CONTROL OF EXPRESSION OF HUMAN snRNA GENES

Justyna Zaborowska

Lincoln College

University of Oxford

A thesis submitted for the degree of Doctor of Philosophy at the University of Oxford, Michaelmas Term 2013

Abstract

In humans, protein-coding genes and most small nuclear (sn)RNA genes are transcribed by RNA polymerase II (pol II).The carboxy-terminal domain (CTD) of the largest subunit of pol II possesses multiple heptapetide repeats of the consensus Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7. Phosphorylation of Ser2, Ser5 and Ser7 mediates the recruitment of transcription and RNA processing factors during the transcription cycle. There are notable differences between snRNA genes and protein-coding genes in terms of mechanisms controlling their expression. Pol II does not appear to make the transition to long-range productive elongation during transcription of snRNA genes, as happens during transcription of protein-coding genes. In addition, recognition of the snRNA gene-type specific 3' box RNA processing element requires initiation from an snRNA gene promoter. These characteristics may, at least in part, be driven by factors recruited to the promoter. Initiation of transcription of most human genes transcribed by pol II requires the formation of a preinitiation complex (PIC) comprising TFIIA, B, D, E, F and H and pol II. The general transcription factor, TFIID is composed of the TATA-binding protein and up to 13 TBP-associated factors (TAFs). Differences in the complement of TAFs might result in differential recruitment of elongation and RNA processing factors. It has already been shown that the promoters of some protein-coding genes do not recruit all the TAFs found in TFIID. Although TAF5, has been shown to be associated with pol II-transcribed snRNA genes, the full complement of TAFs associated with these genes remained unclear. Here I show, using a ChIP and siRNA-mediated knockdown approach, that the TBP/TAF complex on snRNA genes differs from that on protein-coding genes. Interestingly, the largest TAF, TAF1 and the core TAFs, TAF10 and TAF4 are not detected on snRNA genes. I propose that this snRNA gene-specific TAF subset plays a key role in gene-type-specific control of expression. In addition, in order to further understand the molecular mechanism underlying the differences between expression of protein-coding genes and snRNA genes, I have investigated the role of RNA pol II-associated protein 2 (RPAP2) in transcription of snRNA genes. Here I show that RPAP2 recognizes the phospho-Ser7 mark on the pol II CTD, siRNA mediated knockdown of RPAP2 causes defects in snRNA gene expression and that RPAP2 is a CTD Ser5 phosphatase. I also present my studies of the mechanism of inhibition of phospho-Ser2 by herpes simplex virus-1 (HSV-1) protein ICP22. Phosphorylation of Ser2 by the positive transcription elongation factor (P-TEFb) is associated with productive transcriptional elongation. However, P-TEFb is not required for elongation of transcription of snRNA genes, but functions only to activate 3' box-directed RNA processing. In addition, there are conflicting data as to whether Cdk9 is acting as a Ser2 kinase during transcription of pol II-transcribed snRNA genes. As ICP22 is thought to inhibit P-TEFb, this protein could provide an alternative means to study P-TEFb function in expression of snRNA genes.

I dedicate this thesis to my daughter, Kalina.

Acknowledgments

My special thanks to my supervisor Prof Shona Murphy for giving me the opportunity to work in her laboratory. She was always there for me and served with the technical guidance and professional advice. I could not have imagined having a better advisor and mentor for my DPhil study. I also want to thank all the past and present members of the Murphy group for providing a very nice working atmosphere. I am very grateful to Mrs Alice Taylor for all the shared early mornings in the lab. A very special thanks to Dr Dawn O'Reilly, Dr Clelia Laitem, Dr Martin Dienstbier, Dr Hadeel Al-Rawaf, Dr Pilar Vazquez-Arango for all helpful discussions and suggestions and for the sleepless nights we were working together before deadlines. I believe our friendship will flourish with the years to come.

I would like to thank Patrycjusz Grobelny and uncle Robert Mosbauer, who at the very beginning of my scientific career were correcting my endless scientific abstracts. Further, I would like to express my deepest gratitude to Gabriela Ersilova for providing me with wine and cakes to keep me going. Many thanks to all my friend that I have met here in Oxford, especially Dr Luc Henry, Dr Alain Ausoni, Dr Kieren Finan, Dr Sarah Brinckman, Dr Olga Liska. They were lovely house mates, companions and compassionate advisors.

I would also like to thank my family for the support they provided me through my entire life and in particular, I must acknowledge my parents Elzbieta and Lech Zaborowscy, brother Rafal Zaborowski and husband Gabriel Klobusovsky. Without their love and encouragement I would not have finished this thesis.

Table of contents

List of figures

List of Tables

Abbreviations

5'ss	5' splice site
3'ss	3' splice site
bp	base pair
BP	branch point
Brd4	Bromodomain-containing protein 4
BSA	Bovine Serum Albumin
Cdk1	cyclin-dependent kinase 1
Cdk7	cyclin-dependent kinase 7
Cdk8	cyclin-dependent kinase 8
Cdk9	cyclin-dependent kinase 9
Cdk12	cyclin-dependent kinase 12
CF	Cleavage factor
ChIP	Chromatin Immunoprecipitation
CMV	cytomegalovirus
CPF	Cleavage and Polyadenylation Factor
CPSF	Cleavage and Polyadenylation Specificity Factor
CstF	Cleavage stimulation Factor
CTD	C-terminal domain
DMSO	dimethyl sulphoxide
DNA	Deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate

DCE	Downstream Core Element
DPE	Downstream Promoter Element
DRB	5,6-dicloro-1-β-D-ribofuranosylbenzimidazole
DSE	Distal Sequence Element
DSIF	DRB-sensitivity inducing factor
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EKFL	erythroid Kruppel-like factor
ENCODE	Encylopedia of DNA elements
EtOH	ethanol
Fcp1	TFIIF-associated CTD phosphatase 1
GST	gluthatione-S-transferase
GTFs	General Transcription Factors
h	hour
ICP22	immediate early protein 22
Inr	initiatior
HATs	Histone Acetyltransferases
HFD	histone fold domain
HIV	human immunodeficiency virus
HSV-1	herpes simplex virus-1
IP	Immunoprecipitation
kDa	kilo Dalton

mRNA	messenger RNA
NaCl	Sodium Chloride
ncRNA	non-coding RNA
NELF	negative elongation factor
NF-1	nuclear factor-Y
NP-40	Nonidet P40
NRO	Nuclear Run-On
N-TEFs	negative elongation factor
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PMSF	Phenylmethylsulfonlyfluoride
PIC	preinitiation complex
Pol II	RNA Polymerase II
PPT	polypyrymidine tract
PSE	Proximal Sequence Element
PTF	PSE-binding transcription factor
P-TEFb	Positive Transcription Elongation factor b
qPCR	quantitative PCR
qRT-PCR	quantitative reverse transcription PCR
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
RPAP2	RNA pol II associated protein

RT	Room Temperature
RT-PCR	Reverse transcription PCR
Rtr1	regulator of transcription 1
SCP1	small CTD phosphatase 1
SDS	Sodium Dodecyl Sulphate
SF1	splicing factor 1
sec	seconds
snRNA	small nuclear RNA
snRNPs	small nuclear RNPs
SYBR	SybrGreen
TAFs	TBP-associated factors
TBP	TATA-binding protein
TBS	Tris Buffered Saline
TFTC	TAF-containing complex
TSS	transcription start site

Chapter 1| Introduction

Transcription is the first step of gene expression and is strictly regulated to ensure appropriate levels of the gene product. Understanding which factors work together to achieve correct regulation of transcription is critical for a complete understanding of gene regulation. This thesis describes the results of my study of the control of transcription of RNA polymerase II (pol II)-dependent genes. The broad scientific context is the molecular mechanisms underlying transcription of human pol II-dependent genes in cell lines, with particular emphasis on the role of modifications of the carboxyl-terminal domain (CTD) of pol II. In addition, the effect of the herpes simplex virus type 1 (HSV-1) immediate early protein 22 (ICP22) on transcription by pol II was investigated, as ICP22 is thought to interact with a key CTD modification enzyme. Accordingly, in the Introduction I give an overview of transcription of protein-coding and small nuclear (sn)RNA genes, as these are the two major groups of independent genes transcribed by pol II in human cells. I also give a review of what is known about the effect of the HSV-1 ICP22 protein on host cell gene expression. In the remaining part of this Introduction, I outline the aims of this thesis.

1.1 Transcription of human genes

Eukaryotic transcription is a stringently-controlled process where DNA acts as a template for synthesis of an RNA copy (Shandilya and Roberts, 2012). In eukaryotes, there are three distinct polymerases that transcribe specific gene types. RNA polymerase I (pol I) transcribes genes encoding 28S, 18S and 5.8S ribosomal (r)RNAs. RNA polymerase II (pol II) transcribes the protein-coding genes and genes encoding small non-coding RNAs, including the independent genes for the snRNAs. RNA polymerase III (pol III) transcribes a range of short genes for non-coding RNAs, including those encoding transfer (t)RNA and the 5S rRNA (Egloff and Murphy, 2008a, Allison et al., 1985). The transcription cycle has been well studied for protein-coding genes and can be divided in three phases: initiation, elongation and termination. During initiation, a transcription-competent RNA polymerase complex is formed at the promoter and synthesizes the first phosphodiester bond of the RNA. Transcription elongation is divided into three stages: promoter escape, promoter-proximal pausing and productive elongation (Saunders et al., 2006). During promoter

escape, pol II breaks its contacts with promoter-sequence elements/promoter-bound factors. During promoter-proximal pausing, pol II pauses in the 5' region of the transcription unit. This stage functions as a checkpoint before proceeding to productive elongation. During productive elongation, polymerase synthesizes the complete RNA transcript. During termination, the RNA transcript is released and the transcription complex dissociates from the DNA template (Saunders et al., 2006).

1.1.1 RNA polymerase II structure and CTD modifications

Pol II is made up of 12 DNA-directed RNA polymerase II subunits named Rpb1-Rpb12 (Cramer, 2002). The carboxyl-terminal domain (CTD) of the largest subunit of pol II (Rpb1) comprises tandemly-repeated heptapeptides with the consensus sequence Tyr-Ser-Pro-Thr-Ser-Pro-Ser ($Y_1S_2P_3T_4S_5P_6S_7$) (Corden, 1990, McCracken et al., 1997, Allison et al., 1985). The number of repeats and the deviation from the consensus sequence varies between organisms. There are 26 repeats in yeast, 45 in Drosophila and 52 in mammals (Corden, 1990). The divergent heptapeptides are found largely in the C-terminal part with the Ser7 being the least conserved residue (Hsin et al., 2011). The minimal functional unit of a yeast pol II CTD is a di-heptapeptide (Stiller and Cook, 2004). The CTD plays a crucial role in transcription and co-transcriptional RNA processing. It serves as a scaffold for the interaction with nuclear factors (Bentley, 2005, Egloff et al., 2007, Jacobs et al., 2004, Phatnani and Greenleaf, 2006). Reversible modification of the residues within a heptapeptide of the CTD generates a CTD code which is essential for the recruitment of transcription and processing factors at different steps of the transcription cycle (Buratowski, 2003). Thus, modification of the CTD is crucial for coordination of transcription and RNA processing events (Saunders et al., 2006, Egloff and Murphy, 2008a). The CTD can potentially be modified by phosphorylation (on tyrosine, threonine and serine), glycosylation (on tyrosine, threonine and serine) and isomerization (proline) (Table 1). The enzymes that modify the CTD include kinases, phosphatases, glycosyltransferases, deglycosylases and peptidyl-propyl *cis/trans* isomerases (Egloff and Murphy, 2008a). As a result of modification, the hypophosphorylated pol II (pol IIa) that initiates transcription is transformed into a hyperphosphorylated pol II (pol IIo) that engages in productive elongation (Payne et al., 1989). Phosphorylation of the CTD is the

best-characterized modification. There are five potential phosphorylation sites in a heptapeptide repeat ($Y_1S_2T_4S_5S_7$) (Zhang and Corden, 1991, Baskaran et al., 1993) (Table 1).

The Ser5 residue of the CTD is phosphorylated by cyclin-dependent kinase 7 (Cdk7; Kin28 in yeast), a component of the general transcription factor TFIIH, and by the cyclin-dependent kinase 8 (Cdk8; Srb10 in yeast), a subunit of the Mediator complex (Komarnitsky et al., 2000, Riedl and Egly, 2000). Phosphorylation on Ser5 (Ser5P) is highest near the promoter and helps recruit enzymes to cap the 5' end of the newly transcribed transcript and the Set1 histone methyltransferase, which is responsible for trimethylation of lysine residue on histone H3 (H3K4) at the 5' end of transcribed genes (Schroeder et al., 2000, Ng et al., 2003, Komarnitsky et al., 2000). In addition, the Ser5P mark is required for recruitment of the Ser5 phosphatase, the regulator of transcription 1 (Rtr1) (Mosley et al., 2009).

Ser2 is phosphorylated by the catalytic subunit (Cdk9) of the positive transcription elongation factor (P-TEFb) complex (Bur1 in yeast) and/or cyclin-dependent kinase 12 (Cdk1; Ctk1 in yeast) (Peterlin and Price, 2006, Bartkowiak et al., 2010, Keogh et al., 2003). In addition, a recent study described the bromodomain-containing protein 4 (Brd4) as a kinase for Ser2 where P-TEFb is not recruited to promoters and in stem cell lines deficient in P-TEFb (Devaiah et al., 2012). Ser2 phosphorylation (Ser2P) occurs later in the transcription cycle and is generally highest towards the 3' end of the genes. In higher eukaryotes, P-TEFb inhibitors affect elongation of transcription of protein-coding genes, splicing and polyadenylation (Bird et al., 2004, Ni et al., 2004). However, Ser2P does not affect elongation of transcription of mammalian snRNA genes, but plays a major role in RNA 3' processing (Medlin et al., 2005). The Ser2P mark also facilitates association of the Set2 histone methyltransferase, which catalyzes methylation of histone H3K36 (Li et al., 2005). In yeast, methylation of histone H3K36 leads to activation of a histone deacetylase complex, which prevents cryptic transcription within coding regions (Buratowski, 2009, Govind et al., 2010).

Modification	Enzyme		References
Tyr1 phosphorylation	c-Abl		(Baskaran et al., 1993, Mayer et al., 2012)
Ser2 phosphorylation	Kinase	Cdk9, Cdk12, Brd4	(Washington et al., 2002, Clemente-Blanco et al., 2011, Lin et al., 2002, Bartkowiak et al., 2010, Blazek et al., 2011, Devaiah et al., 2012, Guillamot et al., 2011)
	Phosphatase	Fcp1, Cdc14b	
Pro3 isomerization	Pin1		(Zhang et al., 2012b)
Thr4 phosphorylation	Kinase	Cdk9 Plk3	(Hsin et al., 2011, Hintermair et al., 2012)
Ser5 phosphorylation	Kinase	Cdk7, Cdk8	(Ramanathan et al., 2001, Washington et al., 2002, Yeo et al., 2003, Krishnamurthy et al., 2004, Glover-Cutter et al., 2009, Egloff et al., 2012b)
	Phosphatase	RPAP2, Scp1, Ssu72, PP1	
Pro6 isomerization	Pin1		(Zhang et al., 2012b, Bataille et al., 2012)
Ser7 phosphorylation	Kinase	Cdk7, Cdk9	(Egloff et al., 2007, Bataille et al., 2012)
	Phosphatase	Ssu72	
Arg7 methylation	CARM1		(Sims et al., 2011)
Ser and Thr glycosylation	OGT OGA		(Kelly et al., 1993, Ranuncolo et al., 2012)

Table 1. Summary of known CTD modifications in mammals.

In addition to Ser2 and Ser5, Ser7 of the CTD is phosphorylated during transcription (Chapman et al., 2007). Ser7 phosphorylation (Ser7P) has been detected on protein-coding genes and snRNA genes (Chapman et al., 2007, Egloff et al., 2007). The role of Ser7P in expression of protein coding genes is not clear. This mark is important for transcription of snRNA genes and for formation of the 3' end of transcripts (Egloff et al., 2012a). The Cdk7 kinase (Kin28 in yeast) was demonstrated to phosphorylate not only Ser5, but also Ser7 (Glover-Cutter et al., 2009). Recently, Tietjen and colleagues, revealed a novel role for yeast Bur1 kinase as a Ser7 kinase that acts on elongating pol II (Tietjen et al., 2010). However, Bataille and colleagues, showed that Bur1 phosphorylates Ser7 only when Kin28 is not functional (Bataille et al., 2012). Despite its role in 3'-processing of snRNA gene transcripts, the Ser7P mark is observed at promoters of all pol II transcribed genes (Glover-Cutter et al., 2009, Kim et al., 2009b, Egloff et al., 2007). Tietjen and colleagues have suggested that Ser7P could facilitate elongation and suppress cryptic transcription (Tietjen et al., 2010).

The phosphorylation of Thr4 and Tyr1 of the pol II CTD was also reported (Hsin et al., 2011, Baskaran et al., 1993). Knockdown experiments confirmed that Polo-like kinase 3 (PLK3) is a Thr4 specific kinase (Hintermair et al., 2012). This modification of the pol II CTD has been linked to 3'end processing of histone transcripts (Hsin et al., 2011). Tyr1 is phosphorylated by the kinase c-Abl in human cells (Baskaran et al., 1993) and this mark inhibits the binding of termination factors to the transcribing pol II within the body of gene in yeast (Mayer et al., 2012).

Dephosphorylation of the pol II CTD residues is essential for regulation of CTD phosphorylation (Hsin and Manley, 2012). The evolutionary-conserved TFIIF-associated CTD phosphatase 1 (Fcp1) dephosphorylates Ser2P preferentially, whereas RNA polymerase II subunit A C-terminal domain phosphatase (Ssu72), the small CTD phosphatase 1 (SCP1) in mammals and the regulator of transcription 1 (Rtr1) in yeast dephosphorylate Ser5P (Krishnamurthy et al., 2004, Reyes-Reyes and Hampsey, 2007, Yeo et al., 2003, Mosley et al., 2009). Fcp1 binds pol II and its activity is stimulated by the general transcription factor TFIIF (Kobor et al., 1999, Kobor et al., 2000). Yeast Fcp1 is localized to the promoter and coding regions of active genes (Cho et al., 2001). Human Fcp1 stimulates elongation and facilitates recycling of pol II (Cho et al., 1999, Mandal et

al., 2002). A recent study in Drosophila demonstrated that Fcp1 is required for efficient transcription of heat shock genes. The decrease of pol II levels in the coding region of the heat shock-induced Hsp70 gene after Fcp1 depletion, suggests that Fcp1 is required to maintain the pool of initiation-competent unphosphorylated pol II (Fuda et al., 2012). The Ssu72 phosphatase was identified as a component of the yeast Cleavage and Polyadenylation Factor (CPF) and found to be required for pre-mRNA and small nucleolar (sno)RNA 3' end formation (Gavin et al., 2002, He et al., 2003). In addition to its phosphatase activity towards Ser5P, Ssu72 was described as a tyrosine phosphatase and was also recently identified as a Ser7P phosphatase (Meinhart et al., 2003, Zhang et al., 2012a). This would suggest that the Ssu72 active site is able to support a variety of CTD modifications. Ssu72 localizes predominantly to the 3' end of genes but can also be found at promoter regions (Ansari and Hampsey, 2005, Nedea et al., 2003). The SCP1 phosphatase is found only in human cells, is regulated by TFIIF and contains a phosphatase domain similar to that of Fcp1 (Yeo et al., 2003). Finally, the Rtr1 phosphatase in yeast was also shown to be a CTD phosphatase responsible for the dephosphorylation of Ser5P in the 5' region of genes (Mosley et al., 2009). Rtr1 deletion causes increased levels of Ser5P in whole-cell extracts and throughout coding regions. In addition, pol II occupancy on genes is reduced and termination defects are observed (Mosley et al., 2009). The human homolog of Rtr1 is RNA Polymerase II Associated Protein 2 (RPAP2). The first insight into RPAP2 function came from a mass spectrometric analysis of the human transcription machinery interactome, which uncovered four proteins that form strong interactions with pol II, including RPAP2 (Jeronimo et al., 2007). At the time of starting work for this thesis, it was not known if RPAP2 acts as a CTD phosphatase as does its homolog in yeast. Work from our lab has demonstrated that RPAP2 knockdown results in an increase Ser5P levels on pol II transcribed genes and RPAP2 possesses Ser5P phosphatase activity (Egloff et al., 2012b).

Taken together, these findings indicate that dynamic phosphorylation and dephosphorylation of the CTD plays an essential role during transcription.

1.1.2 Transcription initiation factor requirement

Synthesis of pre-messenger RNA by pol II involves assembly of a transcription pre-initiation complex (PIC) at the promoter (Hampsey, 1998). The PIC comprises pol II and a set of general transcription factors (GTFs) that include TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH (Table 2). TFIID is a central player in the assembly of the PIC. TFIID is a multiprotein complex comprising the TATA-binding protein (TBP) and an associated set of 13 evolutionarily-conserved TBP-associated factors (TAFs), and plays a critical role in transcription initiation.

Factor	Function	References
TFIIA	Enhances PIC assembly Stabilizes TBP-TATA box interactions Coactivator	(Kobayashi et al., 1995, Clemens et al., 1996, Kang et al., 1995, Thomas and Chiang, 2006)
TFIIB	Stabilizes TFIID promoter binding Role in recruitment of pol II/TFIIF Role in start site selection	(Orphanides et al., 1996, Hahn, 2004, Thomas and Chiang, 2006, Pardee et al., 1998, Pinto et al., 1994)
TFIID	Role in recognition of core promoter elements Core promoter binding factor Coactivator Posses enzymatic activities (protein kinase, histone acetyltransferase, ubiquitin-activating/conjugating enzyme)	(Dunphy et al., 2000, Sengupta et al., 2009, Maile et al., 2004, Shao et al., 2005, Chiang and Roeder, 1995, Thomas and Chiang, 2006)
TFIIE	Role in recruitment of TFIIH Essential in transition from initiation to elongation	(Flores et al., 1989, Holstege et al., 1996)
TFIIF	Role in recruitment of pol II, TFIIE, TFIIH Role in start site selection Increases the specificity and efficiency of pol II transcription	(Flores et al., 1991, Orphanides et al., 1996, Hampsey, 1998, Thomas and Chiang, 2006)
TFIIH	Posses enzymatic activities (ATPase activity, helicase activity, kinase activity, E3 ubiquitin ligase activity) Required for transcription initiation and promoter clearance Responsible for phosphorylating Ser5 of pol II CTD	(Goodrich and Tjian, 1994, Thomas and Chiang, 2006, Zurita and Merino, 2003, Glover-Cutter et al., 2009)

Table 2. Function of the human general transcription factors.

TFIID: A TBP-TAF complex

The multi-subunit general transcription factor TFIID was initially identified as a chromatographic fraction that is required to reconstitute a basal pol II transcription reaction *in vitro* (Matsui et al., 1980). Interestingly, it was shown that TFIID binds to the TATA-box and a 38 kDa subunit possessing TATA box-binding activity termed TATA-binding protein (TBP) was characterized (Buratowski et al., 1988). TBP can replace TFIID in basal transcription, but cannot support activated transcription, indicating that the TFIID complex contains additional activities required for regulated transcription. A set of tightly-bound polypeptides associated with TBP were identified during immunopurification of TFIID using antibodies against epitope-tagged TBP. It was then determined that TFIID is a complex containing TBP and TBP-associated factors for pol II (TAFs). Significantly, TAFs were shown to play a critical role in activator-responsive transcription (Cavallini et al., 1989, Dynlacht et al., 1991, Matsui et al., 1980, Parker and Topol, 1984, Tanese et al., 1991). Consequently, TAFs have been subjected to intensive study over the last decade. TAFs from *S.cerevisiae*, *S.pombe*, *C.elegans*, *D.melanogaster* and *H. Sapiens* have been identified (Tora, 2002). Some TAFs are also present in distinct human transcriptional regulatory multiprotein complexes such as the SPT3-TAF9-GCN5 containing complex (STAGA; human SAGA), the TBP-free TAF-containing complex (TFTC) and the p300/CBP associated factor complex (PCAF) (see Table 3) (Grant et al., 1998, Ogryzko et al., 1998, Brand et al., 1999, Martinez et al., 2001, Demeny et al., 2007).

COMPLEXES		
hSAGA	**hTFTC**	**hPCAF**
	TAF2	
	TAF4	
	TAF5	
TAF5L	TAF5L	TAF5L
	TAF6	
TAF6L	TAF6L	TAF6L
TAF9	TAF9	TAF9
	TAF9b	
TAF10	TAF10	TAF10
TAF12	TAF12	TAF12

Table 3. Comparison between the complement of TAFs in TAF-containing complexes.

X-ray crystallography and biochemical experiments have shown that nine TAFs contain a sequence motif homologous to histones, the histone fold domain (HFD) (Birck et al., 1998, Gangloff et al., 2001a). The HFD contains a long α-helix flanked on each side by a coil loop and a short α-helix. This domain mediates many of the subunit interactions within the TFIID complex (Birck et al., 1998, Gangloff et al., 2001a). The histone fold-containing TAFs form the heterodimers, TAF6/9, TAF4/12, TAF 8/10, TAF3/10 and TAF11/13. The histone-like heterodimers associate with TAF5 to form the core subcomplex, which interacts with a subcomplex composed of TAF1, TAF2, TAF7 and TBP to form TFIID (Cler et al., 2009, Wright et al., 2006, Gangloff et al., 2001b). The structure of TFIID appears to be dynamic, and it has been shown that differences in subunit composition affect TFIID conformation. For example, the incorporation of TAF4b, a TAF4 paralogue, into TFIID, induces an open conformation at the lobe involved in interaction with TFIIA and various transcriptional activators (Liu et al., 2008). TAF2 may not always be present in TFIID, and in its absence significant domain reorganization within TFIID is

observed (Papai et al., 2009). Therefore, it has been suggested that TFIID has a plastic structure that can be reorganized in response to activator signals, which would help TFIID to stably associate with a wide variety of promoters (Cler et al., 2009). Some TAFs are essential for assembly and integrity of TFIID. There is good evidence that TAF4, TAF5 and TAF10 play primary roles in TFIID assembly. Knockdown of specific TFIID subunits in Drosophila Schneider S2 cells indicate that TBP, TAF1, TAF2, and TAF11 contribute very little to complex stability. In contrast, knockdown of TAF4 leads to degradation of most of the TAFs (Wright et al., 2006). In yeast, temperature-sensitive mutations in the WD40 repeats of TAF5 causes a variety of allele-specific defects in the ability to support transcription and maintain the structure of the TFIID (Durso et al., 2001). In mammalian cells, experiments have shown that TAF10 is necessary for TFIID stability. Indeed, TAF10 gene disruption in mice leads to disassembly of TFIID (Mohan et al., 2003).

Function of TAFs

The core promoter of genes is defined as the minimal DNA sequence required for accurate transcription initiation *in vitro* (Roeder, 1996). One of the common core promoter elements in human genes is a TATA box located approximately 25 bp upstream of the transcription start site. The consensus sequence TATA(A/T)A(A/T)(A/G) is recognized directly by TBP (Hernandez, 1993). An additional core promoter element is the initiator (Inr). The Inr comprises a pyrymidine-rich sequence, PyPyAN(T/A)PyPy, and overlaps the transcription start site in some promoters (Kaufmann and Smale, 1994). The downstream promoter element (DPE) is located between +28 and +34 in many Drosophila TATA-less promoters and it functions cooperatively with the Inr for the binding of TFIID in the absence of a TATA box. The DPE contains the consensus sequence (A/G)G(A/T)CGTG. In the human TAF7 promoter a DPE is located between +29 and +35 (Kutach and Kadonaga, 2000, Smale and Kadonaga, 2003, Thomas and Chiang, 2006). Some promoters contain an additional core promoter element situated downstream of the transcription start site called the downstream core element (DCE). The DCE was discovered in the human β-globin promoter. It contains three sub-elements with core sequences of CTTC, CTGT and AGC (Lee et al., 2005, Purnell et al., 1994). TAFs are thought

to significantly widen the range of promoter recognition by TFIID. Firstly, association of TBP with TAFs helps to overcome chromatin barriers (Thomas and Chiang, 2006). Secondly, contacts between TAFII-Inr, TAFII-DPE and TAFII-DCE enables recognition of TATA-less promoters (Basehoar et al., 2004). TAF1/TAF2 recognizes the Inr sequence (Chalkley and Verrijzer, 1999). Human TAF6 and TAF9 exhibit DPE-binding specificity (Sengupta et al., 2009, Shao et al., 2005). UV photo-cross-linking results demonstrate that TAF1 interacts with the DCE sub-element in a sequence-dependent manner (Lee et al., 2005). Thus, absence of any of those TAFs will eliminate interactions at specific core promoter elements. Conversely, the TAF composition of TFIID may be dependent on the sequence of the core-promoter. Interestingly, it has been shown recently that erythroid Kruppel-like factor (EKLF) interacts with TAF9, but this interaction is not required for activation of all target promoters, which instead depends on the particular promoter architecture. This further supports the notion that the involvement of TAFs in gene transcription depends, at least in part, on promoter architecture (Sengupta et al., 2009). In addition, covalent modifications of histones can be recognized by structural domains of TAFs to modulate promoter recognition. For example, TAF1 contains two tandem bromodomain modules that bind selectively to multiply acetylated histone H4 peptides and the TAF3 PHD finger is highly selective for H3K4me3. Thus, modified histone tails interacting with TAFs might help recruit TFIID to specific chromatin-bound promoters (Jacobson et al., 2000, Vermeulen et al., 2007). All this, suggests that TAFs can serve as promoter selectivity factors that can orient TFIID on the DNA. Moreover, TAFs may be differentially required to activate genes due to differences in the sequence and chromatin structure of core promoters that can affect TFIID binding.

Transcriptional activation can be facilitated in several ways by DNA-binding activators. One mechanism implicates interaction of activators with components of PIC in enhanced recruitment and stabilization of GTFs at promoters. A number of reports demonstrate that specific TAFs directly interact with certain activators. For example, Drosophila TAF4 interacts with Sp1 (Dynlacht et al., 1991), human TAF7 interacts with Sp1, YY1, USF, CTF, adenovirus E1A protein and HIV-1 Tat (Chiang and Roeder, 1995, Fornage et al., 2011), N-TF-1 interacts with TAF2 and TAF6 (Chen et al., 1994), TAF9 interacts with the activation domain of Herpes viral protein 16 (VP16) (Kim et al., 2009a) and TAF1, TAF3, TAF6 and

TAF9 interact with the tumour suppressor protein p53 (Allende-Vega et al., 2007, Bereczki et al., 2008, Thut et al., 1995). These interactions suggest that TAFs can function as coactivators to bridge interactions between DNA-bound activators and the PIC. In agreement with the coactivator model, it has been demonstrated that mutations in activation domains of Sp1 that interact with TAFs affect activation of transcription (Gill et al., 1994), emphasizing that activator-TAF interaction is an important mechanism of transcriptional activation. Studies of regulation by Drosophila activators Bicoid and Hunchback suggest that activator-TAF contacts can direct synergistic transcriptional activation. Bicoid binds TAF4 and Hunchback binds TAF6. Activation by those factors requires the inclusion of their target TAFs into the TFIID complex since they fail to activate transcription mediated by TBP or a TBP-TAF1 complex alone (Sauer et al., 1995, Verrijzer and Tjian, 1996). Together these experiments indicate that activators may target different TAFs to recruit TFIID to promoters. In addition, holo-TFIID may integrate different signals from different activators.

The TAF1 subunit of TFIID possesses histone acetyltransferase (HAT) activity that acetylates histone H3 and H4 (Dunphy et al., 2000). Recruitment of HAT activity to a promoter results in acetylation of core histones and therefore contributes to overcoming chromatin barriers. This in turn facilitates stable binding of TFIID and other basal factors to the core promoter. In addition to histones, other components of the basal transcription machinery, such as TFIIH and TFIIF, may be acetylated by TAF1 (Imhof et al., 1997). HAT activity of TAF1 is also required for cell cycle progression (Dunphy et al., 2000). It has also been demonstrated that Drosophila TAF1 phosphorylates histone H2B and this phosphorylation is critical for transcriptional activation events that promote cell cycle progression and development. In addition, TAF1, as a histone-specific ubiquitin-activating/conjugating enzyme, mediates monoubiquitination of histone H1 *in vitro* (Pham and Sauer, 2000, Maile et al., 2004). The enzymatic activities of TAF1 that modify histones therefore contribute to the importance of this protein in gene transcription. Indeed, inhibition of the HAT activity of TAF1 represses transcription of MHC class I genes. Importantly, TAF7 regulates TAF1 function by binding to TAF1 and inhibiting its HAT activity (Gegonne et al., 2001). In addition, TAF1 binds to acetylated histones by two tandem bromodomains located in the C-terminal region, which may stabilize the binding

of TFIID to the core promoter (Naar et al., 2001). The enzymatic activities of TAF1 may both facilitate TFIID binding to core promoters and alter the chromatin environment to allow favourable PIC assembly.

As mentioned above, TAF7, a 55-kDa TFIID component, binds to TAF1 and negatively regulates its HAT activity. It has been reported that TAF7 dissociates from the TFIID complex upon addition of pol II to the assembling PIC. This release is concomitant with phosphorylation of TAF1, which reduces the binding of TAF1 to TAF7. Additional studies have shown that TAF7 functionally interacts with the TFIIH and the elongation factor P-TEFb. Interaction of TAF7 with TFIIH and P-TEFb inhibits their kinase activities. In addition, TAF7 inhibits steps after PIC assembly in *in vitro* transcription reactions. Accordingly, it was proposed that TAF7 acts as a checkpoint regulator and delays transcription until the PIC complex is assembled (Gegonne et al., 2008, Gegonne et al., 2006). Notably, a similar interaction occurs between TAF7 and the HIV transactivator protein Tat. Like TAF7, Tat interacts with TAF1 and P-TEFb. In addition, TAF7 and Tat inhibit TAF1 HAT enzymatic activity and activate P-TEFb autophosphorylation. In addition, TAF7 and Tat modulate phosphorylation of CTD by TFIIH and P-TEFb (Gegonne et al., 2006, Kobor and Greenblatt, 2002, Zhou et al., 2001).

Despite the diverse roles in transcription activation, TFIID is much more complex than previously thought (Goodrich and Tjian, 2010). Some TAFs have gene paralogues, which play an important role in pol II-dependent transcription. They are often expressed in a tissue- or cell-type-specific manner. The TAF4 paralogue, TAF4b is expressed in B lymphocytes and a unique TAF4b-containing TFIID complex can be isolated from these cells. TAF4b is also essential for proper development and function of the mouse ovary, where it mediates the transcription of a subset of genes required for proper folliculogenesis (Dikstein et al., 1996, Freiman et al., 2001). TAF5L (a TAF5 paralogue) and TAF7L (a TAF7 paralogue) also play important roles in developmental regulation of gene expression. Both are required for spermatogenesis and probably interact with germ cell-specific transcription cofactors to regulate transcription during development (Hiller et al., 2001, Pointud et al., 2003). Interestingly, TAFs are not expressed in some differentiated cell types including hepatocytes (D'Alessio et al., 2011). Furthermore, embryonic stem

cells contain a noncanonical TFIID complex lacking several core subunits (Maston et al., 2012).

To sum up, TAFs are critical components of the machinery responsible for the regulated transcription of eukaryotic genes transcribed by pol II and their biological significance has made them the target of extensive study. TAFs can function as promoter recognition factors, as coactivators and enzymatic modifiers of other proteins. Some are important for the stability of complexes they form. The importance of these functions may vary from promoter to promoter. Interaction between TAFs and core promoters may cause allosteric effects in TFIID that modulate its activity. TAFs are targets of activators and different activator-coactivator combinations may selectively regulate transcription. TAF1 acetylation of histones and monoubiquitination of histone H1 may facilitate and stabilize binding of TFIID. The existence of multiple TAF- and TBP-containing complexes may indicate that gene specificity is ensured by interaction of combinations of activators with a range of different basal transcription machineries.

1.1.3 Transcription elongation

The main rate-limiting step of transcription elongation is promoter-proximal pausing. Two negative transcription elongation factors (N-TEFs), DRB sensitivity-inducing factor (DSIF) and negative elongation factor (NELF), are recruited to the elongation complex in the promoter proximal region to induce pol II pausing (Yamada et al., 2006, Andrulis et al., 2000, Yamaguchi et al., 1999). Release from the pause site is mediated by the action of the Cdk9 subunit of P-TEFb, which phosphorylates DSIF, NELF and Ser2 of the pol II CTD (Yamada et al., 2006, Saunders et al., 2006). Phospho-NELF leaves the elongation complex and phospho-DSIF turns into a transcriptional activator, which remains associated with pol II (Gomes et al., 2006, Yamada et al., 2006). Accordingly, genome-wide studies demonstrate that NELF occupancy is mostly restricted to promoter proximal regions while DSIF is distributed over the coding regions (Rahl et al., 2010). Pol II can enter productive elongation with the action of P-TEFb (Marshall et al., 1996). The role of P-TEFb action in stimulating productive elongation was elucidated using 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB), which is a purine nucleoside analogue that inhibits

Cdk9 activity (Marshall and Price, 1995). Treatment of cells with DRB results in the production of shortened transcripts, diagnostic of the inability of pol II to transcribe through the pol II stalling site in the absence of P-TEFb activity (Marshall and Price, 1992). P-TEFb is composed of Cdk9 kinase and a cyclin. Cdk9 is a Ser/Thr proline-directed kinase and it is ubiquitously expressed. It exists in two isoforms Cdk9-42 and Cdk9-55 due to two transcription start sites (Marshall and Price, 1995). Both isoforms localize to the nucleus (Peng et al., 1998b). The ratio of Cdk9-42/Cdk9-55 is cell type-dependent (Liu and Herrmann, 2005). Cdk9 shares 47% homology with other Cdk's (Marshall and Price, 1995). Cdk9 interacts with T-type cyclins – T1, T2a, T2b and Cyclin K (Peng et al., 1998a, Fu et al., 1999) (see Table 4). However, recently several mass spectrometry studies failed to demonstrate that Cyclin K is associated with Cdk9 (Barboric et al., 2009, Bezstarosti et al., 2010). In addition, a recent study revealed that Cyclin K associates with Cdk12 and Cdk13 (Blazek et al., 2011). The various complexes are cell-type specific and differentiation and activation dependent (Peng et al., 1998b). Interestingly, the function of Cdk9 is modified by various viral gene products. The human immunodeficiency virus Tat protein interacts with P-TEFb and changes the specificity of phosphorylation of the pol II CTD (Herrmann and Rice, 1995, Marshall et al., 1996). Tamrakar and colleagues showed that human cytomegalovirus infection induces the hyperphosphorylation of pol II CTD and this is associated with changes in activity, localization and abundance of Cdk9 and Cdk7 (Tamrakar et al., 2005). Kaposi's sarcoma-associated herpesvirus K-cyclin interacts with Cdk9 and stimulates Cdk9-mediated phosphorylation of the p53 tumor suppressor (Chang and Li, 2008). In 2008, Durand and Roizman concluded that Cdk9 plays an important role in the optimization of expression of genes regulated by HSV-1 infected cell protein 22 (ICP22) (Durand and Roizman, 2008).

Cdk	Cyclin partner	Kinase activity on the pol II CTD
Cdk7	CycH	Ser5 and Ser7
Cdk8	CycC	CTD
Cdk9 (42kDa)	CycT1 CycT2a/b	Ser2
Cdk9 (55kDa)	CycT1	Ser2
Cdk11 (46kDa)	CycL1 CycL2	
Cdk11 (58kDa)	CycL1 CycL2 CycD3	
Cdk11 (110kDa)	CycL1 CycL2	CTD
Cdk12	CycK	Ser2
Cdk13	CycK	CTD

Table 4. Transcription-cycle related Cdks.
Adapted from Kohoutek and Blazek (2012).

After the transition to productive elongation, pol II is affected by general elongation factors (Reines et al., 1999). Isolated elongation factors can increase the processivity of pol II *in vitro*. Examples of such elongation factors are TFIIS, DSIF, CSB, Elongins, ELL, NELF, Fcp1, TFIIF (Sims et al., 2004). Some elongation factors, such as Set1, Set2, PAF, have no direct impact on the catalytic activity of pol II. Others that include FACT and Spt6, act by altering nucleosome structure to allow passage of pol II (Hartzog et al., 2002). Elongation complexes are not uniform in their composition. The rate of elongation may be important to provide sufficient time for RNA processing (Reinberg and Roeder, 1987).

1.1.4 mRNA processing

In eukaryotes, before gene transcripts are transported to the nucleus they have to undergo processing events. mRNA processing involves the addition of a cap structure at the 5'end, splicing out of introns and synthesis of polyadenylated (poly(A)) tail at the 3' end. Interestingly, the CTD of pol II couples transcription and mRNA processing (Hirose and Manley, 2000). Truncation of the CTD causes defects in capping, polyadenylation and splicing (McCracken et al., 1997).

Pre-mRNA splicing

mRNA is synthesized as longer precursor RNAs (pre-mRNAs) in the nucleus. Before it is exported to cytoplasm, the non-protein-coding introns are removed and protein-coding exons are ligated in the process known as RNA splicing. RNA splicing is performed by a multi-protein complex named the spliceosome. The spliceosome is a highly dynamic complex that is assembled in a stepwise manner. It consists of UsnRNPs U1, U2, U4, U5 and U6 and a large number of additional proteins (Chen and Manley, 2009, Wahl et al., 2009). There are four conserved sequence elements that are recognized by the spliceosome: the 5' splice site (5'ss), the branch point (BP), the polypyrymidine tract (PPT) and the 3' splice site (3'ss) (Figure 1-1). The first step of RNA splicing requires binding of U1 snRNP to the 5' splice site; 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 (Ruskin et al., 1988, Berglund et al., 1998).

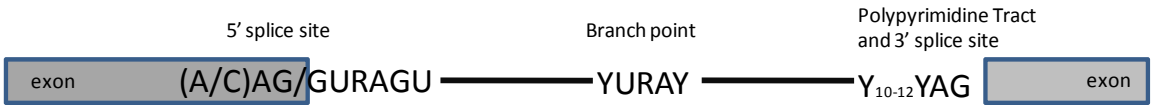


Figure 1-1. Sequences required for splicing.

Consensus sequences for the 5' splice site, branch point, polypyrimidine tract and 3' splice site are noted. R is a purine residue and Y is a pyrimidine residue. Exons are represented by grey boxes and introns by horizontal lines.

The next step is ATP-dependent; SF1 is displaced by U2 snRNA, which base pairs with pre-mRNA at the branch point (Zhuang and Weiner, 1989). Subsequently, the U4, U5 and U6 snRNPs are recruited via interactions between U5 snRNP and the flanking exons and U6 snRNP with U2 snRNP (Wu and Manley, 1991). Upon dissociation of U1 snRNP and U4 snRNP, the spliceosome becomes catalytically active. Introns excision occurs in two chemical steps. Firstly, 5' splice site cleavage and lariat formation occurs. The first base of the intron forms a lariat with the adenosine at the branch site. Next, a phosphodiester bond is formed between the last base of the upstream exon and the first base of the downstream exon. As a result of this reaction, the exons are joined and the introns are released (Konarska et al., 1985, Proudfoot et al., 2002). As mentioned before, the pol II CTD coordinates transcription with pre-mRNA splicing (Bentley, 2005). The CTD provides a molecular platform for binding of splicing regulatory factors (Morris and Greenleaf, 2000) ensuring spliceosome assembly on the emerging pre-mRNA. The CTD plays a critical role in the nuclear distribution of splicing components. Overexpression of the large subunit of pol II induces nuclear reorganization of splicing factors (Du and Warren, 1997). Furthermore, the rate of transcription elongation can influence splicing. Pol II with a reduced elongation rate can stimulate the inclusion of alternative exons (Kornblihtt et al., 2004).

mRNA 3' end processing

The 3'ends of most mRNAs are formed by RNA co-transcriptional cleavage followed by polyadenylation. Cleavage is directed by a conserved AAUAAA sequence (polyA site), and a GU rich sequence (DSE) (Gil and Proudfoot, 1984). The largest subunit of the Cleavage and Polyadenylation Specificity Factor (CPSF), CPSF160, recognizes the AAUAAA signal. CPSF160 links polyadenylation to transcription by interacting with both TFIID and with the pol II CTD (Dantonel et al., 1997). Another component of the mRNA polyadenylation factory is the Cleavage stimulation Factor (CstF). The 64kDa subunit of CstF binds to the GU signal of the RNA (Proudfoot et al., 2002). Additional proteins assemble with CPSF160 and CstF64. To direct the cleavage reaction, two proteins are essential: the Cleavage Factor I and II (CF I and CF II) (Takagaki et al., 1989). Endonucleolytic cleavage targets a CA

sequence between the AAUAAA and the DSE sequence. Following cleavage, the poly-A polymerase (PAP) enzyme adds up to 200 A residues to the RNA 3' end produced by cleavage. The efficiency of 3' cleavage depends on the strength of the poly(A) site (Eggermont and Proudfoot, 1993).

1.1.5 Structure of the β-actin gene

For the research described herein, I have used the β-actin gene as a model protein-coding gene. β-actin is ubiquitously and abundantly expressed in all mammalian non-muscle cells. It is the major component of the microfilamentous structures. Cytoskeletal actin is implicated in intracellular movement of organelles and cell motility (Ng et al., 1985). The β-actin protein-coding gene is a single copy gene with a transcription unit of up to 5 kb, encoding transcripts that are spliced and polyadenylated (Figure 1-2) (Gromak et al., 2006). The β-actin gene is located on chromosome 7 (Ng et al., 1985) and contains five introns (Nakajima-Iijima et al., 1985).

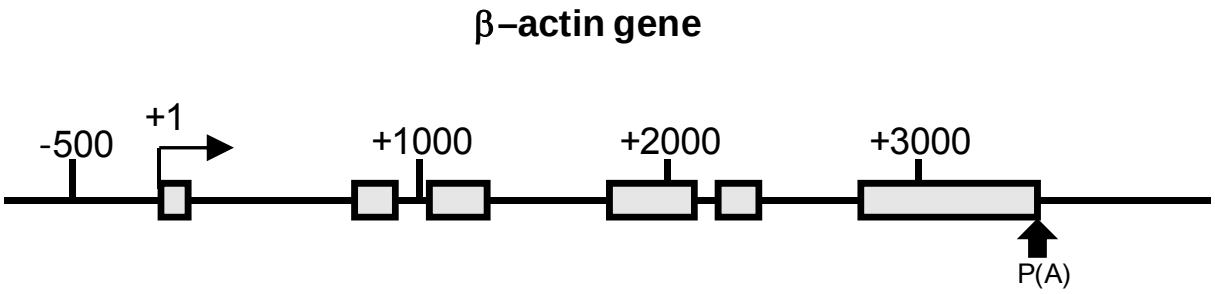


Figure 1-2. Schematic of the β-actin gene.

