Connections between RNA Polymerase II transcriptional regulation, nucleosome arrangement and DNA damage

Jonathan Foxwell

University of Oxford
Sir William Dunn School of Pathology
St Edmund Hall

Thesis submitted for the degree of DPhil

Trinity Term 2017

Acknowledgements

Particular thanks go to Professor Nicholas Proudfoot, in whose laboratory this work was conducted. Nick has been an inspiration and a great help – always fascinated by the science, and calm and collected about everything.

Dr Konstantina Skourti-Stathaki and Dr Takayuki Nojima for their tireless supervision and comments. This thesis would not have been possible without Nadina stepping up at the last moment to help with my introduction to the lab and to provide early research directions and training. Taka has been an incredible help, with his work providing the background for so much of this thesis, and with his personal guidance.

Specific thanks go to Dr Michael Tellier and Dr Margarita Schlackow for all their work on bioinformatic analysis, Dr Ashish Dhir for help with immunofluorescence, Dr Andrew Bassett for help designing and cloning gRNAs, and Dr Fumiko Esashi and her lab for suggestions and information on Rad51, and the antibodies. Also to Svenja Hester, Dr Sabrina Liberatori and the rest of the proteomics facility for processing many samples and answering many questions. Thank you also to Dr Kinga Kamieniarz-Gdula for her guidance with bioinformatics.

NJP lab members past and present have created a wonderful working environment. In particular, Sue-Mei Tan-Wong, “Malevolent” Adam Volanakis

and Esther Griesbach for fun and help when needed throughout the last four years, and for permanently warping my sense of humour.

Thanks also to my friends from Clare, Teddy Hall MCR and the Sunday “Fud” group. Even Martin.

To Dr Anja Mizdrak (because there’s only one). Anja, thank you for everything over the last three years, for keeping me happy and sane, and occasionally upside down in river water. Hvala, i volim te jako puno.

And lastly, to my family. My parents David and Jane, and sisters Harriet and Thea, for all their support and kindness, over the last four years and all the ones before that, despite my inability to communicate.

This DPhil was funded by the Sir William Dunn School Departmental (EPA) Studentship and the MRC.

All experiments presented in this thesis were carried out by Jonathan Foxwell, with one exception. The data presented on page 132, Chapter 6, figures 6.6-8, were generated by Dr Takayuki Nojima.

Abstract

Transcription by RNA polymerase II (RNAPII) is a central process in gene expression and so subject to many layers of regulation: post-translational modifications of the RNAPII C-terminal domain (CTD), transcription factor association, chromatin compaction and the control of polymerase access to DNA. Transcription can also be altered when DNA damage is encountered.

My thesis addressed different aspects of these regulatory mechanisms. Firstly I studied the role of RNA:DNA hybrids (R-loops) in pause-type termination, and assessed whether data based on the *β-actin* gene could be extended to further genes. I also investigated whether DNA damage repair (DDR) factor Rad51 is present over R-loop associated pause regions. Secondly I studied the roles of tyrosine and threonine phosphorylation (Y1P and T4P) of the RNAPII CTD, by testing their association with DDR and transcription termination respectively. I also developed a modified MNase-seq protocol using mNET-seq isolation conditions to allow close comparison between chromatin configuration and the nascent RNA signal.

The outcome of my research was that I found the *PKM* gene exhibits the same basic characteristics of pause-type termination, including R-loop formation. However, little evidence for heterochromatin formation as observed for *β-actin* termination was identified. I also showed that Rad51 accumulates at R-loop regions in pause-type termination, but does not appear to induce R-loops. More likely Rad51 is recruited by downstream DDR factors. Y1P is associated with factors involved in

UV DDR. My studies showed that UV damage causes a global RNAPII elongation defect with altered chromatin associated histone marks and defective recruitment of elongation factors including the PAF1 complex and SPT6 to the polymerase. Finally my studies on T4P-associated termination cis-elements indicated that positions of cleavage-independent termination are formed relative to mRNA 3' ends but not based on specific DNA sequence. Also cleavage factor depletion increased mononucleosome signals in these termination regions.

Contents

Acknowledgements ii
Abstract v
Table of Contents 1
Abbreviations 4

1. Introduction 8
   1.1 Chromatin 9
   1.2 Eukaryotic transcription 11
      1.2.1 RNA polymerase II 11
      1.2.2 The transcription cycle 11
      1.2.3 Transcription termination models 16
      1.2.4 Other species of RNA synthesized by RNAPII 18
   1.3 RNAPII CTD and links to transcriptional processing 20
      1.3.1 The C-terminal domain of RNAPII 20
      1.3.2 Methods to study RNA transcription and CTD profiles 22
      1.3.3 mNET-seq 23
      1.3.4 Established roles and profiles for different RNAPII modifications 24
      1.3.5 CTD Tyrosine 1 and Threonine 4 phosphorylation 26
   1.4 R-loops 30
      1.4.1 Introduction to R-loops 30
      1.4.2 Regulation of R-loops 31
      1.4.3 Features favouring R-loop formation 32
      1.4.4 R-loops and transcriptional regulation 34
      1.4.5 Deleterious effects of R-loops 36
      1.4.6 Rad51 and R-loops 38
   1.5 DNA damage 39
      1.5.1 Ultraviolet radiation 39
      1.5.2 Nucleotide excision repair 40
      1.5.3 UV and transcription 42
      1.5.4 γH2AX as a marker of DNA damage 43
   1.6 Aims 44

2. Materials and Methods 45
   2.1 Tissue Culture 46
      2.1.1 Growth and maintenance 46
      2.1.2 Plasmid transfections and RNAi 46
         2.1.2.1 HeLa 46
         2.1.2.2 eHAP 47
         2.1.2.3 Clonal selection 48
      2.1.3 BO2 inhibitor treatment 48
      2.1.4 UV irradiation 48

