	0.43	0	2	0.22

HSPA6;HSPA7	2	2	0.29	2	5	0.56
INTS6	0	3	0.43	0	4	0.44
HNRNPH1	0	0	0.00	0	3	0.33
SNRPD3	0	1	0.14	0	1	0.11
HSPA5	1	2	0.29	2	6	0.67
RPS25	0	2	0.29	0	3	0.33
RPS14	0	1	0.14	0	3	0.33
SNRPB;SNRPN	0	2	0.29	0	2	0.22
RPL18	0	3	0.43	0	3	0.33
CTBP1	0	1	0.14	0	2	0.22
RPL11	0	2	0.29	0	2	0.22
H3F3C;HIST2H3A	0	0	0.00	0	1	0.11
SLC25A5;SLC25A4	0	2	0.29	0	1	0.11
DDRGK1	0	2	0.29	0	3	0.33
STRAP	0	2	0.29	0	3	0.33
SMN1	0	0	0.00	0	2	0.22
RPS18	0	3	0.43	0	3	0.33
POLR2H	0	1	0.14	0	1	0.11
SNRPD2	0	0	0.00	0	1	0.11
IGF2BP1	0	3	0.43	0	3	0.33
POLR2E	0	0	0.00	0	2	0.22
HSPD1	0	5	0.71	0	4	0.44
RPL7A	0	2	0.29	0	5	0.56
POLR2B	0	1	0.14	0	4	0.44
RPL38	0	0	0.00	0	1	0.11
RPL4	0	1	0.14	0	2	0.22
RPS19	0	1	0.14	0	3	0.33
RPL22	0	1	0.14	0	2	0.22
S100A8	0	1	0.14	1	1	0.11
HSP90AB1;HSP90AA1	0	1	0.14	2	1	0.11
RPL27	0	1	0.14	0	1	0.11
RPL7	0	0	0.00	0	1	0.11
MAGOH	0	0	0.00	0	1	0.11
RPL13	0	0	0.00	0	1	0.11
CACTIN	0	1	0.14	0	2	0.22
DDX17	0	1	0.14	0	1	0.11
EEF1A1P5;EEF1A1	2	3	0.43	1	2	0.22
GEMIN6	0	1	0.14	0	2	0.22
PUF60	0	0	0.00	1	1	0.11
RPS4X;RPS4Y2;RPS4Y	0	0	0.00	0	2	0.22
AZGP1	0	1	0.14	1	1	0.11
INTS8	0	1	0.14	0	2	0.22
SRSF7	0	0	0.00	0	2	0.22
GSDMA	1	0	0.00	0	2	0.22
RPS24	0	0	0.00	0	1	0.11
POLDIP3	0	0	0.00	0	2	0.22
SEMG1;SEMG2	0	0	0.00	0	1	0.11
RIF1	0	1	0.14	0	2	0.22

GEMIN2	0	0	0.00	0	1	0.11
RPL31	0	0	0.00	0	1	0.11
DDX41	0	1	0.14	0	2	0.22
RPL12	0	0	0.00	0	1	0.11
CSTF2	0	1	0.14	0	1	0.11
TRA2B	0	1	0.14	0	1	0.11
RPLP0;RPLP0P6	0	1	0.14	0	1	0.11
SNRPB2	0	0	0.00	0	1	0.11
INTS5	0	1	0.14	0	1	0.11
RPL10A	0	1	0.14	0	1	0.11
GEMIN7	0	0	0.00	0	1	0.11
RPL27A	0	1	0.14	0	1	0.11
RPL35	0	0	0.00	0	1	0.11
RPL14	0	1	0.14	0	1	0.11
RPL3	0	1	0.14	0	1	0.11
PCMT1	0	0	0.00	1	2	0.22
PRDX1	1	1	0.14	1	1	0.11
ENO1	0	1	0.14	2	1	0.11
RPL23A	0	1	0.14	0	1	0.11
RPS9	0	1	0.14	0	2	0.22
ZG16B	0	1	0.14	0	1	0.11
RPL6	0	2	0.29	0	2	0.22
MRI1	1	1	0.14	0	1	0.11
HIST1H1C	0	0	0.00	0	1	0.11
POLR2J3;POLR2J2	0	0	0.00	0	1	0.11
SRSF10	0	0	0.00	0	1	0.11
HNRNPU	0	1	0.14	0	1	0.11
DDX26B	0	2	0.29	0	1	0.11
SUB1	0	1	0.14	1	1	0.11
RPS23	0	1	0.14	0	1	0.11
FXR1	0	1	0.14	0	1	0.11
SNRPA1	0	0	0.00	0	1	0.11
RPS15A	0	0	0.00	0	1	0.11
DDX39B	0	1	0.14	0	1	0.11
RPL35A	0	0	0.00	0	1	0.11
CAT	1	0	0.00	0	1	0.11
PPP2R1A	0	1	0.14	0	2	0.22
POLR2G	0	0	0.00	0	1	0.11
RPL26;RPL26L1	0	0	0.00	0	1	0.11
AP1B1;AP2B1	0	0	0.00	1	1	0.11
ZC3H11A	0	0	0.00	0	1	0.11
ATAD3B	0	16	2.29	0	15	1.67
HNRNPC	0	4	0.57	0	4	0.44
DDX5	0	2	0.29	0	1	0.11
YBX1;YBX2	0	0	0.00	0	1	0.11
RPS13	0	0	0.00	0	1	0.11
HNRNPF	0	0	0.00	0	1	0.11
DDX39A	0	1	0.14	0	1	0.11

MAGOHB	0	0	0.00	0	1	0.11
CTBP2	0	1	0.14	0	1	0.11
PABPC4	0	1	0.14	0	1	0.11
IGHG1;IGHG3;IGHG2	0	1	0.14	0	0	0.00
RBM8A	0	1	0.14	0	0	0.00
PRPF19	0	1	0.14	0	0	0.00
ALYREF	0	0	0.00	0	0	0.00
ERH	0	0	0.00	0	0	0.00
PRSS3	0	0	0.00	0	0	0.00
EIF3A	0	1	0.14	0	0	0.00
FAM214A	1	0	0.00	1	0	0.00
SMU1	0	0	0.00	1	0	0.00
TBC1D10C	0	1	0.14	0	0	0.00
RPS6	0	2	0.29	0	0	0.00
HSPB1	1	0	0.00	0	0	0.00
HNRNPA2B1	0	1	0.14	0	0	0.00
F10	1	0	0.00	0	0	0.00
PKM	1	0	0.00	1	0	0.00
LYZ	0	1	0.14	1	0	0.00
RPL19	0	1	0.14	0	0	0.00
RPL8	0	0	0.00	0	0	0.00
CALML5	0	0	0.00	1	0	0.00
SERPINB12	1	0	0.00	1	0	0.00
RPL10L;RPL10	0	0	0.00	0	0	0.00
LCN1P1;LCN1	1	0	0.00	1	0	0.00
RPS7	0	1	0.14	0	0	0.00
RPL15	0	1	0.14	0	0	0.00
HBA1	1	0	0.00	1	0	0.00
LDHA;LDHAL6A;LDHC	0	0	0.00	0	0	0.00
POLR2C	0	1	0.14	0	0	0.00
IGKV2D-30;IGKV2D-2	0	1	0.14	0	0	0.00
EEF2	0	1	0.14	0	0	0.00
CTSD	1	0	0.00	0	0	0.00
PLOD2	0	1	0.14	0	0	0.00
SLIT2	1	0	0.00	1	0	0.00
AIFM1	0	0	0.00	0	0	0.00
PKIA	0	0	0.00	0	0	0.00
LRRC10	0	0	0.00	1	0	0.00
EXOC8	0	0	0.00	1	0	0.00
SHROOM3	0	0	0.00	0	0	0.00
GPN1	0	1	0.14	0	0	0.00

