ROLE OF R-LOOPS IN PAUSE-DEPENDENT TRANSCRIPTIONAL TERMINATION OF RNA POLYMERASE II

Konstantina Skourti-Stathaki

University of Oxford
Sir William Dunn School of Pathology
and
St. Catherine's College

Thesis submitted for the degree of Doctor of Philosophy at the University of Oxford

Trinity Term 2012

To my family

Abstract

Transcription termination of RNA polymerase II (Pol II) in mammals requires a functional poly(A) signal and either downstream pause sites or co-transcriptional cleavage (CoTC) sequences together with 3'transcript degradation by the nuclear 5'-3' exonuclease Xrn2. However the molecular mechanism of pause-dependent transcriptional termination is not yet fully understood. This thesis investigates the molecular role of R-loop structures in pause-dependent transcriptional termination of mammalian genes. The results described in Chapters 3 and 4 indicate that nascent transcripts form RNA/DNA hybrid structures (R-loops) behind elongating Pol II and are especially prevalent over G-rich pause sites positioned downstream of gene poly(A) signals. Senataxin, a helicase protein and the human homologue of the yeast Sen1, acts to resolve these R-loop structures and by so doing allows access of Xrn2 at 3' cleavage poly(A) sites. This ultimately leads to efficient Pol II termination. In effect R-loops formed over G-rich pause sites, followed by their resolution by senataxin, are required for efficient pause-dependent transcriptional termination. In addition to this, the 3' end processing factor, Pcf11 is also involved in this process. Experiments presented in the final part of this study reveal a link between R-loops and RNAi-dependent H3K9me2 formation over G-rich termination regions. Overall my results suggest that R-loop structures and the H3K9me2 repressive mark over pause regions are important features of Pol II pause-dependent transcriptional termination of mammalian genes.

Acknowledgements

I would first like to thank Nick Proudfoot for giving me the opportunity and freedom to carry out this project and for guiding my research throughout these years. I will always admire his insight and inspiration and I will be grateful for his help and support in all my research efforts, both ‘lab’ and general. I also wish to thank Natasha Gromak for her genuine help, support and supervision throughout the first years of my DPhil research. I have learned a lot working with her and for this I will always be grateful. A big thanks go to all the members for the NJP lab, past and present Proudfeet, for making the days of these years most enjoyable.

I would also like to thank my friends in Oxford for all these evenings of fun, discussions, encouragement and support. Special thanks also go to my friends in Greece for their constant support and encouragement despite the distance.

Finally, I would truly like to thank my family, especially my mum, dad and Eirini for always being here for me. I know I could not have done it without you. Thank you for teaching me not to give up even when things turn tough. Maybe now I can say that I am not studying anymore after all these years!

TABLE OF CONTENTS

ABSTRACT **i**

ACKNOWLEDGEMENTS **ii**

TABLE OF CONTENTS **iii**

ABBREVIATIONS **vi**

1. INTRODUCTION **1**

1.1 Co-transcriptional 3' end pre-mRNA processing 3

1.2 Pol II CTD and 3' end processing 8

1.3 Pol II transcriptional termination 13

1.4 R-loops 23

1.5 RNAi-dependent heterochromatin formation and transcriptional gene silencing 33

1.6 Everything is 'interconnected'-Experimental aims 42

2. MATERIALS AND METHODS **44**

2.1 Reagents 45

2.2 Enzymes 46

2.3 Antibodies 47

2.4 Primers 48

2.5 Tissue Culture 51

2.5a Growth and maintenance 51

2.5b Plasmid transfections 51

2.5c RNA interference 52

2.6 Mammalian RNA isolation 52

2.7 Polymerase Chain Reaction (PCR) 53

2.7a Standard PCR 53

2.7b Quantitative Real-Time PCR 53

2.8 Steady-state RNA analysis 54

2.9 Nascent RNA analysis 54

2.9a Nuclear-Run-On analysis (NRO) 54

2.9b Bromo-Uridine Triphosphate (Br-UTP) NRO analysis 58

2.10 Chromatin Immunoprecipitation (ChIP) analysis 59
2.11 DNA Immunoprecipitation (DIP) analysis 62
2.12 Protein extraction and Western blot analysis 63
2.13 Co-Immunoprecipitation (Co-IP) analysis 64
3. THE ROLE OF HUMAN SENATAXIN PROTEIN AND R-LOOP STRUCTURES IN PAUSE-DEPENDENT TRANSCRIPTIONAL TERMINATION OF TRANSFECTED GENE CONSTRUCTS 66
3.1 Introduction 67
3.2 Senataxin promotes transcriptional termination on transfected gene constructs 69
3.3 R-loops form over transfected gene constructs 71
3.4 RNase H1 over-expression resolves R-loops and causes transcriptional read-through *in vivo* 76
3.5 Discussion 80
4. THE MECHANISM OF R-LOOP DEPENDENT TRANSCRIPTIONAL TERMINATION IN ENDOGENOUS HUMAN GENES: THE ROLE OF SENATAXIN AND PCF11 83
4.1 Introduction 84
4.2 Senataxin is involved in transcriptional initiation and termination of endogenous human β-actin gene 85
4.3 R-loops are formed over endogenous β-actin gene 90
4.4 Senataxin is required for Xrn2 recruitment to promote efficient termination 91
4.5 Senataxin interacts with Xrn2 *in vivo* 91
4.6 Senataxin depletion affects Pcf11 recruitment and Ser2P levels 92
4.7 Pcf11 depletion affects senataxin recruitment and R-loop distribution profile of endogenous β-actin gene 96
4.8 Senataxin binds endogenous γ-actin gene but is not involved in its transcriptional termination 98
4.9 R-loops form over endogenous human γ-actin gene 101
4.10 Discussion 102

5. THE ROLE OF HETEROCHROMATIN IN R-LOOP DEPENDENT TERMINATION PAUSING OF MAMMALIAN GENES 107

5.1 Introduction 108

5.2 The H3K9me2 mark is present at the termination region of endogenous human β-actin and γ-actin genes 110

5.3 Antisense transcription, Dicer and Ago1 are detected over the pause element of human β-actin gene 113

5.4 Loss of R-loops affects antisense transcription and Dicer recruitment over the pause element of human β-actin, but it does not affect H3K9me2 levels 117

5.5 Ago1 protein and antisense transcription are detected over the termination region of mouse β-actin 120

5.6 The H3K9me2 mark is present over the termination region of mouse β-actin and it is reduced to a 50% level in Ago2 KO cells 122

5.7 Depletion of Ago2 enhances Ago1 recruitment to the mouse β-actin gene 125

5.8 R-loops are formed over the mouse β-actin gene. Depletion of Ago2 partially affects their distribution 127

5.9 Loss of R-loops and partial loss of the H3K9me2 mark affect transcriptional termination of endogenous mouse β-actin gene 129

5.10 The H3K9me3 mark is present at the 5' end of mouse β-actin gene and is not affected by Ago2 depletion 132

5.11 Depletion of Suv39h1/h2 methyltransferases does not affect H3K9me2 and H3K9me3 of endogenous mouse β-actin gene 135

5.12 Discussion 137

6. GENERAL DISCUSSION 144

7. REFERENCES 158

Abbreviations

A	Adenine
AID	Activation induced deaminase
ALS4	Amyotrophic lateral sclerosis type 4
AOAII	Ataxia ocular apraxia type II
Ago1	Argonaute 1
Ago2	Argonaute 2
ATM	Ataxia Telangiectasia Mutated
ATP	Adenosine Trisphosphate
BER	Base excision repair
bp	Base pair
Br	bromo
Br-UTP NRO	Bromo-Uridine TriPhosphate Nuclear Run-On
C	Cytosine
Cdk	Cyclin dependent kinase
cDNA	complementary DNA
C. elegans	*Caenorhabditis elegans*
Chp1	Chromodomain protein 1
CFI	Cleavage Factor I
CFIm	mammalian Cleavage Factor I
ChIP	Chromatin Immunoprecipitation
CID	CTD Interaction Domain
Co-IP	Co-ImmunoPrecipitation

CoTC	Co-Transcriptional Cleavage
CPF	Cleavage and Polyadenylation Factor
CPSF	Cleavage and Polyadenylation Specificity Factor
CstF	Cleavage stimulatory Factor
CSR	Class Switch Recombination
CTD	Carboxy-Terminal Domain
CTP	Cytosine Triphosphate
CUTs	Cryptic Unstable Transcripts
dC	deoxy-cytosine
DIP	DNA ImmunPrecipitation
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	Deoxyribonucleotide triphosphate
DMEM	Dulbecco's Modified Eagle Medium
DSB	Double-Strand Break
DSE	Downstream Sequence Element
dsDNA	double stranded DNA
dsRNA	double stranded RNA
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic Acid (disodium salt)
G	Guanine
g	g-force
GFP	Green Fluorescent Protein

GTP	Guanosine Triphosphate
H3	Histone 3
H3K9	Histone 3 Lysine 9
HEPES	N-2-Hydroxyethylpiparizine-N'-2-ethanesulphonic acid
HIV	Human Immunodeficiency virus
HKMT	Histone lysine methyl transferase
hnRNP	heterogeneous nuclear Ribonucleo Protein
hr/hrs	hour/hours
HP1	Heterochromatin protein 1
Ig	Immunoglobulin
IGHMBP2	Immunoglobulin μ-Binding Protein 2
IP	Immunoprecipitation
kb	kilobase
KD	knock-down
kDa	kiloDalton
KO	knock-out
min(s)	minute(s)
miRNA	microRNA
ml	millilitre
mRNA	messenger RNA
mRNPs	messenger ribonucleotide proteins
MSA	Mouse Serum Albumin
Nab	Nuclear polyadenylated RNA binding protein

ncRNA	non coding RNA
NELF	Negative Elongation Factor
NLS	Nuclear Localisation Signal
Nrd1	Nuclear pre-mRNA down regulation
NRO	Nuclear-Run-ON
nt(s)	nucleotide(s)
NTP	Nucleotide Triphosphate
ORF	Open Reading Frame
PABP	Poly(A) Binding Protein
PAGE	PolyAcrylamide Gel Electrophoresis
PAP	Poly(A) Polymerase
PBS	Phosphate Buffered Saline
Pcf11	Polyadenylation cleavage factor 11
PCR	Polymerase Chain Reaction
PIPES	Piperazine-N, N'-bis 2-ethanesulphonic acid
PMSF	Phenylmethylsulfonyl Fluoride
Pol	Polymerase
Poly(A)	Polyadenylation
PROMPTs	Promoter upstream Transcripts
PSF	Protein associated Splicing Factor
pTEFb	Positive Transcription Elongation Factor b
PTGS	Post-Transcriptional Gene Silencing
qPCR	quantitative PCR

Rai1	Rat interacting protein 1
rDNA	ribosomal DNA
RDRC	RNA-Dependent RNA polymerase Complex
rG	ribo-guanine
RISC	RNA-Induced Silencing Complex
RITS	RNA-Induced Transcriptional Silencing
RNA	Ribonucleic Acid
RNAi	RNA interference
RNase	Ribonuclease
rNTPs	Ribonucleotide Trisphosphates
rRNA	ribosomal RNA
RT	Reverse Transcription
S. cerevisiae	*Saccharomyces cerevisiae*
SD	Standard Deviation
SDS	Sodium Dodecyl Sulfate
Senataxin	SETX
Ser2P	CTD Serine 2 phosphorylated
Ser5P	CTD Serine 5 phosphorylated
Ser7P	CTD Serine 7 phosphorylated
SET	Su(var)3-9-Enhancer of *zeste*-Trithorax
siRNA	small interfering RNA
SMN	Spinal Muscular Atrophy
snoRNA	small nucleolar RNA

snRNA	small nuclear RNA
S. pombe	*Schizosaccharomyces pombe*
ssDNA	single stranded DNA
Ssu72	Suppressor of Sua7, gene 2
Suv39h1	Suppressor of variegation 3-9 homologue 1
Suv39h2	Suppressor of variegation 3-9 homologue 2
T	Thymine
TAM	Transcription Associated Mutagenesis
TAR	Transcription Associated Recombination
TAT	Trans-activator of transcription
TGS	Transcriptional Gene Silencing
Top 1	Topoisomerase I
TREX	TRancription-EXport complex
tRNA	transfer RNA
TSS	Transcription Start Site
Tyr1P	CTD Tyrosine 1 phosphorylated
U	Uracil
USE	Upstream Sequence Element
UTP	Uridine Triphosphate
UTR	Untranslated Region
WT	Wild Type

CHAPTER 1

Introduction

54 years have passed since Francis Crick proposed the central dogma of molecular biology in 1958. The dogma is a framework that describes the sequential flow of information that leads to gene expression in living organisms. In other words the central dogma states that 'DNA makes RNA, RNA makes protein, protein makes us' as often expressed.

Since then, there has been an extensive investigation of the steps involved in this flow of information. An important conclusion is that all steps in gene expression are interconnected and co-ordinately regulated (Moore and Proudfoot, 2009). Regulation of gene expression is essential for normal cellular development and often occurs at the level of transcription (Kuehner et al. 2011). In eukaryotes transcription is performed by three functionally and structurally related DNA-dependent RNA polymerases (Cramer et al. 2008). RNA polymerase I (Pol I) transcribes ribosomal RNAs (rRNAs). RNA polymerase II (Pol II) is responsible for the transcription of all protein-coding genes and many non-coding RNAs such as spliceosomal small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), microRNA (miRNA) precursors and cryptic unstable transcripts (CUTs). RNA polymerase III (Pol III) transcribes non-coding RNAs including transfer RNAs (tRNAs), 5S rRNA, and U6 spliceosomal snRNA. Transcription is a highly controlled process which involves a large number of specific factors (Richard and Manley, 2009). It can be divided in three main steps: initiation, elongation and termination. While initiation and elongation (or how transcription starts and proceeds) have been extensively studied (Sims et al. 2004), there is comparatively scant information for the ultimate step in the transcription cycle, termination.

This thesis focuses on the molecular process of Pol II transcriptional termination in mammalian cells and particularly on the mechanism of pause-dependent termination. My thesis plan was originally designed to further investigate the mechanism of Pol II pausing in transcriptional termination of human genes and especially the role of human senataxin protein in this process. However throughout the course of my studies it became evident that other processes and factors are involved and that transcriptional termination is intricately tied to other stages of gene expression. In the following pages I have tried to cover the most important concepts that have influenced my thinking during my graduate research work.

1.1 Co-transcriptional 3' end pre-mRNA processing

Pre-mRNA processing consists of 5' end capping, splicing and 3' end processing. 3' end processing of Pol II transcripts occurs through poly(A) site cleavage and coupled polyadenylation. For the purpose of this study, this introduction will focus only on the poly(A)-dependent 3' end processing.

3' end processing is an important process in the maturation of all mRNAs and is coupled to splicing and transcriptional termination (Hirose and Manley, 2000; Proudfoot et al. 2002; Buratowski, 2005; Proudfoot, 2011). The central sequence element required is the poly(A) signal and was firstly identified in human cells where it consists of the highly conserved sequence motif, AAUAAA (Proudfoot and Brownlee, 1976). However other sequence motifs have been shown to be necessary for efficient mRNA 3' end cleavage and polyadenylation, such as the cleavage site located ~15-30 nucleotides downstream the poly(A) signal which usually contains the CA dinucleotide (Sheets et al.

1990) and a less conserved GU/U rich element (or downstream sequence element, DSE) positioned ~20-30 nucleotides downstream the cleavage site (Gil and Proudfoot, 1984; Gil and Proudfoot, 1987; McLauchlan et al. 1985). In some cases a U-rich sequence element (or upstream sequence element, USE) located immediately upstream the poly(A) signal can also facilitate 3' end processing (Danckwardt et al. 2007;Venkataraman et al. 2005; Moreira et al. 1995).

In mammals, where poly(A) signals are better defined, biochemical studies from the Manley and Keller laboratories have identified two major protein complexes which promote 3' end cleavage and polyadenylation: the cleavage and polyadenylation specificity factor (CPSF) and the cleavage stimulator factor (CstF) (Figure 1.1). *In vitro* systems have shown that these two protein complexes recognise the poly(A) signal and the DSE element which in turn promotes RNA cleavage between these two *cis* elements. CPSF consists of five subunits: CPSF-30, CPSF-73, CPSF-100, CPSF-160 and Fip1 (Murthy and Manley, 1992; Kaufmann et al. 2004), all of which promote cleavage and polyadenylation (Christofori and Keller, 1988). The endonucleolytic cleavage, before polyadenylation occurs, is catalysed by CPSF-73 (Ryan et al. 2004; Mandel et al. 2006). CPSF-160 recognises and directly binds the poly(A) signal (Keller et al. 1991; Murthy and Manley, 1995). CstF complex contains three subunits including CstF-50, CstF-77 and CstF-64. The latter subunit directly recognises the DSE element. CstF complex is required for 3' end cleavage but not for polyadenylation (Takagaki et al. 1989). The connecting bridge between the two major complexes is thought to be the protein symplekin which appears to function as a scaffold between CPSF-160 and CstF-64 (Hofmann et al. 2002; Takagaki and Manley, 2000). Other factors were also shown to be

required for in vitro 3' end processing, including the mammalian cleavage factors Im and IIm (CFIm/IIm).

CFIIm complex contains the highly conserved Pcf11 and Clp1 proteins. Clp1 protein is thought to interact with CFIm and CPSF complexes (de Vries et al. 2000). The poly(A) polymerase (PAP) specialises in polyadenylating the cleaved pre-mRNA, which in turn is protected from exonucleolytic degradation by association with the poly(A) binding protein (PABP). PABP is also required for correct and efficient poly(A) tail synthesis (Minvielle-Sebastia et al. 1997; Colgan and Manley, 1997).

3' end pre-mRNA processing, a seemingly simple two step mechanism including an endonucleolytic cleavage followed by poly(A) tail synthesis, appears to be a rather complex and cooperative process (Shi et al. 2009). For example, CstF-64 binds the DSE sequence element, while CstF-77 binds CPSF-160 and PAP. The CstF-50 subunit is thought to bind the C-terminal domain (CTD) of the Pol II largest subunit. CPSF-73 is responsible for pre-mRNA cleavage, while CPSF-30 and CPSF-100 maintain the interaction of the complex with Pol II. CPSF-160 directly recognises the poly(A) signal and binds CstF-77 and CPSF-100. CPSF-160 also interacts with Fip1, CFIm and PAP. The scaffold protein symplekin further reinforces these interactions. All these proteins interact to form a complex network between themselves, the transcribing Pol II and the pre-mRNA (Gross and Moore, 2001; Murthy and Manley, 1995; Shatkin and Manley, 2000; Takagaki et al. 1992; Takagaki and Manley, 1997; Zhao et al. 1999, Millevoi and Vagner, 2010; Mandel et al. 2008).

In *S. cerevisiae*, 3'end pre-mRNA cleavage and polyadenylation seem to be more variable than in mammalian genes. In essence the *cis* elements required for 3' processing

are not as well defined as in mammals. However the mechanism itself is highly conserved and most of the key players have mammalian homologues as well (Figure 1.1). The initial step of pre-mRNA 3' end formation, the pre-mRNA cleavage, requires the cleavage factor (CFI) and the cleavage and polyadenylation factor (CPF). The first complex contains Rna14 (the yeast homologue of CstF-77), Rna15 (CstF-64), Pcf11, Clp1 and Hrp1, while CPF complex contains homologues to the mammalian CPSF complex. Subsequent pre-mRNA polyadenylation is mediated by the poly(A) binding protein (Pab1) and the poly(A) polymerase (Pap1) (Minvielle-Sebastia et al. 1994; 1997; McCracken et al. 1997; Kessler et al. 1996).

Given the complexity of the interactions between different protein complexes, coupled with the remarkable complexity of the protein complexes themselves, it is perhaps not surprising that transcription and 3' end processing of the nascent RNA are coupled. Indeed this could provide a mechanism to enhance the fidelity and efficiency of this process, as well as its regulation. The question that arises then is: how is this coupling facilitated? To answer this, I will discuss in the next section the role and importance of the Pol II CTD.

1.2 Pol II CTD and 3' end pre-mRNA processing

The Pol II CTD contains multiple heptad repeats with the consensus sequence Tyr_1-Ser_2-Pro_3-Thr_4-Ser_5-Pro_6-Ser_7, varying in number from 26 in yeast to 52 in vertebrates. This domain can undergo extensive post-translational modifications and plays important roles at all steps of the transcription process. Particularly the phosphorylation state of the CTD has been documented to enhance and modulate its activities in transcription and processing of the nascent RNA (Phatnani and Greenleaf, 2006; Egloff and Murphy, 2008; Buratowski, 2009; Perales and Bentley, 2009).

It is however striking that such a simple protein structure, YSPTSPS repeats, can orchestrate multiple and complex events during the transcription process. In addition to this, many proteins are thought to bind to this domain. How does such a simple structure interact with so many targets? The answer is a term, introduced by Buratowski in 2003, the 'CTD code'. Instead of carrying all players throughout the transcription cycle, the CTD undergoes different phosphorylation patterns which subsequently enable it to interact with each factor at the appropriate time (Buratowski, 2003; Egloff and Murphy, 2008; Buratowski, 2009).

Pol II initially enters the transcription cycle with an unphosphorylated CTD (Pol IIA). Serine 5 is then phosphorylated by Cdk7 (Kin28 in yeast), a component of TFIIH general transcription factor. A generalised model depicts that Serine 5 phosphorylation is enriched at the 5' ends of genes, and peaks around transcription start sites, to facilitate promote release and recruitment of capping factors (Komarnitsky et al. 2000; Schroeder et al. 2000). In addition to this, Serine 5 phosphorylation of the CTD also recruits the histone methyltransferase Set1 that is responsible for trimethylation of Lysine 4 on

histone H3 (H3K4me3). This modification is seen near promoters of active genes (Hampsey and Reinberg, 2003). As Pol II elongates, Serine 5 phosphorylation levels drop as Serine 5 is gradually dephosphorylated. The yeast Ssu72 protein, a component of the cleavage and polyadenylation factor (CPF), was shown to be a CTD phosphatase with preference for phosphorylated Serine 5 (Ganem et al. 2003; Krishnamurthy et al. 2004). Furthermore depletion of Ssu72 leads to increased levels of phosphorylated Serine 5 towards the 3' ends of genes (Bataille et al. 2012). Rtr1 (RPAP2 in humans) has also been shown to dephosphorylate Serine 5 *in vitro* and *in vivo* (Mosley et al. 2009; Egloff et al. 2011). However the structure of Rtr1 does not reveal an active site and enzymatic assays with Rtr1 or RPAP2 and CTD substrates did not show any activity (Xiang et al. 2012). The specificity of another phosphatase, the yeast Fcp1, is also unclear as it is able to dephosphorylate both Serine 5 and Serine 2 (Lin et al. 2002), although it appears to prefer Serine 2 (Cho et al. 2001; Hausmann and Shuman, 2002; Ghosh et al. 2008).

It should also be noted at this point that the CTD can also be phosphorylated on Serine 7 by Cdk7 (Egloff et al. 2007; Akhtar et al. 2009; Glover-Cutter et al. 2009; Kim et al. 2009; Boeing et al. 2010). Chromatin immunoprecipitation (ChIP) analysis in yeast showed a similar distribution profile of phosphorylated Serine 5 and Serine 7 (Glover-Cutter et al. 2009; Kim et al. 2009), indicating that Serine 5 and Serine 7 are both phosphorylated at the beginning of the transcription cycle. However studies in human cells showed that higher levels of Serine 7 phosphorylation were detected towards the 3' region as compared to the promoter region (Chapman et al. 2007). Overall it could be concluded that profiles of Serine 7 phosphorylation vary with some genes having discrete peaks at the 5' and/or 3' end (Kim et al. 2010).

As Pol II elongates towards the 3' end, Serine 5 phosphorylation is gradually removed and Serine 2 is increasingly phosphorylated. Usually Serine 2 phosphorylation begins to saturate at ~600 nucleotides downstream the transcription start site (TSS) and gradually decreases downstream the poly(A) signal (Mayer et al. 2010). The CTD of elongating Pol II is phosphorylated by the positive elongation factor pTEFb, containing the kinase Cdk9 (Ctk1 in yeast) and cyclin T. pTEFb also phosphorylates two elongation factors, DRB sensitivity inducing factor (DSIF) and negative elongation factor (NELF). After transcription initiation, the two factors pause elongating Pol II downstream the TSS (Yamaguchi et al. 1999; Wu et al. 2003). Elongation proceeds after pTEFb phosphorylates Serine 2 and the two factors which in turn causes their dissociation from Pol II complex (Renner et al. 2001; Fujinaga et al. 2004; Cheng and Price, 2007). As Pol II transcribes towards the poly(A) signal, Serine 2 phosphorylated CTD facilitates recruitment and/or stabilisation of polyadenylation factors. This couples transcription and mRNA 3' end formation. Initial studies showed that deletion of the CTD has no effects on transcription of a reporter plasmid but it was required for efficient 3' processing (McCracken et al. 1997; Hirose and Manley, 1998). The CTD requirement in proper 3' end processing is also conserved in yeast (Licatalosi et al. 2002). A number of 3' processing factors interact with the CTD, indicating its role as a scaffolding factor. For example, Pcf11 binds the CTD via its N'-terminal interaction domain (CID) in a way that is enhanced by Serine 2 phosphorylation (Barilla et al. 2001; Licatalosi et al. 2002; Meinhart and Cramer, 2004). Genome-wide ChIP experiments showed that enrichment of 3' end processing factors followed Serine 2 phosphorylation peaks, further suggesting that the poly(A) signal in the nascent RNA and CTD Serine 2 phosphorylation contribute

to the recruitment of the polyadenylation complex (Kim et al. 2010; Mayer et al. 2010; 2012). However one question still remains: since Serine 2 phosphorylation is also present during the elongation state, why 3' end processing and termination factors are only recruited when Pol II transcribes towards the 3' region? In other words, how does the elongation-termination transition occur?

Recently, another modification, the phosphorylation of Tyrosine 1, was shown to play a role in productive elongation (Mayer et al. 2012). In essence Mayer and colleagues have shown that Tyrosine 1 CTD phosphorylation blocks the binding of termination factors to the CTD, including the 3' processing factor Pcf11, whereas it facilitates the binding of the Spt6 elongation factor. Interestingly the levels of Serine 2 phosphorylation correlate with phosphorylated Tyrosine 1 across the body of the gene but phosphorylated Tyrosine 1 levels drop sharply immediately after the poly(A) signal (Mayer et al 2012). A simplified explanation for this could be that across the body of the gene Tyrosine 1 phosphorylation masks Serine 2 phosphorylation enabling the recruitment of only elongation factors. As Pol II approaches the poly(A) signal, a Tyrosine 1 phosphatase is recruited while Serine 2 phosphorylation still remains. This could then be the binding platform for 3' processing and termination factors which in turn sets the elongation-termination transition (Mayer et al. 2012).

1.3 Pol II transcriptional termination

Stopping Pol II is not an easy task. Once in processive elongation, Pol II elongates at 4.3kbp/min (70bp/sec) (Darzacq et al. 2007) over a diverse sequence landscape that may extend over 1Mbp in vertebrates. This predicts that the mechanisms leading to efficient Pol II termination are complex and remain as yet incompletely understood. For years, termination, the cessation of RNA synthesis and the release of Pol II from the DNA template, remained the least studied step of the transcription cycle. Since the discovery that a functional poly(A) signal is required for termination (Proudfoot, 1989), considerable progress has been made in understanding this ultimate stage of transcription.

Pol II transcriptional termination is a crucial process for the survival of probably all organisms. Failure to terminate may result in read-through transcription. In addition to this, recycling of the Pol II nuclear pool for re-initiation may be limited due to failure of its release from the DNA template. Finally termination has been shown to be important for optimal protein production (West and Proudfoot, 2009). These observations reflect the need to not only terminate Pol II transcription successfully, but also to tightly regulate this process. In this section I will discuss the main concepts of Pol II termination always having in mind the interplay between transcription and 3' end processing.

As mentioned above, an early insight into the mechanism of Pol II termination was the discovery that a functional poly(A) signal is required for efficient termination in mammalian cells (Proudfoot, 1989). This was the first indication that 3' end processing and termination are coupled processes. Subsequent experiments mainly in *S.cerevisiae* revealed that a number of cleavage factors are also required for efficient termination (Birse et al. 1998; Dichtl et al. 2002; Ganem et al. 2003; Nedea et al. 2003; Kim et al.

2010; Zhang et al. 2012). Consistent with this, in mammalian cells CPSF-73 and CstF-64 have been shown to be associated with the 3' end of genes (Swinburne et al. 2006; Glover-Cutter et al. 2008).

Two models were initially proposed to explain the connection between 3' end processing and transcriptional termination. The first model, known as the allosteric/anti-terminator model, suggests that Pol II termination occurs in response to conformational changes in the Pol II complex due to transcription through the poly(A) site (Logan et al. 1987). The second 'torpedo' model, proposes that Pol II termination occurs in response to degradation of the 3' flanking transcript by a 5'-3' exonuclease (Connelly and Manley, 1988). This latter model has been supported by the demonstration that such a nuclease is indeed required for efficient termination, Xrn2 in humans (West et al. 2004) and Rat1 in yeast (Kim et al. 2004). However, recent evidence shows that the termination mechanism reflects a combination of both models (Luo et al. 2006; Kaneko et al. 2007b; West et al. 2008).

In yeast, Rat1 is activated by its co-factor, Rai1 (Stevens and Poole, 1995; Xiang et al. 2009) and both are associated with the 3'end of genes (Kim et al. 2004). However deletion of *CTK1* (the kinase that phosphorylates the Serine 2 of the CTD) does not affect their recruitment suggesting that the Rat1/Rai1 complex is associated with gene 3' regions in a Serine 2 phosphorylation-independent manner (Kim et al. 2004). This led to the hypothesis that recruitment of Rat1 might be mediated through 3'-end processing factors. Indeed, the Serine 2 phosphorylation CTD binding protein, Rtt103, co-purifies with Rat1 and is also associated with the 3' ends of genes (Kim et al. 2004). Possibly Rtt103 plays a role in termination as the *C. elegans* homologue was identified as

termination factor through an RNAi genome-wide screen (Cui et al. 2008). In addition to this, another CID-containing protein, the cleavage and polyadenylation factor Pcf11, is also required for Rat1 association with active genes (Luo et al. 2006). Consistent with this observation, genome-wide profile of yeast Pcf11 overlaps with that of Rat1 (Kim et al. 2010).

Pcf11 is perhaps the clearest example of a 3' end processing factor which also helps Pol II termination, as well as mediating 3' end processing. Depletion or mutation of Pcf11 leads to 3' end processing and termination defects (Licatalosi et al. 2002; Sadowski et al. 2003; Meinhart and Cramer, 2004; Zhang et al. 2005). Pcf11 contains a CID domain and has been shown to preferentially bind Serine 2 phosphorylated CTD (Barilla et al. 2001; Licatalosi et al. 2002; Ahn et al. 2004; Lunde et al. 2010). *In vitro* studies suggested that both yeast and fly Pcf11 can disassociate Pol II and nascent RNA from the DNA template by bridging the CTD to RNA (Zhang et al. 2005; Zhang and Gilmour, 2006). However the cleavage activity, and not the CTD binding ability, was shown to be required for termination (Kim et al. 2006). In humans Pcf11 is required to stimulate transcript degradation by Xrn2 *in vivo* (West and Proudfoot, 2008) and is implicated in the cross-talk between 3' end processing and transcriptional initiation (Mapendano et al. 2010). Overall it could be concluded that although Pcf11 has received special attention over the last years as a 3' processing factor also involved in termination, its precise role in termination remains unclear.

Although Rat1 is implicated in transcriptional termination, Rat1 activity alone is not sufficient to promote efficient termination. Depletion of Rat1 inhibits but does not completely abolish termination (Kim et al. 2004) and digestion of Pol II-associated RNA

is not sufficient to dismantle enongation complexes (Kireeva et al. 2000). Furthermore the 5'-3' cytoplasmic exonuclease Xrn1, which is homologous to Rat1, is able to complement the exonuclease activity of a Rat1-deficient strain but not the termination defect (Luo et al. 2006). Thus, while degradation of the nascent 3' product is important, other activities must also be required to promote termination.

The above observations led to the discovery that the Pol II termination mechanism requires not only Rat1 but also Sen1 to be fully efficient (Kawauchi et al. 2008). The RNA/DNA helicase Sen1 is also involved in the processing of tRNA, snRNAs and snoRNAs, as well as in transcription-coupled DNA repair (Rasmussen and Culbertson, 1998; Steinmetz et al. 2006; Ursic et al. 1997). Sen1, Nrd1 and Nab3 proteins form the Nrd1 complex; the major factor in promoting sno/snRNA termination. A single amino-acid mutation in the C-terminal Sen1 helicase domain compromises its function, resulting in altered genome-wide distribution of Pol II of both coding and non-coding genes (Steinmetz et al. 2006). Sen1 binds to the CTD phosphorylated on Serine 2 (Ursic et al. 2004; Chinchilla et al. 2012), which possibly facilitates its recruitment to multiple coding as well as non-coding genes, where it tends to accumulate towards the 3' end (Chinchilla et al. 2012). Recently, it was also shown that the helicase domain of Sen1 is required for efficient termination. Thus *sen1-1* mutation caused not only a termination defect but also accumulation of R-loops and subsequent transcription-associated DNA recombination defects. This observation led to the proposed model that R-loop resolution by Sen1 could allow time for Rat1 to 'catch up' with Pol II and eventually degrade the 3' cleaved product so promoting efficient termination (Mischo et al. 2011).

The human homologues of Rat1 and Sen1 are Xrn2 and senataxin respectively, and their function in termination appears to be conserved (West et al. 2004; Kaneko et al., 2007b; Suraweera et al. 2009; Chapter 3; Chapter 4; Skourti-Stathaki et al. 2011). Depletion of Xrn2 results in a termination defect and accumulation of 3' RNA of the β-globin transcript (West et al. 2004). However the effect observed on β-globin pre-mRNA termination is specific to a co-transcriptional cleavage (CoTC) activity and not to the poly(A) signal. The CoTC sequence (discussed below) undergoes a rapid cleavage that exposes a free 5' RNA which in turn serves a substrate for Xrn2 (West et al. 2004). Like Rat1 in yeast, Xrn2 also associates with 3' end processing factors. Kaneko et al (2007) showed that Xrn2 interacts with p54/PSF, a multifunctional protein dimmer that binds the CTD (Emili et al. 2002; Rosonina et al. 2005) and was identified as a component of the 3' processing complex (Shi et al. 2009). p54/PSF may facilitate termination by recruiting Xrn2 to 3' regions, since accumulation of the downstream cleaved RNA and termination defects were detected *in vitro* and *in vivo* after p54 siRNA-mediated knock-down (Kaneko et al., 2007b).

Human senataxin protein has also been implicated in transcriptional termination (Suraweera et al. 2009; Chapter 3; Chapter 4; Skourti-Stathaki et al. 2011; Padmanabhan et al. 2012). Senataxin was initially identified when mutations causing ataxia oculomotor apraxia 2 (AOA2) and amyotrophic lateral sclerosis type 4 (ALS4) were mapped to the *SETX* gene (James and Talbot, 2006; Palau and Espinos, 2006). These diseases result in progressive degeneration of motor neurons in the brain and spinal cord, progressive muscle weakness and atrophy. *SETX* encodes a 302.8 kD widely expressed protein containing an N-terminal putative protein-protein interaction domain and a C-terminal

DEAD-box helicase domain, followed by a nuclear localisation signal (NLS) (Chen et al. 2004). Most senataxin mutations found in AOA2/ALS4 families, either cause premature protein termination or interfere with the function of the helicase or N-terminal protein interaction domains (Chen et al. 2004; Criscuolo et al. 2006; Duquette et al. 2005; Fogel and Perlman, 2006; Moreira et al. 2004). However, the precise mechanism of toxicity caused by these mutations and manifested in AOA2/ALS4 patients remains to be elucidated. Since senataxin localises to both nucleus and cytoplasm, unlike Sen1 which is a nuclear specific protein, it may possess additional cytoplasmic roles in RNA metabolism and translational regulation (Chen et al. 2006). The domain structure of senataxin is similar to Sen1. Furthermore, its C-terminal helicase domain is homologous to helicase domains of human immunoglobulin μ-Binding Protein 2 (IGHMBP2) and regulator of nonsense transcripts UPF1 proteins. Interestingly, mutations in IGHMBP2 have been previously found to cause autosomal recessive Spinal Muscular Atrophy (SMN) (Grohmann et al. 2001). At a cellular level mutations in senataxin cause hypersensitivity to single-strand DNA damage, but not to ionizing radiation, as observed in ataxia-telangiectasia mutated (ATM) deficient cells (Chen et al. 2006; Suraweera et al. 2007). Similar to its yeast counterpart, senataxin interacts with Pol II and other RNA processing factors, such as poly(A)-binding proteins 1 and 2 (PABP1/2), hnRNPs, SAP155 and SMN, pointing towards a role for senataxin in pre-mRNA processing as well as termination (Suraweera et al. 2009).

Mammalian Pol II termination requires *cis* acting DNA elements, as well as dedicated proteins. So far, two major classes of terminator elements located downstream the poly(A) signal have been described to enhance the termination process: CoTC and

pause elements (Figure 1.3). The first class may occur at more distal 3' flanking regions, as compared to pause elements, and is referred to as a co-transcriptional cleavage (CoTC) sequence (Dye and Proudfoot, 2001). This element is an 850 bp AT-rich DNA sequence located approximately 900 bp downstream the human β-globin poly(A) site and was shown to be required for efficient Pol II termination (Dye and Proudfoot, 2001). CoTC RNA sequences are highly unstable and undergo rapid cleavage as soon as they are produced (West et al. 2008). This provides an entry site for Xrn2 to degrade the nascent transcript (West et al. 2004) which in turn leads to release of Pol II from the DNA template (West et al. 2008). In this type of termination a second cleavage event occurs, this time at the poly(A) site, which triggers the release of mRNA from Pol II (West et al. 2008). CoTC-mediated termination has been shown to increase mRNA levels, as 3' end processing of the mRNA (still in association with Pol II) can occur in the nucleoplasm away from nuclear RNA degradation machinery (West et al. 2009). CoTC appears to be a specific termination signal for β- and ε-globin genes (Dye and Proudfoot, 2001) and the mouse serum albumin (MSA) gene (West et al. 2006). However it has not yet been identified in other genes.

The second class of termination elements is the transcription pause sites. ChIP analysis of human genes shows a high density of Pol II at promoter regions which decreases across the gene body and increases again downstream the poly(A) signal (Glover-Cutter et al. 2008). This increased Pol II density at the 3' end, as compared to the gene body, could indicate transcriptional pausing prior to termination. Indeed, some mammalian genes including α2-globin and γ-globin genes contain sequences downstream the poly(A) signal that act to slow down elongating Pol II and appear to play a role in 3'

end processing and termination (Enriquez-Harris et al. 1991; Plant et al. 2005). Another pause site, defined as the G-rich MAZ termination element, facilitates termination of the human liver-specific C2 complement gene (Ashfield et al. 1994; Yonaha and Proudfoot, 1999). This sequence is a potential binding site for the transcription factor MAZ and promotes efficient termination (Gromak et al. 2006). The human β-actin gene also possesses a G-rich pause element downstream the poly(A) signal which similarly acts to promote transcriptional termination, through Xrn2-mediated degradation of the downstream transcript exposed by poly(A) site cleavage (Gromak et al. 2006). It should be noted that in this type of termination process, a single cleavage event occurs at the poly(A) signal which provides the substrate for Xrn2. Pause elements are thought to promote termination by slowing down elongating Pol II and allowing more time for Xrn2 to degrade the nascent transcript, catch up with Pol II and efficient termination to occur (Gromak et al. 2006). Interestingly, genome-wide analysis revealed that G-rich sequences placed immediately downstream of poly(A) signals are relatively common in mammalian genes (Salisbury et al. 2006), suggesting that these elements may play an essential function in regulation of gene expression. Furthermore recent results show that such G-rich elements are associated with the formation of R-loops (Chapter 3; Chapter 4; Skourti-Stathaki et al. 2011). However the molecular mechanism of pausing in mammalian Pol II termination remains unclear.

1.4 R-loops

R-loops are structures where an RNA strand hybridises with the DNA template strand, leaving the non-template DNA strand single-stranded. In essence R-loops are RNA/DNA hybrids plus an unprotected single-stranded DNA strand. The R-loop structure was first characterised over 36 years ago (Thomas et al. 1976) and the first demonstration that R-loops exist *in vivo* came in 1995 with studies described by Drolet and colleagues. They showed that R-loop formation occurs in a bacterial cell and is a consequence of the transcription process (Drolet et al. 1995). In general where R-loops have been described *in vivo*, the RNA strand is generated by Pol II transcribing a C-rich DNA template so that a G-rich transcript is generated. Interestingly, several studies have shown that R-loops are formed preferentially when the non-template strand is G-rich (Reaban et al. 1994; Li and Manley, 2005; Ginno et al. 2012). This could be explained by the fact that rG:dC hybrids are particularly stable (Sugimoto et al. 1995). In addition to this, G quadruplex structures formed on the non-template DNA strand during transcription of the immunoglobulin class switch region may contribute to R-loop formation (Duquette et al. 2004).

It is as yet unclear how R-loops are generated. According to the 'extended RNA/DNA hybrid' model the RNA/DNA hybrid duplex could be the result of an extension of the usual 8 bp RNA/DNA hybrid (Westover et al. 2004) within the transcription bubble as Pol II elongates. The second model suggests that the RNA/DNA hybrid could arise by threading back the RNA before the two strands of the DNA duplex re-anneal (called the 'thread back model'). Crystallographic structure of Pol II reveals that DNA and RNA molecules exit through different channels (Westover et al. 2004),

pointing to the latter model for the mechanism of R-loop formation. According to extensive *in vitro* studies from the Lieber laboratory, R-loops depend on three features (Aguilera and Garcia-Muse, 2012): high G density (Roy and Lieber, 2009), negative supercoiling and DNA nicks (Roy et al. 2010). Initial R-loop formation is favoured by G clusters and DNA nicks downstream of the promoter on the non-template DNA strand, whereas subsequent RNA/DNA hybrid elongation and stabilisation are enhanced by high G density and negative supercoiling.

Once formed, R-loops are particularly stable. RNA/DNA associations are thermodynamically more stable than DNA/DNA interactions for reasons that still remain unclear (Roberts and Crothers, 1992). This enhanced stability might be due to the structure of the RNA/DNA hybrid that is thought to adopt a conformation that is an intermediate between the A-form of a dsRNA and the B-form of a DNA duplex (Shaw and Arya, 2008). Another possibility is that G quadruplexes formed on the single-stranded non-template strand (Duquette et al. 2004) stabilise the R-loop structure.

Stable R-loops have been reported *in vivo* at prokaryotic origins of replication (Baker and Kornberg, 1988; Carles-Kinch and Kreuzer, 1997; Lee and Clayton, 1996; Masukata and Tomizawa, 1984; 1990), the mitochondrial origin of replication (Xu and Clayton, 1996) and mammalian immunoglobulin (Ig) class switch region in activated B lymphocytes (Yu et al. 2003). In the latter case R-loop formation is involved in facilitating class switch recombination (CSR) that generates diverse antibody isotypes.

Transcription can be a 'risky' process. Newly synthesised transcripts could potentially hybridise with the template DNA strand, leaving the non-template DNA strand unpaired (Figure 1.4). ssDNA could then be susceptible to lesions and

transcription-associated mutagenesis (TAM) or transcription-associated recombination (TAR). Indeed, the deleterious effects on genome integrity, derived from R-loops formed during transcription, have been documented in various observations (Aguilera, 2002). Given that formation of R-loops is an evolutionary conserved mechanism (Li and Manley, 2005; Aguilera and Garcia-Muse, 2012), it is perhaps not surprising that different organisms have employed diverse mechanisms to protect their genomes (Li and Manley, 2006; Aguilera and Garcia-Muse, 2012).

In *S. cerevisiae*, transcription-induced R-looping was first documented with the characterisation of the THO complex, suggested to be involved in transcriptional elongation (Aguilera, 2002). This complex consists of four nuclear proteins (Hrp1, Tho2, Mft1 and Thp2) and is associated with the transcription-export complex (TREX), containing Tex1 and the mRNA export factors Sub2 and Yra1 (Chavez et al. 2000; Strasser et al. 2002). The physical interaction of THO complex with TREX directly links mRNA packaging with RNA export. Mutations affecting THO/TREX have been shown to affect transcriptional elongation, proper mRNA export and recombination (Strasser et al. 2002; Rondon et al. 2003). A distinctive phenotype of yeast THO mutants (firstly identified in the *HRP1* and *THO2* genes) is their transcription-associated hyper-recombination phenotype (Aguilera and Klein, 1990; Piruat and Aguilera, 1998) and this is directly associated with R-loop formation (Huertas and Aguilera, 2003). The THO/TREX complex is responsible for packaging of pre-mRNA with RNA binding proteins. As the nascent transcript emerges from elongating Pol II, it may have two immediate fates. Either it is co-transcriptionally packaged into messenger ribonucleotide proteins (mRNPs) and exported through the nuclear pore or it may invade the DNA

duplex behind the elongating Pol II to form R-loop structures. R-loops could then interfere with DNA replication, induce ssDNA breaks or become recombination intermediates. Consequently, it could be argued that efficient mRNA packaging into mRNPs prevents R-loop formation and in turn transcription-associated recombination and DNA damage (Huertas and Aguilera, 2003; Moore and Proudfoot, 2009; Mischo et al. 2011; Aguilera and Garcia-Muse, 2012). However recently it has been reported that even with normal mRNP biogenesis, accumulation of R-loops can still occur (Mischo et al. 2011) suggesting that R-loops form at a much higher frequency than once thought. DNA topoisomerase I (Top I) is an evolutionary conserved factor that suppresses R-loop formation (Drolet et al. 1995; Tuduri et al. 2009; El Hage et al. 2010). This is possibly due to the abilility of topoisomerase I to relax negative supercoiling. In the absence of Top 1, negative supercoiling accumulates behind elongating Pol II to promote opening of DNA which in turn facilitates reannealing between nascent RNA and DNA template strand and subsequent R-loop formation (Drolet et al. 1995; Tuduri et al. 2009; El Hage et al. 2010).