2.2 Isolation of chromatin, ChrRNA, nucleoplasmic and total RNA 49
2.3 PCR 50
2.3.1 RT-PCR (cDNA synthesis) 50
2.3.2 Standard PCR 50
2.3.3 qPCR 50
2.4 Chromatin immunoprecipitation 51
2.5 DNA:RNA immunoprecipitation 52
2.6 Protein extraction 53
2.6.1 SDS buffer 53
2.6.2 Acid histone extraction 53
2.7 Western blot analysis 53
2.8 MNase digestion and RNAPII/chromatin precipitation 54
2.9 Mass spectrometry 55
2.9.1 FASP 55
2.9.2 Processing and analysis 56
2.10 NetMNase-seq 57
2.10.1 Protocol 57
2.10.2 Human genome annotation 57
2.10.3 Data processing 58
2.11 Immunofluorescence 59
2.12 Antibodies 60
2.13 Primers 62
2.14 siRNA duplex sequences 67
2.15 Plasmids and gRNAs 67

3. R-loops and Rad51 69
3.1 Introduction 70
3.2 *PKM* as a model gene for pause-type termination 71
3.3 Rad51 is present at pause regions 78
3.4 Rad51 does not effect R-loop formation 79
3.5 Discussion 83

4. Development of NetMNase-seq 85
4.1 Introduction 86
4.2 Establishment of chromatin isolation and digestion 88
4.3 Size selection and library preparation 89
4.4 First NetMNase-seq analysis 89
4.5 Addition of immunoprecipitation stage 93
4.6 Sequencing results for RNAPII immunoprecipitation 94
4.7 Discussion 101

5. CTD Threonine 4 phosphorylation 105
5.1 Introduction 106
5.2 Factors associated with T4P 107
5.3 CRISPR-Cas9 employed to generate mutant strains lacking “failsafe” region 108

5.4 "Failsafe" termination appears to be positionally determined 112
5.5 CPSF73 depletion alters nucleosome coverage downstream of the pA site 116
5.6 Discussion 119

6. CTD Tyrosine 1 phosphorylation 123
6.1 Introduction 124
6.2 Repeat mass spectrometry identified the same proteins associated specifically with Y1P polymerase 126
6.3 Western blot analysis shows no indication that BAZ1b is a CTD kinase 127
6.4 UV damage 128
6.5 Y1P and damage foci do not co-localise 131
6.6 UV damage causes a global elongation defect 133
6.7 UV damage increases the association of the PAF complex with RNAPII 136
6.8 BAZ1b association with 3' ends does not alter upon UV damage 140
6.9 PAF1C components are lost from chromatin at the 3' ends of genes 140
6.10 UV irradiation causes a loss of histones and specific histone marks 143
6.11 BAZ1b and SNF2h depletion causes nucleosome remodelling 145
6.12 Discussion 150

7. General Discussion 157
7.1 Summary 158
7.2 Future experiments 163

References 166

Appendices 187
1. Mass spectrometry results 188
1.1 Phospho-threonine 4 results 188
1.2 Phospho-tyrosine 1 results 189
1.3 Total RNAPII +/- UV irradiation results 195
2. Sequencing results 199

Abbreviations

APA	Alternative Poly-adenylation
Ago1/2	Argonaute 1/2
BER	Base excision repair
Br-UTP	Bromo-uridine triphosphate
BSA	Bovine Serum Albumin
CDK	Cyclin-dependent kinase
cDNA	Complementary DNA
CFIm	Cleavage Factor I (mammalian)
CFIIm	Cleavage Factor II (mammalian)
ChIP	Chromatin Immunoprecipitation
CLIP	Cross-linking Immunoprecipitation
CoTC	Co-transcriptional cleavage
CPA	Cleavage and poly-adenylation
CPD	Cyclobutane pyrimidine dimer
CPSF	Cleavage and Poly-adenylation Specificity Factor
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CstF	Cleavage Stimulation Factor
CTD	C-terminal domain
DMEM	Dulbecco's Modified Eagle's Medium
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DRIP	DNA:RNA Immunoprecipitation

DSE	Downstream sequence element
DSIF	DRB sensitivity inducing factor
dsRNA	Double-stranded RNA
DTT	Dithiothreitol
ECL	Electrochemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EGS	Ethylene glycol bis(succinimidyl succinate)
eRNA	Enhancer RNA
FACT	Facilitates chromatin transcription
FASP	Filter Aided Sample Preparation
GRO-seq	Global nuclear run-on sequencing
gRNA	Guide RNA
H3	Histone 3
H3K4/9/36	Histone 3 lysine 4/9/36
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HRMA	High resolution melt assay
Ig	Immunoglobulin
IMDM	Iscove's Modified Dulbecco's Medium
IP	Immunoprecipitation
kDa	kilo-Dalton
lincRNA	Long intergenic non-coding RNA
miRNA	microRNA
MNase	Micrococcal nuclease
mNET-seq	Mammalian Nascent-elongating transcript sequencing

mRNA	messenger RNA
mRNP	messenger Ribonucleoprotein
MS	Mass spectrometry
NDR	Nucleosome depleted region
NELF	Negative elongation factor
NER	Nucleotide excision repair
pA	Poly-adenylation
PAP	Poly(A)-polymerase
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PIC	Pre-initiation complex
PIPES	Piperazine-N,N′-bis(2-ethanesulfonic acid)
PMSF	Phenylmethylsulfonyl fluoride
PNK	Polynucleotide 5-hydroxyl kinase
PROMPT	Promoter upstream transcript
PRO-seq	Precision nuclear run-on sequencing
Pu	Purine
Py	Pyrimidine
qPCR	Quantitative Polymerase Chain Reaction
RNA	Ribonucleic acid
RNAPII	RNA Polymerase II
RNase	Ribonuclease
rRNA	Ribosomal RNA
SDS	Sodium dodecyl sulphate

snRNA	small nuclear RNA
ssDNA	single stranded DNA
S2P	Phospho-serine 2
S5P	Phospho-serine 5
TFII	Transcription factor 2
tRNA	transfer RNA
TES	Transcription End Site – usually equivalent to pA site
TSS	Transcription Start Site
T4P	Phospho- threonine 4
UTR	Untranslated region
UV	Ultraviolet light
WT	Wild type
XP	Xeroderma Pigmentosum
Y1P	Phospho-tyrosine 1

Chapter 1

Introduction

1. Introduction

1.1 Chromatin

Eukaryotic DNA is packaged with protein as chromatin. This allows condensation of chromosomes, protects genetic information from damage, facilitates chromosome separation during cell division, and controls access by DNA replication and RNA transcription machinery.