Table S.5.2. Int11 IP Mass spectrometry list in HEK293 cells DMSO and NM conditions.
The ratio of the proteins by Int11 is shown for comparison.

Gene names	293 DMSO IgG	293 DMSO Int11	293 DMSO Int11 ratio	293 NM IgG	293 NM Int11	293 NM Int11 ratio
ATAD3A	0	24	2.67	0	26	3.71
CPSF1	0	20	2.22	0	22	3.14
FIP1L1	0	7	0.78	0	8	1.14
VIM	1	11	1.22	1	14	2.00
WDR33	0	10	1.11	0	12	1.71
CPSF3	0	9	1.00	0	8	1.14
CPSF2	0	13	1.44	0	13	1.86
CPSF4	0	4	0.44	0	3	0.43
RCN2	0	4	0.44	0	4	0.57
RELA	0	7	0.78	0	6	0.86
DDX20	0	6	0.67	0	5	0.71
INTS1	0	16	1.78	0	20	2.86
HNRNPC	0	1	0.11	0	1	0.14
GEMIN4	0	3	0.33	0	4	0.57
CPSF3L	0	9	1.00	0	7	1.00
P4HA1	0	3	0.33	0	3	0.43
P4HB	0	5	0.56	0	5	0.71
INTS9	0	6	0.67	0	4	0.57
INTS4	0	9	1.00	0	9	1.29
HIST1H2AH	0	1	0.11	0	1	0.14
INTS7	0	6	0.67	0	5	0.71
POLR2A	0	3	0.33	0	5	0.71
HIST1H2BL	0	2	0.22	0	1	0.14
YBX3	0	1	0.11	0	2	0.29
PABPC1;PABPC3	0	2	0.22	0	3	0.43
UFL1	0	1	0.11	0	7	1.00
HNRNPK	0	2	0.22	0	1	0.14
RPSA	0	2	0.22	0	3	0.43
SNRNP70	0	3	0.33	0	4	0.57
SYMPK	0	1	0.11	0	0	0.00
INTS3	0	3	0.33	0	2	0.29
RPS3	0	3	0.33	0	4	0.57
IKZF3	0	0	0.00	0	0	0.00
NAV1	0	3	0.33	0	5	0.71
SRRM2	0	1	0.11	0	5	0.71
INTS2	0	2	0.22	0	2	0.29
HNRNPM	0	0	0.00	0	0	0.00
CXorf56	0	0	0.00	0	1	0.14
RPLP2	0	2	0.22	0	1	0.14
NEFM	0	0	0.00	0	0	0.00
RPS20	0	0	0.00	0	2	0.29
C2CD2L	0	0	0.00	0	0	0.00
GNB2L1	0	3	0.33	0	3	0.43
INTS12	0	1	0.11	0	1	0.14
RPS8	0	2	0.22	0	1	0.14

HSPA6;HSPA7	1	2	0.22	1	5	0.71
INTS6	0	2	0.22	0	3	0.43
HNRNPH1	0	2	0.22	0	1	0.14
SNRPD3	0	1	0.11	0	1	0.14
HSPA5	1	1	0.11	0	4	0.57
RPS25	0	0	0.00	0	0	0.00
RPS14	0	1	0.11	0	1	0.14
SNRPB;SNRPN	0	1	0.11	0	2	0.29
RPL18	0	2	0.22	0	2	0.29
CTBP1	0	1	0.11	0	2	0.29
RPL11	0	0	0.00	0	0	0.00
H3F3C;HIST2H3A	0	0	0.00	0	0	0.00
SLC25A5;SLC25A4	0	1	0.11	0	1	0.14
DDRGK1	0	1	0.11	0	2	0.29
STRAP	0	1	0.11	0	1	0.14
SMN1	0	2	0.22	0	1	0.14
RPS18	0	3	0.33	0	3	0.43
POLR2H	0	1	0.11	0	0	0.00
SNRPD2	0	1	0.11	0	2	0.29
IGF2BP1	0	1	0.11	0	1	0.14
POLR2E	0	0	0.00	0	0	0.00
HSPD1	0	4	0.44	0	2	0.29
RPL7A	0	1	0.11	0	2	0.29
POLR2B	0	1	0.11	0	2	0.29
RPL38	0	0	0.00	0	0	0.00
RPL4	0	0	0.00	0	0	0.00
RPS19	0	0	0.00	0	1	0.14
RPL22	0	1	0.11	0	1	0.14
S100A8	0	0	0.00	0	1	0.14
HSP90AB1;HSP90AA1	1	0	0.00	0	0	0.00
RPL27	0	0	0.00	0	1	0.14
RPL7	0	1	0.11	0	1	0.14
MAGOH	0	0	0.00	0	0	0.00
RPL13	0	0	0.00	0	0	0.00
CACTIN	0	2	0.22	0	2	0.29
DDX17	0	1	0.11	0	1	0.14
EEF1A1P5;EEF1A1	1	1	0.11	0	2	0.29
GEMIN6	0	0	0.00	0	0	0.00
PUF60	0	0	0.00	0	0	0.00
RPS4X;RPS4Y2;RPS4Y	0	0	0.00	0	0	0.00
AZGP1	0	1	0.11	1	0	0.00
INTS8	0	2	0.22	0	1	0.14
SRSF7	0	0	0.00	0	0	0.00
GSDMA	0	0	0.00	0	0	0.00
RPS24	0	0	0.00	0	0	0.00
POLDIP3	0	0	0.00	0	0	0.00
SEMG1;SEMG2	0	0	0.00	0	0	0.00
RIF1	0	0	0.00	0	1	0.14