Given the potential harmful effects of stable R-loops, cells use a range of dedicated factors to actively remove them once formed. First of all, the RNase H enzymes have been suggested to cleave the RNA of RNA/DNA hybrids (Stein and Hausen, 1969; Hausen and Stein, 1970; Cerritelli and Crouch, 2009). In most organisms, there are two types of RNase H. Eukaryotic RNase H1 consists of one single protein with the N-terminal domain being responsible for binding to the RNA/DNA hybrid (hence named hybrid binding domain-HBD) and the C-terminal domain containing the RNase H active site (Cerritelli and Crouch, 2009). RNase H2 is composed of three different

proteins with the RNase 2A being the catalytic subunit. RNase H1 and H2 may have different *in vivo* substrates due to their differences in hybrid hydrolysis. RNases H have also been implicated in nuclear and mitochondrial DNA replication by generating and removing RNA primers (Cerritelli and Crouch, 2009). Second, Sen1 helicase acts to remove R-loops and to prevent genomic instability by R-loop mediated DNA damage (Mischo et al. 2011). Two recent studies from Mischo et al (2011) and Stirling et al (2012) reveal a very interesting connection: termination factors such as Sen1 (Mischo et al. 2011) and 3' end processing factors such as Pcf11 and Rna15 (Stirling et al. 2012) both act to protect the heavily transcribed yeast genome from R-loop mediated genomic instability. Based on these observations, a potential link between genomic instability and 3' end processing/termination factors is apparent. My thesis research directly addresses this interesting connection in a mammalian context, I will therefore next describe what is so far known about R-loops in mammalian cells, where genes are significantly longer and the presence of introns is almost ubiquitous.

In higher vertebrates the ASF/SF2, a serine-arginine-rich (SR) protein that regulates the first steps of splicing, appears to interconnect between pre-mRNA processing and genomic instability (Li and Manley, 2005). Li and Manley demonstrated in chicken DT40 cells and human HeLa cells that depletion of ASF/SF2 causes the nascent transcript to form R-loops which in turn promote DNA rearrangements mediated by double-strand breaks (DSBs) (Li and Manley, 2005). In essence the model proposed is that ASF/SF2 is co-transcriptionally loaded into the nascent pre-mRNA via the phosphorylated CTD of Pol II not only to promote splicing but also to prevent R-loop formation and subsequent genomic instability (Li and Manley, 2005; Aguilera, 2005).

This is a clear example of connecting transcription-induced R-looping, pre-mRNA processing, Pol II CTD and transcription-associated genomic instability. The co-transcriptional R-loop induced genomic instability observed in DT40 cells with the loss of ASF/SF2 factor (Li and Manley, 2005; Li and Manley, 2006) resembles the phenotype of yeast THO mutants (Huertas and Aguilera, 2003). Until recently the involvement of human THO/TREX complex in genomic instability was not clearly identified. A study from Dominguez-Sanchez and colleagues (2011), however, revealed that the interplay between mRNP biogenesis and genomic instability is indeed conserved from yeast to humans. In essence depletion of human THO complex results in accumulated DNA breaks and consequent genomic instability which is dependent on R-loop formation (Dominguez-Sanchez et al. 2011). In addition to this, genome-wide data suggest that RNA processing factors could help prevent genomic instability (Paulsen et al. 2009), further pointing towards a powerful interplay between pre-mRNA processing and R-loop dependent genomic instability.

How R-loops lead to genomic instability still remains unclear. Possibly the unpaired DNA strand resulting from R-loop formation is more susceptible to DNA damage such as spontaneous deamination of dC to dU, leading to double-strand breaks and recombination. Indeed ssDNA is more vulnerable to DNA damage than dsRNA (Aguilera, 2002; Li and Manley, 2006). Thus, in one possible model, accumulation of R-loops could make certain regions of the genome more prone to DNA-damaging agents by increasing the occurence of single-stranded regions. In a second potential model, a protein recognising R-loop structures could be involved in initiating the generation of mutagenesis. One possible candidate is the activation-induced cytidine deaminase (AID).

AID is an enzyme that promotes Ig heavy-chain CSR and hypermutation in B lymphocytes. It functions by deaminating cytosines into uracils on transcribed single-strand target DNA sequences (Muramatsu et al. 2000; Revy et al. 2000). R-loops forming behind elongating Pol II can provide the single-strand DNA substrate for this enzyme (Yu et al. 2003). The generated U:G mismatch could then be replicated creating two daughter species, one of which will undergo a C→T transition mutation. However dU could also be processed by the base excision repair (BER) components such as uracil-DNA glycosylase and abasic endonuclease (APE). These enzymes remove the uracil creating a DNA nick or leaving an abasic site (a site of base loss). Replication past the abasic site will result in random incorporation of any of the four nucleotides, possibly leading to further mutations. DNA nicks could also be converted to DSBs which are recognised by the recombination-mediated repair machinery ultimately leading to CSR (Di Noia and Neuberger, 2002). AID is specifically expressed in activated B cells and in chicken DT40 cells (which are B cell derived). This raises the question of how R-loop formation leads to genomic instability in AID-lacking cells. DSBs have also been observed in HeLa cells upon depletion of ASF/SF2 (Li and Manley, 2005), suggesting that other proteins could function analogously to AID in other cell types to initiate R-loop induced genomic instability. Alternatively spontaneous dC→dU mutation may occur albeit at low levels. An additional possible scenario suggests that transcriptional R-loops induce genomic instability by interfering with DNA replication (Aguilera, 2002; Houlard et al. 2011; Gan et al. 2011; Aguilera and Garcia-Muse, 2012). This transcription-dependent replication fork collisions have been shown to induce TAR or DNA breaks (Prado and Aguilera, 2005; Gottipati et al. 2008; Boubakri et al. 2010). R-loop induced

replication blockage could occur by unrepaired DNA lesions in the ssDNA unpaired strand derived from R-loop formation or by RNA/DNA hybrids themselves forming behind elongating Pol II and somehow restricting transcription. Such blocked replication forks could then generate DNA lesions and DSBs in the newly synthesised DNA. These would induce recombination-mediated repair which in turn could lead to chromosome rearrangements and genomic instability (Aguilera and Garcia-Muse, 2012).

So far I have described the potential deleterious effects of R-loops on genome integrity, except in the unusual case of antibody genes where R-loops have been put to a positive use. However R-loops may possess further beneficial features. As mentioned above, Mischo et al (2011) have shown that R-loop structures occur naturally, even when proper mRNA packaging occurs. In addition to this, it is of particular note that R-loops accumulate in the Sen1 mutant that lacks the helicase activity. The helicase domain of Sen1 is also responsible for efficient Pol II termination. Even though this paper focuses on the role of Sen1 in R-loop mediated DNA damage and does not directly consider the potential association of R-loops with Pol II termination, some connection is implied. Another interesting observation that could further support the potential connection of R-loops with Pol II termination comes from Huppert and colleagues (2008). Genome-wide analysis has previously revealed that G-rich sequences immediately downstream the poly(A) signal are relatively common in mammalian genes (Salisbury et al. 2006). Recent genome-wide bioinformatics analysis has also shown that promoters and 3' regions of genes are enriched in G-quadruplex-forming sequences (Huppert et al. 2008). Interestingly, 3'UTR G-quadruplexes are particularly prevalent in cases where a second gene is placed in close proximity, suggesting that G-quadruplexes may be involved in

transcriptional termination (Huppert et al. 2008). Given that G-rich sequences could act to stabilize R-loops, it is tempting to speculate that R-loops could be an important aspect of Pol II termination.

Ginno and colleagues (2012) have recently presented evidence for widespread R-loop formation over the 5' UTR regions downstream of CpG promoters in the human genome (Ginno et al. 2012). These genomic regions are GC rich and have strong positive GC skew (template strand having excess of C versus G residues). Interestingly, R-loop formation at promoter CpG islands may be directly involved in the maintenance of the unmethylated state of CpG promoters either by recruiting the protective H3K4me3 mark or the DNA demethylation complex (Ginno et al. 2012). In accordance with this model, AID complex is also found at H3K4me3 promoter proximal sequences including CpG islands (Yamane et al. 2011). Finally R-loops formed over the centromeric repeats in fission yeast have been shown to mediate RNAi-dependent heterochromatin formation (Nakama et al. 2012). This last study is of particular interest since it shows that R-loops could be potentially involved in silencing the centromeric DNA repeats. These observations raise interesting questions for further investigation as described in my thesis since they suggest that R-loops could control an epigenetic mark directly or indirectly. The possibility that R-loops facilitate the formation of a repressive epigenetic mark to induce gene silencing is intriguing. To further discuss this potential connection, the next section of the introduction covers the mechanism of RNAi-dependent heterochromatin formation.

1.5 RNAi-dependent heterochromatin formation and transcriptional gene silencing

Emil Heitz, in 1928, laid one of the keystones of cytogenetics: the distinction between heterochromatin and euchromatin based on the differential compaction of chromatin during interphase (Passarge, 1979; Grewal and Jia, 2007; Reyes-Turcu and Grewal, 2012). Euchromatin is less condensed and more 'easily' transcribed, whereas heterochromatin is highly condensed and generally an obstacle to the transcriptional machinery (Huisinga et al. 2006). Heterochromatin can be divided in two subcategories: constitutive and facultative heterochromatin. The first class contains highly repetitive DNA sequences, such as clusters of satellites and transposable elements, found at centromeres and telomeres. These regions remain condensed and stably silenced throughout the cell cycle (Grewal and Jia, 2007). Facultative heterochromatin, on the other hand, is more dynamic and is often associated with developmentally regulated genes, where the chromatin state can change in response to cellular signals and gene activity (Trojer and Reinberg, 2007). One of the signatures of heterochromatin is its ability to influence gene expression in a region-specific, sequence-independent way (Grewal and Jia, 2007).

In studies on gene expression, two fields are dominant: RNA and chromatin. Even though traditionally they are considered unrelated, an increasing body of evidence suggests that RNA and chromatin are linked. Indeed, non coding RNAs (ncRNAs) are involved in the control of epigenetic regulation, chromosomal dynamics, normal cellular differentiation and organism development. These processes include dosage compensation in flies and mammals, imprinting and RNA intereference (RNAi)-mediated gene

silencing (Morey and Avner, 2004; O'Neill, 2005; Bernstein and Allis, 2005; Grewal and Jia, 2007; Grewal, 2010).

RNA intereference (RNAi) is a mechanism by which double-stranded RNAs suppress gene transcripts in a sequence-dependent manner. The mechanism of RNAi has received special attention the last decade since its initial discovery in *C. elegans* by Mello and Fire. Fire and colleagues showed that double-stranded RNA is the trigger of gene silencing (Fire et al. 1998). RNAi adopts different strategies in different organisms. In fission yeast RNAi usually represses mRNA transcription (transcriptional gene silencing-TGS) whereas in mammals it principally mediates post-transcriptional gene silencing (PTGS) through mRNA degradation or translational inhibition (Grewal, 2010; Bartel, 2009). TGS is usually associated with DNA methylation and histone tail modifications such as histone H3 lysine 9 dimethylation (H3K9me2) or trimethylation (H3K9me3). Interestingly, recent studies in *C. elegans* have shown that there is a connection between cytoplasmic PTGS and nuclear TGS mechanisms. In essence miRNAs can re-enter the nucleus and be targeted to nascent transcripts. This results in recruitment of histone methyltransferases and ultimately heterochromatin formation and TGS (Gu et al. 2012)

It has been long recognised that RNAi mechanisms occur in a sequence-dependent manner (Hannon, 2002; Meister and Tuschl, 2004; Beirnstein and Allis, 2005). Usually the pathway starts with the formation of double-stranded RNA (dsRNA), which can be produced by bidirectional transcription, transcription of an inverted repeat or 'hairpin' sequence or can be physically introduced to cells by injection, or even feeding dsRNA-expressing bacteria (Kennerdell and Carthew, 1998; Montgomery and Fire, 1998; Tabara et al., 1998; Timmons and Fire, 1998; Waterhouse et al., 1998; Clemens et al.,

2000; Billy et al., 2001; Tomari and Zamore, 2005; Carthew and Sonthheimer, 2009). dsRNA so formed can then be processed by RNAse III type ribonucleases, Dicer and Drosha to generate short interfering RNAs (siRNAs) or micro RNAs (miRNAs) respectively (Carthew and Sonthheimer, 2009). In PTGS, these small RNAs are then loaded onto a dedicated protein complex called RISC (RNA-induced silencing complex). The argonaute protein encodes the nuclease activity of this complex which may cleave the mRNA targeted by siRNA or miRNA-primed RISC (Liu et al. 2004; Meister et al. 2004; Song et al. 2004; Hutvagner and Simard, 2008).

The mechanism of siRNA-directed TGS has been studied extensively in *S. pombe*, particularly in centromeric regions (Figure 1.5). Using this organism as a model system, several groups have shown that siRNAs play an important role in regulating heterochromatin formation. Earliest experiments tested loss-of-function effects of various RNAi genes, including dicer (dcr1), argonaute 1(ago1) and RNA-dependent RNA polymerase (rdrp1) (Hall et al. 2002; Volpe et al. 2002). Interestingly, disruption of these genes led to a decrease of gene silencing of a *ura4* marker gene, inserted into the repeats of the centromere. In addition to this, a subsequent loss of H3K9 methylation was observed (Volpe et al. 2002). Similar experiments performed at the mating-type locus (*mat*), containing homologous sequences to the centromeric repeats, illustrated that the RNAi apparatus is also involved in silencing this locus (Hall et al. 2002). Since then, an increasing body of evidence has emerged to prove that RNAs play a critical role in TGS. Perhaps the key milestone in our understanding of the TGS mechanism came from the biochemical purification of complexes that contain small RNAs involved in this process. The siRNA-primed RNA-induced transcriptional silencing (RITS) contains Ago1 (the

catalytic subunit of the complex), Chp1 (a chromodomain-containing protein) and Tas3. The siRNA in this complex targets it to heterochromatic gene loci with sequence homology to the siRNA ultimately resulting in gene silencing (Verdel et al. 2004). RITS is recruited to heterochromatic loci in a manner dependent on binding to H3K9 methylation via the Chp1 chromodomain (Noma et al. 2004; Cam et al. 2005; Petrie et al. 2005; Schalch et al. 2009) and facilitates targeting of Rdp1 (Motamedi et al. 2004; Sugiyama et al. 2005). The collective model arising from these observations suggests a powerful role for RNAi machinery in the establishment of heterochromatin at centromeres, telomeres and repetitive loci. During RNAi-mediated heterochromatin formation in fission yeast, ncRNAs are transcribed from heterochromatic repeats by Pol II (Djupedal et al, 2005; Kato et al. 2005). Primary siRNAs are then presumably generated through processing of hairpin-like structures from ncRNAs (Djupedal et al. 2009; Grewal, 2010). siRNAs are thought to be used as guiding molecules for the binding of RITS to the DNA sequence or the nascent transcript. In turn, engaged RITS complex recruits RNA-dependent RNA polymerase complex (RDRC) to generate double-stranded RNA (Motamedi et al. 2004) which is then processed into siRNA by Dicer (Verdel et al. 2004; Colmenares et al. 2007) to initiate another round of heterochromatin formation representing a self-enforcing loop to maintain silencing at these loci. In parallel, RITS recruits Clr4 histone methyltransferase to methylate H3K9 which attracts the Swi6 and Chp2 heterochromatin proteins (fission yeast homologues of Heterochromatin Protein1-HP1). This ultimately results in the formation of heterochromatin (Verdel et al. 2004; Zhang et al. 2008; Bayne et al. 2010).

Binding of RITS complex to the nascent transcript or the DNA target appears to be a key step for the RNAi-mediated heterochromatin formation in fission yeast. Since Pol II transcription of the targeted loci seems to be coupled with siRNA generation (Kato et al. 2005) and in turn siRNA production is connected to recruitment of RITS at heterochromatin (Verdel et al. 2004), it is perhaps unsurprising that heterochromatic ncRNAs have been suggested to remain on chromatin and function as a binding platform for the RNAi apparatus (Cam et al. 2009). However two important questions remain unanswered. How do ncRNAs remain bound to chromatin? Does RITS target the gene transcript or the DNA strand? Very recently an intriguing model was proposed for the association of ncRNA derived from centromeric repeats with chromatin and its function in heterochromatin formation (Nakama et al. 2012). Nakama and colleagues suggest that ncRNA transcribed from heterochromatin remains bound to chromatin via the formation of an R-loop and that the so formed RNA/DNA hybrid itself is involved in the heterochromatin formation. In essence, RNA/DNA hybrid foci co-localise with centromeric heterochromatic regions and over-expression of RNase H decreases these foci which in turn disrupts heterochromatin formation. Given that with R-loop structures the single-strand non-template DNA remains unpaired, they predict that RITS complex could target the single-stranded 'unprotected' DNA or alternatively the chromatin-associated RNA to generate heterochromatin formation (Nakama et al. 2012). Even though this study does not discuss whether R-loop formation is the trigger of the heterochromatin formation rather than the consequence, these observations connect tightly the field of transcription and the field of heterochromatin raising interesting questions for further investigation. Do R-loops facilitate transcriptional gene silencing in

other regions of the genome? Is this phenomenon conserved in mammalian cells? Could R-loop 'hot-spots' be regions of RNAi-dependent heterochromatin assembly?

TGS in mammalian cells, analogous to this process in *S. pombe*, still remains a mystery, unlike PTGS, which has been well established. Mammalian TGS has only recently being identified and the view that there is only PTGS in mammals still remains strong. However an emerging number of studies have pointed towards the possibility that mammalian RNAi-mediated TGS also occurs. siRNAs targeted to promoter sequences can induce H3K9 methylation in various mammalian cell lines (Kawasaki and Taira, 2004; Morris et al. 2004). In addition to this, Allo and colleagues (2009) have suggested siRNA-mediated epigenetic control of alternative RNA splicing. In essence siRNAs targeted to intronic sequences of the human fibronectin gene can affect alternative splicing in human cells. This process is dependent on Ago1 and it is mediated by H3K9me2 formation which leads to Pol II pausing and subsequent alternative exon inclusion (Allo et al. 2009). One of the highly controversial topics in mammalian TGS field is the 'nuclear life' of RNA silencing components. A valuable contribution to this came from investigation of a potential nuclear form of the RISC complex. Indeed, Ohrt and colleagues (2008) demonstrated that two different RISC complexes exist; one is present in the cytoplasm and the other is in the nucleus likely composed of Ago2 protein and siRNA (Ohrt et al. 2008). Furthermore human Ago1 and Ago2 (members of the large family of argonaute proteins) have also been shown to have nuclear localization and to be involved in TGS (Kim et al. 2006; Janowski et al. 2006).

The trigger of TGS is siRNAs generated from double-stranded RNA. In mammalian cells, double-stranded RNA could form naturally as a consequence of

endogenous overlapping transcripts. Accumulating evidence for the existence of ncRNAs (Mattick, 2009), including natural antisense transcripts (NATs) (Werner and Sayer, 2009; Faghihi and Wahlestedt, 2009), further supports this hypothesis. Indeed NATs have been demonstrated to induce TGS in both the human α-globin gene cluster (Tufarelli et al. 2003) and the murine sodium/phosphate transporter gene (Carlile et al. 2009).

The ultimate consequence of TGS is heterochromatin formation and gene silencing. Among the various epigenetic marks observed during TGS, methylation of H3K9 is perhaps the most conserved for heterochromatin formation and transcriptional silencing (Jenuwein, 2006). The mammalian Suv39h enzymes, homologues of the *S. pombe* Clr4, were the first histone lysine methyltransferases (HKMTs) discovered (Rea et al. 2000). Su(var)3-9-Enhancer of *zeste*-Trithorax (SET) domain is demonstrated to be the catalytic domain for histone lysine methylation (Rea et al. 2000) and has been shown to create a binding site for the chromodomain of HP1 protein (Lachner et al. 2001; Lachner and Jenuwein, 2002). Even though the specification of each H3K9 HKMTs still remains unclear (Collins et al. 2005; Kubicek et al. 2007; Shinkai and Tachibana, 2011), mammalian Suv39h1/h2 HKMTs appear to catalyse the most stable state of H3K9 methylation on pericentric heterochromatin, trimethylation of H3K9 (H3K9me3), whereas G9a and GLP are major H3K9me1 and H3K9me2 HKMTs of euchromatin (Jenuwein et al. 1998; Tachibana et al. 2001; 2002; 2005; Peters et al. 2003; Rice et al. 2003; Fodor et al. 2010). G9a and GLP exist predominantly as a G9a/GLP heteromeric complex that appears to be a functional methyltransferase in vivo (Tachibana et al. 2005; 2008) inducing gene repression within euchromatin (Tachibana et al. 2002).

Over the last decade a flurry of papers have shed light on an 'RNA-guided pathway for the epigenome' (Jenuwein, 2002). Even though the field of RNAi-mediated TGS is highly controversial and many of its aspects still remain unclear, it seems plausible that this phenomenon represents a critical interrelationship between RNA and chromatin.

1.6 'Everything is interconnected' - Experimental aims

One of the models for mammalian Pol II transcriptional termination is the pause-dependent termination. Detailed studies from this laboratory have shown that G-rich transcription pause sites act to slow down Pol II and have been identified in human β-actin gene (Gromak et al. 2006). In addition to this, genome-wide analysis has recently revealed that G-rich sequences downstream the poly(A) signal are relatively common in mammalian genes (Salisbury et al. 2006). The yeast RNA/DNA helicase, Sen1, is known to cooperate with Xrn2/Rat1 to promote efficient termination (Kawauchi et al. 2008) and its function is likely associated with the resolution of R-loops via its helicase domain (Mischo et al. 2011). R-loops and pre-mRNA processing events have been shown to be tightly linked (Li and Manley, 2005; Aguilera, 2005; Dominguez-Sanchez et al. 2011; Stirling et al. 2012). Finally it was recently demonstrated that R-loops formed over centromeric repeats could act to facilitate RNAi-dependent heterochromatin formation (Nakama et al. 2012).

At the start of my DPhil thesis project, the molecular function of human senataxin (the human homologue of Sen1) in transcription was unclear. Therefore my first experimental aim was to uncover its function in pause-dependent Pol II termination. In

the first Results chapter (Chapter 3) I describe the experiments performed to investigate the role of senataxin in the regulation of gene transcription and whether its molecular function is associated with the R-loop resolution. To address this question, I initially analysed two previously described gene constructs which contained G-rich pause sites (Gromak et al., 2006). A key observation of these studies was that R-loops are required for pause-dependent transcriptional termination. Therefore I further investigated the molecular mechanism of R-loop dependent transcriptional termination in an endogenous system and the role of senataxin and the 3' end processing factor, Pcf11, in this process (Chapter 4). The results obtained from Chapters 3 and 4 reveal that R-loop formation at the termination region facilitates Pol II pausing. Experiments presented in Chapter 5 are aimed to further study how R-loops mediate termination pausing in mammalian genes, adding a new player in this complex inter-connected network, RNAi-dependent H3K9me2 formation.

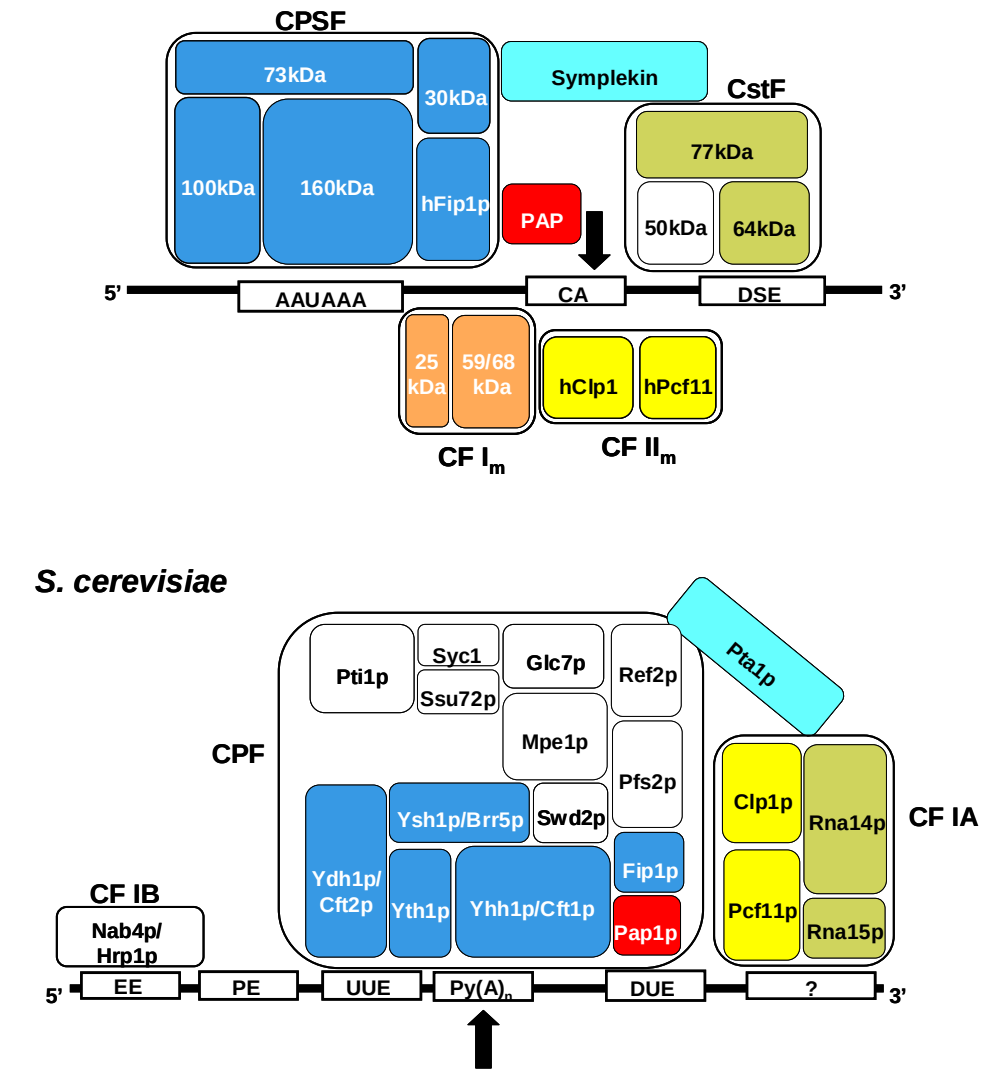


Figure 1.1 Mammalian and *S. cerevisiae* cleavage and polyadenylation factors.
Inter-connections between 3' end processing complexes and their points of contact with the poly(A) signal are indicated. Mammalian and *S. cerevisiae* pre-mRNA shown as black lines with cis-acting elements labelled as white boxes. Site of cleavage is indicated by black arrow. Homologous factors between species are denoted with the same colour whereas non-homologous factors are shown in white. For example, CPSF-160, -100, -73 and -30 kDa subunits are homologous to Yhh1p, Ydh1p, Ysh1p and Yth1p respectively. CstF-77 and -64 kDa subunits correspond to Rna14p and Rna15p respectively. The scaffold protein symplekin (Pta1p in yeast) reinforces the interactions between the two major 3' end processing complexes, CPSF and CstF. Diagram is not drawn to scale and is adapted from Proudfoot, 2004.

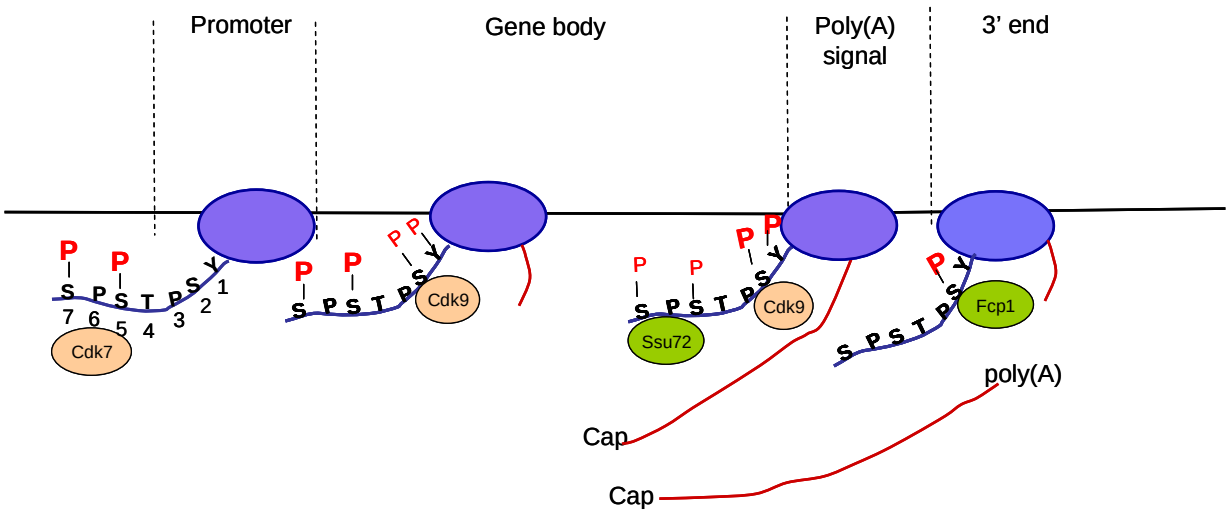
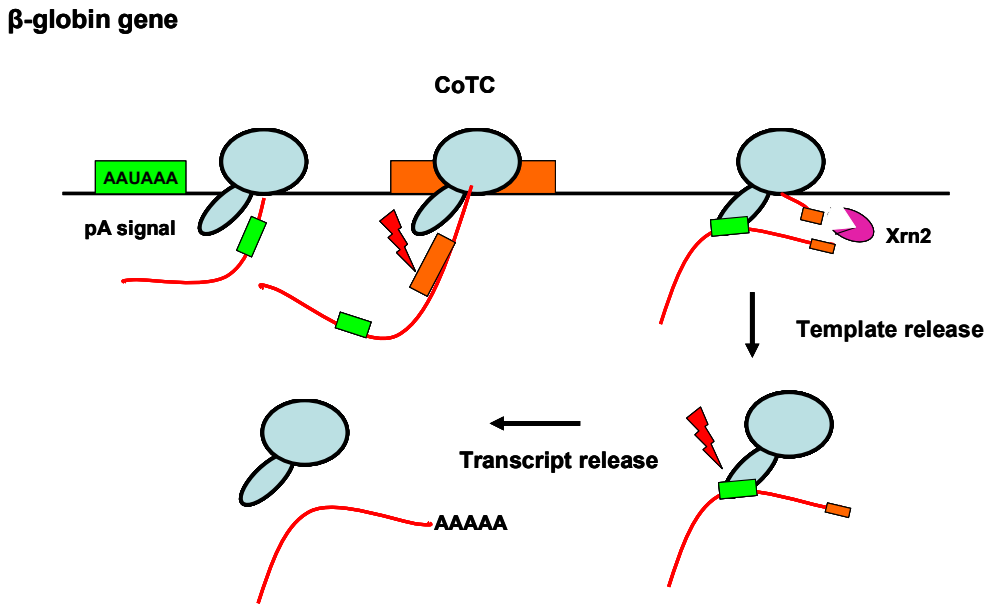


Figure 1.2 Dynamic phosphorylation of the Pol II CTD coordinates the transcription cycle. During transcriptional initiation, Cdk7 kinase phosphorylates Serine 5 and Serine 7. As elongation commences, Serine 2 positions at the CTD of elongating Pol II are phosphorylated by Cdk9, while the phosphate groups of Serine 5 and Serine 7 are gradually removed by Ssu72 phosphatase. CTD Tyrosine 1 residue is also phosphorylated, however the specific kinase remains unidentified. As Pol II approaches the poly(A) signal, the Tyrosine 1 phosphates are removed while phosphorylated Serine 2 still remains. Finally, as Pol II nears termination Serine 2 phosphates are removed by phosphatases such as Fcp1 regenerating unphosphorylated Pol II that can re-enter a further round of transcription. Diagram not drawn to scale.

A


β-globin gene
CoTC
AAUAAA
pA signal
Xrn2
Template release
Transcript release
AAAAA

B

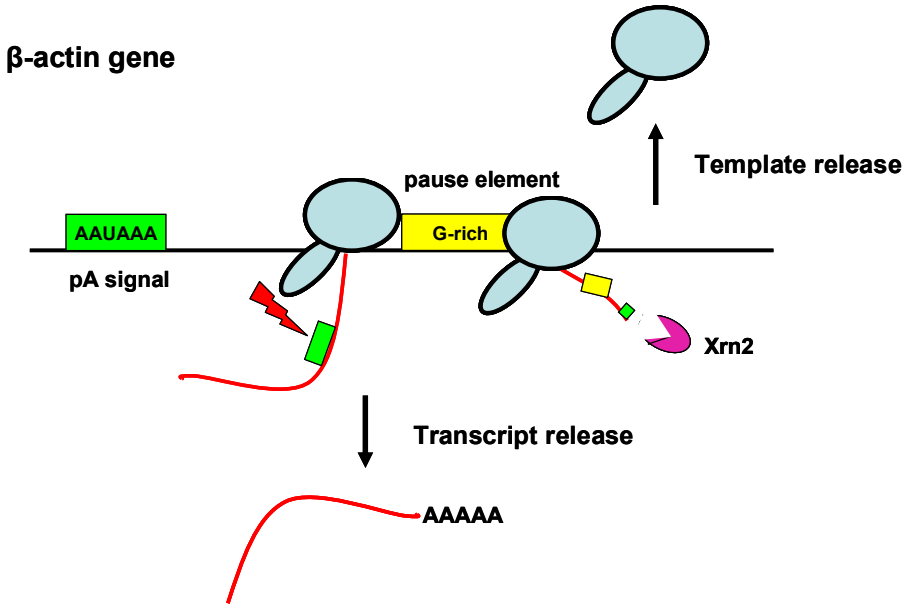

β-actin gene
Template release
pause element
AAUAAA
G-rich
pA signal
Xrn2
Transcript release
AAAAA

Figure 1.3 Models for mammalian CoTC-dependent and pause-dependent Pol II transcriptional termination

A) CoTC-dependent termination of the β-globin gene. DNA is shown as solid line and RNA as red line. Elongating Pol II is shown as a blue oval with CTD appendage. As Pol II transcribes the CoTC element (orange box) located downstream the poly(A) signal (green box), rapid cleavage occurs over the CoTC transcript providing entry sites for the 5'-3' exonuclease Xrn2 (shown in purple) to degrade the nascent transcript. This in turn leads to release of Pol II from the DNA template with the pre-mRNA still attached. A second cleavage event occurs in the nucleoplasm at the poly(A) signal to trigger the release of mRNA from Pol II.

B) Pause-dependent termination of the β-actin gene. DNA, RNA and Pol II are denoted as above. In this type of termination, only one cleavage event occurs at the poly(A) signal (green box) which triggers transcript release and generates the substrate for the exonuclease Xrn2. The G-rich pause element (yellow box) is thought to promote termination by slowing-down elongating Pol II and allowing more time for Xrn2 to degrade the nascent transcript, 'catch-up' with Pol II leading to its release from the DNA template. Diagram is not drawn to scale.

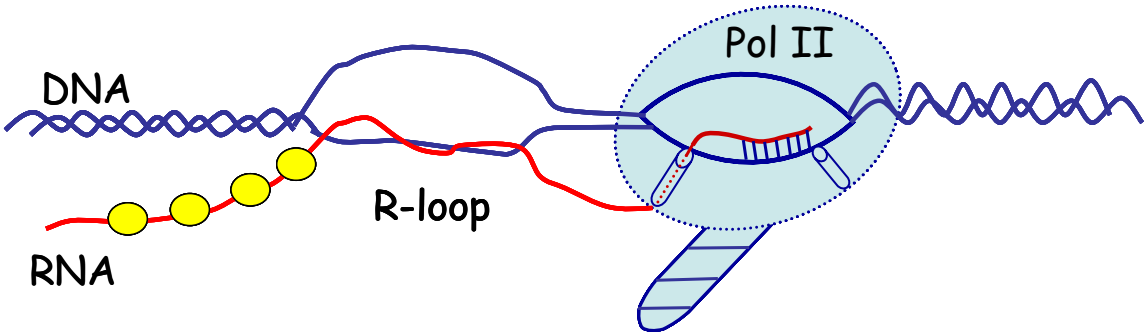


Figure 1.4 R-loops form behind elongating Pol II during transcription
DNA is shown as blue lines and RNA is shown as red line. Nascent transcripts behind elongating Pol II can invade the DNA duplex and re-hybridise with DNA template strand. This so called R-loop structure comprises the RNA/DNA hybrid plus unpaired DNA non-template strand. Efficient packaging into mRNP particles (yellow) restricts R-loop formation. Diagram courtesy of Dr. Natalia Gromak.

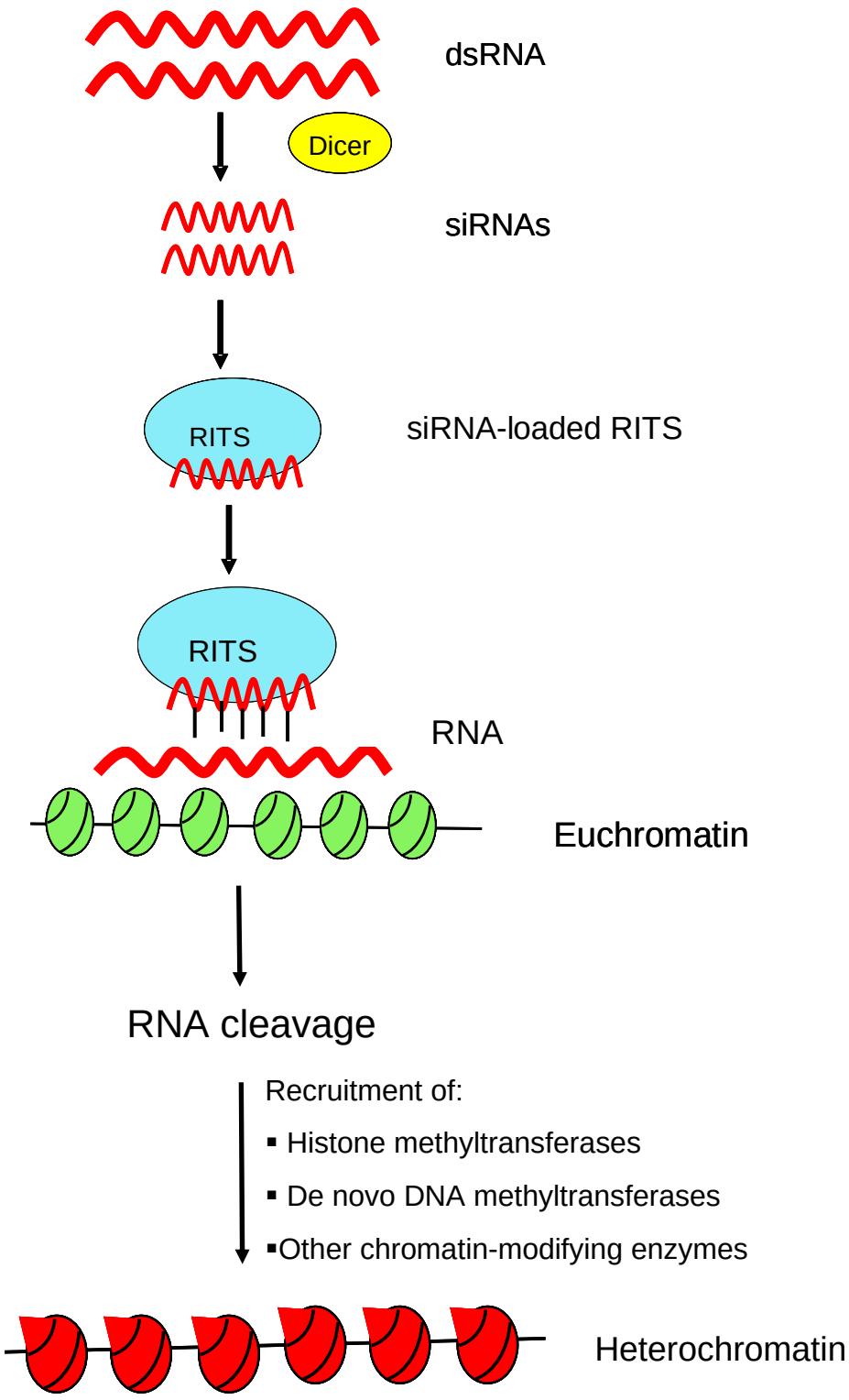

dsRNA
Dicer
siRNAs
RITS
siRNA-loaded RITS
RITS
RNA
Euchromatin
RNA cleavage
Recruitment of:
▪ Histone methyltransferases
▪ De novo DNA methyltransferases
▪Other chromatin-modifying enzymes
Heterochromatin

Figure 1.5 Model for RNAi-dependent gene silencing in *S. pombe*
The mechanism of siRNA-directed TGS in fission yeast involves initial processing of long double-strand RNAs into siRNAs by the RNase-III enzyme, Dicer. This activates the Argonaute-containing RNA-induced transcriptional silencing (RITS) complex. siRNA-activated RITS associates then with DNA target sequence either directly or through the nascent transcript. Binding of RITS to the gene target induces the recruitment of histone methyltransferases, DNA methyltransferases and other chromatin-modifying enzymes. This ultimately results in the formation of heterochromatin and subsequent gene silencing. Diagram is not drawn to scale.

CHAPTER 2

Materials and Methods

2.1 Reagents

Most materials were obtained from either Sigma-Aldrich or VWR for BDH products and were 'AnalaR' grade. Materials obtained from other distributors are listed below:

[α^{32}P]rUTP3000Ci/mmol (10mCi/ml)	GE healthcare
5x Transcription buffer	Promega
Acrylamide (30:1 Protogel)	National Diagnostics
Agarose	Melford
Bradford reagent	Biorad
Deoxyribonucleotide triphosphates (dNTPs)	Invitrogen
Dithiothreitol (DTT)	Invitrogen
Dulbecco's Modified Eagle Medium	PAA
Dulbecco's PBS	PAA
ECL Western Blotting Detection kit	GE healthcare
Foetal Calf Serum	Gibco BRL
Glycogen	Roche
HEPES	Melford
PVDF protein transfer membrane	GE healthcare
Hybond-NX nitrocellulose	GE healthcare
Leupeptin	Roche
Lipofectamine 2000	Invitrogen
Maxi-prep/PCR purification kits	Qiagen

Non Essential Amino Acids	PAA
OPTIMEM	Gibco BRL
Penicillin/Streptomycin	PAA
Pepstatin	Roche
PIPES	Melford
Protein A/G agarose (salmon sperm blocked)	Upstate/Millipore
Proteinase Inhibitor cocktail Complete (EDTA-free)	Roche
Ribonucleotide triphosphates (rNTPs)	Roche
Trizol	Invitrogen
Trypsin	Gibco BRL

2.2 Enzymes

Enzymes were obtained from the following suppliers and were used according to the manufacturer's instructions.

DNaseI RNase-free	Roche
Proteinase K	Roche
RNase A	Roche
RNase H	NEB
RNase OUT	Invitrogen
S1 Nuclease	Fermentas
SYBR Green (Real Time PCR reaction mix)	Qiagen

SuperScript III reverse transcriptase	Invitrogen

2.3 Antibodies

Anti-Actin	Sigma
Anti-Mouse	Sigma
Anti-Rabbit	Sigma
Anti-Senataxin	Bethyl Laboratories
Anti-Xrn2	Gromak et al., 2006
Anti-Pol II H224	Santa-Cruz
Anti-Pcf11	Santa-Cruz
Anti-Pcf11	(Zhang and Gilmour 2006)
Anti-flag	Sigma
Anti-GFP	Abcam
Anti-Ser2P	Covance
Anti-BrdU	Sigma
Anti-H3K9me2	Abcam
Anti-H3K9me3	Abcam
Anti-Ago1	Millipore
Anti-Dicer (13D6)	Abcam
Anti-H3	Abcam
S9.6	mouse hybridoma cell line- kind gift from Richard Gibbons

2.4 Primers

Primers were ordered from Sigma-Aldirch as lyophilised powders and diluted in water to 100Mm stocks. Stocks were kept in -20^{o}C.