The basic proteins in chromatin are histones. These are highly alkaline proteins that assemble into octameric complexes consisting of two copies of an H2A-H2B heterodimer and a H3-H4 tetramer that form a core around which DNA is wound. Approximately 146 base pairs of DNA wrap round this octamer in a left-handed super-helical turn and the histone linker protein H1 binds DNA in place (figure 1.1). The DNA-histone octamer particle is known as a nucleosome. Linker DNA sections (20-80bp) connect nucleosomes in a "beads on a string" model (Luger et al. 1997). Different sections of DNA are then packaged into higher-order structures, with higher compaction being associated with heterochromatin, in comparison to euchromatin. Heterochromatin is less accessible to other protein factors, and is therefore more rarely transcribed, and often replicated later in DNA synthesis than euchromatin (Kim et al. 2003). The exact structure of these higher order compactions of DNA is still much debated (Robinson et al. 2006; Wu et al. 2016).

Histones can be modified in a number of ways. Firstly, a number of variant histone genes exist, encoding modified core protein sequences. These variant histones often have specialised roles. For example the Histone 2A variant H2A.X has roles in DNA damage signalling (Paull et al. 2000). Both canonical and variant histones can also be modified post-translationally. This occurs primarily on the N-terminal "tails" – less structured regions that extend from the core of the nucleosome. Residues can be modified in numerous ways – phosphorylation, methylation, acetylation, and ubiquitinylation to name just those most relevant to this thesis. These modifications can act in concert or antagonistically to attract many other protein factors, and so help to regulate compaction, gene expression, DNA replication and many other processes. This is known as the "histone code" (Jenuwein and Allis 2001).

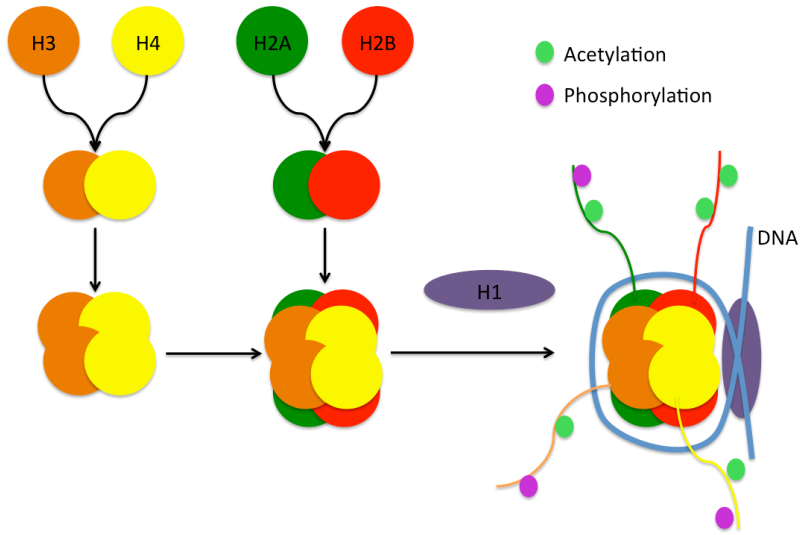


Figure 1.1 Diagram of canonical nucleosome assembly. A H3-H4 tetramer and two H2A-H2B histone dimers constitute the core. Approximately 146 bases of DNA wraps ~1.6 times round the core, and is stabilized by the H1 linker histone. N-terminal tails from the four core histone proteins extend from the main body and can be post-translationally modified in numerous ways – phosphorylation and acetylation shown.

1.2 Eukaryotic transcription

1.2.1 RNA Polymerase II

There are three DNA-dependent RNA polymerases in mammalian cells, each with specialized functions. RNA polymerase I transcribes ribosomal RNA, with the exception of 5S rRNA. RNA polymerase III transcribes 5S rRNA, transfer RNA (tRNA) and a number of other small RNAs (Paule and White 2000). Of most relevance to this thesis, RNA Polymerase II (RNAPII) catalyzes the transcription of DNA into mRNA and most snRNA and miRNA (see section 1.2.4 for more information).

RNAPII is a 550kDa complex with a relatively conserved structure from yeasts to humans. It consists of twelve subunits, the largest of which is RBP1. In combination with a number of other subunits RBP1 creates the polymerase groove that binds and unwinds DNA, and catalyzes RNA polymerization using one DNA strand as a template. Upstream of this transcription bubble, a loop structure within the polymerase called the lid appears to act as a wedge, driving apart the DNA and RNA strands. RNA is guided into a separate "exit channel" to the DNA, and this prevents re-annealing (Hahn 2004).

1.2.2 The transcription cycle

The generation of an mRNA involves a series of different steps, all of which require strict regulation and timing to coordinate productive RNA synthesis, and to prevent aberrant transcription. Each stage in transcription also offers

opportunity for different levels of fine-tuning of gene expression (Shandilya and Roberts 2012).

Transcription begins with RNAPII being loaded onto chromatin at a gene promoter. Areas of transcription initiation are generally nucleosome free or at least nucleosome depleted, allowing transcription factor and RNAPII access. In order to ensure transcription initiates in the correct locations, a number of general transcription factors bind to core promoter sequence elements and directly interact with RNAPII to increase specificity and directionality for DNA binding and unwinding. This complex of transcription factors and RNAPII is known as the Pre-initiation complex (PIC) (Luse 2014). Additional activator protein factors can bind to other DNA elements known as enhancers that can be located up- or downstream of the initiation site - often significant distances away - making contact with the PIC by physical remodeling and looping out of intervening DNA (figure 1.2) (Mora et al. 2016).

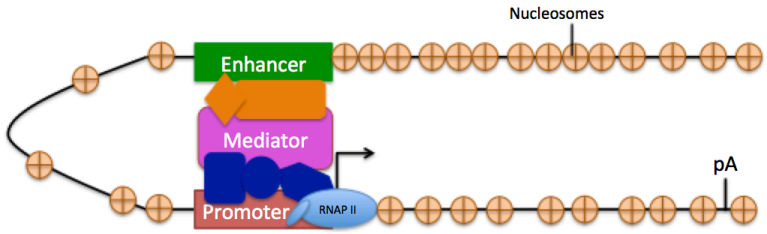


Figure 1.2 Promoter-enhancer interactions and transcription initiation. Nucleosome depleted regions allow the binding of transcription factors (navy) to promoter sequences and activator proteins (orange) to enhancers. These regions are brought into close proximity via physical looping of chromatin and the action of the Mediator complex. Together, general transcription factors and RNAPII form the Pre-initiation complex and contribute to the directionality of transcription.