The arrow represents the transcription start site and exons are denoted with boxes. The polyadenylation signal P(A) is indicated with an arrow. Non-coding regions are denoted by a horizontal line. +1 represents the first base of the β-actin coding region.

The human β-actin gene promoter is particularly strong and can be characterized by the presence of the TATA box (-29 to -23), CCAAT box (-91 to -87) and CCArGG box (-62 to -53) (Ng et al., 1985, Danilition et al., 1991). The CCAAT box and proximal CCArGG box are binding sites for the transcription factors nuclear factor Y (NF-Y) and serum response factor p67SRF, respectively (Danilition et al., 1991).

1.2 Small Nuclear (sn)RNAs

Human snRNA genes are short, intron-less and encode transcripts that are non-polyadenylated (Egloff et al., 2008). They were named after their nuclear localization and their small size. snRNAs are required for splicing of pre-mRNA (Hernandez, 2001, Wahl et al., 2009) and histone 3'end formation (Schaufele et al., 1986, Bond et al., 1991, Dominski and Marzluff, 2007). They are exemplified by the well-characterized U2 snRNA genes. Approximately 20 U2 snRNA genes are clustered in 6.2 kb tandem repeats on chromosome 17q21-q22 (Figure 1-3) (Van Arsdell and Weiner, 1984, Westin et al., 1984). The U2 snRNA genes contain compact yet powerful TATA-less promoters (Schramm et al., 2000), providing a good model system to study gene regulation. The essential proximal sequence element PSE is located 50 bp upstream of the transcription start site of the U2 snRNA gene. This element defines the start site of transcription and recruits the snRNA-specific PSE-binding transcription factor PTF (also known as SNAPc and PBP) (Sadowski et al., 1993). The distal sequence element DSE, located 220 bp upstream of the start site of transcription, functions as an enhancer-like element to activate transcription and comprises binding sites for the trans-activating factors Oct-1 and Sp1 (Murphy et al., 1992, Mattaj et al., 1985, Tanaka et al., 1988). In addition, the human U2 snRNA gene promoter contains three Sp1 binding sites (Ares et al., 1987). Interestingly, not all activators are compatible with the snRNA promoters. For example, the viral activator VP16, which enhances transcription of mRNA genes, cannot activate transcription in the pol II-transcribed snRNA genes (Das et al., 1995). The 3'box, the snRNA gene-specific RNA 3' end processing element, is located 9-19 nucleotides downstream of the 3'end of the region encoding the mature snRNA (Uguen and Murphy, 2003). The primary transcript of

the U2 genes extends for up to 1kb (Medlin et al., 2003, Egloff et al., 2008) [O'Reilly et al., 2013 in press].

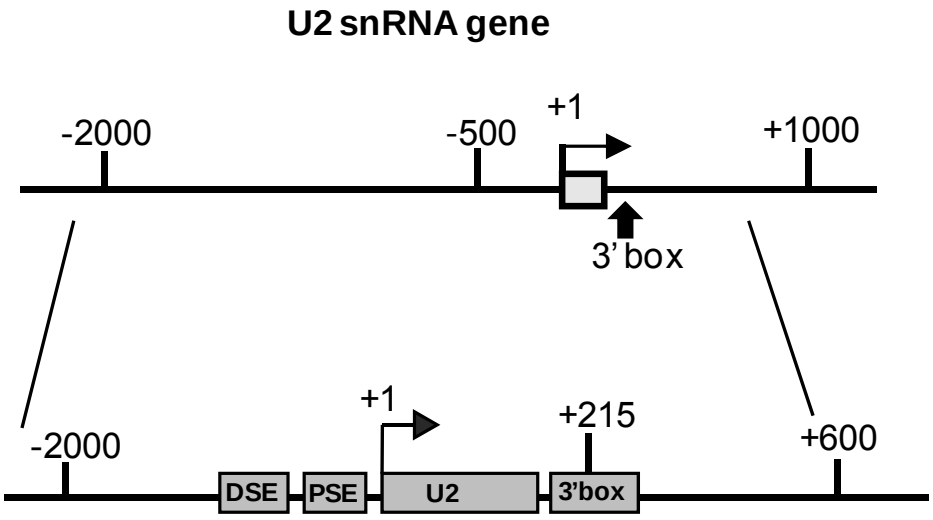


Figure 1-3. Schematic of the U2 snRNA gene.

Above the detailed schematic of the U2 snRNA gene, is shown a schematic of the U2 snRNA gene on the same scale as the β-actin gene schematic in Figure 1. The arrow represents the transcription start site. The distal sequence element (DSE), the proximal sequence element (PSE) in the promoter, the snRNA-encoding region (U2) and the 3'box RNA processing element are denoted with boxes. +1 represents the first base of the U2 snRNA coding region.

1.2.1 Transcription of snRNA genes

Transcription of snRNA genes has been extensively characterized (Egloff et al., 2009). The GTFs TFIIA, TFIIB, TFIIF, TFIIE and TFIIH are implicated in transcription of snRNA genes (Kuhlman et al., 1999, Glover-Cutter et al., 2009). Interestingly, recombinant TBP is necessary for *in vitro* transcription from these TATA-less promoters but cannot be substituted by TFIID (Bernues et al., 1993). In addition, the efficiency of *in vitro* transcription of snRNA genes is very low, particularly in light of their robust transcription *in vivo* (Gunderson et al., 1990). Interestingly, during expression of snRNA genes, the 3' box is only recognized if transcription is initiated from a pol II-dependent snRNA gene promoter (Hernandez and Weiner, 1986, Egloff et al., 2008, de Vegvar et al., 1986). Furthermore, inhibition of Cdk9 does not inhibit transcription of snRNA genes but leads to incorrect RNA 3' end formation. This suggests that P-TEFb does not function as an

elongation factor for snRNA genes as it does for protein coding genes (Medlin et al., 2003). This highlights significant differences in the control of elongation of transcription of snRNA and protein-coding genes.

1.2.2 3’ end formation of snRNAs

Unlike mRNAs, snRNAs are not polyadenylated. 3’-end formation of U2 snRNA depends on the 3’box RNA processing element. This element directs the production of a 3’-extended pre-snRNAs. The short nonpolyadenylated 3’-extended precursors are exported to the cytoplasm for further 3’ trimming and incorporation into small nuclear ribonuleoparticles (snRNPs) (Kleinschmidt and Pederson, 1987, Uguen and Murphy, 2003). Correct 3’ end formation also depends on the presence of an snRNA promoter and the CTD of pol II (Hernandez, 1985, Uguen and Murphy, 2003, de Vegvar et al., 1986). This indicates a close link between snRNA gene transcription and RNA 3’ end processing. Medlin and colleagues demonstrated that P-TEFb and subsequently the phosphorylation of Ser2 of pol II CTD is needed for co-transcriptional recognition of the 3’ box and accurate 3’ end formation of snRNAs (Medlin et al., 2003). Interestingly, Ser7 of the pol II CTD and its phosphorylation are specifically required for snRNA gene transcription and correct 3’ end formation (Egloff et al., 2007, Chapman et al., 2007). Ser7 phosphorylation is also critical for recruitment of an snRNA gene-specific complex, named Integrator. Integrator is a macromolecular 12-subunit complex that binds to the phosphorylated CTD of pol II (Baillat et al., 2005, Egloff and Murphy, 2008a, Egloff et al., 2007). It is associated with U1 and U2 snRNA genes, where it directs 3’end formation by recognizing the 3’ box (Baillat et al., 2005). The Int9 and Int11 subunits of the Integrator complex have homology to the CPSF-73 and CPSF-100 subunits of the Cleavage and Polyadenylation Specificity Factor (Dominski et al., 2005b). CPSF-73 is the endoribonuclease involved in pre-mRNA and replication-activated histone pre-mRNA 3’ end cleavage (Dominski et al., 2005a). CPSF-100 interacts with CPSF-73 and modulates its activity. Int9 and Int11 are therefore thought to be responsible for the first cleavage during the formation of the 3’ end of pre-snRNAs (Baillat et al., 2005). Loss of Integrator causes a defect in transcription in addition to a defect in 3’ processing, (Egloff et al., 2007) [O’Reilly et al., 2013 in press]. Recently, Otani and colleagues suggested that processing of U1 and U2 snRNAs may be

involved in cell differentiation steps (Otani et al., 2013). In their study they demonstrated that expression of the Int6 and Int11 subunits of Integrator are upregulated during adipose differentiation. After completion of differentiation the levels return to basal levels. This suggests a critical role for Integrator in adipose differentiation. The double phosphorylation mark Ser2P/Ser7P is necessary for efficient binding of Integrator to the CTD (Egloff et al., 2010). This further highlights the role of the pol II CTD in pre-snRNA processing.

The described differences to mRNA protein-coding genes make snRNA genes an interesting model to study the molecular mechanisms of pol II transcription.

1.3 Modification of the pol II CTD by the HSV-1 ICP22 protein

Herpesviruses are significant human pathogens associated with a range of mild to severe diseases. Although herpes simplex virus (HSV) infections are generally self-limiting in healthy individuals, they can cause sarcoma, encephalitis, meningitis and keratitis and pose a significant mortality risk in infants and immuno-compromised adults (Knipe and Cliffe, 2008, Galdiero et al., 2013). HSV remains latent in the host for life after infection and can be reactivated. A complete understanding of HSV infection and reactivation from latency requires a detailed knowledge of the molecular mechanisms involved in the regulation of viral replication. During HSV-1 productive infection, three classes of genes are sequentially expressed; first, the immediate early genes (α-genes; IE genes), then, the early genes (β-genes) and finally, the late genes (γ-genes; L genes) (Weir, 2001, Honess and Roizman, 1974). The transcription of α-genes is initiated by the viral tegument protein VP16 and the host cell proliferation regulator HCF-1 with the transcription factor Oct-1 (Wysocka and Herr, 2003). The five immediate early proteins (ICP0, ICP4, ICP22, ICP27 and ICP47) are the first viral proteins to be synthesized in the infected cell (Bastian et al., 2010). The studies described here centre on the role of immediate early protein 22 (ICP22).

1.3.1 The properties of ICP22

ICP22 is encoded by the HSV-11 U_S1 gene and comprises 420 amino acid residues (Figure 1-3). Viral deletion mutants without the U_S1 gene have a reduced capacity to establish latency in animal models. In addition, deletion of this gene causes viral growth defects in some cell lines (Sears et al., 1985, Poffenberger et al., 1994). The Us1 gene encoding the ICP22 protein was the first gene deleted by genetic engineering using selectable markers (Post and Roizman, 1981). The predicted molecular weight of ICP22 is 46,522 Da and it migrates with a size of 68 kDa (Ackermann et al., 1985). In addition, a second protein called $U_S1.5$ is also expressed from the U_S1 gene (Ogle and Roizman, 1999). Several translation start sites are reported for $U_S1.5$ - at residues 171, 147 and 90 (Poon et al., 2000, Ogle and Roizman, 1999, Bowman and Schaffer, 2009). The distinct functions of ICP22 and $U_S1.5$ are not completely understood. Homologs of ICP22 are encoded by all alphaherpesviruses. Recently, Kolb and colleagues, identified a 63 residue motif that is conserved in all alphaherpesvirus U_S1 homologs (Figure 1-4) (Kolb et al., 2011). Interestingly, predicted cyclin binding sites were identified in this region (Figure 1-4).

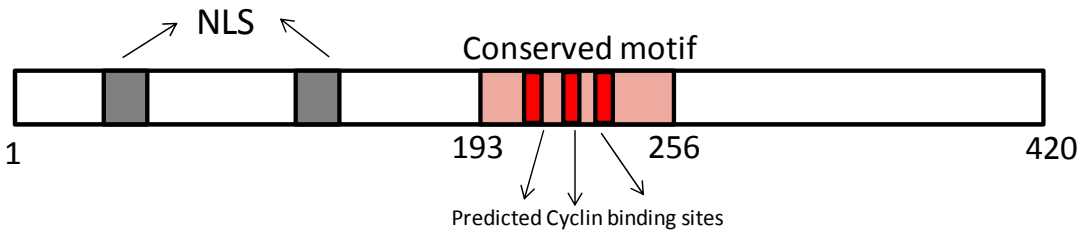


Figure 1-4. Schematic of the ICP22 protein.
Nuclear localization signals (NLS), the conserved 63-residue motif and predicted cyclin binding sites are indicated.

For many years, study of the ICP22 protein was limited because of difficulties with its expression. In 2007, Fraser and Rice first demonstrated the use of the CMV promoter in the pcDNA3 plasmid for expression of ICP22 (Fraser and Rice, 2007). In 2009, Bowman and colleagues determined that ICP22 can be expressed from its own promoter when transfected cells are infected with an ICP22 null virus or cotransfected with an ICP0 expression vector (Bowman et al., 2009). ICP22 is localized in the nucleus in small nuclear

bodies (Stelz et al., 2002, Leopardi et al., 1997). The function of small dense nuclear bodies is not clear but it was suggested that they are involved in transcriptional regulation (Prod'hon et al., 1996). The ICP22 protein contains two independent nuclear localization signals; NLS1 maps to ICP22 amino acid position 16-31 and NLS2 to amino acid position 118-131 (Stelz et al., 2002) (Figure 1-4). U_S1.5 protein localization is different from ICP22. It is distributed in the cytoplasm and nucleus (Cun et al., 2006), which suggests it may have an alternative function.

ICP22 is heavily post-translationally modified by phosphorylation of serine and tyrosine residues, guanylylation and adenylation (Brandt and Kolb, 2003, Blaho et al., 1993, Mitchell et al., 1997, Mitchell et al., 1994, Asai et al., 2007, Blaho et al., 1997). These modifications include phosphorylation by viral kinases U_S3 and U_L13 (Purves et al., 1993, Purves and Roizman, 1992). It has been shown that tyrosine 116 is a potential phosphorylation site and is associated with virulence (Brandt and Kolb, 2003). Multiple isoforms of the protein are found in infected cells. Kolb and colleagues analysed the amino acid sequence variability in the herpesvirus U_S1 homologs (Kolb et al., 2011). The variability is the highest in the amino-terminal third of the protein.

1.3.2 The function of ICP22

The HSV-1 ICP22 protein performs several functions in infected cells, including regulation of HSV-1 gene expression. It is responsible for the efficient expression of some late genes (Purves et al., 1993). However, some studies indicate that ICP22 acts as a transcription inhibitor (Prod'hon et al., 1996, Cun et al., 2006). In a recent study, it has been demonstrated that ICP22 represses transcription of the viral α-, β- and γ-genes by blocking the recruitment of P-TEFb to promoters. In contrast, VP16 overcomes the inhibitory effects on α-gene transcription (Guo et al., 2012). Interestingly, viruses with deletions and insertions in the ICP22 gene are attenuated in the central nervous system of mice (Sears et al., 1985). Importantly, the ICP22 protein is also involved in regulating the modification of pol II (Rice et al., 1994, Rice et al., 1995) and Cdk9 might be involved in this process (Durand and Roizman, 2008). Finally, the ICP22 protein was shown to

decrease cyclin A and B levels and increase cyclin-dependent kinase 1 (Cdk1) activity (Advani et al., 2000).

The role of ICP22 in the modification of RNA polymerase II

As noted previously, it was shown that ICP22 alters the CTD phosphorylation state of pol II (Fraser and Rice, 2007, Rice et al., 1995, Spencer et al., 1997). ICP22 mediates two distinct effects on pol II: the induction of an intermediately-migrating form of pol II (pol Iii) and the loss of pol II forms phosphorylated on Ser2. Pol IIi replaces the normal hyperphosphorylated form of pol II in HSV-1 infected cells. Pol Iii exhibits increased electrophoretic mobility therefore has been named pol IIi for "intermediately migrating". The induction of pol IIi requires the ICP22 protein and another viral factor, U_L13 (Long et al., 1999). The loss of pol II phosphorylated on Ser2 is triggered by ICP22 (Rice et al., 1994, Fraser and Rice, 2007). Bastian and colleagues demonstrated that 240-340 residues of ICP22 are required for pol II modification (Bastian et al., 2010). In addition, later in infection, another pathway involving ICP27 contributes to the loss of Ser2 (Rice et al., 1995, Fraser and Rice, 2005, Fraser and Rice, 2007). Durand and colleagues demonstrated by pull-down experiments that Cdk9 is bound to ICP22 (Durand et al., 2005). They concluded that the interaction is at least partially dependent on the presence of the U_S3 kinase. Further studies confirmed that ICP22, as well as VP16, can form a complex with P-TEFb in an *in vivo* transient expression system and in HSV-1 infected cells (Guo et al., 2012).

Interestingly, it was suggested that Cdk9 plays an important role in the optimization of expression of genes regulated by ICP22 (Durand and Roizman, 2008). The Cdk9 inhibitor, DRB causes a reduction in the accumulation of ICP22 and ICP22-regulated proteins U_L41, U_L38 and U_S11. Similar results were obtained after siRNA-mediated knockdown of Cdk9. Cdk9 is in a complex with pol II in wild-type virus-infected cells; in cells infected with a ΔICP22 or ΔUL13 mutant, Cdk9 and pol II do not co localize, although they are in close proximity. Accordingly, it was suggested that Cdk9 recruits ICP22 into the transcriptional complex (Durand and Roizman, 2008).

1.4 Aims of the thesis

The snRNA genes transcribed by pol II are ubiquitously and highly expressed and provide an ideal model to study mechanisms underlying transcription and co-transcriptional RNA processing. The overall aim of this study is to understand the fundamental mechanisms of snRNA gene transcription, focusing particularly on the factors that make up the PIC at snRNA promoters and CTD modification enzymes involved in expression of snRNA genes. Chapter 2 presents the results of experiments aimed to characterize the snRNA gene-specific TAF complex that may play a central role in facilitating gene-type-specific RNA processing. In Chapter 3, I describe the results of investigating the role of the human RPAP2 and Ssu72 proteins in snRNA gene expression. Chapter 4 presents the experiments exploring the phenomenon of inhibition of CTD Ser2 phosphorylation by the HSV-1 ICP22 protein, which is thought to interact with P-TEFb. I was particularly interested to determine whether Cdk9 is acting as a Ser2 kinase during expression of snRNA genes, which has been a subject of controversy in recent years. Understanding how CTD modifications function and how they are regulated is crucial to understanding the molecular mechanisms controlling gene expression. In Chapter 5, I will discuss the implications of my findings and future aims.

Chapter 2| A novel TBP-TAF complex on RNA polymerase II-transcribed snRNA genes

2.1 Introduction

The snRNA genes transcribed by pol II, which includes the genes for the U1 and U2 spliceosomal snRNAs, are structurally different from protein-coding genes (Figure 2-1). These genes are short, the transcripts are intronless and 3' end formation is directed by the snRNA gene-specific 3'box RNA 3'end processing element rather than a poly(A) signal. The promoters of these genes minimally comprise an essential Proximal Sequence Element (PSE), with some properties of a TATA box and an enhancer-like Distal Sequence Element (DSE) (Egloff et al., 2009) (Figure 2-1B). The GTFs TFIIA, TFIIB, TFIIF, TFIIE and TFIIH are implicated in transcription of snRNA genes (Kuhlman et al., 1999, Glover-Cutter et al., 2009). At the time of starting work for this chapter, only TAF5 had been shown to be present at the promoters of snRNA genes (Oelgeschlager, 2002) and the complement of TAFs associated with these genes was unclear. Bernues and colleagues demonstrated that recombinant TBP is necessary for *in vitro* transcription from U1 snRNA gene promoters but cannot be substituted by TFIID (Bernues et al., 1993). The complement of factors that make up the PIC at snRNA promoters is therefore only partially understood. Interestingly, during expression of snRNA genes, pol II does not appear to make the transition to productive elongation during transcription of these genes, as happens during transcription of protein-coding genes (Egloff et al., 2009). In addition, the 3' box is only recognized if transcription is initiated from a pol II-dependent snRNA gene promoter (Egloff and Murphy, 2008b, Hernandez, 1985, Lund et al., 1985). These gene type-specific properties may, at least in part, be due to differences in factors recruited by the promoter, including TAFs.

In order to better understand the molecular basis for the differential control of transcription and RNA processing in expression of snRNA and protein-coding genes, I have carried out a comparative analysis of the TAF association with the well-characterized U2 snRNA and β-actin protein-coding genes in HeLa cells. HeLa cells were the cell line of my choice because they are of human origin, grow well in cell culture and are widely used for gene expression studies.

2.2 Results

2.2.1 Pol II and TBP associate with the β-actin and U2 snRNA genes

The β-actin protein-coding gene is a single copy gene with a transcription unit of up to 5 kb, encoding transcripts that are spliced and polyadenylated (Figure 2-1A) (Gromak et al., 2006). Approximately 20 U2 snRNA genes are clustered in 6.2 kb tandem repeats on chromosome 17q21-q22 (Figure 2-1B) (Van Arsdell and Weiner, 1984, Westin et al., 1984). The essential PSE is located 50 bp upstream of the transcription start site of the U2 snRNA gene and it binds a snRNA-specific PSE-binding transcription factor PTF (also known as SNAPc and PBP) (Sadowski et al., 1993, Yoon et al., 1995). The DSE, located 220 bp upstream of the start site of transcription, functions as an enhancer-like element to activate transcription and comprises binding sites for the trans-activating factors Oct-1 and Sp1 (Murphy et al., 1992, Mattaj et al., 1985, Tanaka et al., 1988). The 3'box, the snRNA gene-specific RNA 3' end processing element, is located 9-19 nucleotides downstream of the 3'end of the region encoding the mature snRNA (Uguen and Murphy, 2003).

I have analyzed the association of pol II and TBP with the human β-actin and the U2 snRNA genes by chromatin immunoprecipitation (ChIP), coupled with quantitative real-time PCR (qPCR) analysis. ChIP localizes pol II across the transcribed region of both genes with the highest levels at the promoter region as previously published (Egloff et al., 2009). Pol II levels remain high over the RNA-encoding region of the U2 snRNA gene (Figure 2-1B), whereas the levels across the β-actin gene steadily reduce towards the poly(A) site (Figure 2-1A). Interestingly, a second peak of pol II is detected on the β-actin gene downstream of the poly(A) site (Figure 2-1A, regions 7 and 8), as already described by Gromak et al. (Gromak et al., 2006). This may be diagnostic of polymerase pausing (Gromak et al., 2006), looping between this region and the promoter (Hampsey et al., 2011) or the presence of a downstream promoter. TBP is highest at the start of transcription on both genes, as expected (Figure 2-1A and 2-1B). Interestingly, a second peak of TBP is detected just upstream of the second peak of pol II on the β-actin gene (Figure 2-1A, region 6), which may suggest the presence of another promoter.

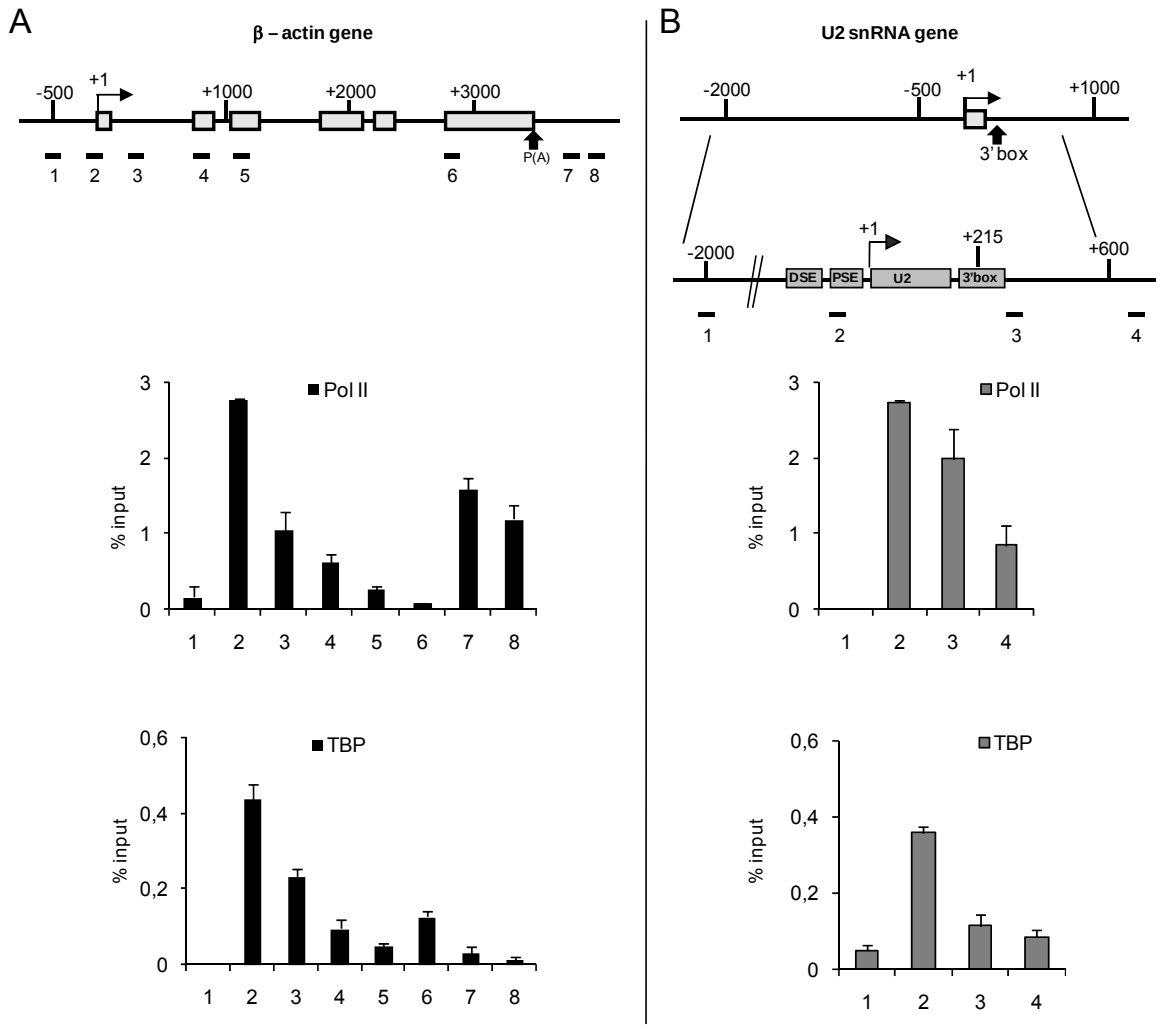


Figure 2-1. Pol II and TBP association with the β-actin and U2 snRNA genes.

(A) A schematic of the β-actin gene is shown at the top of this and subsequent figures. The polyadenylation signal P(A) is indicated with an arrow. The arrow represents the transcription start site and exons are denoted by boxes. Horizontal lines underneath the gene schematics mark the position of the quantitative real-time PCR (qPCR) primer pairs. Primer pairs indicated here were used in all further analyses. The results of qPCR analysis of the output of ChIP using anti-pol II or anti-TBP antibodies is shown under the schematic. Error bars indicate the standard deviation of means from at least three independent experiments in this and subsequent figures. (B) A detailed schematic of the U2 snRNA gene is shown at the top of this and subsequent figures. Above this, is shown a schematic of the U2 snRNA gene on the same scale as the β-actin gene schematic in (A). The arrow represents the transcription start site. The distal sequence element (DSE), the proximal sequence element (PSE) in the promoter, the snRNA-encoding region (U2) and the 3' box processing element are denoted by boxes. Horizontal lines underneath the gene schematics mark the position of the quantitative real-time PCR (qPCR) primer pairs. Primer pairs indicated here were used in all further analyses. The results of qPCR analysis of the output of ChIP using anti-pol II or anti-TBP antibodies is shown under the schematic. Error bars indicate the standard deviation of means from at least three independent experiments in this and subsequent figures.

2.2.2 TAF6, TAF9, TAF11 and TAF13 are associated with the promoters of the β-actin and U2 snRNA genes

Next, I have analysed the association of TAF6, TAF9, TAF11 and TAF13 with the human β-actin and the U2 snRNA genes by ChIP, coupled with qPCR analysis. TAF6, TAF9, TAF11 and TAF13 are associated with both the U2 snRNA and β-actin genes (Figure 2-2A and 2-2B). Analysed TAFs on the β-actin gene and on the U2 snRNA gene have a similar profile to TBP on the corresponding gene, ie, they peak at the promoter region. Interestingly, all of these TAFs also show an increase in association with region 6 of the β-actin gene, which is where the second peak of TBP is located (Figure 2-2A). As mentioned before, TAF6 forms a heterodimer with TAF9 and TAF11 forms a heterodimer with TAF13 (Gangloff et al., 2001a, Gangloff et al., 2001b, Leurent et al., 2002). As these TAFs normally heterodimerize, they interact with U2 snRNA gene promoter in similar way.

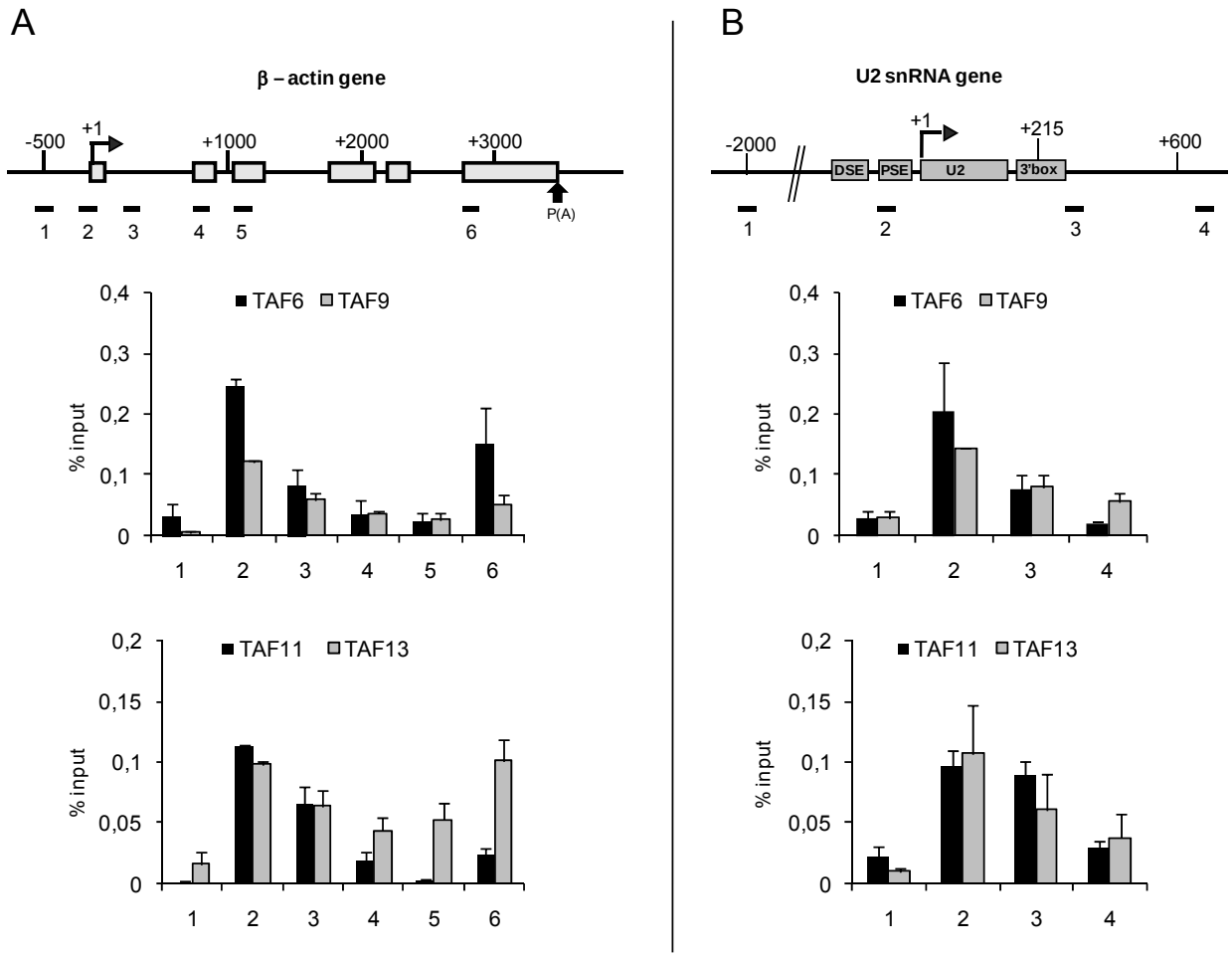


Figure 2-2. TAF6, TAF9, TAF11 and TAF13 are associated with the β-actin (A) and U2 snRNA (B) genes.

The results of qPCR analysis of the ChIP output using anti-TAF6, anti-TAF9, anti-TAF-11 or anti-TAF13 antibodies are shown.

2.2.3 Differential association of TAF5 and TAF8 peak with the β-actin and U2 snRNA genes

ChIP analysis, using antibodies to TAF5 and TAF8 indicate that these factors peak at the promoter of the β-actin gene, as expected for protein-coding genes (Figure 2-3A). Consistent with previous data for other TAF subunits, they also show an increase in association with region 6 of the β-actin gene. TAF5 and TAF8 show a different distribution on the U2 snRNA gene and instead peak at the 3' box, suggesting that these TAFs interact differently with the β-actin and U2 snRNA genes (Figure 2-3B), indicating possible differences in the conformation of the TBP-TAF complex on different gene-types.

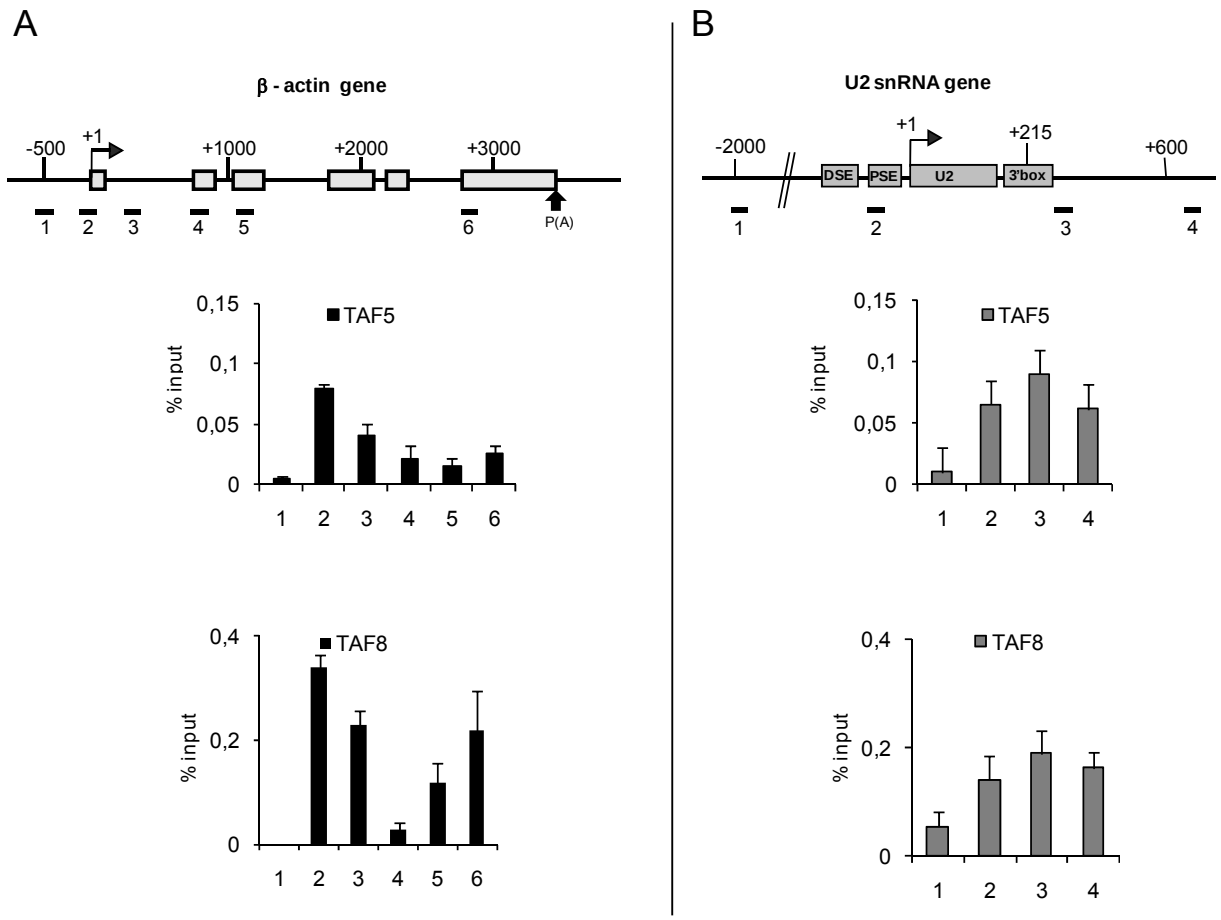


Figure 2-3. TAF5 and TAF8 show different profiles on the β-actin (A) and U2 snRNA (B) genes.

The results of qPCR analysis of the ChIP output using anti-TAF5 or anti-TAF8 antibodies are shown. Error bars indicate the standard deviation of means from at least three independent experiments.

2.2.4 TAF1, TAF2, TAF3, TAF4, TAF7, TAF10 and TAF12 are absent from the U2 snRNA gene.

Next, I have analyzed the association of TAF1, TAF2, TAF3, TAF4, TAF7, TAF10 and TAF12 with the human β-actin and the U2 snRNA genes by ChIP, coupled with qPCR analysis. Although TAF1, TAF2, TAF3, TAF4, TAF7, TAF10 and TAF12 are readily detected peaking at the β-actin gene promoter or in the case of TAF3, just downstream, none of these TAFs are detected on either the promoter or gene body of the U2 snRNA gene (Figure 2-4A and 2-4B). This may indicate that these TAFs are absent from the PIC on the U2 snRNA gene or that their conformation precludes antigen recognition by the specific antibodies. Interestingly, TAF1, TAF2, TAF3, TAF4, TAF10 and TAF12, in common with TBP, show an increase in association with region 6 of the β-actin gene. Interestigly, TAF7 does not. TAF1 and TAF7 are also detected at the promoter of the β-actin gene by ChIP-Seq (data accessed through ENCODE) (Johnson et al., 2007), in accordance with my results. There are no obvious peaks at region 6 in this data set, which may be due to differences in experimental conditions or due to limitations of ChIP-Seq in this region.

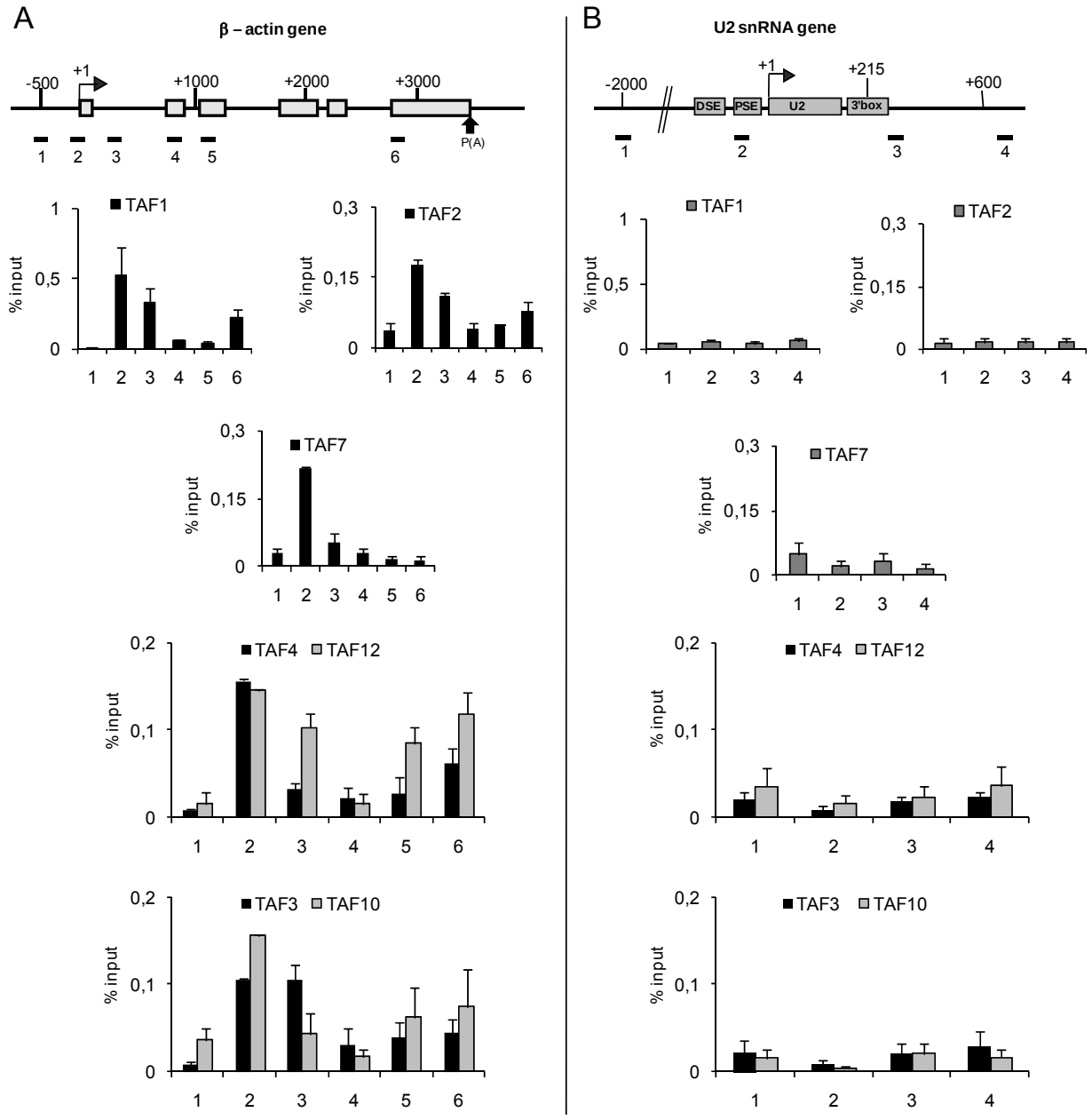


Figure 2-4. TAF1, TAF2, TAF3, TAF4, TAF7, TAF10 and TAF12 are associated with the β-actin gene (A) but not the U2 snRNA gene (B).

The results of qPCR analysis of the ChIP output using anti-TAF1, anti-TAF2, anti-TAF3, anti-TAF4, anti-TAF7, anti-TAF10 or anti-TAF12 antibodies are shown. The results for TAF4/TAF12 and TAF3/TAF10 are merged as they are known to form heterodimers.

2.2.5 Knockdown of TAF6 and TAF13 affects the level of pol II on the β-actin and U2 snRNA genes

In order to determine whether TAF6 and TAF13 association with the β-actin and U2 snRNA genes reflects a functional role in expression of both genes, I have knocked down the level of TAF6 and TAF13 using siRNAs (Figure 2-5A) and assayed pol II occupancy by ChIP (Figure 2-5B and 2-5C). Knockdown of TAF6 has the surprising effect of increasing the level of pol II on both genes. This result indicates that TAF6 plays an important role in transcription of both protein-coding and snRNA genes, but appears to inhibit, rather than favor, recruitment of pol II. In contrast, knockdown of TAF13 causes a decrease in the association of pol II with the β-actin and U2 snRNA genes, indicating that this TAF is required for transcription of both genes.

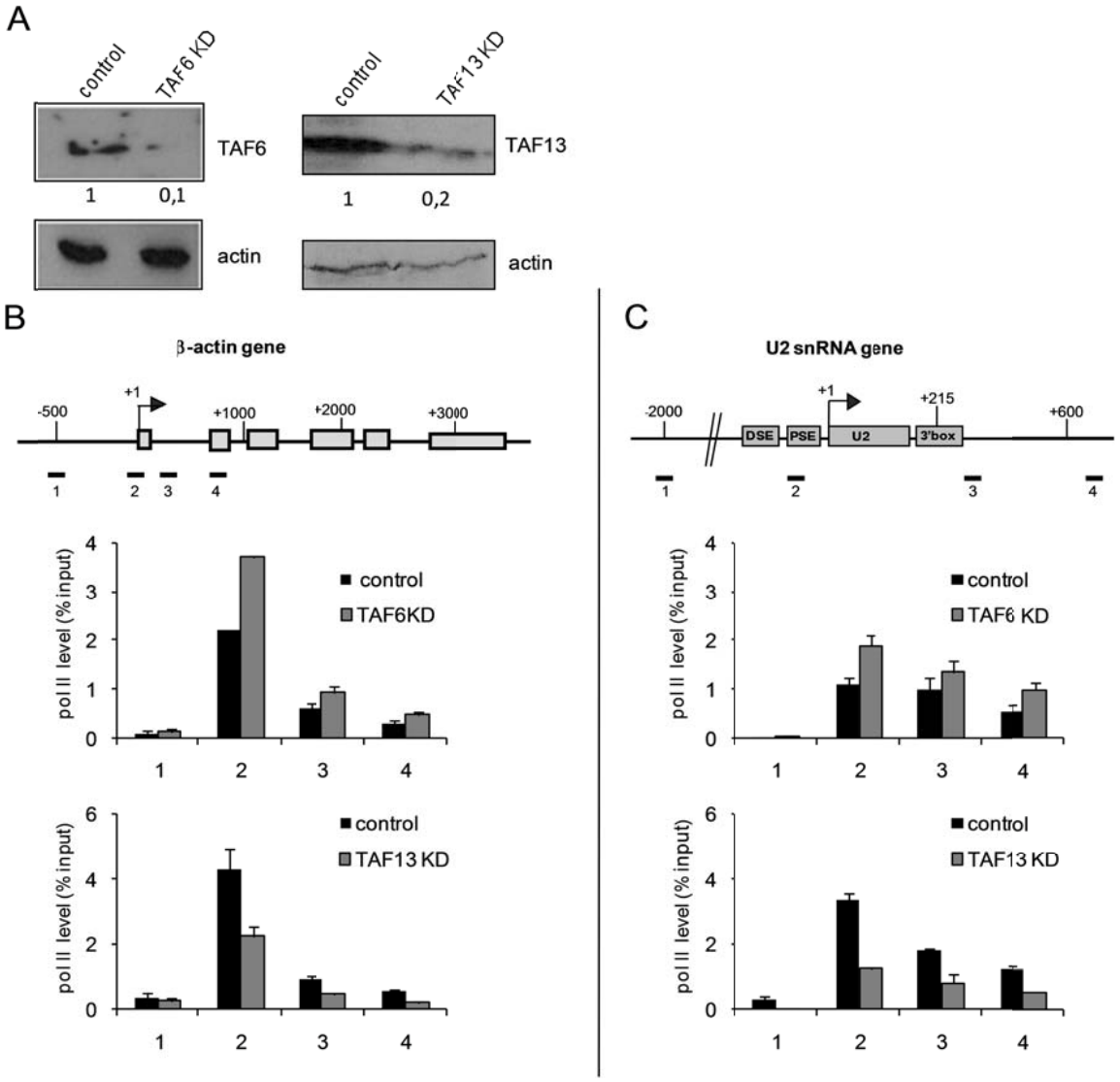


Figure 2-5. Knockdown of TAF6 and TAF13 affects the pol II level on the β-actin and U2 gene.