Table 1

Name	**Sequence (5'→ 3')**
Human β-actin gene	
prom (F)	GAG GGG AGA GGG GGT AAA
prom (R)	AGC CAT AAA AGG CAA CTT TCG
in1 (F)	CGG GGT CTT TGT CTG AGC
in1 (R)	CAG TTA GCG CCC AAA GGA C
in3 (F)	TAA CAC TGG CTC GTG TGA CAA
in3 (R)	AAG TGC AAA GAA CAC GGC TAA
in5 (F),	GGA GCT GTC ACA TCC AGG GTC
in5 (R)	TGC TGA TCC ACA TCT GCT GG
5'pause (F)	TTA CCC AGA GTG CAG GTG TG
5'pause (R)	CCC CAA TAA GCA GGA ACA GA
pause (F)	GGG ACT ATT TGG GGG TGT CT
pause (R)	TCC CAT AGG TGA AGG CAA AG
C (F)	TGG GCC ACT TAA TCA TTC AAC
C (R)	CCT CAC TTC CAG ACT GAC AGC
D (F)	CAG TGG TGT GGT GTG ATC TTG
D (R)	GGC AAA ACC CTG TAT CTG TGA
F (F)	CCA TCA CGT CCA GCC TAT TT
F (R)	TGT GTG AGT CCA GGA GTT GG
Human γ-actin gene	
prom (F)	GGA AAG ATC GCC ATA TAT GGA C
prom (R)	TCA CCG GCA GAG AAA CGC GAC
intr1 (F)	CCG CAG TGC AGA CTT CCG AG
intr1 (R)	CGG GCG CGT CTG TAA CAC GG
ex5(F)	GTG ACA CAG CAT CAC TAA GG
ex5 (R)	ACA GCA CCG TGT TGG CGT
ex6 (F)	GTC AGG GTT GGA AAG TCC AAG
ex6 (F)	AAA TGC AAA CCG CTT CCA AC

A (F)	CAA CCT GGA AGT ATT GGG TAG
A (R)	CTT TCG ATT TGG AAG AAA TCG
B (F)	GGG TCA AGG GAT CGT TCT G
B (R)	GCC TGG AGC TCA GTA AGC
C (F)	GAG GTT TGA GAC TGC AGT GAG
C (R)	CAG ACA TAA TTT TGT GGG GTT TG
D (F)	GAG ACA GAG TCT TGC TCT GTT G
D (R)	TGT AGT CCC AGC TAC TCG GG
β-globin based constructs	
5’ ex1 (F)	GAG TTA GCT CAC TCA TTA GGC
5’ ex1 (R)	CTG TGT GAA ATT GTT ATC CGC
ex1 (F)	GAA CCC ACT GCT TAA GCC TC
ex1 (R)	CCT CAC CAC CAA CTT CAT CC
ex2 (F)	TTG GAC CCA GAG GTT CTT TG
ex2 (R)	GAG CCA GGC CAT CAC TAA AG
ex3 (F)	AGC TCC TGG GCA ACG TGC TGG
ex3 (R)	TTG TGG GCC AGG GCA TTA GCC
3’pA (F)	AAA AGG GAA TGT GGG AGG TC
3’pA (R)	AGC CTC ACC TTC TTT CAT GG
5’pause (F)	CTG CAA ACA GCT AAT GCA CA
5’pause (R)	TGA ATC CTT TTC TGA GGG ATG
pause (F)	GGG ACT ATT TGG GGG TGT CT
pause (R)	GTT TTC CCA GTC ACG ACG TT
3’MAZ (F)	GGA AAC TAT TAC TCA AAG GG
3’MAZ (R)	GTT TTC CCA GTC ACG ACG TT
A (F)	TTG CCT TCC TGT TTT TGC TC
A (R)	CCG CTG TTG AGA TCC AGT TC
intr1-ex2 (F)	GGA GAC CAA TAG AAA CTG GGC
5S rRNA	
5S (F)	AGC GTC TAC GGC CAT ACC
5S (R)	GGT ATT CCC AGG CGG TCT C
SETX	
SETX (F)	CTT CAT CCT CGG ACA TTT GAG
SETX (R)	TTA ATA ATG GCA CCA CGC TTC

Centromere 9	
Cen9 (F)	CCT ATG ATG AAA AAG GTA ATA TCT TC
Cen9 (R)	CTG AAA GCG CTT AAA ACG TCC
PCF11	
PCF11 (F)	CAT CCA TCA TGT TAT GAA GAT TAT C
PCF11 (R)	TCG TTC TTC CTT AAC TTT GGG AC
Mouse β-actin gene	
prom (F)	GAG GGG AGA GGG GGT AAA
prom (R)	GAA GCT GTG CTC GCG G
in1 (F)	TAG GGT CGC GGT GTG
in1 (R)	GGT TCC CAA TAC TGT GTA CTC
in3 (F)	AAG TGG CCT TGG AGT GTG
in3 (R)	CAA GGA GTG CAA GAA CAC AGC
in5 (F)	CAC CAT TCA CCA TCT TGT C
in5 (R)	TGA TCC ACA TCT GCT GG
pause (F)	AGG TTG GGT GGG TGT
pause (R)	AGA CTT CAT TAA AGG GCC TGT
pause2 (F)	CAG GAA GGT GTG AAG CCTA TGG
pause2 (R)	GGG CTG ATT TCC AGG A
C (F)	TCA GGC TGG GCA CTA TAG
C (F)	CAG TTC TGG CTG TCT TGG
D (F)	TCG GTG AGT GGT CTA ATA AGC
D (R)	AGA TGG AGG AGG GAA C
Mouse major sateliites	
majsat (F)	AAA TAC ACA CTT TAG GAC G
majsat (R)	TCA AGT GGA TGT TTC TCA TT

2.5 Tissue Culture

2.5a Growth and maintenance

HeLa and HEK293 cells were grown in DMEM media supplemented with 10% FCS, 100 units/ml penicillin and streptomycin at 37^{o}C with 5% CO_2. For MEF cells, DMEM was also supplemented with 1% non essential amino acids. Cells were passaged at 90% confluency.

2.5b Plasmid transfections

HIV promoter-driven β-globin/MAZ_4, mMAZ_4, MAZ_4ΔpA, ΔCoTC, β-globin/β-actin plasmids were co-transfected with TAT plasmid into HeLa cells (except experiment in Figure 3.3B, where TAT was omitted) as described previously (Dye and Proudfoot, 1999; Gromak et al. 2006; West et al. 2004). 1x10^6 cells were grown in 10 cm plates and transfected with 10 μg of reporter plasmid DNA and 1 μg of TAT plasmid plus 12 μl of Lipofectamine 2000 according to manufacturer's instructions. The cells were harvested 24 hr after transfection.

GFP-RNase H1 plasmid (Cerritelli et al. 2003) was a kind gift from R. J. Crouch (NIH, Bethesda, Maryland USA). Xrn2-flag plasmid (Kaneko et al. 2007b) was a kind gift from James Manley (Fairchild Centre, New York, USA). 20 μg of over-expression plasmids were transfected to 1x10^6 cells. The cells were harvested 48 hr after transfection.

2.5c RNA interference

One day 1, $1x10^6$ cells were split per well in a 24-well plate in DMEM media supplemented with 10% FCS (without streptomycin/penicillin). On day 2, in tube A, 3 µl of Lipofectamine 2000 were mixed with 12 µl of OPTIMEM and incubated for 7 mins. In tube B, 3 µl of a 20mM stock of siRNA duplex or siRNA buffer (mock sample) and 47 µl of OPTIMEM were also mixed. The mixtures from tube A and tube B were combined together and incubated at room temperature for 25 mins. Next, 35 µl of OPTIMEM was added and the mixture was added into a well. On day 3, cells growing in a well of a 24-well plate were split into a well of a 6-well plate. On day 4, cells were transfected as before with the addition of plasmids, if necessary. On day 5, cells were split to 10 cm dish or were let to recover by adding fresh media. On day 6, cells were harvested.

mRNA target sequence for *SETX* siRNA 1 was 5'-GCCAGAUCGUAUACAAUUA-3' (Dharmacon), for *SETX* siRNA 2 5'-CAATGGGTGGAAGCCGTCGTCAAAT-3' (Invitrogen) and for *PCF11* siRNA 5'-GGAUAGAAUUUGAGGAGAU-3' (Dharmacon).

2.6 Mammalian RNA isolation

Cells were washed with PBS and were harvested by adding 1 ml of Trizol reagent. RNA was then isolated following the manufacturer's instructions.

2.7 Polymerase Chain Reaction (PCR)

2.7a Standard PCR

50-100 ng of template was amplified in presence of 0.5 μM primer pairs, 0.2 mM dNTPs, 1.75 mM $MgCl_2$, Taq buffer and Taq polymerase. Primer annealing temperatures were chosen to be approximately 5 ^{o}C lower than the primer melting temperature. Primers were designed to have similar melting temperatures. Cycling conditions were determined according to manufacturer's guidelines.

2.7b Quantitative Real-Time PCR

Real-Time PCR is used to amplify and simultaneously quantify a DNA molecule. As opposed to standard PCR where the product of the reaction is detected at its amplification end point, the amplified DNA is detected as the reaction progresses in real time. Real-Time PCR measures the increase of fluorescence generated by the fluorescent dye SYBR Green which intercalates into the double-strand DNA product as it is generated. The general principle of DNA quantification by Real-Time PCR relies on plotting fluorescence against the number of cycles on a logarithmic scale. Comparison of samples to DNA standards of known concentrations allow relative quantification of samples DNA concentration.

Quantitative Real-Time PCR was performed using the QuantiTest SYBR green PCR master mix in a Corbett Rotorgene 3000 machine. Reactions were performed in triplicates to minimise experimental error and quantitated using the Rotorgene 3000

software. A SYBR Green melting curve was performed simultaneously to confirm the quality of the PCR product.

2.8 Steady-state RNA analysis

Total RNA was isolated using Trizol reagent and reverse transcribed with SuperScript III Reverse Transcriptase using either 50 pmol of oligo-dT or 5 pmol of a gene-specific primer, following the manufacturers' instructions.

Spliced β-globin mRNA was detected with ex1(F)/ex3(R) primers. The termination efficiency of the β-globin-based gene constructs was calculated as a ratio of read-through transcript A (PCR with AF/AR) and correctly initiated transcript B (PCR with intr1-ex2 (F)/ex2R), using qRT-PCR analysis. The termination efficiency of the endogenous human γ-actin gene was calculated as a ratio of the read-through transcript A (PCR with AF/AR), B (PCR with BF/BR), C (PCR with CF/CR) or D (PCR with DF/DR) and correctly initiated transcript ex5 (PCR with ex5F/ex5R). PCR cycles were performed at 94°C for 2 min, followed by 25 cycles of 94°C for 30 s, 58°C for 30 s, and 72°C for 30 s.

2.9 Nascent RNA analysis

2.9a Nuclear Run-On (NRO) analysis

Positions of the NRO probes were described previously (Ashe et al. 1997; Dye and Proudfoot, 1999; West et al. 2004) and are shown in Figure 3.4C. .NRO filters were prepared using a BoRad slot blot apparatus. ssDNA M13 probes resuspended in 5x SSPE were applied to 5x SSPE-washed nitrocellulose membrane. NRO filters were cross-linked

for 2 mins with 254 nm UV-light. The filters were pre-hybridised overnight at 42°C in a rotating hybridisation oven

HIV promoter-driven β-globin/MAZ_4 was co-transfected with TAT plasmid into HeLa cells. RNaseH1-GFP plasmid was transfected simultaneously when necessary. Cells were harvested in cold PBS and pelleted by centrifugation at 500 g for 5 mins. Cell pellets were resuspended in 4 ml HLB/NP40 (cell lysis), incubated on ice for 5 mins and then underplayed with 1 ml of HLB/NP40/sucrose solution (nuclear lysis). Lysed cells were then centrifuged for 5 mins. Nuclear pellets were resuspended in 2x transcription buffer and transferred to 1.5 ml tube. 10mM mix of rATP, rGTP, rCTP was added followed by 60μCi [α^{32}P]rUTP. The *in vitro* transcription reaction was performed for 30 mins at 30°C. Total RNA was isolated using Trizol reageant according to manufacturer's instructions. The RNA was then resuspended in 60 μl of water, partially hydrolysed on ice for 5 mins by addition of 15 μl of 1 M sodium hydroxide and finally neutralised by 30 μl of 0.1 M Tris/0.1 M HCl. Prepared RNA was hybridised to nitrocellulose filters at 42°C overnight. After hybridisation filters were washed in 1x SSPE/0.1% SDS at room temperature for 20 mins. Hybridisation signals were visualised by autoradiography and quantitated using PhosphoImager.

Solutions

HLB/NP40

Tris HCl, pH 7.5	10mM
Sodium Chlolide	10mM
Magnesium Chloride	10mM

NP40	0.5% (v/v)

HLB/NP40/sucrose

10% sucrose was added to HLB/NP40 solution

2x Transcription buffer

Tris HCl Ph 7.9	40mM
Potassium Chloride	300mM
Magnesium Chloride	10mM
Glycerol	40% (v/v)
Prior to use add:	
DTT	2mM

20x SSPE

Sodium Cloride	3.6M
Sodium dihydrogen phosphate	200mM
EDTA	20mM

Adjusted to pH 7.4 by addition of sodium hydroxide

Denaturing solution

Sodium Chloride	3.6M
Sodium hydroxide	0.5M

Neutralising solution

Tris HCl pH 7.2	0.5M
Sodium Chloride	1.5M
EDTA	1mM

100X Denhardts solution

Ficoll	2%
Polyvinylpyrrolidone	2%
Bovine serum albumin	2%

Pre-hybridisation solution

SSPE	6x
Formamide	50% (v/v)
SDS	0.1%
Denehardts	5x
tRNA	100 µg/ml

Hybridisation solution

SSPE	7x
Formamide	58% (v/v)
SDS	0.12%
Denehardts	5.8x
tRNA	100 µg/ml

2.9b Bromo-Uridine Triphosphate (Br-UTP) NRO analysis

The Br-UTP NRO was carried out largely as the standard NRO and as described previously (Lin et al. 2008) with some modifications. Nuclear pellets were resuspended in Transcription Buffer and 10 mM mix of rATP, rCTP, rGTP, and Br-UTP or rUTP (in the control samples). The NRO reaction was performed at 30°C for 30 min. Total RNA was isolated using Trizol reageant according to manufacturer's instructions and treated with RNase-free DNase I. 2 µl of anti-BrU antibody were pre-incubated with 30 µl of Protein G agarose beads and 10 µg tRNA per sample for 1 hr at 4°. The beads were washed three times with RSB-100 buffer and resuspended in 150 µl RSB-100 with 40 U RNase-OUT and 5 µg of glycogen. Total RNA was added to the beads and incubated for additional 1 hr at 4°C. Then the beads were washed three times with RSB-100 buffer. RNA-bound to the beads was extracted with Trizol reagent followed by two DNase I treatments. The RT reaction was performed using SuperScript III Reverse Transcriptase following the manufacturer's instructions. The PCR mixture contained QuantiTest SYBR Green PCR master mix, 2 µl of template cDNA, and primers from the Table 1. Cycling parameters were 95°C for 15 min, followed by 50 cycles of 95°C for 15 s, 58°C for 20 s, and 72°C for 20 s. The data was presented as a fold of enrichment of BrU-RNA over the U-RNA produced over a specific probe.

Solutions

HLB/N40, HLB/NP40/sucrose and 2x Transcription buffer were prepared as described in NRO analysis.

RSB-100

Tris HCl pH 7.4	10mM
Sodium Chloride	100mM
Magnesium Chloride	2.5mM
Triton X-100	0.4%

2.10 Chromatin Immunoprecipitation (ChIP) analysis

ChIP analysis was carried out as previously described (Kadener et al. 2002; West et al. 2004) with some modifications. Cells were fixed for 15 mins by addition of 36.5% methanol-stabilised formaldehyde solution, cross-linked was quenched by adding 1.32 ml of 1 M glycine and cells were washed and harvested in PBS. Cells were lysed in 300 µl of Cell lysis buffer and incubated for 10 mins on ice. The lysed cells were then centrifuged to pellet nuclei. The nuclear pellet was finally resuspended in 400 µl of Nuclear lysis buffer. Sonication was performed in 1.5 ml tubes in a diagenode Bioruptor for 15 mins at medium power with pulses of 30 secs. Lysates were pre-cleared with 50 µl of Protein A/G agarose beads for 1 hr at 4°C and then immuno-precipitated with 3-5 µg of antibody overnight. Washes were performed with buffers A, B, C and 2xD and elution twice with elution buffer. Eluted samples were reverse crosslinked for 4-5 hr at 65°C with 0.3 M NaCl and 3 µg/ml RNase A. Proteinase K treatment was performed for 2 hr at 45°C with 10 mM EDTA, 40 mM Tris-HCl pH 6.5 AND 20 µg proteinase K. The chromatin was purified through Qiagen PCR-clean up columns. The immuno-precipitated, non-precipitated, and input DNAs were used as templates for qPCR. The

PCR mixture contained QuantiTect SYBR green PCR master mix, 2 ul of the template DNA and primers from the Table 1. Cycling parameters were 95°C for 15 min, followed by 50 cycles of 95°C for 15 s, 58°C for 20 s, and 72°C for 20 s. ChIP signal was subtracted for signal arising from non-precipitated sample and then divided to input value. Final concentrations are shown as a % of the input value.

Solutions

Cell lysis buffer

NP40	0.5%
Potassium Chloride	85mM
PIPES pH 8	5mM
PMSF	1mM
Leupeptin	1 μg/ml
Pepstatin	1 μg/ml

Nuclear lysis buffer

SDS	1%
EDTA	10mM
Tris HCl pH 8	50mM
PMSF	1mM
Leupeptin	1 μg/ml
Pepstatin	1 μg/ml

IP dilution buffer

SDS	0.01%
EDTA	1.2mM
Tris HCl pH 8	16.7mM
Triton X-100	1.1%
Sodium Chloride	167mM
PMSF	1mM
Leupeptin	1 μg/ml
Pepstatin	1 μg/ml

Wash buffer A

SDS	0.1%
Triton X-100	1%
EDTA	2mM
Tris HCl pH 8	20mM
Sodium Chloride	150mM

Wash buffer B

SDS	0.1%
Triton X-100	1%
EDTA	2mM
Tris HCl pH 8	20mM
Sodium Chloride	500mM

Wash buffer C

Lithium Chloride	0.25M
NP40	1%
Sodium deoxycholate	1%
Tris HCl pH 8	10mM
EDTA	1mM

Wash buffer D

10:1 Tris HCl pH 8: EDTA

Elution buffer

SDS	1%
Sodium hydrogen carbonate	0.1M

2.11 DNA Immunoprecipitation (DIP) analysis

DIP analysis was performed without a cross-linking step, following the ChIP protocol with some modifications. After the nuclear lysis reaction, extracts were incubated with 30 μg of proteinase K at 55°C for 3 hr and genomic DNA was isolated. Following sonication, DIP analysis was carried out using antibody, recognising RNA/DNA hybrids, purified from S9.6 hybridoma cell lines (kind gift from Richard Gibbons, Boguslawski et al.1986). The immuno-precipitated, non-precipitated, and input DNAs were used as templates for qPCR. DIP RNase-sensitivity analysis was carried out following the genomic isolation and prior to immuno-precipitation step with the addition

of 200 U of S1 Nuclease or 10 U of RNase H. 100 µl nuclease digestion reaction contained 10x reaction buffer suggested by the manufacturer's instructions, and it was performed for 2 hr at 37°C.

Solutions

Solutions used were the same as in ChIP analysis.

2.12 Protein extraction and Western blot analysis

Cells were washed with PBS. Cell pellets were resuspended in RIPA lysis buffer. Cell lysis was performed for 20 mins on ice. Protein lysate was recovered by centrifugation and protein-containing supernatant was kept. Protein concentration was measured by Bradford assay. Protein samples were boiled for 10 mis with SDS/beta-mercaptoethanol protein loading dye and then loaded on a SDS/PAGE gel. Protein samples were then transferred to a PVDF membrane using semi-dry transfer apparatus. Ponceau staining was confirmed by Ponceau staining. Membrane was blocked in TBS 5% dry milk overnight at 4°C. Next the primary antibody was added (1:1000 for anti-senataxin and anti-Pf11; 1:2000 for anti-GFP and anti-actin; 1:500 for anti-flag) for 1.5 hr at room temperature. Three washes with PBS were performed and a secondary antibody (1:4000 dilution for horseradish peroxidise coupled anti-mouse or anti-rabbit antibody) was added and incubated for 1 hr. The membrane was finally washed for three times in PBS. Signal was detected using an ECL detection kit.

Solutions

RIPA lysis buffer

Tris HCl pH 7.5	50mM
Sodium Chloride	150mM
NP40	150mM
Deoxycholate	0.5%
SDS	0.1%
Glycerol	10%
PMSF	1mM

2.13 Co-Immunoprecipitation (Co-IP) analysis

Cells were washed with PBS. Cell pellets were resuspended in 1 ml of RSB-100 and cells were incubated for 10 mins on ice. Sonication was performed for 3 mins to further lyse the cells in a diagenode Bioruptor at medium power with pulses of 30 secs. Cell debri were pelleted by centrifugation and the cell lysates were collected. 20 µl were then saved as input while the remaining sample was immuno-precipitated with 3 µg of antibody for 1.5 hr at 4°C. 15 µl of protein A/G agarose beads were then added and incubated for 1.5 hr. The beads were washed three times with RSB-100 buffer. Co-immunoprecipitated complex remained at beads was finally separated by SDS-PAGE and detected by Western blot analysis.

Solutions

RSB-100

Tris HCl pH 7.4	10mM
Sodium Chloride	100mM
Magnesium Chloride	2.5mM
Triton X-100	0.4%
PMSF	1mM
Leupeptin	1 μg/ml
Pepstatin	1 μg/ml

CHAPTER 3

The role of human senataxin protein and R-loop structures in pause-dependent transcriptional termination of transfected gene constructs

3.1 Introduction

As detailed in the Introduction, efficient Pol II termination is dependent on both a functional poly(A) signal and downstream terminator sequences (Connelly and Manley, 1988; West et al. 2004; West et al. 2008; Whitelaw and Proudfoot, 1986). So far, two classes of terminator sequences have been identified in human genes: co-transcriptionally cleaved (CoTC) RNA sequences and transcription pause sites. The human β-globin gene contains a CoTC element, located ~1kb downstream of the poly(A) site, which is essential for efficient Pol II transcriptional termination (Dye and Proudfoot, 1999; 2001). Alternatively, some terminator sequences may act as transcription pause sites that slow down elongating Pol II and promote its transcriptional termination. Such G-rich pause elements have been identified in C2 complement gene (MAZ_4) (Ashfield et al. 1994). The human β-actin gene also possesses a G-rich pause element downstream of the poly(A) signal which promotes transcriptional termination, through degradation of the downstream transcript by the 5'-3' exonuclease Xrn2 (Gromak et al. 2006). Interestingly, genome-wide analysis revealed that G-rich sequences placed immediately downstream of poly(A) signals are relatively common in mammalian genes (Salisbury et al. 2006), suggesting that these elements may play an essential function in regulation of gene expression.

Recently, it has been demonstrated that Sen1 protein, a RNA/DNA helicase, is required to promote Pol I and Pol II transcriptional termination in *S. cerevisiae* (Kawauchi et al. 2008). The research to date carried out on human senataxin is limited. Senataxin interacts with Pol II and other RNA processing factors, such as polyA-binding proteins 1 and 2 (PABP1/2), suggesting its role in transcriptional regulation and mRNA

processing (Suraweera et al. 2009). However, the mechanism of senataxin function in transcription is not clear.

The majority of senataxin mutations found in AOA2/ALS4 disorders, either cause premature protein termination or interfere with the function of the helicase or N-terminal protein interaction domain of the protein (Chen et al. 2004; Moreira et al. 2004; Duquette et al. 2005; Criscuolo et al. 2006; Fogel and Perlman, 2006), suggesting their importance in senataxin function. Furthermore, the function of yeast Sen1 in transcription is likely to be associated with Sen1 RNA/DNA helicase activity, which acts to resolve RNA/DNA hybrids (R-loops), formed between nascent RNA and template DNA strand behind the elongating Pol II during transcription (Mischo et al. 2011).

With the results presented in this Chapter, I wished to investigate the role of senataxin in the regulation of gene transcription and whether its molecular function is associated with the R-loop resolution. To address this, I firstly analysed gene constructs with different pause termination sequences. The use of transiently tranfected gene plasmids might be considered not physiological as this is an exogenous system and that may not fully reflect the mechanism of an endogenous gene system. However the benefit of working with gene constructs is the ability to manipulate the system and use plasmids with *cis* mutations affecting transcriptional termination.

Here, I used two previously described gene constructs containing the β-actin pause element or MAZ_4 pause sequences, cloned downstream of the β-globin poly(A) signal instead of its termination CoTC element (Gromak et al. 2006) (Figure 3.1B). Transcription of these gene constructs is driven by the efficient HIV promoter, which is fully activated by co- expression of TAT trans-activator. Both the β-actin pause element

and MAZ_4 G-rich pause sequences have been previously shown to promote poly(A) site- and Xrn2- dependent transcriptional termination (Gromak et al. 2006).

3.2 Senataxin promotes transcriptional termination on transfected gene constructs

To investigate if senataxin is involved in transcriptional termination, I initially employed RNA interference (RNAi) to deplete senataxin in HeLa cells. The cells were transfected with a siRNA duplex specifically targeted to *SETX* mRNA. As control, mock tranfections were performed in parallel where the duplex was omitted. Senataxin mRNA and protein levels were depleted to a 15% level using *SETX* siRNA 1, as measured by qRT-PCR and Western blot analyses (Figure 3.1A). Western blot analysis (Figure 3.1A, right panel carried out by N. Gromak) was performed using an anti-senataxin antibody (Bethyl Laboratories Inc.) which selectively detects the 303 kDa senataxin protein in mock-treated HeLa cells.

Next we tested if the transcriptional termination profile of gene constructs transiently transfected into HeLa cells is affected in senataxin-depleted cells. The termination efficiency of β-globin/β-actin and β-globin/MAZ_4 constructs in senataxin-deficient cells was assessed using quantitative RT-PCR (qRT-PCR) analysis. In this experiment I measured the amount of read-through transcript A and the amount of correctly initiated transcript B (Figure 3.1B). The ratio of A/B represents the efficiency of transcriptional termination. As demonstrated in Figure 3.1B, senataxin depletion causes 1.8-2.2 fold increase in the amount of the read-through transcript A for both β-globin/β-actin and β-globin/MAZ4 constructs, pointing towards a role for senataxin in

transcriptional termination. To avoid potential off-target effects during RNAi procedure, I also used *SETX* siRNA 2, targeting a different region of *SETX* mRNA. The efficiency of the knock-down was measured by qRT-PCR (Figure 3.2, left panel). As demonstrated in Figure 3.2 (right panel) performing qRT-PCR, senataxin depletion with *SETX* siRNA 2 causes a transcriptional termination defect of β-globin/MAZ_4 construct.

3.3 R-loops form over transfected gene constructs

A potentially deleterious consequence of transcription is the tendency of nascent transcripts to interact with the DNA template behind the elongating Pol II, forming RNA/DNA hybrids (R-loops). Senataxin protein possesses a C-terminal RNA/DNA helicase domain, therefore its function may be associated with its RNA/DNA helicase activity, possibly resolving R-loop structures formed during transcription. In order to investigate the potential role of senataxin in R-loop resolution, initially I tested whether RNA/DNA hybrids are formed over the coding regions of human genes. DNA immuno-precipitation (DIP) was employed, which relies on the immuno-precipitation of non-crosslinked purified RNA/DNA hybrids using S9.6 antibody that recognizes RNA/DNA duplexes (Boguslawski et al. 1986; El Hage et al. 2010). As shown in Figure 3.3A, R-loops are formed over the transcription units of β-globin/β-actin and β-globin/MAZ_4 constructs, showing some enrichment in RNA/DNA hybrid signal at the 5'end of each transcription unit. To verify that the detected R-loop signal is dependent on active transcription, I also performed DIP analysis on HeLa cells transfected with β-globin/β-actin gene construct without the trans-activator TAT, which is required for full HIV promoter activity. In this experiment I observed a reduction of R-loop signal over the

coding regions of the gene (Figure 3.3B, left panel). Some R-loop signal still remained in -TAT samples possibly due to significant levels of 'leaky' transcription observed in HeLa cells, transfected with β-globin/β-actin gene construct. As shown in Figure 3.3B (right panel), reduced mRNA levels are still detected in -TAT samples. These results confirm the transcriptional dependency of R-loops.

Next I sought to investigate if cis-mutations affecting transcriptional termination of β-globin/MAZ_4 construct can affect the R-loop distribution profile. Here I used constructs containing either mutated poly(A) signal or mutant MAZ sequences ($mMAZ_4$), that do not promote transcriptional pausing. Both of these mutations were previously shown to cause transcriptional termination defects (Gromak et al., 2006). Both mutations decreased the levels of R-loop signal across the β-globin/MAZ_4construct (Figure 3.3C), suggesting that R-loops depend on functional poly(A) signals and G-rich pause elements.

To investigate the role of senataxin in R-loop resolution over these transfected gene constructs, I employed DIP analysis across β-globin/β-actin and β-globin/MAZ_4 gene constructs in senataxin-deficient cells. Interestingly, a decrease in R-loops over the coding regions of the gene constructs was detected combined with a significant increase upstream of the β-actin and MAZ_4 pause elements (Figure 3.3D). These results suggest a potential role of senataxin in R-loop resolution that may be connected to transcriptional termination.

3.4 RNase H1 over-expression resolves R-loops and causes transcriptional read-through *in vivo*

RNase H is an enzyme that specifically degrades the RNA in RNA/DNA hybrids that form during replication and DNA repair. Both in bacteria (*Escherichia coli*) and eukaryotes there are two types of RNAse H, however eukaryotic RNases H are larger and more complex. RNase H1 consists of one single protein, whereas eukaryotic RNase H2 is a complex of three proteins (Cerritelli and Crouch, 2009).

R-loops can be removed by cleavage using RNase H1 or H2, which have different cleavage patterns but similar catalytic mechanisms and overlapping functions (Arudchandran et al. 2000).

Taking advantage of the RNase H1 specificity, I over-expressed GFP-tagged RNase H1 plasmid (Cerritelli et al. 2003) in HeLa cells to investigate if R-loops are directly associated with the transcriptional termination process. The level of RNase H1 over-expression was measured by Western blot using anti-GFP antibody as compared to endogenous actin (Figure 3.4A, carried out by N. Gromak). Then I investigated if over-expression of RNase H1 *in vivo* can affect transcriptional termination of the β-globin/β-actin and β-globin/MAZ_4 constructs. As controls I used constructs either lacking termination elements (ΔCoTC) or containing mutant MAZ sequences ($mMAZ_4$) (Dye and Proudfoot, 1999; Gromak et al. 2006). The termination efficiency was measured by qRT-PCR with primers positioned over regions A and B (see Figure 3.1B for primer positions). As demonstrated in Figure 3.4B, progressive over-expression of RNase H1 caused an increase in the amount of read-through transcript A of globin/β-actin and β-globin/MAZ_4, but not globin/$mMAZ_4$ and

ΔCoTC plasmids. This suggests a specific effect of RNase H1 on constructs containing functional terminator G-rich pause elements.

To further confirm this result, the amount of active transcription was measured on the β-globin/MAZ_4 construct in mock-treated and RNase H1-over-expressing HeLa cells, using nuclear run on analysis (Figure 3.4C). NRO analysis involves isolation of HeLa cell nuclei in such a way that *in vivo* transcription is stalled but transcribing polymerases remain actively bound to their DNA templates. These can resume transcription after addition of ^{a32}P-UTP and ribonucleotides *in vitro.* The nascent radiolabelled RNA is then extracted and hybridized to single-strand DNA probes across the gene. Here I employed a number of single-stranded anti-sense DNA probes (cloned into the single strand M13 phage DNA), covering β-globin/MAZ_4 regions of promoter (P), exon 3 (B3), downstream of poly(A) site (B4), downstream of the termination pause element (A), and upstream of the promoter (U3) . M13 probe (M13 vector DNA) represents the background. High levels of transcription are usually detected over the gene body probes P, B3 and B4. Signals above the M13 probe background over probes A and U3 define the degree of transcriptional termination. In mock-treated HeLa cells an efficient transcriptional termination of the β-globin/MAZ_4 construct was observed as indicated by low signals, over the background, over probes A and U3. Interestingly, over-expression of 2 μg of RNase H1 caused a two fold increase in the amount of read-through transcription, observed over probe A. A slight increase in the amount of read-through transcription over the probe U3 was also observed, however the effect was not as dramatic as over probe A. In addition, I also detected read-through defect in cells depleted of senataxin protein, as

indicated by a 3-4 fold increase in A and U3 signals. These results provide further support for the specific role of senataxin and RNase H1 in transcriptional termination.

To investigate the molecular mechanism of RNase H1 function in transcriptional termination, we studied the R-loop profile of the β-globin/MAZ_4 construct following RNase H over-expression. Using DIP analysis in Figure 3D, we show that over-expression of RNase H1 in HeLa cells diminishes R-loop levels over both gene and termination regions of the β-globin/MAZ_4 construct. These data suggest that R-loops are involved in the process of transcriptional termination, and their resolution by RNase H1 induces transcriptional read-through. Next I tested if RNase H1 over-expression and its ability to resolve R-loops increase the termination defect that was observed in senataxin-deficient cells. As shown in Figure 3.4E, RNase H1 over-expression enhanced transcript read-through in β-globin/MAZ_4 construct in both mock-treated and senataxin-depleted cells. These results suggest that R-loops are required for transcriptional termination but once formed; they have to be resolved by senataxin.

3.5 Discussion

Experiments based on the transfected gene constructs in this chapter have revealed that senataxin plays a role in transcriptional termination. Using qRT-PCR and NRO approaches, I demonstrate that senataxin-deficient HeLa cells have transcription termination defects (Figures 3.1B, 3.2 and 3.4C). The initial hypothesis is that the function of senataxin protein in human cells may be related to its RNA/DNA helicase activity, and in particular, its ability to unwind RNA/DNA hybrids formed between

nascent RNA transcript and single stranded DNA template formed behind the elongating Pol II.

Using DNA-immuno-precipitation (DIP) technique, I also show that R-loops can be efficiently detected over transcribed regions of the transfected gene constructs (Figures 3.3A). Figures 3.3B and 3.3C indicate that formation of R-loops is dependent upon active transcription and the presence of either a functional poly(A) signal or termination G-rich pause elements. Previous work from our lab has demonstrated that due to a strong termination defect, both of these construct have a reduced level of transcription. This results in decreased R-loop levels over these gene regions as observed in this experiment. This effect is similar to that seen for the construct transfected without TAT (Figure 3.3B). The fact that poly(A) signal and termination pause elements are both associated with efficient transcriptional termination and the formation of R-loops, implies that R-loops represent a signature for a properly defined transcription unit, containing a functional promoter and terminator elements.

Interestingly, senataxin depletion caused accumulation of R-loops downstream the poly(A) signal, supporting its function in R-loops resolution in termination regions (Figure 3.3D). In addition, experiments presented on Figures 3.4C and 3.4D indicate that RNase H1 over-expression *in vivo* resolves R-loops, resulting in transcriptional read-through. In effect R-loops formed over pause elements can be considered as an essential component of the termination process. But once they are formed, their removal by senataxin is also required, since senataxin depletion and RNase H1 over-expression have additive defects on transcriptional termination (Figure 3.4E).

The fact that R-loops play an important role in the process of transcriptional termination raises interesting issues for further investigation. So far, it has been suggested that transcription-induced R-looping is an evolutionary conserved process that may have deleterious consequences (Li and Manley, 2005; Li and Manley, 2006). Extensive formation of RNA/DNA hybrids causes genomic instability in yeast and vertebrate cells (Li and Manley, 2005, Mischo et al., 2011). The results presented here in contrast imply a positive role for R-loops that need to be formed for efficient pause-dependent termination to occur. So far, two types of termination elements have been identified in mammals-the CoTC and pause elements. In the future, it would be interesting to investigate whether R-loops form in the CoTC–dependent termination process or are specific for pause-mediated termination.

Experiments presented in this chapter have revealed that human senataxin protein and R-loop structures are important features of the pause-dependent transcriptional termination. The above experiments rely on analysis of transfected gene constructs with different termination sequences. To validate these data using an endogenous gene system (in native chromatin) I have also investigated the role of senataxin in termination of the endogenous β-actin and γ-actin genes in HeLa cells as presented in the following chapter.

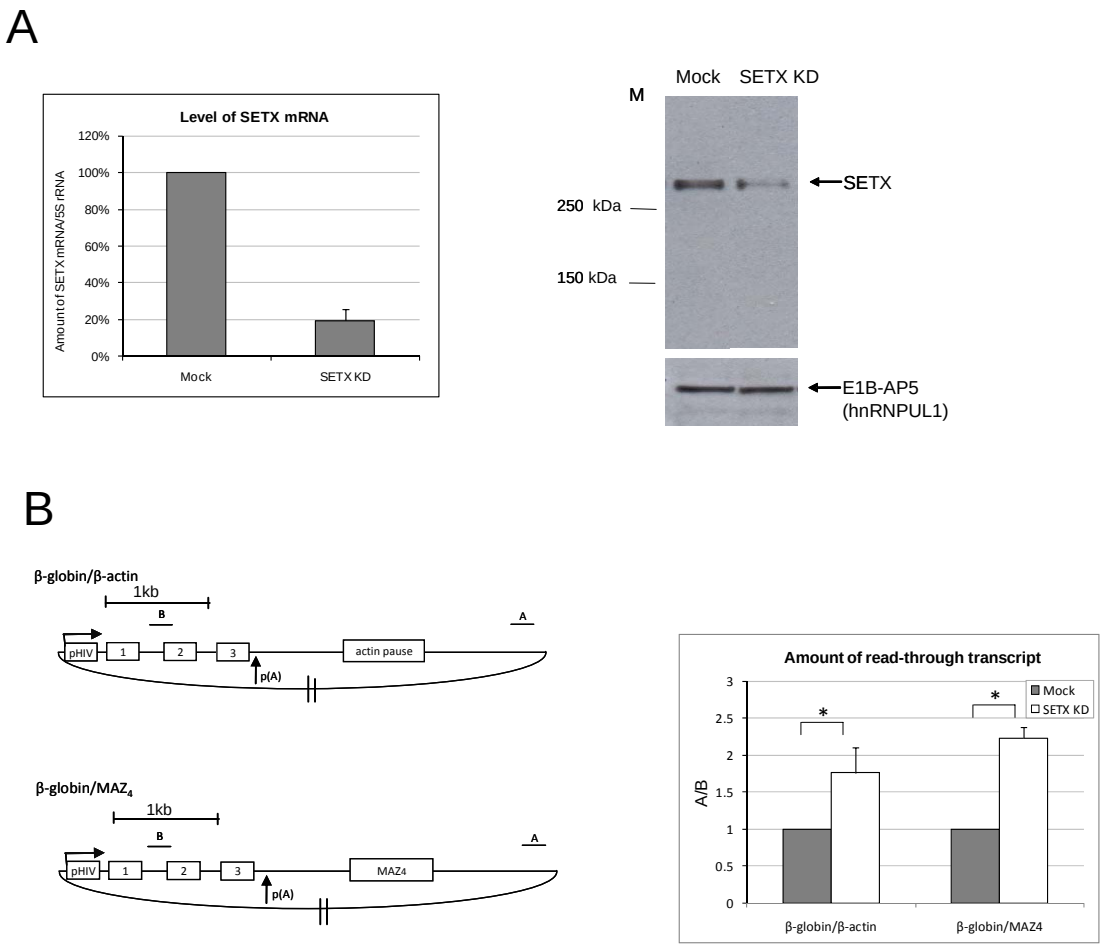


Figure 3.1 Transcriptional termination defect of constructs with G-rich pause sites in senataxin-deficient HeLa cells.

A. RNAi-mediated depletion of human senataxin in HeLa cells. Left panel: qRT-PCR analysis of mock- and senataxin siRNA-treated HeLa cells. Senataxin was depleted to a 15% level as compared to the level of 5S rRNA used as a control. Right panel: Western blot analysis of 60 μg of protein extracts from mock- and senataxin siRNA-treated HeLa cells. Western blot was analysed using anti-senataxin antibody. E1B-AP5 protein was used as a loading control.

B. Left panel: Diagram of β-globin/β-actin and β-globin/MAZ_4 gene constructs, with exons shown as boxes and pause elements labelled as 'actin pause' and 'MAZ_4'. Scale in gene constructs is indicated. Plasmid backbone is not drawn to scale. Position of poly(A) signal is indicated with an arrow. A and B denote positions of PCR products used to calculate termination efficiency of the transfected constructs (A/B). Right panel: Termination efficiency of β-globin/β-actin and β-globin/MAZ_4 constructs from mock-treated and senataxin siRNA-treated cells using qRT-PCR analysis. Termination efficiency was calculated as a ratio of RT-PCR product A, representing read-through RNA, vs RT-PCR product B, representing the level of efficient transcriptional initiation from three independent experiments; values are +/- SD. The A/B ratio in mock-treated cells was taken as 1.

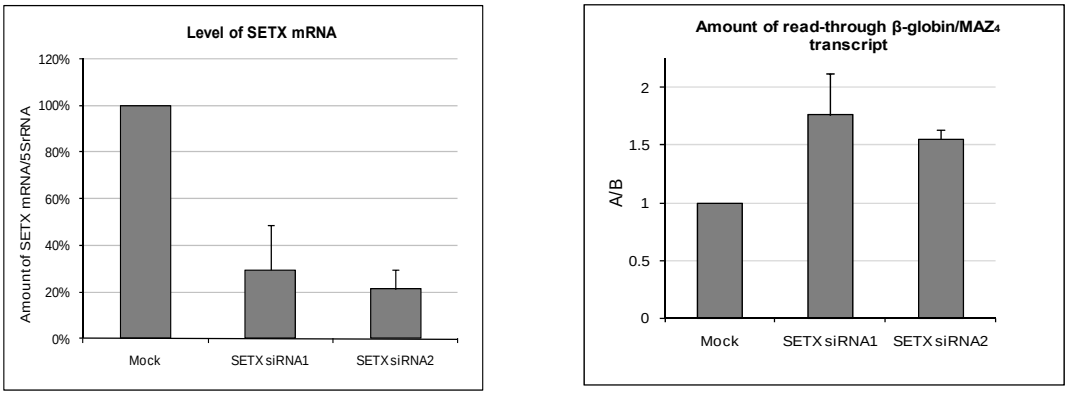
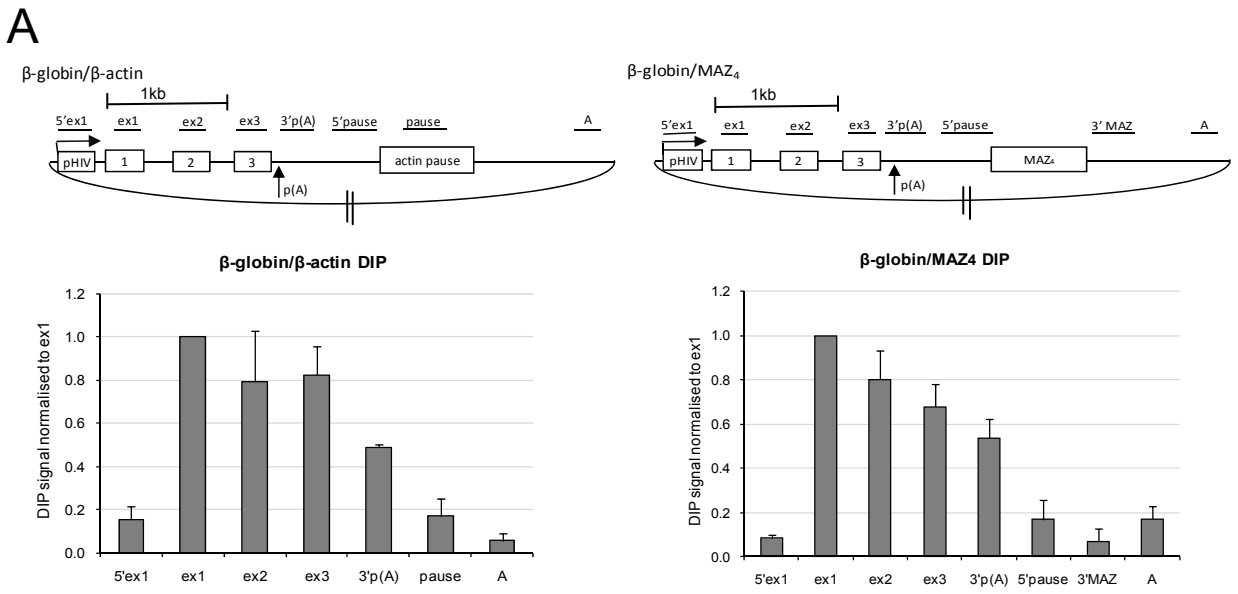


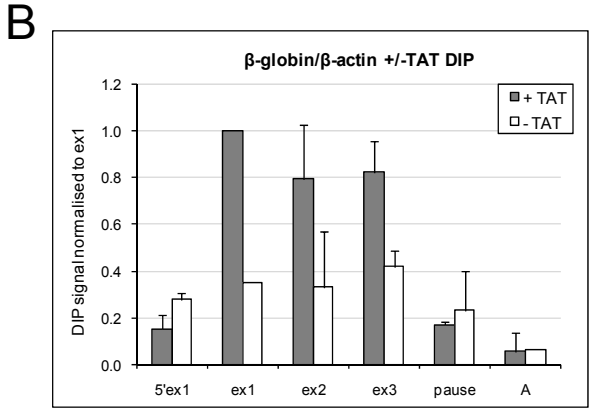
Figure 3.2 Transcriptional termination defect of β-globin/MAZ_4 construct in HeLa cells treated with SETX siRNA 2.

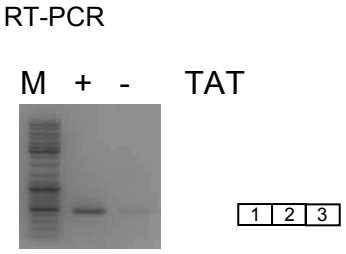
RNAi-mediated depletion of human senataxin in HeLa cells using *SETX* siRNA duplexes 1 and 2. Left panel: qRT-PCR analysis of mock- and senataxin siRNA-treated HeLa cells. Bars represent average values +/- SD from three independent biological experiments. 5S rRNA was used as an internal control. The amount of *SETX* mRNA in mock cells was taken as 100%. Right panel: Termination efficiency of β-globin/MAZ_4 construct from mock-treated and cells treated with *SETX* siRNA 1 and 2, using qRT-PCR analysis on the total RNA. Amount of read-through transcript was calculated as a ratio of RT-PCR product A, representing read-through RNA, vs RT-PCR product B, representing the level of efficient transcriptional initiation.The A/B ratio in mock-treated cells was taken as 1. Bars represent average values +/- SD from three independent biological experiments.