After an open transcription bubble is formed, the polymerization of the first few (5-10) RNA nucleotides is catalyzed. The growing RNA molecule sterically clashes with a region of a transcription factor (TFIIB) known as the B finger that extends into the RNAPII active site (Westover et al. 2004). A number of these small RNA moieties are released in rounds of abortive transcription. Loaded polymerase bound here in a futile cycle provides a level of regulation allowing fast changes in expression in response to stimuli. In order to enter the full elongation stage of transcription, RNAPII must escape the promoter and PIC by displacing TFIIB, allowing extension of the RNA polymerization. On clearing the promoter the 5' end of RNA exiting the polymerase is capped with a methylated guanine nucleotide connected to the RNA by a 5'-5' triphosphate linkage. This not only provides resistance to exonucleases but is also a required modification for subsequent stages in gene expression (Rau et al. 1996; Cho et al. 1997).

Once escaped from the PIC, promoter proximal pausing often leads to an accumulation of RNAPII close to the initiation site. This is functionally different from the abortive initiation mentioned above in that it generally occurs 20-60bp downstream of the TSS (Transcription Start Site), RNAPII is stably engaged with the DNA and RNA molecules, and efficient elongation can be resumed by the depletion/inhibition of negative regulators of elongation (DSIF and NELF). 5' pausing is generally considered a means of fine tuning expression and allowing for quick changes in expression level, rather than a form of gene silencing. Promoter proximal pausing also contributes to maintenance of nucleosome depleted regions (NDRs) at promoters, perhaps through direct competition for DNA binding. Pausing may also provide a platform for, and checkpoint of,

efficient RNA capping and RNA processing factor recruitment before main stage elongation (Adelman and Lis 2012). RNAPII is released from pausing by phosphorylation of the DSIF/NELF complex, and RNAPII itself, by P-TEFb (Peterlin and Price 2006).

During the elongation phase, RNAPII processes downstream along the gene, synthesizing the bulk of the RNA molecule. During this phase a number of regulatory steps can be combined to co-transcriptionally modify the RNA sequence and to recruit factors for later RNA processing and export. For example, pre-mRNA sequences are often spliced, removing introns, allowing for multiple isoforms of one gene to be expressed through alternative splicing. What processing events occur can be affected by the speed of elongation, which in turn can be altered by such factors as chromatin compaction (Jonkers and Lis 2015). In order for DNA to be transcribed it must be unwound from around histone octamers. These must in some way be returned or replaced (for example by the FACT complex), and post-translational modifications maintained, after RNAPII has moved further downstream (Das and Tyler 2013).

Transcription ends in cleavage and termination. RNAPII dissociates from chromatin, and mRNA is cleaved between a canonical upstream AAUAAA poly-adenylation (pA) sequence and a downstream (DSE) GU-rich sequence. An RNA chain of adenine molecules (poly(A)) is then added to the 3' end of the cleaved mRNA. This is poly-adenylation (Colgan and Manley 1997).

Cleavage and poly-adenylation (CPA) is regulated by a very large protein complex consisting of a series of sub-complexes: CPSF, CstF, CFIm and CFIIm (figure 1.3) (Chan et al. 2011). CPSF, the Cleavage and poly-adenylation specificity factor, contains subunits responsible for binding the above-mentioned pA signal, as well as CPSF73, an endonuclease responsible for cleaving the mRNA precursor. CPSF also recruits other factors including the poly(A) polymerase (PAP), to couple cleavage and poly(A) addition. The Cleavage stimulation factor (CstF) complex binds the DSE and helps to define the exact cleavage site. Cleavage factors (CFIm and CFIIm) bind further upstream elements further enhancing the CPA process as well as regulating alternative poly-adenylation (Tian and Manley 2013; Hardy and Norbury 2016).

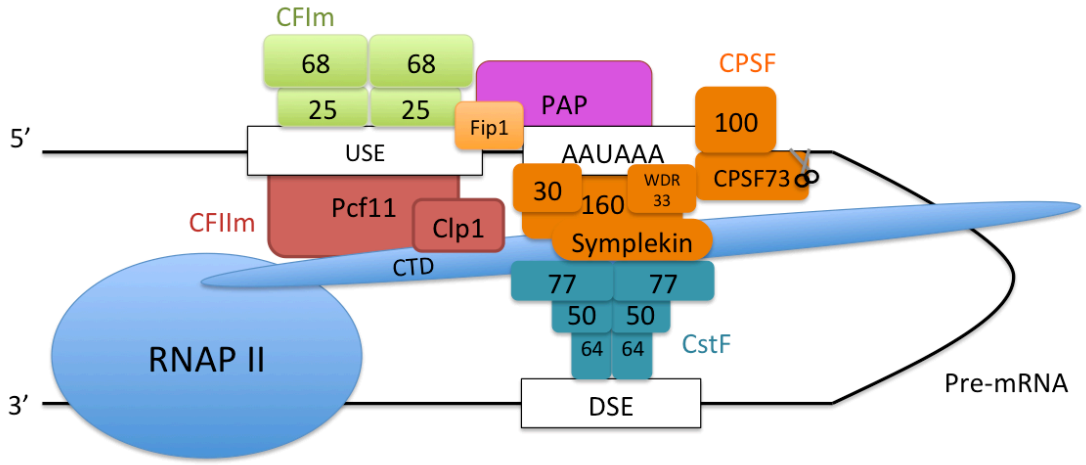


Figure 1.3 Schematic of the CPA complex showing subunits of CPSF, CstF, CFIm and CFIIm and their general positions binding to the AAUAAA poly-adenylation signal and the upstream (USE) and downstream (DSE) sequence elements on mRNA. Most subunit names are an indication of molecular weight (e.g. CPSF100 is approximately 100kDa). Adapted from (Chan et al. 2014; Davis and Shi 2014; Schonemann et al. 2014)

Again, steps in termination provide more levels of gene regulation. mRNA can be poly-adenylated at alternative sites, leading to different length mRNA isoforms,

which can have different stabilities and internal regulatory elements (miRNA binding sites for example). Failure to terminate can also have major consequences for the cell, with excessive read-through leading to aberrant expression of downstream genes (Grosso et al. 2015).