GEMIN2	0	2	0.22	0	0	0.00
RPL31	0	0	0.00	0	0	0.00
DDX41	0	0	0.00	0	0	0.00
RPL12	0	0	0.00	0	1	0.14
CSTF2	0	0	0.00	0	2	0.29
TRA2B	0	0	0.00	0	0	0.00
RPLP0;RPLP0P6	0	1	0.11	0	1	0.14
SNRPB2	0	0	0.00	0	0	0.00
INTS5	0	1	0.11	0	2	0.29
RPL10A	0	0	0.00	0	1	0.14
GEMIN7	0	0	0.00	0	1	0.14
RPL27A	0	0	0.00	0	1	0.14
RPL35	0	0	0.00	0	0	0.00
RPL14	0	1	0.11	0	1	0.14
RPL3	0	1	0.11	0	2	0.29
PCMT1	0	1	0.11	0	2	0.29
PRDX1	1	1	0.11	1	1	0.14
ENO1	0	0	0.00	0	0	0.00
RPL23A	0	0	0.00	0	1	0.14
RPS9	0	0	0.00	0	0	0.00
ZG16B	0	0	0.00	0	0	0.00
RPL6	0	1	0.11	0	0	0.00
MRI1	1	1	0.11	1	1	0.14
HIST1H1C	0	1	0.11	0	0	0.00
POLR2J3;POLR2J2	0	0	0.00	0	0	0.00
SRSF10	0	0	0.00	0	0	0.00
HNRNPU	0	0	0.00	0	1	0.14
DDX26B	0	0	0.00	0	0	0.00
SUB1	0	0	0.00	0	1	0.14
RPS23	0	0	0.00	0	0	0.00
FXR1	0	0	0.00	0	0	0.00
SNRPA1	0	0	0.00	0	1	0.14
RPS15A	0	0	0.00	0	1	0.14
DDX39B	0	0	0.00	0	0	0.00
RPL35A	0	0	0.00	0	0	0.00
CAT	0	1	0.11	0	0	0.00
PPP2R1A	0	0	0.00	0	1	0.14
POLR2G	0	0	0.00	0	1	0.14
RPL26;RPL26L1	0	0	0.00	0	0	0.00
AP1B1;AP2B1	0	0	0.00	0	0	0.00
ZC3H11A	0	0	0.00	0	0	0.00
ATAD3B	0	16	1.78	0	19	2.71
HNRNPC	0	1	0.11	0	1	0.14
DDX5	0	1	0.11	0	1	0.14
YBX1;YBX2	0	1	0.11	0	1	0.14
RPS13	0	0	0.00	0	0	0.00
HNRNPF	0	1	0.11	0	0	0.00
DDX39A	0	0	0.00	0	0	0.00

Table S.5.3. List of proteins that changed the phosphorylation status under inhibition of CDK12/13 in Int11 IP mass spectrometry list in 12/13 as cells NM conditions.

No change was seen in HEK293 or 12/13 as cells DMSO.

Gene names	Amino acid	Sequence window
CPSF4	S	TQPPAKQSNNPPLQRSSSLIQLTSQNSSPNQ
CPSF4	S	QPPAKQSNNPPLQRSSSLIQLTSQNSSPNQQ
CPSF4	S	PPAKQSNNPPLQRSSSLIQLTSQNSSPNQQR
WDR33	S	_____MATEIGSPPRFFHMPRFQHQAP

References

1. Grummt I. Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus. Genes & development. 2003;17(14):1691-702.
2. Russell J, Zomerdijk JC. RNA-polymerase-I-directed rDNA transcription, life and works. Trends in biochemical sciences. 2005;30(2):87-96.
3. Jawdekar GW, Henry RW. Transcriptional regulation of human small nuclear RNA genes. Biochimica et biophysica acta. 2008;1779(5):295-305.
4. White RJ. Transcription by RNA polymerase III: more complex than we thought. Nature reviews Genetics. 2011;12(7):459-63.
5. Shi X, Sun M, Liu H, Yao Y, Song Y. Long non-coding RNAs: a new frontier in the study of human diseases. Cancer Lett. 2013;339(2):159-66.
6. Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, et al. MicroRNA genes are transcribed by RNA polymerase II. The EMBO journal. 2004;23(20):4051-60.
7. Matera AG, Terns RM, Terns MP. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. Nature reviews Molecular cell biology. 2007;8(3):209-20.
8. Tarn WY, Steitz JA. Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge. Trends in biochemical sciences. 1997;22(4):132-7.
9. Marzluff WF, Wagner EJ, Duronio RJ. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nature reviews Genetics. 2008;9(11):843-54.
10. Culver GM. Sno-capped: 5' ends of preribosomal RNAs are decorated with a U3 SnoRNP. Chemistry & biology. 2002;9(7):777-9.
11. Egloff S, O'Reilly D, Murphy S. Expression of human snRNA genes from beginning to end. Biochemical Society transactions. 2008;36(Pt 4):590-4.
12. Hernandez N. Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. The Journal of biological chemistry. 2001;276(29):26733-6.
13. Hung KH, Stumph WE. Regulation of snRNA gene expression by the Drosophila melanogaster small nuclear RNA activating protein complex (DmSNAPc). Critical reviews in biochemistry and molecular biology. 2011;46(1):11-26.
14. O'Reilly D, Dienstbier M, Cowley SA, Vazquez P, Drozdz M, Taylor S, et al. Differentially expressed, variant U1 snRNAs regulate gene expression in human cells. Genome research. 2013;23(2):281-91.
15. Lindgren V, Ares M, Jr., Weiner AM, Francke U. Human genes for U2 small nuclear RNA map to a major adenovirus 12 modification site on chromosome 17. Nature. 1985;314(6006):115-6.
16. Marz M, Kirsten T, Stadler PF. Evolution of spliceosomal snRNA genes in metazoan animals. Journal of molecular evolution. 2008;67(6):594-607.
17. de Vegvar HE, Lund E, Dahlberg JE. 3' end formation of U1 snRNA precursors is coupled to transcription from snRNA promoters. Cell. 1986;47(2):259-66.
18. Hernandez N, Weiner AM. Formation of the 3' end of U1 snRNA requires compatible snRNA promoter elements. Cell. 1986;47(2):249-58.
19. Murphy S, Yoon JB, Gerster T, Roeder RG. Oct-1 and Oct-2 potentiate functional interactions of a transcription factor with the proximal sequence element of small nuclear RNA genes. Molecular and cellular biology. 1992;12(7):3247-61.
20. Janson L, Bark C, Pettersson U. Identification of proteins interacting with the enhancer of human U2 small nuclear RNA genes. Nucleic acids research. 1987;15(13):4997-5016.
21. Schaub M, Myslinski E, Schuster C, Krol A, Carbon P. Staf, a promiscuous activator for enhanced transcription by RNA polymerases II and III. The EMBO journal. 1997;16(1):173-81.