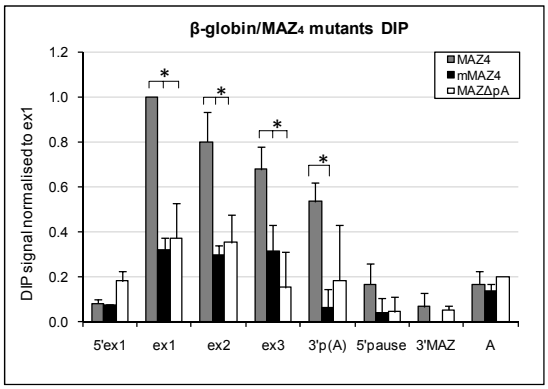

A
β-globin/β-actin
1kb
5'ex1
ex1
ex2
ex3
3'p(A)
5'pause
pause
A
pHIV
1
2
3
actin pause
p(A)
β-globin/MAZ4
1kb
5'ex1
ex1
ex2
ex3
3'p(A)
5'pause
3' MAZ
A
pHIV
1
2
3
MAZ4
p(A)

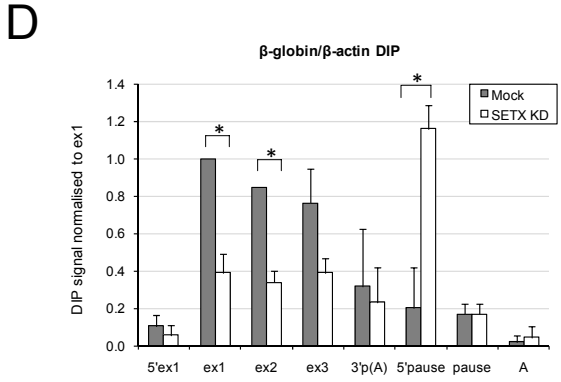
β-globin/β-actin DIP
DIP signal normalised to ex1
1.2
1.0
0.8
0.6
0.4
0.2
0.0
5'ex1
ex1
ex2
ex3
3'p(A)
pause
A

β-globin/MAZ4 DIP
DIP signal normalised to ex1
1.2
1.0
0.8
0.6
0.4
0.2
0.0
5'ex1
ex1
ex2
ex3
3'p(A)
5'pause
3'MAZ
A


B
β-globin/β-actin +/-TAT DIP
DIP signal normalised to ex1
+ TAT
- TAT
1.2
1.0
0.8
0.6
0.4
0.2
0.0
5'ex1
ex1
ex2
ex3
pause
A


RT-PCR
M + - TAT
1 2 3


C
β-globin/MAZ4 mutants DIP
DIP signal normalised to ex1
MAZ4
mMAZ4
MAZΔpA
*
1.2
1.0
0.8
0.6
0.4
0.2
0.0
5'ex1
ex1
ex2
ex3
3'p(A)
5'pause
3'MAZ
A


D
β-globin/β-actin DIP
DIP signal normalised to ex1
Mock
SETX KD
*
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
5'ex1
ex1
ex2
ex3
3'p(A)
5'pause
pause
A

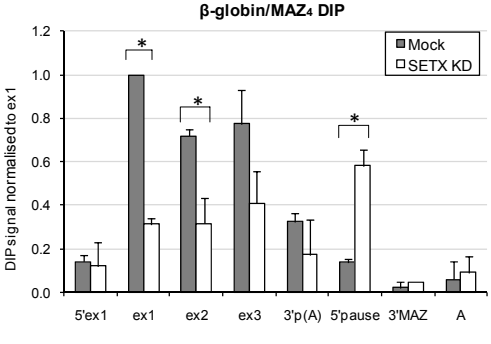

β-globin/MAZ4 DIP
DIP signal normalised to ex1
Mock
SETX KD
*
1.2
1.0
0.8
0.6
0.4
0.2
0.0
5'ex1
ex1
ex2
ex3
3'p(A)
5'pause
3'MAZ
A

Figure 3.3 R-loops form over transfected gene constructs and depend on transcription, a functional poly(A) signal and G-rich pause elements.

A. Top panels: Diagram of β-globin/β-actin (left) and β-globin/MAZ4 (right) gene constructs as in Figure 3.1B. Positions of primers used in DIP analysis are shown above the diagram. Bottom panels: DIP analysis of β-globin/β-actin (left) and β-globin/MAZ4 (right) constructs. R-loops are detected over the transcription unit of the constructs, with some enrichment over their 5' ends.

B. R-loops are dependent on active transcription. Left panel: R-loop profile of cells transfected with β-globin/β-actin gene construct +/- trans-activator TAT. Right panel: RT-PCR analysis of the spliced mRNA detected from HeLa cells transfected with β-globin/β-actin construct +/- TAT. M indicates the molecular weight DNA markers.

C. R-loop profile of β-globin/MAZ_4 (grey bars), mMAZ_4 (black bars) and MAZ_4ΔpA (white bars) gene constructs.

D. R-loop profile of β-globin/β-actin (left panel) and β-globin/MAZ_4 (right panel) gene constructs in mock-treated (grey bars) and senataxin siRNA-treated (white bars) cells. Upon senataxin depletion R-loops decrease over the coding region of the gene constructs and accumulate upstream of the pause elements.

DIP profiles are based on average values +/- SD from three independent biological experiments. * indicates statistical significance ($p<0.05$), based on unpaired, two-tailed distribution Student's t test. Scale in gene constructs is indicated. Plasmid backbone is not drawn to scale.

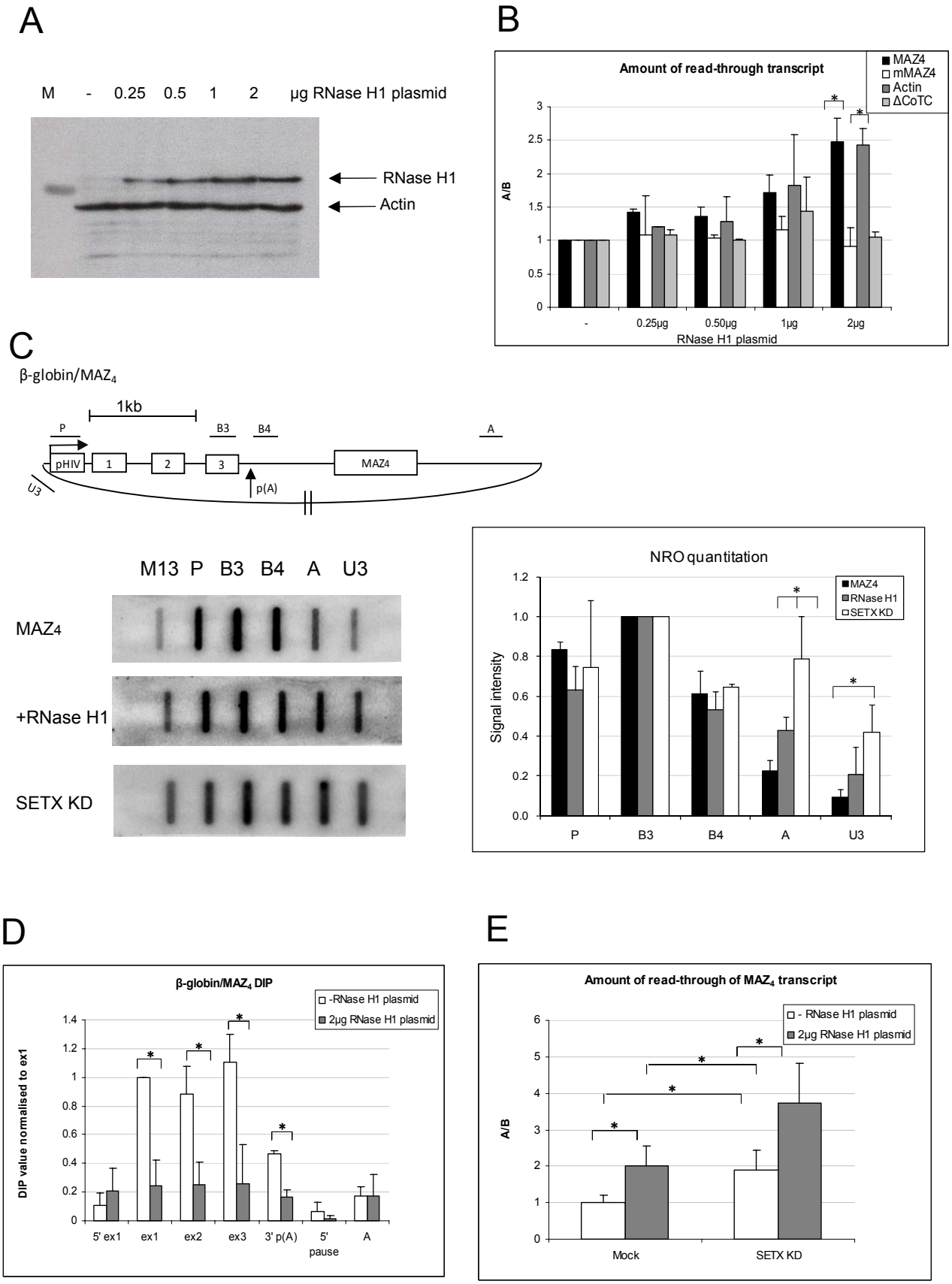

A
M - 0.25 0.5 1 2 µg RNase H1 plasmid
RNase H1
Actin
B
Amount of read-through transcript
MAZ4
mMAZ4
Actin
ΔCoTC
A/B
RNase H1 plasmid
C
β-globin/MAZ4
1kb
P
B3
B4
A
pHIV
1
2
3
MAZ4
U3
p(A)
M13 P B3 B4 A U3
MAZ4
+RNase H1
SETX KD
NRO quantitation
MAZ4
RNase H1
SETX KD
Signal intensity
P
B3
B4
A
U3
D
β-globin/MAZ4 DIP
-RNase H1 plasmid
2µg RNase H1 plasmid
DIP value normalised to ex1
5' ex1
ex1
ex2
ex3
3' p(A)
5' pause
A
E
Amount of read-through of MAZ4 transcript
- RNase H1 plasmid
2µg RNase H1 plasmid
A/B
Mock
SETX KD

Figure 3.4 RNase H1 resolves R-loops and promotes transcriptional read-through.

A. Western blot analysis of 30 μg of protein extracts obtained from HeLa cells co-transfected with increasing amounts (0-2 μg) of GFP-RNase H1 expression plasmid. Western blot was analysed using anti-GFP antibody. Actin was used as loading control.

B. Amount of read-through transcript detected in β-globin/β-actin, β-globin/MAZ_4, β-globin/$mMAZ_4$ and ΔCoTC constructs in mock HeLa cells (sample -) and HeLa cells, cotransfected with increasing amounts of GFP-RNase H1 plasmid. The amount of read-through transcripts was calculated as a ratio of A/B, using qRT-PCR analysis (see Figure 1B for primer positions). The A/B ratio in mock-treated cells (sample -) was taken as 1. Increased RNase H1 over-expression plasmid caused an increase in read-through β-globin/β-actin and β-globin/MAZ_4 transcripts, but not β-globin/$mMAZ_4$ and ΔCoTC controls.

C. Top panel: Diagram of β-globin/MAZ_4 gene construct with positions of NRO probes indicated above the diagram. Bottom left and right panels: NRO analysis of β-globin/MAZ_4 construct in mock-treated cells (black bars on the right panel), HeLa cells, cotransfected with 2 μg of GFP-RNase H1 plasmid (grey bars) and senataxin siRNA-treated cells (white bars). Position of the NRO probes is indicated above the filters. M13 represents a background probe. Right panel: NRO signals were quantified by PhosphoImager, where the background (signal from probe M13) was subtracted from each hybridization signal followed by signal normalization to probe B3, taken as 1. All bars represent average values +/- SD from three independent biological experiments.

D. DIP analysis of β-globin/MAZ4 construct following over-expression of 2μg of GFP-RNase H1 plasmid in HeLa cells. Positions of primers used in DIP analysis are shown in Figure 3.3A. DIP profiles are based on average values +/- SD from three independent biological experiments. DIP signal was normalized to exon 1 signal of β-globin/MAZ_4 construct without GFP-RNase H1 over-expression. RNase H1 over-expression causes a reduction of R-loops both over gene and termination regions of β-globin/MAZ_4 construct.

E. Amount of read-through β-globin/MAZ_4 transcript in mock-treated and senataxin depleted HeLa cells with or without over-expression of 2 μg of GFP-RNase H1 plasmid. The amount of read-through transcript was calculated as a ratio of A/B, using qRT-PCR analysis (for primer positions see Figure 1B). The A/B ratio in mock-treated cells (-RNase H1 sample) was taken as 1. RNase H1 over-expression increased read-through transcripts both in mock-treated and senataxin-depleted cells.

Average qRT-PCR values +/- SD from four independent biological experiments are presented in Figures 3.4B and E. * indicates statistical significance ($p<0.05$), based on unpaired, two-tailed distribution Student's t test. Scale in gene constructs is indicated. Plasmid backbone is not drawn to scale.

CHAPTER 4

The mechanism of R-loop dependent transcriptional termination in endogenous human genes: The role of senataxin and Pcf11

4.1 Introduction

Two models were initially proposed to account for the mechanism of transcriptional termination in protein-coding genes. The first model, known as the allosteric/anti-terminator model, suggests that Pol II termination occurs in response to conformational changes in the Pol II complex due to transcription through the poly(A) site (Logan et al. 1987). The second 'torpedo' model, proposes that Pol II termination occurs in response to degradation of the 3' flanking transcript (Connelly and Manley, 1988). This latter model has been supported by the demonstration that the human 5'-3' exonuclease Xrn2 and yeast homologue Rat1 are required for efficient termination (West et al. 2004; Kim et al. 2004). However, recent evidence shows that the termination mechanism reflects a combination of both models (Luo et al. 2006; Kaneko et al. 2007b; West et al. 2008).

The function of the exonuclease XRN2/Rat1 in the termination process has been studied quite extensively (Kaneko et al. 2007b; Kim et al. 2004; Luo et al. 2006; Xiang et al. 2009). ChIP analysis shows that Rat1 is recruited to the 3' end of genes, possibly through interactions with 3' end processing factors (Kim et al. 2004). Indeed, Rat1 recruitment to the 3' end of genes is decreased by inactivation of the 3' end processing factor Pcf11 (Luo et al. 2006). In addition to this, Rat1 is required for recruitment of Pcf11 and Rna15 (the yeast homologue of CstF64) to the 3' end of genes (Luo et al. 2006). Pcf11 is a 3' processing factor that contains a conserved CTD-interacting domain (CID) and it is able to bind nascent RNA (Licatalosi et al. 2002; Sadowski et al. 2003; Meinhart and Cramer, 2004; Zhang et al. 2005; Zhang and Gilmour, 2006). In

mammalian cells, depletion of Pcf11 reduces termination and stabilises the 3' pre-mRNA cleavage product (West and Proudfoot, 2008).

Recently it has been demonstrated that Sen1, the yeast homologue of human senataxin, cooperates with Rat1 to promote efficient transcriptional termination in *S. cerevisiae* (Kawauchi et al., 2008).

An important conclusion from chapter 3 is that R-loops are required for pause-dependent transcriptional termination. In this chapter, I will present experiments performed to further investigate the molecular mechanism of R-loop dependent transcriptional termination by studying the endogenous β-actin gene and particularly, the role of senataxin and Pcf11 in this process. Furthermore, the last part of this chapter is focused on the potential role of senataxin in transcriptional termination of the endogenous γ-actin gene.

4.2 Senataxin is involved in transcriptional initiation and termination of endogenous human β-actin gene

To validate the results obtained from the plasmids as described in Chapter 3, I also investigated the role of senataxin in termination of the endogenous β-actin gene in HeLa cells. In contrast to transiently transfected plasmids, transcriptional profiles of endogenous genes can be reproducibly analysed by ChIP experiments, employing a number of PCR primers located over the coding and 3'flanking regions. ChIP on transfected plasmids is difficult to perform possibly because the large amount of untranscribed plasmid trapped at the nuclear membrane gives a high background ChIP signal. However DIP experiments performed on transfected plasmids in Chapter 3 did not

exhibit the same problem. A possible explanation for this discrepancy could be that DIP experiments rely on the immuno-precipitation of purified genomic DNA as opposed to the immuno-precipitation of chromatin employed in ChIP experiments.

Previously a 3'flanking pause element involved in transcriptional termination of the human β-actin gene has been identified (Gromak et al. 2006). As demonstrated in Figure 4.1A (left panel), I observe Pol II enrichment based on ChIP analysis over the promoter of β-actin gene and around its pause element downstream of the poly(A) signal. Accumulated Pol II levels over β-actin gene are likely to be associated with pausing at the promoter and terminator regions and are resonant with gene loop conformations as previously described (Tan-Wong et al. 2008). Pol II detected over the coding region was lower than over the promoter, suggesting that during transcription only a subfraction of Pol II molecules undergo efficient elongation as generally found for protein-coding genes (Core et al. 2008; Nechaev et al. 2010).

To determine if senataxin binds the β-actin gene, I performed ChIP analysis using anti-senataxin antibody, which specifically recognizes the 303 kDa senataxin protein in HeLa cells (Figure 3.1A). I demonstrate that senataxin binds at the highest level upstream of the β-actin pause element (Figure 4.1A, right panel). The association of senataxin with β-actin gene suggests an involvement in transcriptional regulation. Therefore I investigated the transcriptional profile of the β-actin gene in senataxin-deficient cells. ChIP analysis revealed that upon senataxin depletion, reduced level of Pol II occurred over the promoter and intron 1 regions of the β-actin gene, implying an effect on transcriptional initiation which leads to a defect in elongation (Figure 4.1B).

Additionally, an increase in Pol II density over the 5'pause and pause regions was observed, indicative of a defect in transcriptional termination.

ChIP analysis provides information about the total Pol II density across a gene. However in contrast to Nuclear Run-On (NRO) analysis, it does not discriminate between actively transcribing and stalled Pol II. To further substantiate a role for senataxin in transcriptional initiation/elongation and termination, I employed a modified NRO procedure using Br-UTP as the labelled nucleotide, called bromo-uridine 5'-triphosphate nuclear run-on (Br-UTP NRO) (Core et al. 2008; Lin et al. 2008). Br-UTP NRO involves isolation of HeLa cell nuclei in such a way that transcribing polymerases remain actively bound to their DNA templates. These can resume transcription after addition of transcription buffer +/-Br-UTP (or UTP in control samples) *in vitro*. The nascent RNA containing incorporated Br-UTP is selected using anti-Br-UTP antibody. As demonstrated in Figure 4.1C (top panel), in mock-treated cells I observed low RNA signals over the promoter and intron 1 regions reflecting a lack of transcription over the promoter and high level of paused Pol II accumulated over intron 1. Both of these primers are positioned over nucleosome-free region, where Pol II pauses prior to efficient transcription elongation (Egloff et al. 2009). I observed high enrichment of nascent transcripts over the intron 3 probe, positioned immediately downstream of the nucleasome-free region, indicative of efficient transcriptional elongation. Based on Br-UTP NRO, I observed a 2-5 fold decrease in nascent RNA levels produced over the gene body in senataxin-depleted cells as compared to mock-treated cells (Figure 4.1C, top panel). In contrast, I detected significant enrichment of the nascent read-through RNA signals over the 5'pause, pause, C and D regions, relative to the gene body (intron 5

probe), in either mock-treated or senataxin knocked-down cells, indicating a potential role for senataxin in transcriptional termination (Figure 4.1C, bottom panel).

Overall these results imply that loss of senataxin distorts the β-actin gene transcription profile, resulting in reduced Pol II levels and nascent transcription in the gene promoter and body. In contrast Pol II accumulates over the pause region where it maintains some transcriptional activity. Experiments in *S. cerevisiae* have shown that Sen1 is required to promote Pol I and Pol II transcriptional termination (Kawauchi et al. 2008). This may point towards the generality of this senataxin-dependent mechanism of transcriptional termination.

4.3 R-loops are formed over endogenous β-actin gene

Next I investigated if R-loops can be formed over the transcribed regions of endogenous β-actin gene similarly to transfected gene constructs (Figure 3.3A). As demonstrated in Figure 4.1D (top panel), R-loops were detected over the transcribed regions of endogenous β-actin gene with their specific enrichment over the intron 1 and immediately downstream of the poly(A) signal. Interestingly, these regions correspond to those enriched for senataxin binding (Figure 4.1A, right panel). To test the specificity of the signal detected by S9.6 antibody in DIPs, I performed nuclease treatment experiments prior to immuno-precipitation step. In particular, I treated the samples with RNase H, recognising DNA/RNA hybrids, and S1 nuclease, recognising ssRNA and ssDNA molecules and RNA/DNA hybrids at high enzyme concentrations. As demonstrated in Figure 4.1D (bottom panel), treatment of the samples with RNase H completely abolished

the R-loop signals, suggesting that they are specific for RNA/DNA hybrids. As expected R-loops were only partially digested by S1 nuclease.

4.4 Senataxin is required for Xrn2 recruitment to promote efficient transcriptional termination

Previously the role of the human 5'-3' exonuclease Xrn2 in the process of transcriptional termination has been demonstrated (West et al. 2004). To confirm that Xrn2 protein binds the termination region of the endogenous β-actin gene, I performed ChIP analysis using anti-Xrn2 antibody. A specific enrichment of Xrn2 signal downstream of the poly(A) signal and upstream of the pause element of the endogenous β-actin gene was evident (Figure 4.1E, grey bars).

Next I tested if Xrn2 recruitment to the termination region of the endogenous β-actin gene is affected in senataxin-deficient cells. Therefore I performed Xrn2 ChIP in senataxin-depleted HeLa cells. As demonstrated in Figure 4.1E, Xrn2 recruitment was substantially reduced in senataxin knock-down cells. These results suggest that senataxin is required for the resolution of R-loops formed downstream of the poly(A) signal, allowing Xrn2 recruitment and efficient Pol II transcriptional termination.

4.5 Senataxin interacts with Xrn2 *in vivo*

The question that arose from the observation that senataxin depletion affects Xrn2 recruitment was whether these two proteins physically interact *in vivo*. To address

this I initially over-expressed flag-tagged Xrn2 (Kaneko et al. 2007b) in HEK293 cells. HEK293 cells were used for this part of the study as preliminary experiments showed that HeLa cells failed to over-express Xrn2. The level of flag-tagged Xrn2 was measured by Western blot as compared to endogenous actin (Figure 4.2A, left panel). The protein levels of senataxin in HEK293 cells were also measured by Western blot (Figure 4.2A, right panel).

Next I performed co-immunoprecipitation experiments on HEK293 cells transiently over-expressing Xnr2-flag. The complex was immuno-precipitated with anti-senataxin antibody, separated by SDS-PAGE and probed with anti-flag antibody. No anti-senataxin antibody was added to the negative control samples. Xrn2-flag was co-immunoprecipitated only when the anti-senataxin antibody was added (Figure 4.2B), suggesting that senataxin and Xrn2 interact *in vivo*.

4.6 Senataxin depletion affects Pcf11 recruitment and Ser2P levels

Previously, it has been shown that Rat1 (the yeast homologue of Xrn2) is involved in the recruitment of the 3' processing factor Pcf11 (Luo et al. 2006). The results presented so far suggest that human senataxin interacts with Xrn2 and affects its recruitment to the 3' end of β-actin gene. To determine if Pcf11 binds β-actin gene, I performed ChIP analysis using anti-Pcf11 antibody. I demonstrate that Pcf11 binds around the pause element (5' pause probe and pause probe) at the highest level (Figure 4.3A, left panel).

To investigate whether Pcf11 is also involved in senataxin-mediated R-loop resolution, I tested if Pcf11 recruitment to the termination region of β-actin gene is affected in senataxin-deficient cells. To address this, I performed Pcf11 ChIP in senataxin-depleted HeLa cells. As demonstrated (Figure 4.3A, right panel), senataxin depletion substantially blocks Pcf11 recruitment.

Gilmour and colleagues have shown that Pcf11 can terminate Pol II *in vitro*; this reaction involves the interaction of Pcf11 with both the CTD and nascent transcript. Furthermore Pcf11 binds phosphorylated Ser2 of the CTD *in vitro* (Barilla et al. 2001; Licatalosi et al. 2002; Ahn et al. 2004; Lunde et al. 2010). To test whether Ser2P levels are also affected upon senataxin depletion, I performed ChIP analysis using anti-Ser2P antibody. I observed high enrichment of Ser2P levels over intron 3 probe (Figure 4.3B, grey bars), where efficient elongation starts at β-actin gene as demonstrated in Figure 4.1C (Br-UTP NRO, grey bars). Ser2P levels are also evident at the termination region of β-actin (Figure 4.3B, grey bars). These data are consistent with the existing literature (Egloff et al. 2009), confirming that Ser2P levels increase into the body of the gene and at the gene's 3' end.

Interestingly, upon senataxin depletion Ser2P levels are reduced (Figure 4.3B). While it is difficult to distinguish cause from effect it is apparent that loss of senataxin influences both Ser2P CTD levels and 3' end processing factor recruitment. Upon senataxin depletion, reduced Ser2P levels at the 3' end of β-actin gene could lead to reduced recruitment of Pcf11 at the termination region.

4.7 Pcf11 depletion affects senataxin recruitment and R-loop distribution profile of endogenous β-actin gene

To investigate whether Pcf11 has a specific role in senataxin-mediated R-loop resolution during pause-dependent termination, I initially depleted Pcf11 from HeLa cells using the RNAi approach. The cells were transfected with a siRNA duplex specifically targeted to PCF11 mRNA. PCF11 mRNA and protein levels were reduced using PCF11 siRNA, as measured by qRT-PCR and Western blot analyses (Figure 4.4A). Western blot analysis was carried out using an anti-Pcf11 antibody (Zhang and Gilmour, 2006) which selectively detects the 173 kDa Pcf11 protein in mock-treated cells (Figure 4.4A, right panel).

Next, I tested if senataxin recruitment to the termination region of the β-actin gene is affected in Pcf11-deficient cells. To this end, I performed senataxin ChIP in Pcf11-depleted HeLa cells. As shown in Figure 4.4B, Senataxin recruitment was impaired in Pcf11 knock-down cells. Taking into consideration that Pcf11 recruitment is also reduced in senataxin-deficient cells (Figure 4.3A, right panel), it could be suggested that senataxin and Pcf11 interact, directly or indirectly, to promote efficient transcriptional termination of the endogenous β-actin gene.

Finally, I sought to investigate if the R-loop distribution profile is affected in Pcf11-depleted cells. DIP analysis revealed that, upon Pcf11 depletion, increased levels of R-loop signals occurred around the pause element, implying that Pcf11 depletion leads to stabilisation of R-loops at the termination region of endogenous β-actin (Figure 4.4C).

4.8 Senataxin binds endogenous γ-actin gene but it is not involved in its transcriptional termination

To further investigate the role of senataxin protein and R-loops in transcriptional termination of endogenous human genes, I decided to study a second highly expressed human gene, γ-actin. As opposed to human β-actin gene, the mechanism of transcriptional termination in γ-actin is uncharacterised.

To determine the transcriptional profile of γ-actin I performed Pol II ChIP analysis, employing a number of qRT-PCR primers covering the coding region and the 3' end of the gene. As demonstrated in the left panel of Figure 4.5A, pausing of Pol II over the γ-actin promoter is especially enhanced. However with the low levels of Pol II that read into the gene body, I observe that termination occurs over the probe D, positioned ~3 kb downstream of the poly(A) signal. A slightly higher Pol II density immediately downstream of the poly(A) site (probe A, Figure 4.5A-right panel). However to investigate whether transcriptional termination of γ-actin is mediated through a pause-dependent termination mechanism requires further analysis.

Next I performed ChIP analysis on γ-actin gene to test for senataxin binding. Interestingly, I observe enrichment of senataxin over the promoter of the γ-actin gene with reduced levels downstream of the poly(A) signal (Figure 4.5B). This result contrasts with the senataxin binding profile for the β-actin gene, where the main peak is over the termination region (Figure 4.1A, right panel). Nevertheless the association of senataxin with γ-actin gene promoter region suggests an involvement in transcriptional regulation. Therefore I investigated the transcriptional profile of the γ-actin gene in senataxin-deficient cells. ChIP analysis reveals that upon senataxin depletion, less Pol II enrichment

occured over the γ-actin promoter, implying an effect on transcriptional initiation (Fig.4.5C, left panel). However I did not observe an increase in Pol II density over the 3'flanking probes (A, B, C and D), indicating that senataxin is not involved in transcriptional termination of human γ-actin gene.

To further validate that senataxin is not required for transcriptional termination of γ-action, I assessed the termination efficiency of γ-actin gene in senataxin-deficient cells, using qRT-PCR analysis. In this experiment the amount of read-through transcript was determined using RT-PCR with primers A, B, C, or D whereas the amount of correctly initiated transcript was measured with RT-PCR ex5. The ratio between the amount of read-through RNA from each RT-PCR product (A-D) versus the amount of gene transcript from ex5 represents the efficiency of transcriptional termination. As demonstrated in the right panel of Figure 4.5C, senataxin depletion does not cause an increase in the amount of the read-through transcripts, suggesting that senataxin is not involved in transcriptional termination of endogenous human γ-actin gene.

4.9 R-loops form over endogenous human γ-actin gene

A possible hypothesis to explain the reason why senataxin is not involved in transcriptional termination of γ-actin and why it does not bind the termination region at the highest level, could be that R-loops do not form over γ-actin gene and therefore senataxin is not required to resolve them at the termination region to promote efficient termination.

To test this hypothesis I performed DIP analysis on γ-actin gene. R-loops were detected over the transcribed regions of endogenous γ-actin gene with their specific

enrichment over the intron 1 and immediately downstream of the poly(A) signal (Figure 4.5D). Interestingly, these regions correspond to the two main peaks of R-loops in endogenous β-actin gene (Figure 4.1D, probes intron 1 and 5' pause). These results point towards the generality of R-loops enrichment over endogenous human genes.

4.10 Discussion

Previous studies have extensively characterized the mechanism of transcriptional termination in mammals (Kaneko et al. 2007b; Luo et al. 2006; West et al. 2004; West et al. 2008). It was demonstrated that the 5'-3' exonuclease Xrn2 recognizes the nascent RNA cleaved at the poly(A) site, promotes its degradation, and so causes Pol II transcriptional termination. This process requires termination elements located downstream of the poly(A) signal such as pause sites or CoTC elements.

This chapter combines with Chapter 3 to describe a molecular function for human senataxin protein in the regulation of gene expression at the level of transcription. Using ChIP, NRO and qRT-PCR analysis, I demonstrate that senataxin-deficient HeLa cells have transcription initiation/elongation and termination defects using either plasmid gene constructs or the endogenous β-actin gene (Figures 3.1B, 3.2, 3.4C, 4.1B and 4.1C). In particular, I observe a decrease in nascent RNA and Pol II density over the gene body and an increase in the amount of read-through RNA and Pol II density downstream of the poly(A) signal in the termination regions. Recent studies have demonstrated an interesting link between 3'end processing and transcriptional initiation (Mapendano et al. 2010). In agreement with this study, my results reveal that senataxin deficiency causes a defect in termination and also influences transcriptional initiation. This is particularly

evident on the endogenous β-gene but less so with the transfected gene constructs, as their expression is driven by strong heterologous HIV promoter. This further supports the idea of factors recycling due to established promoter-terminator 'gene-loop' interactions *in vivo*. In the future, it will be interesting to test whether senataxin is involved in establishing or maintaining this gene conformation.

R-loop structures can be efficiently detected on transfected gene constructs and endogenous genes (Figures 3.3A, 4.1D and 4.5C). Based on results described at Chapter 3, R-loops are an essential component of the pause-dependent transcriptional termination (Figure 3.4B and 3.4C) and are dependent on functional poly(A) signal and G-rich pause elements (Figure 3.3C). The data of this chapter have uncovered that R-loops are highly enriched at the beginning and at the termination regions of endogenous human genes (Figures 4.1D and 4.5D). R-loops are strongly associated genome-wide with G-rich promoters (Ginno et al. 2012) and have been suggested to serve to anchor Pol II and allow for a capping-dependent transcription elongation checkpoint control (Kaneko et al. 2007a). It could be predicted that R-loops downstream the poly(A) signal of a gene may equally serve to pause the polymerase at this position. Indeed, based on results presented so far, I suggest that the formation of R-loops in transcriptional pause regions is essential to cause Pol II pausing downstream of the poly(A) site prior to termination. This pausing could facilitate the recruitment of Xrn2 at the termination region. However these R-loops have to be subsequently resolved by senataxin to release the nascent RNA and so allow its Xrn2-mediated degradation, which ultimately results in efficient Pol II transcriptional termination. Senataxin depletion stabilizes R-loop structures downstream the poly(A) signal (Figure 3.3D) and severely affects Xrn2 recruitment at the termination

region (Figure 4.1E). Overall it could be predicted that a fine balance of R-loop structures is needed for efficient pause-dependent termination in human genes.

Senataxin was previously found to interact with Pol II and various RNA binding factors, such as poly(A)-binding proteins 1 and 2 (PABP1/2), hnRNPs , SMN and SAP155 (Suraweera et al. 2009). As determined in Figure 4.2B, senataxin interacts *in vivo* with Xrn2. Therefore, senataxin may be involved in recognition of poly(A) signal and/or recruitment of other termination/RNA-processing factors. Transcriptional termination in humans requires 3'-end processing factor Pcf11, previously found to terminate stalled Pol II complexes *in vitro* and stimulate transcript degradation by Xrn2 *in vivo* (West and Proudfoot, 2008; Zhang et al. 2005; Zhang and Gilmour, 2006). Pcf11 is also implicated in the cross-talk between 3'end processing and transcriptional initiation (Mapendano et al. 2010). Results of this chapter suggest that Pcf11 is involved in senataxin-mediated R-loop resolution during transcriptional termination. In essence, I observed that depletion of either Pcf11 or senataxin affects the other's recruitment at the termination region of β-actin (Figure 4.3A and Figure 4.4B). A possible explanation for the reduced Pcf11 recruitment upon senataxin depletion could be that the increased levels of R-loops (as a result of senataxin depletion) could mask the poly(A) signal. Another interesting observation is that the levels of R-loops formed at the termination region of β-actin increase upon Pcf11 depletion (Figure 4.4C). This effect could be due to the increased levels of transcription observed upon Pcf11 knock-down (West and Proudfoot, 2008) or due to reduced recruitment of senataxin at the termination region followed by subsequent increase in R-loops levels. It would be interesting to investigate in the future if other 3' processing factors, such as CstF64, show similar effects on R-loop distribution

profile or whether Pcf11 has a specific role in R-loop dependent transcriptional termination.

Upon senataxin depletion, Serine 2 phosphorylation levels of the Pol II CTD are also affected (Figure 4.3B). Reduced Serine 2 phosphorylation levels at the 3' end of β-actin gene could explain the reduced recruitment of Pcf11 at the termination region, since possibly the interaction of Pcf11 CID domain with the Pol II CTD is affected. Serine 2 is phosphorylated by the Cdk9 kinase subunit of the positive elongation factor pTEFb. Therefore it could be speculated that Cdk9 levels are also reduced upon senataxin knock-down, however further investigation would be required.

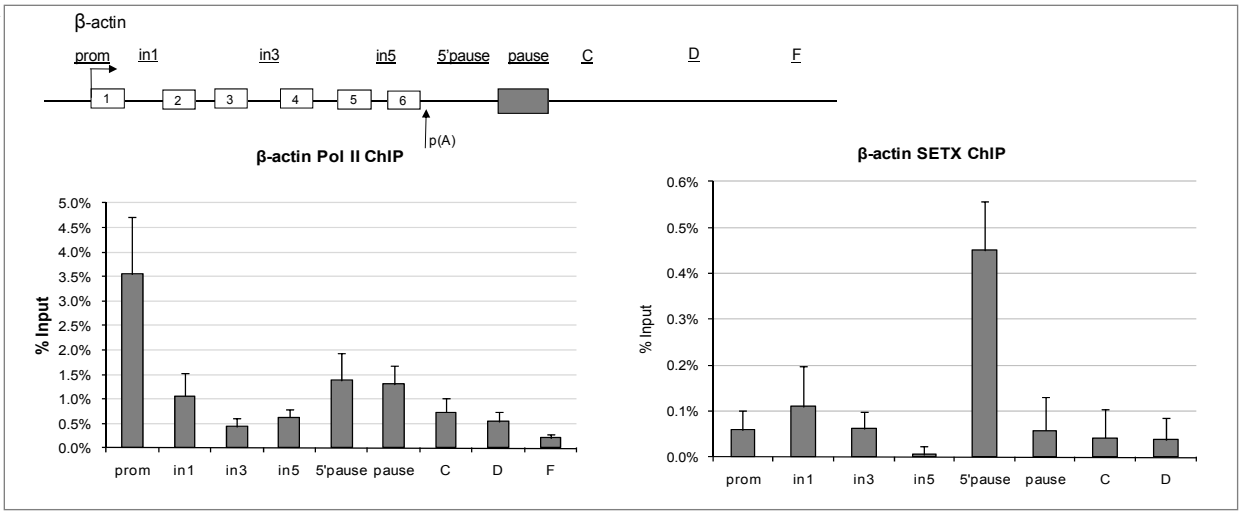
In my studies on the endogenous γ-actin gene, I observed the highest peak of senataxin binding over the intron 1 (Figure 4.5B). Furthermore, senataxin does not affect transcriptional termination of γ-actin gene even though R-loops form at the termination region (Figure 4.5D). It is not clear what the molecular function of senataxin is at the beginning of γ-actin gene. Recently it has been demonstrated that *S.cerevisiae* Rat1p is recruited towards gene 5'ends in order to terminate Pol II molecules engaged in the production of uncapped transcripts (Jimeno-Gonzalez et al. 2010). It is possible that mammalian Xrn2 protein is involved in a similar process. Possibly senataxin found at the beginning of human genes co-operates with Xrn2 in this process. Alternatively senataxin, like Sen1, could be involved in termination and degradation of non-coding RNAs (called CUTs in *S.cerevisiae*). It has been shown that Nrd1-Sen1 complex interacts with the nuclear exosome to promote degradation of CUTs (Vasiljeva and Buratowski, 2006). Kapranov and colleagues have shown that transcription in mammalian genomes is surprisingly widespread and that many transcripts do not correspond to coding genes or

stable noncoding transcripts (Kapranov et al. 2007). Therefore suppression of cryptic transcription by co-transcriptional targeting of termination and degradation machineries might be even more important in higher eukaryotes. To this end, the potential interaction between senataxin and Dis3 (the nuclear endonuclease component of the mammalian exosome) could be investigated.

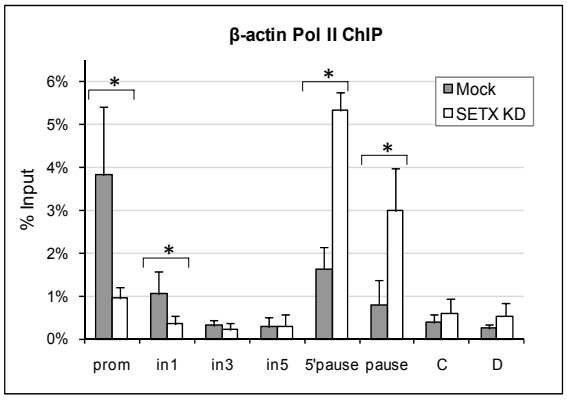
Overall an important conclusion derives from the experiments presented in Chapter 3 and 4. R-loops formed over the G-rich pause regions are necessary for Pol II to pause downstream the poly(A) site. Senataxin is needed to resolve R-loop structures and so allow 5'-3' exonuclease Xrn2 to degrade the nascent transcript from the site of poly(A) cleavage and catch up with paused Pol II, causing its efficient transcriptional termination. Possibly Pcf11 is also involved in this process, however further investigation is needed towards this end.

Pausing of Pol II prior to termination appears to be significant for efficient termination of at least a subset of genes. My results suggest so far that R-loop formation at the termination region facilitates Pol II pausing. Experiments presented in Chapter 5 are aimed at further understanding the role of R-loops in pause-dependent transcriptional termination of mammalian genes.

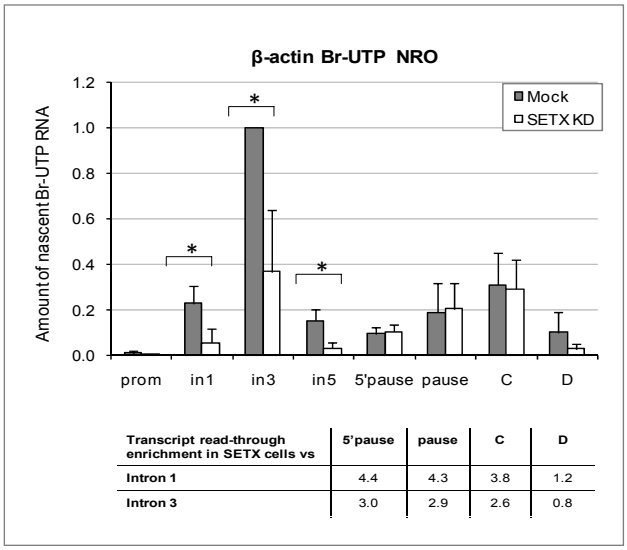
A



B

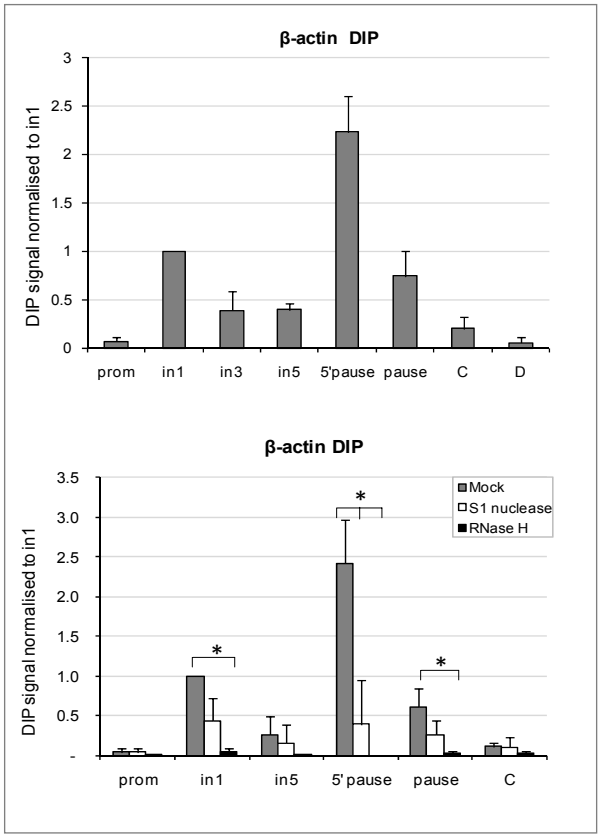


C



Transcript read-through enrichment in SETX cells vs	5'pause	pause	C	D
Intron 1	4.4	4.3	3.8	1.2
Intron 3	3.0	2.9	2.6	0.8

D



E

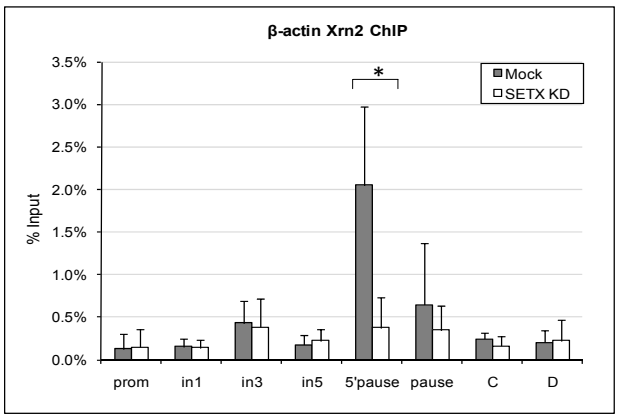


Figure 4.1 Senataxin is required for Pol II transcription and Xrn2 recruitment to the termination regions of the endogenous β-actin gene.

A. Diagram of β-actin gene, with exons shown as white boxes and pause element shown as grey box. Positions of primers used in ChIP analysis are shown above the diagram. ChIP analysis was carried out with Pol II (left panel) or senataxin (right panel) antibody.

B. Pol II ChIP analysis of β-actin gene in mock-treated (grey bars) and senataxin siRNA-treated (white bars) cells.

C. Top panel: Br-UTP NRO analysis of β-actin gene in mock-treated (grey bars) and senataxin siRNA-treated (white bars) cells. Bottom panel: Amount of nascent Br-UTP RNA was normalised to intron 3 probe in mock-treated cells. All bars represent average values +/- SD from at least three independent biological experiments. Enrichment of read-through transcripts for probes '5' pause', 'pause', 'C' and 'D' in senataxin-depleted cells, as compared to mock cells, was calculated relative to intron 1 and intron 3 signals.

D. Top panel: DIP analysis of β-actin gene using RNA/DNA hybrids-specific S9.6 antibody, based on average values +/- SD from at least three independent biological experiments. DIP signal was normalised to intron 1 signal. Bottom panel: R-loops formed over β-actin gene are sensitive to RNase H, and only partially to S1 nuclease digestion. DIP samples were either untreated (grey) or treated with 10U of RNase H (black bars) or 200U of S1 nuclease (white bars).

E. Xrn2 ChIP of β-actin in mock-treated (grey bars) and senataxin depleted (white bars) HeLa cells. Recruitment of Xrn2 to the region downstream of the poly(A) site is blocked in senataxin-deficient HeLa cells.

ChIP values are based on average values +/- SD from at least three independent biological experiments. * indicates statistical significance ($p<0.05$), based on unpaired, two-tailed distribution Student's t -test.