1.2.3 Transcriptional termination models

Termination requires a functional pA signal. Two models have been suggested to explain how the pA sequence triggers termination. In the allosteric model, RNAPII somehow senses the transcription of the AAUAAA sequence, and this triggers a conformational change in the enzyme, perhaps due to association with the CPA complex. This leads to eventual release of RNAPII from DNA. Prior cleavage at the pA is not necessary (Zhang et al. 2015).

Alternatively, the "torpedo" model suggests that cleavage of the pre-mRNA at the pA (or further downstream) allows 5'-3' exonucleases (Xrn2 and Rat1 in mammals and yeast respectively) to degrade the further RNA produced by a still transcribing polymerase. Once these nucleases catch up to RNAPII, they could trigger release from template (Fong et al. 2015; Proudfoot 2016).

The torpedo model can be assisted by termination sequence elements downstream of the pA site. Two forms of termination element have been described in mammals – the co-transcriptionally cleaved (CoTC) element, and the transcription pause site (figure 1.4). In pause-type termination, pA site cleavage is thought to occur first, and the still transcribing polymerase then

slows down at a downstream pause element, allowing the exonuclease to catch up and displace polymerase from the template (Gromak et al. 2006; Skourti-Stathaki et al. 2011). Termination involving CoTC elements is believed to include a different transcript cleavage at a labile CoTC sequence, involving seemingly random sequences interspersed with short A/T homopolymer tracts, often kilobases downstream of the pA site. This is then followed by degradation of the short remaining nascent transcript by Xrn2, displacing RNAPII from the template (Nojima et al. 2013; White et al. 2013). Cleavage at the pA site and release of the mRNA transcript from RNAPII association occurs subsequently.

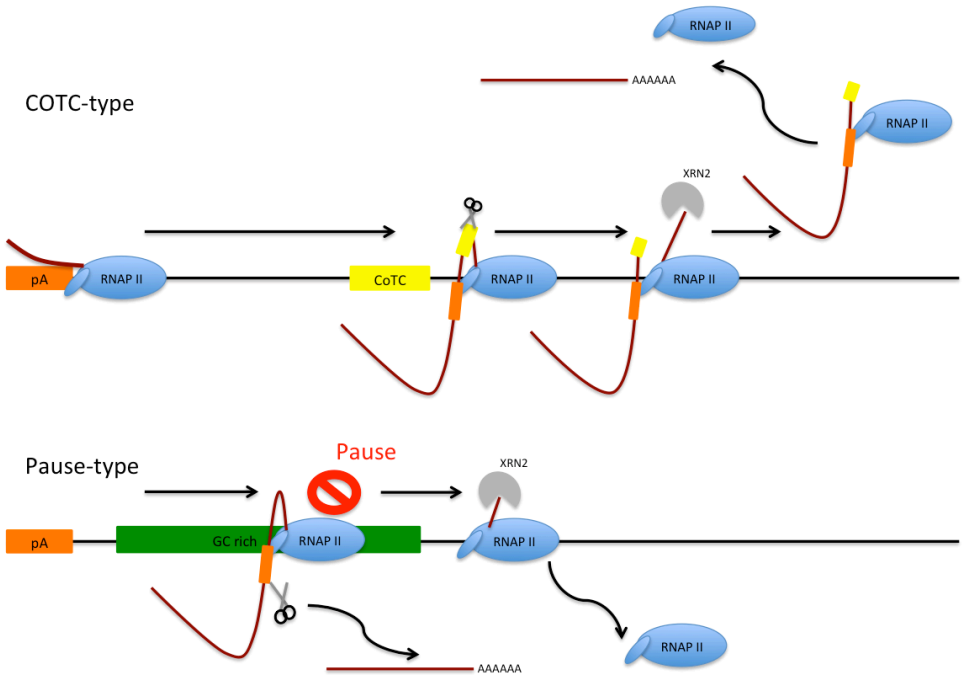


Figure 1.4 Diagram of the difference between termination using the CoTC-type element and the pause-type. In pause-type termination, after transcription of the pA signal, RNAPII enters a GC rich region that leads to retardation or pausing of the polymerase, allowing cleavage at the pA site to occur before RNAPII has transcribed significant lengths of downstream RNA. XRN2 can then start degrading this 3' fragment and catch up to the polymerase, causing RNAPII release from chromatin. In CoTC-type termination, cleavage first occurs at an element further downstream, allowing XRN2 entry and degradation of the RNA emerging from the polymerase exit channel. RNAPII is then released from chromatin into the nucleoplasm, with the pre-mRNA in association with the CTD and bound CPA factors. Cleavage at the pA site then occurs.

While archetypal example genes have been described for both types of element, there is no simple sequence motif for either, and it may be the case that many genes have weaker versions and/or use both elements. Likewise the allosteric and torpedo models may not be entirely mutually exclusive – conformational changes in the polymerase at the pA site may be reinforced by encounters with exonucleases.

1.2.4 Other species of RNA synthesized by RNAPII

In mammals RNAPII transcribes a number of other classes of RNA that often differ radically in their processing at different stages in the transcription cycle. Firstly, canonical (replication-dependent) histone mRNAs are intronless and are not poly-adenylated. Different termination factors recognize a conserved 3' UTR stem loop and downstream purine rich region, and trigger cleavage by CPSF73 between the two (Marzluff et al. 2008).

In addition to mRNA, RNAPII also transcribes a number of non-coding RNAs (ncRNA). These are generally categorized according to their size, place of origin relative to mRNA genes, and the steps in their production (figure 1.5) (St Laurent et al. 2015).

Long non-coding RNAs (lncRNA) include enhancer RNAs (eRNA), Promoter Upstream Transcripts (PROMPTs) and long intergenic non-coding RNAs (lincRNA). As the names suggest, eRNAs and PROMPTs originate in the locality of mRNA gene enhancers and promoters, respectively (Kowalczyk et al. 2012;

Jensen et al. 2013). Nucleosome-depleted regions associated with transcription factor and RNAPII binding in these regions probably allow a certain amount of transcription initiation in an antisense direction at promoters and bi-directionally at enhancers. LincRNAs are produced from independent transcription units, although these are difficult to map and show little conservation and sequence similarity. There is evidence that some lincRNAs are processed in a similar fashion to mRNAs: capping, splicing and poly-adenylation (Derrien et al. 2012). However other measures of co-transcriptional processing suggest processing of most lincRNAs is significantly different or less efficient than mRNA (Schlackow et al. 2017). Whether most lncRNAs have a function or are in general aberrant transcripts is still highly debated.