(A) Western blot analysis of HeLa whole-cell extracts from control cells or cells transfected with an siRNA specific for TAF6 and TAF13. Antibodies used are noted on the right. Actin was used as a loading control. Quantification of signal strength is noted below each line. The results of qPCR analysis of the output of ChIP using anti-pol II antibodies before (control) or after TAF6 knockdown (TAF6 KD) and TAF13 knockdown (TAF13 KD) for the β-actin (B) and U2 snRNA gene (C) are shown. Error bars indicate the standard deviation of means from at least three independent experiments.

2.2.6 Knockdown of TAF4 or TAF7 specifically affects the level of pol II on the β-actin gene

In order to test the possibility that some TAFs are not recruited to the U2 snRNA gene because they are not required for transcription (Figure 2-4), I have performed siRNA-mediated knockdown of TAF4 and TAF7 (Figure 2-6). Western blot analysis confirmed that siRNA knockdown of TAF4 and TAF7 was efficient (Figure 2-6A and 2-6B). In TAF4 and TAF7 KD cells the levels of pol II on the β-actin gene decrease when compared to the levels observed in control cells as expected (Figure 2-6C). In contrast, there is no drop of pol II on the U2 snRNA genes when either TAF4 or TAF7 are knocked down (Figure 2-6D). These results indicate that neither TAF4 nor TAF7 is involved in transcription of the U2 snRNA gene.

Taken together, these results support the notion that, in contrast to the β-actin gene, there is only a subset of the full TFIID TAF complement on U2 snRNA genes. The novel snRNA-TAF complex, which I have named snTAFc, comprises only TAF5, TAF6, TAF8, TAF9, TAF11 and TAF13.

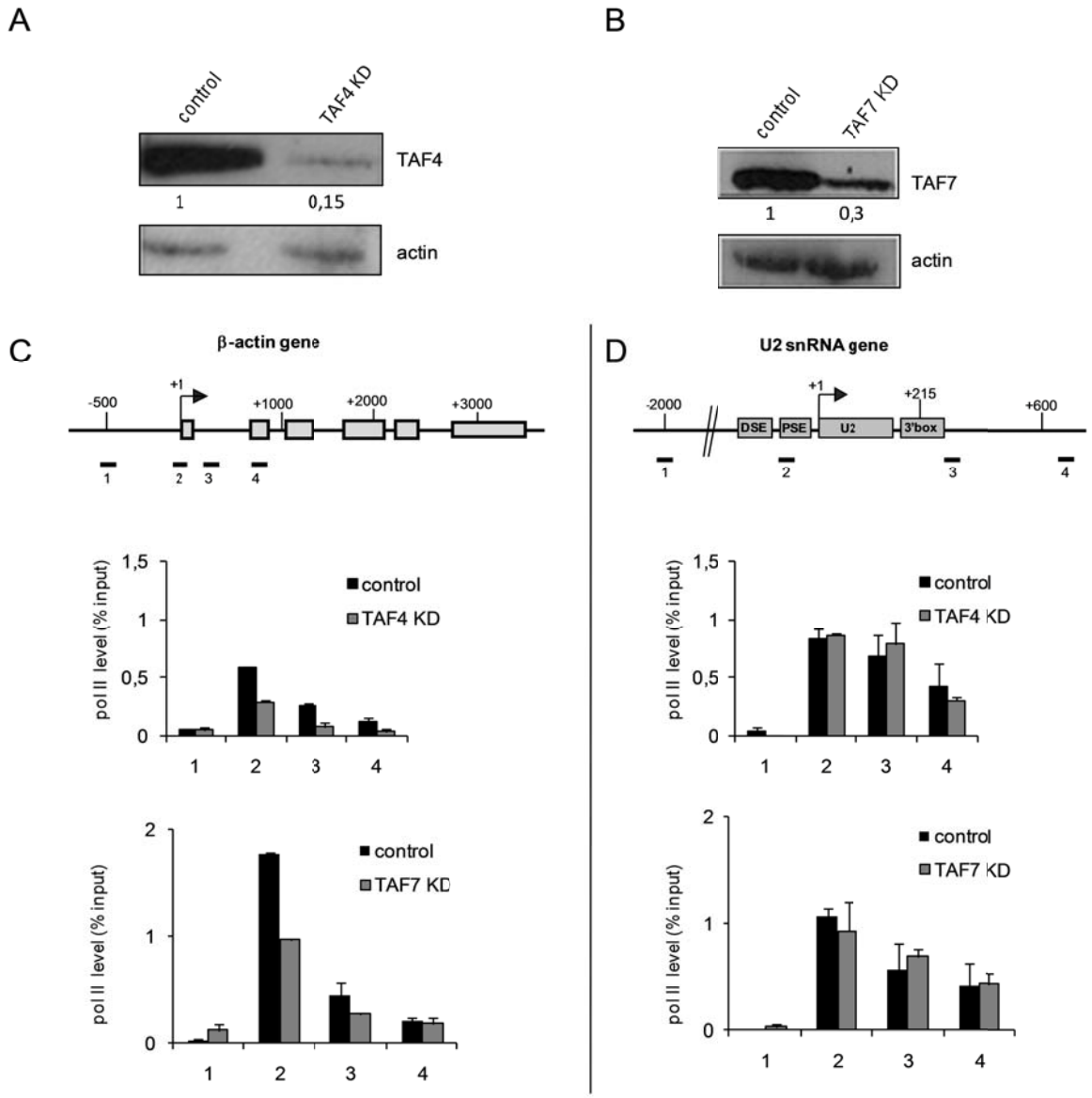


Figure 2-6. Knockdown of TAF4 or TAF7 specifically affects the level of pol II on the β-actin gene.

(A) Western blot analysis of HeLa whole-cell extracts from control cells or cells transfected with a siRNA specific for TAF4 as indicated. Antibodies used are noted at the right. Quantification of signal strength is noted below each line. (B) Western blot analysis of HeLa whole-cell extracts from control cells or cells transfected with a siRNA specific for TAF7. Antibodies used are noted at the right. The results of qPCR analysis of the ChIP output using anti-pol II antibodies before (control) or after TAF4 (TAF4 KD) or TAF7 (TAF7 KD) knockdown for the β-actin (C) and U2 snRNA gene (D) are shown. Error bars indicate the standard deviation of means from at least three independent experiments.

2.2.7 CPSF is specifically associated with the β-actin gene

TAFs are implicated in the recruitment of the cleavage-polyadenylation specificity factor, CPSF to protein-coding genes through TFIID. TAF5, TAF7 and TAF12 interact strongly with CPSF160 (Dantonel et al., 1997). The transcripts from snRNA genes are not polyadenylated and 3' box-dependent 3' end formation requires the Integrator complex rather than the factors involved in poly(A) site-dependent cleavage and polyadenylation (Hernandez, 1985, Baillat et al., 2005, Egloff et al., 2012a). Although TAF5 is detected on the U2 snRNA genes, TAF7 and TAF12 are not (Figures 2-3B and 2-4B). Accordingly, I have analyzed the association of CPSF160 with both the β-actin and U2 snRNA genes (Figure 2-7).

CPSF160 is readily detected by ChIP associated with the β-actin gene, close to the position of the poly(A) signal at +3560 as would be expected since cleavage occurs within this region (Figure 2-7A) (Egloff and Murphy, 2008a). CPSF160 is, however, not detected at the promoter region of the β-actin gene, indicating either that interaction between CPSF and TFIID on this gene is not stable or that crosslinking of CPSF160 in this region is not efficient due to lack of contact with or proximity to promoter DNA. In contrast, CPSF73 has been shown to be associated both with the poly(A) site and promoter regions of the c-myc and GAPDH protein-coding genes (Glover-Cutter et al., 2008). CPSF160 is not detected at any region of the U2 snRNA gene (Figure 2-7B). These results are consistent with a gene-type-specific role of TAF7 and TAF12 in helping to recruit CPSF to protein-coding genes and indicate that TAF5 alone is not sufficient. In line with these results, CPSF73 is not detected on the U2 snRNA gene (Glover-Cutter et al., 2008).

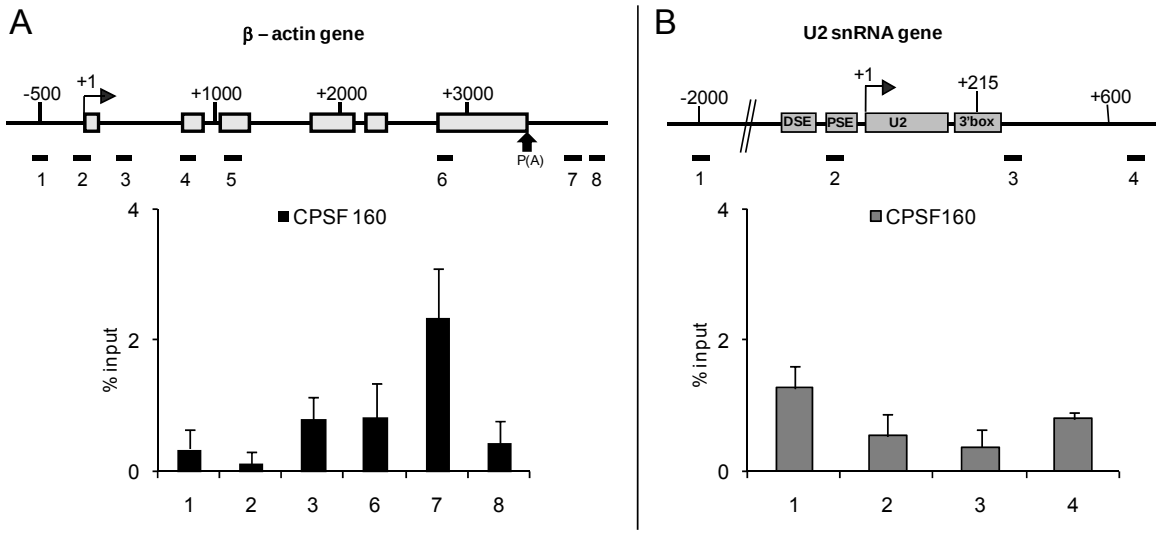


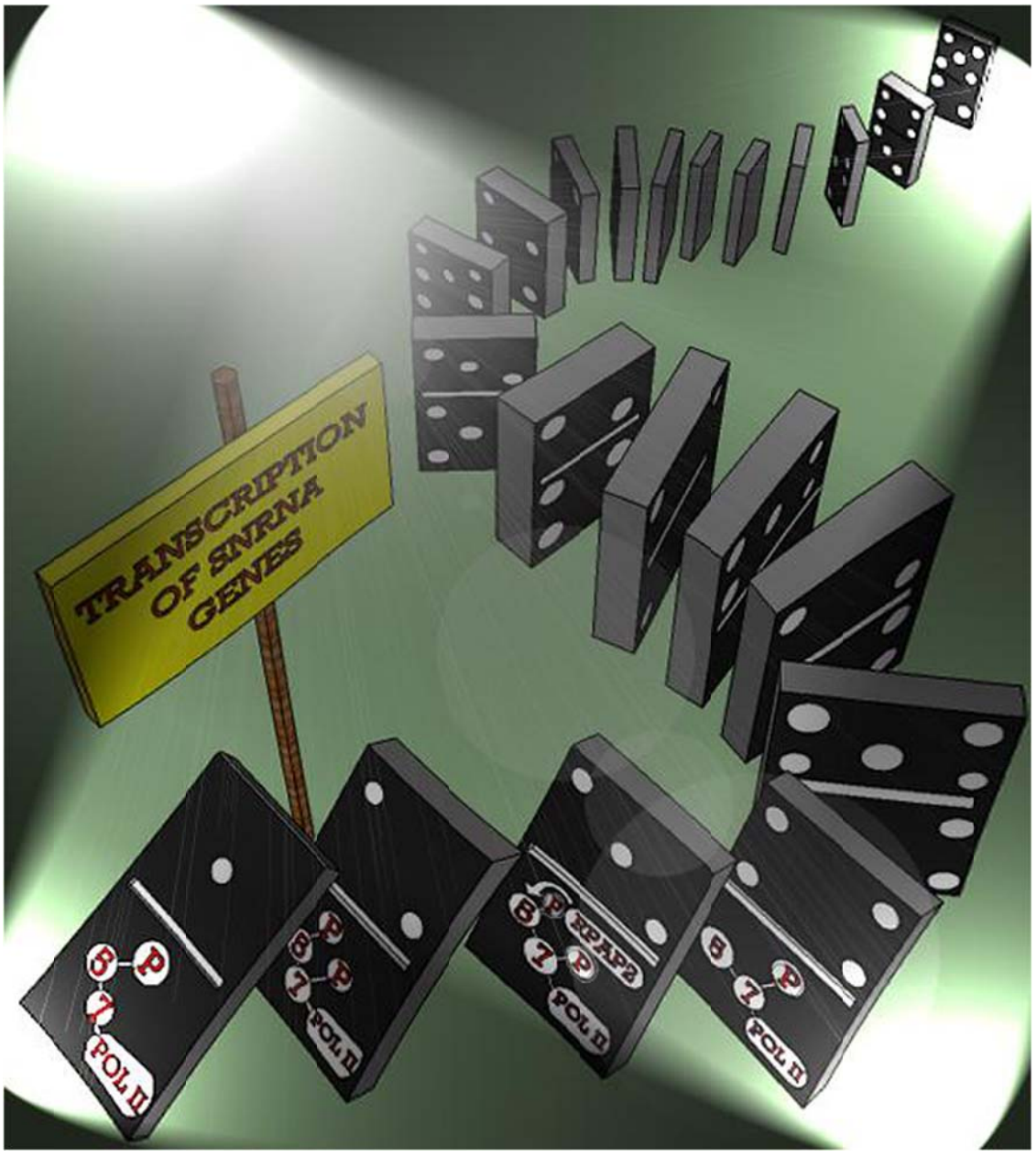
Figure 2-7. CPSF160 is specifically associated with the β-actin gene.

The results of the qPCR analysis of the ChIP output using anti-CPSF160 antibody on the β-actin (A) and U2 snRNA gene (B) are shown. Error bars indicate the standard deviation of means from at least three independent experiments.

2.3 Conclusions

My data suggest that a subcomplex of TFIID, containing TBP, TAF5, TAF6, TAF8, TAF9, TAF11, and TAF13 is loaded onto snRNA genes. Alternatively, TFIID may exist in different conformations on both the U2 snRNA and β-actin genes. However, the existence of a novel subcomplex on the snRNA genes is supported by the finding that knocking down TAF4 and TAF7 only affects β-actin and not affect pol II occupancy on U2 snRNA gene. TAFs exist as components of a number of distinct complexes, but, interestingly, the proposed snTAFc does not contain the classical core TAFs (TAF4 and TAF10) and lacks the highest-molecular weight TAFs, TAF1 and TAF2. A complete TFIID complex does not appear to be required for transcription initiation from snRNA gene promoters since knockdown of key TFIID subunits does not arrest transcription of the U2 snRNA gene (Figure 2-6). My ChIP analysis of a complete set of TAFs on the U2 snRNA and β-actin genes therefore provides further evidence for different TFIID compositions at different promoters. These results shed new light on the composition of TAF complexes on two different types of genes transcribed by pol II. TBP-TAF complexes might therefore make important contributions to the differences in control of transcription elongation and RNA processing between snRNA and protein-coding genes. Transcripts from snRNA genes, in contrast to mRNA genes, are not polyadenylated. Dantonel and colleagues showed that cleavage-polyadenylation specificity factor CPSF, associates with the PIC through TFIID (Dantonel et al., 1997). Neither TAF7 nor TAF12, which interact strongly with CPSF-160, is recruited to the U2 snRNA gene promoter (Figure 2-7). The snRNA gene-specific TAF complex may therefore play a central role in facilitating gene-type-specific RNA processing.

Chapter 3| CTD phosphatases on snRNA genes



3.1 Introduction

On the snRNA gene, TBP associates with the snRNA gene specific factor PTF and with snTAFc, as I have demonstrated in the previous chapter (Zaborowska et al., 2012, Egloff et al., 2008). This may play a role in setting the elongation properties of pol II on snRNA genes, possibly due to lack of recruitment of key elongation factors and may facilitate recognition of the gene specific 3' end processing element called 3'box. The snRNA 3'box RNA processing element is recognized by the snRNA gene-specific Integrator 3'end processing complex (Egloff et al., 2012b, Baillat et al., 2005). Integrator complex is recruited to snRNA genes through the pol II CTD. Phosphorylated Ser7 (Ser7P) is a key determinant for the recruitment of the Integrator complex to snRNA genes (Baillat et al., 2005, Egloff et al., 2007). A study by Jeronimo and colleagues revealed previously unknown components of pol II transcription complexes including an association between RNA pol II – associated protein 2 (RPAP2) and the Integrator complex (Jeronimo et al., 2007). This finding raised the possibility that RPAP2 protein may participate in expression of snRNA genes.

In human, the RPAP2 protein contains 612 amino acids with an evolutionary conserved domain of unknown function called DUF408 at the N terminus (Figure 3-1). This domain, which contains four conserved cysteine residues, is essential for the function of the *Saccharomyces cerevisiae* homologue of RPAP2, Rtr1 (Mosley et al., 2009). The UniProt entry for RPAP2 predicts several features besides the DUF408 domain: a putative coiled coil in the N-terminal region, and a transmembrane domain in the C-terminus (Figure 3-1). Interestingly, the yeast Rtr1 protein has been shown to act as CTD Ser5 phosphatase (Mosley et al., 2009, Gibney et al., 2008). Before starting work for this chapter, work in the Murphy laboratory had demonstrated that RPAP2 specifically recognizes the Ser7P mark on the pol II CTD and interacts with subunits of the snRNA gene-specific Integrator complex. Prompted by these findings, I have investigated the role of the RPAP2 protein in snRNA gene expression. Accordingly, I have analyzed RPAP2 association with snRNA genes and analysed how RPAP2 depletion affects snRNA gene expression. In addition, I have investigated the role of RPAP2 in recruitment of the Integrator complex to snRNA genes. Furthermore, I performed studies to determine whether RPAP2 acts as a Ser5 phosphatase, as does its homolog Rtr1 in yeast. As a complement to this study, I have

also analysed the role of the well-characterized Ser5 CTD phosphatase Ssu72 in expression of snRNA genes. All the experiments described in this chapter were performed in either HeLa cells or 293 human embryonic kidney cells as they are both cell lines of human origin, that grow well in cell culture and are widely used for gene expression studies.

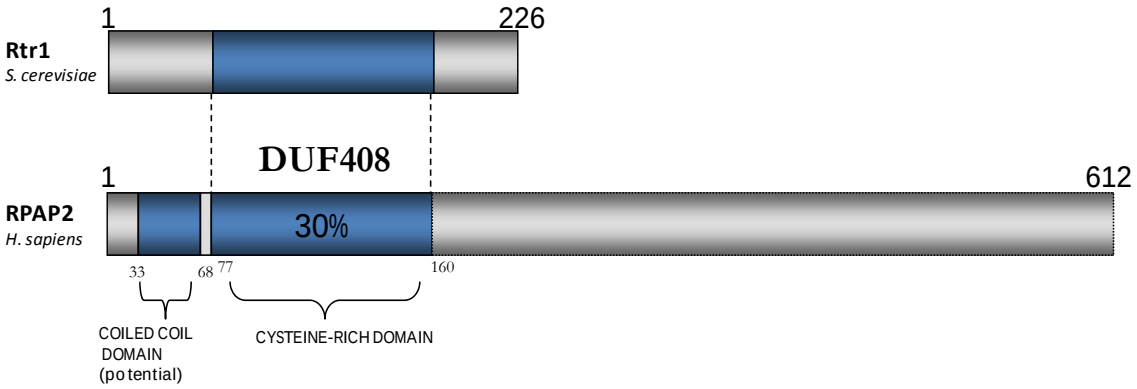


Figure 3-1. RPAP2 – RNA Pol II associated protein.
Human (*H. sapiens*) RPAP2 and yeast (*S. cerevisiae*) Rtr1 are shown on the schematic. Percent identity within the DUF408 domain is indicated. The conserved cysteine-rich domain and potential coiled-coil are highlighted.

3.2 Results

3.2.1 RPAP2 is recruited to snRNA genes during the transcription cycle

The work in this chapter has been carried out on the U2 snRNA genes which have been used extensively as models to study transcription of pol II-transcribed snRNA genes (Egloff et al., 2008). At the time of this study, there was no commercial antibody for RPAP2 available to use for ChIP. Accordingly, in order to investigate the function of RPAP2 in snRNA gene expression, the RPAP2 protein was tagged with three tandem Myc epitopes at the C-terminus (construct was generated by Dr Shona Murphy). Ectopic expression of Myc-RPAP2 was conducted by transfecting the expression vector Myc-RPAP2 into HeLa cells. RPAP2 protein was detected by Western blot analysis with the commercial anti-RPAP2 antibody for use in Western blot. As expected, Western blot analysis detected expression of two protein products; the ectopically expressed Myc-RPAP2 and endogenous RPAP2 (Figure 3-2). HeLa cells transfected with empty vector were used as a negative control and actin was used as a loading control. Result show that Myc-RPAP2 is expressed at approximately the same level as endogenous protein (Figure 3-2).

To determine whether RPAP2 is recruited to snRNA genes, I have analyzed its association with U2 snRNA genes by ChIP using an antibody to the Myc epitope. Figure 3-3 shows a model of the U2 snRNA gene with the relative locations of the PCR primers used for analysis. ChIP analysis indicate that RPAP2 is associated with the U2 snRNA gene, specifically over the 3' box (Figure 3-3, primer pair 3).

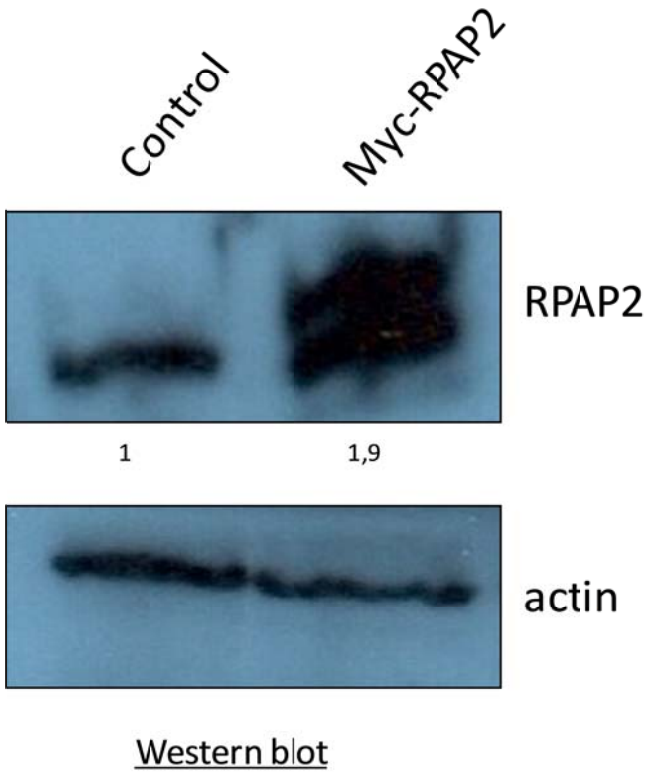


Figure 3-2. Myc-RPAP2 is expressed in transfected HeLa cells.
Extracts from HeLA cells transfected with empty vector (Control) and Myc-RPAP2 transfected (Myc-RPAP2) were analyzed by Westernblot using antibodies against RPAP2 and actin. Actin was used as a loading control. Quantification of signal strength is noted below each line.

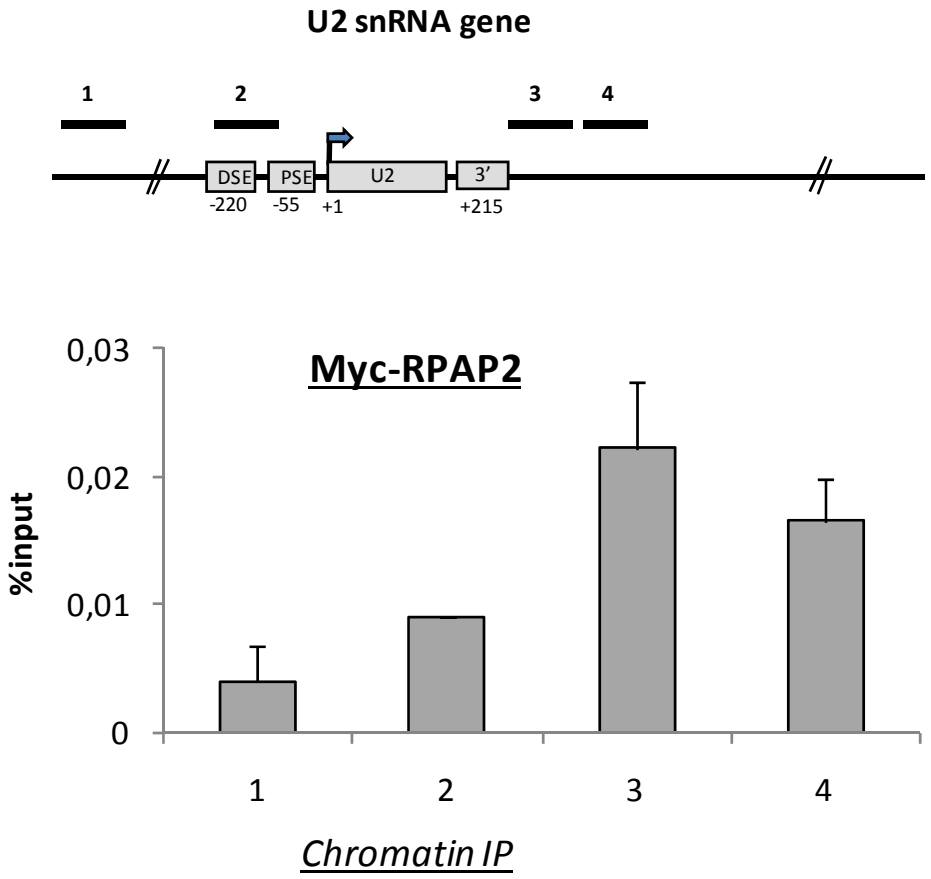


Figure 3-3. Myc-RPAP2 association with the U2 snRNA gene.
A schematic of the U2 snRNA gene is shown at the top. Horizontal lines above the gene schematics mark the position of the quantitative real-time PCR (qPCR) primer pairs. The arrow represents the transcription start site. The distal sequence element (DSE), the proximal sequence element (PSE) in the promoter, the snRNA-encoding region (U2) and the 3'box processing element (3') are denoted with boxes. The results of qPCR analysis of the output of ChIP using anti-Myc antibody is shown at the bottom. Error bars indicate the standard deviation of means from three independent experiments in this and subsequent figures.

3.2.2 RPAP2 interacts with the phospho-CTD through phospho-Ser7

It is well established that the CTD of pol II is a binding platform for nuclear factors that can influence the different steps of pol II transcription and RNA processing (Buratowski, 2009). Pol II CTD Ser7P has been detected in protein coding genes and the pol II-transcribed snRNA genes (Chapman et al., 2007, Egloff et al., 2007). While the role of Ser5P and Ser2P have been described extensively, Ser7P has so far no clear function in expression of protein-coding genes. Mutation of Ser7 to alanine does not significantly affect mRNA production from tested genes (Egloff et al., 2007, Chapman et al., 2007). However, Ser7P is essential for efficient transcription and 3' end processing of snRNA genes (Egloff et al., 2007). Previous work in the Murphy lab indicated that RPAP2 preferentially associates with Ser7-phosphorylated pol II. In order to determine whether RPAP2 binds CTD Ser7P directly, I have tested the ability of biotinylated CTD peptides to interact with bacterially expressed GST-RPAP2. The peptides used comprise only two heptapeptide repeats. As shown in Figure 3-4, recombinant RPAP2 interacts with an *in vitro* phosphorylated CTD peptide carrying two repeats, but does not interact with nonphosphorylated peptides. This is in accordance with a demonstration in yeast that the basic unit of the CTD is constituted by two repeats (Stiller and Cook, 2004). To assess the role of Ser7 in this interaction, the same experiment was performed using the peptide containing alanine instead of serine at position 7. Interestingly, mutation of Ser7 abrogates the interaction of RPAP2 with CTD. This result suggest that RPAP2 interacts directly with the CTD, that two CTD repeats are sufficient for binding and that phosphorylation on Ser7 of the pol II CTD is required for this interaction.

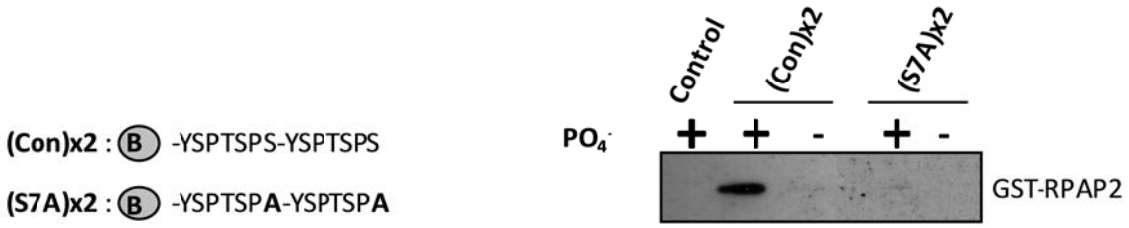


Figure 3-4. Ser7 of the pol II CTD is required for interaction with RPAP2.
The schematic of the biotinylated peptides used in the experiment is shown on the left. The binding of GST-RPAP2 to the biotinylated (B), phosphorylated (PO_4+) and non-phoshorylated (PO_4-) peptides was monitored by Western blot. Beads incubated with RPAP2 were used as a control.

3.2.3 RPAP2 is required for proper expression of snRNA genes

As stated before, Ser7P is critical for transcription and 3' end processing of snRNA genes (Egloff et al., 2007). My previous experiments (Figure 3-3 and 3-4) indicate that RPAP2 is recruited to snRNA genes and that CTD Ser7P is required for interaction with RPAP2. Accordingly, I asked the question of whether RPAP2 is required for snRNA gene expression. For this purpose, I have reduced the level of RPAP2 in HeLa cells using siRNA-mediated knockdown and determined the effect on pol II level by ChIP. The efficiency of the knockdown was verified by measuring the level of the protein by Westernblot using an anti-RPAP2 antibody, which confirmed that the level of RPAP2 was considerably reduced following knockdown (Figure 3-5). ChIP data clearly shows that pol II occupancy is reduced after RPAP2 depletion (Figure 3-6), suggesting that RPAP2 is required for efficient expression.

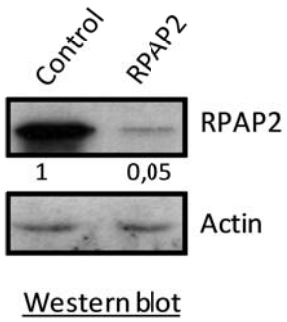


Figure 3-5. Efficiency of RPAP2 knockdown.
Western blot analysis of whole-cell extracts from control cells or cells transfected with an siRNA specific for RPAP2 (RPAP2 KD). Antibodies used are noted on the right. Quantification of signal strength is noted below each line.

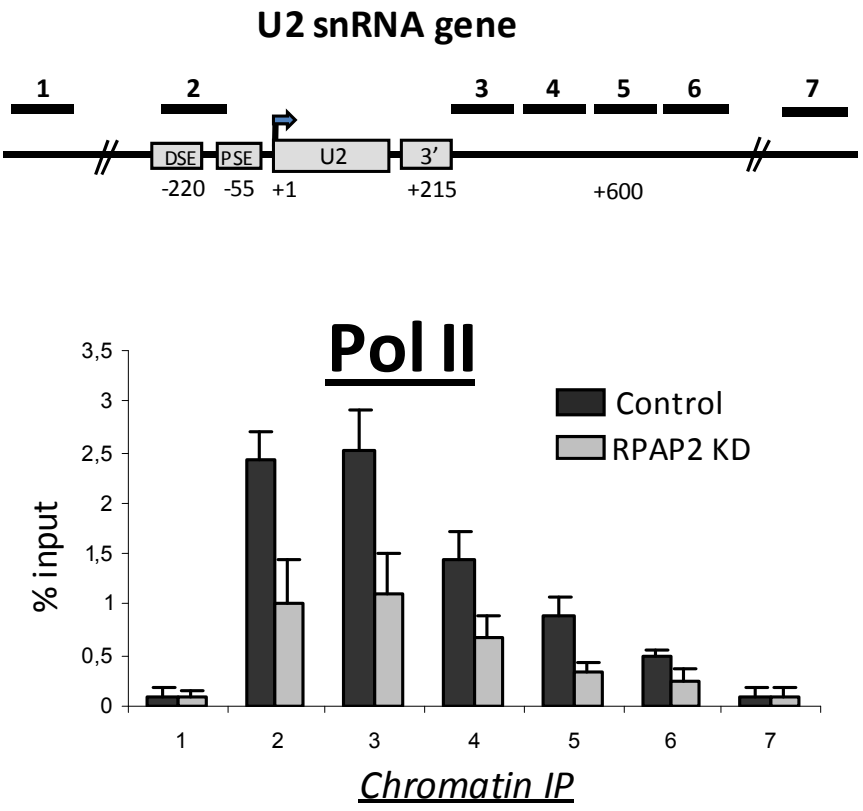


Figure 3-6. RPAP2 is required for proper snRNA gene expression.
The results of qPCR analysis of the output of ChIP using anti-pol II antibody is shown under the schematic. The regions amplified are noted on the schematic above. Error bars indicate the standard deviation of means from at least three independent experiments.

Analysis of the effect of RPAP2 knockdown on processing of the U2 snRNA gene transcripts by RNase protection (Figure 3-7) indicates that 3' box recognition is also affected following RPAP2 depletion. There are three RNase protection products detected – corresponding to the U2 pre-snRNA (Pre-U2), stable mature U2 snRNA (U2) and the readthrough product that has escaped 3' box-directed processing (U2RT) (Egloff et al., 2007, Medlin et al., 2005). The ratio of Pre-U2 to mature U2 is reduced to approximately 50% after RPAP2 knockdown confirming the transcription defect observed by ChIP. Furthermore, the ratio of RT to Pre-U2 increases approximately 3-fold indicating a defect in snRNA 3'end maturation after RPAP2 knockdown.

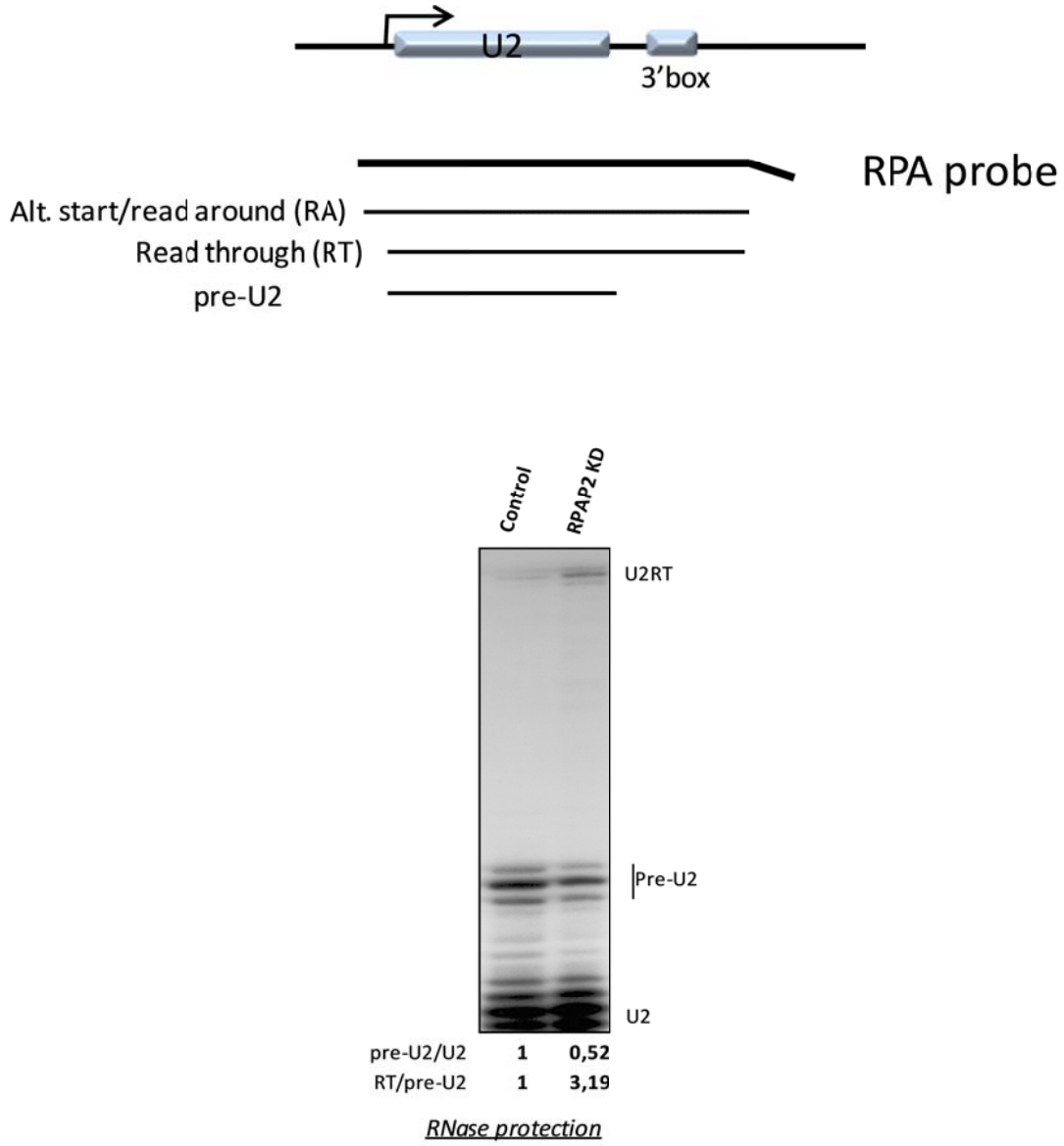


Figure 3-7. Defect in snRNA 3' end processing after RPAP2 knockdown.
RNase protection analysis of U2 transcripts in control cells and cells transfected with an siRNA specific for RPAP2 (RPAP2 KD). Location of the RNase Protection Analysis (RPA) probe and the expected products of the RPA are noted on the schematic. Position of the pre-U2 snRNA and unprocessed read-through (RT) U2 snRNA is noted on the right. Relative amount of correct 3' end and readthrough (RT) is also noted below each lane.

Previous work in the Murphy laboratory indicated that RPAP2 interacts with the Integrator complex. In addition, it had been shown that the Integrator complex associates with snRNA genes to activate transcription and direct 3' box processing (Egloff et al., 2010). It is tempting to speculate that RPAP2 may be involved in Integrator recruitment to U2 snRNA genes. To address this, I have knocked down RPAP2 (Figure 3-5) and looked at the Integrator subunit Int11 occupancy on U2 snRNA gene by ChIP analysis before and after RPAP2 depletion. Figure 3-8 indicates that Int11 is stably recruited to U2 snRNA gene later in transcription cycle with its highest level nearest to the 3' box (Figure 3-8; probe 3). Interestingly, the recruitment of the Int11 on U2 snRNA genes is affected after RPAP2 knockdown (Figure 3-8).

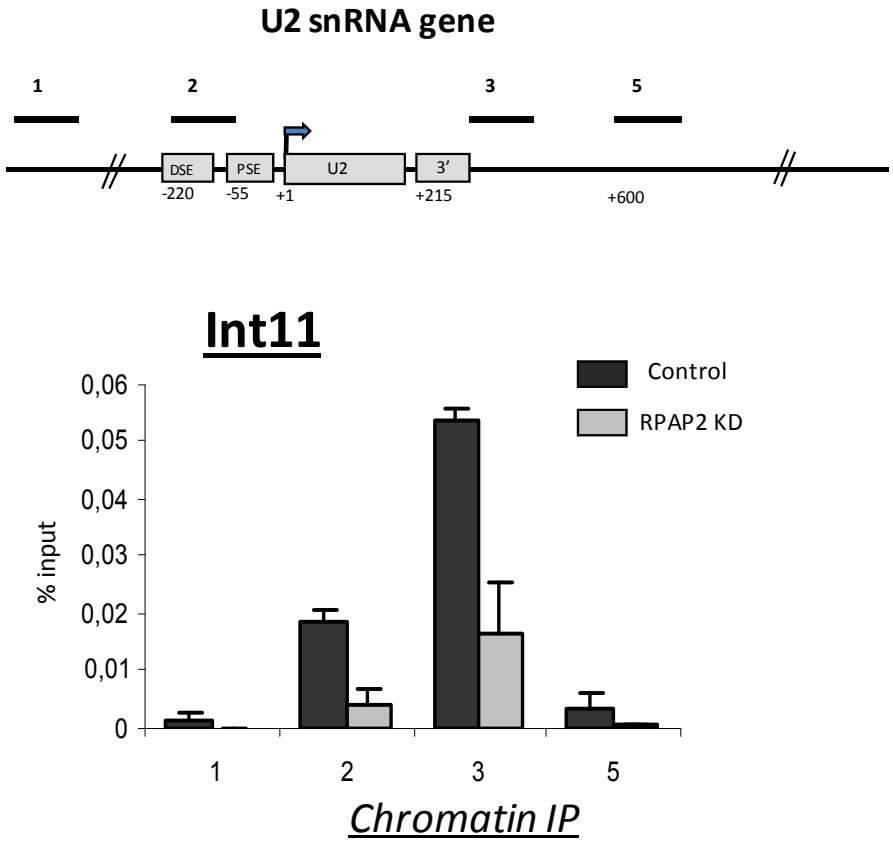


Figure 3-8. RPAP2 is required for the recruitment of the Int11 to U2 snRNA genes.
The results of qPCR analysis of the output of ChIP using antibody against the Int11 subunit of Integrator complex before and after RPAP2 knockdown. The regions amplified are noted on the schematic.

3.2.4 RPAP2 is a CTD phosphatase

I have demonstrated that RPAP2 is required for proper expression of pol II - transcribed snRNA genes (Figure 3-5 and 3-7) and that it binds to the CTD phosphorylated on Ser7P (Figure 3-4). Interestingly, the yeast homolog of RPAP2, Rtr1, is a CTD phosphatase (Mosley et al., 2009). Mosley and colleagues demonstrated that Ser5P accumulates in whole-cell extracts and throughout mRNA gene coding regions after deletion of Rtr1 (Mosley et al., 2009). My next step was therefore to determine whether human RPAP2 also functions as a CTD phosphatase. Accordingly, I have analyzed the Ser2P, Ser5P and Ser7P levels in whole-cell extracts using Western blot after RPAP2 knockdown (Figure 3-9). The efficiency of the knockdown was verified by measuring the level of the protein by Western blot analysis. siRNA-mediated RPAP2 depletion results in an increase in the level of Ser5P but does not affect the levels of Ser2P and Ser7P, suggesting that RPAP2 is also a Ser5 phosphatase in human (Figure 3-9). In addition, I looked at Ser5P levels on the U2 snRNA gene following RPAP2 knockdown (Figure 3-10). After knockdown of RPAP2, the level of Ser5P increases on U2 snRNA genes throughout the whole transcribed region (Figure 3-10A). As the pol II levels decrease after RPAP2 knockdown (Figure 3-6), the ratio of Ser5P to pol II is therefore highly increased upon siRNA mediated RPAP2 depletion (Figure 3-10B). This result suggests that RPAP2 dephosphorylates Ser5P on the actively transcribing pol II.

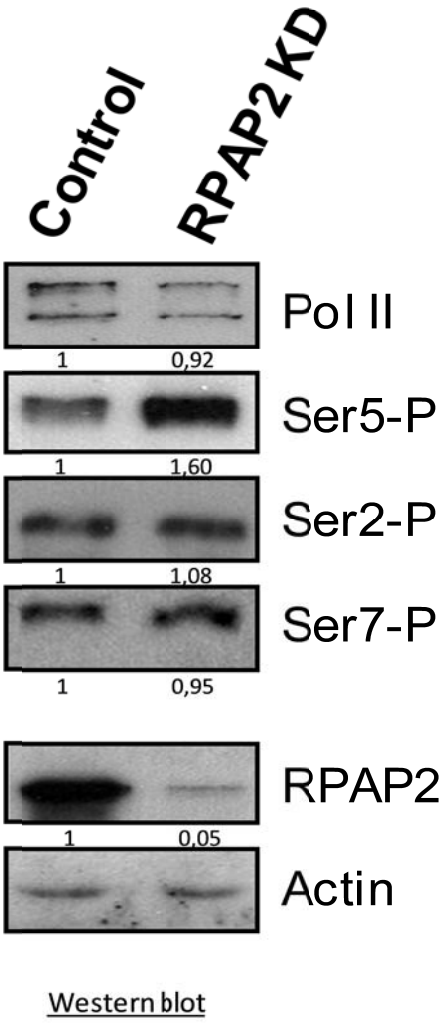
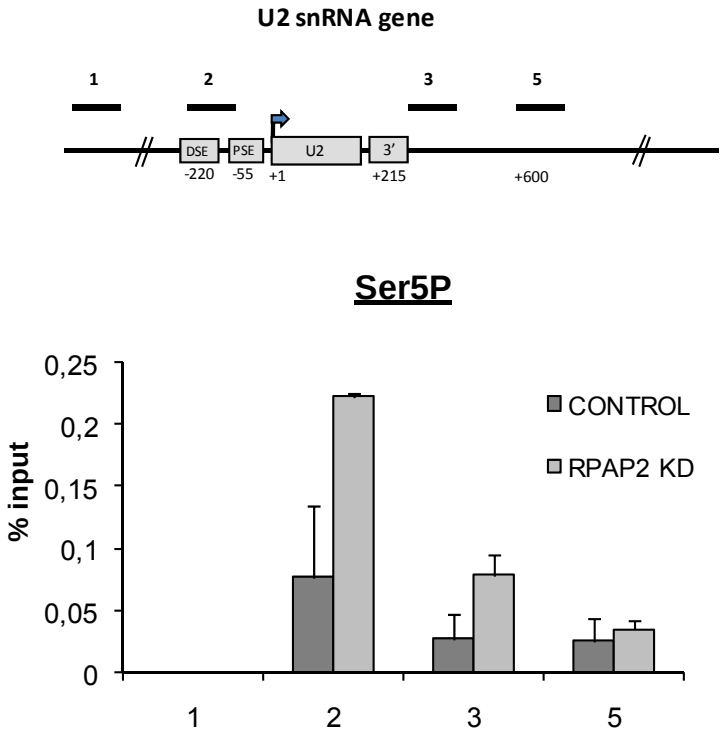


Figure 3-9. RPAP2 depletion results in an increase in the levels of Ser5P.
Western blot analysis of whole-cell extracts from control cells or cells transfected with an siRNA specific for RPAP2 (RPAP2 KD). Antibodies used are noted on the right. Quantification of signal strength is noted below each line.