22. Janson L, Weller P, Pettersson U. Nuclear factor I can functionally replace transcription factor Sp1 in a U2 small nuclear RNA gene enhancer. Journal of molecular biology. 1989;205(2):387-96.
23. Sadowski CL, Henry RW, Lobo SM, Hernandez N. Targeting TBP to a non-TATA box cis-regulatory element: a TBP-containing complex activates transcription from snRNA promoters through the PSE. Genes & development. 1993;7(8):1535-48.
24. Waldschmidt R, Wanandi I, Seifart KH. Identification of transcription factors required for the expression of mammalian U6 genes in vitro. The EMBO journal. 1991;10(9):2595-603.
25. Henry RW, Sadowski CL, Kobayashi R, Hernandez N. A TBP-TAF complex required for transcription of human snRNA genes by RNA polymerase II and III. Nature. 1995;374(6523):653-6.
26. Murphy S. Differential in vivo activation of the class II and class III snRNA genes by the POU-specific domain of Oct-1. Nucleic acids research. 1997;25(11):2068-76.
27. Christova R, Oelgeschlager T. Association of human TFIID-promoter complexes with silenced mitotic chromatin in vivo. Nature cell biology. 2002;4(1):79-82.
28. 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.
29. Bieniossek C, Papai G, Schaffitzel C, Garzoni F, Chaillet M, Scheer E, et al. The architecture of human general transcription factor TFIID core complex. Nature. 2013;493(7434):699-702.
30. Takahashi H, Takigawa I, Watanabe M, Anwar D, Shibata M, Tomomori-Sato C, et al. MED26 regulates the transcription of snRNA genes through the recruitment of little elongation complex. Nature communications. 2015;6:5941.
31. Kuhlman TC, Cho H, Reinberg D, Hernandez N. The general transcription factors IIA, IIB, IIF, and IIE are required for RNA polymerase II transcription from the human U1 small nuclear RNA promoter. Molecular and cellular biology. 1999;19(3):2130-41.
32. Glover-Cutter K, Larochelle S, Erickson B, Zhang C, Shokat K, Fisher RP, et al. TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. Molecular and cellular biology. 2009;29(20):5455-64.
33. Zhao X, Pendergrast PS, Hernandez N. A positioned nucleosome on the human U6 promoter allows recruitment of SNAPc by the Oct-1 POU domain. Molecular cell. 2001;7(3):539-49.
34. Boyd DC, Greger IH, Murphy S. In vivo footprinting studies suggest a role for chromatin in transcription of the human 7SK gene. Gene. 2000;247(1-2):33-44.
35. Boyd DC, Pombo A, Murphy S. Interaction of proteins with promoter elements of the human U2 snRNA genes in vivo. Gene. 2003;315:103-12.
36. Stunkel W, Kober I, Seifart KH. A nucleosome positioned in the distal promoter region activates transcription of the human U6 gene. Molecular and cellular biology. 1997;17(8):4397-405.
37. O'Reilly D, Kuznetsova OV, Laitem C, Zaborowska J, Dienstbier M, Murphy S. Human snRNA genes use polyadenylation factors to promote efficient transcription termination. Nucleic acids research. 2014;42(1):264-75.
38. Yuan CC, Zhao X, Florens L, Swanson SK, Washburn MP, Hernandez N. CHD8 associates with human Staf and contributes to efficient U6 RNA polymerase III transcription. Molecular and cellular biology. 2007;27(24):8729-38.
39. Allen BL, Taatjes DJ. The Mediator complex: a central integrator of transcription. Nature reviews Molecular cell biology. 2015;16(3):155-66.
40. Lariviere L, Plaschka C, Seizl M, Wenzeck L, Kurth F, Cramer P. Structure of the Mediator head module. Nature. 2012;492(7429):448-51.
41. Soutourina J, Wydau S, Ambroise Y, Boschiero C, Werner M. Direct interaction of RNA polymerase II and mediator required for transcription in vivo. Science. 2011;331(6023):1451-4.

42. Seizl M, Lariviere L, Pfaffeneder T, Wenzeck L, Cramer P. Mediator head subcomplex Med11/22 contains a common helix bundle building block with a specific function in transcription initiation complex stabilization. Nucleic acids research. 2011;39(14):6291-304.
43. Lariviere L, Geiger S, Hoeppner S, Rother S, Strasser K, Cramer P. Structure and TBP binding of the Mediator head subcomplex Med8-Med18-Med20. Nature structural & molecular biology. 2006;13(10):895-901.
44. Egloff S, O'Reilly D, Chapman RD, Taylor A, Tanzhaus K, Pitts L, et al. Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. Science. 2007;318(5857):1777-9.
45. Akhtar MS, Heidemann M, Tietjen JR, Zhang DW, Chapman RD, Eick D, et al. TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. Molecular cell. 2009;34(3):387-93.
46. 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. The EMBO journal. 2005;24(23):4154-65.
47. Baillat D, Hakimi MA, Naar 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):265-76.
48. Wada T, Takagi T, Yamaguchi Y, Watanabe D, Handa H. Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription in vitro. The EMBO journal. 1998;17(24):7395-403.
49. Yamaguchi Y, Takagi T, Wada T, Yano K, Furuya A, Sugimoto S, et al. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell. 1999;97(1):41-51.
50. Cuello P, Boyd DC, Dye MJ, Proudfoot NJ, Murphy S. Transcription of the human U2 snRNA genes continues beyond the 3' box in vivo. The EMBO journal. 1999;18(10):2867-77.
51. Smith ER, Lin C, Garrett AS, Thornton J, Mohaghegh N, Hu D, et al. The little elongation complex regulates small nuclear RNA transcription. Molecular cell. 2011;44(6):954-65.
52. Smith E, Lin C, Shilatifard A. The super elongation complex (SEC) and MLL in development and disease. Genes & development. 2011;25(7):661-72.
53. Hutten S, Chachami G, Winter U, Melchior F, Lamond AI. A role for the Cajal-body-associated SUMO isopeptidase USPL1 in snRNA transcription mediated by RNA polymerase II. Journal of cell science. 2014;127(Pt 5):1065-78.
54. Chen J, Ezzeddine N, Waltenspiel B, Albrecht TR, Warren WD, Marzluff WF, 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):2148-56.
55. Egloff S, Al-Rawaf H, O'Reilly D, Murphy S. Chromatin structure is implicated in "late" elongation checkpoints on the U2 snRNA and beta-actin genes. Molecular and cellular biology. 2009;29(14):4002-13.
56. Yamamoto J, Hagiwara Y, Chiba K, Isobe T, Narita T, Handa H, et al. DSIF and NELF interact with Integrator to specify the correct post-transcriptional fate of snRNA genes. Nature communications. 2014;5:4263.
57. Gardini A, Baillat D, Cesaroni M, Hu D, Marinis JM, Wagner EJ, et al. Integrator regulates transcriptional initiation and pause release following activation. Molecular cell. 2014;56(1):128-39.
58. Stadelmayer B, Micas G, Gamot A, Martin P, Malirat N, Koval S, et al. Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. Nature communications. 2014;5:5531.
59. Skaar JR, Ferris AL, Wu X, Saraf A, Khanna KK, Florens L, et al. The Integrator complex controls the termination of transcription at diverse classes of gene targets. Cell research. 2015;25(3):288-305.

60. Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, et al. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature. 2011;479(7371):74-9.
61. Laitem C, Zaborowska J, Tellier M, Yamaguchi Y, Cao Q, Egloff S, et al. CTCF regulates NELF, DSIF and P-TEFb recruitment during transcription. Transcription. 2015;6(5):79-90.
62. Nguyen VT, Kiss T, Michels AA, Bensaude O. 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. Nature. 2001;414(6861):322-5.
63. Yang F, Hsu P, Lee SD, Yang W, Hoskinson D, Xu W, et al. The C terminus of Pcf11 forms a novel zinc-finger structure that plays an essential role in mRNA 3'-end processing. Rna. 2017;23(1):98-107.
64. Lee TI, Young RA. Transcription of eukaryotic protein-coding genes. Annual review of genetics. 2000;34:77-137.
65. Marshall NF, Peng J, Xie Z, Price DH. Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. The Journal of biological chemistry. 1996;271(43):27176-83.
66. Hu D, Smith ER, Garruss AS, Mohaghegh N, Varberg JM, Lin C, et al. The little elongation complex functions at initiation and elongation phases of snRNA gene transcription. Molecular cell. 2013;51(4):493-505.
67. Anwar D, Takahashi H, Watanabe M, Suzuki M, Fukuda S, Hatakeyama S. p53 represses the transcription of snRNA genes by preventing the formation of little elongation complex. Biochimica et biophysica acta. 2016;1859(8):975-82.
68. Egloff S, Vitali P, Tellier M, Raffel R, Murphy S, Kiss T. The 7SK snRNP associates with the little elongation complex to promote snRNA gene expression. The EMBO journal. 2017.
69. Jonkers I, Lis JT. Getting up to speed with transcription elongation by RNA polymerase II. Nature reviews Molecular cell biology. 2015;16(3):167-77.
70. 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. The Journal of biological chemistry. 2010;285(27):20564-9.
71. Weiner AM. E Pluribus Unum: 3' end formation of polyadenylated mRNAs, histone mRNAs, and U snRNAs. Molecular cell. 2005;20(2):168-70.
72. Burglin TR, Affolter M. Homeodomain proteins: an update. Chromosoma. 2016;125(3):497-521.
73. Kheirallah AK, de Moor CH, Faiz A, Sayers I, Hall IP. Lung function associated gene Integrator Complex subunit 12 regulates protein synthesis pathways. BMC Genomics. 2017;18(1):248.
74. Skaar JR, Richard DJ, Saraf A, Toschi A, Bolderson E, Florens L, et al. INTS3 controls the hSSB1-mediated DNA damage response. The Journal of cell biology. 2009;187(1):25-32.
75. Li Y, Bolderson E, Kumar R, Muniandy PA, Xue Y, Richard DJ, et al. HSSB1 and hSSB2 form similar multiprotein complexes that participate in DNA damage response. The Journal of biological chemistry. 2009;284(35):23525-31.
76. Zhang F, Ma T, Yu X. A core hSSB1-INTS complex participates in the DNA damage response. Journal of cell science. 2013;126(Pt 21):4850-5.
77. Huang J, Gong Z, Ghosal G, Chen J. SOSS complexes participate in the maintenance of genomic stability. Molecular cell. 2009;35(3):384-93.
78. Gu P, Deng W, Lei M, Chang S. Single strand DNA binding proteins 1 and 2 protect newly replicated telomeres. Cell research. 2013;23(5):705-19.
79. Baillat D, Wagner EJ. Integrator: surprisingly diverse functions in gene expression. Trends in biochemical sciences. 2015;40(5):257-64.
80. Egloff S, Zaborowska J, Laitem C, Kiss T, Murphy S. Ser7 phosphorylation of the CTD recruits the RPAP2 Ser5 phosphatase to snRNA genes. Molecular cell. 2012;45(1):111-22.

81. Ong CT, Corces VG. CTCF: an architectural protein bridging genome topology and function. Nature reviews Genetics. 2014;15(4):234-46.
82. 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 genetics. 2008;4(7):e1000138.
83. Neuman de Vegvar HE, Dahlberg JE. Initiation and termination of human U1 RNA transcription requires the concerted action of multiple flanking elements. Nucleic acids research. 1989;17(22):9305-18.
84. Guiro J, O'Reilly D. Insights into the U1 small nuclear ribonucleoprotein complex superfamily. Wiley interdisciplinary reviews RNA. 2015;6(1):79-92.
85. Marzluff WF, Brown DT, Lobo S, Wang SS. Isolation and characterization of two linked mouse U1b small nuclear RNA genes. Nucleic acids research. 1983;11(18):6255-70.
86. Lund E, Kahan B, Dahlberg JE. Differential control of U1 small nuclear RNA expression during mouse development. Science. 1985;229(4719):1271-4.
87. Sontheimer EJ, Steitz JA. Three novel functional variants of human U5 small nuclear RNA. Molecular and cellular biology. 1992;12(2):734-46.
88. Chen L, Lullo DJ, Ma E, Celniker SE, Rio DC, Doudna JA. Identification and analysis of U5 snRNA variants in Drosophila. Rna. 2005;11(10):1473-7.
89. Kim JL, Nikolov DB, Burley SK. Co-crystal structure of TBP recognizing the minor groove of a TATA element. Nature. 1993;365(6446):520-7.
90. Kim Y, Geiger JH, Hahn S, Sigler PB. Crystal structure of a yeast TBP/TATA-box complex. Nature. 1993;365(6446):512-20.
91. Wright KJ, Marr MT, 2nd, Tjian R. TAF4 nucleates a core subcomplex of TFIID and mediates activated transcription from a TATA-less promoter. Proc Natl Acad Sci U S A. 2006;103(33):12347-52.
92. Cavallini B, Faus I, Matthes H, Chipoulet JM, Winsor B, Egly JM, et al. Cloning of the gene encoding the yeast protein BTF1Y, which can substitute for the human TATA box-binding factor. Proc Natl Acad Sci U S A. 1989;86(24):9803-7.
93. Dynlacht BD, Hoey T, Tjian R. Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. Cell. 1991;66(3):563-76.
94. Matsui T, Segall J, Weil PA, Roeder RG. Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. J Biol Chem. 1980;255(24):11992-6.
95. Parker CS, Topol J. A Drosophila RNA polymerase II transcription factor binds to the regulatory site of an hsp 70 gene. Cell. 1984;37(1):273-83.
96. Tanese N, Pugh BF, Tjian R. Coactivators for a proline-rich activator purified from the multisubunit human TFIID complex. Genes Dev. 1991;5(12A):2212-24.
97. Buratowski S, Hahn S, Guarente L, Sharp PA. Five intermediate complexes in transcription initiation by RNA polymerase II. Cell. 1989;56(4):549-61.
98. Imbalzano AN, Zaret KS, Kingston RE. Transcription factor (TF) IIB and TFIIA can independently increase the affinity of the TATA-binding protein for DNA. The Journal of biological chemistry. 1994;269(11):8280-6.
99. Berroteran RW, Ware DE, Hampsey M. The sua8 suppressors of Saccharomyces cerevisiae encode replacements of conserved residues within the largest subunit of RNA polymerase II and affect transcription start site selection similarly to sua7 (TFIIB) mutations. Molecular and cellular biology. 1994;14(1):226-37.
100. Goodrich JA, Tjian R. Transcription factors IIE and IIH and ATP hydrolysis direct promoter clearance by RNA polymerase II. Cell. 1994;77(1):145-56.
101. Perales R, Bentley D. "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. Molecular cell. 2009;36(2):178-91.
102. Pandit S, Wang D, Fu XD. Functional integration of transcriptional and RNA processing machineries. Curr Opin Cell Biol. 2008;20(3):260-5.