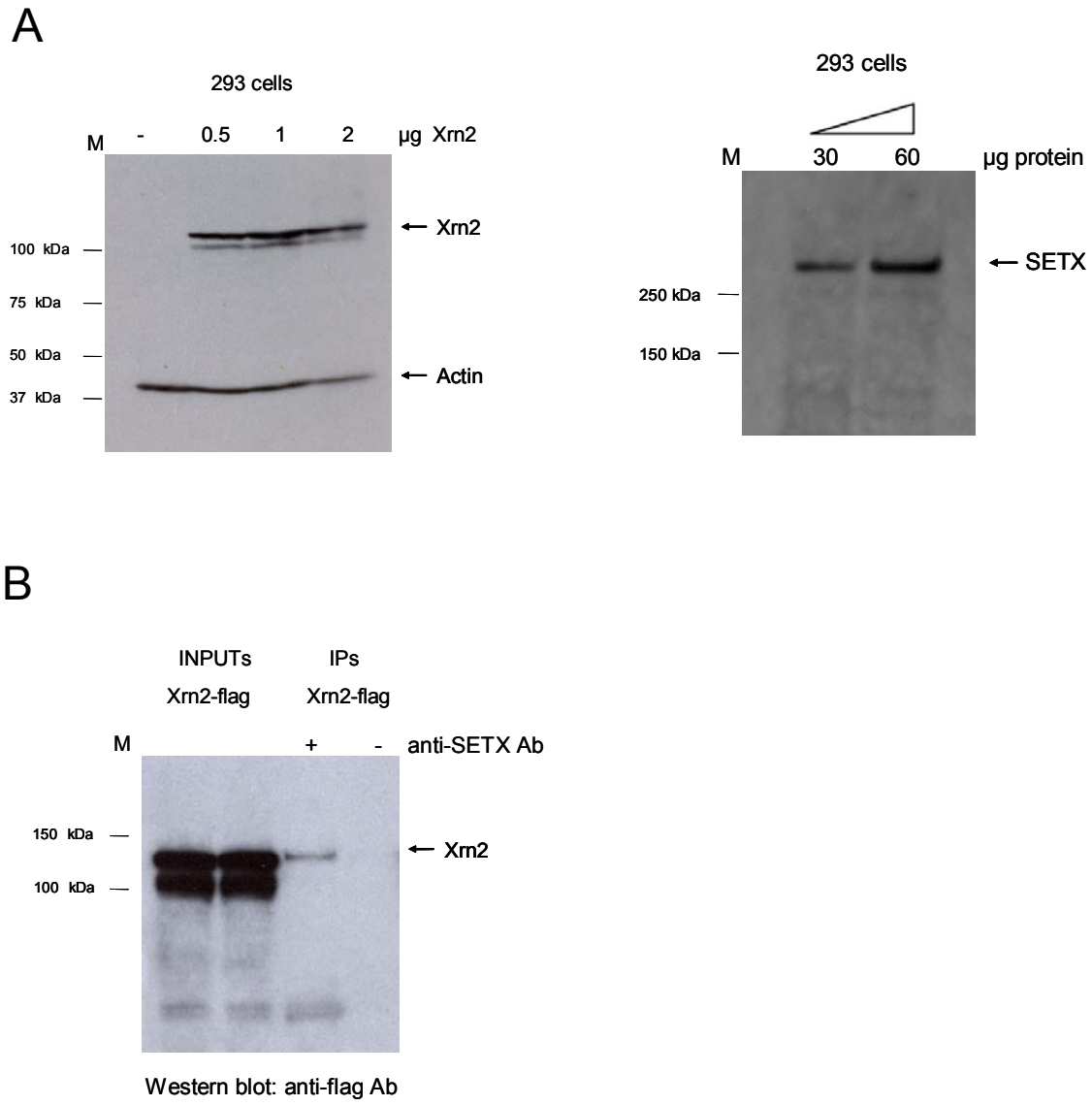


Figure 4.2 Senataxin interacts with Xrn2 *in vivo*

A. Left panel: Western blot analysis of 30 µg of protein extracts obtained from HEK293 cells co-transfected with increasing amounts (0-2 µg) of Xrn2-flag expression plasmid. Western blot was analysed using anti-flag antibody. Actin was used as loading control. Right panel: Western blot analysis of 30 µg and 60µg of protein extracts obtained from HEK293 cells. Western blot was analysed using anti-senataxin antibody.

B. Co-immunoprecipitation analysis of senataxin with Xrn2 in HEK293 cells transiently over-expressing Xrn2 protein. Immunoprecipitated complexes with anti-senataxin antibody were seperated by SDS-PAGE and probed with anti-flag antibody. The experiment has been repeated three times and the pattern obtained is reproducible.

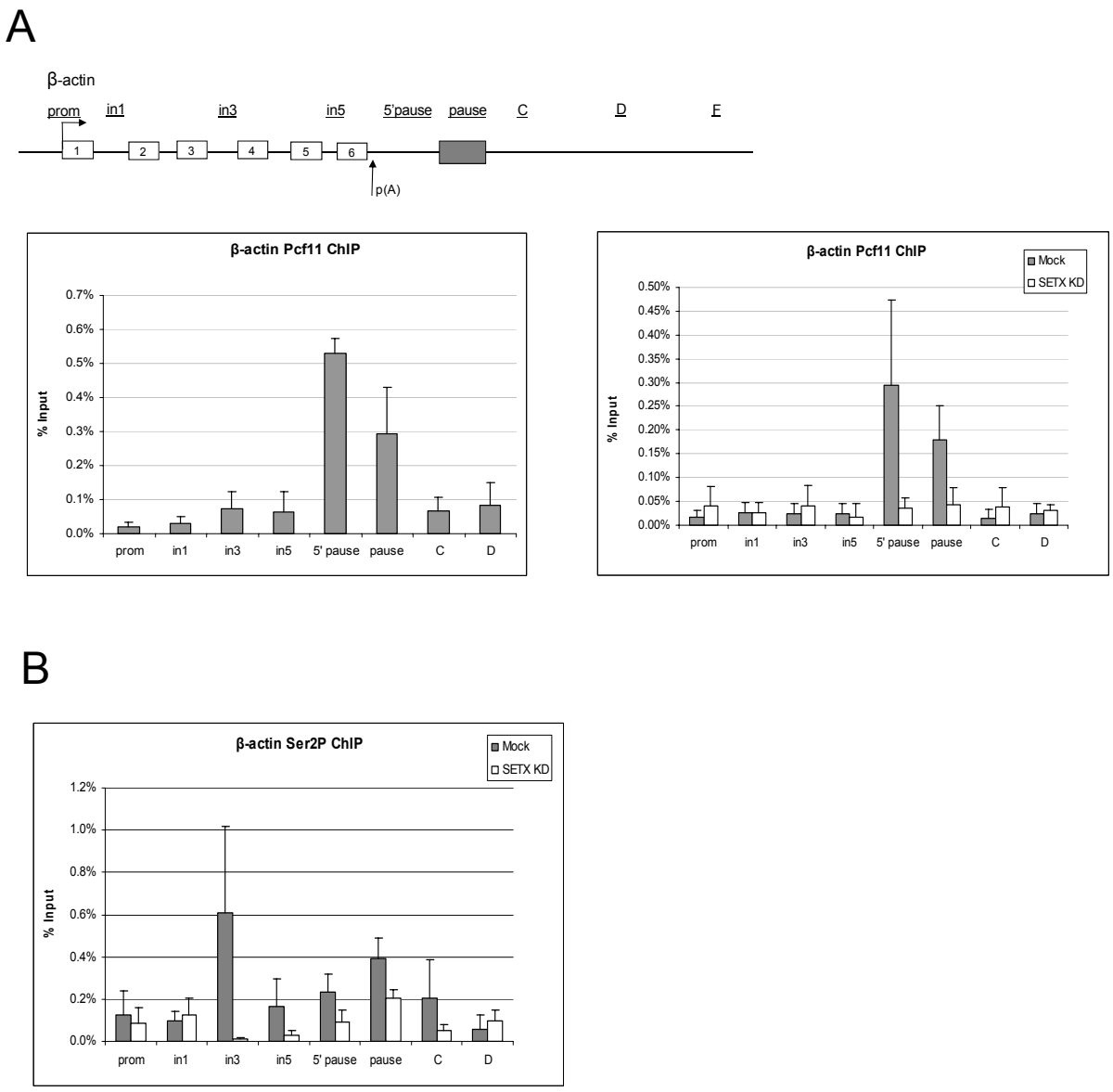


Figure 4.3 Senataxin depletion affects Pcf11 recruitment and Ser2 phosphorylation levels at the termination region of endogenous β-actin gene.

A. Left panel: Pcf11 ChIP analysis performed on endogenous β-actin gene. Pcf11 binds β-actin gene around the region of the pause element. Right panel: Pcf11 recruitment in mock-treated (grey bars) and senataxin-depleted (white bars) HeLa cells.

B. Ser2 phosphorylation ChIP analysis of β-actin gene in mock-treated (grey bars) and senataxin siRNA-treated (white bars) cells. In mock-treated cells, Ser2 phosphorylation of the CTD takes place over in3, where productive elongation starts, and further downstream the β-actin gene.

ChIP profiles are based on average values +/- SD from at least three independent biological experiments.

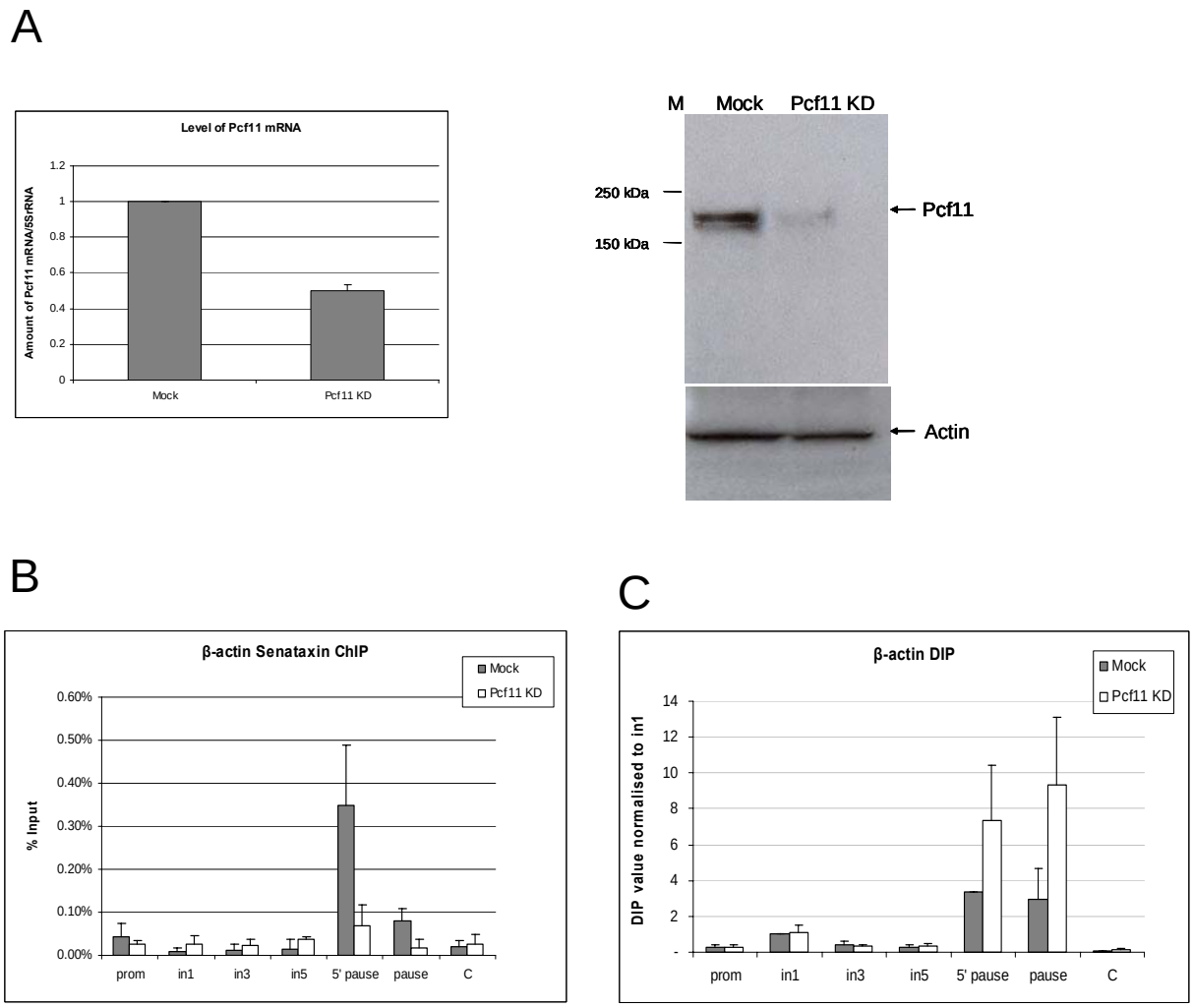


Figure 4.4 Pcf11 depletion affects senataxin recruitment and R-loop distribution profile of endogenous β-actin gene.

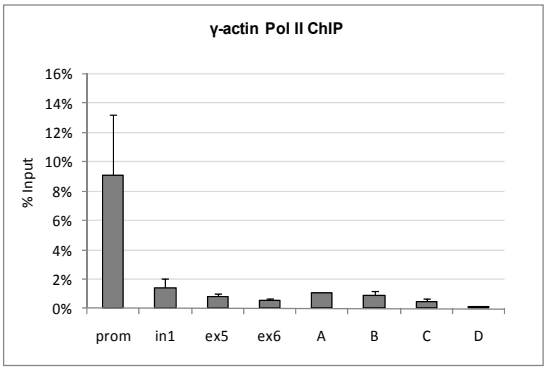
A. RNAi-mediated depletion of human Pcf11 in HeLa cells. Left panel: qRT-PCR analysis of mock- and Pcf11 siRNA-treated HeLa cells. Pcf11 was depleted to a 50% level as compared to the level of 5S rRNA used as a control. Right panel: Western blot analysis of 60 μg of protein extracts from mock- and Pcf11 siRNA-treated HeLa cells. Western blot was analysed using anti-Pcf11 antibody. Actin protein was used as a loading control.

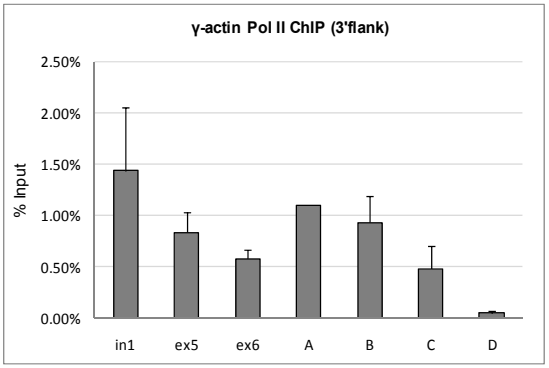
B. Senataxin ChIP analysis of β-actin gene in mock-treated (grey bars) and Pcf11 siRNA-treated (white bars) cells, based on average values +/- SD from at least three independent biological experiments. Senataxin recruitment upstream of the pause element of endogenous β-actin gene is reduced in Pcf11 knockdown cells.

C. R-loop profile analysis of endogenous β-actin gene in mock-treated (grey bars) and Pcf11-depleted (white bars) cells. DIP signal was normalised to in1 signal of β-actin gene in mock-treated cells. DIP profile is based on average values +/- SD from three independent biological experiments.

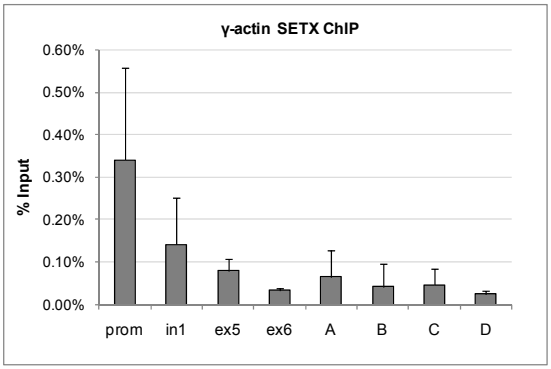
A


γ-actin
prom
in1
ex5
ex6
A
B
C
D
1
2
3
4
5
6
p(A)

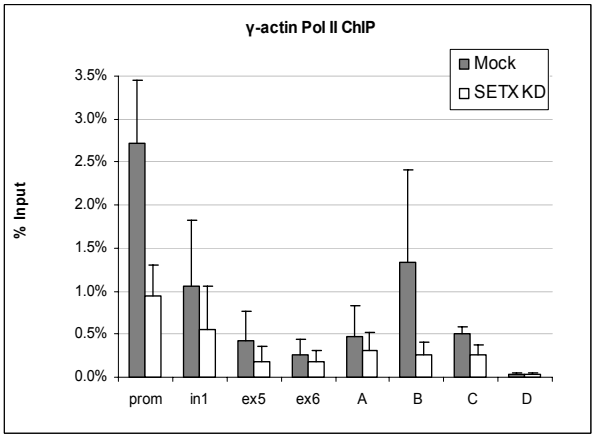

γ-actin Pol II ChIP
% Input
16%
14%
12%
10%
8%
6%
4%
2%
0%
prom
in1
ex5
ex6
A
B
C
D

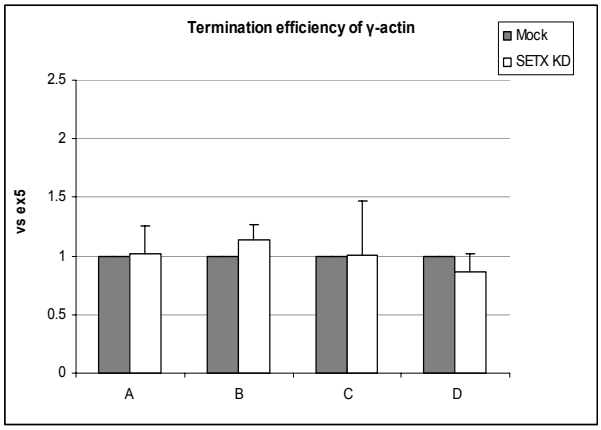

γ-actin Pol II ChIP (3'flank)
% Input
2.50%
2.00%
1.50%
1.00%
0.50%
0.00%
in1
ex5
ex6
A
B
C
D

B


γ-actin SETX ChIP
% Input
0.60%
0.50%
0.40%
0.30%
0.20%
0.10%
0.00%
prom
in1
ex5
ex6
A
B
C
D

C


γ-actin Pol II ChIP
Mock
SETX KD
% Input
3.5%
3.0%
2.5%
2.0%
1.5%
1.0%
0.5%
0.0%
prom
in1
ex5
ex6
A
B
C
D


Termination efficiency of γ-actin
Mock
SETX KD
vs ex5
2.5
2
1.5
1
0.5
0
A
B
C
D

D

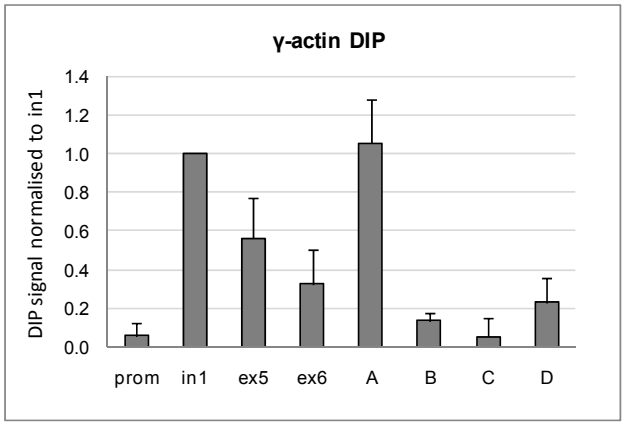


Figure 4.5 Senataxin binds endogenous γ-actin gene but it is not required for Pol II transcriptional termination.

A-B. Diagram of γ-actin gene, with exons shown as white boxes and pause element shown as grey box. Positions of primers used in ChIP analysis are shown above the diagram. ChIP analysis was carried out with Pol II (left panel) or senataxin (right panel) antibody. Right panel of Figure 4.5A shows probes downstream of γ-actin in1 probe on more extended Y axis. ChIP profiles are based on average values +/- SD from at least three independent biological experiments.

C. Left panel: Pol II ChIP analysis of γ-actin gene in mock-treated (grey bars) and senataxin siRNA-treated (white bars) cells. Right panel: Termination efficiency of γ-actin gene from mock-treated and senataxin siRNA-treated cells using qRT-PCR analysis on total RNA. Termination efficiency was calculated as a ratio of RT-PCR product A, B, C, D (read-through RNA) versus RT-PCR product ex5 (gene transcript) from three independent experiments; values are +/- SD. The ratios in mock-treated cells were taken as 1.

D. DIP analysis of γ-actin gene using RNA/DNA hybrids-specific S9.6 antibody, based on average values +/- SD from at least three independent biological experiments. DIP signal was normalised to intron 1 signal.

CHAPTER 5

The role of H3K9me2 mark in R-loop dependent termination pausing of mammalian genes

5.1 Introduction

All steps in gene expression are thought to be interconnected and co-ordinately regulated (Moore and Proudfoot, 2009). Small interfering RNAs (siRNAs) trigger transcriptional gene silencing (TGS) in human cells through heterochromatin formation at DNA target sequences. This process is called RNA interference (RNAi)-dependent gene silencing and is one of the key players in regulation of gene expression. RNAi-dependent gene silencing involves recruitment of chromatin-modifying enzymes, resulting in dimethylation and or trimethylation of histone H3 Lys9 (H3K9me2, H3K9me3), trimethylation of histone H3 Lys27 (H3K27me3), DNA methylation and histone deacetylation (Grewal, 2010).

The mechanism of siRNA-directed TGS in fission yeast involves initially the processing of long double-strand RNAs into 21-25 nucleotides long siRNAs by the RNase-III enzyme, Dicer (Bernstein et al. 2001; Carthew and Sontheimer, 2009). This activates the Argonaute-containing RNA-induced transcriptional silencing (RITS) complex. siRNA-activated RITS complex associates then with DNA target sequences (Verdel et al. 2004). In mammalian cells there are several Argonaute proteins. Argonaute-2 (Ago2) has the most well-established role in RNAi, as it is the catalytic subunit of the RNA-induced silencing complex (RISC) that drives the mRNA cleavage (Hutvagner and Simard, 2008). Argonaute-1 (Ago1) is 80% identical to Ago2. Ago1 and Ago2 are both present in the nucleus and they have both been shown to be involved in transcriptional silencing (Kim et al. 2006; Janowski et al. 2006).

The process of transcriptional gene silencing has not been well established and the mechanistic details are not yet clear. For instance, it is not well understood whether RITS

complex targets gene transcripts or single-stranded DNA template. ncRNAs transcribed from the centromeric repeats or mRNAs transcribed from active genes are assumed to remain on chromatin and function as a platform for the RNAi machinery (Cam et al. 2009) but until recently there was no direct evidence to support this hypothesis.

Interestingly, Nakama and colleagues demonstrated that in fission yeast ncRNA transcribed from centromeric repeats is associated with chromatin via the formation of a RNA/DNA hybrid. In essence, RNA/DNA hybrid foci co-localise with centromeric heterochromatic regions and over-expression of RNase H decreases these foci leading to disruption of heterochromatin formation. Based on their results, it could be predicted that RITS complex could target to the chromatin-associated RNA or the single-stranded 'unprotected' DNA produced by the RNA/DNA hybrid to generate heterochromatin formation (Nakama et al. 2012).

In Chapter 3 and Chapter 4, I show that in a human gene with a G-rich pause element downstream the poly(A) signal, a certain degree of R-loops need to be formed for Pol II to pause prior to termination. However it still remains to be elucidated how R-loops facilitate termination pausing. The connection between R-loops and heterochromatin formation in fission yeast (Nakama et al. 2012) raises an interesting question; could R-loops that form around 3' positioned pause elements of human genes induce formation of repressive marks to facilitate pausing prior to efficient termination?

In my final results chapter I describe an investigation into the possibility that the molecular mechanism of pausing in transcriptional termination of mammalian genes is mediated by formation of the H3K9me2 mark over the termination regions. In the first part of this chapter, I present experiments to investigate whether repressive marks and

key players of the RNAi machinery are present at the termination region of human β-actin and γ-actin genes and whether these are associated with R-loop formation. In the second part, I describe experiments performed on the mouse β-actin gene using available mouse cell lines with constitutive knock-out of central RNAi factors to further investigate the potential role of the H3K9me2 mark in R-loop dependent termination pausing.

5.2 The H3K9me2 mark is present at the termination region of endogenous human β-actin and γ-actin genes

One of the main repressive epigenetic marks is H3K9me2. To investigate the occurrence of the H3K9me2 mark at the termination regions of endogenous human genes, I performed H3K9me2 ChIP analysis on human β-actin and γ-actin genes. As demonstrated in Figures 5.1A and 5.2A, H3K9me2 mark is present at the termination regions of both genes. As a positive control in this part of the study, I used centromeric repeats of human chromosome 9. Human centromeres contain large arrays of repetitive DNA (α-satellites) packaged into constitutive heterochromatin.

To rule out the possibility that the H3K9me2 levels observed are due to the higher H3 occupancy at the termination regions of β-actin and γ-actin genes, I performed ChIP analysis on β-actin and γ-actin genes using anti-H3 antibody. ChIP analysis revealed that in β-actin gene there is less H3 occupancy downstream the poly(A) signal and over the pause element as compared to the gene body (probes 5' pause and pause-Figure 5.1B). Furthermore, I observed lower H3 levels over the promoter and intron 1 regions. Both these primers are positioned over a known nucleosome-free region where Pol II pauses after efficient elongation (Egloff et al. 2009). In γ-actin gene there is a slight drop of H3

levels immediately downstream the poly(A) signal (probe A, Figure 5.2B), but overall it could be concluded that H3 occupancy is more widespread in γ-actin gene as compared to the β-actin gene.

Next I normalised the H3K9me2 signals to the total H3 levels over β-actin and γ-actin genes. The ratio of H3K9me2 versus the total H3 levels represents the H3K9me2 occupancy that is not dependent on the presence of the histone H3. As demonstrated in Figures 5.1C and 5.2C, H3K9me2 levels show specific enrichment immediately downstream the poly(A) signal of both β-actin (5' pause amplicon) and γ-actin (A amplicon) genes. Interestingly, both these regions are also enriched for R-loop signals (Figures 4.1D and 4.5C). This points towards the hypothesis that the H3K9me2 repressive mark is present at the R-loop associated termination regions of human genes.

5.3 Antisense transcription, Dicer and Ago1 are detected over the pause element of human β-actin gene

Detailed investigations have provided insights into the multi-step process of heterochromatin formation. An emerging theme is that heterochromatin and RNAi machinery act broadly across the genome at gene-active regions to regulate gene expression especially in the well studied yeast *S.pombe* (Gullerova and Proudfoot, 2008; Zofall et al. 2009; Zhang et al. 2011). Therefore there is a prominent view that processing of RNAs by the RNAi machinery is coupled to the loading of heterochromatin factors (Matzke and Birchler, 2005; Grewal and Jia, 2007; Buhler and Moazed, 2007; Reyes-Turcu and Grewal, 2012). However recent studies have revealed that transcription-dependent but RNAi-independent pathways also operate in the assembly of

heterochromatin (Reyes-Turcu et al. 2011; Zofall et al. 2011; Reyes-Turcu and Grewal, 2012).

In order to verify whether the presence of H3K9me2 mark over the termination region of human β-actin gene is RNAi-dependent, I tested for the presence of central RNAi factors around the pause element of human β-actin gene. For the following experiments I focus on the human β-actin gene where the 3'-flanking pause element is well characterised (Gromak et al. 2006).

Processing of double-stranded RNAs is the trigger for the RNAi-dependent gene silencing. A source for the generation of double-stranded RNAs could be the production of antisense transcripts that could then hybridise with the nascent RNA to form double-stranded RNAs. To address this point, I investigated whether antisense transcription is detectable over the termination region of human β-actin gene by RT-PCR analysis. cDNA was synthesised from total RNA isolated from HeLa cells, using forward primers located over the body of the gene (intron 5 amplicon), downstream the poly(A) signal (5' pause amplicon), over the pause element (pause amplicon) and downstream the pause element (C amplicon). These cDNAs were then amplified by qRT-PCR. As shown in Figure 5.3A, localised antisense transcripts were detected over the pause element of β-actin gene, indicating that possibly double-strand RNA could be formed over the pause element. This might in turn trigger the RNAi pathway and subsequently H3K9me2 formation over the termination region of human β-actin gene.

Next I tested whether Dicer and Ago1 are detectable on chromatin over the termination region of β-actin gene. As demonstrated in Figures 5.3C and 5.3D, Dicer and Ago1 were specifically enriched over the pause element of β-actin gene. Interestingly,

antisense transcription, Dicer protein, Ago1 protein and H3K9me2 mark are all co-localised downstream the poly(A) signal and over the pause element of β-actin gene (Figures 5.1C, 5.3A, 5.3B and 5.3C). This implies that dsRNA may act to recruit RNAi machinery that in turn promotes the assembly of H3K9me2 repressive mark over the termination region of human β-actin gene.

5.4 Loss of R-loops affects antisense transcription and Dicer recruitment over the pause element of human β-actin, but it does not affect H3K9me2 levels

In chapter 4, I show that R-loops are specifically enriched immediately downstream the poly(A) signal of human β-actin gene (Figure 4.1D). Interestingly, this region is just upstream the pause element of β-actin gene where antisense RNA, Dicer, Ago1 and the H3K9me2 mark are detected. An interesting question that arises is whether R-loops play a role or even facilitate the recruitment of RNAi factors and H3K9me2 formation over the termination region of β-actin gene.

To this end I over-expressed GFP-tagged RNase H1, an enzyme that specifically degrades the RNA in RNA/DNA hybrids, in HeLa cells. In chapter 3, I show that over-expression of RNase H1 diminishes R-loop levels (Figure 3.4D). Taking this into consideration, I sought to investigate whether loss of R-loops could affect the detected H3K9me2 levels. To address this I studied the H3K9me2 profile of the β-actin gene following RNase H1 over-expression. Using ChIP analysis (Figure 5.4A) I showed that over-expression of RNase H1 does not affect H3K9me2 levels over the termination

region of β-actin gene. This result suggests that either R-loops are not involved in the H3K9me2 formation over the termination region of β-actin gene or that the repressive mark is already established throughout cell divisions and therefore it is too stable to be affected by the loss of R-loops. Thus I decided to investigate whether other RNAi factors are affected upon RNase H1 over-expression.

I initially tested if antisense RNA detected over the pause element of human β-actin gene is dependent on R-loop formation. Following RNase H1 over-expression, I observed a decrease in the amount of antisense transcription detected over the β-actin pause element (Figure 5.4B). Interestingly, antisense RNA produced over the promoter and intron 1 is unaffected by RNase H1 over-expression even though R-loops are enriched over the intron 1 region (Figure 4.1D). This result could suggest that the R-loops formed at the beginning of β-actin gene do not trigger the antisense transcription detected.

Next I investigated if Dicer recruitment over the pause element of β-actin gene is affected by R-loop resolution. To this end I performed Dicer ChIP analysis following RNase H1 over-expression. As shown in Figure 5.4C, Dicer recruitment was significantly reduced upon RNase H1 over-expression.

These results suggest that R-loops could facilitate production of antisense RNA by exposing the 'unprotected' DNA strand to other Pol II molecules. Antisense RNA could then hybridise with the sense RNA to form double-strand RNA that in turn could trigger Dicer recruitment to the β-actin pause element with subsequent assembly of the RNAi machinery.

5.5 Ago1 protein and antisense transcription are detected over the termination region of mouse β-actin

So far I have shown that the H3K9me2 mark is present over the termination region of human β-actin gene (Figure 5.1C). Based on detection of antisense transcription, Dicer and Ago1 recruitment over the pause element, I predict that the H3K9me2 formation over the termination region of human β-actin gene is RNAi-dependent (Figure 5.3). Furthermore this process seems to be dependent on the formation of R-loops downstream the poly(A) signal (Figure 5.4B and 5.4C).

The question that rises at this point is the molecular function of the H3K9me2 mark over the termination region of β-actin and whether it facilitates efficient termination. To address this I decided to use the mouse model system where constitutive knock-outs of key RNAi players have been generated.

To validate some of the data I obtained with the human gene system, I initially tested if Ago1 protein binds the termination region of endogenous mouse β-actin at the high levels. Employing a number of PCR primers located over the coding and 3'-flanking region, I performed Ago1 ChIP analysis on mouse β-actin gene using mouse embryonic fibroblast (MEF) chromatin. As shown in Figure 5.5A, Ago1 is recruited downstream the poly(A) of mouse β-actin gene.

Next I investigated whether antisense RNA can be detected over the termination region of mouse β-actin gene similar to human β-actin gene. RT-PCR analysis of total RNA from MEF cells revealed that antisense RNA is detected over the intron 1 region and immediately downstream the poly(A) signal (Figure 5.5B). Antisense RNA over the

intron 1 region could be associated with promoter-upstream transcripts (PROMPTs-Preker et al. 2011).

5.6 The H3K9me2 mark is present over the termination region of mouse β-actin gene and it is reduced to a 50% level in Ago2 KO cells

To investigate the molecular function of H3K9me2 mark over the termination region of mammalian genes, I decided to use the Ago2 KO MEFs (Liu et al. 2004). Ago2 protein is shown to be a principal catalytic enzyme of the mammalian RNAi pathway (Liu et al. 2004). Therefore I sought to examine whether knock-out of Ago2 can affect the H3K9me2 profile over the termination region of mouse β-actin gene.

As shown (Figure 5.6A, left panel) in mouse WT cells (MEFs) I observed an enrichment of H3K9me2 mark downstream the poly(A) signal. In Ago2 KO cells a reduced amount of H3K9me2 was detected, indicating that the H3K9me2 mark is affected upon Ago2 depletion, but not completely abolished. As a positive control for these experiments performed on mouse β-actin, I used repetitive DNA sequences of the mouse centromeres known as major satellites in pericentric heterochromatin (Figure 5.6A, right panel).

To normalise for the total H3 levels, I initially performed H3 ChIP on mouse β-actin gene (Figure 5.6B, left panel). Next I corrected the H3K9me2 signals (Figure 5.6A, left panel) to the total H3 levels. The ratio of H3K9me2 versus the total H3 represents the H3K9me2 occupancy that is not dependent on the presence of the histone H3. As

demonstrated in Figures 5.6C in WT cells I observed higher H3K9me2 occupancy immediately downstream the poly(A) signal and over the termination region (amplicons pause, pause2 and C) of mouse β-actin as compared to the H3K9me2 levels observed over the gene body. In Ago2-depleted cells the H3K9me2 signals over the termination region were reduced, indicating a role for Ago2 in H3K9me2 formation over the mouse β-actin gene. H3K9me2 levels were also reduced in Ago2 knock-out cells over the mouse major satellites (Figure 5.6C, right panel).

The data presented so far (Figures 5.5 and 5.6) are consistent with the human β-actin results (Figures 5.1C, 5.3A and 5.3C), pointing towards the evolutionary conservation of this RNAi-dependent H3K9me2 formation over G-rich termination regions of mammalian genes.

5.7 Depletion of Ago2 enhances Ago1 recruitment to the mouse β-actin gene

Argonaute proteins are key components in RNAi-silencing pathway (Hutvagner and Simard, 2008). In mouse and human species Ago1 and Ago2 have been shown to a play a role in heterochromatin silencing (Janowski et al. 2006; Kim et al. 2006).

Over the endogenous mouse β-actin, Ago1 binds the termination region (Figure 5.5A) and Ago2 seems to be involved, to some extent, in the H3K9me2 formation over the termination region (Figure 5.6C). Since H3K9me2 profile over the mouse β-actin is only partially affected upon Ago2 knock-out, Ago1 could possibly compensate for Ago2 depletion. To address this point I performed Ago1 ChIP in Ago2 knock-out cells. As shown (Figure 5.7, left panel) Ago1 recruitment to mouse β-actin is enhanced upon Ago2

depletion. Interestingly, Ago1 recruitment is not facilitated in Ago2 knock-out cells over the major satellites (Figure 5.7, right panel), indicating that Ago1 does not compensate for Ago2 depletion in the centromeres.

These results could indicate that Ago1 and Ago2 have similar roles in the H3K9me2 assembly over the termination region of endogenous mouse β-actin gene. Furthermore it could be also predicted that Ago2 has an important role in mouse β-actin gene and therefore Ago1 compensates for Ago2 depletion.

5.8 R-loops are formed over the mouse β-actin gene. Depletion of Ago2 partially affects their distribution

Next I investigated if R-loops can be formed over the transcribed region of mouse β-actin gene similar to human β-actin gene (Figure 4.1D). As demonstrated (Figure 5.8, left panel) R-loops were detected over the coding region of mouse β-actin gene, with their specific enrichment immediately upstream the poly(A) signal (intron 5 probe). In contrast to the human β-actin gene where R-loops were specifically enriched over intron 1 and upstream the pause element, R-loop signals in the mouse β-actin gene seem to be more widespread. In Ago2 knock-out cells DIP signals are reduced over most of the amplicons. R-loops seem to remain unaffected over pause2 and C probes which are located at the 3' flanking region (Figure 5.8, left panel). One possible explanation for this result could be that reduced R-loop levels in Ago2 knock-out cells are connected to the reduced H3K9me2 levels observed (Figure 5.6C).

R-loops are also detected over the mouse major satellites (Figure 5.8, right panel). Similarly, experiments in fission yeast have shown that R-loop foci co-localise with

centromeric repeats (Nakama et al. 2012). This points towards the generality of a potential R-loop dependent heterochromatin formation over centromeric repeats.

5.9 Loss of R-loops and partial loss of the H3K9me2 mark affect transcriptional termination of endogenous mouse β-actin gene

The data obtained so far suggest that R-loops and the H3K9me2 mark are present over the termination region of mouse β-actin gene (Figures 5.6 and 5.8). Results presented in chapter 3 revealed that R-loops are an important feature of pause-dependent transcriptional termination (Figure 3.4). To investigate if both R-loops and H3K9me2 mark are needed for efficient transcriptional termination of mouse β-actin gene, I performed Pol II ChIP on mouse β-actin gene under the following conditions: in untreated WT cells where both the H3K9me2 mark and R-loops are present over the termination region, in WT cells over-expressing RNaseH1-GFP plasmid where H3K9me2 mark is present but the R-loop signals are decreased, in Ago2 knock-out cells where H3K9me2 mark is reduced but R-loops are formed, in Ago2 knock-out cells over-expressing RNaseH1-GFP plasmid where both the H3K9me2 mark and R-loops are reduced.

As demonstrated (Figure 5.9), in WT untreated cells I observed Pol II enrichment over the promoter of mouse β-actin gene. Pol II detected over the coding region was lower than over the promoter, similar to human β-actin and γ-actin genes (Figures 4.1A and 4.5A). Pol II termination occurs over probe D located ~3 kb downstream the poly(A)

signal. Over-expression of RNase H1 in WT cells causes an increase in Pol II density across the promoter, gene body and immediately downstream the poly(A) signal (probes pause and pause2). However, Pol II density is not significantly increased over C and D probes positioned further downstream the 3' flanking region, suggesting that resolution of R-loops affects Pol II density across mouse β-actin gene, but it is not enough to cause a strong termination defect.

In Ago2-depleted cells Pol II density over the termination region is unaffected as compared to WT cells, indicating that the H3K9me2 mark still detected in combination with the formation of R-loops is enough to promote efficient transcriptional termination of mouse β-actin gene. Over-expression of RNaseH1-GFP on Ago2 knock-out cells causes an increase in Pol II density across the promoter, gene body and termination region. Interestingly, I observed and increase in Pol II density over probes C and D, indicative of a defect in transcriptional termination. In contrast to over-expression of RNase H1 in WT cells where the mark is present, over-expression of RNase H1 in Ago2 knock-out cells where the mark is reduced affects transcriptional termination.

This result suggests that only combined loss of R-loops and H3K9me2 mark can affect transcriptional termination of mouse β-actin gene. In conclusion my results predict that R-loops and the H3K9me2 mark formed over the termination region of mouse β-actin gene are both essential components of efficient pause-dependent Pol II termination.

5.10 The H3K9me3 mark is present at the 5' end of mouse β-actin gene and is not affected by Ago2 depletion

Another repressive mark that I have not studied so far is H3K9me3. To test if H3K9me3 mark is also present at the termination region of mouse β-actin gene and is affected by Ago2 knock-out similar to H3K9me2 mark, I performed H3K9me3 ChIP on mouse β-actin gene (Figure 5.10A, left panel). Surprisingly, in WT cells H3K9me3 is highly enriched over intron 1 and intron 3 but not over the termination region. Furthermore in Ago2 knock-out cells H3K9me3 remains unaffected. To normalise for the total H3 levels, I initially performed ChIP analysis using anti-H3 antibody (Figure 5.10B) and then I divided the H3K9me3 signals to the total H3 levels (Figure 5.10C). According to the H3K9me3 profile obtained after the normalisation, in WT cells H3K9me3 occupancy is at the highest level over intron 1. Upon Ago2 depletion H3K9me3 levels over intron 1 and across the mouse β-actin gene remain unaffected. Interestingly, H3K9me3 levels are reduced in Ago2 knock-out cells over the major satellites, suggesting that in mouse β-actin gene Ago2 does not play a role in H3K9me3 formation.

The above results are not consistent with the H3K9me2 data (Figure 5.6C). H3K9me3 is present at the beginning of β-actin gene whereas H3K9me2 forms over the termination region, suggesting that the repressive mark that possibly facilitates R-loop dependent termination pausing is specifically the H3K9me2 mark and not H3K9me3.

5.11 Depletion of Suv39h1/h2 methyltransferases does not affect H3K9me2 and H3K9me3 levels of endogenous mouse β-actin gene

Depletion of Ago2 protein reduced H3K9me2 levels over the termination region of mouse β-actin gene, but it did not completely abolish them (Figure 5.6C). Furthermore I observed a termination defect in Ago2 knock-out cells over-expressing RNaseH1-GFP, but the defect was mostly localised over C and D probes located further downstream the 3'-flanking region (Figure 5.9). In order to further substantiate the molecular function of H3K9me2 mark in R-loop dependent termination pausing, I decided to use the mouse Suv39h1/h2 double knock-out cell line (Rea et al. 2000). Suv39h1 and Suv39h2 are lysine histone methyltransferases (HKMTs) with a SET-containing domain (Rea et al. 2000). Biochemically, both Suv39h1 and G9a, which is another HKMT, can catalyse mono-, di- and trimethylation reactions on H3K9 in mammalian cells (Collins et al. 2005; Kubicek at al. 2007; Shinkai and Tachibana, 2011).

To investigate whether Suv39h1/h2 enzymes affect H3K9me2 levels detected over the termination region of mouse β-actin, I performed H3K9me2 ChIP analysis in Suv39h1/h2 knock-out cells. As shown (Figure 5.11A, left panel), I did not observe a decrease in H3K9me2 signals, suggesting that Suv39h1 and Suv39h2 are not the methyltransferases of H3K9me2 in mouse β-actin gene. However H3K9me2 levels were significantly decreased in Suv39h1/h2 double deficient cells over the major satellites indicating that Suv39h1 and Suv39h2 play a role in H3K9me2 formation on pericentric heterochromatin.

Investigations on Suv39h1/h2 null depleted cells have shown that Suv39h1 and Suv39h2 are crucial HKMTs for H3K9me3 on pericentric heterochromatin (Peters et al. 2003; Rea et al. 2003). To test if Suv39h1 and Suv39h2 are HKMTs for H3K9me3 on mouse β-actin gene, I performed H3K9me3 ChIP on mouse β-actin in WT and Suv39h1/h2 double knock-out cells (Figure 5.11B, left panel). Unexpectedly I observed an increase in H3K9me3 levels over intron 1 and intron 3 probes upon Suv39h1/h2 double depletion, indicating that Suv39h1 and Suv39h2 are probably not the HKMTs for H3K9me3 on mouse β-actin gene. In contrast to the endogenous β-actin gene, upon depletion of Suv39h1 and Suv39h2, H3K9me3 levels are completely abolished over the major satellites (Figure 5.11B, right panel). This result is consistent with the existing literature (Peters et al. 2003; Rea et al. 2003), further confirming that Suv39h1 and Suv39h2 catalyse H3K9me3 on pericentric heterochromatin.

5.12 Discussion

Experiments presented in this Chapter reveal a link between R-loop structures and the RNAi pathway and suggest a mechanism for the regulation of pause-dependent transcriptional termination. My data also suggest the presence of an evolutionary conserved mechanism for the H3K9me2 assembly over G-rich termination regions of mammalian genes.

In human β-actin gene I observed formation of the H3K9me2 mark over the pause element (Figure 5.1C), but also the recruitment of the key RNAi factors, Ago1 and Dicer (Figures 5.3B and 5.3C). Furthermore antisense RNA was detected over the β-actin pause element (Figure 5.3A). Interestingly, R-loops are specifically enriched immediately

upstream the pause element (5' pause probe, Figure 4.1D). Resolution of R-loops by RNaseH1 over-expression affects antisense transcription specifically generated from the pause element and Dicer recruitment over the pause element of human β-actin gene (Figures 5.4B and 5.4C). It is notable that RNase H1 over-expression does not affect antisense transcription generated from the promoter and intron 1 regions and that the RNAi machinery is not present at the beginning of human β-actin gene. An interesting link between R-loops formed specifically over the termination region and the H3K9me2 mark could then be predicted.

Indeed, based on results obtained from the human model, I suggest that the formation of R-loops upstream transcriptional pause regions leaves the non-template DNA strand unprotected and prone to be the template for antisense transcription as Pol II elongates further downstream. The question how antisense transcription is generated remains to be elucidated. Given that two Pol II molecules would collide if they both transcribed in opposite orientations, one possible scenario could be that antisense transcription could occur while Pol II has already moved further downstream the pause regionownstream the poly(A) signal prior to termination, another Pol II molecule could transcribe in the antisense orientation having as template the 'unprotected' DNA strand. Then antisense RNA would be produced from the pause element which could trigger Dicer recruitment over the pause element to process double-stranded RNA generated from the hybridisation of sense and antisense transcripts. The recruitment of other RNAi factors, such as Ago1 protein, would then be enhanced leading finally to the recruitment of HKMTs for the formation of H3K9me2 heterochromatin mark.

Results presented in Chapter 3 and Chapter 4 suggest that R-loop formation at the termination region facilitates Pol II pausing. In this final Chapter, I added a new player in the R-loop dependent termination process, the H3K9me2 repressive mark. H3K9me2 forms preferentially over the termination region of human β-actin and γ-actin genes (Figures 5.1C and 5.2C). The molecular function of its presence requires further investigation. However I suggest that R-loop structures once they are formed and before their subsequent resolution by senataxin could trigger heterochromatin assembly in an RNAi-dependent manner. R-loop structures alone might not be sufficient to pause Pol II at the termination region. In transfected gene constructs I previously showed that resolution of R-loops by over-expression of RNase H1 causes termination defect (Figure 3.4). In the endogenous β-actin gene, over-expression of RNase H1 does not cause termination defect (data not shown). One possible explanation for this observation is that in endogenous genes two factors are required to facilitate Pol II pausing prior to efficient termination, R-loop structures and the H3K9me2 mark. Therefore pausing could be facilitated by the heterochromatin mark when R-loops are resolved. Due to high levels of transcription in gene constructs, chromatin formation and packaging could be different and H3K9me2 formation could be not favoured.