A



B

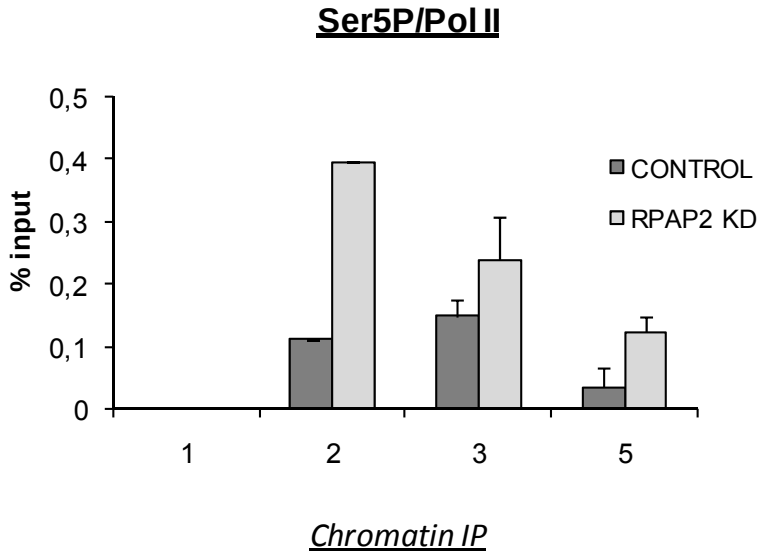


Figure 3-10. The Ser5P levels are affected on the pol II transcribing the U2 snRNA gene after RPAP2 KD.

(A) qPCR quantitation of ChIP analysis of U2 snRNA gene after RPAP2 knockdown using an antibody against Ser5P. (B) qPCR quantitation of ChIP analysis of U2 snRNA genes after RPAP2 knockdown using an antibody against Ser5P relative to the pol II signal. Primers used are noted on the schematic of the U2 snRNA gene above.

3.2.5 Recruitment of RPAP2 to protein-coding genes does not require Ser7 of the CTD

Mosley and colleagues found Rtr1 associated with protein-coding genes (Mosley et al., 2009). I also tested whether RPAP2 is recruited to protein coding genes, using the β-actin gene as a model protein-coding gene. Having confirmed expression of Myc-RPAP2 in HeLa cells (Figure 3-2), I have analyzed its association with the β-actin gene by ChIP using an antibody to the three tandem Myc epitopes at the C-terminus. Figure 3-11 shows a model of the β-actin gene, as previously noted. ChIP analysis indicates that RPAP2 is associated with the β-actin gene (Figure 3-11).

To asses RPAP2 function on protein-coding genes, I have knockdown RPAP2 and looked on the levels of pol II and Ser5P pol II by ChIP (Figure 3-12). Interestingly, RPAP2 knockdown causes a reduction in pol II levels at beginning of the gene and an accumulation of Ser5P throughout coding regions of β-actin gene (Figure 3-12).

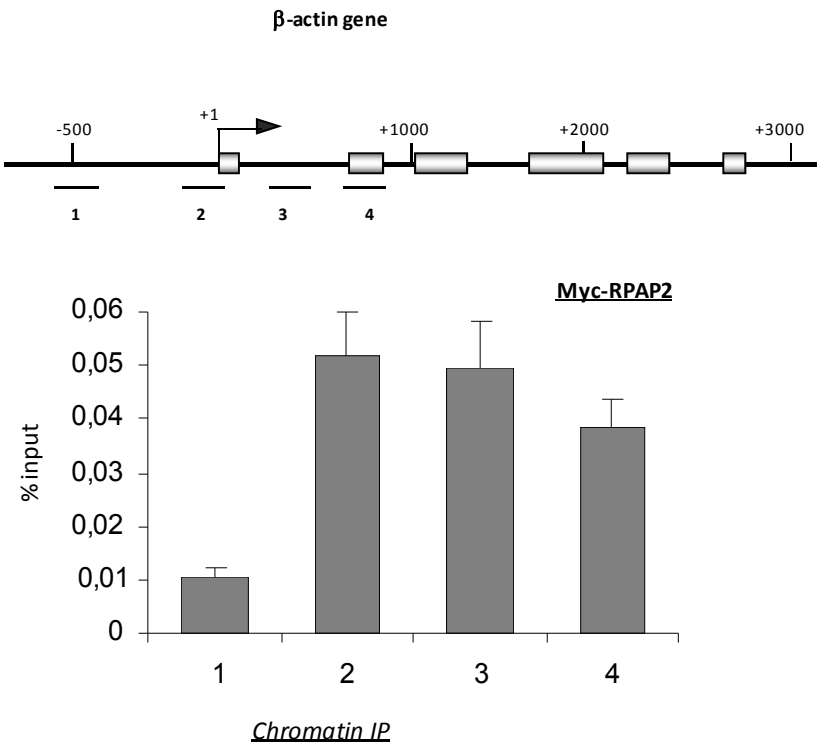


Figure 3-11. RPAP2 is associated with β-actin protein coding gene.
A schematic of the β-actin gene is shown at the top. The arrow represents the transcription start site and exons are denoted with boxes. Non-coding regions are denoted by a horizontal line. The results of qPCR analysis of the output of ChIP using anti-Myc antibody against Myc-RPAP2 is shown in the graph under the schematic. The regions amplified are noted on the schematic. Error bars indicate the standard deviation from three independent experiments in this and subsequent figures.

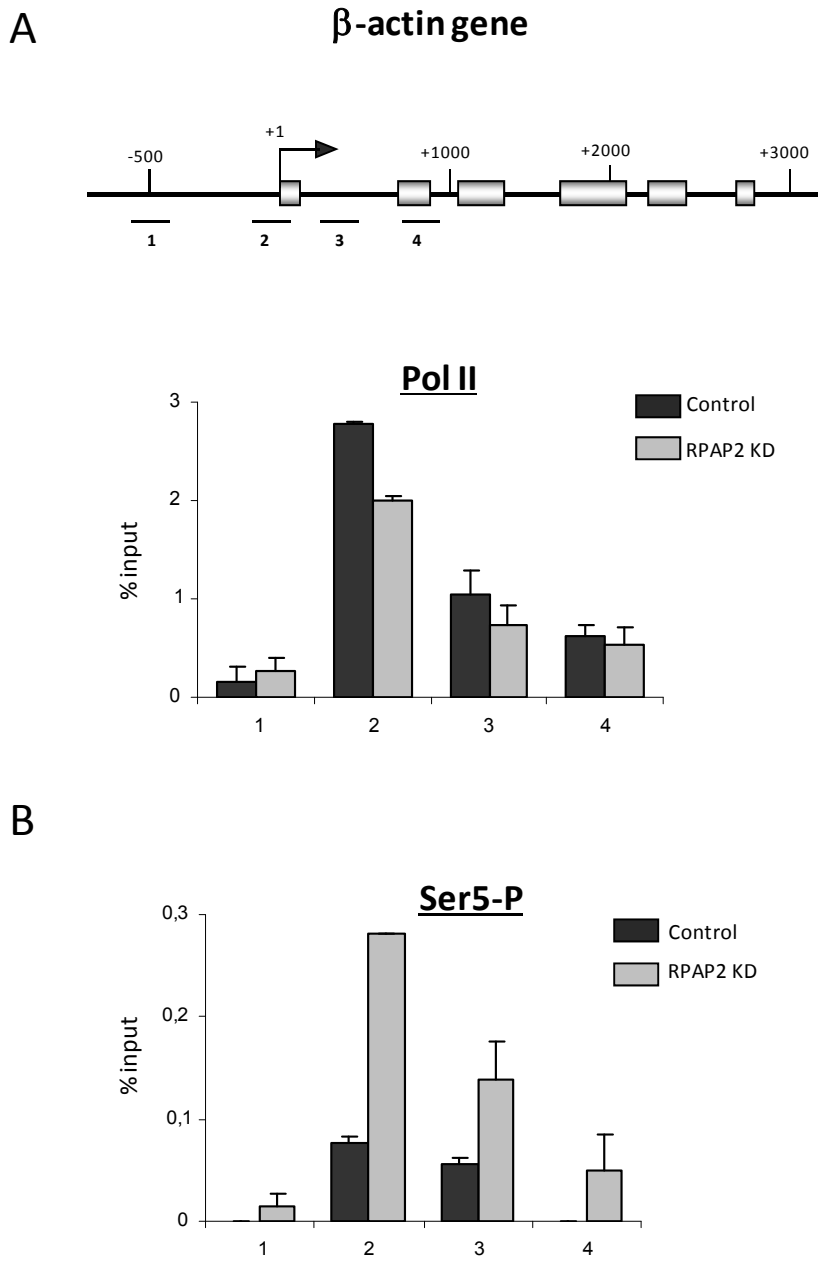


Figure 3-12. The effect of RPAP2 knockdown on pol II and Ser5P on pol II transcribing β-actin protein coding gene.

(A) The results of qPCR analysis of the output of ChIP using anti-pol II antibody is shown under the schematic. The regions amplified are noted on the schematic. (B) qPCR quantitation of ChIP analysis of β-actin gene after RPAP2 knockdown using an antibody against Ser5P. Primers used are noted on the schematic. Error bars indicate the standard deviation of means from at least three independent experiments in this figure.

These results raise the question if recruitment of RPAP2 to protein coding genes requires Ser7 of the pol II CTD and why mutation of Ser7 specifically affects expression of snRNA genes (Egloff et al., 2007). In order to address this question, I have used a complementation system where the endogenous pol II is turned over by α-amanitin treatment and replaced by an α-amanitin resistant pol II. I have carried out ChIP of Myc-RPAP2 on the U2 snRNA and β-actin genes after complementation of α-amanitin treated cells with either consensus 48 CTD repeats (Con)48 or 48 CTD repeats with Ser7 mutated to alanine (S7A)48 pol II. For CTD complementation, mutations are introduced into consensus (Con) CTD repeats in an α-amanitin-resistant pol II large subunit (Egloff et al., 2007, Chapman et al., 2005, Chapman et al., 2007). The consensus 48 CTD repeats properly supports viability (Chapman et al., 2007). The large subunit of endogenous pol II is turned over after α-amanitin addition. As indicated in Figure 3-13, RPAP2 is recruited to the β-actin gene when Ser7 is mutated to alanine. In contrast, recruitment of RPAP2 to U2 snRNA gene is affected by mutation of Ser7. This would suggest that recruitment of RPAP2 to protein-coding genes does not require Ser7 of the CTD.

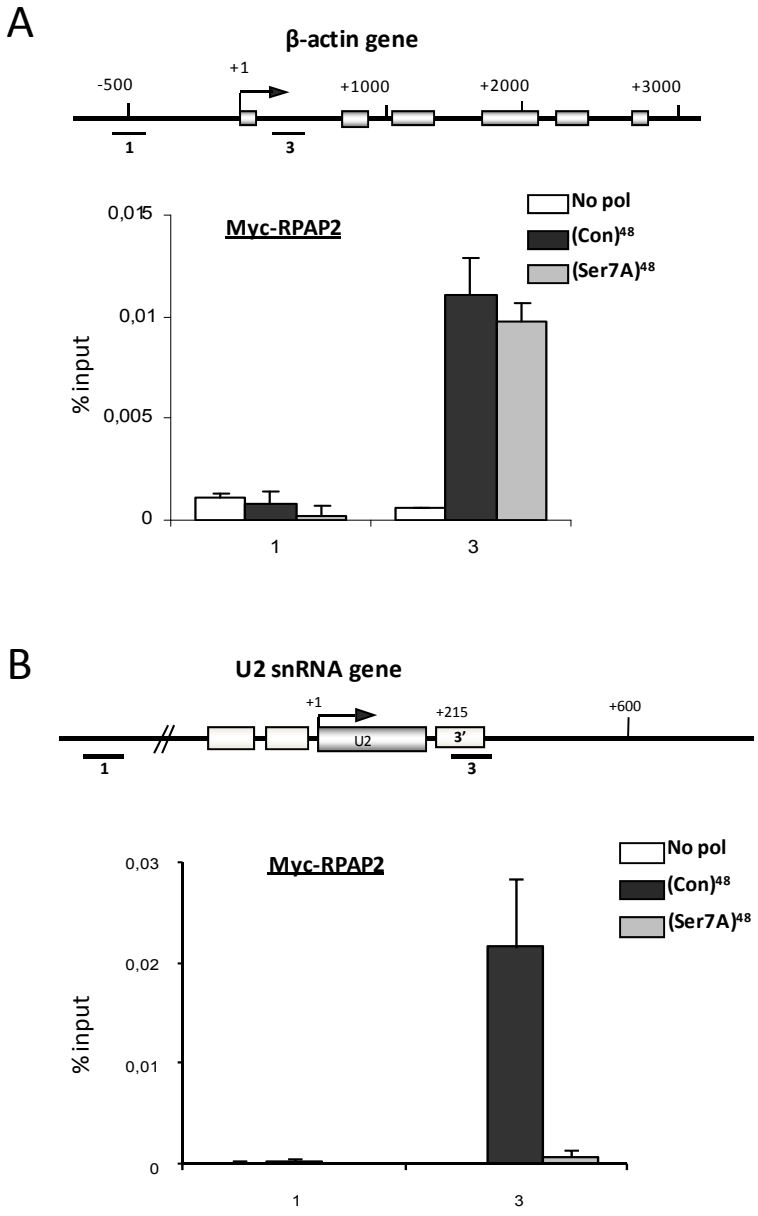


Figure 3-13. Recruitment of RPAP2 to protein-coding genes does not require Ser7 of the CTD.
(A) qPCR quantitation of ChIP analysis of the β-actin gene using antibodies to Myc-RPAP2 after complementation with $(Con)^{48}$ or $(Ser7A)^{48}$ as indicated. (Con) designates consensus CTD heptapeptides. (B) qPCR quantitation of ChIP analysis of the U2 snRNA using antibodies to Myc-RPAP2 after complementation with $(Con)^{48}$ or $(Ser7A)^{48}$ as indicated. Error bars indicate the standard deviation of means from at least three independent experiments.

3.2.6 Ssu72 is involved in human snRNA transcription

Ssu72 is well-characterized CTD Ser5P phosphatase and a component of the yeast cleavage and polyadenylation factor (CPF) complex (Dichtl et al., 2002, Krishnamurthy et al., 2004, Reyes-Reyes and Hampsey, 2007, Steinmetz and Brow, 2003). In addition, a recent paper identified Ssu72 as a Ser7 phosphatase (Bataille et al., 2012). ChIP assays demonstrate that Ssu72 is predominantly enriched at the 3'end of genes but also occupies the promoter region (Ansari and Hampsey, 2005, Nedea et al., 2003, Bataille et al., 2012). My finding suggest that RPAP2 is a Ser5P phosphatase on U2 snRNA genes (Figure 3-9 and Figure 3-10). However, this may not be the only Ser5P phosphatase acting on these genes. Other phosphatases may also be involved and Ssu72 is a candidate. I have therefore analyzed by ChIP the association of the Ssu72 with the U2 snRNA gene and the β-actin gene as a model protein-coding gene. Interestingly, Ssu72 is found at the promoter and at the end of the transcription unit of U2 snRNA gene (i.e. 800bp after the transcription start site). For the β-actin gene, Ssu72 is found at the promoter region of the β-actin gene (Figure 3-14). More probes at the 3' end of the β-actin gene need to be analyzed to confirm previously published data indicating that Ssu72 is associated as well with 3' end of the genes.

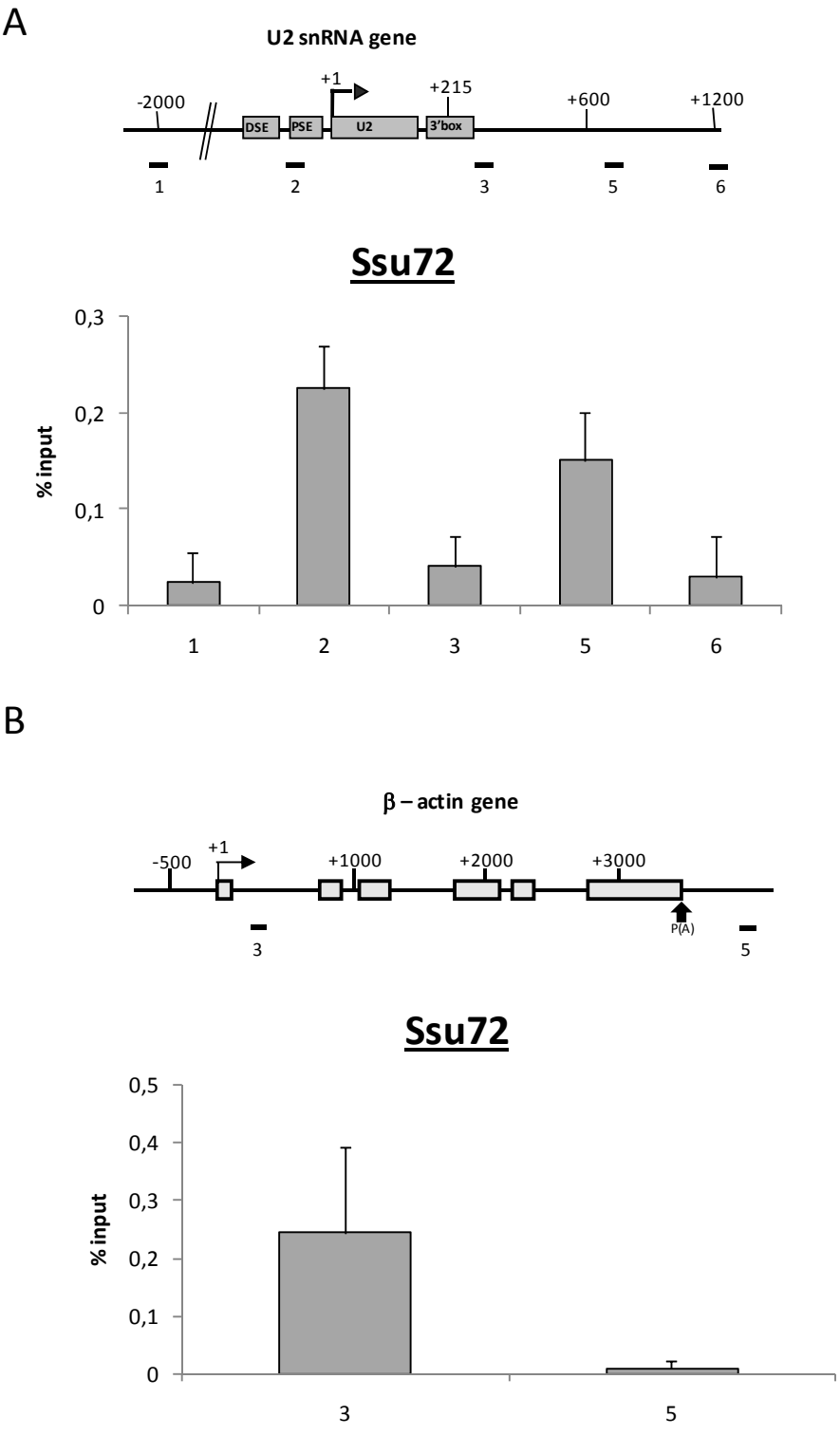


Figure 3-14. Ssu72 is associated with the U2 snRNA and β-actin genes.
(A) qPCR quantitation of ChIP analysis of U2 snRNA gene using an antibody against Ssu72. Primers used are noted on the schematic of the U2 snRNA gene above. (B)) qPCR quantitation of ChIP analysis of β-actin gene using an antibody against Ssu72. Primers used are noted on the schematic of the β-actin gene above. Error bars indicate the standard deviation of means from at least three independent experiments.

I have wondered whether Ssu72 affects pol II occupancy on the genes. For this purpose, the effects of Ssu72 depletion on the pol II occupancy across the U2 snRNA gene and β-actin protein-coding genes were determined by ChIP after siRNA-mediated knockdown. The efficiency of the knockdown was verified by Western blot (Figure 3-15).

ChIP analysis shows that the pol II occupancy decreases in the promoter region of the U2 snRNA gene after Ssu72 knockdown (Figure 3-16A). On the β-actin gene, instead there is no drop of pol II observed on the promoter following Ssu72 depletion (Figure 3-16B). Interestingly, more pol II is detected at the end of the transcription unit of the U2 snRNA gene (Figure 3-16B, primer pair 5) and the β-actin gene (Figure 3-16A, primer pair 5). This observation suggests that the activity of Ssu72 is required for proper transcription termination. In addition, the lower levels of pol II within the body of the U2 snRNA gene suggests that there is a defect in transcription initiation. This indicates that when Ssu72 is depleted, CTD remains in a hyperphosphorylated state and is unable to effectively terminate transcription and recycle back to the promoter to be reassembled into the PIC complex. This is in accordance with the findings that the phosphatase activity of Ssu72 is required for proper pol II termination at some loci *in vivo* (Mosley et al., 2009, Krishnamurthy et al., 2004, Ganem et al., 2003).

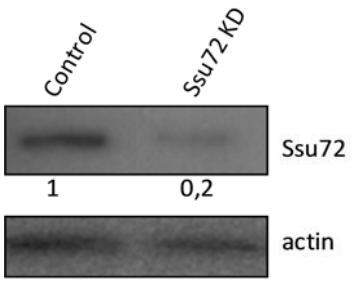


Figure 3-15. Efficiency of Ssu72 knockdown.
Western blot analysis of whole-cell extracts from control cells (Control) or cells transfected with an siRNA specific for Ssu72 (Ssu72 KD). The antibodies used are noted on the right.

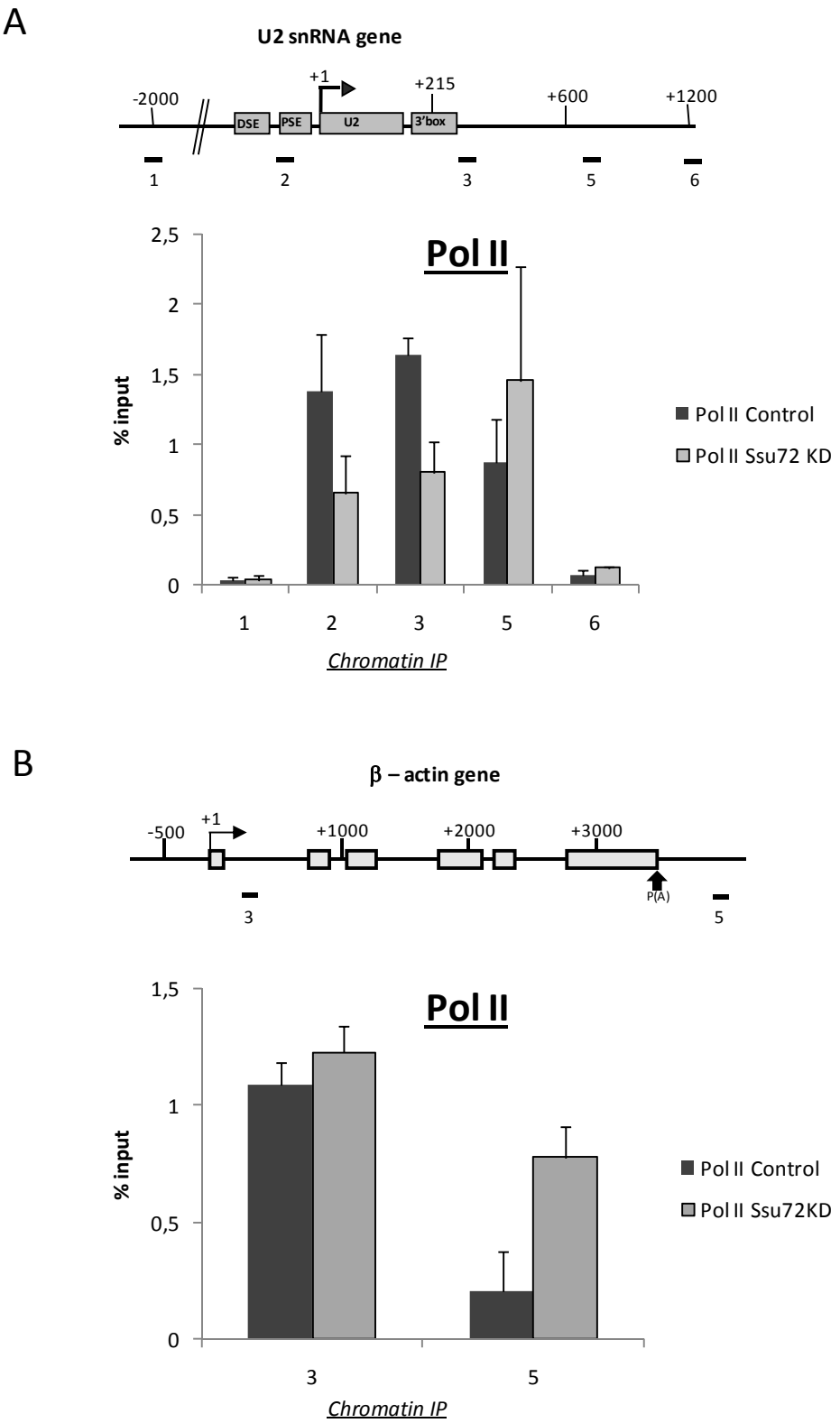


Figure 3-16. Pol II occupancy after Ssu72 knockdown on the U2 snRNA and β-actin genes.
(A) and (B) The results of qPCR analysis of the output of ChIP using anti-pol II antibody is shown under the schematic for the U2 snRNA and β-actin genes respectively. The regions amplified are noted on the schematic. Error bars indicate the standard deviation of means from at least three independent experiments.

3.2.7 Results suggest that Ser5P is removed in two waves along the U2 snRNA gene

In order to determine the effects of Ssu72 depletion on the levels of DNA-associated Ser5P pol II, I carried out ChIP analyses in the Ssu72-knockdown cells and compared the levels of Ser5P related to pol II to those found in control cells (Figure 3-17). The levels of Ser5P pol II increase in the Ssu72 knockdown cells when compared to the levels observed in control cells. The biggest differences are seen toward the 3' end of the U2 snRNA gene (Figure 3-17, primer pairs 5 and 6). This result suggests that Ssu72 is a Ser5 phosphatase that removes the phosphate form of Ser5 on U2 snRNA genes. Worth noticing is the fact, that the profile of the increase of Ser5P levels after RPAP2 and Ssu72 depletion along the U2 snRNA gene differs (compare Figure 3-10 and Figure 3-17). Bataille and colleagues suggested that Ser5P is removed in two waves in yeast, first by the RPAP2 homolog Rtr1 and later by Ssu72 (Bataille et al., 2012). In accordance with their findings, my results suggest that on the U2 snRNA gene, Ser5P is dephosphorylated first by RPAP2 and later by Ssu72.

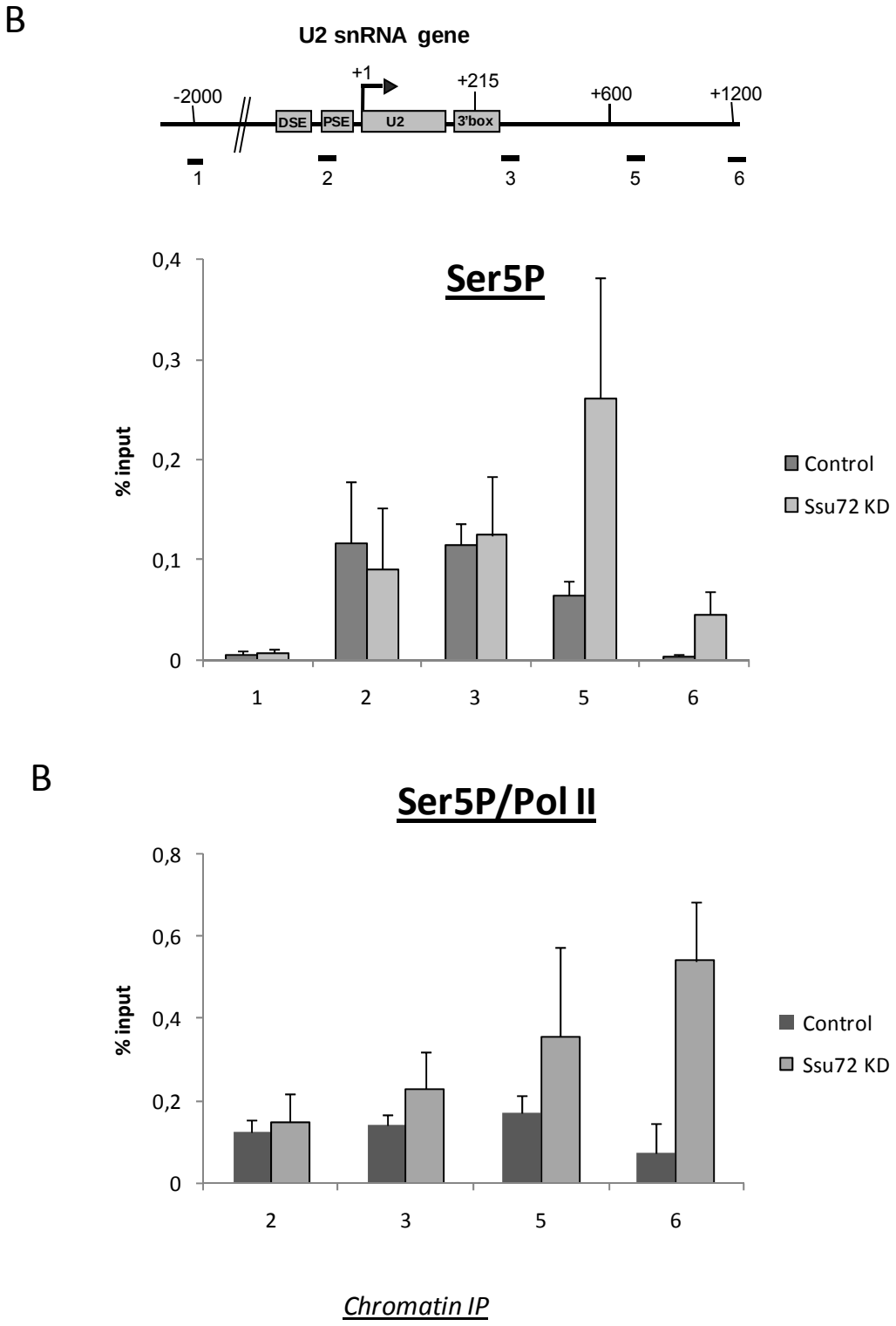


Figure 3-17. The effect of Ssu72 knockdown on Ser5P pol II transcribing U2 snRNA gene.
(A) The results of qPCR analysis of the output of ChIP using anti-Ser5P antibody is shown under the schematic of the U2 snRNA gene. The regions amplified are noted on the schematic. (B) qPCR quantitation of ChIP analysis of Ser5P levels related to pol II associated with the U2 snRNA gene. Primers used are noted on the schematic above. Error bars indicate the standard deviation of means from at least three independent experiments.

3.3 Conclusions

In this study, I have shown that the RPAP2 protein is associated with both protein-coding genes and snRNA genes. In addition, I have demonstrated that recombinant RPAP2 interacts directly with the CTD and that Ser7 and phosphorylation of the pol II CTD is required for this interaction. Interestingly, RPAP2 knockdown reduces transcription of the β-actin and U2 snRNA gene and the efficiency of 3' box processing of the U2 snRNA gene. siRNA-mediated depletion of RPAP2 causes also loss of the Int11 subunit of the Integrator complex from the U2 snRNA gene. These defects are similar to those observed when Ser7 of the CTD is mutated to alanine (Egloff et al., 2007). Importantly, Ser5P levels increase *in vivo* after RPAP2 knockdown, indicating that RPAP2 might act as a Ser5 phosphatase. Moreover, RPAP2 is detected on protein-coding genes, but the association is Ser7P-independent. Taken together, these results suggest that during transcription of snRNA genes, Ser7 phosphorylation facilitates recruitment of the RPAP2 protein, to dephosphorylate Ser5P and in turn to recruit Integrator. Interestingly, RPAP2 recruitment to the β-actin gene is not affected by mutation of Ser7, implicating a mechanism alternative to Ser7 phosphorylation in recruitment of RPAP2 to protein-coding genes. In addition, the phosphatase Ssu72 also plays an important role in transcription and in the regulation of CTD phosphorylation on the U2 snRNA and β-actin genes.

Some of these results were included in a publication in Molecular Cell in 2012 (Egloff et al., 2012b) and NAR in 2013 *[O'Reilly in press]*. Based on our findings, we proposed a model in which recruitment of RPAP2 to snRNA genes through CTD Ser7P results in stable association of Integrator and removal of Ser5P. This generates the characteristic pattern of CTD phosphorylation found on actively transcribed snRNA genes (Egloff et al., 2012b).

Chapter 4| Inhibition of CTD Ser2 phosphorylation by the HSV-1 ICP22 protein

4.1 Introduction

During Herpes simplex virus type 1 (HSV1) replication, the first viral proteins to be expressed are the immediate early proteins (IE) (Honess and Roizman, 1974, Sears and Roizman, 1990). The studies described here centre on the role of an immediate early protein 22 (ICP22). ICP22 is known to inhibit expression of some viral genes that are transcribed by pol II after the early stages of infection, which may be an important step in establishing latency. ICP22 has also been shown to inhibit expression of some host genes transcribed by pol II, which may contribute to infectivity (Advani et al., 2000, Cun et al., 2006, Guo et al., 2012). Ectopic expression of ICP22 in human cells has been shown to cause a change in the pattern of phosphorylation of the CTD of the large subunit of pol II, where phosphorylation of Ser2 of the YSPTSPS repeat of the pol II CTD is specifically inhibited (Rice et al., 1995, Fraser and Rice, 2007). Phosphorylation of Ser2 by the Cdk9 kinase subunit of Positive Elongation Factor b (P-TEFb) is associated with productive elongation of transcription and RNA processing (Egloff et al., 2012b) and ICP22 is therefore thought to interfere with the elongation process. There are some indications that ICP22 interacts with P-TEFb through Cdk9 or one of its cyclin partners (cyclin T1, T2 or K) and it has been proposed that ICP22 directly inhibits the activity of Cdk9 (Rice and Davido, 2013). Recently, Kolb and colleagues identified a 63 residues motif that is conserved in all alphaherpesvirus U_S1 homologs. This sequence is located in the central part of ICP22 protein. Interestingly, authors identified predicted cyclin binding sites that reside in this region (Kolb et al., 2011). Our laboratory has worked for many years on of the kinase complex, P-TEFb in transcription regulation and RNA processing (Medlin et al., 2003, Medlin et al., 2005, Egloff et al., 2007, Egloff et al., 2010). We routinely use a range of P-TEFb inhibitors in our studies (Medlin et al., 2003, Medlin et al., 2005, Egloff et al., 2009). However, these have limitations in that each drug inhibits more than a one cellular kinase. Therefore, recently I have started to explore the mechanistic insights of inhibition of CTD Ser2 phosphorylation by the HSV-1 ICP22 protein. As ICP22 is thought to inhibit P-TEFb, this could potentially provide an alternative means to study P-TEFb function. However, it has not yet been definitively demonstrated that ICP22 interacts with P-TEFb directly or inhibits its kinase activity. In addition, there are conflicting data as to whether Cdk9 functions as a Ser2 kinase on pol II transcribed snRNA genes. Data from our lab

indicate that Cdk9 is recruited to U2 snRNA genes (Egloff et al., 2009). This is in conflict with the data published by Glover-Cutter and colleagues in 2008, who found that there is no Cdk9 on the U2 snRNA genes (Glover-Cutter et al., 2008). Furthermore, the effect of ICP22 on expression of the host cell genome is not clear. I have therefore initiated a study investigating the mechanism of ICP22 inhibition of CTD phosphorylation by expressing ectopically ICP22 in HeLa cells and looking at its effect on the U2 snRNA gene and PLK2 protein-coding gene. PLK2 was chosen as it has been tested in our laboratory in response to Cdk9 inhibitors.

Almost all previous studies of ICP22 have relied on comparison between cells infected with wild-type HSV or HSV with mutated versions of ICP22. Since the expression of many viral genes are dependent on ICP22, it is difficult to determine what is a direct/or indirect effect of ICP22. My approach allows the characterization of truncated forms of ICP22 without the need to introduce mutations directly into the viral genome. The purpose of this study is to understand the mechanism by which ICP22 triggers the loss of phosphorylated form of Ser2 pol II CTD in order to determine the role of Cdk9 in expression of snRNA genes.

4.2 Results

4.2.1 Myc - ICP22 is successfully expressed in transfected HeLa cells

In order to study the functional properties of the ICP22 protein, I have constructed a vector expressing Myc-tagged ICP22. The ICP22 ORF was amplified by PCR and cloned in frame into a pcDNA3 plasmid containing three tandem repeat Myc epitope tags to generate a vector with Myc-ICP22 driven by the cytomegalovirus (CMV) promoter (constructs were generated by Dr Shona Murphy). Whole-cell lysates were prepared fifteen hours after transfection with an empty vector (Control) and the Myc-ICP22 vector (Figure 4-1). Myc-ICP22 levels were determined by Western blot analysis with an anti-Myc epitope antibody. The band migrating at approximately 68 kDa is the expected size for Myc-ICP22 (Figure 4-1). As mentioned before, the ICP22 gene (U_S1) also encodes another protein, known as $U_S1.5$ (Carter and Roizman, 1996). The $U_S1.5$ protein lacks the N-terminal part of ICP22. As the tag is located at the C-terminus of ICP22, the antibody used detects this smaller Myc-$U_S1.5$ protein, which migrates approximately at 40 kDa. These data are consistent with the results of Fraser and Rice (2007) who found that that ICP22 is efficiently expressed when transcription is driven by the CMV promoter (Fraser and Rice, 2007).

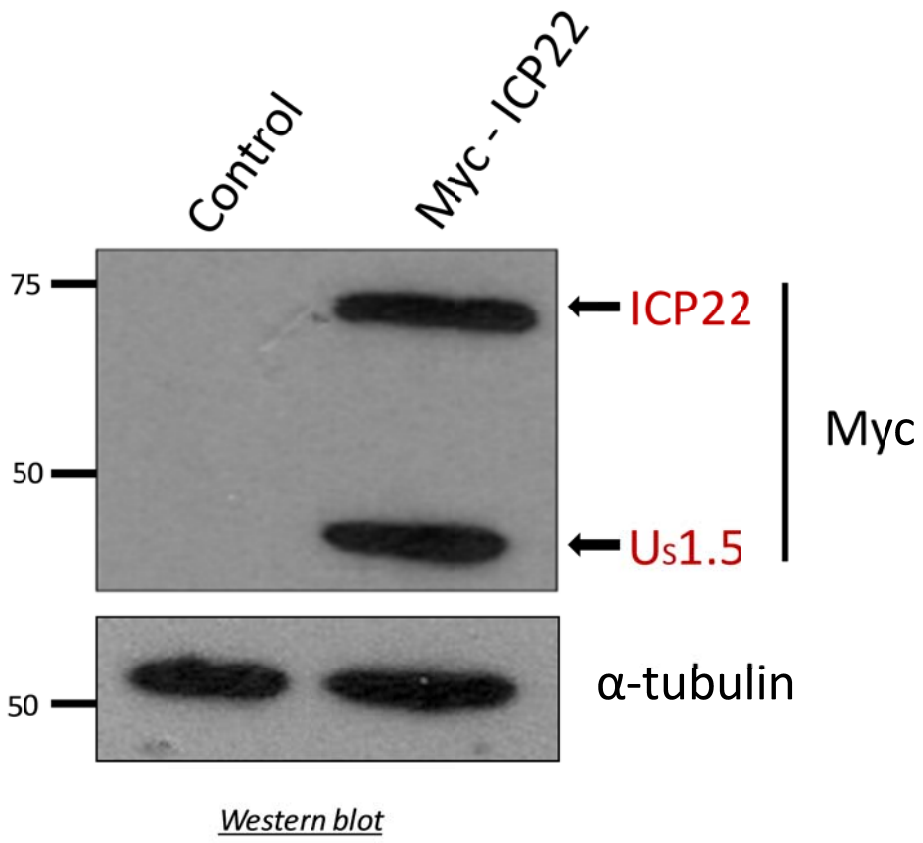


Figure 4-1. The U_S1 gene expresses ICP22 and $U_S1.5$ protein *in vivo*.
The Myc-ICP22 construct was transfected into HeLa cells (Myc-ICP22). Whole cell extracts were analyzed by Western blot using an anti-Myc antibody (Myc). Extract from cells transfected with an empty pcDNA3 vector was used as a control. α-tubulin was used as a loading control. The arrows indicate the bands corresponding to the expressed proteins noted on the right.

4.2.2 Transient transfection of Myc-ICP22 in HeLa cells specifically targets phosphorylation of Ser2 of the pol II CTD.

In attempt to further characterize the function of ICP22, the phosphorylation status of the pol II CTD was investigated by Western blot using antibodies specific to phospho-Ser2 (Ser2P), phospho-Ser5 (Ser5P) and phospho-Ser7 (Ser7P) after transfection of HeLa cells with Myc-ICP22. After equalization of the loading control, CTCF, it was found that ICP22 specifically targets Ser2P of the pol II CTD. Ser2P is consistently lost, whereas the levels of Ser5P and Ser7P are unaffected (Figure 4-2). This is consistent with previously published results showing that transient expression of ICP22 is sufficient to trigger the loss of Ser2P pol II (Fraser and Rice, 2007). This result further demonstrates that ICP22 specifically regulates global levels of Ser2P in the cell.

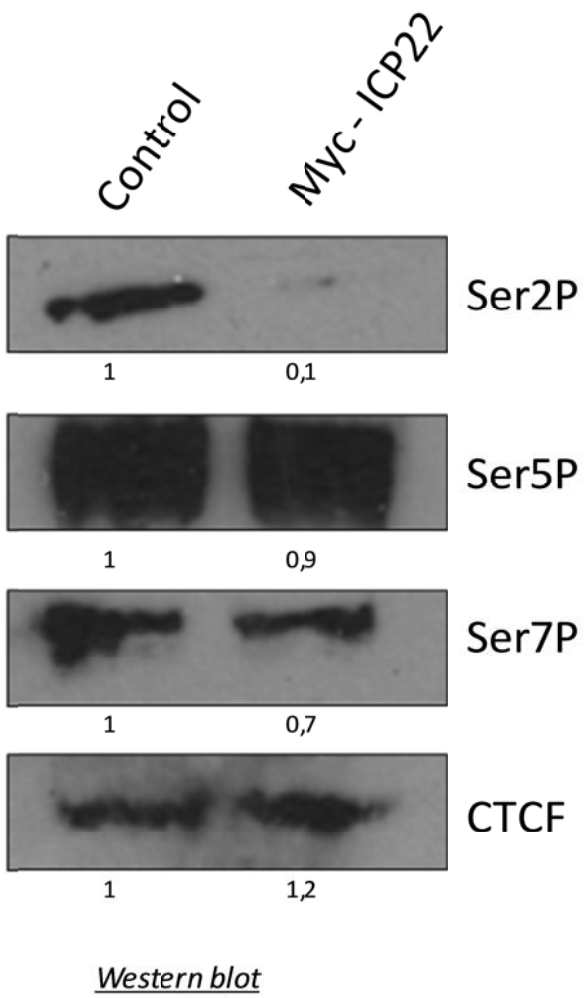


Figure 4-2. Myc-ICP22 specifically triggers loss of Ser2P in HeLa cells.
The Myc-ICP22 construct was expressed in HeLa cells and its effect on pol II was analyzed by Western blot using antibodies directed against the different phosphorylated forms of pol II. Extract from cells transfected with an empty pcDNA3 vector was used as a control. CTCF was used as a loading control. Quantification of signal strength is denoted below each line.

4.2.3 ICP22 is recruited to cellular genes

To test whether ICP22 is recruited to protein-coding genes during transcription, I have analyzed its association with the *PLK2* gene by ChIP using antibodies to Myc epitope at the C terminus of ICP22. Figure 4-3 shows a model of the *PLK2* gene with the relative locations of the PCR primers used for analysis. This gene was chosen for this and further analysis, because it is one of many that has been tested in our laboratory in response to Cdk9 inhibitors. Results from our laboratory indicate that the drugs have a profound effect on the ability of pol II to make the transition to productive elongation after initiation, presumably due to inhibition of P-TEFb *[Laitem and Murphy, unpublished results]*. My data indicates that after transfection of the Myc-ICP22 vector into HeLa cells, Myc-ICP22 is detected by ChIP on the *PLK2* promoter region. Therefore, Myc-ICP22 is recruited to *PLK2* and it is preferentially associated with the promoter of the gene (Figure 4-3; probe 2).

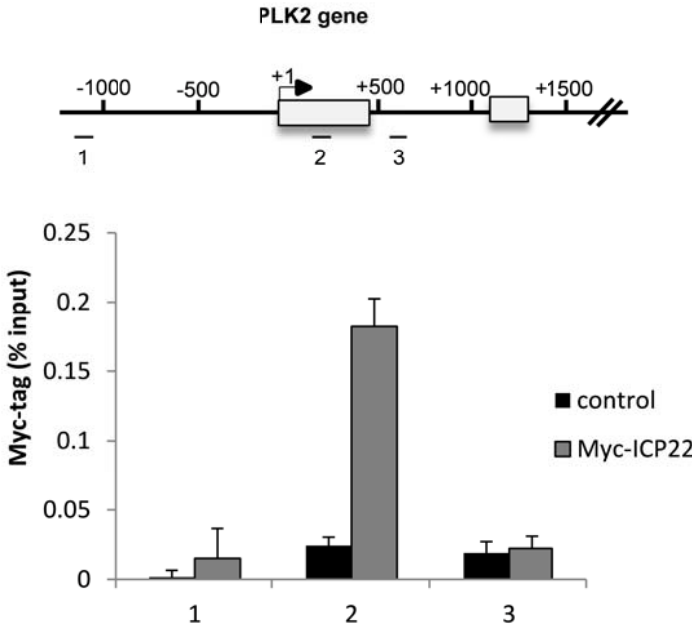


Figure 4-3. Myc-ICP22 associates with the *PLK2* gene.
A schematic of the first 1500 bp of the *PLK2* gene is shown at the top. Exons are denoted with boxes, introns with lines. Horizontal lines below the gene schematic mark the position of the quantitative real-time PCR (qPCR) primer pairs used in the ChIP analysis. The arrow represents the transcription start site. HeLa cells were transfected with an empty pcDNA3 vector (Control) and a pcDNA3 vector expressing Myc-ICP22 (Myc-ICP22). The results of qPCR analysis of the output of ChIP using anti-Myc antibody is shown at the bottom. Error bars indicate the standard deviation of means from at least three independent experiments in this and subsequent figures.