103. Malik S, Guermah M, Yuan CX, Wu W, Yamamura S, Roeder RG. Structural and functional organization of TRAP220, the TRAP/mediator subunit that is targeted by nuclear receptors. Molecular and cellular biology. 2004;24(18):8244-54.
104. Taatjes DJ, Tjian R. Structure and function of CRSP/Med2; a promoter-selective transcriptional coactivator complex. Molecular cell. 2004;14(5):675-83.
105. Liu Y, Ranish JA, Aebersold R, Hahn S. Yeast nuclear extract contains two major forms of RNA polymerase II mediator complexes. The Journal of biological chemistry. 2001;276(10):7169-75.
106. Taatjes DJ, Naar AM, Andel F, 3rd, Nogales E, Tjian R. Structure, function, and activator-induced conformations of the CRSP coactivator. Science. 2002;295(5557):1058-62.
107. Daniels DL, Ford, M., Schwinn, M. K., Benink, H., Galbraith, M. D., Amunugama, R., ... & Miki, F. . Mutual exclusivity of MED12/MED12L, MED13/13L, and CDK8/19 paralogs revealed within the CDK-Mediator kinase module. Journal of Proteomics & Bioinformatics. 2013.
108. Muncke N, Jung C, Rudiger H, Ulmer H, Roeth R, Hubert A, et al. Missense mutations and gene interruption in PROSIT240, a novel TRAP240-like gene, in patients with congenital heart defect (transposition of the great arteries). Circulation. 2003;108(23):2843-50.
109. Mukhopadhyay A, Kramer JM, Merkx G, Lugtenberg D, Smeets DF, Oortveld MA, et al. CDK19 is disrupted in a female patient with bilateral congenital retinal folds, microcephaly and mild mental retardation. Human genetics. 2010;128(3):281-91.
110. Baek HJ, Kang YK, Roeder RG. Human Mediator enhances basal transcription by facilitating recruitment of transcription factor IIB during preinitiation complex assembly. The Journal of biological chemistry. 2006;281(22):15172-81.
111. Johnson KM, Carey M. Assembly of a mediator/TFIID/TFIIA complex bypasses the need for an activator. Current biology : CB. 2003;13(9):772-7.
112. Jishage M, Malik S, Wagner U, Uberheide B, Ishihama Y, Hu X, et al. Transcriptional regulation by Pol II(G) involving mediator and competitive interactions of Gdown1 and TFIIF with Pol II. Molecular cell. 2012;45(1):51-63.
113. Esnault C, Ghavi-Helm Y, Brun S, Soutourina J, Van Berkum N, Boschiero C, et al. Mediator-dependent recruitment of TFIIH modules in preinitiation complex. Molecular cell. 2008;31(3):337-46.
114. Guermah M, Tao Y, Roeder RG. Positive and negative TAF(II) functions that suggest a dynamic TFIID structure and elicit synergy with traps in activator-induced transcription. Molecular and cellular biology. 2001;21(20):6882-94.
115. Johnson KM, Wang J, Smallwood A, Arayata C, Carey M. TFIID and human mediator coactivator complexes assemble cooperatively on promoter DNA. Genes & development. 2002;16(14):1852-63.
116. Marr MT, 2nd, Isogai Y, Wright KJ, Tjian R. Coactivator cross-talk specifies transcriptional output. Genes & development. 2006;20(11):1458-69.
117. Takahashi H, Parmely TJ, Sato S, Tomomori-Sato C, Banks CA, Kong SE, et al. Human mediator subunit MED26 functions as a docking site for transcription elongation factors. Cell. 2011;146(1):92-104.
118. Xu M, Sharma P, Pan S, Malik S, Roeder RG, Martinez E. Core promoter-selective function of HMGA1 and Mediator in Initiator-dependent transcription. Genes & development. 2011;25(23):2513-24.
119. Meyer KD, Lin SC, Bernecky C, Gao Y, Taatjes DJ. p53 activates transcription by directing structural shifts in Mediator. Nature structural & molecular biology. 2010;17(6):753-60.
120. Kim YJ, Bjorklund S, Li Y, Sayre MH, Kornberg RD. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. Cell. 1994;77(4):599-608.