Smaller ncRNAs transcribed by RNAPII include most small nuclear RNAs (snRNA), small nucleolar RNAs (snoRNA) and micro RNAs (miRNA). snRNAs are transcribed from independent promoters, and capped like mRNA. However termination is driven by a 3' box sequence element, and further processing of pre-snRNAs occurs in the cytoplasm, with 3' trimming to form a 3' stem loop and cap hyper-methylation. These RNAs form snRNP complexes with protein factors and function in the spliceosome, the machinery that removes introns from pre-mRNA, and in histone 3' end processing (Guiro and Murphy 2017).

snoRNAs have roles in chemical modification of other RNA species. The most common role of miRNAs is post-transcriptional regulation of gene expression (Ambros 2004). Many snoRNAs and around 50% of miRNAs are excised from introns, 3' UTRs or even exons of mRNAs, although some snoRNAs have been

found to have independent transcription units (Tycowski and Steitz 2001; Filipowicz and Pogacic 2002; Scott and Ono 2011; Yates et al. 2013). Other miRNAs are encoded in intergenic clusters (Rodriguez et al. 2004; Macfarlane and Murphy 2010).

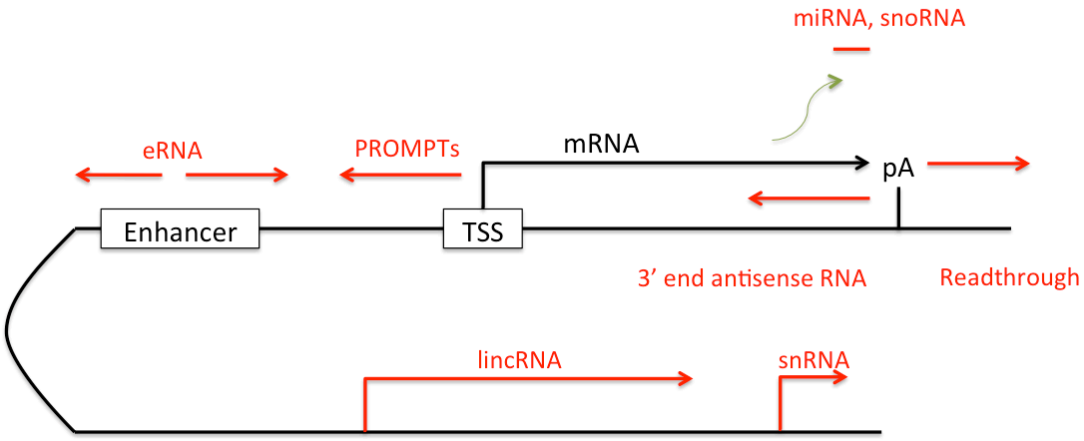


Figure 1.5 The RNA zoo. RNAPII transcription generates many forms of RNA, often classified by their point of origin in the genome relative to mRNA. PROMPTs are upstream antisense transcripts generated from the same promoter as mRNA. eRNAs are transcribed bidirectionally from open chromatin at mRNA enhancers. miRNA and most snoRNAs are excised from pre-mRNA. lincRNA and snRNA generally have separate, independent transcription units. Failure to terminate efficiently can lead to gene read-through at 3' ends. Antisense RNA can also be produced at 3' ends of mRNA, perhaps triggered by R-loop formation.

1.3 RNAPII CTD and links to transcriptional processing

1.3.1 The C-terminal domain of RNAPII

RBP1, the largest subunit of RNAPII, contains a C-terminal domain (CTD) "tail" which consists of multiple repeats of a consensus heptad amino acid sequence (YSPTSPS), that are referred to as tyrosine 1, serine 2, proline 3 and so on. In *S. cerevisiae* there are 26 repeats, whereas the mammalian CTD entails 52 repeats.

Not all heptads conform exactly to the consensus sequence. Most eukaryotic CTD sequences contain a limited diversity of sequence, with more divergence toward the C-terminal portion. Mammals are an extreme case, with the second half of the CTD containing only three heptads with exactly the above sequence (Corden 1990). The extent of divergence is different for different residues - both prolines and the tyrosine amino acid are highly conserved throughout the heptad repeats, whereas serine 7 is the most variable residue. There is also considerable sequence variation between even closely related species (Corden 2013).

The RNAPII "tail" is relatively unstructured and dynamic in shape; it cannot be visualized in X-ray crystallography studies of polymerase structure. Whilst exact information about the conformations the CTD takes *in vivo* is hard to discern, there is evidence that post-translational modifications (PTMs) to the CTD (particularly phosphorylation) can alter its compaction and binding to protein factors. Many of the CTD residues can be dynamically and reversibly modified in a number of different ways (Eick and Geyer 2013; Zaborowska et al. 2016). All three serine, as well as the tyrosine and threonine residues, can be phosphorylated. Glycosylation of the threonine and serine residues has also been detected, as well as methylation and ubiquitinylation of specific non-consensus lysines in mammals (Sims et al. 2011). Cis-trans isomerization of proline residues can alter CTD conformational extension and protein binding capacity (Zhang et al. 2012). Multiple modifications of a single repeat within the CTD appear to be rare. For example, most phosphorylated repeats have been found to be mono or less frequently di- phosphorylated (Schuller et al. 2016).

It is believed that this dynamic modification allows for efficient, regulated recruitment of RNA processing factors with the CTD acting as a scaffold, and PTMs can therefore be considered to act as a "code" linking transcription with coupled RNA-processing, in a similar manner to the histone "code" (Egloff and Murphy 2008; Eick and Geyer 2013). The CTD sits close to the RNA exit channel, ideally placed for orchestrating co-transcriptional RNA processing (Cramer et al. 2001).

1.3.2 Methods to study RNA transcription and CTD profiles

A number of different protocols have been developed to map unstable nascent RNA production in mammals, or the genomic location of different modifications of the RNAPII CTD. These include GRO- and PRO- seq, which employ nuclear run-on and labelling with modified nucleotides (Mahat et al. 2016; Gardini 2017). Native elongating transcript sequencing (NET-seq) maps nascent RNA transcripts in yeast by precipitation of a FLAG-tagged polymerase (Churchman and Weissman 2012). ChIP-seq of RNAPII can provide some evidence of occupancy of different CTD modifications. However, these techniques do not provide information directly connecting RNAPII phosphorylation state to RNA processing events.