4.2.4 Ser2P is lost from engaged pol II after ICP22 transfection

The effects of ICP22 protein on transcription of cellular genes are not clear. In 1988, it was demonstrated that repression of transcription by HSV1 requires the expression of ICP22 (Kemp and Latchman, 1988). However, Spencer and colleagues found that ICP22 has little effect on the transcription of several selected host genes and suggested that the shutoff of host gene transcription is triggered by other viral components (Spencer et al., 1997). Therefore, further studies are required to determine whether ICP22 inhibits the ability of pol II to transcribe cellular genes. My system facilitates the study of the effect of ICP22 on host cell transcription. Since ICP22 is recruited to at least one cellular gene (Figure 4-3), I have asked whether ICP22 inhibits the ability of pol II to transcribe the gene. I have looked at pol II occupancy on *PLK2* in cells transfected with a pcDNA3 vector expressing Myc-ICP22 and compared it to levels observed in cells transfected with the empty pcDNA3 vector. The data presented in Figure 4-4 demonstrates that after ICP22 expression pol II levels increase at the promoter region (Figure 4-4, probe 2) and then drastically decrease in the gene body (Figure 4-4, probe 3). Thus, ICP22 appears to repress transcription by pol II of at least one protein-coding gene. Interestingly, in the presence of two drugs, KM05382 (KM) and DRB, known to inhibit the Cdk9 activity of P-TEFb (Mancebo et al., 1997), the pol II profile on *PLK2* looks very similar *[Laitem and Murphy, unpublished observations]*. This suggests that ICP22 has the same effect as P-TEFb inhibitors. ICP22 may affect gene expression, directly or indirectly, by inhibiting Cdk9 activity.

Next, I asked whether Ser2P of pol II CTD was affected on actively-transcribing pol II in the presence of ICP22 protein. HeLa cells were transfected with an empty pcDNA3 vector (Control) and the Myc-ICP22 vector (Myc-ICP22). The results of qPCR analysis of the output of ChIP using anti-Ser2P antibody are shown in Figure 4-5. Consistent with Cdk9 inhibition, ChIP indicates that Ser2P levels are greatly reduced in the gene body of *PLK2* following expression of ICP22.

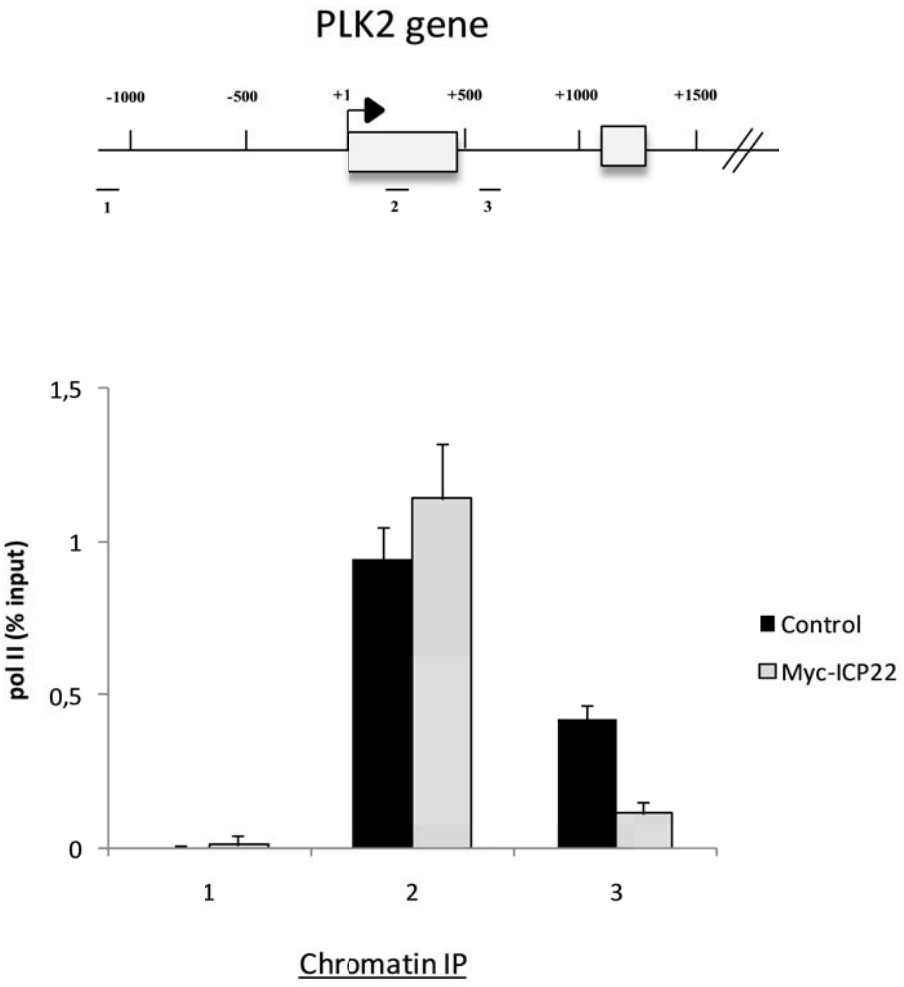


Figure 4-4. ICP22 affects pol II levels on the *PLK2* gene.
A schematic of the first 1500 bp of the *PLK2* gene is shown at the top. Horizontal lines below the gene schematics mark the position of the quantitative real-time PCR (qPCR) primer pairs. The arrow represents the transcription start site. HeLa cells were transfected with an empty pcDNA3 vector (Control) and a Myc-ICP22 vector (Myc-ICP22). The results of qPCR analysis of the output of ChIP using anti-pol II antibody are shown at the bottom. Error bars indicate the standard deviation of means from at least three independent experiments in this and subsequent figures.

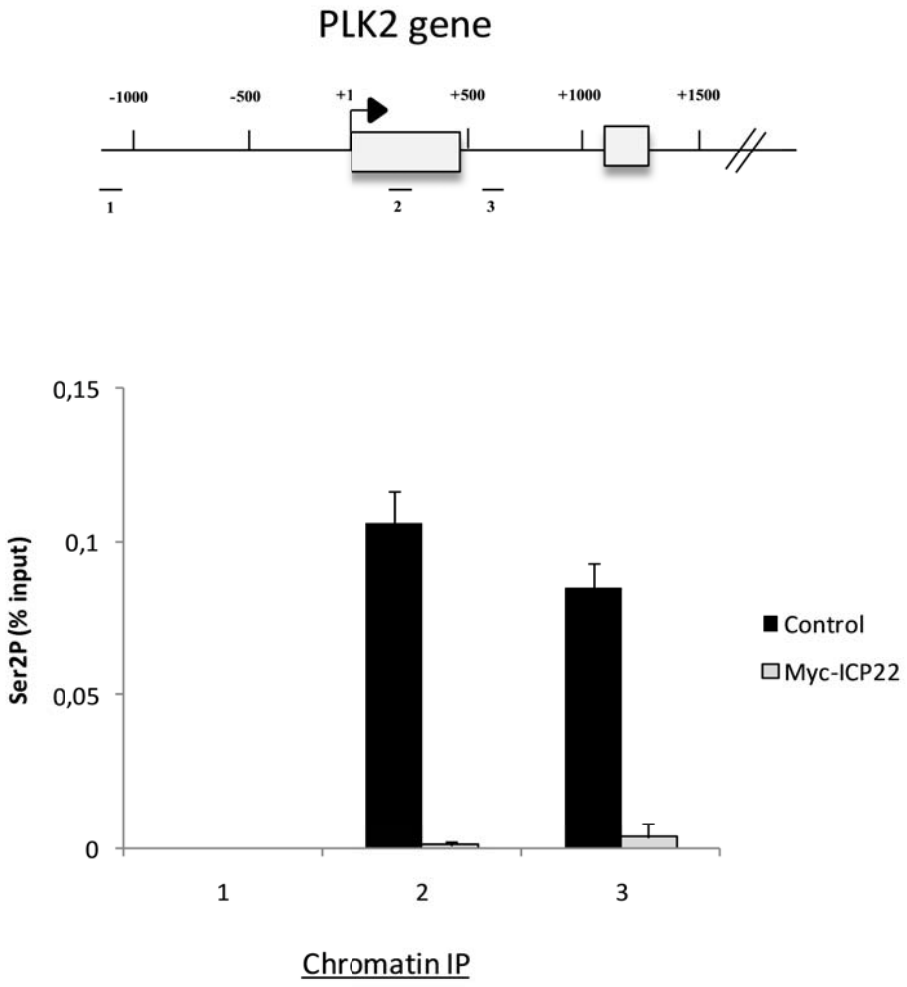


Figure 4-5. The Ser2P pol II levels are affected on the *PLK2* gene following transfection of Myc-ICP22.

A schematic of the first 1500 bp *PLK2* gene is shown at the top. Horizontal lines above the gene schematics mark the position of the quantitative real-time PCR (qPCR) primer pairs. The arrow represents the transcription start site. HeLa cells were transfected with an empty pcDNA3 vector (Control) and a Myc-ICP22 vector (Myc-ICP22). The results of qPCR analysis of the output of ChIP using anti-Ser2P antibody is shown at the bottom.

4.2.5 ICP22-derived constructs were successfully expressed in transfected HeLa cells

To determine which part of ICP22 mediates the loss of Ser2P, I have constructed a series of expression plasmids encoding truncated forms of ICP22. A schematic of the generated constructs is shown in Figure 4-6 A. All the constructs were prepared with a C-terminal three tandem repeat Myc epitope tag. Next, I have performed Western blot analysis, using anti-Myc antibody, to confirm that constructs express the protein of the expected size. Plasmids were transfected in HeLa cells, whole cells extract prepared after 15h from transfection and analyzed using anti-Myc antibody (Figure 4-6B). Western blot analysis of the HeLa whole cell extract transfected with pcDNA3 encoding for Myc-ICP22 shows two proteins - a 68 kD protein corresponding to ICP22 and 40 kD protein corresponding to $U_S1.5$ as expected. In cells transfected with pcDNA3 vectors encoding clone 1 (residues 240-340), clone 3 (residues 171-320) and clone 4 (residues 193-256), I detect a band of the expected size. Interestingly, after transfection of the vector encoding clone 2 (residues 171-420), a main product is running at the same molecular weight as $U_S1.5$ (40 kDa). The protein $U_S1.5$ was reported to initiate at codon 147 and later it was reported by the same group to initiate at codon 171 (Carter and Roizman, 1996, Ogle and Roizman, 1999). Recently, another group concluded that $U_S1.5$ initiates at codon 90 (Bowman and Schaffer, 2009). Therefore, the extent of the $U_S1.5$ is still not clear. In my system, $U_S1.5$ appears to initiate at or close to codon 171 as protein comprising aa 171 to 420 (clone 2) migrates similarly to that of the $U_S1.5$ protein expressed from the full length Myc-ICP22 plasmid. Western blot analysis of the HeLa whole cell extracts after transfection of the vector encoding for clone 5 (residues 90-420) detect a protein corresponding to the expected size and also an additional product corresponding to the 40 kDa $U_S1.5$ protein. Taken together, the results show that the construct generated express proteins of the expected sizes. Apart from clone 1, proteins are expressed roughly at similar levels. The multiple forms of proteins expressed from some of the clones might be due to posttranslational modification by cellular enzymes or cleavage targeted.

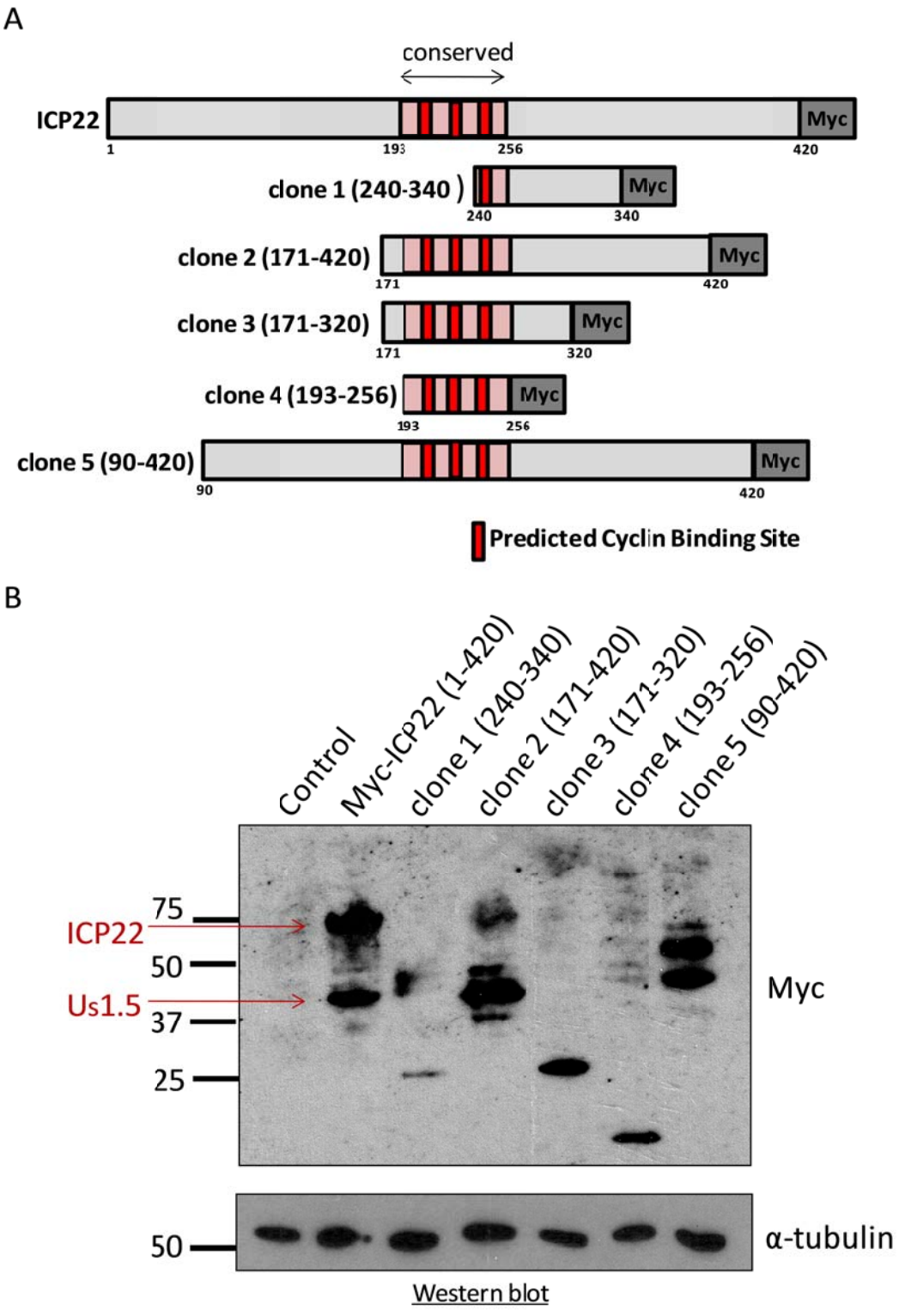


Figure 4-6. Truncated Myc-ICP22 proteins are successfully expressed in HeLa cells.
(A) The schematic of truncated Myc-ICP22 derived constructs. The dark grey box represents the C-terminal three tandem repeats of the Myc epitope. The conserved 63-residue motif, predicted cyclin binding sites and aa size are indicated. (B) The indicated constructs were expressed in HeLa cells and analysed by Western blot using an anti-myc antibody. Extract from cells transfected with the pcDNA3 plasmid was used as a control. α-tubulin was used as loading control.

4.2.6 The region of ICP22 encoding the cyclin binding sites (clone 4) retains the ability to inhibit Ser2P of pol II CTD

I next investigated which part of ICP22 is responsible for the loss of Ser2P. I have analysed the effect on CTD phosphorylation by measuring Ser2P in whole-cell extracts after transfection in HeLa cells plasmids encoding Myc-ICP22 and the different truncated forms of the ICP22 protein (indicated in Figure 4-6). The proteins are expressed at roughly similar levels (Figure 4-7 A and 4-7 B). My results show that all truncated forms of ICP22 trigger the loss of Ser2P pol II (Figure 4-7 C and 4-7 D). The results are in accordance with previous study which demonstrated by immunofluorescence, that the C-terminal half of ICP22 (residues 240-340; in this study clone 1) is sufficient to trigger loss of Ser2P pol II in transfected kidney epithelial cells of the African Green Monkey (Vero cells) (Bastian and Rice, 2009). Here, I have confirmed by Western blot that clone 1 (residues 240-340) also triggers the loss of Ser2P in transiently transfected HeLa cells.

Furthermore it can be concluded that clone 4 (residues 193-256) triggers the loss of Ser2P pol II as efficiently as full length ICP22**.** As mentioned before, the 63 residues of clone 4 (residues 193-256) contains a motif conserved in all α-herpesvirus U_s1 homologs. It contains three predicted cyclin binding sites (Kolb et al., 2011) as opposed to one in clone 1 (residues 240-340). This data suggest that the 63 residues conserved domain may play a significant role regulating ICP22 function blocking Ser2P of pol II.

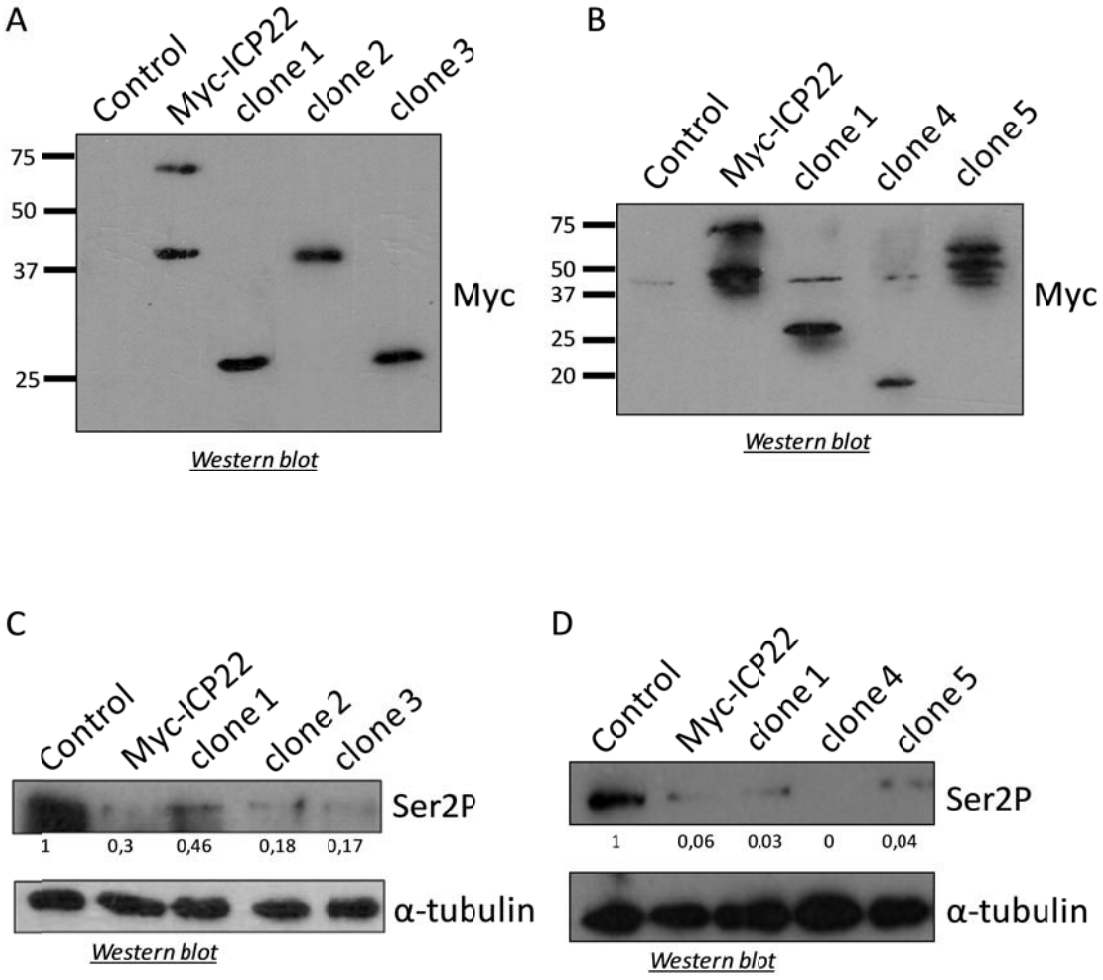


Figure 4-7. Clone 4 is the shortest form of ICP22 that triggers the loss of Ser2P pol II as efficiently as full-length ICP22.

(A) and (B) The indicated constructs were expressed in HeLa cells and expression analysed by Western blot using anti-Myc antibody. Extract from cells transfected with empty pcDNA3 vector was used as a control. (C) and (D) The Myc-ICP22 constructs indicated on the figure were expressed in HeLa cells and their effect on Ser2P was analysed by Western blot using an antibody directed against Ser2P (H5). The numbers indicate the quantitation of Ser2P signal strength. Extract from cells transfected with empty pcDNA3 vector was used as a control. α-tubulin was used as a loading control.

4.2.7 Loss of cyclin binding sites impair ICP22 activity *in vivo*

Previously, I demonstrated that clone 4 comprising ICP22 residues 193-256 efficiently triggers loss of Ser2P (Figure 4-7). Next, I constructed a clone (clone 6, Δ193-256, Figure 4-8 A) that has an in-frame deletion of these residues in order to confirm that this is the sequence of ICP22 responsible for Ser2P loss. To facilitate analysis, the expressed protein has, as previously, a Myc C-terminal epitope tag. HeLa cells were transfected with empty pcDNA3 vector and pcDNA3 vector encoding Myc-ICP22 and clone 6 (Δ193-256). Cells were harvested after 15 h and analysed by Western blot using a Myc-tag specific antibody. The Western blot analysis shows that the construct expresses a protein of the expected size (Figure 4-8 B). I next tested whether expressed clone 6 affects endogenous Ser2P levels. As previously, transfected HeLa cells were harvested after 15h and analysed by Western blot using a Ser2P-specific antibody. From the result, I can conclude that expression of the clone 6 (Δ193-256) in HeLa cells does not affect Ser2P level in whole-cell extracts (Figure 4-8 B). This results support the notion that residues 193-256 of ICP22 contain the core residues involved in loss of modification of the pol II CTD.

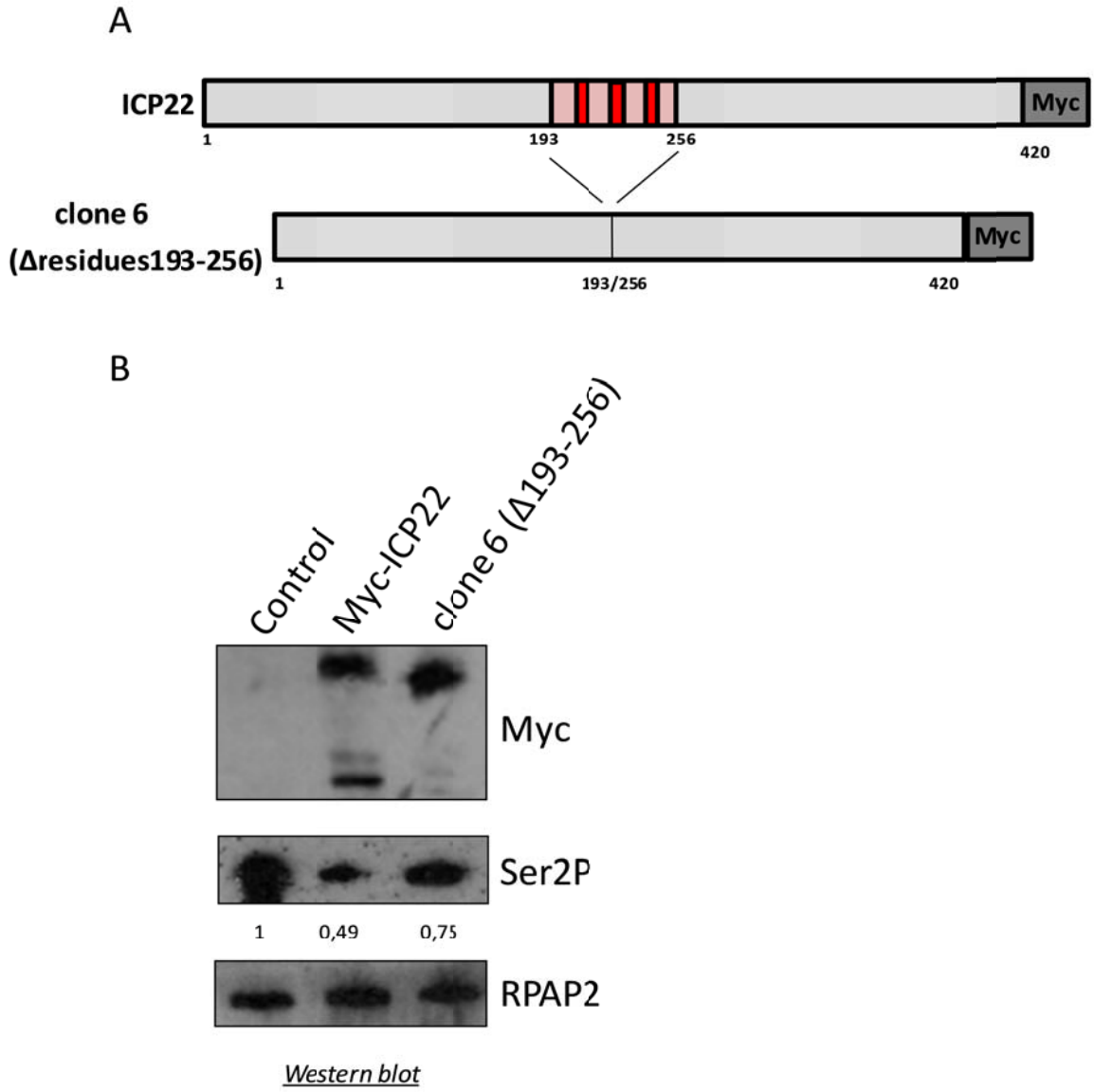


Figure 4-8. Cyclin domain sites are required for Myc-ICP22 activity.
(A) The schematic of Myc-ICP22 and clone 6 (Δ193-256) constructs. The dark grey box represents the C-terminal three tandem Myc epitope. The residues of ICP22 missing in clone 6 are indicated clone. (B) The constructs were transfected in HeLa cells and proteins level were analysed by Western blot using anti-Myc, anti-Ser2P (3E10) and anti-RPAP2 antibody. Extract from cells transfected with empty pcDNA3 plasmid was used as a control. RPAP2 was used as loading control. Quantification of signal strength is noted below each line.

4.2.8 Binding cyclin domain sites are crucial for the loss of Ser2P mark on actively transcribing pol II

I have demonstrated previously that expression of Myc-ICP22 in HeLa cells interferes with the pol II occupancy on *PLK2* gene by affecting the Ser2P levels associated with the gene (Figure 4-4 and Figure 4-5). Next step was to determine if its truncated form, clone 4 (residues 193-256), can function in the same way and also causes loss of only Ser2P. Accordingly, I transfected HeLa cells with an empty pcDNA3 vector (Control), with Myc-ICP22 vector (Myc-ICP22) and the truncated form (clone 4 and clone 6) and determined their effect on pol II, Ser2P and Ser5P levels by ChIP. After transfection of clone 6, the levels of pol II, Ser2P and Ser5P are not affected in comparison to control cells, as expected (Figure 4-9 and Figure 4-10). In contrast, transfection of clone 4 has a similar affect as transfection with full length Myc-ICP22; i.e., the pol II levels slightly increase in the promoter proximal region and then drop on probe 3. ChIP analysis also indicates that Ser2P levels in the *PLK2* gene body are also greatly reduced after transfection of HeLa cells with vector expressing Myc-ICP22 and clone 4. As expected, the levels of Ser5P are not affected (Figure 4-9 and Figure 4-10).

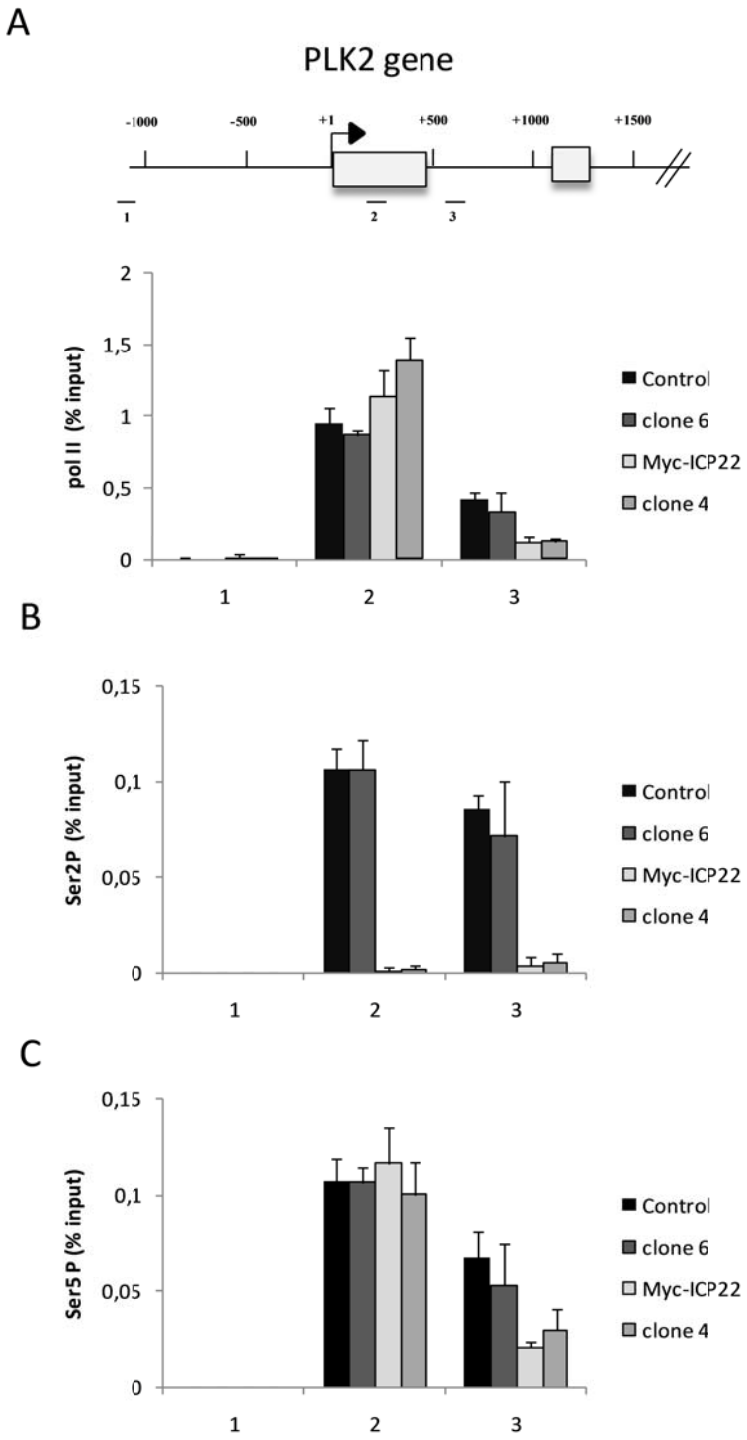


Figure 4-9. Pol II and Ser2P levels are affected on *PLK2* gene after expression of ICP22 and clone 4, but not clone 6.

A schematic of the first 1500 bp of the *PLK2* gene is shown at the top. Horizontal lines above the gene schematics mark the position of the quantitative real-time PCR (qPCR) primer pairs. The arrow represents the transcription start site. qPCR quantitation of ChIP analysis of the *PLK2* gene after transfection with Myc-ICP22 vector or the truncated forms (clone 4 and clone 6), using an antibody against (A) pol II, (B) Ser2P and (C) Ser5P. Error bars indicate the standard deviation of means from at least three independent experiments in this and subsequent figures.

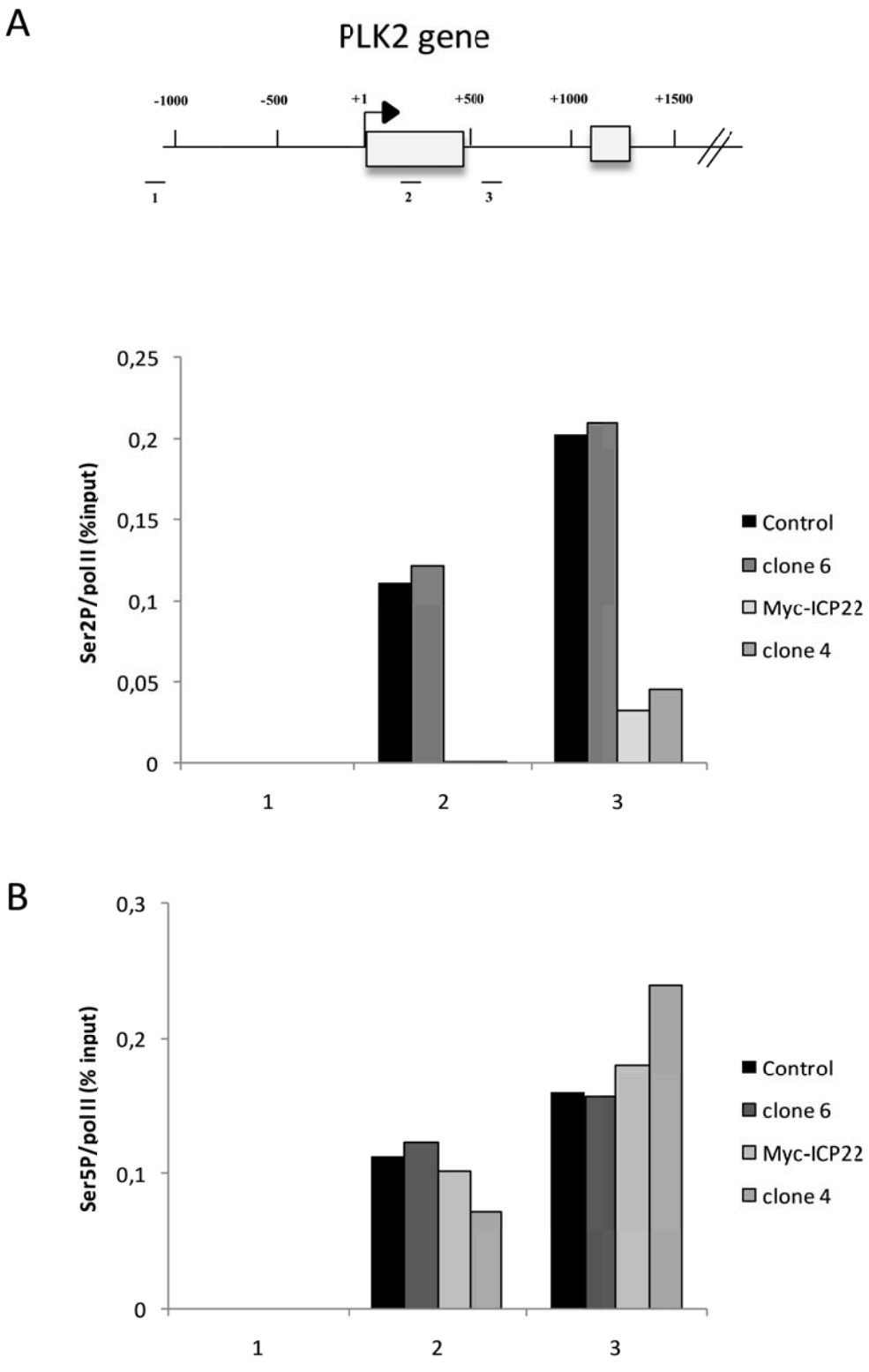


Figure 4-10. The Ser2P, but not Ser5P levels are affected on the *PLK2* gene after transfection with ICP22 and clone 4.

qPCR quantitation of ChIP analysis of the *PLK2* gene after transfection with Myc-ICP22 vector and the truncated forms (clone 4 and clone 6) using an antibody (A) against Ser2P relative to the pol II signal and (B) against Ser5P relative to the pol II signal. Primers used are noted on the schematic of the *PLK2* gene above.

4.2.9 Cdk9 is a Ser2P kinase on U2 snRNA genes

There is conflicting data as to whether Cdk9 is a Ser2P kinase for snRNA genes. Glover-Cutter and coworkers argue that Cdk9 is not recruited to snRNA genes (Glover-Cutter et al., 2008). However, ChIP data from our lab indicates that Cdk9 is recruited to U2 snRNA genes (Egloff et al., 2009) *[and Laitem and Murphy, unpublished results]*. To resolve this uncertainty, I have used my system to determine whether ICP22 transfection and the possible inhibition of Cdk9 has any consequences for transcription of the U2 snRNA genes. Accordingly, I have performed ChIP analysis of pol II, Ser2P and Ser5P on the U2 snRNA genes after transfection of HeLa cells with vector encoding Myc-ICP22 or its derived constructs. My results indicate that transfection of the Myc-ICP22, clone 4 and clone 6 do not affect pol II occupancy on U2 snRNA genes (Figure 4-11 A). This is consistent with published data, which show that the loss of Ser2P has no effect on the U2 snRNA gene expression (Medlin et al., 2003). After transfection of clone 6, the levels of pol II, Ser2P and Ser5P are not affected in comparison to control cells (Figure 4-11 and 4-12). As expected, the Ser2P levels decrease after expression of Myc-ICP22 and clone 4 (Figure 4-11 B and Figure 4-12 A), although not as drastically as on *PLK2* gene. Ser5P levels instead are maintained which is consistent with Cdk9 inhibition (Figure 4-11 C and Figure 4-12 B).

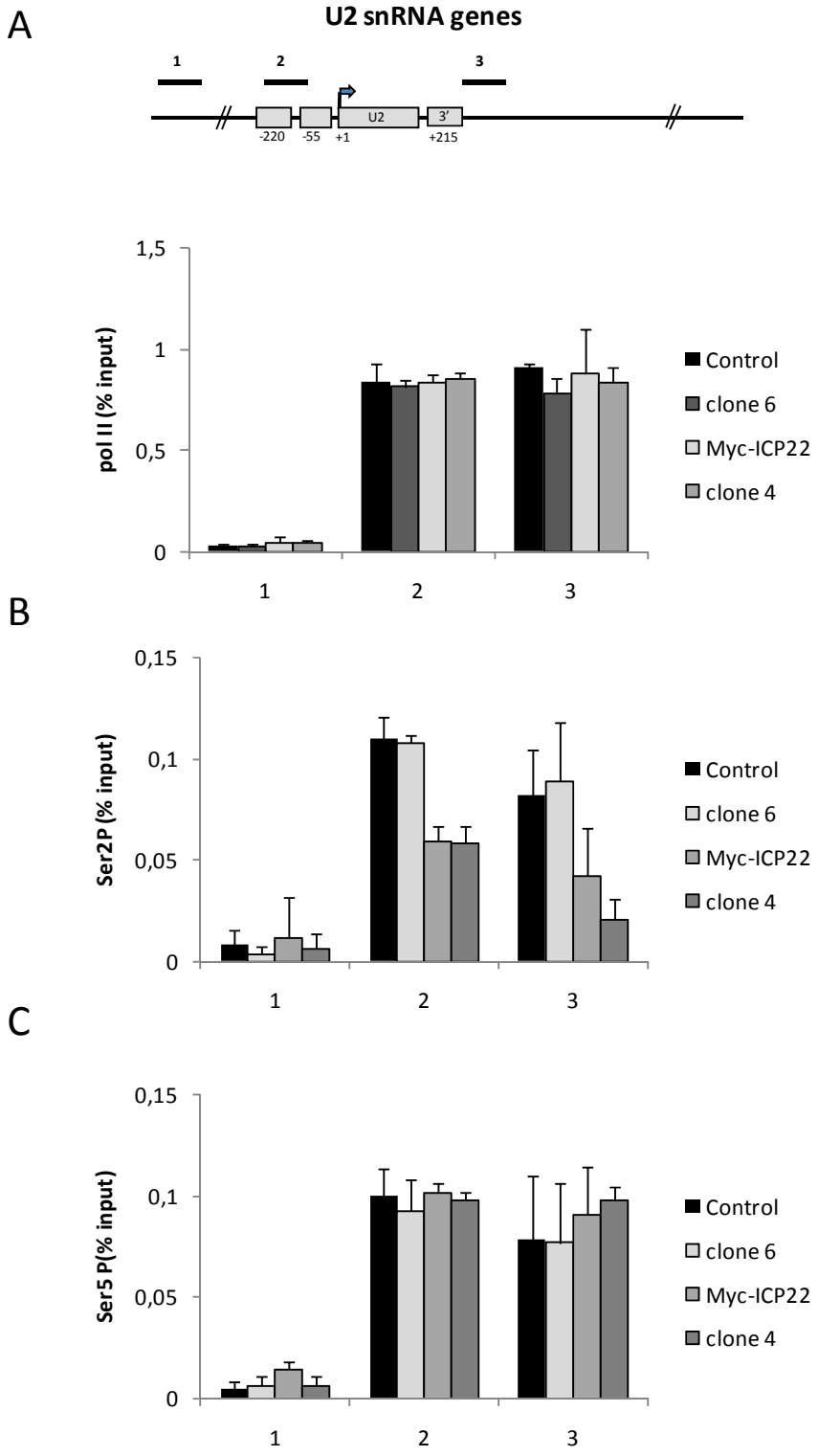


Figure 4-11. The Ser2P levels are affected on the U2 snRNA gene after transfection with Myc-ICP22 expressing vector and clone 4, but not clone 6.

A schematic of the U2 snRNA gene is shown at the top. Horizontal lines above the gene schematics mark the position of the quantitative real-time PCR (qPCR) primer pairs. The arrow represents the transcription start site. qPCR quantitation of ChIP analysis of U2 snRNA gene after transfection with Myc-ICP22 and its truncated forms (clone 4 and clone 6) using an antibody against (A) pol II, (B) Ser2P and (C) Ser5P.

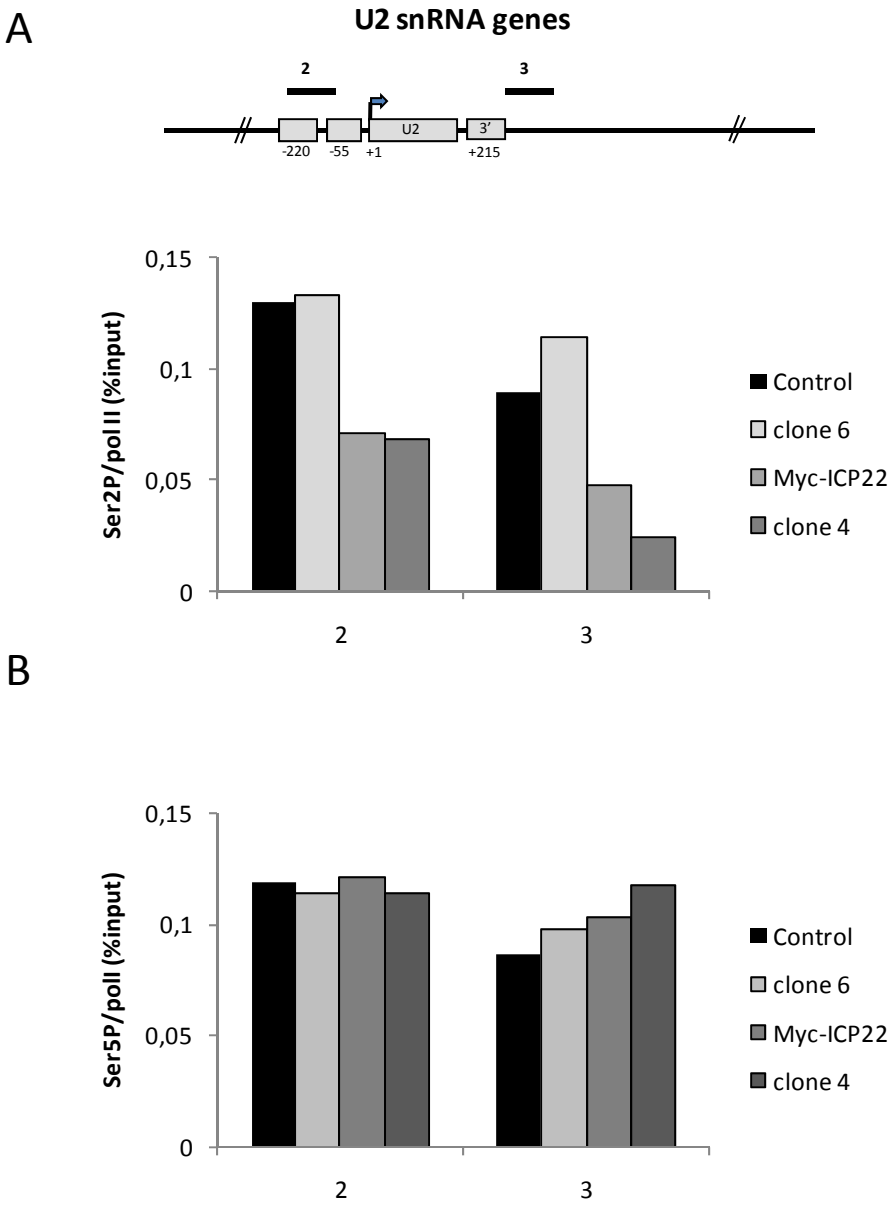


Figure 4-12. The Ser2P, but not Ser5P levels are affected on the U2 snRNA gene after transfection with ICP22 and clone 4.

qPCR quantitation of ChIP analysis of U2 snRNA gene after transfection with Myc-ICP22 and its truncated forms (clone 4 and clone 6) using an antibody (A) against Ser2P relative to the pol II signal and (B) against Ser5P relative to the pol II signal. Primers used are noted on the schematic of the U2 snRNA gene above.

4.2.10 Loss of Cdk9 does not affect the recognition of 3'box

Using kinase inhibitors, Medlin and co-workers demonstrated that Ser2P on the pol II CTD is important for correct 3' end formation of snRNA gene transcripts (Medlin et al., 2005). This suggested that CTD kinases targeted by kinase inhibitors, including P-TEFb, could contribute to efficient snRNA gene expression. To assess whether ICP22 affects 3' box-directed RNA processing, RNase protection of U2 snRNA gene transcripts was carried out. Surprisingly, the transfection of vectors encoding Myc-ICP22, clone 4 and clone 6 does not abolish the recognition of the 3' box (Figure 4-13). Furthermore, targeted knockdown of Cdk9 subunit of P-TEFb using siRNA mediated depletion, does not result in the loss of 3' box directed processing (Figure 4-14; by Dr Clelia Laitem).