Another interesting point is whether the H3K9me2 mark is stable throughout cell divisions and epigenetically inherited. Based on results demonstrated in Figure 5.4A, it could be speculated that even though R-loops could trigger the RNAi pathway and the subsequent H3K9me2 formation, the mark is stable throughout cell divisions and is not affected by their resolution. Therefore it could be suggested that R-loops facilitate H3K9me2 establishment but not maintenance.

The mouse model system has revealed interesting observations concerning this R-loop mediated H3K9me2 formation. Results obtained in mouse β-actin gene so far correspond to the data observed in human β-actin gene: in both cases there is R-loop formation downstream the poly(A) signal (Figures 4.1D and 5.8), antisense transcription detected immediately downstream the 3' end R-loop enrichment (Figures 5.3A and 5.5B), Ago1 recruitment over the region where antisense transcription occurs (Figures 5.3C and 5.5A) and finally H3K9me2 mark localised over the termination region (Figures 5.1C and 5.6C). These observations point towards the conclusion that R-loop mediated H3K9me2 formation over G-rich termination regions is a mammalian conserved mechanism that facilitates Pol II pausing prior to efficient termination. It is particularly notable that in both mouse and human β-actin genes, the R-loop structures and the RNAi machinery are specifically localised downstream the poly(A) signal and around G-rich sequence elements.

The advantage of using the mouse model system is the availability of central RNAi factors knock-out cell lines. Using the Ago2 knock-out cell line I observed that Ago2 protein plays a role in the H3K9me2 formation since upon its depletion, the mark is reduced to a 50% level (Figure 5.6C). One possible explanation for why I did not observe a full reduction of the mark could be the enhanced recruitment of Ago1 protein following depletion of Ago2 (Figure 5.7, left panel). Interestingly, the same effect was not observed over the major satellites, indicating that in the mouse β-actin gene Ago2 and Ago1 compensate for each other and upon one's absence the other one 'takes over'.

R-loops are formed over the transcribed region of mouse β-actin but they seem to be more widespread as compared to the human β-actin gene (Figures 4.1D and 5.8, left

panel). Nevertheless their main enrichment is detected immediately upstream the poly(A) signal and upstream the region where antisense RNA, Ago1 and H3K9me2 are detected, analogous to the human β-actin gene. In Ago2 knock-out cells, R-loop signals are slightly affected but still detected (Figure 5.8, left panel).

R-loops are also formed over the mouse major satellites (Figure 5.8, right panel). Nakama and colleagues have recently shown that in fission yeast R-loop foci co-localise with centromeric repeats (Nakama et al. 2012). Therefore it could be speculated that R-loops formed over centromeric repeats facilitate heterochromatin formation in an evolutionary conserved manner.

My initial hypothesis was that R-loop structures and the H3K9me2 mark are both needed to facilitate Pol II pausing prior to efficient termination. Based on this it could be predicted that loss of R-loops and loss of the mark would result in inefficient pausing and subsequent termination defect. Indeed, Pol II ChIP analysis on mouse β-actin gene revealed that Ago2 knock-out cells over-expressing RNase H1 show increased levels of Pol II density over two amplicons positioned further downstream the termination region as compared to WT cells (amplicons C and D, Figure 5.9). This result suggests that efficient Pol II termination of endogenous mouse β-actin requires the formation of R-loop structures and H3K9me2 repressive mark over the termination region. However Ago2 knock-out results in only partial loss of the H3K9me2 mark, I could predict that complete loss of the H3K9me2 mark combined with resolution of R-loops would lead to stronger termination defect. Towards this end further investigation is required.

H3K9me3 mark is also detected over the mouse β-actin gene but unexpectedly it is particularly enriched at the beginning of the gene and is not affected by Ago2 depletion

(Figure 5.10C). Thus I suggest that the histone mark that plays a role in pause-dependent transcriptional termination is H3K9me2, and not H3K9me3. A possible explanation for the presence of H3K9me3 at the beginning of mouse β-actin gene could be the prevention of internal initiation from cryptic start sites. H3K9me3 could close chromatin to prevent transcription initiation. Indeed it has been documented that H3K9me3 mark is repressive when found in the promoter region but activating when found over the gene body (Berger, 2007; Kouzarides, 2007; Li et al. 2007; Pfluger and Wagner, 2007). Nevertheless I could conclude that two repressive histone marks are detected over the mouse β-actin gene that have a different localisation pattern and are not similarly affected in the absence of Ago2 protein (Figures 5.6C and 5.10C).

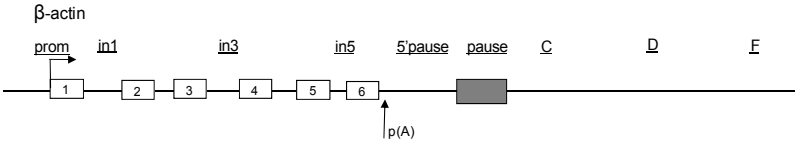
Suv39h1 and Suv39h2 are HKMTs that catalyse mono-, di- and trimethylation reactions on H3K9 in mammalian cells (Collins et al. 2005; Kubicek at al. 2007; Shinkai and Tachibana, 2011). Based on this I decided to use the mouse Suv39h1/h2 double knock-out cell line in order to further investigate the molecular function of H3K9me2 mark in R-loop dependent termination pausing. Experiments presented on Figure 5.11A show that Suv39h1 and Suv39h2 do not catalyse H3K9me2 on endogenous mouse β-actin gene. In major satellites H3K9me2 mark is reduced significantly upon Suv39h1/h2 double knock-out, indicating that these two HKMTs play a role in H3K9me2 formation on pericentric heterochromatin (Figure 5.11A, right panel). G9a/KMT1C and GLP/Eu-HMTase1/KMT1D are members of the Suv39 subgroup of SET-containing proteins and together have been shown to be HKMTs for H3K9me1 and H3K9me2 (Jenuwein et al. 1998; Tachibana et al. 2001; 2002; 2005). Since Suv39h1 and Suv39h2 seem to not catalyse H3K9me2 on mouse β-actin gene, in the future it would be interesting to

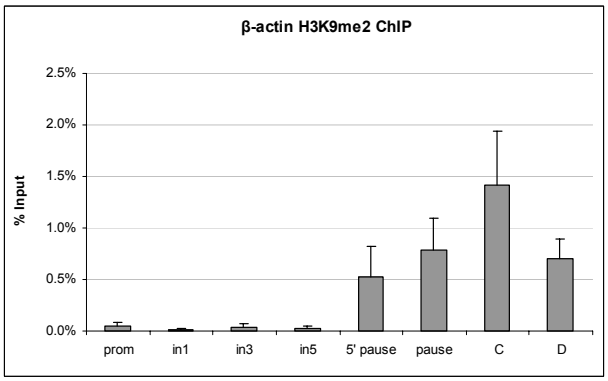
investigate whether G9a/GLP are the crucial HKMTs for H3K9me2 on mouse β-actin gene.

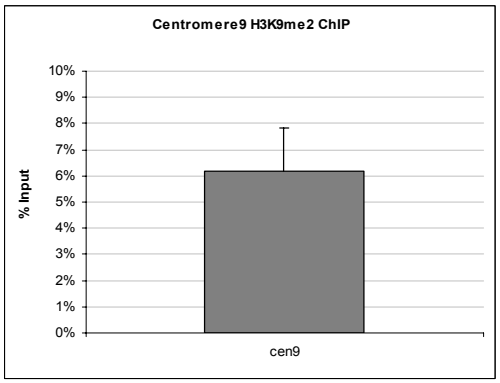
Interestingly, Suv39h1 and Suv39h2 are not the HKMTs for H3K9me3 either (Figure 5.11B, left panel), even though double knock-out of Suv39h1/h2 completely abolishes H3K9me3 mark on major satellites (Figure 5.11B, right panel). The above observations suggest that Suv39h1 and Suv39h2, as previously shown, catalyse H3K9me3 on major satellites (Tachibana et al. 2001; 2002; Peters et al. 2003; Rea et al. 2003) but interestingly there is another methyltransferase responsible for the H3K9me3 mark on endogenous mouse β-actin gene. Even though H3K9me3 is probably not involved in the R-loop-dependent termination process of mouse β-actin gene, it would be interesting to identify in the future the HKMT for H3K9me3 on β-actin gene and to also investigate the molecular function of H3K9me3 mark at the beginning of the gene.

Experiments presented in this Chapter add a new molecular layer to the mechanism, presented in Chapter 3 and Chapter 4, of R-loop dependent transcriptional termination in mammals. Overall my results suggest that the formation of R-loops in transcriptional pause regions could facilitate the formation of H3K9me2 mark in an RNAi-dependent manner. R-loop structures and the H3K9me2 mark formed over the termination region could anchor Pol II and so promote efficient termination.

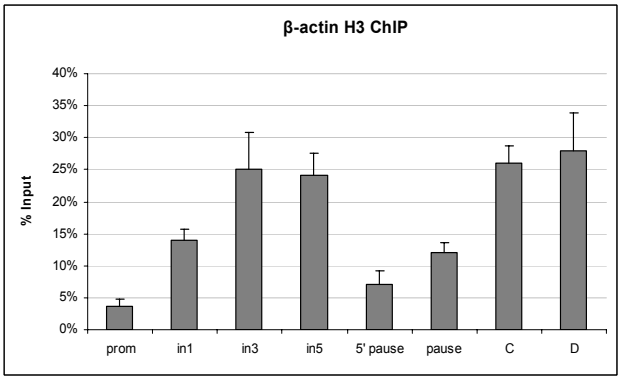
A

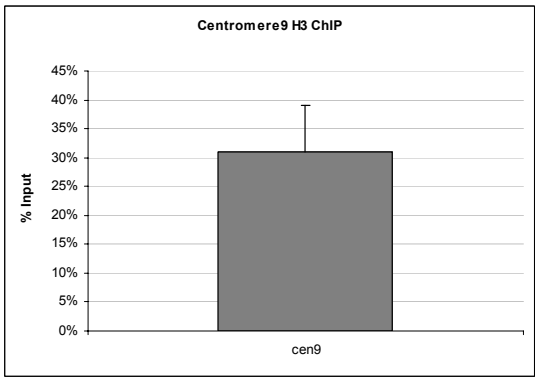

β-actin
prom
in1
in3
in5
5'pause
pause
C
D
F
1
2
3
4
5
6
p(A)


β-actin H3K9me2 ChIP
% Input
2.5%
2.0%
1.5%
1.0%
0.5%
0.0%
prom
in1
in3
in5
5' pause
pause
C
D

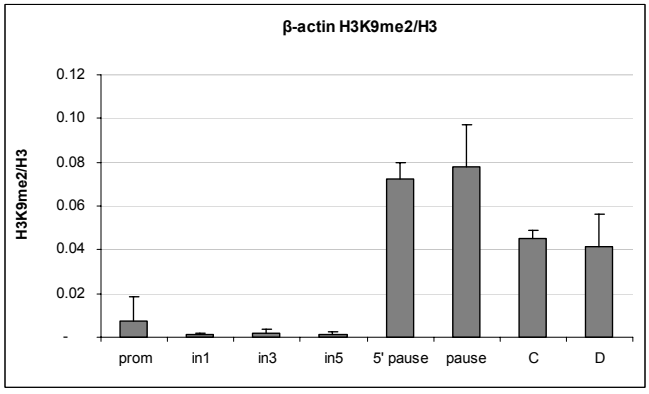

Centromere9 H3K9me2 ChIP
% Input
10%
9%
8%
7%
6%
5%
4%
3%
2%
1%
0%
cen9

B


β-actin H3 ChIP
% Input
40%
35%
30%
25%
20%
15%
10%
5%
0%
prom
in1
in3
in5
5' pause
pause
C
D


Centromere9 H3 ChIP
% Input
45%
40%
35%
30%
25%
20%
15%
10%
5%
0%
cen9

C


β-actin H3K9me2/H3
H3K9me2/H3
0.12
0.10
0.08
0.06
0.04
0.02
-
prom
in1
in3
in5
5' pause
pause
C
D

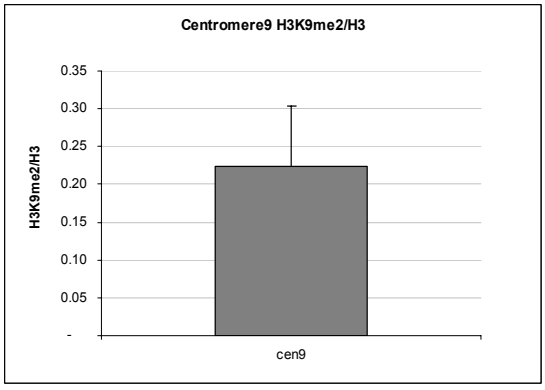

Centromere9 H3K9me2/H3
H3K9me2/H3
0.35
0.30
0.25
0.20
0.15
0.10
0.05
-
cen9

Figure 5.1 H3K9me2 mark is present at the termination region of endogenous human β-actin gene.

A. Left panel: H3K9me2 ChIP analysis performed on endogenous β-actin gene. H3K9me2 forms over the termination region of β-actin. Right panel: H3K9me2 ChIP analysis performed on human centromere 9 (positive control).

B. Left panel: H3 ChIP analysis performed on endogenous β-actin gene. Downstream the poly(A) signal and over the pause element of β-actin there is a nucleosome-free region. Right panel: H3 ChIP analysis performed on human centromere 9.

C. Left panel: Ratio of H3K9me2 signal versus H3 signal on β-actin. According to the normalisation of H3K9me2 signal to total H3 signal, H3K9me2 mark peaks downstream the poly(A) signal and around the pause element of human β-actin gene. Right panel: Ratio of H3K9me2 signal versus H3 signal on human centromere 9.

ChIP profiles are based on average values +/- SD from at least three independent biological experiments.

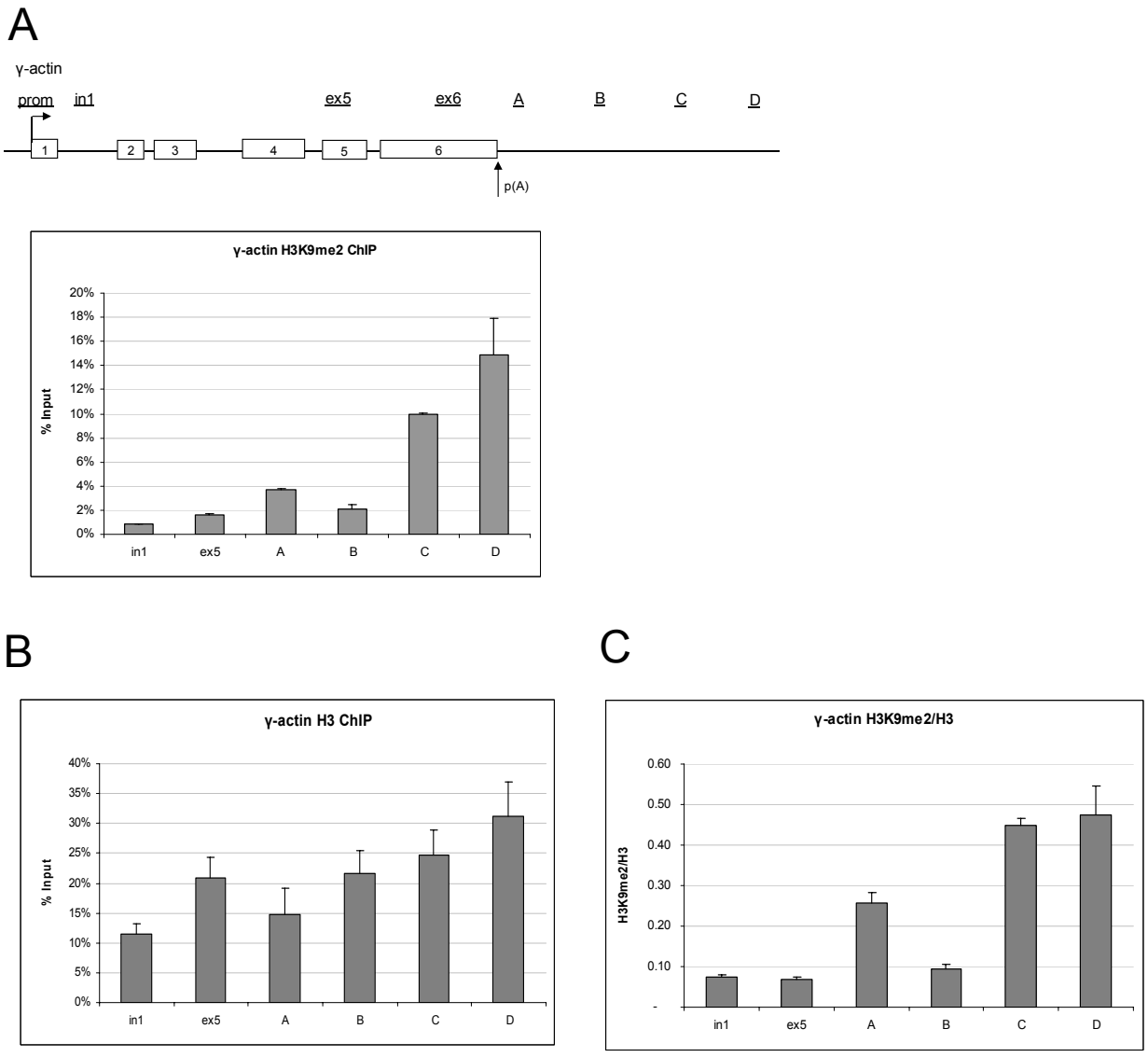


Figure 5.2 H3K9me2 mark is present at the termination region of endogenous human γ-actin gene.

A. H3K9me2 ChIP analysis performed on endogenous γ-actin gene. H3K9me2 forms over the termination region of γ-actin.

B. H3 ChIP analysis performed on endogenous β-actin gene.

C. Ratio of H3K9me2 signal versus H3 signal on γ-actin. According to the normalisation of H3K9me2 signal to total H3 signal, H3K9me2 mark peaks over the termination region of human γ-actin gene.

ChIP profiles are based on average values +/- SD from at least three independent biological experiments.

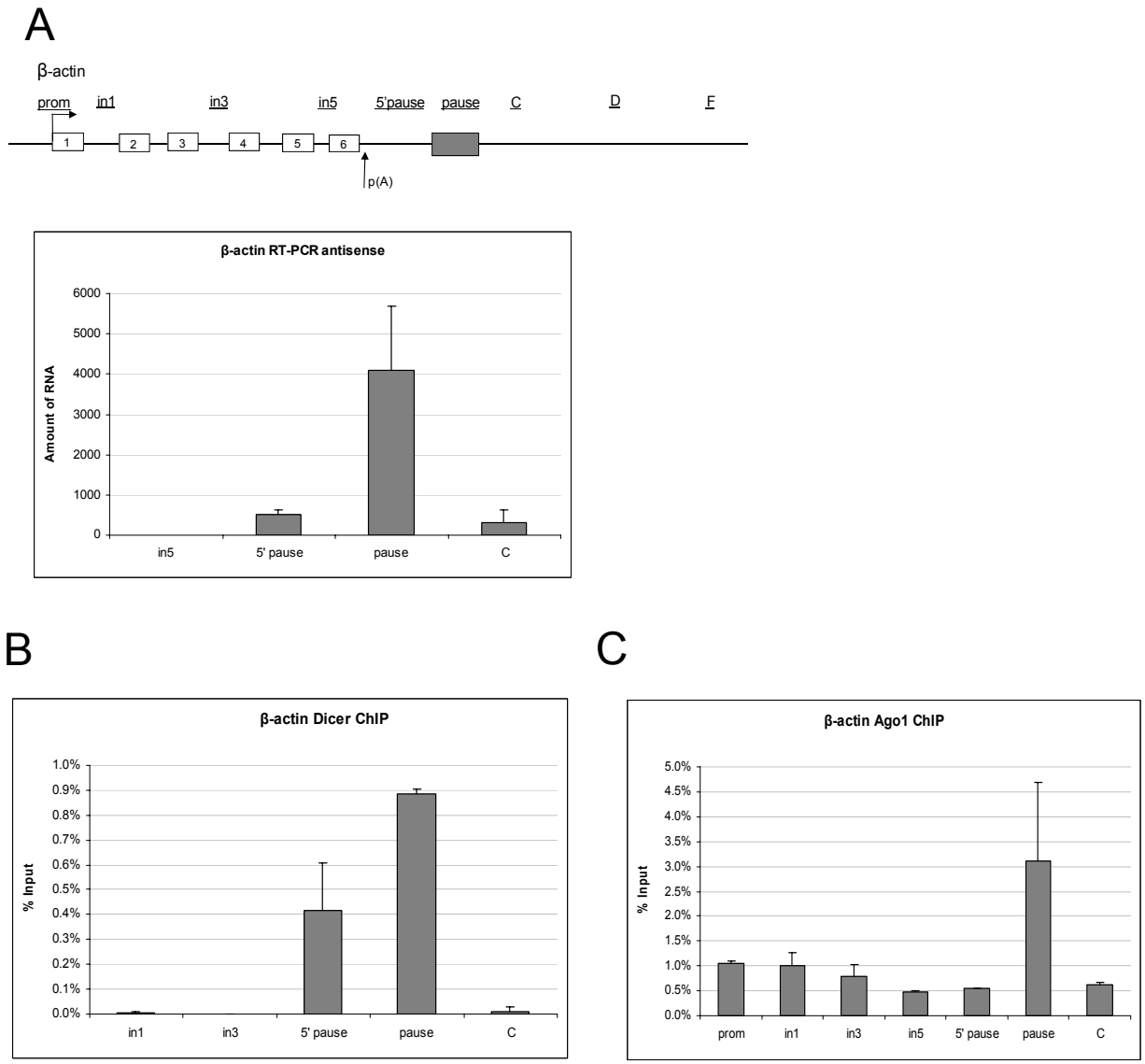


Figure 5.3 Detection of antisense transcription and recruitment of Dicer and Ago1 proteins over the pause element of human β-actin gene.

A. RT-PCR analysis of total RNA performed on endogenous β-actin gene. RT reaction was performed with region-specific forward primer. Antisense transcription is detected over the pause element of human β-actin gene. Average qRT-PCR values +/- SD from four independent biological experiments are presented.

B. Dicer ChIP analysis performed on endogenous β-actin gene. Dicer protein is recruited downstream the poly(A) signal and over the pause element of β-actin gene.

C. Ago1 ChIP analysis of β-actin gene. Ago1 protein is recruited over the pause element of β-actin at the highest level.

ChIP profiles are based on average values +/- SD from at least three independent biological experiments.

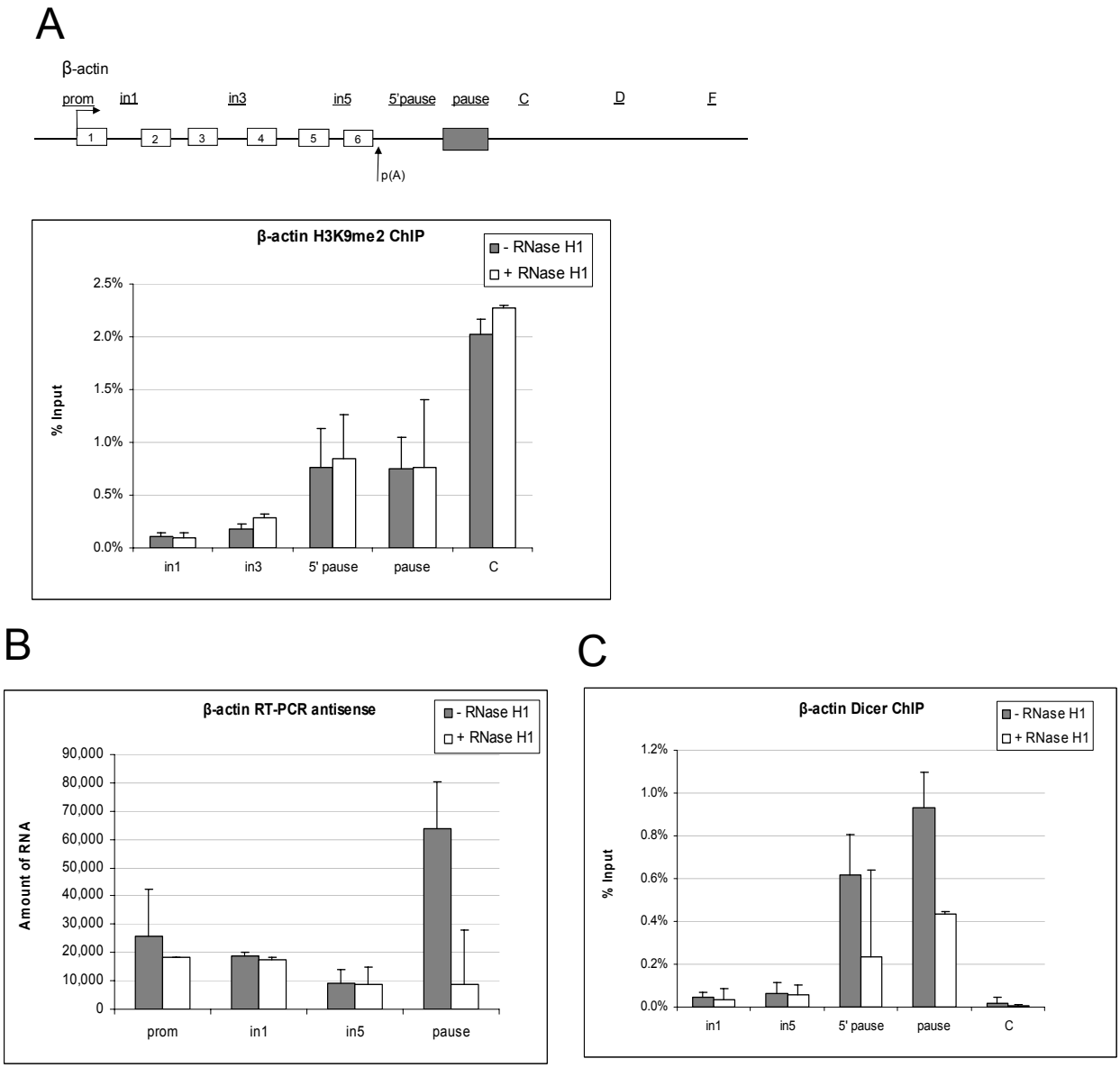


Figure 5.4 RNase H1 over-expression does not affect H3K9me2 levels, but it affects antisense transcription and Dicer recruitment over the pause element of human β-actin gene.

A. H3K9me2 ChIP analysis of endogenous β-actin gene following over-expression of 2 μg of RNaseH1-GFP plasmid in Hela cells. Upon RNase H1 over-expression, H3K9me2 levels over the termination region of β-actin remain unaffected. ChIP profile is based on average values +/- SD from three independent biological experiments.

B. RT-PCR of total RNA performed on endogenous β-actin gene with or without over-expression of 2 μg of RNaseH1-GFP plasmid. Over-expression of RNaseH1 leads to reduction of antisense transcript produced over the pause element of β-actin. Average qRT-PCR values +/- SD from three independent biological experiments are presented.

C. Dicer ChIP analysis of β-actin gene following over-expression of 2 μg of RNaseH1-GFP plasmid. Dicer recruitment to the pause element of human β-actin gene is reduced upon over-expression of RNase H1. ChIP profiles are based on average values +/- SD from three independent biological experiments.

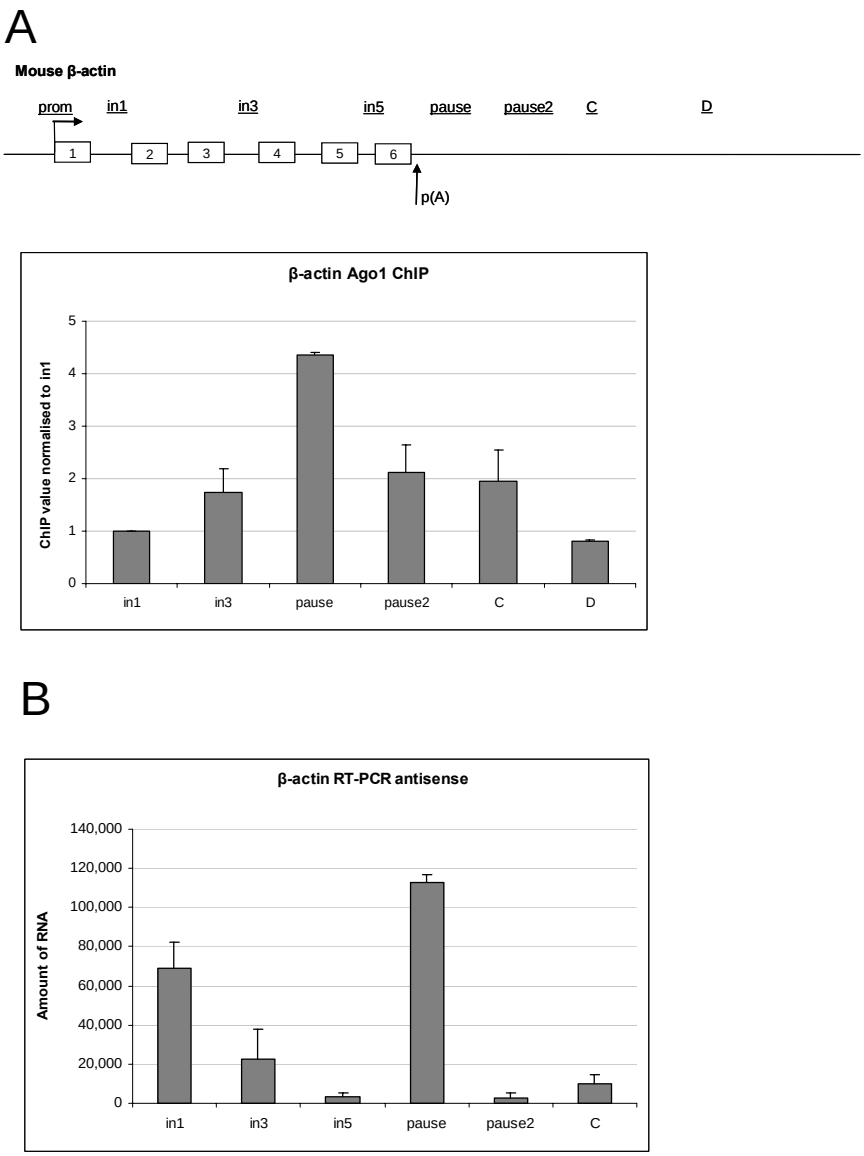
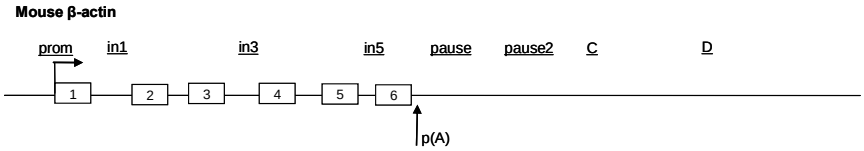


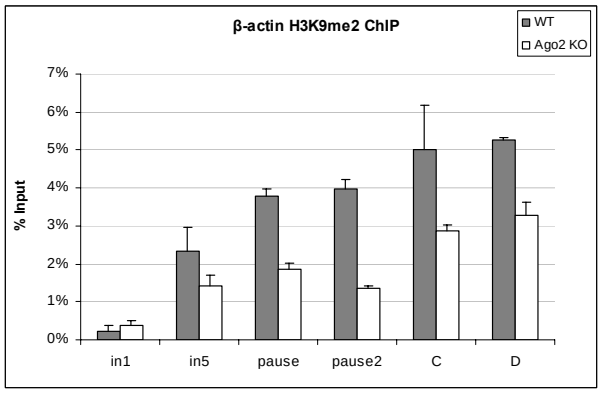
Figure 5.5 Ago1 protein and antisence transcription are detected over the termination region of endogenous mouse β-actin.

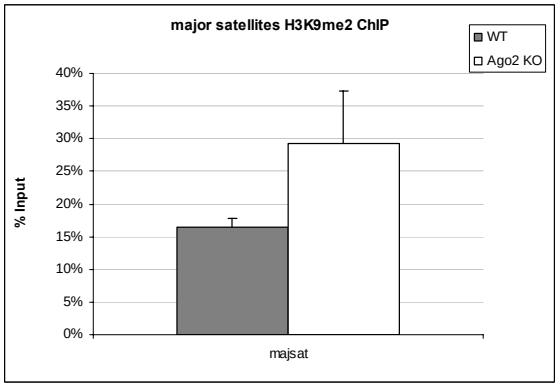
A. Diagram of mouse β-actin gene, with exons shown as white boxes. Positions of primers used in ChIP are shown above the diagram. Bottom panel: Ago1 ChIP analysis analysis performed on endogenous mouse β-actin gene in MEFs. Ago1 is recruited downstream the poly(A) signal of β-actin at the highest level. ChIP profile is based on average values +/- SD from three independent biological experiments.

B. RT-PCR analysis of total RNA from MEF cells performed on endogenous β-actin gene. RT reaction was performed with region-specific forward primer. Antisense transcription is detected over the intron 1 and immediately downstream the poly(A) signal of mouse β-actin gene. Average qRT-PCR values +/- SD from four independent biological experiments are presented.

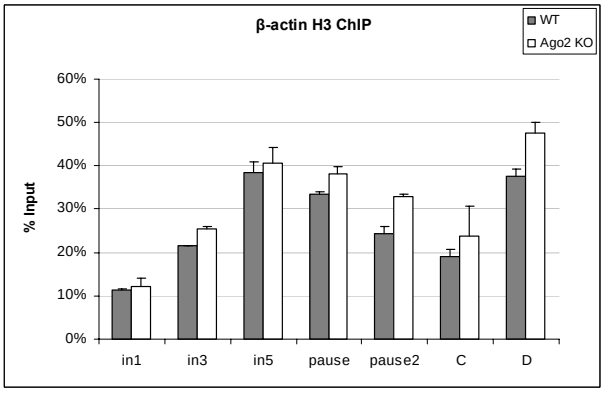
A

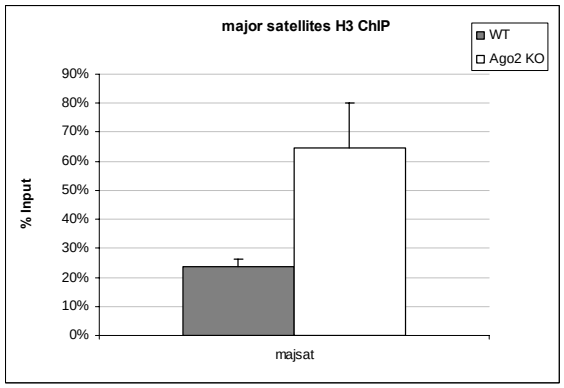

Mouse β-actin
prom
in1
in3
in5
pause
pause2
C
D
1
2
3
4
5
6
p(A)


β-actin H3K9me2 ChIP
WT
Ago2 KO
% Input
7%
6%
5%
4%
3%
2%
1%
0%
in1
in5
pause
pause2
C
D

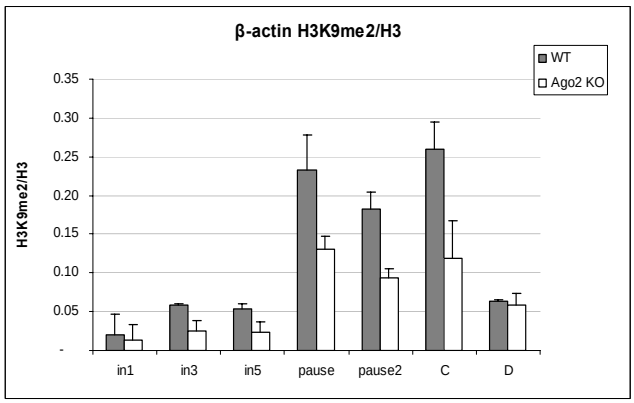

major satellites H3K9me2 ChIP
WT
Ago2 KO
% Input
40%
35%
30%
25%
20%
15%
10%
5%
0%
majsat

B


β-actin H3 ChIP
WT
Ago2 KO
% Input
60%
50%
40%
30%
20%
10%
0%
in1
in3
in5
pause
pause2
C
D


major satellites H3 ChIP
WT
Ago2 KO
% Input
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
majsat

C


β-actin H3K9me2/H3
WT
Ago2 KO
H3K9me2/H3
0.35
0.30
0.25
0.20
0.15
0.10
0.05
-
in1
in3
in5
pause
pause2
C
D

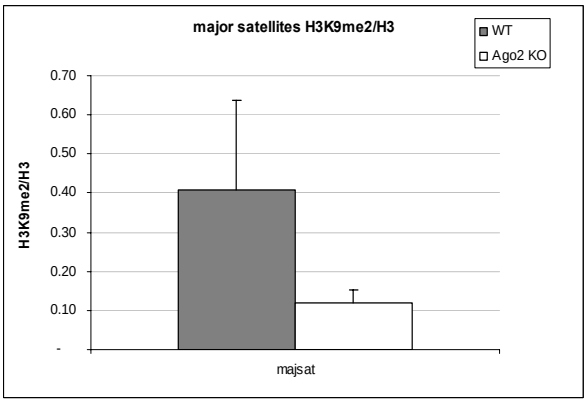

major satellites H3K9me2/H3
WT
Ago2 KO
H3K9me2/H3
0.70
0.60
0.50
0.40
0.30
0.20
0.10
-
majsat

Figure 5.6 The H3K9me2 mark is present at the termination region of endogenous mouse β-actin gene and is reduced to a 50% level in Ago2 KO cells.

A. Diagram of mouse β-actin gene, with exons shown as white boxes. Positions of primers used in ChIP are shown above the diagram. Left panel: H3K9me2 ChIP analysis performed on endogenous mouse β-actin gene in WT and Ago2 KO cells. H3K9me2 forms over the termination region of mouse β-actin at the highest level. In Ago2-depleted cells, H3K9me2 occupancy is reduced. Right panel: H3K9me2 ChIP analysis performed on the mouse major satellites in WT and Ago2 KO cells.

B. Left panel: H3 ChIP analysis of mouse β-actin in WT and Ago2 KO cells. Right panel: H3 ChIP analysis performed on major satellites.

C. Left panel: Ratio of H3K9me2 signal versus H3 signal on mouse β-actin in WT and Ago2 KO cells. According to the normalisation of H3K9me2 signal to total H3 signal, in WT cells the H3K9me2 mark peaks immediately downstream the poly(A) signal. Upon Ago2 depletion, H3K9me2 is reduced to a 50% level. Right panel: Ratio of H3K9me2 signal versus H3 signal on major satellites in WT and Ago2 KO cells.

WT cells are represented as grey bars and Ago2 KO cells are represented as white bars. ChIP profiles are based on average values +/- SD from at least three independent biological experiments.

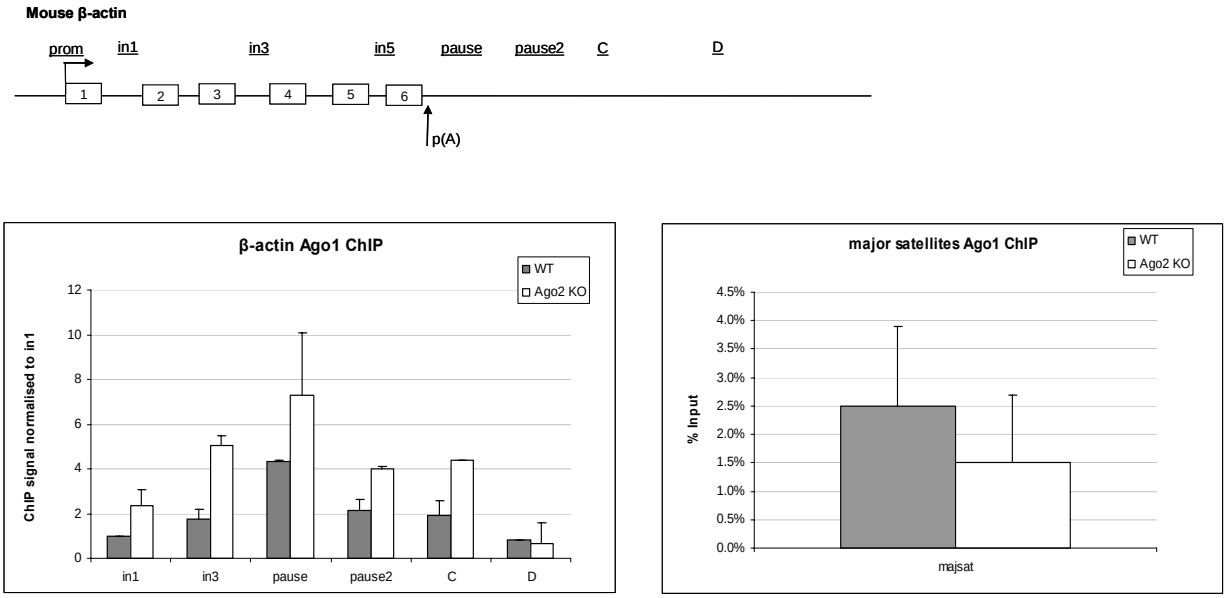


Figure 5.7 Depletion of Ago2 protein enhances recruitment of Ago1 to the mouse β-actin gene.

Left panel: Ago1 ChIP analysis performed on endogenous mouse β-actin gene in WT (grey bars) and Ago2 KO cells (white bars). Recruitment of Ago1 protein to the mouse β-actin gene is increased in Ago2 KO cells. Right panel: Ago1 ChIP analysis of major satellites. ChIP profile is based on average values +/- SD from at least three independent biological experiments.

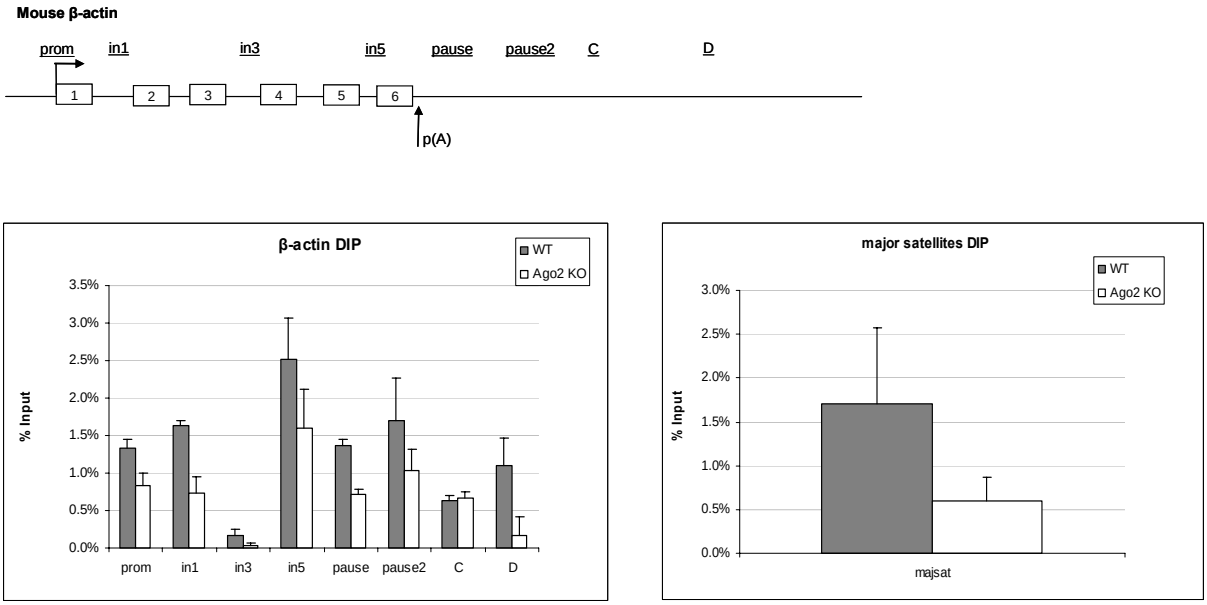


Figure 5.8 R-loops are formed over endogenous mouse β-actin. Depletion of Ago2 partially affects their distribution.

Left panel: DIP analysis on mouse β-actin in WT (grey bars) and Ago2 KO (white bars) cells. analysis performed on endogenous mouse β-actin gene in WT (grey bars) and Ago2 KO cells (white bars). R-loops are detected over the transcribed region of mouse β-actin. In Ago2 KO cells, R-loop distribution is partially affected. Right panel: R-loop profile of major satellites. R-loops form over the mouse major satellites.

R-loop profile is based on average values +/- SD from at least three independent biological experiments.