1.3.3 mNET-seq

Our laboratory has recently developed a technique to generate genomic profiles linking CTD modification and nascent RNA transcription/processing. This is called mammalian Nascent Elongating Transcript sequencing (mNET-seq). It is a protocol to allow for isolation of RNAPII isoforms and associated RNA with endogenous RNAPII antibodies, and can provide information on processing events such as splicing and miRNA processing (Nojima et al. 2015; Nojima et al. 2016; Schlackow et al. 2017).

The mNET-seq protocol (Nojima et al. 2016) involves gentle lysis of cells to collect nuclei, and then precipitation of chromatin with elongating polymerase still bound in 1 M Urea, 300 mM NaCl. This chromatin fraction is then solubilized by a micrococcal nuclease (MNase) digestion. MNase will digest any exposed, unprotected DNA or RNA, and so only RNA protected by protein binding, such as that within the RNAPII active site is left intact. Polymerase and associated RNA/protein complexes are then purified using RNAPII antibodies that can be specific for different subunits or CTD modifications. 5' hydroxyl groups formed by MNase digestion are phosphorylated with Polynucleotide 5'-hydroxyl kinase (PNK) to allow directional sequencing, and the RNA is then extracted. During library preparation, the 3' hydroxyl from the last nucleotide in the RNAPII active site is ligated to the 3' linker, prior to cDNA synthesis and PCR amplification. Library sequencing allows single nucleotide resolution of the position of RNAPII with different post-translational modifications. In addition, other 3' OH ends, such as those generated in splicing and miRNA cleavage and then protected from

digestion within the co-immunoprecipitated spliceosome or the microprocessor respectively, can also be detected by mNET-seq. This gives information on which CTD modifications are associated with these co-transcriptional processing events. mNET-seq does not employ any cross-linking. It therefore avoids potential introduction of artifacts by crosslinking steps, in procedures such as ChIP (formaldehyde) and CLIP (UV).

mNET-seq can also be adapted to study the protein complexes that co-immunoprecipitate with transcribing RNAPII. After RNAPII pull-down, samples can be processed by Filter Aided Sample Preparation (FASP) and analysed by Mass Spectrometry (MS) to detect proteins significantly associated with different CTD modifications.

1.3.4 Established roles and profiles for different RNAPII modifications

Probably the best-characterised CTD residue modifications involved in co-transcriptional processing are phosphorylations of serine residues 2 and 5 (S2P and S5P). A general picture has emerged of their role in the transcription cycle, with evidence largely derived from budding yeast data (Buratowski 2009).

RNAPII is believed to be loaded onto chromatin at gene promoters in an unphosphorylated state (Phatnani and Greenleaf 2006). S5 is subsequently phosphorylated, for example by CDK7, which is part of the transcription initiation factor TFIIH. S5 is therefore phosphorylated as part of transcription

initiation at the TSS, and promotes the recruitment of capping enzymes that modify the 5' end of mRNA. Phosphorylation of S5 also disrupts the Mediator complex binding to RNAPII, releasing factors that stabilize complexes associated with the promoter (Sogaard and Svejstrup 2007). Elongating RNAPII is then phosphorylated at S2 residues by the CDK9 subunit of P-TEFb (Laitem et al. 2015).

Whilst RNAPII transcribes along genes in the elongation phase, S5 phosphorylation is removed by Ssu72 (Komarnitsky et al. 2000; Krishnamurthy et al. 2004; Reyes-Reyes and Hampsey 2007). This means S5P signal tends to peak at gene 5' ends, whilst S2P signal tends to increase towards the 3' end. Dephosphorylation of S5 and phosphorylation of S2 helps to recruit cleavage and poly-adenylation factors such as the CPSF and CF complexes necessary for transcriptional termination and mRNA poly-adenylation (Licatalosi et al. 2002; Meinhart and Cramer 2004; Davidson et al. 2014). S2P is eventually dephosphorylated by Fcp1 (Meinhart et al. 2005).

Recent mNET-seq data has however, suggested that the above picture is at best somewhat oversimplified when applied to mammalian transcription (Nojima et al. 2015). Low or unphosphorylated RNAPII signal was detected over the TSS region, with a noticeable lack of S5P signal. Most S5P signal appears to be distributed over the gene body rather than the 5' end, and in particular this modification is associated with spliced exons and especially the 5' splice site. This exonic signal is probably due to the occurrence of a second 3'-OH generated during splicing, and protected from MNase digestion by the spliceosome.

Consistently with the above model, S2P is predominantly associated with gene ends (transcription end sites or TES), but is also present across the 5' region of genes and its distribution can be skewed towards the TSS by depletion of cleavage and poly-adenylation (CPA) factors. Whilst some of the differences between these and earlier results could be attributed to the relative advantages of mNET-seq outlined above, it is also likely that the CTD code is significantly different in Mammalia compared to lower eukaryotes. Both the increase in heptad repeat number and variation within repeats allows for increased specialization of certain marks for different processes. The low number of introns in budding yeast may mean S5P association with splicing is less necessary or less detectable in *S. cerevisiae*.

There may be a possibility that mNET-seq does not detect the promoter proximal peak of S5P seen using other methods as it cannot detect signal from the first ~20 bases of transcription as the RNA has yet to emerge from the exit channel and so cannot be phosphorylated by PNK. Very small RNA moieties would also be lost during gel size selection. However, other modification signal peaks are visible at the 5' end of genes, and so mNET-seq would likely detect S5P too if it was present.

1.3.5 CTD Tyrosine 1 and Threonine 4 phosphorylation

The consequences of S2 and S5 phosphorylation have been extensively studied, but less is known about the role of Y1 and T4 phosphorylation, especially in higher eukaryotes. It is becoming increasingly clear that the roles of these

residues differ considerably between yeast and mammals (Yurko and Manley 2017).