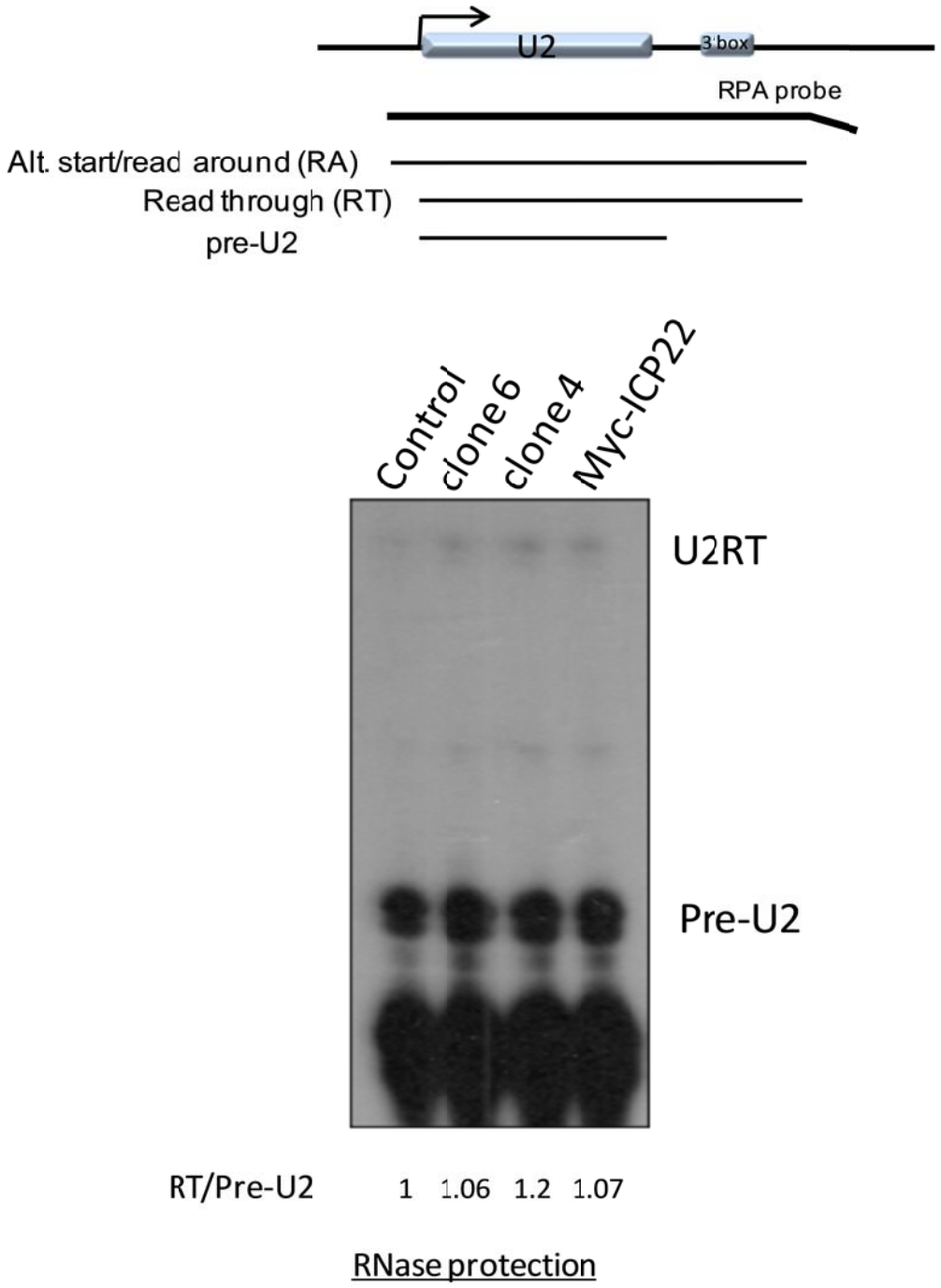


Figure 4-13. Expression of Myc-ICP22 and Myc-ICP22 derived construct does not affect 3' end processing.

RNase protection analysis of U2 transcripts in control cells and cells transfected with Myc-ICP22 and the derived constructs (clone 4 and clone 6). Location of the RNase Protection Analysis (RPA) probe and the expected products of the RPA are noted on the schematic. Position of the pre-U2 snRNA and unprocessed read-through (RT) U2 snRNA are noted on the right. Ratio of RT/Pre-U2 is also noted below each lane.

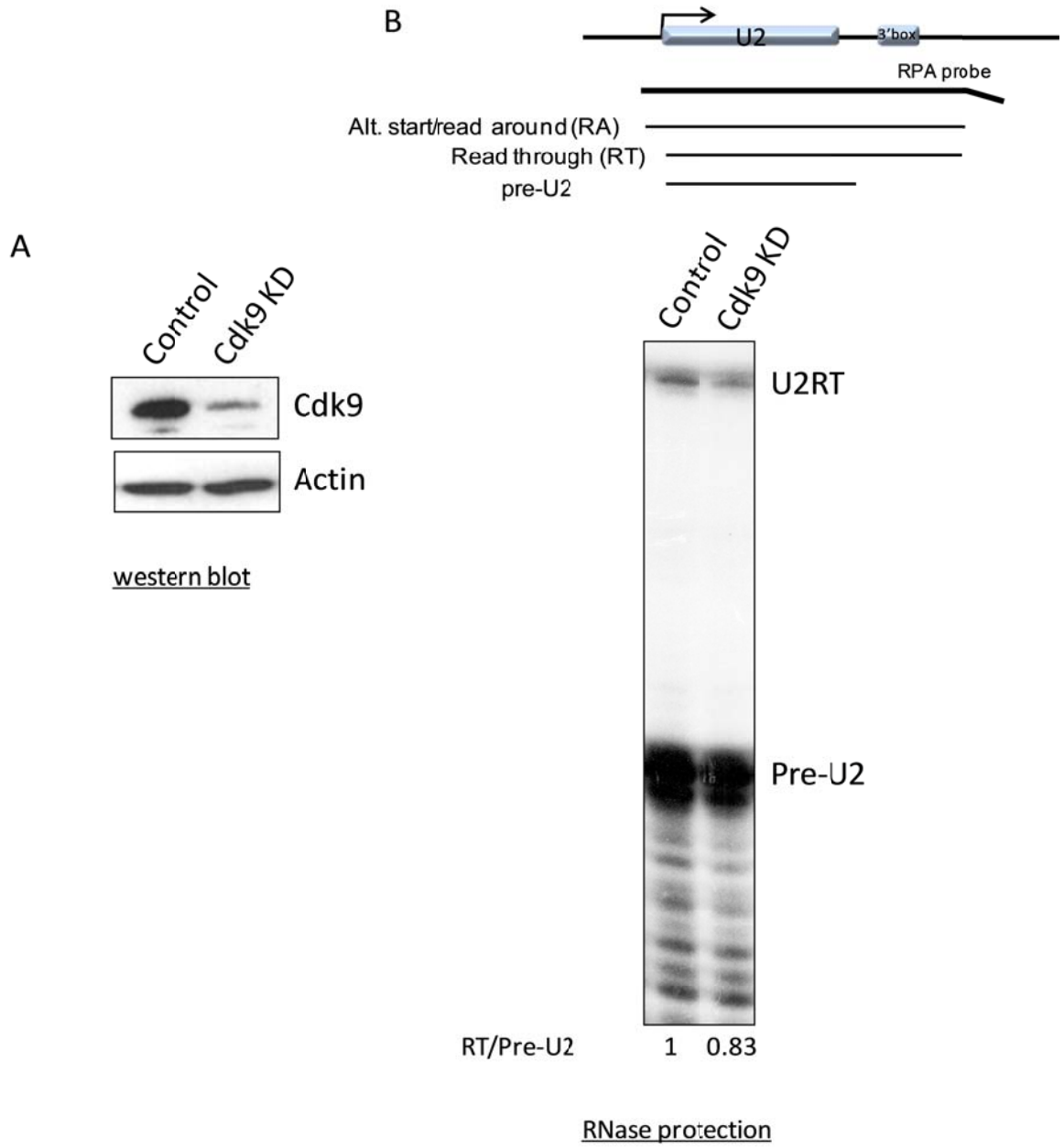


Figure 4-14. Cdk9 knockdown does not affect recognition of the U2 snRNA 3' box.
(A) Western blot analysis of whole-cell extracts from control cells or cells transfected with an siRNA specific for Cdk9 (Cdk9 KD). Antibodies used are noted on the right. (B) RNase protection analysis of U2 transcripts in control cells and cells transfected with an siRNA specific for Cdk9 (Cdk9 KD). Location of the RNase Protection Analysis (RPA) probe and the expected products of the RPA are noted on the schematic. Position of the pre-U2 snRNA and unprocessed read-through (RT) U2 snRNA are noted on the right. Ratio of RT/Pre-U2 is also noted below each lane.

4.3 Conclusions

In this study I have started to explore the mechanism of inhibition of CTD Ser2 phosphorylation by the HSV-1 ICP22 protein. Importantly, the recombinant viruses lacking the gene encoding this immediate early protein are defective in pathogenicity in mouse models and establishment of latency (Sears et al., 1985). It has been suggested that ICP22 interacts directly with P-TEFb to inhibit kinase activity (Durand and Roizman, 2008, Rice and Davido, 2013). As mentioned above (Chapter 1) P-TEFb is required to overcome the influence of negative elongation factors and allows polymerase to enter productive elongation (Nechaev and Adelman, 2011). As one of the main research interests of my laboratory is the function of Cdk9 in CTD phosphorylation and transcription elongation, I have carried out some initial studies on the effect of ectopic expression of ICP22 on expression of pol II-transcribed genes in HeLa cells. I have determined that ICP22 with a Myc epitope tag is well-expressed in HeLa cells and specifically causes loss of phosphorylation of Ser2 of the CTD. I have demonstrated that the C-terminal part of the ICP22 protein is essential for pol II modification. This is consistent with previously published data. In order to identify the residues in ICP22 that are required for modification of pol II, I made several ICP22 truncated constructs. My results reveal that the functional region of ICP22 required for inhibition of CTD Ser2 phosphorylation is between amino acids 193-256.

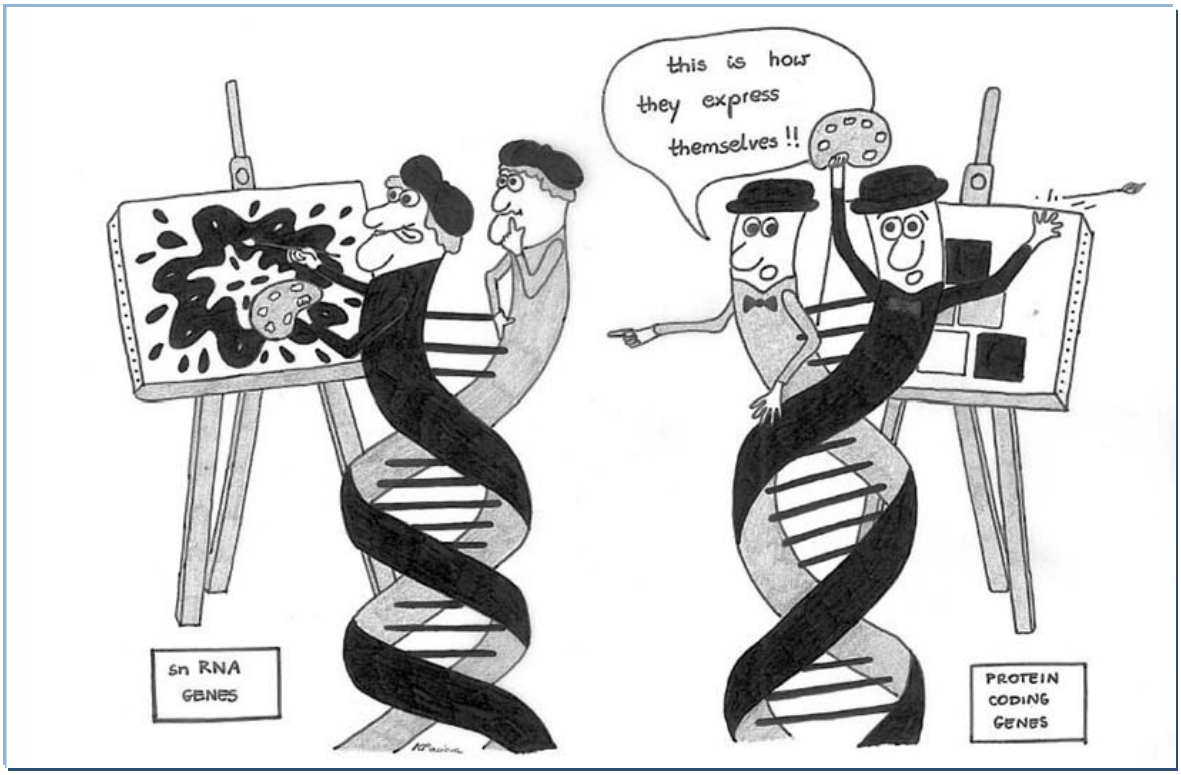
Taken together, my results recapitulate the effects of ICP22 expression in other systems and I therefore now have an ideal system to investigate the molecular mechanisms underlying inhibition of CTD phosphorylation by ICP22. My preliminary ChIP studies suggest that Ser2P on pol II transcribing *PLK2* and U2 snRNA gene is affected by ICP22 expression.

Surprisingly, knocking down Cdk9 (Figure 4-14, by Dr Clelia Laitem, unpublished data) or ectopic expression of ICP22 (Figure 4-13) is not sufficient to abolish recognition of the U2 snRNA 3’ box. These results might indicate that Cdk9 is not knocked down enough to see the effect on 3’ box recognition and that the Cdk9 associated with the U2 snRNA gene is not efficiently inhibited by ICP22. A low level of Ser2P remains on pol II transcribing the U2 snRNA gene after Myc-ICP22 expression (see Figure 4-11 and Figure 4-12) and this

may suffice for recognition of the 3'box. Alternatively, the loss of 3' box-directed RNA processing caused by Cdk9 inhibitors (Medlin et al., 2005), could be the result of inhibiting a different kinase.

My system and preliminary data are critical in establishing strategies that can be used to decipher the mechanism of action of ICP22 protein on cellular genes. Identification of the factors targeted by ICP22 is crucial to understand the molecular mechanism by which ICP22 affects Ser2P. For this purpose, I have constructed vectors expressing full length GST-ICP22 and truncated form of ICP22 with residues 193-256 (GST-clone 4) in order to perform pull-downs to identify interaction partners of ICP22. The intimate interaction of ICP22 with the cellular machinery provides an excellent tool to uncover mechanisms by which HSV viruses can so successfully coexists with humans.

Chapter 5| Discussion



5.1 A sub-complex of TFIID is loaded onto snRNA genes

5.1.1 Recruitment of a U2 snRNA gene-specific complex is essential for proper snRNA gene expression

TAFs are critical components of the machinery responsible for the transcription of eukaryotic genes transcribed by pol II, and their biological significance has made them the target of extensive study. Understanding the interplay between TAFs and other components of the transcription machinery provides important insights into how transcriptional regulation helps direct the growth and development of eukaryotic organisms. TAFs can function as promoter recognition factors, as coactivators and as enzymatic modifiers of other proteins. Some are important for the stability of complexes they form. Interestingly, the importance of these functions may vary from promoter to promoter. Taken together, my data suggest that a subcomplex of TFIID, containing TBP, TAF5, TAF6, TAF8, TAF9, TAF11, and TAF13 is loaded onto snRNA genes (Figure 5-1). Alternatively, TFIID may exists in different conformations on both the U2 and β-actin genes. However, the existence of a novel subcomplex on the snRNA genes is supported by the finding that knocking down TAF4 and TAF7 only affects β-actin and not U2 snRNA gene transcription. TAFs exist as components of a number of distinct complexes, but, interestingly, the proposed snTAFc does not contain the classical core TAFs (TAF4 and TAF10) and lacks the highest-molecular weight TAFs, TAF1 and TAF2 (Figure 2-4). A complete TFIID complex does not appear to be required for transcription initiation from snRNA gene promoters since knockdown of key TFIID subunits does not arrest transcription of the U2 snRNA gene (Figure 2-6). My ChIP analysis of a complete set of TAFs on the U2 snRNA and β-actin genes therefore provides further evidence for different TFIID compositions at different promoters (Zaborowska et al., 2012).

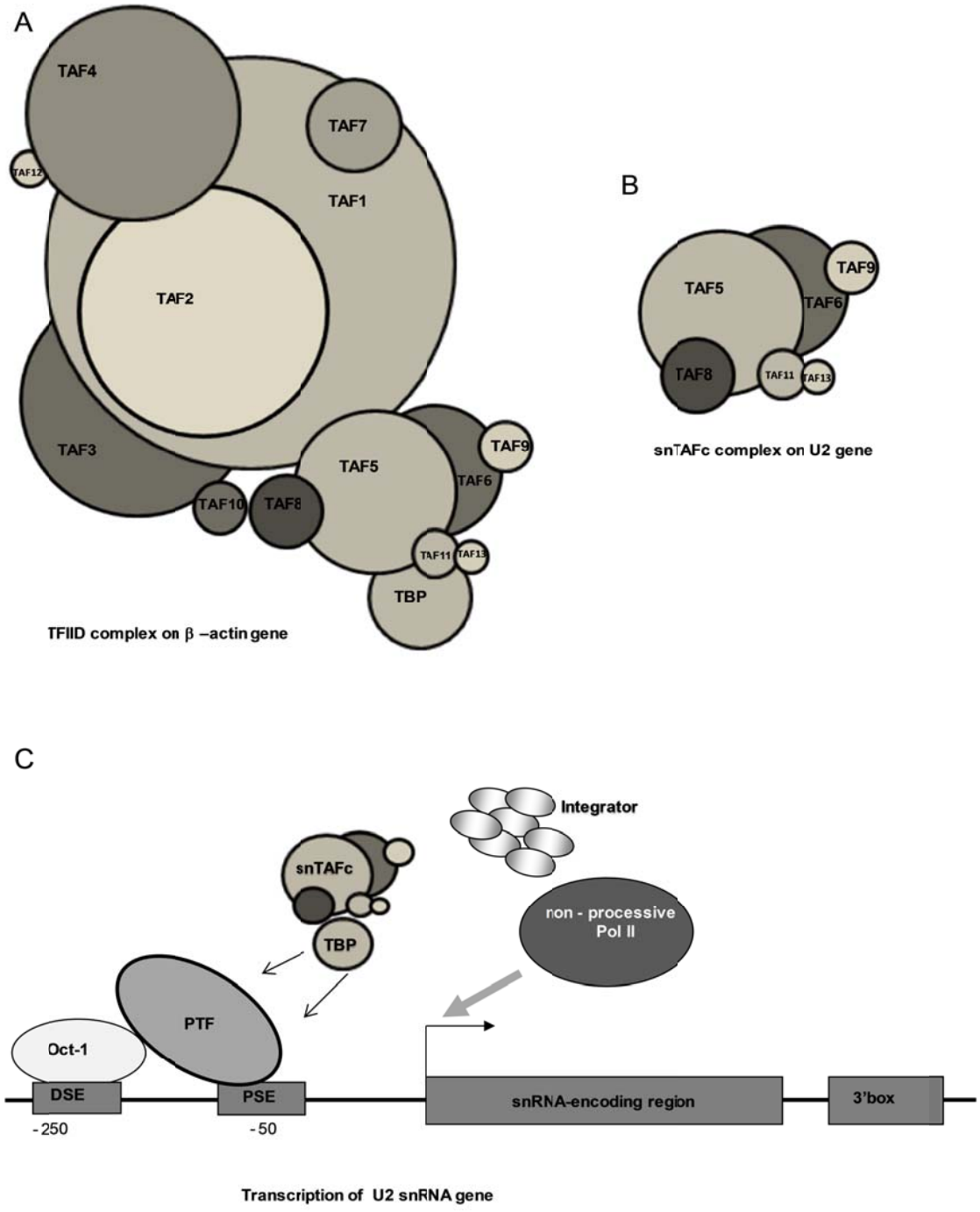


Figure 5-1. Recruitment of a U2 snRNA gene-specific TBP-TAF complex is essential for proper snRNA gene expression.

(A) On the β-actin gene, the histone fold containg TAFs (TAF3, TAF4, TAF6, TAF8, TAF9, TAF10, TAF11, TAF12 and TAF13) form heterodimers and interact with TAF1, TAF2, TAF5, TAF7 and TBP to form TFIID. Some of the TAFs, e.g., TAF10 may be present in more than one copy (Cler et al., 2009). (B) On the U2 snRNA gene, TAF5, TAF6, TAF8, TAF9, TAF11 and TAF13 associate to form an snRNA gene-specific-TAF complex (snTAFc). (C) On the snRNA gene, TBP associates with PTF and with snTAFc. This may facilitate specific recruitment of the Integrator complex and plays a role is setting the elongation properties of pol II, possibly due to the lack of recruitment of key elongation factors.

Interestingly, TAF1 is not present on the U2 snRNA gene (Figure 2-4). TAF1 has histone acetyltransferase (HAT) activity that acetylates histone H3 and H4 (Dunphy et al., 2000) also acts as a histone-specific ubiquitin-activating/conjugating enzyme to mediate monoubiquitination of histone H1 *in vitro* (Pham and Sauer, 2000, Maile et al., 2004). These activities contribute to the importance of this TAF in gene transcription. For example, transcription of the MHC class I genes is dependent on the HAT activity of TAF1 (Howcroft et al., 2003, Weissman et al., 1998, Gegonne et al., 2001). TAF7 binds to TAF1 and negatively regulates its HAT activity (Gegonne et al., 2001). It has been reported that TAF7 dissociates from the TFIID complex upon recruitment of pol II to the assembling PIC. This release is concomitant with phosphorylation of TAF1, which reduces the binding of TAF1 to TAF7. Additional studies have shown that TAF7 functionally interacts with TFIIH and elongation factor P-TEFb and inhibits their kinase activities. Accordingly, it was proposed that TAF7 acts as a checkpoint regulator and delays transcription until the PIC complex is assembled (Gegonne et al., 2008, Gegonne et al., 2006). As both TAF1 and TAF7 are absent from the U2 snRNA gene, neither the enzymatic activities of TAF1 not the kinase interactions of TAF7 are required for expression of this gene. TAF1 may not be necessary for transcription of the U2 snRNA genes as the transcription unit is nucleosome depleted (Egloff et al., 2009).

TAF composition is at least partially core promoter-dependent (Green, 2000). The core promoter of genes is defined as the minimal DNA sequence required for accurate transcription initiation *in vitro* (Roeder, 1996). One of the common core promoter elements in human protein-coding genes is a TATA box located approximately 25 bp upstream of the transcription start site. The consensus sequence TATA(A/T)A(A/T)(A/G) is recognized directly by TBP (Hernandez, 1993). Other core promoter elements, such as the initiator (Inr), the downstream promoter element (DPE), and the downstream core element (DCE) may also be present (Kaufmann and Smale, 1994, Smale and Kadonaga, 2003, Lee et al., 2005, Purnell et al., 1994). TAFs are thought to significantly widen the range of promoter recognition by TFIID. Firstly, TAF association with TBP helps to overcome chromatin barriers (Thomas and Chiang, 2006). Secondly, contacts between TAFs and Inrs, DPEs and DCEs can facilitate promoter recognition (Basehoar et al., 2004). TAF1/TAF2 recognizes the Inr sequence (Chalkley and Verrijzer, 1999). In addition, UV

photo-cross-linking results demonstrate that TAF1 interacts with the DCE sub-element in a sequence-dependent manner (Lee et al., 2005). Thus, the absence of TAF1 and TAF2 may reflect the absence of an Inr sequence in addition to the lack of functional requirement for these TAFs. In contrast, TAF6 and TAF9, which are found on the U2 snRNA gene, exhibit DPE-binding specificity, (Shao et al., 2005, Sengupta et al., 2009) although there is no DPE in U2 snRNA promoter. TAFs may be differentially required for transcription of the snRNA and β-actin genes or differentially recruited due to differences in the DNA sequence, chromatin structure at the promoters and the complement of other factors at the promoter.

A number of reports demonstrate that some TAFs directly interact with activators (Allende-Vega et al., 2007, Bereczki et al., 2008). Those interactions suggest that TAFs can function as co-activators to bridge DNA-bound activators and the PIC. In agreement with this model, it has been demonstrated that mutations of Sp1 that disrupt interaction with TAFs affect activation of transcription (Gill et al., 1994). Thus, activators may target different TAFs to recruit TFIID to promoters. In addition, holo-TFIID may integrate different signals from different activators. Interestingly, VP16 and p53 activate transcription of protein-coding genes but not snRNA genes (Das et al., 1995), although TAF6 and TAF9 are present on U2 snRNA genes (Figure 2-2). In addition, Sp1 can activate transcription of snRNA genes (Das et al., 1995) although TAF4 and TAF7 are absent. This indicates that not all activators are compatible with snRNA gene promoters, that TAF6 and TAF9 are not sufficient to mediate activation by VP16 and p53 and that activation by Sp1 occurs in the absence of interaction with TAF4 and TAF7.

It is particularly striking that TAF4 and TAF10 are missing from the U2 snRNA gene promoter, as these play primary roles in TFIID assembly. Knockdown of Drosophila TAF4 leads to degradation of most of the TAFs (Wright et al., 2006) and TAF10 gene disruption in mice leads to disassembly of TFIID (Mohan et al., 2003). TAF5, which is associated with the U2 snRNA genes, is also implicated in TFIID structure. Thus, snTAFs must not require TAF4 and TAF10 for assembly and stability. TBP is recruited to snRNA genes by PTF - as an integral part of SNAPc (Henry et al., 1995). The snTAFc may be recruited as a new TBP-free complex, with a different composition from TFTC, STAGA, PCAF/GCN5 (see Table 3). Alternatively, snTAFc may not be stable off DNA and may assemble on the snRNA gene

promoter through interactions with PTF and TBP. Recognition of the snRNA gene-specific 3' box RNA 3' end processing element only occurs if transcription is initiated from a pol II-dependent snRNA gene promoter (Egloff et al., 2009, Hernandez, 1985), implicating specialized transcription complexes in the processing reaction. In addition, pol II initiated at the promoter of snRNA genes does not appear to make the transition to productive elongation and lack the H3K363m mark of productive elongation (Egloff et al., 2009). The selective recognition of the 3' box and the lack of processivity of pol II transcribing snRNA genes might therefore be, in some way, connected to the specific TAF composition of the PIC.

Transcripts from snRNA genes, in contrast to mRNA genes, are not polyadenylated. Dantonel and coworkers showed that cleavage-polyadenylation specificity factor CPSF, associates with the PIC through TFIID (Dantonel et al., 1997). Neither TAF7 nor TAF12, which interact strongly with CPSF-160, is recruited to the U2 snRNA gene promoter (Figure 2-7). The snRNA gene-specific TAF complex may therefore play a central role in facilitating gene-type-specific RNA processing.

Based on my results, I propose a model in which recruitment of a subset of TAFs to snRNA genes underpins a cascade of events critical for proper gene expression (Figure 5-1 C). PTF binding to the PSE facilitates stable recruitment of TBP, which in turn helps to recruit an snRNA gene-specific subset of TAFs (snTAFc) involved in specific recruitment of, for example, the Integrator complex. In addition, snTAFc may lack TAFs involved in recruiting elongation factors, thereby helping to determine the elongation potential of pol II.

The significance of the presence of the second peak of TBP and several TAFs within the gene but upstream of the poly(A) site of the β-actin gene is not clear. This may be diagnostic for a gene loop. However, loops are generally found between the promoter and the pol(A) site/region of transcription termination (Hampsey et al., 2011). In addition, the second peak of pol II does not coincide with the second peak of TBP and TAFs. Thus the second peak of pol II may reflect pausing at the end of the gene and the TBP/TAF peak may mark the position of a second promoter where an inactive PIC lacking TAF7 has assembled.

5.1.2 Future directions

It is well documented that there is a coupling between the snRNA gene promoter and the 3' box (Hernandez and Weiner, 1986, Dahlberg and Schenborn, 1988). Interestingly, it was demonstrated that human immunodeficiency virus (HIV) encoded elongation factor (Tat) together with the transactivation response element (TAR) can disturb this link (Ratnasabapathy et al., 1990). The productive elongation of transcription of HIV templates requires binding of the Tat to the TAR sequence at the start of HIV transcripts. Tat helps to recruit P-TEFb, which comprises Cdk9 kinase and Cyclin T, which overcomes the effect of negative elongation factors (Peterlin and Price, 2006). Interestingly, a number of studies have shown that TAFs are dispensable for Tat function and Cyclin T1 is implicated in directing assembly of a PIC containing TBP but no TAFs (Raha et al., 2005). P-TEFb functions in expression of the short intron-less snRNA genes only to activate recognition of the 3' box (Medlin et al., 2005). Recruitment of Tat to snRNA genes through a TAR sequence placed downstream of the site of initiation inhibits co-transcriptional recognition of the 3' box RNA processing element and may alter the elongation properties of pol II (Ratnasabapathy et al., 1990). The 3' box is recognized in transcripts from a construct with an snRNA promoter followed by the TAR sequence, 3' box and PA signal. However, when Tat is transfected, transcripts are polyadenylated and the 3' box is not recognized. The lenght of the transcription unit is altered from 1kb, as described for U2snRNA gene, to over 3 kb [*Kollakowski and Murphy, unpublished observations*]. As mentioned in Chapter 1 (Section 1.2), the CTD of pol II is essential for efficient 3' box-directed processing and Ser2 phosphorylation by P-TEFb is required to recruit Integrator, the snRNA gene-specific RNA processing complex (Medlin et al., 2005, Baillat et al., 2005, Egloff et al., 2007, Egloff et al., 2010). It is possible that TAR/Tat alters 3' box recognition by affecting the PIC complex on the snRNA gene promoter and/or modifying CTD phosphorylation by P-TEFb. In addition, it has been shown that TAR/Tat activates elongation of transcription from snRNA gene promoters in transient transfection experiments [*Kollakowski and Murphy, unpublished observations*] suggesting a link between the transcription elongation properties of pol II transcribing U2 snRNA genes and coupling between the snRNA promoter and 3' box. It would be of particular interest to answer the question of how events at the promoter can influence elongation

properties and how HIV Tat modifies the transcription elongation properties of the U2 snRNA gene to become similar to those of protein-coding genes. Future work could determine whether HIV-1 Tat affects the complement of TAFs within the U2 PIC. I have already generated stable cell lines (Flp-In T-Rex Core Kit for 293 cells) with either a marked U2 gene containing a fragment of globin gene sequence (U2G) or a U2 gene containing the HIV TAR sequence just downstream of the transcription start site (U2TAR). The effect of HIV-1 TAR/Tat on the composition of the PIC in stable cell lines could be analysed by ChIP with and without ectopic expression of Tat. It would also be interesting to analyse the recruitment of Integrator and transcription elongation factors to these constructs in the presence and absence of Tat.

5.2 Role of RPAP2 in transcription of snRNA genes

5.2.1 Recruitment of RPAP2 to snRNA genes through Ser7 phosphorylation underpins a cascade of events critical for proper snRNA gene expression

The first insights to RPAP2 function came from a systematic analysis that determined the composition and organization of the soluble pol II machinery (Jeronimo et al., 2007). Interestingly, the RPAP2 protein was found to interact with Integrator in addition to pol II, which raised the possibility that it plays a role in transcription of snRNA genes. In addition, Rtr1, the yeast homolog of RPAP2, was identified as a Ser5P phosphatase (Gibney et al., 2008). My results demonstrate that RPAP2 specifically and directly interacts with the Ser7P form of the pol II CTD. RPAP2 knockdown causes loss of Integrator from snRNA genes, reduces the rate of transcription and impairs 3' end processing. Furthermore, my results demonstrate that RPAP2 regulates the global level of Ser5P in the cell and dephosphorylates Ser5P on actively transcribing pol II. Based on these results, I conclude that recruitment of RPAP2 to snRNA genes through Ser7P results both in stable association of Integrator with snRNA genes and removal of Ser5P (Figure 5-2) (Egloff et al., 2012b).

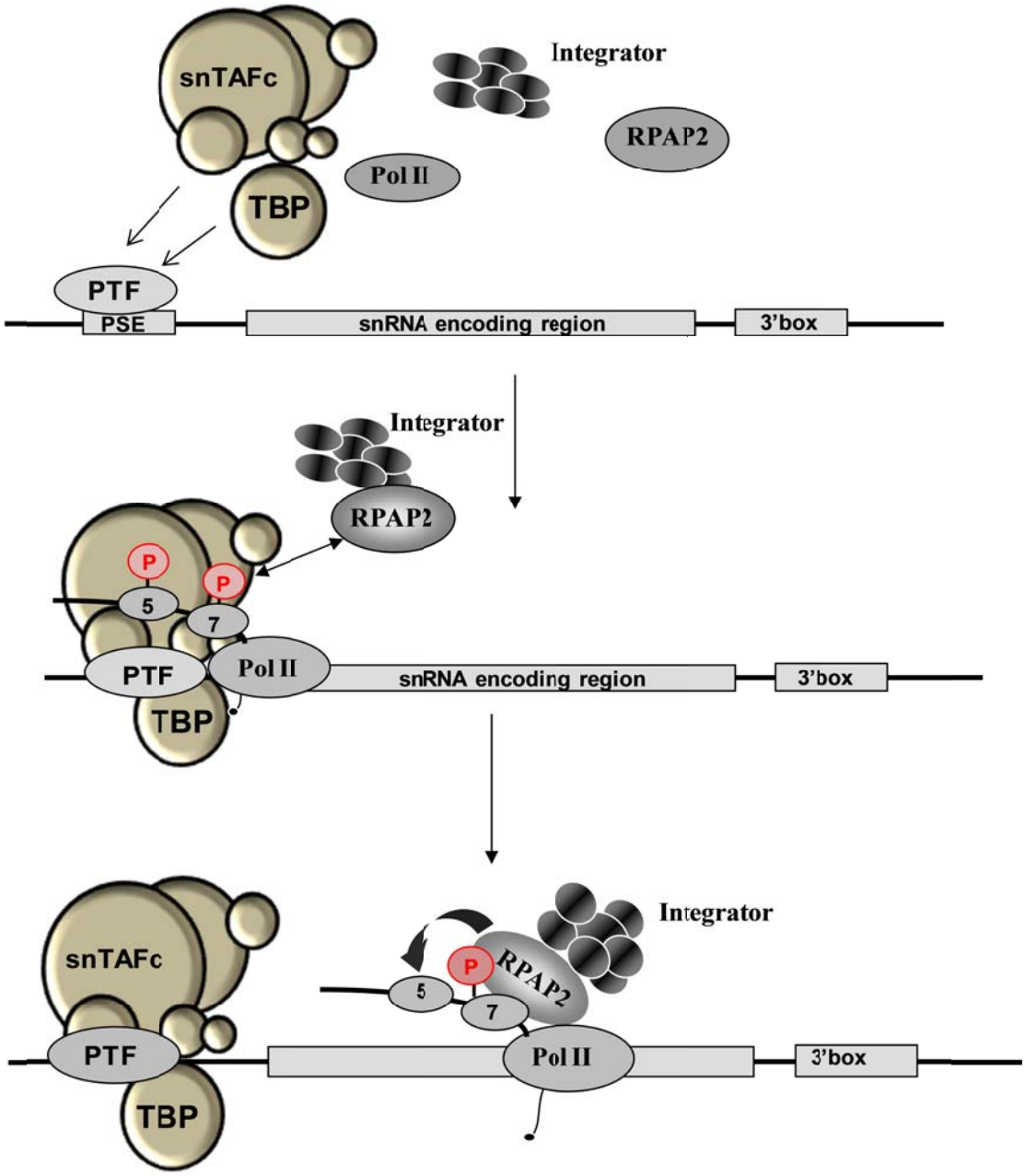


Figure 5-2. Recruitment of RPAP2 to snRNA genes through Ser7 phosphorylation underpins a cascade of events critical for proper snRNA gene expression.
On snRNA genes, TBP associates with PTF and with snTAFc. The pol II CTD is first phosphorylated on Ser5 and then on Ser7. Subsequently, RPAP2 is recruited to snRNA genes and tethers Integrator to snRNA genes. RPAP2 dephosphorylates Ser5P of the pol II CTD facilitating transcription.

Interestingly, the defect in transcription of snRNA genes caused by RPAP2 knockdown may occur due to the failure to dephosphorylate Ser5P. Ser5 hypophosphorylation has been proposed to reduce the rate of transcription reinitiation by pol II or result in persistent association of the capping enzyme (Reyes-Reyes and Hampsey, 2007). In addition, in yeast, knockdown of Rtr1 and Ssu72 causes a transcription defect (Reyes-Reyes and Hampsey, 2007, Gibney et al., 2008, Mosley et al., 2009). My results demonstrate that Ssu72 is recruited to U2 snRNA gene and β-actin protein-coding gene. Furthermore, Ssu72 knockdown causes pol II to accumulate at the 3' end of both U2 snRNA and β-actin genes, which is indicative of termination defects. However, after Ssu72 siRNA-mediated depletion there is no 3'end processing defect on snRNA genes. This would suggest that Ssu72 functions primarily as a transcription terminator rather than 3'end processing factor for snRNA genes *[O'Reilly et al., NAR; in press]*.

5.2.2 Future directions

Interestingly, the recruitment of RPAP2 to a protein-coding gene is not affected by mutation of Ser7. It would be of interest to decipher a mechanism by which RPAP2 is recruited to protein-coding genes and perform ChIP-seq to determine the occupancy of RPAP2 on mRNA genes. It is worth noting that there is some evidence that argues against the phosphatase activity of RPAP2. According to a recent paper, there is a marked absence of an active site in both His-tagged and GST-tagged *Kluyveromyces lactis* Rtr1 (Xiang et al., 2012). Given these conflicting data, it is important to determine whether human RPAP2 is truly the Ser5P phosphatase. The next step towards resolving the uncertainty surrounding the phosphatase activity of RPAP2 is to solve the crystal structure of human RPAP2 in the presence of the CTD.

5.3 Inhibition of Ser2 phosphorylation by the HSV-1 ICP22 protein

5.3.1 ICP22 specifically inhibits Ser2 phosphorylation on transcriptionally-active pol II transcribed genes

It has been shown that expression of ICP22 causes the loss of phosphorylation of Ser2 (Fraser and Rice, 2005), a mark associated with productive elongation (Egloff and Murphy, 2008a). PTEFb is thought to be a major CTD Ser2 kinase (Egloff and Murphy, 2008a) and it has been suggested that ICP22 interacts directly with P-TEFb to inhibit the kinase activity (Durand et al., 2005). Acyclovir and related nucleoside analogue anti-Herpes drugs are effective and widely available. However, resistance to these drugs is becoming a major public health problem (Galdiero et al., 2013). In addition, current antiviral strategies for HSV affect only lytic infection (Knipe and Cliffe, 2008). There is therefore a pressing need for the identification of potential new Herpesvirus drug targets to facilitate the development of the next generation of anti-Herpesvirus drug. Here, I have initiated a study of the HSV-1 ICP22 protein, which is a determinant of infectivity and latency (Poffenberger et al., 1994, Orlando et al., 2006) and could be a good future drug target. As there are close homologues to the HSV-1 ICP22 in all alpha-Herpesvirus, including HSV-2 and Varicella zoster, any drugs that target HSV-1 ICP22 may also be effective against these other Herpesviruses. My studies were focused on the mechanism of action of ICP22 inhibition of CTD phosphorylation. I have determined that ICP22 with a Myc epitope tag is well-expressed in HeLa cells and specifically causes loss of phosphorylation of Ser2 of the CTD without markedly affecting phosphorylation of Ser5 and Ser7. In addition, I have shown that the region between amino acids 193 to 256 is sufficient to cause loss of Ser2 phosphorylation. The specific loss of Ser2 phosphorylation recapitulates the effect of ICP22 expression in other systems. My ChIP studies on a *PLK2* gene suggests that transcription and CTD Ser2 phosphorylation on transcribing pol II are drastically affected by expression of either full-length ICP22 or the truncated form comprising amino acids 193 to 256 (clone 4). My current working model is that ICP22 interacts directly with either the Cdk9 or cyclin component of P-TEFb to inhibit the kinase activity. The ICP22 region involved in the binding of P-TEFb would be within residues 193-

256. However, the pol II occupancy on U2 snRNA genes is not affected by expression of ICP22. This is in line with previously published results that inhibition of Cdk9 does not inhibit transcription of snRNA genes but leads to incorrect RNA 3'end formation (Medlin et al., 2003). The Ser2P mark is not required for the expression of the p53-regulated p21 gene, histones genes and snRNA genes. Thus, this mark is not universally needed for RNA synthesis (Tietjen et al., 2010, Gomes et al., 2006, Medlin et al., 2005). Furthermore, my results suggest that Ser2P levels on U2 snRNA genes do not drop as drastically as on the *PLK2* gene. This may suggest either that P-TEFb is not the only acting Ser2P kinase on snRNA genes or that inhibition of Ser2 phosphorylation is not the result of ICP22 simply inhibiting the activity of CDK9. For example, ICP22 may not inhibit all the functions of P-TEFb, as P-TEFb-dependant RNA processing of snRNA gene transcripts is not affected. Alternatively, ICP22 can only access P-TEFb to effectively inhibit the kinase activity if P-TEFb association with the genes is not too stable to block access or ICP22 selectively inhibits P-TEFb function on some genes through sequestration, perhaps to another part of the nucleus. In these latter cases, only genes where P-TEFb does not interact very stably with the transcription complex would be affected.

5.3.2 Future directions

There is much more to be discovered about ICP22. Identification of the host cell factors targeted by ICP22 is critical to understanding the molecular mechanism by which ICP22 affects Ser2P. Therefore, the next step would be to perform interaction studies with the full-length ICP22 and the shorter form that triggers the loss of Ser2 phosphorylation to identify interaction partners, using a proteomics approach. For example, epitope-tagged versions of the full-length ICP22 and the short functional conserved region could be used to pull down interacting host proteins, followed by proteomic identification. Further, to test whether CDK9 or one of its cyclin partners directly interacts with ICP22, interaction between recombinant GST-tagged forms of the short truncated form of ICP22 and full-length ICP22 with recombinant P-TEFb should be tested. It would also be informative to analyse the effect of ICP22 on the expression of host genes genome-wide. The results of these experiments will underpin the development of a judicious strategy to determine the molecular mechanism of ICP22 action. In addition, these preliminary analyses will

facilitate the elucidation of the structural detail of the molecular interaction between ICP22 and host partner proteins using, for example, crystallography, which will, in turn, inform drug-screening strategies.

Recent work indicates that not only Cdk9, but also Cdk12 mediates Ser2 phosphorylation of pol II CTD (Bartkowiak et al., 2010). Future studies should address whether Cdk12 function is affected by ICP22.

The proposed research will help elucidate the molecular mechanisms underlying viral strategies of infection and propagation and allow the judicious design of therapeutics. As ICP22 appears to act at the level of transcription, this research will also shed light on the control of expression of human genes at this level. As such, the results of this research will be of interest to scientists working in the field of regulation of gene expression, in addition to virologists and clinicians.

NOVELTY

My work has highlighted important factors and regulatory mechanisms specific to snRNA gene transcription. The demonstration that the promoter of the U2 snRNA gene is recognized by a specific set of TBP-associated factors and the role of RPAP2 and Ssu72 in snRNA gene expression is both novel and significant. Parts of the thesis were based on published manuscripts (Zaborowska et al., 2012, Egloff et al., 2012b) and [*O'Reilly et al. NAR in press*].

My follow up of my initial study of the mechanism of action of ICP22 may reveal a novel mechanism by which ICP22 can manipulate host cell gene expression.

Chapter 6| Materials and Methods

6.1 Materials

6.1.1 Antibodies

Antibodies were stored at 4^{o}C or at -20^{o}C

Antibodies	**Company**
Anti-RNA Polymerase II (sc-N20; rabbit polyclonal)	Santa Cruz
Anti-RNA Polymerase II (sc-H224; rabbit polyclonal)	Santa Cruz
Anti-TBP (sc-421; mouse monoclonal)	Santa Cruz
Anti-TAF1 (sc-735; mouse monoclonal)	Santa Cruz
Anti-TAF2 (NBP1-21371; rabbit polyclonal)	Novus Biologicals
Anti-TAF3 (A302-360A; rabbit polyclonal)	Bethyl laboratories
Anti-TAF4 (sc-736; mouse monoclonal)	Santa Cruz
Anti-TAF7 (sc-101167; mouse monoclonal)	Santa Cruz
Anti-TAF10 (sc-102125; rabbit polyclonal)	Santa Cruz
Anti-TAF5, anti-TAF6, anti-TAF8, anti-TAF9, anti-TAF11, anti-TAF12, anti-TAF13 (all rabbit polyclonal)	Kind gift from Robert Roeder (Raha et al., 2005)
Anti-CPSF160 (sc-28872; rabbit polyclonal)	Santa Cruz
Anti-RNA polymerase CTD phosphorylated on Serine 2 (3E10), 5 (3EB) and 7 (4E12)	Kind gift from Dirk Eick (Chapman et al., 2007)

(all rat monoclonal)	
Anti-actin (sc-1615; goat polyclonal)	Santa Cruz
Anti-RPAP2 (17401-1-AP; rabbit polyclonal)	Proteintech
Anti-Myc epitope (sc-764X; rabbit polyclonal)	Santa Cruz
Anti-Myc epitope (ab1932; goat polyclonal)	Abcam
Goat anti - mouse HRP (1858413)	Pierce
Donkey anti-goat HRP (sc-2020)	Santa Cruz
Goat anti-rabbit HRP (12-348)	Millipore

6.1.2 Antibiotics

- Ampicillin (100 μg/ml; Sigma Aldrich)

6.1.3 Buffers and solutions

- Phosphate Buffered Saline (PBS) - 140 mM NaCl, 3.4 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4

- Tris Buffered Saline - 150 mM NaCl, 40 mM Tris-HCl - pH7.4

- TBE Buffer - 10 mM Tris, 250 mM orthoborate, 2 mM EDTA

- Cell lysis buffer – 5 mM PIPES pH8.0, 85 mM KCl, 0.5% NP40

- Nuclear lysis buffer – 50 mM Tris-HCl pH 8.1, 10 mM EDTA, 1% SDS

- Dialysis Buffer – 2 mM EDTA, 50 mM Tris-HCl pH 8.0, 0.2 % Sarkosyl

- Dilution buffer – 0.01% SDS, 1.1 % Triton, 1.2 mM EDTA, 16.7 mM Tris-HCl pH 8.1, 167 mM NaCl

- Wash buffer 1 (Rabbit polyclonal antibodies) – 100 mM Tris-HCl pH 9.0, 500 mM LiCl, 1% NP40, 1% Deoxycholic acid

- Wash buffer 2 (Rat or mouse monoclonal antibodies) – 100 mM Tris-HCl pH 8.0, 250 mM LiCl, 0.5% NP40, 0.5% Deoxycholic acid

- Elution buffer – 50 mM NHCO3, 1% SDS

- Proteinase K Buffer (5×) - 50 mM Tris-HCl pH 7.5, 25 mM EDTA, 1.2%SDS

- EB buffer – 10 mM Tris-HCl, pH8.1

- NETN buffer – 0,5% NP-40, 20mM Tris-HCl pH8, 100mMNaCl, 1mM EDTA

- HEGN buffer – 20mM HEPES pH7,6, 10% glycerol, 0,2% NP-40, 0,1mM EDTA, 1mM DTT

- SDS running buffer (25mM Tris, 250mM glycine, 1% SDS)

6.1.4 Reagents

Chemical reagents were purchased from VWR International or Sigma, apart from the reagents listed below:

- Trizol (Invitrogen)

- RNase-DNase-free water (Qiagen)

- LipofectAMINE 2000 (Invitrogen)

- Dimethylsulfoxide (DMSO; Merck)

- Precision plus protein Dual Colour Standards (Bio-RAD)

- Dithiothreitol (DTT; Roche)

ChIP materials:

- BSA (10 mg/ml; New England Biolabs)

- Herring Sperm DNA (10 mg/ml; Sigma)

- Formaldehyde (Sigma)

- Protease inhibitor cocktail (Roche)

- Staphlococcus aureus cell membranes (Calbiochem)

Kits and enzymes:

- Plasmid mini and maxi kits (Qiagen)

- Qiaquick PCR purification kit (Qiagen)

- QIAGEN Plasmid Mini Kit (Qiagen)

- QIAGEN Plasmid Maxi Kit (Qiagen)

- QuantiTect SYBR Green PCR master mix (Qiagen)

- Superscript III for cDNA synthesis (Invitrogen)

- DNase 1 (Promega)

- Proteinase K (Promega)

PCR Oligonucleotides:

All oligonucleotides used were synthesized by Sigma Aldrich. ChIP primers were designed to give a PCR product of approximately 100 bp in length. Sequences of primers used in all assays are listed 5' to 3' in the following pages.