121. Boeing S, Rigault C, Heidemann M, Eick D, Meisterernst M. RNA polymerase II C-terminal heptarepeat domain Ser-7 phosphorylation is established in a mediator-dependent fashion. The Journal of biological chemistry. 2010;285(1):188-96.
122. Nair D, Kim Y, Myers LC. Mediator and TFIIH govern carboxyl-terminal domain-dependent transcription in yeast extracts. The Journal of biological chemistry. 2005;280(40):33739-48.
123. Lai F, Gardini A, Zhang A, Shiekhattar R. Integrator mediates the biogenesis of enhancer RNAs. Nature. 2015;525(7569):399-403.
124. Peterlin BM, Price DH. Controlling the elongation phase of transcription with P-TEFb. Molecular cell. 2006;23(3):297-305.
125. Gilmour DS. Promoter proximal pausing on genes in metazoans. Chromosoma. 2009;118(1):1-10.
126. Luo Z, Lin C, Guest E, Garrett AS, Mohaghegh N, Swanson S, et al. The super elongation complex family of RNA polymerase II elongation factors: gene target specificity and transcriptional output. Molecular and cellular biology. 2012;32(13):2608-17.
127. Ruskin B, Zamore PD, Green MR. A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. Cell. 1988;52(2):207-19.
128. Berglund JA, Abovich N, Rosbash M. A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. Genes Dev. 1998;12(6):858-67.
129. Zhuang Y, Weiner AM. A compensatory base change in human U2 snRNA can suppress a branch site mutation. Genes Dev. 1989;3(10):1545-52.
130. Wu JA, Manley JL. Base pairing between U2 and U6 snRNAs is necessary for splicing of a mammalian pre-mRNA. Nature. 1991;352(6338):818-21.
131. Fica SM, Tuttle N, Novak T, Li NS, Lu J, Koodathingal P, et al. RNA catalyses nuclear pre-mRNA splicing. Nature. 2013;503(7475):229-34.
132. Bentley DL. Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. Curr Opin Cell Biol. 2005;17(3):251-6.
133. Morris DP, Greenleaf AL. The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. The Journal of biological chemistry. 2000;275(51):39935-43.
134. Kornblihtt AR, de la Mata M, Fededa JP, Munoz MJ, Nogues G. Multiple links between transcription and splicing. RNA. 2004;10(10):1489-98.
135. Proudfoot NJ. Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. Science. 2016;352(6291):aad9926.
136. Clerici M, Faini M, Aebersold R, Jinek M. Structural insights into the assembly and polyA signal recognition mechanism of the human CPSF complex. eLife. 2017;6.
137. Sun Y, Zhang Y, Hamilton K, Manley JL, Shi Y, Walz T, et al. Molecular basis for the recognition of the human AAUAAA polyadenylation signal. Proceedings of the National Academy of Sciences of the United States of America. 2018;115(7):E1419-E28.
138. Dantonel JC, Murthy KG, Manley JL, Tora L. Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA. Nature. 1997;389(6649):399-402.
139. Mandel CR, Kaneko S, Zhang H, Gebauer D, Vethantham V, Manley JL, et al. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. Nature. 2006;444(7121):953-6.
140. Dominski Z, Yang XC, 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. Molecular and cellular biology. 2005;25(4):1489-500.
141. Takagaki Y, Manley JL. RNA recognition by the human polyadenylation factor CstF. Molecular and cellular biology. 1997;17(7):3907-14.
142. Rosonina E, Kaneko S, Manley JL. Terminating the transcript: breaking up is hard to do. Genes & development. 2006;20(9):1050-6.

143. Kim M, Krogan NJ, Vasiljeva L, Rando OJ, Nedea E, Greenblatt JF, et al. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature. 2004;432(7016):517-22.
144. Zaborowska J, Egloff S, Murphy S. The pol II CTD: new twists in the tail. Nature structural & molecular biology. 2016;23(9):771-7.
145. Albrecht TR, Wagner EJ. snRNA 3' end formation requires heterodimeric association of integrator subunits. Molecular and cellular biology. 2012;32(6):1112-23.
146. Bartkowiak B, Liu P, Phatnani HP, Fuda NJ, Cooper JJ, Price DH, et al. CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. Genes & development. 2010;24(20):2303-16.
147. Blazek D, Kohoutek J, Bartholomeeusen K, Johansen E, Hulinkova P, Luo Z, et al. The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. Genes & development. 2011;25(20):2158-72.
148. Liang K, Gao X, Gilmore JM, Florens L, Washburn MP, Smith E, et al. Characterization of human cyclin-dependent kinase 12 (CDK12) and CDK13 complexes in C-terminal domain phosphorylation, gene transcription, and RNA processing. Molecular and cellular biology. 2015;35(6):928-38.
149. Mattaj IW. Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. Cell. 1986;46(6):905-11.
150. Nojima T, Gomes T, Grosso ARF, Kimura H, Dye MJ, Dhir S, et al. Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. Cell. 2015;161(3):526-40.
151. Nojima T, Rebelo K, Gomes T, Grosso AR, Proudfoot NJ, Carmo-Fonseca M. RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-transcriptional Splicing. Molecular cell. 2018;72(2):369-79 e4.
152. Mayer A, Lidschreiber M, Siebert M, Leike K, Soding J, Cramer P. Uniform transitions of the general RNA polymerase II transcription complex. Nature structural & molecular biology. 2010;17(10):1272-8.
153. Chapman RD, Heidemann M, Albert TK, Mailhammer R, Flatley A, Meisterernst M, et al. Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. Science. 2007;318(5857):1780-2.
154. Descostes N, Heidemann M, Spinelli L, Schuller R, Maqbool MA, Fenouil R, 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.
155. Harlen KM, Trotta KL, Smith EE, Mosaheb MM, Fuchs SM, Churchman LS. Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. Cell reports. 2016;15(10):2147-58.
156. Hintermair C, Heidemann M, Koch F, Descostes N, Gut M, Gut I, et al. Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. The EMBO journal. 2012;31(12):2784-97.
157. Hsin JP, Sheth A, Manley JL. RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing. Science. 2011;334(6056):683-6.
158. Bartkowiak B, Greenleaf AL. Phosphorylation of RNAPII: To P-TEFb or not to P-TEFb? Transcription. 2011;2(3):115-9.
159. Yoh SM, Cho H, Pickle L, Evans RM, Jones KA. The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. Genes & development. 2007;21(2):160-74.
160. Bono F, Gehring NH. Assembly, disassembly and recycling: the dynamics of exon junction complexes. RNA biology. 2011;8(1):24-9.
161. Licatalosi DD, Geiger G, Minet M, Schroeder S, Cilli K, McNeil JB, et al. Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. Molecular cell. 2002;9(5):1101-11.