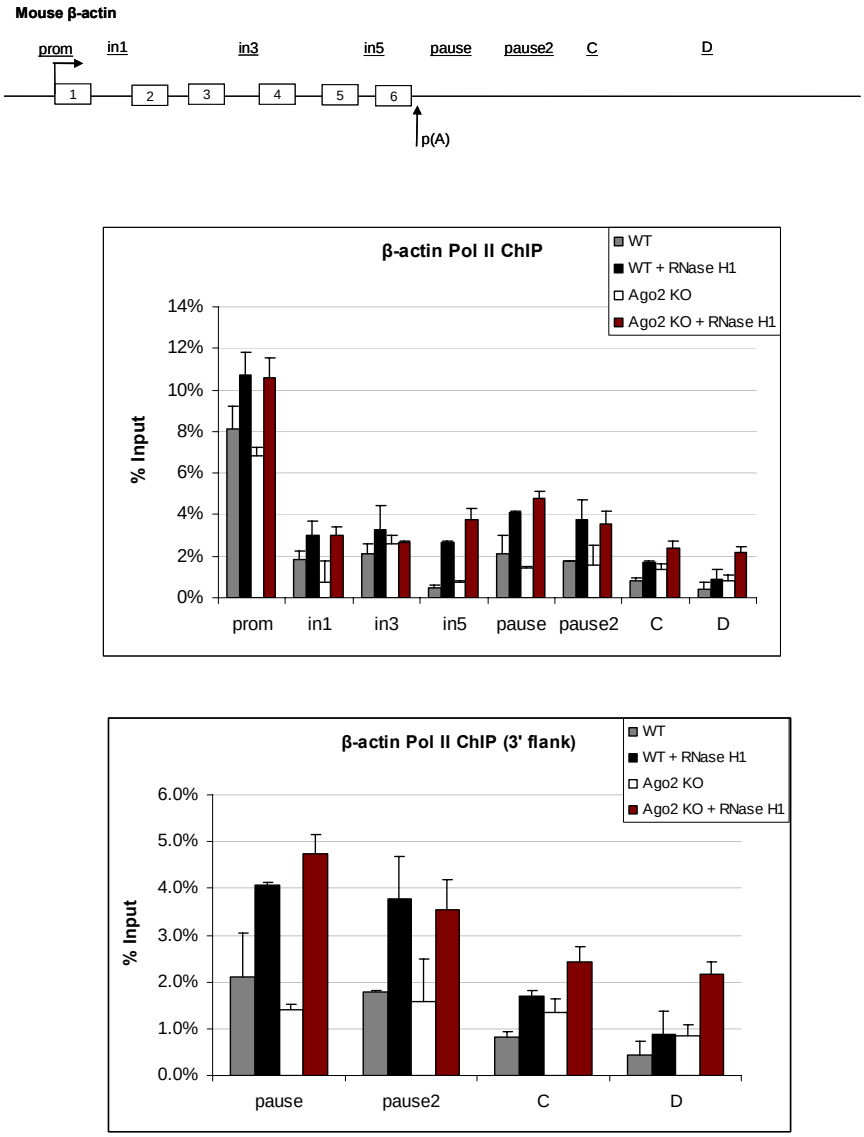
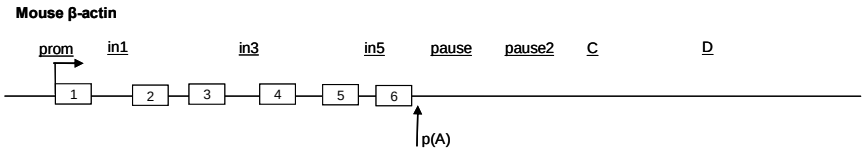


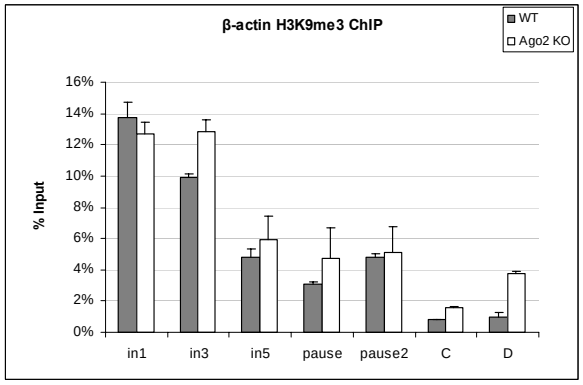
Figure 5.9 RNase H1 over-expression and Ago2 depletion affects transcriptional termination of endogenous mouse β-actin gene.

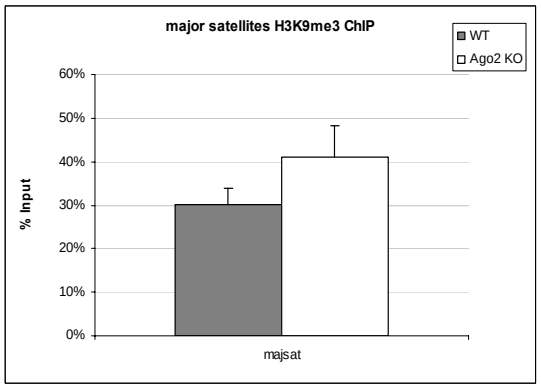
Pol II ChIP analysis of mouse β-actin in WT cells (grey bars), WT cells over-expressing 2 μg of RNaseH1-GFP plasmid (black bars), Ago2 KO cells (white bars) and Ago2 KO cells over-expressing RNaseH1-GFP plasmid (brown bars). Over-expression of RNase H1 in Ago2 KO cells causes increase in Pol II density particularly over C and D probes of mouse β-actin. Bottom panel shows probes downstream the poly(A) signal on more extended Y axis.

ChIP profile is based on average values +/- SD from four independent biological experiments.

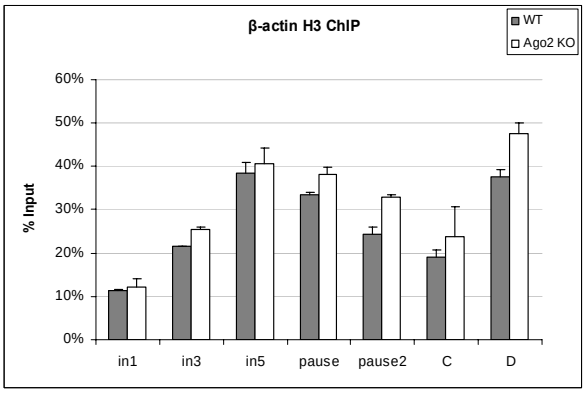
A

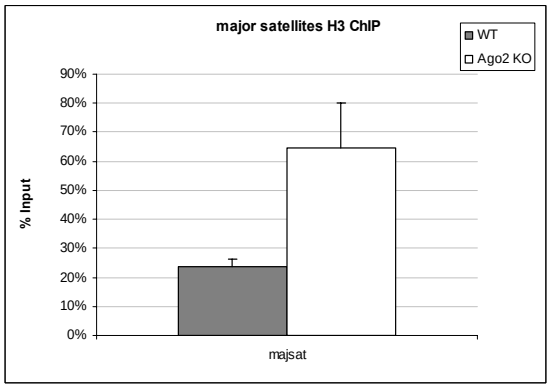

Mouse β-actin
prom
in1
in3
in5
pause
pause2
C
D
1
2
3
4
5
6
p(A)


β-actin H3K9me3 ChIP
WT
Ago2 KO
% Input
16%
14%
12%
10%
8%
6%
4%
2%
0%
in1
in3
in5
pause
pause2
C
D

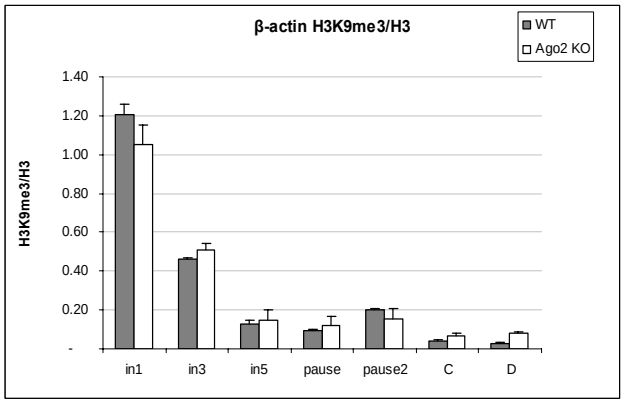

major satellites H3K9me3 ChIP
WT
Ago2 KO
% Input
60%
50%
40%
30%
20%
10%
0%
majsat

B


β-actin H3 ChIP
WT
Ago2 KO
% Input
60%
50%
40%
30%
20%
10%
0%
in1
in3
in5
pause
pause2
C
D


major satellites H3 ChIP
WT
Ago2 KO
% Input
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
majsat

C


β-actin H3K9me3/H3
WT
Ago2 KO
H3K9me3/H3
1.40
1.20
1.00
0.80
0.60
0.40
0.20
-
in1
in3
in5
pause
pause2
C
D

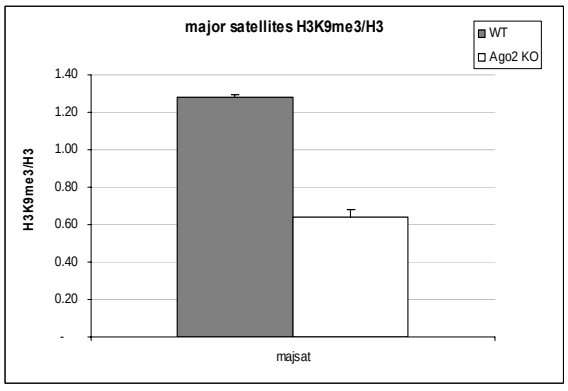

major satellites H3K9me3/H3
WT
Ago2 KO
H3K9me3/H3
1.40
1.20
1.00
0.80
0.60
0.40
0.20
-
majsat

Figure 5.10 The H3K9me3 mark is present at the beginning of endogenous mouse β-actin gene and is not affected in Ago2 KO cells.

A. Left panel: H3K9me3 ChIP analysis performed on endogenous mouse β-actin gene in WT and Ago2 KO cells. H3K9me3 forms over intron 1 of mouse β-actin at the highest level. In Ago2-depleted cells, H3K9me3 levels are not affected. Right panel: H3K9me3 ChIP analysis performed on the mouse major satellites in WT and Ago2 KO cells.

B. Left panel: H3 ChIP analysis of mouse β-actin in WT and Ago2 KO cells. Right panel: H3 ChIP analysis performed on major satellites.

C. Left panel: Ratio of H3K9me3 signal versus H3 signal on mouse β-actin in WT and Ago2 KO cells. According to the normalisation of H3K9me3 signal to total H3 signal, in WT cells the H3K9me3 mark peaks at the intron 1 of mouse β-actin. In Ago2 KO cells H3K9me3 levels remain unaffected. Right panel: Ratio of H3K9me3 signal versus H3 signal on major satellites in WT and Ago2 KO cells.

WT cells are represented as grey bars and Ago2 KO cells are represented as white bars. ChIP profiles are based on average values +/- SD from at least three independent biological experiments.

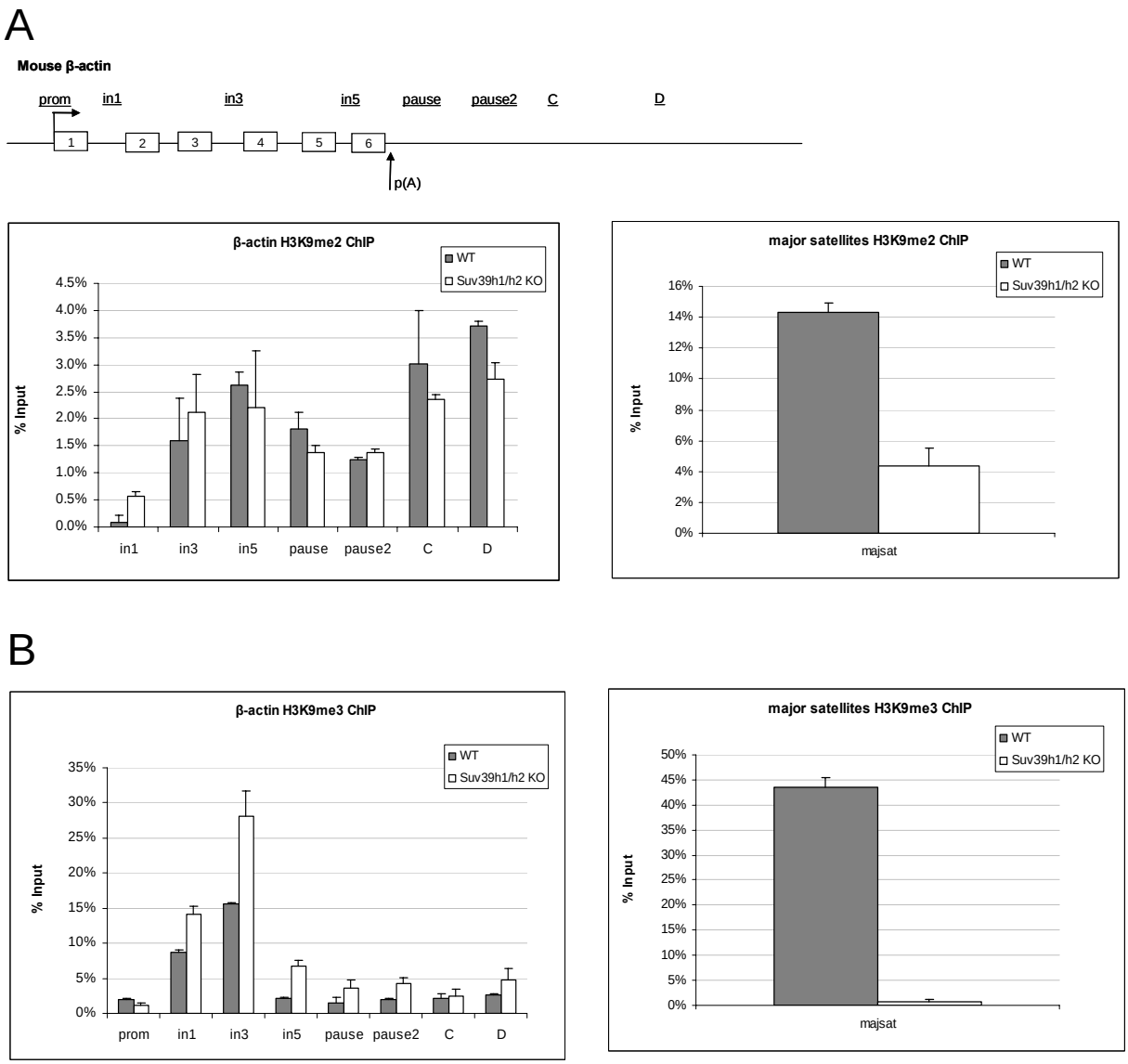


Figure 5.11 Suv39h1/h2 KO does not affect H3K9me2 and H3K9me3 levels of endogenous mouse β-actin gene.

A. Left panel: H3K9me2 ChIP analysis performed on endogenous mouse β-actin gene in WT and Suv39h1/h2 KO cells. H3K9me2 levels over the endogenous β-actin are not affected upon double knock-out of Suv39h methyltransferases. Right panel: H3K9me2 ChIP analysis performed on the mouse major satellites in WT and Suv39h1/h2 KO cells. H3K9me2 levels are reduced in Suv39h1/2 KO cells over mouse major satellites.

B. Left panel: H3K9me3 ChIP analysis of mouse β-actin gene in WT and Suv39h1/h2 KO cells. In endogenous β-actin, H3K9me3 levels over intron 1 and intron 3 are not reduced in Suv39h1/h2 KO cells. Right panel: H3K9me3 ChIP analysis of the mouse major satellites in WT and Suv39h1/h2 KO cells. Upon double knock-out of Suv39h, H3K9me3 levels over the major satellites are almost completely abolished.

WT cells are represented as grey bars and Suv39h1/h2 KO cells are represented as white bars. ChIP profiles are based on average values +/- SD from at least three independent biological experiments.

CHAPTER 6

General Discussion

Efficient Pol II termination of mammalian genes is dependent on both a functional poly(A) signal and downstream terminator sequences (Connelly and Manley, 1988; West et al. 2004; West et al. 2008; Whitelaw and Proudfoot, 1986). Two classes of terminator sequences have been identified in human genes: co-transcriptionally cleaved (CoTC) RNA sequences and transcription pause sites. The function of XRN2/Rat1 in the termination process has been studied extensively (Kaneko et al. 2007b; Kim et al. 2004; Luo et al. 2006; Xiang et al. 2009). Transcription pause sites act to slow down elongating Pol II and have been identified in α2- and γ-globin genes (Enriquez-Harris et al. 1991; Plant et al. 2005). Another pause site, defined as the G rich MAZ termination element, facilitates termination of the human C2 complement gene (Ashfield et al. 1994). The human β-actin gene also possesses a G-rich pause element downstream of the poly(A) signal which similarly acts to promote transcriptional termination, through Xrn2-mediated degradation of the downstream transcript exposed by poly(A) site cleavage (Gromak et al. 2006). Interestingly, genome-wide analysis revealed that G-rich sequences placed immediately downstream of poly(A) signals are relatively common in mammalian genes (Salisbury et al. 2006), suggesting that these elements may have an essential function in regulation of gene expression. However it still remains to be elucidated how G-rich pause sites facilitate termination by slowing down elongating Pol II.

In this thesis I have investigated the mechanism of pause-dependent transcriptional termination in mammalian genes. The three Result chapters combine to describe a molecular function for R-loop structures, senataxin protein and the H3K9me2 mark in the regulation of gene expression at the level of transcription.

Using ChIP, NRO and qRT-PCR approaches, I demonstrate that senataxin-deficient HeLa cells have transcription initiation/elongation and termination defects (Figures 3.1B, 3.2, 3.4C, 4.1B and 4.1C). In particular, I observe a decrease in nascent RNA and Pol II density over the gene body and an increase in the amount of read-through RNA and Pol II density downstream of the poly(A) signal in the termination regions. I hypothesized that the function of senataxin protein in human cells may be related to its RNA/DNA helicase activity, and in particular, its ability to unwind RNA/DNA hybrids formed between nascent RNA transcript and single stranded DNA template formed behind the elongating Pol II. Using DNA-immuno-precipitation (DIP) technique, I show that R-loops can be efficiently detected over transcribed regions of the endogenous β-actin gene and transfected gene constructs (Figures 3.3A and 4.1D). Furthermore, the RNA/DNA hybrid signature of R-loops was confirmed by RNase H treatment, specifically targeting their degradation (Figure 4.1D, bottom panel). Based on the transfected gene constructs, I show that formation of R-loops is dependent upon active transcription and the presence of a functional poly(A) signal and termination G-rich pause elements (Figures 3.3B and 3.3C). Interestingly, senataxin depletion caused accumulation of R-loops downstream the poly(A) signal, supporting its function in R-loops resolution in termination regions (Figure 3.3D). I also demonstrate that RNase H1 over-expression *in vivo* resolves R-loops, resulting in transcriptional read-through (Figures 3.4B, 3.4C and 3.4D). In effect R-loops formed over pause elements can be considered as an essential component of the termination process. But once they are formed, their removal by senataxin is also required. Furthermore, senataxin depletion severely diminished Xrn2 recruitment to the termination regions (Figure 4.1E).

It has been demonstrated that the senataxin mutations observed in ALS4/AOA2 patients cause defects in DNA damage repair (Chen et al. 2006; Suraweera et al. 2007). Significantly, deletion of Sen1 in *S. cerevisiae* causes transcription-associated DNA recombination defects, connected with transient accumulation of R-loop structures (Mischo et al. 2011). These findings suggest that R-loop formation is a frequent event during transcription and that Sen1 acts to prevent their accumulation and the associated genome instability events. However, it remains to be established whether the tendency of transcription to induce R-loop formation is a general feature of all eukaryotic genomes that requires a range of dedicated helicases to resolve these potentially harmful structures.

Recently it was shown that senataxin is involved in the process of splicing regulation (Suraweera et al. 2009). However, the molecular mechanism of this process is not clear, and might be related to the defect in Pol II elongation or defects in interaction with RNA splicing factors. Interestingly, previous studies have demonstrated that defects in splicing induced by the depletion of SR splicing factor ASF/SF2 correlated with the increase in R-loop levels, resulting in genome instability (Li and Manley, 2005). In addition, capping enzyme was shown to promote R-loop formation (Kaneko et al. 2007a). Mutations in senataxin protein cause AOA2/ALS4 neurodegenerative disorders. Even though senataxin is ubiquitously expressed, neuronal genes display high levels of alternative splicing. Possibly defects in senataxin function are more severely manifested in neuronal tissues. As compared to yeast Sen1, which has exclusively nuclear localization, senataxin is also detected in the cytoplasm, where it may have additional functions. Future structural studies on this protein and the analysis of its cytoplasmic roles, may further clarify the role of senataxin in human cells.

Overall results presented in Chapters 3 and 4 suggest that the formation of R-loops in transcriptional pause regions is essential to cause Pol II pausing downstream the poly(A) site prior to termination. However, these R-loops have to be subsequently resolved by senataxin to release the nascent RNA and so allow its Xrn2-mediated degradation, which ultimately results in efficient Pol II transcriptional termination. Based on Figures 4.3 and 4.4, the 3'-end processing factor Pcf11 may also be involved in this process. In essence using ChIP and DIP approaches I show that Pcf11 and senataxin 'communicate' directly or indirectly and that depletion of Pcf11 affects the R-loop distribution profile. It is not clear whether Pcf11 has a specific role in R-loop dependent transcriptional termination or whether depletion of 3' end processing factors alters the R-loop distribution profile. However, it can be speculated that nascent RNA, freed from R-loops by senataxin, may provide a binding platform for Pcf11 to hold Pol II prior to Xrn2-mediated termination. A possible sequence of events could be that in a gene that possesses a G-rich pause site a certain degree of R-loops need to be formed for Pol II to pause prior to efficient termination. As soon as Pol II pauses, R-loops have then to be resolved by senataxin for recruitment of Pcf11 at the termination region and subsequent transcript degradation by Xrn2.

An interesting question that arose during the course of my studies was how R-loops facilitate transcriptional pausing prior to their resolution by senataxin. Given the fact that Pol II molecule is a very large protein, it could be argued that hybrids between the nascent RNA and the template DNA strand are not enough to anchor Pol II. The connection between R-loops and heterochromatin formation over centromeric repeats in fission yeast (Nakama et al. 2012) raised the possibility that R-loops formed over the

pause elements of human genes could induce formation of repressive marks to facilitate pausing. In addition to this Allo and colleagues have shown that siRNAs directed against an intronic sequence could affect alternative splicing in human cells. This is mediated by H3K9me2 formation which could lead to Pol II pausing and subsequent alternative exon inclusion (Allo et al. 2009). Furthermore it was demonstrated recently that alternative splicing of the human *CD44* gene is regulated via H3K9me3 formation and HP1 recruitment (Saint-Andre et al. 2011). Indeed experiments presented in Chapter 5 reveal a link between R-loop structures formed over G-rich termination region of mammalian genes and the RNAi pathway. I observe specific enrichment of antisense transcription, Dicer protein, Ago1 protein and the H3K9me2 heterochromatin mark around the pause element of human β-actin gene (Figures 5.1C and 5.3). The sequence of events in this process is of particular interest. Do R-loops trigger H3K9me2 assembly or does the H3K9me2 formation lead to formation of R-loops over the termination region?

Even though one could argue that all steps in gene expression are inter-connected and therefore the exact sequence of events is not clear, based on results demonstrated on Figure 5.4, I could hypothesise that R-loops initially trigger the RNAi pathway and the subsequent H3K9me2 formation. In essence I show that over-expression of RNase H1 and subsequent reduction in R-loop signals causes a decrease in the amount of antisense RNA over the pause element and also reduced Dicer recruitment (Figures 5.4B and 5.4C). Therefore I suggest that formation of R-loops upstream the pause regions leaves the non-template DNA strand unprotected and prone to be the template for antisense transcription. While Pol II pauses downstream the poly(A) signal, another Pol II molecule could transcribe in the antisense orientation producing antisense RNA that

could hybridise with the sense transcript to form double-strand RNA. This event could then trigger Dicer recruitment and other RNAi factors and finally the formation of H3K9me2 mark. The observation that over-expression of RNase H1 does not affect the distribution profile of H3K9me2 mark over the termination region raises interesting questions (Figure 5.4A). Even though my results suggest that R-loops over the termination region facilitate antisense transcription and Dicer recruitment, I could speculate that H3K9me2 mark, once its initial R-loop mediated formation is then stable throughout cell divisions and its maintenance is not affected by R-loops resolution.

Another interesting point is that even though R-loops form at the beginning of the gene as well as the 3' end, the RNAi apparatus and the H3K9me2 is present only at the termination region. Furthermore RNase H1 over-expression does not affect antisense transcription observed over the promoter of human β-actin gene (Figure 5.4B). Based on these results I could hypothesise that R-loops formed over the termination regions of human genes may have a different function to the ones formed over the 5' ends and particularly facilitate pausing either on their own (possibly in the case of transfected gene constructs where over-expression of RNase H1 is enough to cause a termination defect) or in combination with the presence of H3K9me2 repressive mark.

In order to make use of murine gene knock-out cells I also investigated the conserved mouse β-actin gene. The main conclusions that I could draw from the experiments performed on this gene are that R-loops form over the termination region (Figure 5.8), the RNAi apparatus is also present over the same region (Figure 5.5) and H3K9me2, but not H3K9me3, is the mark that is specifically enriched over the termination region possibly playing a role in pause-dependent transcriptional termination

(Figures 5.6 and 5.10). Finally Ago2 depletion causes partial loss of H3K9me2 mark (Figure 5.6) and in combination with resolution of R-loops causes a termination defect in mouse β-actin gene (Figure 5.9). The latter result suggests that loss of R-loops and loss of H3K9me2 mark result in inefficient pausing and subsequent termination defect. Therefore it could be predicted that in a gene that terminates through the pause-dependent mechanism both R-loop structures and the H3K9me2 mark are needed for Pol II to pause prior to termination.

Gemone-wide data of Pol II profile and chromatin modifications of human β-actin are now publically available (Figure 6.1). According to these data, Pol II is specifically enriched over the promoter and gradually decreases towards the poly(A) signal. A second peak is detected over the pause element, whereas Pol II levels drop approximately 2 kb downstream of the poly(A) signal. This is consistent with the Pol II profile I observe over this gene (Figure 4.1A). Active histone marks (such as H3K4me3, H3K36me3 and H3K9ac) are highly associated with the promoter and ORF of β-actin gene. In the future it would be interesting to investigate the occupancy of such active histone marks and compare them with the H3K9me2 and H3K9me3 levels I observe over human and mouse β-actin (Figures 5.1, 5.6 and 5.10). Repressive histone marks (H3K9me3, H3K27me3 and H4K20me1) show generally lower enrichment over the human β-actin gene. Slight occupancy of H3K9me3 modification is detected over the 5' end of human β-actin (Figure 6.1), which is consistent with the results I have obtained for H3K9me3 formation over the mouse β-actin gene (Figure 5.10). The H3K9me2 ChIP-seq data are not yet available, but one could hypothesise that these levels would also be lower

than those of the active marks. One reason for this could be that the repressive histone marks are highly associated with the centromeres and other heterochromatic repetitive sequences and therefore the sequencing depth is not sufficient enough to capture specific enrichment in euchromatic regions (active genes).

Combining results from the three Result chapters I could suggest a model for efficient Pol II pause-dependent termination in mammalian genes (Figure 6.2). In a gene that possesses a G-rich pause element downstream the poly(A) signal, a certain degree of R-loops forms at the termination region. This facilitates the generation of an antisense transcript over the pause element. Given that two Pol II molecules could collide if they transcribed at the same time in opposite orientations and that the dsDNA enters Pol II through one unique channel, the antisense transcript could be generated possibly either by another Pol II molecule which utilises the unprotected DNA single strand as its template at a different time point or by the same Pol II molecule which transcribes in the opposite orientation. Antisense RNA could then hybridise with the sense transcript to form double-strand RNA which is processed by the ribonuclease Dicer. Next the RNAi apparatus is recruited leading to H3K9me2 formation over the termination region. R-loop structures and the H3K9me2 mark combine to facilitate Pol II pausing prior to termination. Once Pol II slows down, senataxin will resolve the R-loops, allowing Pcf11 binding to the termination region. Finally Xrn2 is recruited to degrade the released nascent transcript, which ultimately results in efficient Pol II transcriptional termination.

A final point that arises is the potential generality of this mechanism. So far I have shown that R-loops and H3K9me2 are essential components for the termination of human and mouse β-actin genes. It should be noted at this point that in human γ-actin gene I also detect R-loops and the H3K9me2 mark over the termination region (Figures 4.5D and 5.2C) even though senataxin is not involved in transcriptional termination of this gene (Figure 4.5C). A possible explanation for this observation could be that another helicase resolves R-loops over the termination region of human γ-actin gene to promote efficient Pol II termination. Given the fact that G-rich sequences downstream the poly(A) signal are common in mammalian genes (Salisbury et al. 2006), in the future it will be important to investigate whether R-loops are particularly enriched at the G-rich 3' ends of mammalian genes, similar to their 5' end enrichment observed recently (Ginno et al. 2012). In addition, it will be important to investigate whether the H3K9me2 mark is also generally present at G-rich termination regions of mammalian genes. Finally it will be interesting to study at a genome-wide level whether the presence of H3K9me2 mark correlates with the formation of R-loops over the termination regions. This could point towards a general mechanism of pausing mediated by R-loop dependent H3K9me2 formation resulting in efficient Pol II transcriptional termination of mammalian genes.

Transcription direction

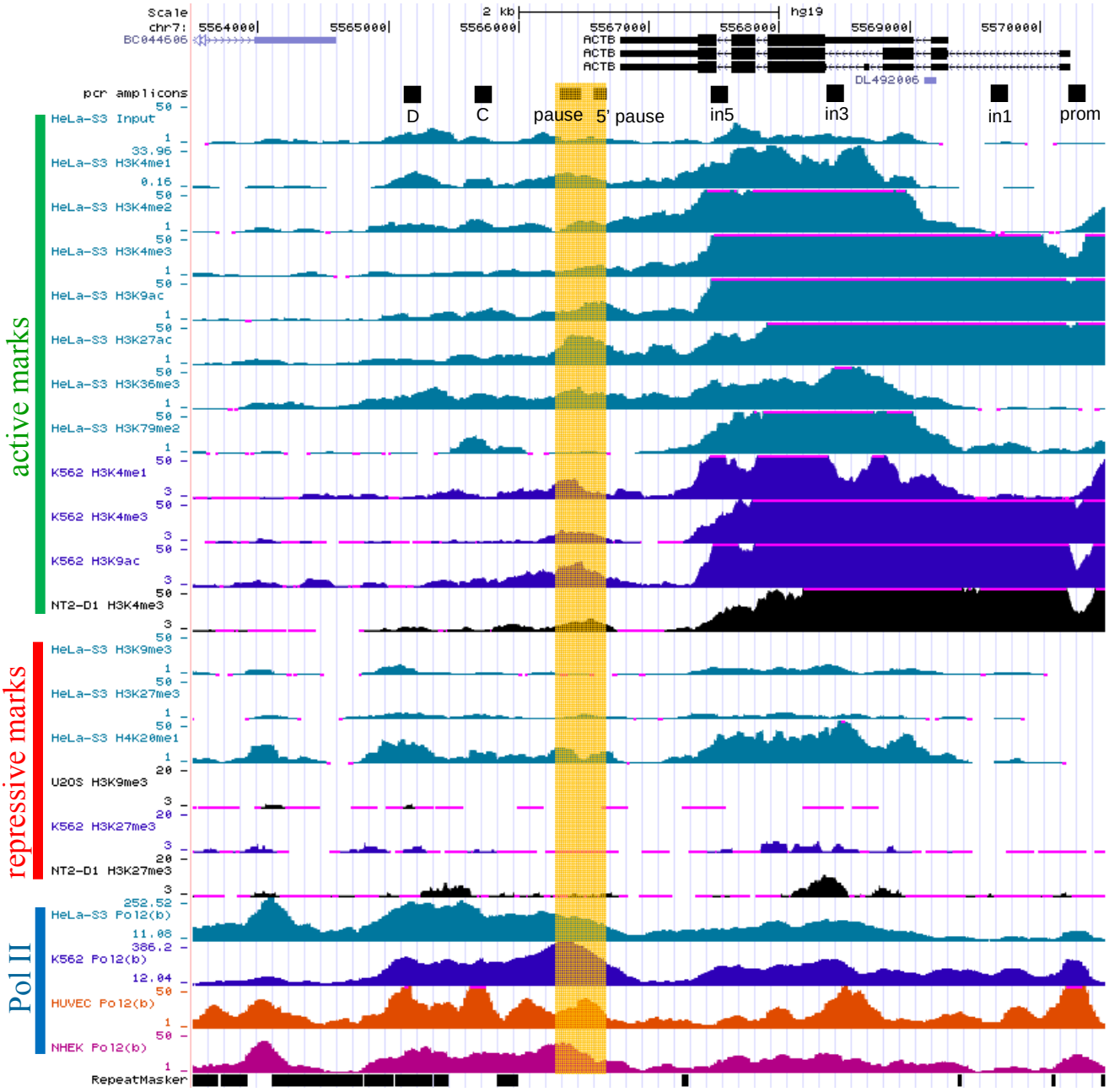
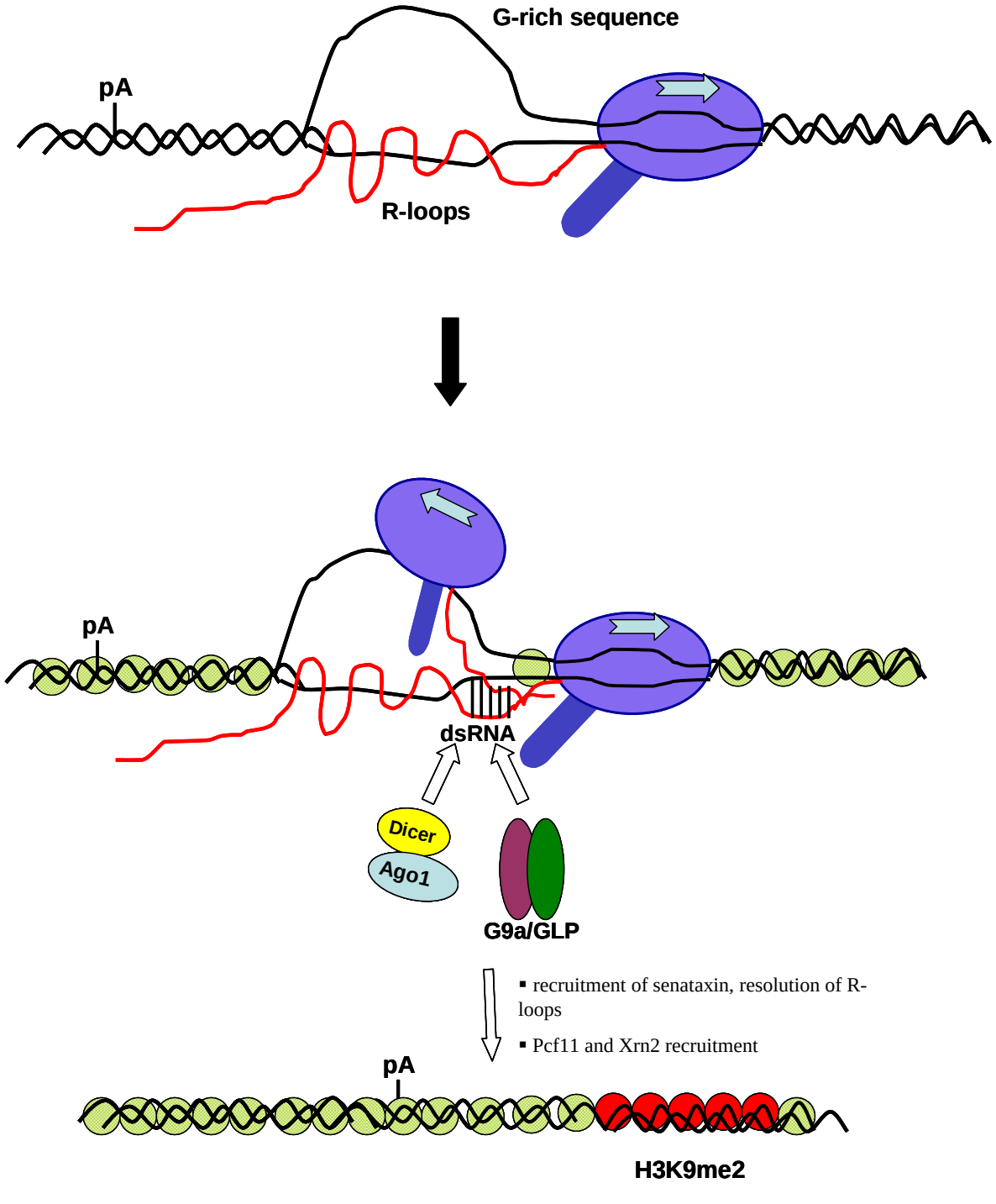


Figure 6.1 Genome-wide data of chromatin modifications and Pol II distribution of human β-actin gene.

ChIP-seq datasets of various human cell lines showing distribution of histone marks and Pol II on the β-actin gene (ENCODE project). Direction of transcription is indicated by a red arrow. Positions of PCR amplicons used in ChIP experiments are highlighted. Histone modifications associated with active transcription are indicated in green, whereas histone marks associated with trascriptional repression are indicated in red. Pol II profiles are indicated in blue.

Active histone marks are highly associated with the promoter and ORF of β-actin gene. Repressive histone marks (H3K9me3, H3K27me3 and H4K20me1) show generally lower enrichment over the human β-actin gene. Pol II is specifically enriched over the promoter and gradually decreases towards the poly(A) signal. A second peak is detected over the pause element. Pol II levels drop approximately 2 kb downstream of the poly(A) signal.



Pol II release-transcriptional termination

Figure 6.2 General model: Role of R-loops and RNAi-dependent H3K9me2 formation in pause-dependent transcriptional termination of mammalian genes.
In a mammalian gene that possesses a G-rich pause element downstream of the poly(A) signal, a certain degree of R-loops forms at the termination region. This facilitates the generation of an antisense transcript either by the same Pol II molecule which transcribes in the opposite orientation or by another Pol II molecule transcribing at a different time point. The antisense transcript can then hybridise with the sense to form dsRNA. This triggers the recruitment of the RNAi factors, Dicer and Ago1. The G9a/GLP methyltransferases are then recruited to catalyse the formation of the H3K9me2 repressive mark. R-loops and H3K9me2 mark facilitate Pol II pausing prior to termination. Once Pol II slows down, senataxin resolves the R-loops allowing Pcf11 recruitment to the termination region. Finally, Xrn2 degrades the released nascent transcript which ultimately results in efficient transcriptional termination.

CHAPTER 7

References

Aguilera, A. 2002. The connection between transcription and genomic instability. *EMBO J* **21**(3): 195-201.

-. 2005. mRNA processing and genomic instability. *Nat Struct Mol Biol* **12**(9): 737-738.

Aguilera, A. and Garcia-Muse, T. 2012. R loops: from transcription byproducts to threats to genome stability. *Mol Cell* **46**(2): 115-124.

Aguilera, A. and Klein, H.L. 1990. HPR1, a novel yeast gene that prevents intrachromosomal excision recombination, shows carboxy-terminal homology to the Saccharomyces cerevisiae TOP1 gene. *Mol Cell Biol* **10**(4): 1439-1451.

Ahn, S.H., Kim, M., and Buratowski, S. 2004. Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Mol Cell* **13**(1): 67-76.

Akhtar, M.S., Heidemann, M., Tietjen, J.R., Zhang, D.W., Chapman, R.D., Eick, D., and Ansari, A.Z. 2009. TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Mol Cell* **34**(3): 387-393.

Allo, M., Buggiano, V., Fededa, J.P., Petrillo, E., Schor, I., de la Mata, M., Agirre, E., Plass, M., Eyras, E., Elela, S.A., Klinck, R., Chabot, B., and Kornblihtt, A.R. 2009. Control of alternative splicing through siRNA-mediated transcriptional gene silencing. *Nat Struct Mol Biol* **16**(7): 717-724.

Ashe, H.L., Monks, J., Wijgerde, M., Fraser, P., and Proudfoot, N.J. 1997. Intergenic transcription and transinduction of the human beta-globin locus. *Genes Dev* **11**(19): 2494-2509.

Ashfield, R., Patel, A.J., Bossone, S.A., Brown, H., Campbell, R.D., Marcu, K.B., and Proudfoot, N.J. 1994. MAZ-dependent termination between closely spaced human complement genes. *EMBO J* **13**(23): 5656-5667.

Baker, T.A. and Kornberg, A. 1988. Transcriptional activation of initiation of replication from the E. coli chromosomal origin: an RNA-DNA hybrid near oriC. *Cell* **55**(1): 113-123.

Barilla, D., Lee, B.A., and Proudfoot, N.J. 2001. Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **98**(2): 445-450.

Bartel, D.P. 2009. MicroRNAs: target recognition and regulatory functions. *Cell* **136**(2): 215-233.

Bataille, A.R., Jeronimo, C., Jacques, P.E., Laramee, L., Fortin, M.E., Forest, A., Bergeron, M., Hanes, S.D., and 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**(2): 158-170.

Bayne, E.H., White, S.A., Kagansky, A., Bijos, D.A., Sanchez-Pulido, L., Hoe, K.L., Kim, D.U., Park, H.O., Ponting, C.P., Rappsilber, J., and Allshire, R.C. 2010. Stc1: a critical link between RNAi and chromatin modification required for heterochromatin integrity. *Cell* **140**(5): 666-677.

Berger, S.L. 2007. The complex language of chromatin regulation during transcription. *Nature* **447**(7143): 407-412.

Bernstein, E. and Allis, C.D. 2005. RNA meets chromatin. *Genes Dev* **19**(14): 1635-1655.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**(6818): 363-366.

Billy, E., Brondani, V., Zhang, H., Muller, U., and Filipowicz, W. 2001. Specific interference with gene expression induced by long, double-stranded RNA in mouse embryonal teratocarcinoma cell lines. *Proc Natl Acad Sci U S A* **98**(25): 14428-14433.

Birse, C.E., Minvielle-Sebastia, L., Lee, B.A., Keller, W., and Proudfoot, N.J. 1998. Coupling termination of transcription to messenger RNA maturation in yeast. *Science* **280**(5361): 298-301.

Boeing, S., Rigault, C., Heidemann, M., Eick, D., and Meisterernst, M. 2010. RNA polymerase II C-terminal heptarepeat domain Ser-7 phosphorylation is established in a mediator-dependent fashion. *J Biol Chem* **285**(1): 188-196.

Boguslawski, S.J., Smith, D.E., Michalak, M.A., Mickelson, K.E., Yehle, C.O., Patterson, W.L., and Carrico, R.J. 1986. Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. *J Immunol Methods* **89**(1): 123-130.

Boubakri, H., de Septenville, A.L., Viguera, E., and Michel, B. 2010. The helicases DinG, Rep and UvrD cooperate to promote replication across transcription units in vivo. *EMBO J* **29**(1): 145-157.

Buhler, M. and Moazed, D. 2007. Transcription and RNAi in heterochromatic gene silencing. *Nat Struct Mol Biol* **14**(11): 1041-1048.

Buratowski, S. 2003. The CTD code. *Nat Struct Biol* **10**(9): 679-680.

-. 2005. Connections between mRNA 3' end processing and transcription termination. *Curr Opin Cell Biol* **17**(3): 257-261.

-. 2009. Progression through the RNA polymerase II CTD cycle. *Mol Cell* **36**(4): 541-546.

Cam, H.P., Chen, E.S., and Grewal, S.I. 2009. Transcriptional scaffolds for heterochromatin assembly. *Cell* **136**(4): 610-614.

Cam, H.P., Sugiyama, T., Chen, E.S., Chen, X., FitzGerald, P.C., and Grewal, S.I. 2005. Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. *Nat Genet* **37**(8): 809-819.

Carles-Kinch, K. and Kreuzer, K.N. 1997. RNA-DNA hybrid formation at a bacteriophage T4 replication origin. *J Mol Biol* **266**(5): 915-926.

Carlile, M., Swan, D., Jackson, K., Preston-Fayers, K., Ballester, B., Flicek, P., and Werner, A. 2009. Strand selective generation of endo-siRNAs from the Na/phosphate transporter gene Slc34a1 in murine tissues. *Nucleic Acids Res* **37**(7): 2274-2282.

Carthew, R.W. and Sontheimer, E.J. 2009. Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136**(4): 642-655.

Cerritelli, S.M. and Crouch, R.J. 2009. Ribonuclease H: the enzymes in eukaryotes. *FEBS J* **276**(6): 1494-1505.

Cerritelli, S.M., Frolova, E.G., Feng, C., Grinberg, A., Love, P.E., and Crouch, R.J. 2003. Failure to produce mitochondrial DNA results in embryonic lethality in Rnaseh1 null mice. *Mol Cell* **11**(3): 807-815.

Chapman, R.D., Heidemann, M., Albert, T.K., Mailhammer, R., Flatley, A., Meisterernst, M., Kremmer, E., and Eick, D. 2007. Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science* **318**(5857): 1780-1782.

Chavez, S., Beilharz, T., Rondon, A.G., Erdjument-Bromage, H., Tempst, P., Svejstrup, J.Q., Lithgow, T., and Aguilera, A. 2000. A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in Saccharomyces cerevisiae. *EMBO J* **19**(21): 5824-5834.

Chen, Y.Z., Bennett, C.L., Huynh, H.M., Blair, I.P., Puls, I., Irobi, J., Dierick, I., Abel, A., Kennerson, M.L., Rabin, B.A., Nicholson, G.A., Auer-Grumbach, M., Wagner, K., De Jonghe, P., Griffin, J.W., Fischbeck, K.H., Timmerman, V., Cornblath, D.R., and Chance, P.F. 2004. DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). *Am J Hum Genet* **74**(6): 1128-1135.

Chen, Y.Z., Hashemi, S.H., Anderson, S.K., Huang, Y., Moreira, M.C., Lynch, D.R., Glass, I.A., Chance, P.F., and Bennett, C.L. 2006. Senataxin, the yeast Sen1p orthologue: characterization of a unique protein in which recessive mutations cause ataxia and dominant mutations cause motor neuron disease. *Neurobiol Dis* **23**(1): 97-108.

Cheng, B. and Price, D.H. 2007. Properties of RNA polymerase II elongation complexes before and after the P-TEFb-mediated transition into productive elongation. *J Biol Chem* **282**(30): 21901-21912.

Chinchilla, K., Rodriguez-Molina, J.B., Ursic, D., Finkel, J.S., Ansari, A.Z., and Culbertson, M.R. 2012. Interactions of Sen1, Nrd1, and Nab3 with multiple phosphorylated forms of the Rpb1 C-terminal domain in Saccharomyces cerevisiae. *Eukaryot Cell* **11**(4): 417-429.

Cho, E.J., Kobor, M.S., Kim, M., Greenblatt, J., and 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**(24): 3319-3329.

Christofori, G. and Keller, W. 1988. 3' cleavage and polyadenylation of mRNA precursors in vitro requires a poly(A) polymerase, a cleavage factor, and a snRNP. *Cell* **54**(6): 875-889.

Clemens, J.C., Worby, C.A., Simonson-Leff, N., Muda, M., Maehama, T., Hemmings, B.A., and Dixon, J.E. 2000. Use of double-stranded RNA interference in Drosophila cell lines to dissect signal transduction pathways. *Proc Natl Acad Sci U S A* **97**(12): 6499-6503.

Colgan, D.F. and Manley, J.L. 1997. Mechanism and regulation of mRNA polyadenylation. *Genes Dev* **11**(21): 2755-2766.

Collins, R.E., Tachibana, M., Tamaru, H., Smith, K.M., Jia, D., Zhang, X., Selker, E.U., Shinkai, Y., and Cheng, X. 2005. In vitro and in vivo analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases. *J Biol Chem* **280**(7): 5563-5570.