Primers used in Chapter 2:

β-actin 1 (F)	GCTGCGGCTGGGTAGGTTTG
β-actin 1 (R)	CACTTAGAAGTCGCAGGACC
β-actin 2 (F)	CCAATCAGCGTGCGCCGTTCCGA

β-actin 2 (R)	GGTGTGGACGGGCGGCGGATC
β-actin 3 (F)	GGGCAACCGGCGGGGTCTTT
β-actin 3 (R)	ACGCAGTTAGCGCCCAAAGG
β-actin 4 (F)	CACAGCGCGCCCGGCTATTC
β-actin 4 (R)	AGCCAGCTCCCCTACCTGGT
β-actin 5 (F)	CCCCATCGAGCACGGCATCGTC
β-actin 5 (R)	CACCTGGGTCATCTTCTCGCGGT
β-actin 6 (F)	CATTGCTCCTCCTGAGCGCAAGTA
β-actin 6 (R)	TTGCGGTGGACGATGGAGGGGCC
β-actin 7 (F)	GGGACTATTTGGGGGTGTCT
β-actin 7 (R)	TCCCATAGGTGAAGGCAAAG
β-actin 8 (F)	TGGGCCACTTAATCATTCAAC
β-actin 8 (R)	CCTCACTTCCAGACTGACAGC
U2 1 (F)	GGAGCGGAGCGTTCTCTGTCTCCCC
U2 1 (R)	AGAGTGTGAGCCCTCATTCACGCCC
U2 2 (F)	ATGAGAGTGGGACGGTGA
U2 2 (F)	CACTTGATCTTAGCCAAAAGG
U2 3 (F)	ACGAGTCCTGTGACGCGCCGGCTTG
U2 3 (R)	CTCCGGGTGGGTCCCATTCCTTTAA
U2 4 (F)	CCTCCCCGCCTCTCCCTCGCTC
U2 4 (R)	GGACAAATAGCCAACGCATGCGG

Primers used in Chapter 3:

β-actin 1 (F)	GCTGCGGCTGGGTAGGTTTG
β-actin 1 (R)	CACTTAGAAGTCGCAGGACC
β-actin 2 (F)	CCAATCAGCGTGCGCCGTTCCGA
β-actin 2 (R)	GGTGTGGACGGGCGGCGGATC
β-actin 3 (F)	GGGCAACCGGCGGGGTCTTT
β-actin 3 (R)	ACGCAGTTAGCGCCCAAAGG
β-actin 4 (F)	CACAGCGCGCCCGGCTATTC

β-actin 4 (R)	AGCCAGCTCCCCTACCTGGT
β-actin 5 (F)	TTACCCAGAGTGCAGGTGTG
β-actin 5 (R)	CCCCAATAAGCAGGAACAGA
U2 1 (F)	GGAGCGGAGCGTTCTCTGTCTCCCC
U2 1 (R)	AGAGTGTGAGCCCTCATTCACGCCC
U2 2 (F)	ATGAGAGTGGGACGGTGA
U2 2 (F)	CACTTGATCTTAGCCAAAAGG
U2 3 (F)	ACGAGTCCTGTGACGCGCCGGCTTG
U2 3 (R)	CTCCGGGTGGGTCCCATTCCTTTAA
U2 4 (F)	GGAGTTCTTCTTTCTCCTGTCTC
U2 4 (R)	AGACAGACAGAGAGACAGAGACAGA
U2 5 (F)	CCTCCCCGCCTCTCCCTCGCTC
U2 5 (R)	GGACAAATAGCCAACGCATGCGG
U2 6 (F)	TGATATAAACAACGGGGGCGCGATG
U2 6 (R)	GCTGTGTAGGGAAGCTGCCTCTCTG
U2 7 (F)	CGGTGTTTGGTTTTCGCTTCCTGTG
U2 7 (R)	GGACACGTATGTTTACTGCAGCAC

Primers used in Chapter 4:

PLK2 1 (F)	AAGTGTCTCCTCTGTACCAGGA
PLK2 1 (R)	GGAATCATGACCAGGAAATGTACGG
PLK2 2 (F)	GGAGCCTTCGAAGAACGGGACA
PLK2 2 (R)	AAGCAAGTCTTGTCCCGCCA
PLK2 3 (F)	ACCGGGGTGTTGGGTGCTAGT
PLK2 3 (R)	ATAGTCCGCAAAAGCTCCATG
PLK2 (F)	CACAGCGCGCCCGGCTATTC
PLK2 (R)	AGCCAGCTCCCCTACCTGGT
U2 1 (F)	GGAGCGGAGCGTTCTCTGTCTCCCC
U2 1 (R)	AGAGTGTGAGCCCTCATTCACGCCC
U2 2 (F)	ATGAGAGTGGGACGGTGA

U2 2 (F)	CACTTGATCTTAGCCAAAAGG
U2 3 (F)	ACGAGTCCTGTGACGCGCCGGCTTG
U2 3 (R)	CTCCGGGTGGGTCCCATTCCTTTAA
U2 4 (F)	CCTCCCCGCCTCTCCCTCGCTC
U2 4 (R)	GGACAAATAGCCAACGCATGCGG

6.1.5 Bacteria and cell lines

- Escherichia coli NM554 competent cells were used to produce plasmid DNA.

- HeLa human cervical cancer adherent cells were used for ChIP, transient transfection and knockdown experiments.

- 293 human embryonic kidney cells were used for complementation assays.

6.1.6 Bacterial and tissue culture media

- NM554 cells were cultured in 2-YT media at 37^{O}C, shaking overnight.

- HeLa cells were grown in complete media consisting of Dulbecco's modified Eagle's media, supplemented with 2 mM L-glutamine (Invitrogen) and 10% heat inactivated foetal bovine serum (Invitrogen).

6.1.7 Vectors

- pGEX-6P-1 from GE Healthcare was used in all cloning experiments to generate GST-RPAP2 and GST-ICP22 (vectors generated by Dr Shona Murphy).

- pcDNA3.1 from Invitrogen was used in cloning experiments to generate Myc-RPAP2 and Myc-ICP22 (vectors generated by Dr Shona Murphy).

6.2 Methods

DNA related methods

6.2.1 Polymerase Chain Reaction (PCR)

DNA products for cloning experiments were amplified by the polymerase chain reaction (PCR). For all PCR assays, a Taq-containing master mix (2×) was used (HotStarTaq Mastermix, QIAGEN), which contains a heat activated Taq enzyme, dNTPs and $MgCl_2$ containing reaction buffer. Typical PCR reactions contained corresponding amount of DNA (100 ng of genomic DNA or 10 ng of plasmid DNA) plus 10 pmoles of each primer in a total volume of 15 µl. For all cloning experiments, 2U of a proofreading enzyme, typically PFU Turbo, was added to each reaction to minimize the error rate. Thermal cycling was carried out as follows: 95^{o}C for 15 min, 25-30 cycles at 94^{o}C for 15 sec, 57^{o}C for 15 sec and 72^{o}C for 20-60 sec, followed by 72^{o}C for 7 min. The products were analysed by electrophoresis on 1-2% agarose gels and phenol/chloroform purified before use in the digest reaction.

6.2.2 Real-time quantitative Polymerase Chain Reaction (qPCR)

Real time PCR (qPCR) was used to measure the accumulation of PCR product during the exponential phase and to quantify the gene transcript levels *in vivo*. A Rotorgene 6000 thermal cycler (Corbett) was used for qPCR using a 2x master mix similar to that use for standard PCR reactions with the addition of SYBR green. Reaction volumes of 10 µl consisted of 5 µl of 2 x master mix, 3 µl ddH2O, 10 picomoles of each primer and 1µl of ChIP DNA as template DNA (usually 1/30th of ChIP DNA). Thermal cycling included: 95^{o}C for 15min, followed by 40 cycles at 94^{o}C for 15sec, 57^{o}C for 15 sec and 72^{o}C for 30 sec. Since SYBR green binds all double stranded (ds) DNA molecules, including artefacts such as primer dimers or non-specific products that may arise from cross hybridization of primer sequences, all amplifications were followed by a 10 minutes melt cycle to confirm that a single PCR product was generated in the reaction. This involves ramping the temperature from 50^{o}C to 99^{o}C by 1^{o}C every 5 seconds. Each amplicon denaturates at a specific temperature and thus a single amplicon should yield a clear peak when the melt

curve is plotted at the end of the run. For ChIP analysis, a standard curve was produced for each primer pair using 1/5 step-wise dilutions of Input DNA sample. Signals from all other samples were quantitated relative to this, resulting in a “percentage input DNA” value for each sample. To adjust for background, the no antibody “percentage input DNA” was subtracted to give the final value. At the end of the run a threshold value is set for all samples and standards over the linear range of the PCR reaction. The Rotorgene software converts the cycle number (Ct) for each standard to copies/µl and generates a standard curve. The Ct values for each sample are quantitated relative to this standard. To allow normalisation of β-actin, U2snRNA and PLK2 transcript levels in HeLa, the concentration (in copies/µl) of the specific amplicon were divided by the total input DNA for each immunoprecipitation, giving the final value plotted on the graphs shown in this thesis.

6.2.3 DNA phenol/chloroform extraction and EtOH precipitation

1/10 3M sodium acetate pH5.2 and equal volume of phenol/chloroform pH 8.0 was added to the samples followed by mixing and centrifugation at 12000 x g for 5 min. Supernatant was recovered and mixed with 2.5x starting volume of EtOH followed by centrifugation at 12000 x g for 5 min. Pellets were resuspended with appropriate volume of water.

6.2.4 Agarose gel electrophoresis

Agarose gel electrophoresis was used to size PCR products used in cloning experiments and sonicated DNA when optimizing ChIP. The DNA samples were mixed with 0.16 volumes of bromophenol blue-containing DNA loading dye and loaded on 1 – 2% agarose gels containing 0.002% Ethidium Bromide (10 mg/ml stock). Electrophoresis was carried out at 100 mA in 1× TBE buffer, also containing 0.002% Ethidium Bromide. DNA bands were visualized under UV light. The 100 bp or 1kB ladder was used a sizing marker.

6.2.5 DNA sequencing

DNA sequencing was carried out to confirm that all constructs received and clones generated were as expected. The sequencing reaction was carried out using mini-prep or maxi-prep (Qiagen) purified plasmid along with the appropriate primer for the clone in question. All sequencing reactions were sourced out to Source BioScience UK limited, Department of Biochemistry, University of Oxford, UK.

6.2.6 Restriction enzyme digestion

The corresponding amount (500 ng – 1 µg) of plasmid DNA or PCR product was used in all enzyme digestion reactions. The required amount of plasmid DNA or PCr product was incubated with the appropriate digestion buffer (1×) and 5 – 10 U of restriction enzyme(s) (never exceeded 10% the volume of the reaction mix). Digestions were incubated at 37^{o}C up to 1 h for plasmid DNA and up to 3 h for PCR products.

6.2.7 DNA ligation

DNA fragments with blunt or overlapping ends were incubated with 1 unit (U) of T4 DNA ligase in 1× ligation buffer for 5 min at RT. Typically, the ratio of insert to vector was 50:10 ng and this were incubated for in a total reaction volume of 10 µl.

6.2.8 Plasmid construction

To make Myc-RPAP2, three tandem Myc epitopes were cloned at the C terminus of the RPAP2-coding region in pcDNA3.1. The level of Myc-RPAP2 relative to endogenous RPAP2 was determined by Western blot analysis (Figure 3-2). To make GST-RPAP2, the coding region of RPAP2 was cloned into pGEX-6P-1 (Egloff et al., 2012b). To make Myc-ICP22, three tandem Myc epitopes were cloned at the C terminus of the ICP22-coding region in pcDNA3.1. To make GST-ICP22, the coding region of ICP22 was cloned into pGEX-6P-1. All constructs were generated by Dr Shona Murphy.

6.2.9 Plasmid DNA purification

Qiagen miniprep and maxiprep kits were used to purify plasmid DNA from 200 µl to 100 ml transformed bacterial cultures, respectively.

Mini-prep

For miniprep analysis, typically 200 µl of 2xYT media, containing the appropriate antibiotic, was inoculated with a single colony and incubated for 6 h at 37^{o}C with vigorous shaking. Plasmid DNA was purified using QIAquick® Miniprep Kit (Qiagen) according to the instructions from manufacturer. In short: cells were lysed by alkaline lysis. The Lysate was then neutralised and adjusted to high-salt binding conditions. The plasmid DNA was adsorbed onto a silica-gel membrane in a spin column, by centrifugation (17,000 x g, 1 min, RT). Contaminations were removed by washing the membrane in a solution containing 80% ethanol. The purified plasmid was then eluted in dH2O. Miniprep DNA preparations were generally used in sequencing reactions to confirm the expected sequence of clones generated in this thesis.

Maxi-prep

For maxi-prep analysis, a 100 ml of 2YT media was inoculated with 100 µl of a 5 ml overnight culture. Plasmid DNA was purified using Plasmid Maxi Kit (Qiagen) according to the manufacturer's protocol. In short: The cells were lysed by alkaline lysis. The Lysate was then neutralised and adjusted to high-salt binding conditions. The plasmid DNA was adsorbed onto a pre-equilibrated silica-gel membrane in a Qiagen Tip-500 column. Cellular contaminants were removed by washing the membrane in a solution containing 80% ethanol. The purified plasmid was eluted in low salt buffer and precipitated with isopropanol. The pellet was resuspended in dH2O. Maxiprep DNA was always re-sequenced to control for human error. This DNA was used in downstream application, including ChIP and transfection assays.

RNA-related methods

6.2.10 Purification of total RNA

Total RNA was isolated from HeLa cells using Trizol reagent according to manufacturer's instructions. Briefly 1 ml of Trizol reagent (GIBCO) was added directly to cells per well of a 6-well plate, resuspended thoroughly and the 1 ml lysates were transferred to eppendorfs (with the aid of cell scrapper). The samples were processed immediately or stored at -80^{o}C. Lysed cells in Trizol reagent were set at RT for 5 min and 0.2 ml of chloroform was added to each sample. Following a vigorous shake for 15 sec at RT samples were centrifuged at 4^{o}C, 12,000 x g for 15 min. In order to precipitate RNA, the aqueous supernatant was carefully transferred into a new Eppendorf tube and the total RNA precipitated in an equal volume of isopropanol for 10 min at RT. Following centrifugation at 4^{o}C, 12,000 x g for 10 min, the supernatant was removed. The pellet was washed in 1 ml of 75% Ethanol (EtOH), centrifuged at 7,500 x g for 5 min at 4^{o}C and air-dried at RT for 5 min. The pellet was resuspended in 20 µl of RNase-DNase-free water. Following this, the RNA sample was DNase treated. Samples were brought to 90µl, 10µl of DNase buffer (Roche) and 1µl of DNase enzyme (Roche) were added, followed incubation at 37^{o}C for 30 minutes. Following this, the equal volume of phenol-chloroform was added to the samples. Samples were mixed, incubated at RT for 5 minutes and centrifuged at 4^{o}C at 12,000 x g for 15 minutes. The supernatant was then mixed with 2.5 x starting volume of 100% EtOH and 1/10 of 3M sodium acetate (NaOAc) ph5.5, incubate don ice for 15 minutes and centrifuges at 4^{o}C at 12,000 x g for 15 minutes. The pellet was washed with 500 µl 75% EtOH, centrifuges at 4^{o}C at 7,500 x g for 5 minutes and resuspended in 20µl of RNase-free water. RNA was quantitated using a Nanodrop spectrophotometer (Thermo scientific).

6.2.11 RNAi-mediated knockdown

RNAi duplexes were made by Dharmacon. HeLa cells were plated into 24-well Costar plates in complete DMEM media. The next day, cells were transfected when 60-70% confluent. 2 µl of siRNA and 1 µl LipofectAMINE2000 were diluted in 50 µl of Opti-MEM 1 (Invitrogen) reduced serum medium and incubated separately for 5 min at RT. siRNA and

lipofectAMINE2000 were combined and incubated for a further 25 minutes at RT. 100 ml of complexes was added to adherent HeLa cells. Media was changed after 6 hours. Cells (60-70% confluent) were split into 6 well plates. Transfection was repeated as above except with 10µl siRNA in 250 µl Opti-MEM and 5µl of LipofectAMINE2000 in 250µl Opti-MEM.

6.2.12 RNase protection assay

RNase protection assays was carried out by Dr Shona Murphy as described in (Cuello et al., 1999). Briefly: 10 to 20 µg of RNA in 30 µl of R-loop buffer (50 mM PIPES, pH 7.4, 1 mM NaEDTA, 500 mM NaCl) were incubated with 1µl of riboprobe (prepared by Dr Shona Murphy) for 10 min at 80^{o}C and hybridized at 56°C overnight. After adding 300µl of RNase protection solution (300 mM NaCl, 10 mM Tris - HCL pH7.4, 5 mM EDTA, 40 µg RNaseA and 1100 U RNaseT1) the samples were incubated at 18^{o}C for 2 h. After adding of 30 µl of Proteinase K mix (1%SDS, 3.5 mg/ml of protease K) the mixture was than incubated for 30 min at 37^{o}C, phenol / chloroform purified, ethanol precipitated and analyzed on a 6% polyacrylamide-urea gel.

Protein-related methods

6.2.13 Cell Culture

Cell culture was performed in a category 2 laminar flow hood. HeLa cells were grown at 37^{o}C in 5% CO2 in sterile medium Dulbecco's Modified Eagle Medium (DMEM) media supplemented with 10% heat-inactivated foetal bovine serum, 2mM L-glutamine and 1x penicilin. Typically, adherent HeLa cells in 75 cm2 flasks were lifted off by incubation with 1 ml of PBS plus 1 mM EDTA for 5 min. This was followed by gentle tapping of the side of the flask. Cells were resuspended in 10 ml of complete DMEM media and passaged into new flasks at a 1:10 dilution.

6.2.14 Freezing

Cells stored at -70^{o}C were resuspended in 1 ml of complete DMEM media, supplemented with 10% foetal bovine serum and 10% DMSO, per 75 cm^2 flask. Final volumes were transferred into a cryotube and placed in a precooled Cryobomb (NALGENE) filled with isopropanol in a -80^{o}C freezer.

6.2.15 Transfection into HeLa cells

Transfections were carried out using Lipofectamine 2000 according to manufacturer's instructions (Invitrogen). The day before transfection, HeLa cells were plated in complete DMEM media, in 6-well Costar dishes at a density of 2 × 10^5 cells per well. At the day of transfection cells were at 80-90% confluence. Each DNA to be transfected was diluted in 50 µl of Opti-MEM (GIBCO) reduced serum medium and incubated for 5 min at RT. LipofectAMINE 2000 was also diluted in Opti-MEM reduced serum medium and incubated for 5 min at RT. Typically 2 µl of Lipofectamine 2000 was used per µg of DNA in 50 µl of Opti-MEM 1 reduced serum medium. The DNA and lipofectamine 2000 were combined and incubated for a further 30 min at RT. 100 ml of complexes was added, by slow dispensing, to the corresponding well. Culture media was changed after 6 h and cells were harvested 24h after.

6.2.16 ChIP analysis

Chromatin immunoprecipitations were carried out as described by Farnham et al. (2000). In short: HeLa cells were grown in 10 mm dishes in 10 ml of DMEM complete media. 1% Formaldehyde was added to fix proteins to nucleic acids. The reaction was stopped by adding glycine to 125 mM. After washing with PBS, the cells were lysed by incubation on ice for 20 min in cell lysis buffer (5 mM PIPES pH 8, 85 mM KCl, 0.5% NP40, protease inhibitor) and the cytoplasmic fraction was separated from the nucleus by spinning at 5000 x g for 5 min. The nuclei were then lysed using nuclear lysis buffer (50 mM Tris-Cl pH 8.1, 10 mM EDTA, 1% SDS, protease inhibitor) and the chromatin was fragmented by 10 sonications for 10 sec at 10 amp using a Stratagen sonicater. Staph A cells were pre-blocked with 1 mg/ml herring sperm DNA and 1 mg/ml BSA overnight at 4^{o}C. After pre-clearing of the sample by incubation with StaphA cells for 15 min at 4^{o}C the sample was

split into four aliquots and IP dilution buffer (0.01% SDS, 1.1% Triton 100, 1.3 mM EDTA, 16.7 mM Tris-Cl pH 8.1, 167 mM NaCl) was added. No-antibody controls were also set up to control for nonspecific binding. Each aliquot was incubated with 1 µg of required antibody O/N at 4^{o}C rotating. The next day StaphA cells were added to the sample and incubated for 15 min at 4^{o}C rotating. Staph A-complexes were pelleted by centrifugation at 4^{o}C for 3 minutes, and the supernatants discarded with the exception of the "no antibody" supernatant. 10% of this was kept as the "input DNA" sample. After washing 2 times with ChIP dialysis buffer (2 mM EDTA, 50 mM Tris-Cl pH 8) and 4 times with ChIP wash buffer (100 mM Tris-Cl pH 9, 500 mM LiCl, 1% NP40, 1% deoxycholic acid) the DNA-antibody complex were eluted with elution buffer (50 mM NaHCO3, 1% SDS) at RT for 15 min. Following the reversion of the cross links by incubation for 4 h at 65^{o}C and RNA digest by incubation with RNase A the DNA was precipitated O/N with 100% ethanol. The following day remaining proteins were digested using proteinase K and the DNA was purified using a PCR purification kit (Qiagen). Finally the precipitated DNA was analyzed by real time PCR (Corbett bioscience).

Preparation of Staph A cells involved the following: 1 g of Staph A cells were washed twice in Dialysis buffer, by centrifugation at 10,000 rpm for 5 min at 4^{o}C, then resuspended in 3 ml of PBS plus 3% SDS and 10% β-mercaptoethanol and boiled for 30 min. Following 2 washes in Dialysis buffer, as above, the final cell pellet was resuspended in 4 ml of Dialysis buffer and divided into 100 µl aliquots, which were stored at -70^{o}C. Prior to use in ChIP assays, 100 µl of the Staph A cells were pre-blocked with 10 µl of 10 mg/ ml Herring sperm DNA plus 10 µl of 10 mg/ml BSA for 3 h rotating at 4^{o}C. Before use, cells were washed twice in Dialysis buffer and resuspended in 100 µl of the same buffer. 10 µl of these cells were used per immunoprecipitation.

6.2.17 CTD complementation assay

293 cells were transfected with α-amanitin-resistant Rpb1 constructs (WT, Con48, or Ser7A48), and endogenous Rpb1 was turned over by the addition of α-amanitin (Medlin et al., 2003). ChIP was carried out 48 hr after the addition of α-amanitin.

6.2.18 CTD binding assay

In vitro phosphorylation of the GST-CTDs and GST-CTD pull-downs were performed as described previously (Egloff et al., 2007). For the CTD peptide binding assay, CTD peptides were synthesized (Cambridge Peptides) with two repeats of YSPTSPS or YSPTSPA and amino-terminal biotinylation. Streptavidin-coated Sepharose beads (20 µl) were resuspended in 0.1 M KCl HEGN, mixed with 10 µg of biotinylated peptides, and incubated at 4^{o}C for 1 hr. After washing four times with 0.1 M KCl HEGN, 1 µg of recombinant RPAP2 protein was added and the mixture was incubated at 4^{o}C for 3 hr. After washing the beads five times with 1 ml of 0.3 M KCl HEGN, protein-bound beads were boiled and loaded directly on SDS-PAGE gels.

6.2.19 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

The separation gels were prepared with the following final concentration: 375 mM Tris-HCl pH 8.8, 0.1% SDS, 8% to 12% acrylamide/bisacrylamide 37.5:1 (Biorad). The stacking gels were prepared with the following final concentration: 125 mM Tris-HCl pH 6.8, 0.1% SDS, 6% acrylamide/bisacrylamide 37.5. The gels were polymerised by adding ammonium persulfate and TEMED to final concentrations of 0.033% and 0.2%, respectively. The gels were allowed to set for at least 30 min. Protein samples were boiled for 5 min at 96^{o}C in 2x loading buffer containing DTT before loading on the gel. All gels were run in SDS running buffer.

6.2.20 Coomassie staining

Coomassie staining was carried out by adding a solution of Coomassie Brilliant blue in 40% methanol and 10% acetic acid to SDS-PAGE gels followed by an incubation of at least 30 min. Gels were destained using a 20% methanol / 10% acetic acid solution.

6.2.21 Western blot

Cells were collected in PBS and centrifuged at 500 g for 3 min at 4^{o}C. The pellet was lysed using 200 µl of 1X Laemmli buffer with 100mM DTT and denatured by boiling at 90^{o}C for 5 minutes. Proteins separated by SDS-PAGE were transferred to a nitrocellulose

membrane (Hyperbond-C Extra, GE Healthcare) using a semi-dry blotter (Transblot, BioRad). The transfer was carried out at a constant voltage of 10 V for 30 min. The membranes were blocked in 5% milk and 0.1% Tween in PBS for 1 h at RT. The primary antibody was added directly to the blotting solution and incubated for 1 h at RT. After 3 washes of 10 min in 0.1% Tween in PBS, the membrane was incubated with HRP-conjugated secondary antibody in 5% milk and 0.1% Tween in PBS for 1 h at RT. After repeating the 3 washes as previously described, the bands were detected by enhanced chemiluminescence (VisualizerTM Westernblot detection kit, Upstate). Chemiluminescence was detected with BioMax LS Film (Kodak) and developed on a Kodak X-Omat 2000 processor.

Bacterial methods

6.2.22 Transformation of competent bacterial cells

Competent NM554 bacterial cells were thawed on ice and gently resuspended. DNA was added to the cells and mixed gently, incubated on ice for 10 min and heat shocked at 37^{o}C for 5 min in a dry heater. The whole cell mixture was plated onto L-Agar plates with appropriate antibiotics and incubated at 37^{o}C overnight.

References

ACKERMANN, M., SARMIENTO, M. & ROIZMAN, B. 1985. Application of antibody to synthetic peptides for characterization of the intact and truncated alpha 22 protein specified by herpes simplex virus 1 and the R325 alpha 22- deletion mutant. *J Virol,* 56, 207-15.

ADVANI, S. J., BRANDIMARTI, R., WEICHSELBAUM, R. R. & ROIZMAN, B. 2000. The disappearance of cyclins A and B and the increase in activity of the G(2)/M-phase cellular kinase cdc2 in herpes simplex virus 1-infected cells require expression of the alpha22/U(S)1.5 and U(L)13 viral genes. *J Virol,* 74, 8-15.

ALLENDE-VEGA, N., SAVILLE, M. K. & MEEK, D. W. 2007. Transcription factor TAFII250 promotes Mdm2-dependent turnover of p53. *Oncogene,* 26, 4234-42.

ALLISON, L. A., MOYLE, M., SHALES, M. & INGLES, C. J. 1985. Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell,* 42, 599-610.

ANDRULIS, E. D., GUZMAN, E., DORING, P., WERNER, J. & LIS, J. T. 2000. High-resolution localization of Drosophila Spt5 and Spt6 at heat shock genes in vivo: roles in promoter proximal pausing and transcription elongation. *Genes Dev,* 14, 2635-49.

ANSARI, A. & HAMPSEY, M. 2005. A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping. *Genes Dev,* 19, 2969-78.

ARES, M., JR., CHUNG, J. S., GIGLIO, L. & WEINER, A. M. 1987. Distinct factors with Sp1 and NF-A specificities bind to adjacent functional elements of the human U2 snRNA gene enhancer. *Genes Dev,* 1, 808-17.

ASAI, R., OHNO, T., KATO, A. & KAWAGUCHI, Y. 2007. Identification of proteins directly phosphorylated by UL13 protein kinase from herpes simplex virus 1. *Microbes Infect,* 9, 1434-8.

BAILLAT, D., HAKIMI, M. A., NAAR, A. M., SHILATIFARD, A., COOCH, N. & SHIEKHATTAR, R. 2005. Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell,* 123, 265-76.

BARBORIC, M., LENASI, T., CHEN, H., JOHANSEN, E. B., GUO, S. & PETERLIN, B. M. 2009. 7SK snRNP/P-TEFb couples transcription elongation with alternative splicing and is essential for vertebrate development. *Proc Natl Acad Sci U S A,* 106, 7798-803.

BARTKOWIAK, B., LIU, P., PHATNANI, H. P., FUDA, N. J., COOPER, J. J., PRICE, D. H., ADELMAN, K., LIS, J. T. & GREENLEAF, A. L. 2010. CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes Dev,* 24, 2303-16.

BASEHOAR, A. D., ZANTON, S. J. & PUGH, B. F. 2004. Identification and distinct regulation of yeast TATA box-containing genes. *Cell,* 116, 699-709.

BASKARAN, R., DAHMUS, M. E. & WANG, J. Y. 1993. Tyrosine phosphorylation of mammalian RNA polymerase II carboxyl-terminal domain. *Proc Natl Acad Sci U S A,* 90, 11167-71.

BASTIAN, T. W., LIVINGSTON, C. M., WELLER, S. K. & RICE, S. A. 2010. Herpes simplex virus type 1 immediate-early protein ICP22 is required for VICE domain formation during productive viral infection. *J Virol,* 84, 2384-94.

BASTIAN, T. W. & RICE, S. A. 2009. Identification of sequences in herpes simplex virus type 1 ICP22 that influence RNA polymerase II modification and viral late gene expression. *J Virol,* 83, 128-39.

BATAILLE, A. R., JERONIMO, C., JACQUES, P. E., LARAMEE, L., FORTIN, M. E., FOREST, A., BERGERON, M., HANES, S. D. & ROBERT, F. 2012. A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. *Mol Cell,* 45, 158-70.

BENTLEY, D. L. 2005. Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. *Curr Opin Cell Biol,* 17, 251-6.

BERECZKI, O., UJFALUDI, Z., PARDI, N., NAGY, Z., TORA, L., BOROS, I. M. & BALINT, E. 2008. TATA binding protein associated factor 3 (TAF3) interacts with p53 and inhibits its function. *BMC Mol Biol,* 9, 57.

BERGLUND, J. A., ABOVICH, N. & ROSBASH, M. 1998. A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev,* 12, 858-67.

BERNUES, J., SIMMEN, K. A., LEWIS, J. D., GUNDERSON, S. I., POLYCARPOU-SCHWARZ, M., MONCOLLIN, V., EGLY, J. M. & MATTAJ, I. W. 1993. Common and unique transcription factor requirements of human U1 and U6 snRNA genes. *EMBO J,* 12, 3573-85.

BEZSTAROSTI, K., GHAMARI, A., GROSVELD, F. G. & DEMMERS, J. A. 2010. Differential proteomics based on 18O labeling to determine the cyclin dependent kinase 9 interactome. *J Proteome Res,* 9, 4464-75.

BIRCK, C., POCH, O., ROMIER, C., RUFF, M., MENGUS, G., LAVIGNE, A. C., DAVIDSON, I. & MORAS, D. 1998. Human TAF(II)28 and TAF(II)18 interact through a histone fold encoded by atypical evolutionary conserved motifs also found in the SPT3 family. *Cell,* 94, 239-49.

BIRD, G., ZORIO, D. A. & BENTLEY, D. L. 2004. RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3'-end formation. *Mol Cell Biol,* 24, 8963-9.

BLAHO, J. A., MITCHELL, C. & ROIZMAN, B. 1993. Guanylylation and adenylylation of the alpha regulatory proteins of herpes simplex virus require a viral beta or gamma function. *J Virol,* 67, 3891-900.

BLAHO, J. A., ZONG, C. S. & MORTIMER, K. A. 1997. Tyrosine phosphorylation of the herpes simplex virus type 1 regulatory protein ICP22 and a cellular protein which shares antigenic determinants with ICP22. *J Virol,* 71, 9828-32.

BLAZEK, D., KOHOUTEK, J., BARTHOLOMEEUSEN, K., JOHANSEN, E., HULINKOVA, P., LUO, Z., CIMERMANCIC, P., ULE, J. & PETERLIN, B. M. 2011. The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. *Genes Dev,* 25, 2158-72.

BOND, U. M., YARIO, T. A. & STEITZ, J. A. 1991. Multiple processing-defective mutations in a mammalian histone pre-mRNA are suppressed by compensatory changes in U7 RNA both in vivo and in vitro. *Genes Dev,* 5, 1709-22.

BOWMAN, J. J., ORLANDO, J. S., DAVIDO, D. J., KUSHNIR, A. S. & SCHAFFER, P. A. 2009. Transient expression of herpes simplex virus type 1 ICP22 represses viral promoter activity and complements the replication of an ICP22 null virus. *J Virol,* 83, 8733-43.

BOWMAN, J. J. & SCHAFFER, P. A. 2009. Origin of expression of the herpes simplex virus type 1 protein U(S)1.5. ***J Virol,*** **83, 9183-94.**

BRAND, M., YAMAMOTO, K., STAUB, A. & TORA, L. 1999. Identification of TATA-binding protein-free TAFII-containing complex subunits suggests a role in nucleosome acetylation and signal transduction. ***J Biol Chem,*** **274, 18285-9.**

BRANDT, C. R. & KOLB, A. W. 2003. Tyrosine 116 of the herpes simplex virus type 1 IEalpha22 protein is an ocular virulence determinant and potential phosphorylation site. ***Invest Ophthalmol Vis Sci,*** **44, 4601-7.**

BURATOWSKI, S. 2003. The CTD code. ***Nat Struct Biol,*** **10, 679-80.**

BURATOWSKI, S. 2009. Progression through the RNA polymerase II CTD cycle. ***Mol Cell,*** **36, 541-6.**

BURATOWSKI, S., HAHN, S., SHARP, P. A. & GUARENTE, L. 1988. Function of a yeast TATA element-binding protein in a mammalian transcription system. ***Nature,*** **334, 37-42.**

CARTER, K. L. & ROIZMAN, B. 1996. The promoter and transcriptional unit of a novel herpes simplex virus 1 alpha gene are contained in, and encode a protein in frame with, the open reading frame of the alpha 22 gene. ***J Virol,*** **70, 172-8.**

CAVALLINI, B., FAUS, I., MATTHES, H., CHIPOULET, J. M., WINSOR, B., EGLY, J. M. & CHAMBON, P. 1989. 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,*** **86, 9803-7.**

CHALKLEY, G. E. & VERRIJZER, C. P. 1999. DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250-TAF(II)150 complex recognizes the initiator. ***EMBO J,*** **18, 4835-45.**

CHANG, P. C. & LI, M. 2008. Kaposi's sarcoma-associated herpesvirus K-cyclin interacts with Cdk9 and stimulates Cdk9-mediated phosphorylation of p53 tumor suppressor. ***J Virol,*** **82, 278-90.**

CHAPMAN, R. D., CONRAD, M. & EICK, D. 2005. Role of the mammalian RNA polymerase II C-terminal domain (CTD) nonconsensus repeats in CTD stability and cell proliferation. ***Mol Cell Biol,*** **25, 7665-74.**

CHAPMAN, R. D., HEIDEMANN, M., ALBERT, T. K., MAILHAMMER, R., FLATLEY, A., MEISTERERNST, M., KREMMER, E. & EICK, D. 2007. Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. ***Science,*** **318, 1780-2.**

CHEN, J. L., ATTARDI, L. D., VERRIJZER, C. P., YOKOMORI, K. & TJIAN, R. 1994. Assembly of recombinant TFIID reveals differential coactivator requirements for distinct transcriptional activators. ***Cell,*** **79, 93-105.**

CHEN, M. & MANLEY, J. L. 2009. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. ***Nat Rev Mol Cell Biol,*** **10, 741-54.**

CHIANG, C. M. & ROEDER, R. G. 1995. Cloning of an intrinsic human TFIID subunit that interacts with multiple transcriptional activators. ***Science,*** **267, 531-6.**

CHO, E. J., KOBOR, M. S., KIM, M., GREENBLATT, J. & BURATOWSKI, S. 2001. Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. ***Genes Dev,*** **15, 3319-29.**

CHO, H., KIM, T. K., MANCEBO, H., LANE, W. S., FLORES, O. & REINBERG, D. 1999. A protein phosphatase functions to recycle RNA polymerase II. ***Genes Dev,*** **13, 1540-52.**

CLEMENS, K. E., PIRAS, G., RADONOVICH, M. F., CHOI, K. S., DUVALL, J. F., DEJONG, J., ROEDER, R. & BRADY, J. N. 1996. Interaction of the human T-cell lymphotropic virus type 1 tax transactivator with transcription factor IIA. *Mol Cell Biol,* 16, 4656-64.

CLEMENTE-BLANCO, A., SEN, N., MAYAN-SANTOS, M., SACRISTAN, M. P., GRAHAM, B., JARMUZ, A., GIESS, A., WEBB, E., GAME, L., EICK, D., BUENO, A., MERKENSCHLAGER, M. & ARAGON, L. 2011. Cdc14 phosphatase promotes segregation of telomeres through repression of RNA polymerase II transcription. *Nat Cell Biol,* 13, 1450-6.

CLER, E., PAPAI, G., SCHULTZ, P. & DAVIDSON, I. 2009. Recent advances in understanding the structure and function of general transcription factor TFIID. *Cell Mol Life Sci,* 66, 2123-34.

CORDEN, J. L. 1990. Tails of RNA polymerase II. *Trends Biochem Sci,* 15, 383-7.

CRAMER, P. 2002. Common structural features of nucleic acid polymerases. *Bioessays,* 24, 724-9.

CUELLO, P., BOYD, D. C., DYE, M. J., PROUDFOOT, N. J. & MURPHY, S. 1999. Transcription of the human U2 snRNA genes continues beyond the 3' box in vivo. *EMBO J,* 18, 2867-77.

CUN, W., HONG, M., LIU, L. D., DONG, C. H., LUO, J. & LI, Q. H. 2006. Structural and functional characterization of herpes simplex virus 1 immediate-early protein infected-cell protein 22. *J Biochem,* 140, 67-73.

D'ALESSIO, J. A., NG, R., WILLENBRING, H. & TJIAN, R. 2011. Core promoter recognition complex changes accompany liver development. *Proc Natl Acad Sci U S A,* 108, 3906-11.

DAHLBERG, J. E. & SCHENBORN, E. T. 1988. The human U1 snRNA promoter and enhancer do not direct synthesis of messenger RNA. *Nucleic Acids Res,* 16, 5827-40.

DANILITION, S. L., FREDERICKSON, R. M., TAYLOR, C. Y. & MIYAMOTO, N. G. 1991. Transcription factor binding and spacing constraints in the human beta-actin proximal promoter. *Nucleic Acids Res,* 19, 6913-22.

DANTONEL, J. C., MURTHY, K. G., MANLEY, J. L. & TORA, L. 1997. Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA. *Nature,* 389, 399-402.

DAS, G., HINKLEY, C. S. & HERR, W. 1995. Basal promoter elements as a selective determinant of transcriptional activator function. *Nature,* 374, 657-60.

DE VEGVAR, H. E., LUND, E. & DAHLBERG, J. E. 1986. 3' end formation of U1 snRNA precursors is coupled to transcription from snRNA promoters. *Cell,* 47, 259-66.

DEMENY, M. A., SOUTOGLOU, E., NAGY, Z., SCHEER, E., JANOSHAZI, A., RICHARDOT, M., ARGENTINI, M., KESSLER, P. & TORA, L. 2007. Identification of a small TAF complex and its role in the assembly of TAF-containing complexes. *PLoS One,* 2, e316.

DEVAIAH, B. N., LEWIS, B. A., CHERMAN, N., HEWITT, M. C., ALBRECHT, B. K., ROBEY, P. G., OZATO, K., SIMS, R. J., 3RD & SINGER, D. S. 2012. BRD4 is an atypical kinase that phosphorylates serine2 of the RNA polymerase II carboxy-terminal domain. *Proc Natl Acad Sci U S A,* 109, 6927-32.

DICHTL, B., BLANK, D., OHNACKER, M., FRIEDLEIN, A., ROEDER, D., LANGEN, H. & KELLER, W. 2002. A role for SSU72 in balancing RNA polymerase II transcription elongation and termination. ***Mol Cell,*** **10, 1139-50.**

DIKSTEIN, R., ZHOU, S. & TJIAN, R. 1996. Human TAFII 105 is a cell type-specific TFIID subunit related to hTAFII130. ***Cell,*** **87, 137-46.**

DOMINSKI, Z. & MARZLUFF, W. F. 2007. Formation of the 3' end of histone mRNA: getting closer to the end. ***Gene,*** **396, 373-90.**

DOMINSKI, Z., YANG, X. C. & MARZLUFF, W. F. 2005a. The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing. ***Cell,*** **123, 37-48.**

DOMINSKI, Z., YANG, X. C., PURDY, M., WAGNER, E. J. & MARZLUFF, W. F. 2005b. A CPSF-73 homologue is required for cell cycle progression but not cell growth and interacts with a protein having features of CPSF-100. ***Mol Cell Biol,*** **25, 1489-500.**

DU, L. & WARREN, S. L. 1997. A functional interaction between the carboxy-terminal domain of RNA polymerase II and pre-mRNA splicing. ***J Cell Biol,*** **136, 5-18.**

DUNPHY, E. L., JOHNSON, T., AUERBACH, S. S. & WANG, E. H. 2000. Requirement for TAF(II)250 acetyltransferase activity in cell cycle progression. ***Mol Cell Biol,*** **20, 1134-9.**

DURAND, L. O., ADVANI, S. J., POON, A. P. & ROIZMAN, B. 2005. The carboxyl-terminal domain of RNA polymerase II is phosphorylated by a complex containing cdk9 and infected-cell protein 22 of herpes simplex virus 1. ***J Virol,*** **79, 6757-62.**

DURAND, L. O. & ROIZMAN, B. 2008. Role of cdk9 in the optimization of expression of the genes regulated by ICP22 of herpes simplex virus 1. ***J Virol,*** **82, 10591-9.**

DURSO, R. J., FISHER, A. K., ALBRIGHT-FREY, T. J. & REESE, J. C. 2001. Analysis of TAF90 mutants displaying allele-specific and broad defects in transcription. ***Mol Cell Biol,*** **21, 7331-44.**

DYNLACHT, B. D., HOEY, T. & TJIAN, R. 1991. Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. ***Cell,*** **66, 563-76.**

EGGERMONT, J. & PROUDFOOT, N. J. 1993. Poly(A) signals and transcriptional pause sites combine to prevent interference between RNA polymerase II promoters. ***EMBO J,*** **12, 2539-48.**

EGLOFF, S., AL-RAWAF, H., O'REILLY, D. & MURPHY, S. 2009. Chromatin structure is implicated in "late" elongation checkpoints on the U2 snRNA and beta-actin genes. ***Mol Cell Biol,*** **29, 4002-13.**

EGLOFF, S., DIENSTBIER, M. & MURPHY, S. 2012a. Updating the RNA polymerase CTD code: adding gene-specific layers. ***Trends Genet,*** **28, 333-41.**

EGLOFF, S. & MURPHY, S. 2008a. Cracking the RNA polymerase II CTD code. ***Trends Genet,*** **24, 280-8.**

EGLOFF, S. & MURPHY, S. 2008b. Role of the C-terminal domain of RNA polymerase II in expression of small nuclear RNA genes. ***Biochem Soc Trans,*** **36, 537-9.**

EGLOFF, S., O'REILLY, D., CHAPMAN, R. D., TAYLOR, A., TANZHAUS, K., PITTS, L., EICK, D. & MURPHY, S. 2007. Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. ***Science,*** **318, 1777-9.**

EGLOFF, S., O'REILLY, D. & MURPHY, S. 2008. Expression of human snRNA genes from beginning to end. ***Biochem Soc Trans,*** **36, 590-4.**

EGLOFF, S., SZCZEPANIAK, S. A., DIENSTBIER, M., TAYLOR, A., KNIGHT, S. & MURPHY, S. 2010. The integrator complex recognizes a new double mark on the RNA polymerase II carboxyl-terminal domain. *J Biol Chem,* 285, 20564-9.

EGLOFF, S., ZABOROWSKA, J., LAITEM, C., KISS, T. & MURPHY, S. 2012b. Ser7 phosphorylation of the CTD recruits the RPAP2 Ser5 phosphatase to snRNA genes. *Mol Cell,* 45, 111-22.

FLORES, O., LU, H., KILLEEN, M., GREENBLATT, J., BURTON, Z. F. & REINBERG, D. 1991. The small subunit of transcription factor IIF recruits RNA polymerase II into the preinitiation complex. *Proc Natl Acad Sci U S A,* 88, 9999-10003.

FLORES, O., MALDONADO, E. & REINBERG, D. 1989. Factors involved in specific transcription by mammalian RNA polymerase II. Factors IIE and IIF independently interact with RNA polymerase II. *J Biol Chem,* 264, 8913-21.

FORNAGE, M., DEBETTE, S., BIS, J. C., SCHMIDT, H., IKRAM, M. A., DUFOUIL, C., SIGURDSSON, S., LUMLEY, T., DESTEFANO, A. L., FAZEKAS, F., VROOMAN, H. A., SHIBATA, D. K., MAILLARD, P., ZIJDENBOS, A., SMITH, A. V., GUDNASON, H., DE BOER, R., CUSHMAN, M., MAZOYER, B., HEISS, G., VERNOOIJ, M. W., ENZINGER, C., GLAZER, N. L., BEISER, A., KNOPMAN, D. S., CAVALIERI, M., NIESSEN, W. J., HARRIS, T. B., PETROVIC, K., LOPEZ, O. L., AU, R., LAMBERT, J. C., HOFMAN, A., GOTTESMAN, R. F., GARCIA, M., HECKBERT, S. R., ATWOOD, L. D., CATELLIER, D. J., UITTERLINDEN, A. G., YANG, Q., SMITH, N. L., ASPELUND, T., ROMERO, J. R., RICE, K., TAYLOR, K. D., NALLS, M. A., ROTTER, J. I., SHARRETT, R., VAN DUIJN, C. M., AMOUYEL, P., WOLF, P. A., GUDNASON, V., VAN DER LUGT, A., BOERWINKLE, E., PSATY, B. M., SESHADRI, S., TZOURIO, C., BRETELER, M. M., MOSLEY, T. H., SCHMIDT, R., LONGSTRETH, W. T., DECARLI, C. & LAUNER, L. J. 2011. Genome-wide association studies of cerebral white matter lesion burden: the CHARGE consortium. *Ann Neurol,* 69, 928-39.