162. Marshall NF, Price DH. Purification of P-TEFb, a transcription factor required for the transition into productive elongation. The Journal of biological chemistry. 1995;270(21):12335-8.
163. Bosken CA, Farnung L, Hintermair C, Merzel Schachter M, Vogel-Bachmayr K, Blazek D, et al. The structure and substrate specificity of human Cdk12/Cyclin K. Nature communications. 2014;5:3505.
164. Zhang T, Kwiatkowski N, Olson CM, Dixon-Clarke SE, Abraham BJ, Greifenberg AK, et al. Covalent targeting of remote cysteine residues to develop CDK12 and CDK13 inhibitors. Nature chemical biology. 2016;12(10):876-84.
165. Poss ZC, Ebmeier CC, Odell AT, Tangpeerachaikul A, Lee T, Pelish HE, et al. Identification of Mediator Kinase Substrates in Human Cells using Cortistatin A and Quantitative Phosphoproteomics. Cell reports. 2016;15(2):436-50.
166. Bishop AC, Ubersax JA, Petsch DT, Matheos DP, Gray NS, Blethrow J, et al. A chemical switch for inhibitor-sensitive alleles of any protein kinase. Nature. 2000;407(6802):395-401.
167. Bartkowiak B, Greenleaf AL. Expression, purification, and identification of associated proteins of the full-length hCDK12/CyclinK complex. The Journal of biological chemistry. 2015;290(3):1786-95.
168. Bartkowiak B, Yan C, Greenleaf AL. Engineering an analog-sensitive CDK12 cell line using CRISPR/Cas. Biochimica et biophysica acta. 2015;1849(9):1179-87.
169. Chirackal Manavalan AP, Pilarova K, Kluge M, Bartholomeeusen K, Rajecky M, Oppelt J, et al. CDK12 controls G1/S progression by regulating RNAPII processivity at core DNA replication genes. EMBO reports. 2019:e47592.
170. Dejardin J, Kingston RE. Purification of proteins associated with specific genomic Loci. Cell. 2009;136(1):175-86.
171. Saksouk N, Barth TK, Ziegler-Birling C, Olova N, Nowak A, Rey E, et al. Redundant mechanisms to form silent chromatin at pericentromeric regions rely on BEND3 and DNA methylation. Molecular cell. 2014;56(4):580-94.
172. Fujita T, Fujii H. Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. Biochemical and biophysical research communications. 2013;439(1):132-6.
173. Tsui C, Inouye C, Levy M, Lu A, Florens L, Washburn MP, et al. dCas9-targeted locus-specific protein isolation method identifies histone gene regulators. Proceedings of the National Academy of Sciences of the United States of America. 2018;115(12):E2734-E41.
174. Liu X, Zhang Y, Chen Y, Li M, Zhou F, Li K, et al. In Situ Capture of Chromatin Interactions by Biotinylated dCas9. Cell. 2017;170(5):1028-43 e19.
175. Kustatscher G, Wills KL, Furlan C, Rappsilber J. Chromatin enrichment for proteomics. Nature protocols. 2014;9(9):2090-9.
176. Myers SA, Wright J, Peckner R, Kalish BT, Zhang F, Carr SA. Discovery of proteins associated with a predefined genomic locus via dCas9-APEX-mediated proximity labeling. Nature methods. 2018;15(6):437-9.
177. Su JM, Lai XM, Lan KH, Li CP, Chao Y, Yen SH, et al. X protein of hepatitis B virus functions as a transcriptional corepressor on the human telomerase promoter. Hepatology. 2007;46(2):402-13.
178. Xu M, Katzenellenbogen RA, Grandori C, Galloway DA. An unbiased in vivo screen reveals multiple transcription factors that control HPV E6-regulated hTERT in keratinocytes. Virology. 2013;446(1-2):17-24.
179. Glasspool RM, Burns S, Hoare SF, Svensson C, Keith WN. The hTERT and hTERC telomerase gene promoters are activated by the second exon of the adenoviral protein, E1A, identifying the transcriptional corepressor CtBP as a potential repressor of both genes. Neoplasia. 2005;7(6):614-22.

180. Medlin JE, Uguen P, Taylor A, Bentley DL, Murphy S. The C-terminal domain of pol II and a DRB-sensitive kinase are required for 3' processing of U2 snRNA. The EMBO journal. 2003;22(4):925-34.
181. Yoon JB, Roeder RG. Cloning of two proximal sequence element-binding transcription factor subunits (gamma and delta) that are required for transcription of small nuclear RNA genes by RNA polymerases II and III and interact with the TATA-binding protein. Molecular and cellular biology. 1996;16(1):1-9.
182. Nojima T, Gomes T, Carmo-Fonseca M, Proudfoot NJ. Mammalian NET-seq analysis defines nascent RNA profiles and associated RNA processing genome-wide. Nature protocols. 2016;11(3):413-28.
183. van Steensel B, de Lange T. Control of telomere length by the human telomeric protein TRF1. Nature. 1997;385(6618):740-3.
184. Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, et al. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell. 2005;123(4):581-92.
185. Nojima T, Tellier M, Foxwell J, Ribeiro de Almeida C, Tan-Wong SM, Dhir S, et al. Deregulated Expression of Mammalian lncRNA through Loss of SPT6 Induces R-Loop Formation, Replication Stress, and Cellular Senescence. Molecular cell. 2018;72(6):970-84 e7.
186. Yaniv M. Chromatin remodeling: from transcription to cancer. Cancer Genet. 2014;207(9):352-7.
187. Reines D, Chamberlin MJ, Kane CM. Transcription elongation factor SII (TFIIS) enables RNA polymerase II to elongate through a block to transcription in a human gene in vitro. The Journal of biological chemistry. 1989;264(18):10799-809.
188. SivaRaman L, Reines D, Kane CM. Purified elongation factor SII is sufficient to promote read-through by purified RNA polymerase II at specific termination sites in the human histone H3.3 gene. The Journal of biological chemistry. 1990;265(24):14554-60.
189. Izban MG, Luse DS. Transcription on nucleosomal templates by RNA polymerase II in vitro: inhibition of elongation with enhancement of sequence-specific pausing. Genes & development. 1991;5(4):683-96.
190. Di Giammartino DC, Manley JL. New links between mRNA polyadenylation and diverse nuclear pathways. Molecules and cells. 2014;37(9):644-9.
191. Clerici M, Faini M, Muckenfuss LM, Aebersold R, Jinek M. Structural basis of AAUAAA polyadenylation signal recognition by the human CPSF complex. Nature structural & molecular biology. 2018;25(2):135-8.
192. King HA, Trotter KW, Archer TK. Chromatin remodeling during glucocorticoid receptor regulated transactivation. Biochimica et biophysica acta. 2012;1819(7):716-26.
193. Ford E, Thanos D. The transcriptional code of human IFN-beta gene expression. Biochimica et biophysica acta. 2010;1799(3-4):328-36.
194. Masliah-Planchon J, Bieche I, Guinebretiere JM, Bourdeaut F, Delattre O. SWI/SNF chromatin remodeling and human malignancies. Annu Rev Pathol. 2015;10:145-71.
195. Fong N, Brannan K, Erickson B, Kim H, Cortazar MA, Sheridan RM, et al. Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. Molecular cell. 2015;60(2):256-67.
196. Krajewska M, Dries R, Grassetti AV, Dust S, Gao Y, Huang H, et al. CDK12 loss in cancer cells affects DNA damage response genes through premature cleavage and polyadenylation. Nature communications. 2019;10(1):1757.
197. Guiro J, Murphy S. Regulation of expression of human RNA polymerase II-transcribed snRNA genes. Open Biol. 2017;7(6).