Colmenares, S.U., Buker, S.M., Buhler, M., Dlakic, M., and Moazed, D. 2007. Coupling of double-stranded RNA synthesis and siRNA generation in fission yeast RNAi. *Mol Cell* **27**(3): 449-461.

Connelly, S. and Manley, J.L. 1988. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev* **2**(4): 440-452.

Core, L.J., Waterfall, J.J., and Lis, J.T. 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**(5909): 1845-1848.

Cramer, P., Armache, K.J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G.E., Dengl, S., Geiger, S.R., Jasiak, A.J., Jawhari, A., Jennebach, S., Kamenski, T., Kettenberger, H., Kuhn, C.D., Lehmann, E., Leike, K., Sydow, J.F., and Vannini, A. 2008. Structure of eukaryotic RNA polymerases. *Annu Rev Biophys* **37**: 337-352.

Criscuolo, C., Chessa, L., Di Giandomenico, S., Mancini, P., Sacca, F., Grieco, G.S., Piane, M., Barbieri, F., De Michele, G., Banfi, S., Pierelli, F., Rizzuto, N., Santorelli, F.M., Gallosti, L., Filla, A., and Casali, C. 2006. Ataxia with oculomotor apraxia type 2: a clinical, pathologic, and genetic study. *Neurology* **66**(8): 1207-1210.

Cui, M., Allen, M.A., Larsen, A., Macmorris, M., Han, M., and Blumenthal, T. 2008. Genes involved in pre-mRNA 3'-end formation and transcription termination revealed by a lin-15 operon Muv suppressor screen. *Proc Natl Acad Sci U S A* **105**(43): 16665-16670.

Danckwardt, S., Kaufmann, I., Gentzel, M., Foerstner, K.U., Gantzert, A.S., Gehring, N.H., Neu-Yilik, G., Bork, P., Keller, W., Wilm, M., Hentze, M.W., and Kulozik, A.E. 2007. Splicing factors stimulate polyadenylation via USEs at non-canonical 3' end formation signals. *EMBO J* **26**(11): 2658-2669.

Darzacq, X., Shav-Tal, Y., de Turris, V., Brody, Y., Shenoy, S.M., Phair, R.D., and Singer, R.H. 2007. In vivo dynamics of RNA polymerase II transcription. *Nat Struct Mol Biol* **14**(9): 796-806.

de Vries, H., Ruegsegger, U., Hubner, W., Friedlein, A., Langen, H., and Keller, W. 2000. Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors. *EMBO J* **19**(21): 5895-5904.

Di Noia, J. and Neuberger, M.S. 2002. Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature* **419**(6902): 43-48.

Dichtl, B., Blank, D., Sadowski, M., Hubner, W., Weiser, S., and Keller, W. 2002. Yhh1p/Cft1p directly links poly(A) site recognition and RNA polymerase II transcription termination. *EMBO J* **21**(15): 4125-4135.

Djupedal, I., Kos-Braun, I.C., Mosher, R.A., Soderholm, N., Simmer, F., Hardcastle, T.J., Fender, A., Heidrich, N., Kagansky, A., Bayne, E., Wagner, E.G., Baulcombe, D.C., Allshire, R.C., and Ekwall, K. 2009. Analysis of small RNA in fission yeast; centromeric siRNAs are potentially generated through a structured RNA. *EMBO J* **28**(24): 3832-3844.

Djupedal, I., Portoso, M., Spahr, H., Bonilla, C., Gustafsson, C.M., Allshire, R.C., and Ekwall, K. 2005. RNA Pol II subunit Rpb7 promotes centromeric transcription and RNAi-directed chromatin silencing. *Genes Dev* **19**(19): 2301-2306.

Dominguez-Sanchez, M.S., Barroso, S., Gomez-Gonzalez, B., Luna, R., and Aguilera, A. 2011. Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. *PLoS Genet* **7**(12): e1002386.

Drolet, M., Phoenix, P., Menzel, R., Masse, E., Liu, L.F., and Crouch, R.J. 1995. Overexpression of RNase H partially complements the growth defect of an Escherichia coli delta topA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I. *Proc Natl Acad Sci U S A* **92**(8): 3526-3530.

Duquette, A., Roddier, K., McNabb-Baltar, J., Gosselin, I., St-Denis, A., Dicaire, M.J., Loisel, L., Labuda, D., Marchand, L., Mathieu, J., Bouchard, J.P., and Brais, B. 2005. Mutations in senataxin responsible for Quebec cluster of ataxia with neuropathy. *Ann Neurol* **57**(3): 408-414.

Duquette, M.L., Handa, P., Vincent, J.A., Taylor, A.F., and Maizels, N. 2004. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev* **18**(13): 1618-1629.

Dye, M.J. and Proudfoot, N.J. 1999. Terminal exon definition occurs cotranscriptionally and promotes termination of RNA polymerase II. *Mol Cell* **3**(3): 371-378.

-. 2001. Multiple transcript cleavage precedes polymerase release in termination by RNA polymerase II. *Cell* **105**(5): 669-681.

Egloff, S. and Murphy, S. 2008. Cracking the RNA polymerase II CTD code. *Trends Genet* **24**(6): 280-288.

Egloff, S., O'Reilly, D., Chapman, R.D., Taylor, A., Tanzhaus, K., Pitts, L., Eick, D., and Murphy, S. 2007. Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science* **318**(5857): 1777-1779.

Egloff, S., Zaborowska, J., Laitem, C., Kiss, T., and Murphy, S. 2012. Ser7 phosphorylation of the CTD recruits the RPAP2 Ser5 phosphatase to snRNA genes. *Mol Cell* **45**(1): 111-122.

El Hage, A., French, S.L., Beyer, A.L., and Tollervey, D. 2011. Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev* **24**(14): 1546-1558.

Emili, A., Shales, M., McCracken, S., Xie, W., Tucker, P.W., Kobayashi, R., Blencowe, B.J., and Ingles, C.J. 2002. Splicing and transcription-associated proteins PSF and p54nrb/nonO bind to the RNA polymerase II CTD. *RNA* **8**(9): 1102-1111.

Enriquez-Harris, P., Levitt, N., Briggs, D., and Proudfoot, N.J. 1991. A pause site for RNA polymerase II is associated with termination of transcription. *EMBO J* **10**(7): 1833-1842.

Faghihi, M.A. and Wahlestedt, C. 2009. Regulatory roles of natural antisense transcripts. *Nat Rev Mol Cell Biol* **10**(9): 637-643.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. 1998. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**(6669): 806-811.

Fodor, B.D., Shukeir, N., Reuter, G., and Jenuwein, T. 2010. Mammalian Su(var) genes in chromatin control. *Annu Rev Cell Dev Biol* **26**: 471-501.

Fogel, B.L. and Perlman, S. 2006. Novel mutations in the senataxin DNA/RNA helicase domain in ataxia with oculomotor apraxia 2. *Neurology* **67**(11): 2083-2084.

Fujinaga, K., Irwin, D., Huang, Y., Taube, R., Kurosu, T., and Peterlin, B.M. 2004. Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. *Mol Cell Biol* **24**(2): 787-795.

Gan, W., Guan, Z., Liu, J., Gui, T., Shen, K., Manley, J.L., and Li, X. 2011. R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev* **25**(19): 2041-2056.

Ganem, C., Devaux, F., Torchet, C., Jacq, C., Quevillon-Cheruel, S., Labesse, G., Facca, C., and Faye, G. 2003. Ssu72 is a phosphatase essential for transcription termination of snoRNAs and specific mRNAs in yeast. *EMBO J* **22**(7): 1588-1598.

Ghosh, A., Shuman, S., and Lima, C.D. 2008. The structure of Fcp1, an essential RNA polymerase II CTD phosphatase. *Mol Cell* **32**(4): 478-490.

Gil, A. and Proudfoot, N.J. 1984. A sequence downstream of AAUAAA is required for rabbit beta-globin mRNA 3'-end formation. *Nature* **312**(5993): 473-474.

-. 1987. Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit beta-globin mRNA 3' end formation. *Cell* **49**(3): 399-406.

Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I., and Chedin, F. 2012. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell* **45**(6): 814-825.

Glover-Cutter, K., Larochelle, S., Erickson, B., Zhang, C., Shokat, K., Fisher, R.P., and 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**(20): 5455-5464.

Gottipati, P., Cassel, T.N., Savolainen, L., and Helleday, T. 2008. Transcription-associated recombination is dependent on replication in Mammalian cells. *Mol Cell Biol* **28**(1): 154-164.

Grewal, S.I. 2010. RNAi-dependent formation of heterochromatin and its diverse functions. *Curr Opin Genet Dev* **20**(2): 134-141.

Grewal, S.I. and Jia, S. 2007. Heterochromatin revisited. *Nat Rev Genet* **8**(1): 35-46.

Grohmann, K., Schuelke, M., Diers, A., Hoffmann, K., Lucke, B., Adams, C., Bertini, E., Leonhardt-Horti, H., Muntoni, F., Ouvrier, R., Pfeufer, A., Rossi, R., Van Maldergem, L., Wilmshurst, J.M., Wienker, T.F., Sendtner, M., Rudnik-Schoneborn, S., Zerres, K., and Hubner, C. 2001. Mutations in the gene encoding immunoglobulin mu-binding protein 2 cause spinal muscular atrophy with respiratory distress type 1. *Nat Genet* **29**(1): 75-77.

Gromak, N., West, S., and Proudfoot, N.J. 2006. Pause sites promote transcriptional termination of mammalian RNA polymerase II. *Mol Cell Biol* **26**(10): 3986-3996.

Gross, S. and Moore, C. 2001. Five subunits are required for reconstitution of the cleavage and polyadenylation activities of Saccharomyces cerevisiae cleavage factor I. *Proc Natl Acad Sci U S A* **98**(11): 6080-6085.

Gu, S.G., Pak, J., Guang, S., Maniar, J.M., Kennedy, S., and Fire, A. 2012. Amplification of siRNA in Caenorhabditis elegans generates a transgenerational sequence-targeted histone H3 lysine 9 methylation footprint. *Nat Genet* **44**(2): 157-164.

Gullerova, M. and Proudfoot, N.J. 2008. Cohesin complex promotes transcriptional termination between convergent genes in S. pombe. *Cell* **132**(6): 983-995.

Hall, I.M., Shankaranarayana, G.D., Noma, K., Ayoub, N., Cohen, A., and Grewal, S.I. 2002. Establishment and maintenance of a heterochromatin domain. *Science* **297**(5590): 2232-2237.

Hampsey, M. and Reinberg, D. 2003. Tails of intrigue: phosphorylation of RNA polymerase II mediates histone methylation. *Cell* **113**(4): 429-432.
Hannon, G.J. 2002. RNA interference. *Nature* **418**(6894): 244-251.
Hausen, P. and Stein, H. 1970. Ribonuclease H. An enzyme degrading the RNA moiety of DNA-RNA hybrids. *Eur J Biochem* **14**(2): 278-283.
Hausmann, S. and Shuman, S. 2002. Characterization of the CTD phosphatase Fcp1 from fission yeast. Preferential dephosphorylation of serine 2 versus serine 5. *J Biol Chem* **277**(24): 21213-21220.
Hirose, Y. and Manley, J.L. 1998. RNA polymerase II is an essential mRNA polyadenylation factor. *Nature* **395**(6697): 93-96.
-. 2000. RNA polymerase II and the integration of nuclear events. *Genes Dev* **14**(12): 1415-1429.
Hofmann, I., Schnolzer, M., Kaufmann, I., and Franke, W.W. 2002. Symplekin, a constitutive protein of karyo- and cytoplasmic particles involved in mRNA biogenesis in Xenopus laevis oocytes. *Mol Biol Cell* **13**(5): 1665-1676.
Houlard, M., Artus, J., Leguillier, T., Vandormael-Pournin, S., and Cohen-Tannoudji, M. 2011. DNA-RNA hybrids contribute to the replication dependent genomic instability induced by Omcg1 deficiency. *Cell Cycle* **10**(1): 108-117.
Huertas, P. and Aguilera, A. 2003. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol Cell* **12**(3): 711-721.
Huisinga, K.L., Brower-Toland, B., and Elgin, S.C. 2006. The contradictory definitions of heterochromatin: transcription and silencing. *Chromosoma* **115**(2): 110-122.
Huppert, J.L., Bugaut, A., Kumari, S., and Balasubramanian, S. 2008. G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res* **36**(19): 6260-6268.
Hutvagner, G. and Simard, M.J. 2008. Argonaute proteins: key players in RNA silencing. *Nat Rev Mol Cell Biol* **9**(1): 22-32.
James, P.A. and Talbot, K. 2006. The molecular genetics of non-ALS motor neuron diseases. *Biochim Biophys Acta* **1762**(11-12): 986-1000.
Janowski, B.A., Huffman, K.E., Schwartz, J.C., Ram, R., Nordsell, R., Shames, D.S., Minna, J.D., and Corey, D.R. 2006. Involvement of AGO1 and AGO2 in mammalian transcriptional silencing. *Nat Struct Mol Biol* **13**(9): 787-792.
Jenuwein, T. 2002. Molecular biology. An RNA-guided pathway for the epigenome. *Science* **297**(5590): 2215-2218.
-. 2006. The epigenetic magic of histone lysine methylation. *FEBS J* **273**(14): 3121-3135.
Jenuwein, T., Laible, G., Dorn, R., and Reuter, G. 1998. SET domain proteins modulate chromatin domains in eu- and heterochromatin. *Cell Mol Life Sci* **54**(1): 80-93.
Jimeno-Gonzalez, S., Haaning, L.L., Malagon, F., and Jensen, T.H. The yeast 5'-3' exonuclease Rat1p functions during transcription elongation by RNA polymerase II. *Mol Cell* **37**(4): 580-587.
Kaneko, S., Chu, C., Shatkin, A.J., and Manley, J.L. 2007a. Human capping enzyme promotes formation of transcriptional R loops in vitro. *Proc Natl Acad Sci U S A* **104**(45): 17620-17625.
Kaneko, S., Rozenblatt-Rosen, O., Meyerson, M., and Manley, J.L. 2007b. The multifunctional protein p54nrb/PSF recruits the exonuclease XRN2 to facilitate

pre-mRNA 3' processing and transcription termination. *Genes Dev* **21**(14): 1779-1789.

Kapranov, P., Willingham, A.T., and Gingeras, T.R. 2007. Genome-wide transcription and the implications for genomic organization. *Nat Rev Genet* **8**(6): 413-423.

Kato, H., Goto, D.B., Martienssen, R.A., Urano, T., Furukawa, K., and Murakami, Y. 2005. RNA polymerase II is required for RNAi-dependent heterochromatin assembly. *Science* **309**(5733): 467-469.

Kaufmann, I., Martin, G., Friedlein, A., Langen, H., and Keller, W. 2004. Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. *EMBO J* **23**(3): 616-626.

Kawasaki, H. and Taira, K. 2004. Induction of DNA methylation and gene silencing by short interfering RNAs in human cells. *Nature* **431**(7005): 211-217.

Kawauchi, J., Mischo, H., Braglia, P., Rondon, A., and Proudfoot, N.J. 2008. Budding yeast RNA polymerases I and II employ parallel mechanisms of transcriptional termination. *Genes Dev* **22**(8): 1082-1092.

Keller, W., Bienroth, S., Lang, K.M., and Christofori, G. 1991. Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3' processing signal AAUAAA. *EMBO J* **10**(13): 4241-4249.

Kennerdell, J.R. and Carthew, R.W. 1998. Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. *Cell* **95**(7): 1017-1026.

Kessler, M.M., Zhao, J., and Moore, C.L. 1996. Purification of the Saccharomyces cerevisiae cleavage/polyadenylation factor I. Separation into two components that are required for both cleavage and polyadenylation of mRNA 3' ends. *J Biol Chem* **271**(43): 27167-27175.

Kim, D.H., Villeneuve, L.M., Morris, K.V., and Rossi, J.J. 2006. Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. *Nat Struct Mol Biol* **13**(9): 793-797.

Kim, H., Erickson, B., Luo, W., Seward, D., Graber, J.H., Pollock, D.D., Megee, P.C., and Bentley, D.L. 2010. Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nat Struct Mol Biol* **17**(10): 1279-1286.

Kim, M., Krogan, N.J., Vasiljeva, L., Rando, O.J., Nedea, E., Greenblatt, J.F., and Buratowski, S. 2004. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* **432**(7016): 517-522.

Kim, M., Suh, H., Cho, E.J., and Buratowski, S. 2009. Phosphorylation of the yeast Rpb1 C-terminal domain at serines 2, 5, and 7. *J Biol Chem* **284**(39): 26421-26426.

Kim, M., Vasiljeva, L., Rando, O.J., Zhelkovsky, A., Moore, C., and Buratowski, S. 2006. Distinct pathways for snoRNA and mRNA termination. *Mol Cell* **24**(5): 723-734.

Kireeva, M.L., Komissarova, N., and Kashlev, M. 2000. Overextended RNA:DNA hybrid as a negative regulator of RNA polymerase II processivity. *J Mol Biol* **299**(2): 325-335.

Komarnitsky, P., Cho, E.J., and Buratowski, S. 2000. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev* **14**(19): 2452-2460.

Kouzarides, T. 2007. Chromatin modifications and their function. *Cell* **128**(4): 693-705.

Krishnamurthy, S., He, X., Reyes-Reyes, M., Moore, C., and Hampsey, M. 2004. Ssu72 Is an RNA polymerase II CTD phosphatase. *Mol Cell* **14**(3): 387-394.

Kubicek, S., O'Sullivan, R.J., August, E.M., Hickey, E.R., Zhang, Q., Teodoro, M.L., Rea, S., Mechtler, K., Kowalski, J.A., Homon, C.A., Kelly, T.A., and Jenuwein, T. 2007. Reversal of H3K9me2 by a small-molecule inhibitor for the G9a histone methyltransferase. *Mol Cell* **25**(3): 473-481.

Kuehner, J.N., Pearson, E.L., and Moore, C. 2011. Unravelling the means to an end: RNA polymerase II transcription termination. *Nat Rev Mol Cell Biol* **12**(5): 283-294.

Lachner, M. and Jenuwein, T. 2002. The many faces of histone lysine methylation. *Curr Opin Cell Biol* **14**(3): 286-298.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**(6824): 116-120.

Lee, D.Y. and Clayton, D.A. 1996. Properties of a primer RNA-DNA hybrid at the mouse mitochondrial DNA leading-strand origin of replication. *J Biol Chem* **271**(39): 24262-24269.

Li, B., Carey, M., and Workman, J.L. 2007. The role of chromatin during transcription. *Cell* **128**(4): 707-719.

Li, X. and Manley, J.L. 2005. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* **122**(3): 365-378.

-. 2006. Cotranscriptional processes and their influence on genome stability. *Genes Dev* **20**(14): 1838-1847.

Licatalosi, D.D., Geiger, G., Minet, M., Schroeder, S., Cilli, K., McNeil, J.B., and Bentley, D.L. 2002. Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. *Mol Cell* **9**(5): 1101-1111.

Lin, P.S., Dubois, M.F., and Dahmus, M.E. 2002. TFIIF-associating carboxyl-terminal domain phosphatase dephosphorylates phosphoserines 2 and 5 of RNA polymerase II. *J Biol Chem* **277**(48): 45949-45956.

Lin, S., Coutinho-Mansfield, G., Wang, D., Pandit, S., and Fu, X.D. 2008. The splicing factor SC35 has an active role in transcriptional elongation. *Nat Struct Mol Biol* **15**(8): 819-826.

Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. 2004. Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**(5689): 1437-1441.

Logan, J., Falck-Pedersen, E., Darnell, J.E., Jr., and Shenk, T. 1987. A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. *Proc Natl Acad Sci U S A* **84**(23): 8306-8310.

Lunde, B.M., Reichow, S.L., Kim, M., Suh, H., Leeper, T.C., Yang, F., Mutschler, H., Buratowski, S., Meinhart, A., and Varani, G. 2010. Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat Struct Mol Biol* **17**(10): 1195-1201.

Luo, W., Johnson, A.W., and Bentley, D.L. 2006. The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. *Genes Dev* **20**(8): 954-965.

Mandel, C.R., Bai, Y., and Tong, L. 2008. Protein factors in pre-mRNA 3'-end processing. *Cell Mol Life Sci* **65**(7-8): 1099-1122.

Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J.L., and Tong, L. 2006. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* **444**(7121): 953-956.

Mapendano, C.K., Lykke-Andersen, S., Kjems, J., Bertrand, E., and Jensen, T.H. 2010. Crosstalk between mRNA 3' end processing and transcription initiation. *Mol Cell* **40**(3): 410-422.

Masukata, H. and Tomizawa, J. 1984. Effects of point mutations on formation and structure of the RNA primer for ColE1 DNA replication. *Cell* **36**(2): 513-522.

-. 1990. A mechanism of formation of a persistent hybrid between elongating RNA and template DNA. *Cell* **62**(2): 331-338.

Mattick, J.S. 2009. The genetic signatures of noncoding RNAs. *PLoS Genet* **5**(4): e1000459.

Matzke, M.A. and Birchler, J.A. 2005. RNAi-mediated pathways in the nucleus. *Nat Rev Genet* **6**(1): 24-35.

Mayer, A., Heidemann, M., Lidschreiber, M., Schreieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., and Cramer, P. 2012. CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science* **336**(6089): 1723-1725.

Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Soding, J., and Cramer, P. 2010. Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol* **17**(10): 1272-1278.

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., and Bentley, D.L. 1997. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**(6614): 357-361.

McLauchlan, J., Gaffney, D., Whitton, J.L., and Clements, J.B. 1985. The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini. *Nucleic Acids Res* **13**(4): 1347-1368.

Meinhart, A. and Cramer, P. 2004. Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature* **430**(6996): 223-226.

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., and Tuschl, T. 2004. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell* **15**(2): 185-197.

Meister, G. and Tuschl, T. 2004. Mechanisms of gene silencing by double-stranded RNA. *Nature* **431**(7006): 343-349.

Millevoi, S. and Vagner, S. 2010. Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Res* **38**(9): 2757-2774.

Minvielle-Sebastia, L., Preker, P.J., and Keller, W. 1994. RNA14 and RNA15 proteins as components of a yeast pre-mRNA 3'-end processing factor. *Science* **266**(5191): 1702-1705.

Minvielle-Sebastia, L., Preker, P.J., Wiederkehr, T., Strahm, Y., and Keller, W. 1997. The major yeast poly(A)-binding protein is associated with cleavage factor IA and functions in premessenger RNA 3'-end formation. *Proc Natl Acad Sci U S A* **94**(15): 7897-7902.

Mischo, H.E., Gomez-Gonzalez, B., Grzechnik, P., Rondon, A.G., Wei, W., Steinmetz, L., Aguilera, A., and Proudfoot, N.J. 2011. Yeast Sen1 helicase protects the genome from transcription-associated instability. *Mol Cell* **41**(1): 21-32.

Montgomery, M.K. and Fire, A. 1998. Double-stranded RNA as a mediator in sequence-specific genetic silencing and co-suppression. *Trends Genet* **14**(7): 255-258.

Moore, M.J. and Proudfoot, N.J. 2009. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* **136**(4): 688-700.

Moreira, A., Wollerton, M., Monks, J., and Proudfoot, N.J. 1995. Upstream sequence elements enhance poly(A) site efficiency of the C2 complement gene and are phylogenetically conserved. *EMBO J* **14**(15): 3809-3819.

Moreira, M.C., Klur, S., Watanabe, M., Nemeth, A.H., Le Ber, I., Moniz, J.C., Tranchant, C., Aubourg, P., Tazir, M., Schols, L., Pandolfo, M., Schulz, J.B., Pouget, J., Calvas, P., Shizuka-Ikeda, M., Shoji, M., Tanaka, M., Izatt, L., Shaw, C.E., M'Zahem, A., Dunne, E., Bomont, P., Benhassine, T., Bouslam, N., Stevanin, G., Brice, A., Guimaraes, J., Mendonca, P., Barbot, C., Coutinho, P., Sequeiros, J., Durr, A., Warter, J.M., and Koenig, M. 2004. Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2. *Nat Genet* **36**(3): 225-227.

Morey, C. and Avner, P. 2004. Employment opportunities for non-coding RNAs. *FEBS Lett* **567**(1): 27-34.

Morris, K.V., Chan, S.W., Jacobsen, S.E., and Looney, D.J. 2004. Small interfering RNA-induced transcriptional gene silencing in human cells. *Science* **305**(5688): 1289-1292.

Mosley, A.L., Pattenden, S.G., Carey, M., Venkatesh, S., Gilmore, J.M., Florens, L., Workman, J.L., and 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**(2): 168-178.

Motamedi, M.R., Verdel, A., Colmenares, S.U., Gerber, S.A., Gygi, S.P., and Moazed, D. 2004. Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* **119**(6): 789-802.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. 2000. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102**(5): 553-563.

Murthy, K.G. and Manley, J.L. 1992. Characterization of the multisubunit cleavage-polyadenylation specificity factor from calf thymus. *J Biol Chem* **267**(21): 14804-14811.

-. 1995. The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. *Genes Dev* **9**(21): 2672-2683.

Nakama, M., Kawakami, K., Kajitani, T., Urano, T., and Murakami, Y. DNA-RNA hybrid formation mediates RNAi-directed heterochromatin formation. *Genes Cells* **17**(3): 218-233.

Nechaev, S., Fargo, D.C., dos Santos, G., Liu, L., Gao, Y., and Adelman, K. 2010. Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. *Science* **327**(5963): 335-338.

Nedea, E., He, X., Kim, M., Pootoolal, J., Zhong, G., Canadien, V., Hughes, T., Buratowski, S., Moore, C.L., and 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**(35): 33000-33010.

Noma, K., Sugiyama, T., Cam, H., Verdel, A., Zofall, M., Jia, S., Moazed, D., and Grewal, S.I. 2004. RITS acts in cis to promote RNA interference-mediated transcriptional and post-transcriptional silencing. *Nat Genet* **36**(11): 1174-1180.

O'Neill, M.J. 2005. The influence of non-coding RNAs on allele-specific gene expression in mammals. *Hum Mol Genet* **14 Spec No 1**: R113-120.

Ohrt, T., Mutze, J., Staroske, W., Weinmann, L., Hock, J., Crell, K., Meister, G., and Schwille, P. 2008. Fluorescence correlation spectroscopy and fluorescence cross-correlation spectroscopy reveal the cytoplasmic origination of loaded nuclear RISC in vivo in human cells. *Nucleic Acids Res* **36**(20): 6439-6449.

Padmanabhan, K., Robles, M.S., Westerling, T., and Weitz, C.J. 2012. Feedback regulation of transcriptional termination by the mammalian circadian clock PERIOD complex. *Science* **337**(6094): 599-602.

Palau, F. and Espinos, C. 2006. Autosomal recessive cerebellar ataxias. *Orphanet J Rare Dis* **1**: 47.

Passarge, E. 1979. Emil Heitz and the concept of heterochromatin: longitudinal chromosome differentiation was recognized fifty years ago. *Am J Hum Genet* **31**(2): 106-115.

Paulsen, R.D., Soni, D.V., Wollman, R., Hahn, A.T., Yee, M.C., Guan, A., Hesley, J.A., Miller, S.C., Cromwell, E.F., Solow-Cordero, D.E., Meyer, T., and Cimprich, K.A. 2009. A genome-wide siRNA screen reveals diverse cellular processes and pathways that mediate genome stability. *Mol Cell* **35**(2): 228-239.

Perales, R. and Bentley, D. 2009. "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. *Mol Cell* **36**(2): 178-191.

Peters, A.H., Kubicek, S., Mechtler, K., O'Sullivan, R.J., Derijck, A.A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y., Martens, J.H., and Jenuwein, T. 2003. Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol Cell* **12**(6): 1577-1589.

Petrie, V.J., Wuitschick, J.D., Givens, C.D., Kosinski, A.M., and Partridge, J.F. 2005. RNA interference (RNAi)-dependent and RNAi-independent association of the Chp1 chromodomain protein with distinct heterochromatic loci in fission yeast. *Mol Cell Biol* **25**(6): 2331-2346.

Pfluger, J. and Wagner, D. 2007. Histone modifications and dynamic regulation of genome accessibility in plants. *Curr Opin Plant Biol* **10**(6): 645-652.

Phatnani, H.P. and Greenleaf, A.L. 2006. Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev* **20**(21): 2922-2936.

Piruat, J.I. and Aguilera, A. 1998. A novel yeast gene, THO2, is involved in RNA pol II transcription and provides new evidence for transcriptional elongation-associated recombination. *EMBO J* **17**(16): 4859-4872.

Plant, K.E., Dye, M.J., Lafaille, C., and Proudfoot, N.J. 2005. Strong polyadenylation and weak pausing combine to cause efficient termination of transcription in the human Ggamma-globin gene. *Mol Cell Biol* **25**(8): 3276-3285.

Prado, F. and Aguilera, A. 2005. Impairment of replication fork progression mediates RNA polII transcription-associated recombination. *EMBO J* **24**(6): 1267-1276.

Proudfoot, N. 2004. New perspectives on connecting messenger RNA 3' end formation to transcription. *Curr Opin Cell Biol* **16**(3): 272-278.

Proudfoot, N.J. 2011. Ending the message: poly(A) signals then and now. *Genes Dev* **25**(17): 1770-1782.

-. 1989. How RNA polymerase II terminates transcription in higher eukaryotes. *Trends Biochem Sci* **14**(3): 105-110.

Proudfoot, N.J. and Brownlee, G.G. 1976. 3' non-coding region sequences in eukaryotic messenger RNA. *Nature* **263**(5574): 211-214.

Proudfoot, N.J., Furger, A., and Dye, M.J. 2002. Integrating mRNA processing with transcription. *Cell* **108**(4): 501-512.

Rasmussen, T.P. and Culbertson, M.R. 1998. The putative nucleic acid helicase Sen1p is required for formation and stability of termini and for maximal rates of synthesis and levels of accumulation of small nucleolar RNAs in Saccharomyces cerevisiae. *Mol Cell Biol* **18**(12): 6885-6896.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., and Jenuwein, T. 2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**(6796): 593-599.

Reaban, M.E., Lebowitz, J., and Griffin, J.A. 1994. Transcription induces the formation of a stable RNA.DNA hybrid in the immunoglobulin alpha switch region. *J Biol Chem* **269**(34): 21850-21857.

Renner, D.B., Yamaguchi, Y., Wada, T., Handa, H., and Price, D.H. 2001. A highly purified RNA polymerase II elongation control system. *J Biol Chem* **276**(45): 42601-42609.

Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Labelouse, R., Gennery, A., Tezcan, I., Ersoy, F., Kayserili, H., Ugazio, A.G., Brousse, N., Muramatsu, M., Notarangelo, L.D., Kinoshita, K., Honjo, T., Fischer, A., and Durandy, A. 2000. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell* **102**(5): 565-575.

Reyes-Turcu, F.E. and Grewal, S.I. 2012. Different means, same end-heterochromatin formation by RNAi and RNAi-independent RNA processing factors in fission yeast. *Curr Opin Genet Dev* **22**(2): 156-163.

Reyes-Turcu, F.E., Zhang, K., Zofall, M., Chen, E., and Grewal, S.I. 2011. Defects in RNA quality control factors reveal RNAi-independent nucleation of heterochromatin. *Nat Struct Mol Biol* **18**(10): 1132-1138.

Rice, J.C., Briggs, S.D., Ueberheide, B., Barber, C.M., Shabanowitz, J., Hunt, D.F., Shinkai, Y., and Allis, C.D. 2003. Histone methyltransferases direct different degrees of methylation to define distinct chromatin domains. *Mol Cell* **12**(6): 1591-1598.

Roberts, R.W. and Crothers, D.M. 1992. Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. *Science* **258**(5087): 1463-1466.

Rondon, A.G., Jimeno, S., Garcia-Rubio, M., and Aguilera, A. 2003. Molecular evidence that the eukaryotic THO/TREX complex is required for efficient transcription elongation. *J Biol Chem* **278**(40): 39037-39043.

Rosonina, E., Ip, J.Y., Calarco, J.A., Bakowski, M.A., Emili, A., McCracken, S., Tucker, P., Ingles, C.J., and Blencowe, B.J. 2005. Role for PSF in mediating transcriptional activator-dependent stimulation of pre-mRNA processing in vivo. *Mol Cell Biol* **25**(15): 6734-6746.

Roy, D. and Lieber, M.R. 2009. G clustering is important for the initiation of transcription-induced R-loops in vitro, whereas high G density without clustering is sufficient thereafter. *Mol Cell Biol* **29**(11): 3124-3133.

Roy, D., Yu, K., and Lieber, M.R. 2008. Mechanism of R-loop formation at immunoglobulin class switch sequences. *Mol Cell Biol* **28**(1): 50-60.

Roy, D., Zhang, Z., Lu, Z., Hsieh, C.L., and Lieber, M.R. 2010. Competition between the RNA transcript and the nontemplate DNA strand during R-loop formation in vitro: a nick can serve as a strong R-loop initiation site. *Mol Cell Biol* **30**(1): 146-159.

Ryan, K., Calvo, O., and Manley, J.L. 2004. Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *RNA* **10**(4): 565-573.

Sadowski, M., Dichtl, B., Hubner, W., and Keller, W. 2003. Independent functions of yeast Pcf11p in pre-mRNA 3' end processing and in transcription termination. *EMBO J* **22**(9): 2167-2177.

Salisbury, J., Hutchison, K.W., and Graber, J.H. 2006. A multispecies comparison of the metazoan 3'-processing downstream elements and the CstF-64 RNA recognition motif. *BMC Genomics* **7**: 55.

Schalch, T., Job, G., Noffsinger, V.J., Shanker, S., Kuscu, C., Joshua-Tor, L., and Partridge, J.F. 2009. High-affinity binding of Chp1 chromodomain to K9 methylated histone H3 is required to establish centromeric heterochromatin. *Mol Cell* **34**(1): 36-46.

Schroeder, S.C., Schwer, B., Shuman, S., and Bentley, D. 2000. Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes Dev* **14**(19): 2435-2440.

Shatkin, A.J. and Manley, J.L. 2000. The ends of the affair: capping and polyadenylation. *Nat Struct Biol* **7**(10): 838-842.

Shaw, N.N. and Arya, D.P. 2008. Recognition of the unique structure of DNA:RNA hybrids. *Biochimie* **90**(7): 1026-1039.

Sheets, M.D., Ogg, S.C., and Wickens, M.P. 1990. Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Res* **18**(19): 5799-5805.

Shi, Y., Di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates, J.R., 3rd, Frank, J., and Manley, J.L. 2009. Molecular architecture of the human pre-mRNA 3' processing complex. *Mol Cell* **33**(3): 365-376.

Shinkai, Y. and Tachibana, M. 2011. H3K9 methyltransferase G9a and the related molecule GLP. *Genes Dev* **25**(8): 781-788.

Sims, R.J., 3rd, Mandal, S.S., and Reinberg, D. 2004. Recent highlights of RNA-polymerase-II-mediated transcription. *Curr Opin Cell Biol* **16**(3): 263-271.

Skourti-Stathaki, K., Proudfoot, N.J., and Gromak, N. 2011. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol Cell* **42**(6): 794-805.

Song, J.J., Smith, S.K., Hannon, G.J., and Joshua-Tor, L. 2004. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* **305**(5689): 1434-1437.

Stein, H. and Hausen, P. 1969. Enzyme from calf thymus degrading the RNA moiety of DNA-RNA Hybrids: effect on DNA-dependent RNA polymerase. *Science* **166**(3903): 393-395.

Steinmetz, E.J., Ng, S.B., Cloute, J.P., and Brow, D.A. 2006a. cis- and trans-Acting determinants of transcription termination by yeast RNA polymerase II. *Mol Cell Biol* **26**(7): 2688-2696.

Steinmetz, E.J., Warren, C.L., Kuehner, J.N., Panbehi, B., Ansari, A.Z., and Brow, D.A. 2006b. Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. *Mol Cell* **24**(5): 735-746.

Stevens, A. and Poole, T.L. 1995. 5'-exonuclease-2 of Saccharomyces cerevisiae. Purification and features of ribonuclease activity with comparison to 5'-exonuclease-1. *J Biol Chem* **270**(27): 16063-16069.

Stirling, P.C., Chan, Y.A., Minaker, S.W., Aristizabal, M.J., Barrett, I., Sipahimalani, P., Kobor, M.S., and Hieter, P. 2012. R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. *Genes Dev* **26**(2): 163-175.

Strasser, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., Rondon, A.G., Aguilera, A., Struhl, K., Reed, R., and Hurt, E. 2002. TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* **417**(6886): 304-308.

Sugimoto, N., Nakano, S., Katoh, M., Matsumura, A., Nakamuta, H., Ohmichi, T., Yoneyama, M., and Sasaki, M. 1995. Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry* **34**(35): 11211-11216.

Sugiyama, T., Cam, H., Verdel, A., Moazed, D., and Grewal, S.I. 2005. RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. *Proc Natl Acad Sci U S A* **102**(1): 152-157.

Suraweera, A., Becherel, O.J., Chen, P., Rundle, N., Woods, R., Nakamura, J., Gatei, M., Criscuolo, C., Filla, A., Chessa, L., Fusser, M., Epe, B., Gueven, N., and Lavin, M.F. 2007. Senataxin, defective in ataxia oculomotor apraxia type 2, is involved in the defense against oxidative DNA damage. *J Cell Biol* **177**(6): 969-979.

Suraweera, A., Lim, Y., Woods, R., Birrell, G.W., Nasim, T., Becherel, O.J., and Lavin, M.F. 2009. Functional role for senataxin, defective in ataxia oculomotor apraxia type 2, in transcriptional regulation. *Hum Mol Genet* **18**(18): 3384-3396.

Swinburne, I.A., Meyer, C.A., Liu, X.S., Silver, P.A., and Brodsky, A.S. 2006. Genomic localization of RNA binding proteins reveals links between pre-mRNA processing and transcription. *Genome Res* **16**(7): 912-921.

Tabara, H., Grishok, A., and Mello, C.C. 1998. RNAi in C. elegans: soaking in the genome sequence. *Science* **282**(5388): 430-431.

Tachibana, M., Matsumura, Y., Fukuda, M., Kimura, H., and Shinkai, Y. 2008. G9a/GLP complexes independently mediate H3K9 and DNA methylation to silence transcription. *EMBO J* **27**(20): 2681-2690.

Tachibana, M., Sugimoto, K., Fukushima, T., and Shinkai, Y. 2001. Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone

methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. *J Biol Chem* **276**(27): 25309-25317.

Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., Ohki, M., Fukuda, M., Takeda, N., Niida, H., Kato, H., and Shinkai, Y. 2002. G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev* **16**(14): 1779-1791.

Tachibana, M., Ueda, J., Fukuda, M., Takeda, N., Ohta, T., Iwanari, H., Sakihama, T., Kodama, T., Hamakubo, T., and Shinkai, Y. 2005. Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. *Genes Dev* **19**(7): 815-826.

Takagaki, Y., MacDonald, C.C., Shenk, T., and Manley, J.L. 1992. The human 64-kDa polyadenylylation factor contains a ribonucleoprotein-type RNA binding domain and unusual auxiliary motifs. *Proc Natl Acad Sci U S A* **89**(4): 1403-1407.

Takagaki, Y. and Manley, J.L. 1997. RNA recognition by the human polyadenylation factor CstF. *Mol Cell Biol* **17**(7): 3907-3914.

-. 2000. Complex protein interactions within the human polyadenylation machinery identify a novel component. *Mol Cell Biol* **20**(5): 1515-1525.

Takagaki, Y., Ryner, L.C., and Manley, J.L. 1989. Four factors are required for 3'-end cleavage of pre-mRNAs. *Genes Dev* **3**(11): 1711-1724.

Thomas, M., White, R.L., and Davis, R.W. 1976. Hybridization of RNA to double-stranded DNA: formation of R-loops. *Proc Natl Acad Sci U S A* **73**(7): 2294-2298.

Timmons, L. and Fire, A. 1998. Specific interference by ingested dsRNA. *Nature* **395**(6705): 854.

Tomari, Y. and Zamore, P.D. 2005. Perspective: machines for RNAi. *Genes Dev* **19**(5): 517-529.

Trojer, P. and Reinberg, D. 2007. Facultative heterochromatin: is there a distinctive molecular signature? *Mol Cell* **28**(1): 1-13.

Tuduri, S., Crabbe, L., Conti, C., Tourriere, H., Holtgreve-Grez, H., Jauch, A., Pantesco, V., De Vos, J., Thomas, A., Theillet, C., Pommier, Y., Tazi, J., Coquelle, A., and Pasero, P. 2009. Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nat Cell Biol* **11**(11): 1315-1324.

Tufarelli, C., Stanley, J.A., Garrick, D., Sharpe, J.A., Ayyub, H., Wood, W.G., and Higgs, D.R. 2003. Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nat Genet* **34**(2): 157-165.

Ursic, D., Chinchilla, K., Finkel, J.S., and Culbertson, M.R. 2004. Multiple protein/protein and protein/RNA interactions suggest roles for yeast DNA/RNA helicase Sen1p in transcription, transcription-coupled DNA repair and RNA processing. *Nucleic Acids Res* **32**(8): 2441-2452.

Ursic, D., Himmel, K.L., Gurley, K.A., Webb, F., and Culbertson, M.R. 1997. The yeast SEN1 gene is required for the processing of diverse RNA classes. *Nucleic Acids Res* **25**(23): 4778-4785.

Vasiljeva, L. and Buratowski, S. 2006. Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Mol Cell* **21**(2): 239-248.

Venkataraman, K., Brown, K.M., and Gilmartin, G.M. 2005. Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. *Genes Dev* **19**(11): 1315-1327.

Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S.I., and Moazed, D. 2004. RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**(5658): 672-676.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., and Martienssen, R.A. 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**(5588): 1833-1837.

Waterhouse, P.M., Graham, M.W., and Wang, M.B. 1998. Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *Proc Natl Acad Sci U S A* **95**(23): 13959-13964.

Werner, A. and Sayer, J.A. 2009. Naturally occurring antisense RNA: function and mechanisms of action. *Curr Opin Nephrol Hypertens* **18**(4): 343-349.

West, S., Gromak, N., and Proudfoot, N.J. 2004. Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* **432**(7016): 522-525.

West, S. and Proudfoot, N.J. 2008. Human Pcf11 enhances degradation of RNA polymerase II-associated nascent RNA and transcriptional termination. *Nucleic Acids Res* **36**(3): 905-914.

-. 2009. Transcriptional termination enhances protein expression in human cells. *Mol Cell* **33**(3): 354-364.

West, S., Proudfoot, N.J., and Dye, M.J. 2008. Molecular dissection of mammalian RNA polymerase II transcriptional termination. *Mol Cell* **29**(5): 600-610.

West, S., Zaret, K., and Proudfoot, N.J. 2006. Transcriptional termination sequences in the mouse serum albumin gene. *RNA* **12**(4): 655-665.

Westover, K.D., Bushnell, D.A., and Kornberg, R.D. 2004. Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell* **119**(4): 481-489.

Whitelaw, E. and Proudfoot, N. 1986. Alpha-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3' end processing in the human alpha 2 globin gene. *EMBO J* **5**(11): 2915-2922.

Wu, C.H., Yamaguchi, Y., Benjamin, L.R., Horvat-Gordon, M., Washinsky, J., Enerly, E., Larsson, J., Lambertsson, A., Handa, H., and Gilmour, D. 2003. NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila. *Genes Dev* **17**(11): 1402-1414.

Xiang, K., Manley, J.L., and Tong, L. 2012. The yeast regulator of transcription protein Rtr1 lacks an active site and phosphatase activity. *Nat Commun* **3**: 946.

Xiang, S., Cooper-Morgan, A., Jiao, X., Kiledjian, M., Manley, J.L., and Tong, L. 2009. Structure and function of the 5'-->3' exoribonuclease Rat1 and its activating partner Rai1. *Nature* **458**(7239): 784-788.

Xu, B. and Clayton, D.A. 1996. RNA-DNA hybrid formation at the human mitochondrial heavy-strand origin ceases at replication start sites: an implication for RNA-DNA hybrids serving as primers. *EMBO J* **15**(12): 3135-3143.

Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H. 1999. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* **97**(1): 41-51.

Yamane, A., Resch, W., Kuo, N., Kuchen, S., Li, Z., Sun, H.W., Robbiani, D.F., McBride, K., Nussenzweig, M.C., and Casellas, R. 2011. Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. *Nat Immunol* **12**(1): 62-69.

Yonaha, M. and Proudfoot, N.J. 1999. Specific transcriptional pausing activates polyadenylation in a coupled in vitro system. *Mol Cell* **3**(5): 593-600.

Yu, K., Chedin, F., Hsieh, C.L., Wilson, T.E., and Lieber, M.R. 2003. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat Immunol* **4**(5): 442-451.

Zhang, D.W., Mosley, A.L., Ramisetty, S.R., Rodriguez-Molina, J.B., Washburn, M.P., and Ansari, A.Z. 2012. 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**(11): 8541-8551.

Zhang, K., Fischer, T., Porter, R.L., Dhakshnamoorthy, J., Zofall, M., Zhou, M., Veenstra, T., and Grewal, S.I. 2011. Clr4/Suv39 and RNA quality control factors cooperate to trigger RNAi and suppress antisense RNA. *Science* **331**(6024): 1624-1627.

Zhang, K., Mosch, K., Fischle, W., and Grewal, S.I. 2008. Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin. *Nat Struct Mol Biol* **15**(4): 381-388.

Zhang, Z., Fu, J., and Gilmour, D.S. 2005. CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11. *Genes Dev* **19**(13): 1572-1580.

Zhang, Z. and Gilmour, D.S. 2006. Pcf11 is a termination factor in Drosophila that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. *Mol Cell* **21**(1): 65-74.

Zhao, J., Hyman, L., and Moore, C. 1999. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol Mol Biol Rev* **63**(2): 405-445.

Zofall, M., Fischer, T., Zhang, K., Zhou, M., Cui, B., Veenstra, T.D., and Grewal, S.I. 2009. Histone H2A.Z cooperates with RNAi and heterochromatin factors to suppress antisense RNAs. *Nature* **461**(7262): 419-422.