FRASER, K. A. & RICE, S. A. 2005. Herpes simplex virus type 1 infection leads to loss of serine-2 phosphorylation on the carboxyl-terminal domain of RNA polymerase II. *J Virol,* 79, 11323-34.

FRASER, K. A. & RICE, S. A. 2007. Herpes simplex virus immediate-early protein ICP22 triggers loss of serine 2-phosphorylated RNA polymerase II. *J Virol,* 81, 5091-101.

FREIMAN, R. N., ALBRIGHT, S. R., ZHENG, S., SHA, W. C., HAMMER, R. E. & TJIAN, R. 2001. Requirement of tissue-selective TBP-associated factor TAFII105 in ovarian development. *Science,* 293, 2084-7.

FU, T. J., PENG, J., LEE, G., PRICE, D. H. & FLORES, O. 1999. Cyclin K functions as a CDK9 regulatory subunit and participates in RNA polymerase II transcription. *J Biol Chem,* 274, 34527-30.

FUDA, N. J., BUCKLEY, M. S., WEI, W., CORE, L. J., WATERS, C. T., REINBERG, D. & LIS, J. T. 2012. Fcp1 dephosphorylation of the RNA polymerase II C-terminal domain is required for efficient transcription of heat shock genes. *Mol Cell Biol,* 32, 3428-37.

GALDIERO, S., FALANGA, A., TARALLO, R., RUSSO, L., GALDIERO, E., CANTISANI, M., MORELLI, G. & GALDIERO, M. 2013. Peptide inhibitors against herpes simplex virus infections. *J Pept Sci,* 19, 148-58.

GANEM, C., DEVAUX, F., TORCHET, C., JACQ, C., QUEVILLON-CHERUEL, S., LABESSE, G., FACCA, C. & FAYE, G. 2003. Ssu72 is a phosphatase essential for transcription termination of snoRNAs and specific mRNAs in yeast. *EMBO J,* 22, 1588-98.

GANGLOFF, Y. G., ROMIER, C., THUAULT, S., WERTEN, S. & DAVIDSON, I. 2001a. The histone fold is a key structural motif of transcription factor TFIID. *Trends Biochem Sci,* 26, 250-7.

GANGLOFF, Y. G., SANDERS, S. L., ROMIER, C., KIRSCHNER, D., WEIL, P. A., TORA, L. & DAVIDSON, I. 2001b. Histone folds mediate selective heterodimerization of yeast TAF(II)25 with TFIID components yTAF(II)47 and yTAF(II)65 and with SAGA component ySPT7. *Mol Cell Biol,* 21, 1841-53.

GAVIN, A. C., BOSCHE, M., KRAUSE, R., GRANDI, P., MARZIOCH, M., BAUER, A., SCHULTZ, J., RICK, J. M., MICHON, A. M., CRUCIAT, C. M., REMOR, M., HOFERT, C., SCHELDER, M., BRAJENOVIC, M., RUFFNER, H., MERINO, A., KLEIN, K., HUDAK, M., DICKSON, D., RUDI, T., GNAU, V., BAUCH, A., BASTUCK, S., HUHSE, B., LEUTWEIN, C., HEURTIER, M. A., COPLEY, R. R., EDELMANN, A., QUERFURTH, E., RYBIN, V., DREWES, G., RAIDA, M., BOUWMEESTER, T., BORK, P., SERAPHIN, B., KUSTER, B., NEUBAUER, G. & SUPERTI-FURGA, G. 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature,* 415, 141-7.

GEGONNE, A., WEISSMAN, J. D., LU, H., ZHOU, M., DASGUPTA, A., RIBBLE, R., BRADY, J. N. & SINGER, D. S. 2008. TFIID component TAF7 functionally interacts with both TFIIH and P-TEFb. *Proc Natl Acad Sci U S A,* 105, 5367-72.

GEGONNE, A., WEISSMAN, J. D. & SINGER, D. S. 2001. TAFII55 binding to TAFII250 inhibits its acetyltransferase activity. *Proc Natl Acad Sci U S A,* 98, 12432-7.

GEGONNE, A., WEISSMAN, J. D., ZHOU, M., BRADY, J. N. & SINGER, D. S. 2006. TAF7: a possible transcription initiation check-point regulator. *Proc Natl Acad Sci U S A,* 103, 602-7.

GIBNEY, P. A., FRIES, T., BAILER, S. M. & MORANO, K. A. 2008. Rtr1 is the Saccharomyces cerevisiae homolog of a novel family of RNA polymerase II-binding proteins. *Eukaryot Cell,* 7, 938-48.

GIL, A. & PROUDFOOT, N. J. 1984. A sequence downstream of AAUAAA is required for rabbit beta-globin mRNA 3'-end formation. *Nature,* 312, 473-4.

GILL, G., PASCAL, E., TSENG, Z. H. & TJIAN, R. 1994. A glutamine-rich hydrophobic patch in transcription factor Sp1 contacts the dTAFII110 component of the Drosophila TFIID complex and mediates transcriptional activation. *Proc Natl Acad Sci U S A,* 91, 192-6.

GLOVER-CUTTER, K., KIM, S., ESPINOSA, J. & BENTLEY, D. L. 2008. RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nat Struct Mol Biol,* 15, 71-8.

GLOVER-CUTTER, K., LAROCHELLE, S., ERICKSON, B., ZHANG, C., SHOKAT, K., FISHER, R. P. & BENTLEY, D. L. 2009. TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. *Mol Cell Biol,* 29, 5455-64.

GOMES, N. P., BJERKE, G., LLORENTE, B., SZOSTEK, S. A., EMERSON, B. M. & ESPINOSA, J. M. 2006. Gene-specific requirement for P-TEFb activity and RNA polymerase II phosphorylation within the p53 transcriptional program. *Genes Dev,* 20, 601-12.

GOODRICH, J. A. & TJIAN, R. 1994. Transcription factors IIE and IIH and ATP hydrolysis direct promoter clearance by RNA polymerase II. *Cell,* 77, 145-56.

GOODRICH, J. A. & TJIAN, R. 2010. Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. *Nat Rev Genet,* 11, 549-58.

GOVIND, C. K., QIU, H., GINSBURG, D. S., RUAN, C., HOFMEYER, K., HU, C., SWAMINATHAN, V., WORKMAN, J. L., LI, B. & HINNEBUSCH, A. G. 2010. Phosphorylated Pol II CTD recruits multiple HDACs, including Rpd3C(S), for methylation-dependent deacetylation of ORF nucleosomes. *Mol Cell,* 39, 234-46.

GRANT, P. A., SCHIELTZ, D., PRAY-GRANT, M. G., STEGER, D. J., REESE, J. C., YATES, J. R., 3RD & WORKMAN, J. L. 1998. A subset of TAF(II)s are integral components of the SAGA complex required for nucleosome acetylation and transcriptional stimulation. *Cell,* 94, 45-53.

GREEN, M. R. 2000. TBP-associated factors (TAFIIs): multiple, selective transcriptional mediators in common complexes. *Trends Biochem Sci,* 25, 59-63.

GROMAK, N., WEST, S. & PROUDFOOT, N. J. 2006. Pause sites promote transcriptional termination of mammalian RNA polymerase II. *Mol Cell Biol,* 26, 3986-96.

GUILLAMOT, M., MANCHADO, E., CHIESA, M., GOMEZ-LOPEZ, G., PISANO, D. G., SACRISTAN, M. P. & MALUMBRES, M. 2011. Cdc14b regulates mammalian RNA polymerase II and represses cell cycle transcription. *Sci Rep,* 1, 189.

GUNDERSON, S. I., KNUTH, M. W. & BURGESS, R. R. 1990. The human U1 snRNA promoter correctly initiates transcription in vitro and is activated by PSE1. *Genes Dev,* 4, 2048-60.

GUO, L., WU, W. J., LIU, L. D., WANG, L. C., ZHANG, Y., WU, L. Q., GUAN, Y. & LI, Q. H. 2012. Herpes simplex virus 1 ICP22 inhibits the transcription of viral gene promoters by binding to and blocking the recruitment of P-TEFb. *PLoS One,* 7, e45749.

HAHN, S. 2004. Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol,* 11, 394-403.

HAMPSEY, M. 1998. Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol Mol Biol Rev,* 62, 465-503.

HAMPSEY, M., SINGH, B. N., ANSARI, A., LAINE, J. P. & KRISHNAMURTHY, S. 2011. Control of eukaryotic gene expression: gene loops and transcriptional memory. *Adv Enzyme Regul,* 51, 118-25.

HARTZOG, G. A., SPEER, J. L. & LINDSTROM, D. L. 2002. Transcript elongation on a nucleoprotein template. *Biochim Biophys Acta,* 1577, 276-86.

HE, X., KHAN, A. U., CHENG, H., PAPPAS, D. L., JR., HAMPSEY, M. & MOORE, C. L. 2003. Functional interactions between the transcription and mRNA 3' end processing machineries mediated by Ssu72 and Sub1. *Genes Dev,* 17, 1030-42.

HENRY, R. W., SADOWSKI, C. L., KOBAYASHI, R. & HERNANDEZ, N. 1995. A TBP-TAF complex required for transcription of human snRNA genes by RNA polymerase II and III. *Nature,* 374, 653-6.

HERNANDEZ, N. 1985. Formation of the 3' end of U1 snRNA is directed by a conserved sequence located downstream of the coding region. *EMBO J,* 4, 1827-37.

HERNANDEZ, N. 1993. TBP, a universal eukaryotic transcription factor? *Genes Dev,* 7, 1291-308.

HERNANDEZ, N. 2001. Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. *J Biol Chem,* 276, 26733-6.

HERNANDEZ, N. & WEINER, A. M. 1986. Formation of the 3' end of U1 snRNA requires compatible snRNA promoter elements. *Cell,* 47, 249-58.

HERRMANN, C. H. & RICE, A. P. 1995. Lentivirus Tat proteins specifically associate with a cellular protein kinase, TAK, that hyperphosphorylates the carboxyl-terminal domain of the large subunit of RNA polymerase II: candidate for a Tat cofactor. *J Virol,* 69, 1612-20.

HILLER, M. A., LIN, T. Y., WOOD, C. & FULLER, M. T. 2001. Developmental regulation of transcription by a tissue-specific TAF homolog. *Genes Dev,* 15, 1021-30.

HINTERMAIR, C., HEIDEMANN, M., KOCH, F., DESCOSTES, N., GUT, M., GUT, I., FENOUIL, R., FERRIER, P., FLATLEY, A., KREMMER, E., CHAPMAN, R. D., ANDRAU, J. C. & EICK, D. 2012. Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. *EMBO J,* 31, 2784-97.

HIROSE, Y. & MANLEY, J. L. 2000. RNA polymerase II and the integration of nuclear events. *Genes Dev,* 14, 1415-29.

HOLSTEGE, F. C., VAN DER VLIET, P. C. & TIMMERS, H. T. 1996. Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. *EMBO J,* 15, 1666-77.

HONESS, R. W. & ROIZMAN, B. 1974. Regulation of herpesvirus macromolecular synthesis. I. Cascade regulation of the synthesis of three groups of viral proteins. *J Virol,* 14, 8-19.

HOWCROFT, T. K., RAVAL, A., WEISSMAN, J. D., GEGONNE, A. & SINGER, D. S. 2003. Distinct transcriptional pathways regulate basal and activated major histocompatibility complex class I expression. *Mol Cell Biol,* 23, 3377-91.

HSIN, J. P. & MANLEY, J. L. 2012. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev,* 26, 2119-37.

HSIN, J. P., SHETH, A. & MANLEY, J. L. 2011. RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing. *Science,* 334, 683-6.

IMHOF, A., YANG, X. J., OGRYZKO, V. V., NAKATANI, Y., WOLFFE, A. P. & GE, H. 1997. Acetylation of general transcription factors by histone acetyltransferases. *Curr Biol,* 7, 689-92.

JACOBS, E. Y., OGIWARA, I. & WEINER, A. M. 2004. Role of the C-terminal domain of RNA polymerase II in U2 snRNA transcription and 3' processing. *Mol Cell Biol,* 24, 846-55.

JACOBSON, R. H., LADURNER, A. G., KING, D. S. & TJIAN, R. 2000. Structure and function of a human TAFII250 double bromodomain module. *Science,* 288, 1422-5.

JERONIMO, C., FORGET, D., BOUCHARD, A., LI, Q., CHUA, G., POITRAS, C., THERIEN, C., BERGERON, D., BOURASSA, S., GREENBLATT, J., CHABOT, B., POIRIER, G. G., HUGHES, T. R., BLANCHETTE, M., PRICE, D. H. & COULOMBE, B. 2007. Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. *Mol Cell,* 27, 262-74.

JOHNSON, D. S., MORTAZAVI, A., MYERS, R. M. & WOLD, B. 2007. Genome-wide mapping of in vivo protein-DNA interactions. *Science,* 316, 1497-502.

KANG, J. J., AUBLE, D. T., RANISH, J. A. & HAHN, S. 1995. Analysis of the yeast transcription factor TFIIA: distinct functional regions and a polymerase II-specific role in basal and activated transcription. *Mol Cell Biol,* 15, 1234-43.

KAUFMANN, J. & SMALE, S. T. 1994. Direct recognition of initiator elements by a component of the transcription factor IID complex. *Genes Dev,* 8, 821-9.

KELLY, W. G., DAHMUS, M. E. & HART, G. W. 1993. RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. *J Biol Chem,* 268, 10416-24.

KEMP, L. M. & LATCHMAN, D. S. 1988. Induction and repression of cellular gene transcription during herpes simplex virus infection are mediated by different viral immediate-early gene products. *Eur J Biochem,* 174, 443-9.

KEOGH, M. C., PODOLNY, V. & BURATOWSKI, S. 2003. Bur1 kinase is required for efficient transcription elongation by RNA polymerase II. *Mol Cell Biol,* 23, 7005-18.

KIM, D. H., LEE, S. H., NAM, K. H., CHI, S. W., CHANG, I. & HAN, K. H. 2009a. Multiple hTAF(II)31-binding motifs in the intrinsically unfolded transcriptional activation domain of VP16. *BMB Rep,* 42, 411-7.

KIM, M., SUH, H., CHO, E. J. & BURATOWSKI, S. 2009b. Phosphorylation of the yeast Rpb1 C-terminal domain at serines 2, 5, and 7. *J Biol Chem,* 284, 26421-6.

KLEINSCHMIDT, A. M. & PEDERSON, T. 1987. Accurate and efficient 3' processing of U2 small nuclear RNA precursor in a fractionated cytoplasmic extract. *Mol Cell Biol,* 7, 3131-7.

KNIPE, D. M. & CLIFFE, A. 2008. Chromatin control of herpes simplex virus lytic and latent infection. *Nat Rev Microbiol,* 6, 211-21.

KOBAYASHI, N., BOYER, T. G. & BERK, A. J. 1995. A class of activation domains interacts directly with TFIIA and stimulates TFIIA-TFIID-promoter complex assembly. *Mol Cell Biol,* 15, 6465-73.

KOBOR, M. S., ARCHAMBAULT, J., LESTER, W., HOLSTEGE, F. C., GILEADI, O., JANSMA, D. B., JENNINGS, E. G., KOUYOUMDJIAN, F., DAVIDSON, A. R., YOUNG, R. A. & GREENBLATT, J. 1999. An unusual eukaryotic protein phosphatase required for transcription by RNA polymerase II and CTD dephosphorylation in S. cerevisiae. *Mol Cell,* 4, 55-62.

KOBOR, M. S. & GREENBLATT, J. 2002. Regulation of transcription elongation by phosphorylation. *Biochim Biophys Acta,* 1577, 261-275.

KOBOR, M. S., SIMON, L. D., OMICHINSKI, J., ZHONG, G., ARCHAMBAULT, J. & GREENBLATT, J. 2000. A motif shared by TFIIF and TFIIB mediates their interaction with the RNA polymerase II carboxy-terminal domain phosphatase Fcp1p in Saccharomyces cerevisiae. *Mol Cell Biol,* 20, 7438-49.

KOLB, A. W., SCHMIDT, T. R., DYER, D. W. & BRANDT, C. R. 2011. Sequence variation in the herpes simplex virus U(S)1 ocular virulence determinant. *Invest Ophthalmol Vis Sci,* 52, 4630-8.

KOMARNITSKY, P., CHO, E. J. & BURATOWSKI, S. 2000. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev,* 14, 2452-60.

KONARSKA, M. M., GRABOWSKI, P. J., PADGETT, R. A. & SHARP, P. A. 1985. Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors. *Nature,* 313, 552-7.

KORNBLIHTT, A. R., DE LA MATA, M., FEDEDA, J. P., MUNOZ, M. J. & NOGUES, G. 2004. Multiple links between transcription and splicing. *RNA,* 10, 1489-98.

KRISHNAMURTHY, S., HE, X., REYES-REYES, M., MOORE, C. & HAMPSEY, M. 2004. Ssu72 Is an RNA polymerase II CTD phosphatase. *Mol Cell,* 14, 387-94.

KUHLMAN, T. C., CHO, H., REINBERG, D. & HERNANDEZ, N. 1999. The general transcription factors IIA, IIB, IIF, and IIE are required for RNA polymerase II transcription from the human U1 small nuclear RNA promoter. *Mol Cell Biol,* 19, 2130-41.

KUTACH, A. K. & KADONAGA, J. T. 2000. The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. *Mol Cell Biol,* 20, 4754-64.

LEE, D. H., GERSHENZON, N., GUPTA, M., IOSHIKHES, I. P., REINBERG, D. & LEWIS, B. A. 2005. Functional characterization of core promoter elements: the downstream core element is recognized by TAF1. *Mol Cell Biol,* 25, 9674-86.

LEOPARDI, R., WARD, P. L., OGLE, W. O. & ROIZMAN, B. 1997. Association of herpes simplex virus regulatory protein ICP22 with transcriptional complexes containing EAP, ICP4, RNA polymerase II, and viral DNA requires posttranslational modification by the U(L)13 proteinkinase. *J Virol,* 71, 1133-9.

LEURENT, C., SANDERS, S., RUHLMANN, C., MALLOUH, V., WEIL, P. A., KIRSCHNER, D. B., TORA, L. & SCHULTZ, P. 2002. Mapping histone fold TAFs within yeast TFIID. *EMBO J,* 21, 3424-33.

LI, M., PHATNANI, H. P., GUAN, Z., SAGE, H., GREENLEAF, A. L. & ZHOU, P. 2005. Solution structure of the Set2-Rpb1 interacting domain of human Set2 and its interaction with the hyperphosphorylated C-terminal domain of Rpb1. *Proc Natl Acad Sci U S A,* 102, 17636-41.

LIN, P. S., DUBOIS, M. F. & DAHMUS, M. E. 2002. TFIIF-associating carboxyl-terminal domain phosphatase dephosphorylates phosphoserines 2 and 5 of RNA polymerase II. *J Biol Chem,* 277, 45949-56.

LIU, H. & HERRMANN, C. H. 2005. Differential localization and expression of the Cdk9 42k and 55k isoforms. *J Cell Physiol,* 203, 251-60.

LIU, W. L., COLEMAN, R. A., GROB, P., KING, D. S., FLORENS, L., WASHBURN, M. P., GELES, K. G., YANG, J. L., RAMEY, V., NOGALES, E. & TJIAN, R. 2008. Structural changes in TAF4b-TFIID correlate with promoter selectivity. *Mol Cell,* 29, 81-91.

LONG, M. C., LEONG, V., SCHAFFER, P. A., SPENCER, C. A. & RICE, S. A. 1999. ICP22 and the UL13 protein kinase are both required for herpes simplex virus-induced modification of the large subunit of RNA polymerase II. *J Virol,* 73, 5593-604.

LUND, E., KAHAN, B. & DAHLBERG, J. E. 1985. Differential control of U1 small nuclear RNA expression during mouse development. *Science,* 229, 1271-4.

MAILE, T., KWOCZYNSKI, S., KATZENBERGER, R. J., WASSARMAN, D. A. & SAUER, F. 2004. TAF1 activates transcription by phosphorylation of serine 33 in histone H2B. *Science,* 304, 1010-4.

MANCEBO, H. S., LEE, G., FLYGARE, J., TOMASSINI, J., LUU, P., ZHU, Y., PENG, J., BLAU, C., HAZUDA, D., PRICE, D. & FLORES, O. 1997. P-TEFb kinase is required for HIV Tat transcriptional activation in vivo and in vitro. *Genes Dev,* 11, 2633-44.

MANDAL, S. S., CHO, H., KIM, S., CABANE, K. & REINBERG, D. 2002. FCP1, a phosphatase specific for the heptapeptide repeat of the largest subunit of RNA polymerase II, stimulates transcription elongation. *Mol Cell Biol,* 22, 7543-52.

MARSHALL, N. F., PENG, J., XIE, Z. & PRICE, D. H. 1996. Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *J Biol Chem,* 271, 27176-83.

MARSHALL, N. F. & PRICE, D. H. 1992. Control of formation of two distinct classes of RNA polymerase II elongation complexes. *Mol Cell Biol,* 12, 2078-90.

MARSHALL, N. F. & PRICE, D. H. 1995. Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *J Biol Chem,* 270, 12335-8.

MARTINEZ, E., PALHAN, V. B., TJERNBERG, A., LYMAR, E. S., GAMPER, A. M., KUNDU, T. K., CHAIT, B. T. & ROEDER, R. G. 2001. Human STAGA complex is a chromatin-acetylating transcription coactivator that interacts with pre-mRNA splicing and DNA damage-binding factors in vivo. *Mol Cell Biol,* 21, 6782-95.

MASTON, G. A., ZHU, L. J., CHAMBERLAIN, L., LIN, L., FANG, M. & GREEN, M. R. 2012. Non-canonical TAF complexes regulate active promoters in human embryonic stem cells. *Elife,* 1, e00068.

MATSUI, T., SEGALL, J., WEIL, P. A. & ROEDER, R. G. 1980. Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. *J Biol Chem,* 255, 11992-6.

MATTAJ, I. W., LIENHARD, S., JIRICNY, J. & DE ROBERTIS, E. M. 1985. An enhancer-like sequence within the Xenopus U2 gene promoter facilitates the formation of stable transcription complexes. *Nature,* 316, 163-7.

MAYER, A., HEIDEMANN, M., LIDSCHREIBER, M., SCHREIECK, A., SUN, M., HINTERMAIR, C., KREMMER, E., EICK, D. & CRAMER, P. 2012. CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science,* 336, 1723-5.

MCCRACKEN, S., FONG, N., YANKULOV, K., BALLANTYNE, S., PAN, G., GREENBLATT, J., PATTERSON, S. D., WICKENS, M. & BENTLEY, D. L. 1997. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature,* 385, 357-61.

MEDLIN, J., SCURRY, A., TAYLOR, A., ZHANG, F., PETERLIN, B. M. & MURPHY, S. 2005. P-TEFb is not an essential elongation factor for the intronless human U2 snRNA and histone H2b genes. *EMBO J,* 24, 4154-65.

MEDLIN, J. E., UGUEN, P., TAYLOR, A., BENTLEY, D. L. & MURPHY, S. 2003. The C-terminal domain of pol II and a DRB-sensitive kinase are required for 3' processing of U2 snRNA. *EMBO J,* 22, 925-34.

MEINHART, A., SILBERZAHN, T. & CRAMER, P. 2003. The mRNA transcription/processing factor Ssu72 is a potential tyrosine phosphatase. *J Biol Chem,* 278, 15917-21.

MITCHELL, C., BLAHO, J. A., MCCORMICK, A. L. & ROIZMAN, B. 1997. The nucleotidylylation of herpes simplex virus 1 regulatory protein alpha22 by human casein kinase II. *J Biol Chem,* 272, 25394-400.

MITCHELL, C., BLAHO, J. A. & ROIZMAN, B. 1994. Casein kinase II specifically nucleotidylylates in vitro the amino acid sequence of the protein encoded by the alpha 22 gene of herpes simplex virus 1. *Proc Natl Acad Sci U S A,* 91, 11864-8.

MOHAN, W. S., JR., SCHEER, E., WENDLING, O., METZGER, D. & TORA, L. 2003. TAF10 (TAF(II)30) is necessary for TFIID stability and early embryogenesis in mice. *Mol Cell Biol,* 23, 4307-18.

MORRIS, D. P. & GREENLEAF, A. L. 2000. The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J Biol Chem,* 275, 39935-43.

MOSLEY, A. L., PATTENDEN, S. G., CAREY, M., VENKATESH, S., GILMORE, J. M., FLORENS, L., WORKMAN, J. L. & WASHBURN, M. P. 2009. Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation. *Mol Cell,* 34, 168-78.

MURPHY, S., YOON, J. B., GERSTER, T. & ROEDER, R. G. 1992. Oct-1 and Oct-2 potentiate functional interactions of a transcription factor with the proximal sequence element of small nuclear RNA genes. *Mol Cell Biol,* 12, 3247-61.

NAAR, A. M., LEMON, B. D. & TJIAN, R. 2001. Transcriptional coactivator complexes. *Annu Rev Biochem,* 70, 475-501.

NAKAJIMA-IIJIMA, S., HAMADA, H., REDDY, P. & KAKUNAGA, T. 1985. Molecular structure of the human cytoplasmic beta-actin gene: interspecies homology of sequences in the introns. *Proc Natl Acad Sci U S A,* 82, 6133-7.

NECHAEV, S. & ADELMAN, K. 2011. Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. *Biochim Biophys Acta,* 1809, 34-45.

NEDEA, E., HE, X., KIM, M., POOTOOLAL, J., ZHONG, G., CANADIEN, V., HUGHES, T., BURATOWSKI, S., MOORE, C. L. & GREENBLATT, J. 2003. Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3'-ends. *J Biol Chem,* 278, 33000-10.

NG, H. H., ROBERT, F., YOUNG, R. A. & STRUHL, K. 2003. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol Cell,* 11, 709-19.

NG, S. Y., GUNNING, P., EDDY, R., PONTE, P., LEAVITT, J., SHOWS, T. & KEDES, L. 1985. Evolution of the functional human beta-actin gene and its multi-pseudogene family: conservation of noncoding regions and chromosomal dispersion of pseudogenes. *Mol Cell Biol,* 5, 2720-32.

NI, Z., SCHWARTZ, B. E., WERNER, J., SUAREZ, J. R. & LIS, J. T. 2004. Coordination of transcription, RNA processing, and surveillance by P-TEFb kinase on heat shock genes. *Mol Cell,* 13, 55-65.

OELGESCHLAGER, T. 2002. Regulation of RNA polymerase II activity by CTD phosphorylation and cell cycle control. *J Cell Physiol,* 190, 160-9.

OGLE, W. O. & ROIZMAN, B. 1999. Functional anatomy of herpes simplex virus 1 overlapping genes encoding infected-cell protein 22 and US1.5 protein. *J Virol,* 73, 4305-15.

OGRYZKO, V. V., KOTANI, T., ZHANG, X., SCHILTZ, R. L., HOWARD, T., YANG, X. J., HOWARD, B. H., QIN, J. & NAKATANI, Y. 1998. Histone-like TAFs within the PCAF histone acetylase complex. *Cell,* 94, 35-44.

ORLANDO, J. S., BALLIET, J. W., KUSHNIR, A. S., ASTOR, T. L., KOSZ-VNENCHAK, M., RICE, S. A., KNIPE, D. M. & SCHAFFER, P. A. 2006. ICP22 is required for wild-type composition and infectivity of herpes simplex virus type 1 virions. *J Virol,* 80, 9381-90.

ORPHANIDES, G., LAGRANGE, T. & REINBERG, D. 1996. The general transcription factors of RNA polymerase II. *Genes Dev,* 10, 2657-83.

OTANI, Y., NAKATSU, Y., SAKODA, H., FUKUSHIMA, T., FUJISHIRO, M., KUSHIYAMA, A., OKUBO, H., TSUCHIYA, Y., OHNO, H., TAKAHASHI, S., NISHIMURA, F., KAMATA, H., KATAGIRI, H. & ASANO, T. 2013. Integrator complex plays an essential role in adipose differentiation. *Biochem Biophys Res Commun,* 434, 197-202.

PAPAI, G., TRIPATHI, M. K., RUHLMANN, C., WERTEN, S., CRUCIFIX, C., WEIL, P. A. & SCHULTZ, P. 2009. Mapping the initiator binding Taf2 subunit in the structure of hydrated yeast TFIID. *Structure,* 17, 363-73.

PARDEE, T. S., BANGUR, C. S. & PONTICELLI, A. S. 1998. The N-terminal region of yeast TFIIB contains two adjacent functional domains involved in stable RNA polymerase II binding and transcription start site selection. *J Biol Chem,* 273, 17859-64.

PARKER, C. S. & TOPOL, J. 1984. A Drosophila RNA polymerase II transcription factor binds to the regulatory site of an hsp 70 gene. *Cell,* 37, 273-83.

PAYNE, J. M., LAYBOURN, P. J. & DAHMUS, M. E. 1989. The transition of RNA polymerase II from initiation to elongation is associated with phosphorylation of the carboxyl-terminal domain of subunit IIa. *J Biol Chem,* 264, 19621-9.

PENG, J., MARSHALL, N. F. & PRICE, D. H. 1998a. Identification of a cyclin subunit required for the function of Drosophila P-TEFb. *J Biol Chem,* 273, 13855-60.

PENG, J., ZHU, Y., MILTON, J. T. & PRICE, D. H. 1998b. Identification of multiple cyclin subunits of human P-TEFb. *Genes Dev,* 12, 755-62.

PETERLIN, B. M. & PRICE, D. H. 2006. Controlling the elongation phase of transcription with P-TEFb. *Mol Cell,* 23, 297-305.

PHAM, A. D. & SAUER, F. 2000. Ubiquitin-activating/conjugating activity of TAFII250, a mediator of activation of gene expression in Drosophila. *Science,* 289, 2357-60.

PHATNANI, H. P. & GREENLEAF, A. L. 2006. Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev,* 20, 2922-36.

PINTO, I., WU, W. H., NA, J. G. & HAMPSEY, M. 1994. Characterization of sua7 mutations defines a domain of TFIIB involved in transcription start site selection in yeast. *J Biol Chem,* 269, 30569-73.

POFFENBERGER, K. L., IDOWU, A. D., FRASER-SMITH, E. B., RAICHLEN, P. E. & HERMAN, R. C. 1994. A herpes simplex virus type 1 ICP22 deletion mutant is altered for virulence and latency in vivo. *Arch Virol,* 139, 111-9.

POINTUD, J. C., MENGUS, G., BRANCORSINI, S., MONACO, L., PARVINEN, M., SASSONE-CORSI, P. & DAVIDSON, I. 2003. The intracellular localisation of TAF7L, a paralogue of transcription factor TFIID subunit TAF7, is developmentally regulated during male germ-cell differentiation. *J Cell Sci,* 116, 1847-58.

POON, A. P., OGLE, W. O. & ROIZMAN, B. 2000. Posttranslational processing of infected cell protein 22 mediated by viral protein kinases is sensitive to amino acid substitutions at distant sites and can be cell-type specific. *J Virol,* 74, 11210-4.

POST, L. E. & ROIZMAN, B. 1981. A generalized technique for deletion of specific genes in large genomes: alpha gene 22 of herpes simplex virus 1 is not essential for growth. *Cell,* 25, 227-32.

PROD'HON, C., MACHUCA, I., BERTHOMME, H., EPSTEIN, A. & JACQUEMONT, B. 1996. Characterization of regulatory functions of the HSV-1 immediate-early protein ICP22. *Virology,* 226, 393-402.

PROUDFOOT, N. J., FURGER, A. & DYE, M. J. 2002. Integrating mRNA processing with transcription. *Cell,* 108, 501-12.

PURNELL, B. A., EMANUEL, P. A. & GILMOUR, D. S. 1994. TFIID sequence recognition of the initiator and sequences farther downstream in Drosophila class II genes. *Genes Dev,* 8, 830-42.

PURVES, F. C., OGLE, W. O. & ROIZMAN, B. 1993. Processing of the herpes simplex virus regulatory protein alpha 22 mediated by the UL13 protein kinase determines the accumulation of a subset of alpha and gamma mRNAs and proteins in infected cells. *Proc Natl Acad Sci U S A,* 90, 6701-5.

PURVES, F. C. & ROIZMAN, B. 1992. The UL13 gene of herpes simplex virus 1 encodes the functions for posttranslational processing associated with phosphorylation of the regulatory protein alpha 22. *Proc Natl Acad Sci U S A,* 89, 7310-4.

RAHA, T., CHENG, S. W. & GREEN, M. R. 2005. HIV-1 Tat stimulates transcription complex assembly through recruitment of TBP in the absence of TAFs. *PLoS Biol,* 3, e44.

RAHL, P. B., LIN, C. Y., SEILA, A. C., FLYNN, R. A., MCCUINE, S., BURGE, C. B., SHARP, P. A. & YOUNG, R. A. 2010. c-Myc regulates transcriptional pause release. *Cell,* 141, 432-45.

RAMANATHAN, Y., RAJPARA, S. M., REZA, S. M., LEES, E., SHUMAN, S., MATHEWS, M. B. & PE'ERY, T. 2001. Three RNA polymerase II carboxyl-terminal domain kinases display distinct substrate preferences. *J Biol Chem,* 276, 10913-20.

RANUNCOLO, S. M., GHOSH, S., HANOVER, J. A., HART, G. W. & LEWIS, B. A. 2012. Evidence of the involvement of O-GlcNAc-modified human RNA polymerase II CTD in transcription in vitro and in vivo. *J Biol Chem,* 287, 23549-61.

RATNASABAPATHY, R., SHELDON, M., JOHAL, L. & HERNANDEZ, N. 1990. The HIV-1 long terminal repeat contains an unusual element that induces the synthesis of short RNAs from various mRNA and snRNA promoters. *Genes Dev,* 4, 2061-74.

REINBERG, D. & ROEDER, R. G. 1987. Factors involved in specific transcription by mammalian RNA polymerase II. Transcription factor IIS stimulates elongation of RNA chains. *J Biol Chem,* 262, 3331-7.

REINES, D., CONAWAY, R. C. & CONAWAY, J. W. 1999. Mechanism and regulation of transcriptional elongation by RNA polymerase II. *Curr Opin Cell Biol,* 11, 342-6.

REYES-REYES, M. & HAMPSEY, M. 2007. Role for the Ssu72 C-terminal domain phosphatase in RNA polymerase II transcription elongation. *Mol Cell Biol,* 27, 926-36.

RICE, S. A. & DAVIDO, D. J. 2013. HSV-1 ICP22: hijacking host nuclear functions to enhance viral infection. *Future Microbiol,* 8, 311-21.

RICE, S. A., LONG, M. C., LAM, V., SCHAFFER, P. A. & SPENCER, C. A. 1995. Herpes simplex virus immediate-early protein ICP22 is required for viral modification of host RNA polymerase II and establishment of the normal viral transcription program. *J Virol,* 69, 5550-9.

RICE, S. A., LONG, M. C., LAM, V. & SPENCER, C. A. 1994. RNA polymerase II is aberrantly phosphorylated and localized to viral replication compartments following herpes simplex virus infection. *J Virol,* 68, 988-1001.

RIEDL, T. & EGLY, J. M. 2000. Phosphorylation in transcription: the CTD and more. *Gene Expr,* 9, 3-13.

ROEDER, R. G. 1996. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci,* 21, 327-35.

RUSKIN, B., ZAMORE, P. D. & GREEN, M. R. 1988. A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell,* 52, 207-19.

SADOWSKI, C. L., HENRY, R. W., LOBO, S. M. & HERNANDEZ, N. 1993. Targeting TBP to a non-TATA box cis-regulatory element: a TBP-containing complex activates transcription from snRNA promoters through the PSE. *Genes Dev,* 7, 1535-48.

SAUER, F., HANSEN, S. K. & TJIAN, R. 1995. DNA template and activator-coactivator requirements for transcriptional synergism by Drosophila bicoid. *Science,* 270, 1825-8.

SAUNDERS, A., CORE, L. J. & LIS, J. T. 2006. Breaking barriers to transcription elongation. *Nat Rev Mol Cell Biol,* 7, 557-67.

SCHAUFELE, F., GILMARTIN, G. M., BANNWARTH, W. & BIRNSTIEL, M. L. 1986. Compensatory mutations suggest that base-pairing with a small nuclear RNA is required to form the 3' end of H3 messenger RNA. *Nature,* 323, 777-81.

SCHRAMM, L., PENDERGRAST, P. S., SUN, Y. & HERNANDEZ, N. 2000. Different human TFIIIB activities direct RNA polymerase III transcription from TATA-containing and TATA-less promoters. *Genes Dev,* 14, 2650-63.

SCHROEDER, S. C., SCHWER, B., SHUMAN, S. & BENTLEY, D. 2000. Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes Dev,* 14, 2435-40.

SEARS, A. E., HALLIBURTON, I. W., MEIGNIER, B., SILVER, S. & ROIZMAN, B. 1985. Herpes simplex virus 1 mutant deleted in the alpha 22 gene: growth and gene expression in permissive and restrictive cells and establishment of latency in mice. *J Virol,* 55, 338-46.

SEARS, A. E. & ROIZMAN, B. 1990. Amplification by host cell factors of a sequence contained within the herpes simplex virus 1 genome. *Proc Natl Acad Sci U S A,* 87, 9441-4.

SENGUPTA, T., COHET, N., MORLE, F. & BIEKER, J. J. 2009. Distinct modes of gene regulation by a cell-specific transcriptional activator. *Proc Natl Acad Sci U S A,* 106, 4213-8.

SHANDILYA, J. & ROBERTS, S. G. 2012. The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. *Biochim Biophys Acta,* 1819, 391-400.

SHAO, H., REVACH, M., MOSHONOV, S., TZUMAN, Y., GAZIT, K., ALBECK, S., UNGER, T. & DIKSTEIN, R. 2005. Core promoter binding by histone-like TAF complexes. *Mol Cell Biol,* 25, 206-19.

SIMS, R. J., 3RD, BELOTSERKOVSKAYA, R. & REINBERG, D. 2004. Elongation by RNA polymerase II: the short and long of it. *Genes Dev,* 18, 2437-68.

SIMS, R. J., 3RD, ROJAS, L. A., BECK, D., BONASIO, R., SCHULLER, R., DRURY, W. J., 3RD, EICK, D. & REINBERG, D. 2011. The C-terminal domain of RNA polymerase II is modified by site-specific methylation. *Science,* 332, 99-103.

SMALE, S. T. & KADONAGA, J. T. 2003. The RNA polymerase II core promoter. *Annu Rev Biochem,* 72, 449-79.

SPENCER, C. A., DAHMUS, M. E. & RICE, S. A. 1997. Repression of host RNA polymerase II transcription by herpes simplex virus type 1. *J Virol,* 71, 2031-40.

STEINMETZ, E. J. & BROW, D. A. 2003. Ssu72 protein mediates both poly(A)-coupled and poly(A)-independent termination of RNA polymerase II transcription. *Mol Cell Biol,* 23, 6339-49.

STELZ, G., RUCKER, E., ROSORIUS, O., MEYER, G., STAUBER, R. H., SPATZ, M., EIBL, M. M. & HAUBER, J. 2002. Identification of two nuclear import signals in the alpha-gene product ICP22 of herpes simplex virus 1. *Virology,* 295, 360-70.

STILLER, J. W. & COOK, M. S. 2004. Functional unit of the RNA polymerase II C-terminal domain lies within heptapeptide pairs. *Eukaryot Cell,* 3, 735-40.

TAKAGAKI, Y., RYNER, L. C. & MANLEY, J. L. 1989. Four factors are required for 3'-end cleavage of pre-mRNAs. *Genes Dev,* 3, 1711-24.

TAMRAKAR, S., KAPASI, A. J. & SPECTOR, D. H. 2005. Human cytomegalovirus infection induces specific hyperphosphorylation of the carboxyl-terminal domain of the large subunit of RNA polymerase II that is associated with changes in the abundance, activity, and localization of cdk9 and cdk7. *J Virol,* 79, 15477-93.

TANAKA, M., GROSSNIKLAUS, U., HERR, W. & HERNANDEZ, N. 1988. Activation of the U2 snRNA promoter by the octamer motif defines a new class of RNA polymerase II enhancer elements. *Genes Dev,* 2, 1764-78.

TANESE, N., PUGH, B. F. & TJIAN, R. 1991. Coactivators for a proline-rich activator purified from the multisubunit human TFIID complex. *Genes Dev,* 5, 2212-24.

THOMAS, M. C. & CHIANG, C. M. 2006. The general transcription machinery and general cofactors. *Crit Rev Biochem Mol Biol,* 41, 105-78.

THUT, C. J., CHEN, J. L., KLEMM, R. & TJIAN, R. 1995. p53 transcriptional activation mediated by coactivators TAFII40 and TAFII60. *Science,* 267, 100-4.

TIETJEN, J. R., ZHANG, D. W., RODRIGUEZ-MOLINA, J. B., WHITE, B. E., AKHTAR, M. S., HEIDEMANN, M., LI, X., CHAPMAN, R. D., SHOKAT, K., KELES, S., EICK, D. & ANSARI, A. Z. 2010. Chemical-genomic dissection of the CTD code. *Nat Struct Mol Biol,* 17, 1154-61.

TORA, L. 2002. A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription. *Genes Dev,* 16, 673-5.

UGUEN, P. & MURPHY, S. 2003. The 3' ends of human pre-snRNAs are produced by RNA polymerase II CTD-dependent RNA processing. *EMBO J,* 22, 4544-54.

VAN ARSDELL, S. W. & WEINER, A. M. 1984. Human genes for U2 small nuclear RNA are tandemly repeated. *Mol Cell Biol,* 4, 492-9.

VERMEULEN, M., MULDER, K. W., DENISSOV, S., PIJNAPPEL, W. W., VAN SCHAIK, F. M., VARIER, R. A., BALTISSEN, M. P., STUNNENBERG, H. G., MANN, M. & TIMMERS, H. T. 2007. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell,* 131, 58-69.

VERRIJZER, C. P. & TJIAN, R. 1996. TAFs mediate transcriptional activation and promoter selectivity. *Trends Biochem Sci,* 21, 338-42.

WAHL, M. C., WILL, C. L. & LUHRMANN, R. 2009. The spliceosome: design principles of a dynamic RNP machine. *Cell,* 136, 701-18.

WASHINGTON, K., AMMOSOVA, T., BEULLENS, M., JEREBTSOVA, M., KUMAR, A., BOLLEN, M. & NEKHAI, S. 2002. Protein phosphatase-1 dephosphorylates the C-terminal domain of RNA polymerase-II. *J Biol Chem,* 277, 40442-8.

WEIR, J. P. 2001. Regulation of herpes simplex virus gene expression. *Gene,* 271, 117-30.

WEISSMAN, J. D., BROWN, J. A., HOWCROFT, T. K., HWANG, J., CHAWLA, A., ROCHE, P. A., SCHILTZ, L., NAKATANI, Y. & SINGER, D. S. 1998. HIV-1 tat binds TAFII250 and represses TAFII250-dependent transcription of major histocompatibility class I genes. *Proc Natl Acad Sci U S A,* 95, 11601-6.

WESTIN, G., LUND, E., MURPHY, J. T., PETTERSSON, U. & DAHLBERG, J. E. 1984. Human U2 and U1 RNA genes use similar transcription signals. *EMBO J,* 3, 3295-301.

WRIGHT, K. J., MARR, M. T., 2ND & TJIAN, R. 2006. TAF4 nucleates a core subcomplex of TFIID and mediates activated transcription from a TATA-less promoter. *Proc Natl Acad Sci U S A,* 103, 12347-52.

WU, J. A. & MANLEY, J. L. 1991. Base pairing between U2 and U6 snRNAs is necessary for splicing of a mammalian pre-mRNA. *Nature,* 352, 818-21.

WYSOCKA, J. & HERR, W. 2003. The herpes simplex virus VP16-induced complex: the makings of a regulatory switch. *Trends Biochem Sci,* 28, 294-304.

XIANG, K., MANLEY, J. L. & TONG, L. 2012. The yeast regulator of transcription protein Rtr1 lacks an active site and phosphatase activity. *Nat Commun,* 3, 946.

YAMADA, T., YAMAGUCHI, Y., INUKAI, N., OKAMOTO, S., MURA, T. & HANDA, H. 2006. P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Mol Cell,* 21, 227-37.

YAMAGUCHI, Y., TAKAGI, T., WADA, T., YANO, K., FURUYA, A., SUGIMOTO, S., HASEGAWA, J. & HANDA, H. 1999. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell,* 97, 41-51.

YEO, M., LIN, P. S., DAHMUS, M. E. & GILL, G. N. 2003. A novel RNA polymerase II C-terminal domain phosphatase that preferentially dephosphorylates serine 5. *J Biol Chem,* 278, 26078-85.

YOON, J. B., MURPHY, S., BAI, L., WANG, Z. & ROEDER, R. G. 1995. Proximal sequence element-binding transcription factor (PTF) is a multisubunit complex required for transcription of both RNA polymerase II- and RNA polymerase III-dependent small nuclear RNA genes. *Mol Cell Biol,* 15, 2019-27.

ZABOROWSKA, J., TAYLOR, A. & MURPHY, S. 2012. A novel TBP-TAF complex on RNA polymerase II-transcribed snRNA genes. *Transcription,* 3, 92-104.

ZHANG, D. W., MOSLEY, A. L., RAMISETTY, S. R., RODRIGUEZ-MOLINA, J. B., WASHBURN, M. P. & ANSARI, A. Z. 2012a. Ssu72 phosphatase-dependent erasure of phospho-Ser7 marks on the RNA polymerase II C-terminal domain is essential for viability and transcription termination. *J Biol Chem,* 287, 8541-51.

ZHANG, J. & CORDEN, J. L. 1991. Identification of phosphorylation sites in the repetitive carboxyl-terminal domain of the mouse RNA polymerase II largest subunit. *J Biol Chem,* 266, 2290-6.

ZHANG, M., WANG, X. J., CHEN, X., BOWMAN, M. E., LUO, Y., NOEL, J. P., ELLINGTON, A. D., ETZKORN, F. A. & ZHANG, Y. 2012b. Structural and kinetic analysis of prolyl-isomerization/phosphorylation cross-talk in the CTD code. *ACS Chem Biol,* 7, 1462-70.

ZHOU, M., NEKHAI, S., BHARUCHA, D. C., KUMAR, A., GE, H., PRICE, D. H., EGLY, J. M. & BRADY, J. N. 2001. TFIIH inhibits CDK9 phosphorylation during human immunodeficiency virus type 1 transcription. *J Biol Chem,* 276, 44633-40.

ZHUANG, Y. & WEINER, A. M. 1989. A compensatory base change in human U2 snRNA can suppress a branch site mutation. *Genes Dev,* 3, 1545-52.

ZURITA, M. & MERINO, C. 2003. The transcriptional complexity of the TFIIH complex. *Trends Genet,* 19, 578-84.