Mammalian Y1P RNAPII signal was shown to accumulate in an antisense direction at promoters and enhancers (Descostes et al. 2014; Hsin et al. 2014a). It may also function in maintaining CTD stability (Hsin et al. 2014a). In yeast however, Y1P signal pattern appears similar to S2P, rising across the gene body, but diminishing before the 3' end (Mayer et al. 2012). It is proposed that Y1P inhibits the recruitment of termination factors before it diminishes around the TES. Y1P in yeast also may recruit the histone chaperone Spt6, which promotes the introduction of histone modifications (H3K36me3) to enhance mRNA processing, and suppression of aberrant transcription initiation from cryptic sites and transcription read-through (Xiao et al. 2003; Carrozza et al. 2005; Joshi and Struhl 2005; Wagner and Carpenter 2012; Burugula et al. 2014).

Y1 is phosphorylated by mammalian c-Abl *in vitro* (Baskaran et al. 1993; Baskaran et al. 1996; Baskaran et al. 1999). There is no known homologue in yeast. The main yeast Y1P phosphatase appears to be Glc7 (Schreieck et al. 2014).

It has been suggested that T4P is necessary for efficient 3' processing of non-poly-adenylated histone transcripts in DT40 chicken cells (Hsin et al. 2014b). This is different to yeast where histones are poly-adenylated. A T4A mutant (substitution in 48 repeats) was shown to be lethal in mammals and to produce a genome-wide elongation defect with elevated promoter proximal signal and

lower 3' RNAPII levels. The same mutation is non-lethal in yeast, where T4 mutation leads to a global downstream shift of pA site selection (Stiller et al. 2000). ChIP signals for T4P are high at 3' end in mammals. T4 phosphorylated RNAPII is also found to associate with centrosomes in M phase in HeLa cells (Hintermair et al. 2016). The distribution is different in yeast, with lower signal after the pA site and suspected roles in nucleosome remodelling/eviction at promoters of certain gene subsets (PHO) (Mayer et al. 2012; Rosonina et al. 2014). There remains uncertainty about how prevalent T4P is in yeast – estimates vary from 2 to 100% levels of S2P (Schuller et al. 2016; Suh et al. 2016).

Several potential kinases have been identified that may phosphorylate threonine 4 in vertebrates. CDK9 inhibition (Hsin et al. 2011) and siRNA depletion (Hsin et al. 2014b) decrease T4P levels. Polo-like kinases Plk1 and 3 have also been shown to phosphorylate T4 residues *in vitro* (Hintermair et al. 2012; Hintermair et al. 2016). Whether these phosphorylate T4 under similar or different conditions is unknown. Fcp1, as stated earlier a phosphatase of S2P, has also been shown to dephosphorylate T4P (Hsin et al. 2014b).

Applying mNET-seq analysis to Y1 phosphorylation shows peaks of signal at the promoter and TES in mammalian mRNA genes (figure 1.6) (Schlackow et al. 2017). Y1P may therefore be a marker of promoter-proximal and perhaps 3' pausing. A strong T4P signal is seen downstream of the pA site, with very little or no signal in the body of the gene (in contrast to lincRNA). This is consistent with ChIP-seq results and may suggest that T4P acts as some form of termination

signal. Following knockdown of CPSF73, which produces a termination defect evident in Chromatin RNA sequencing (ChrRNA-seq) datasets, the T4P peak visible in mNET-seq shifts downstream. This may indicate a form of CPSF73-independent “failsafe” termination.

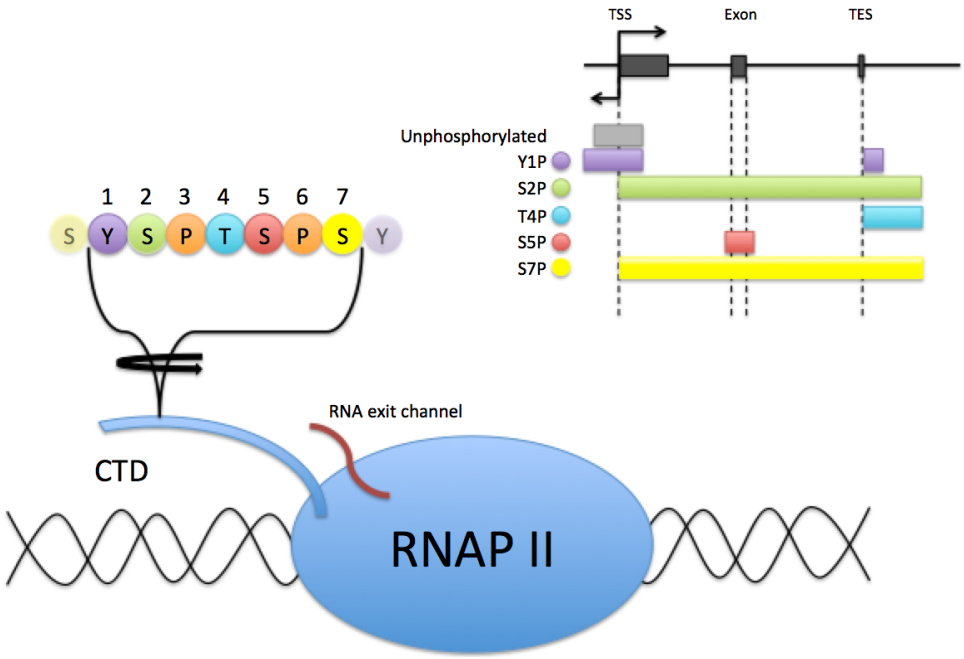


Figure 1.6 Schematic of one canonical heptad repeat in the RNAPII CTD, and summary of phosphorylation signals of corresponding residues across a gene as seen with mNET-seq. The C-terminal “tail” of RNAPII extends from the enzyme in proximity to the RNA exit channel, ideally positioned for orchestrating assembly of RNA processing factors. Phosphorylation of different CTD residues varies over a gene body. Unphosphorylated CTD and Y1P signal is most prominent at gene TSSs. S2P and S7P signal are seen across genes. S5P signal is associated with exons due to a second 3’OH generated during splicing. T4P signal is highest in transcription termination regions. CTD profile data can be found in (Schlackow et al. 2017)

In seeking to further define the roles of CTD Y1P, our laboratory has recently used the mass spectrometry variation on mNET-seq to study the protein factors associated with this mark, in comparison to total RNAPII, S5P and S2P. S5P has been found to be significantly associated with splicing factors when compared to

other marks. Y1P fractions, in contrast, are significantly enriched for a number of factors with roles in chromatin remodeling and NER DNA repair (Takayuki Nojima, unpublished data). This may suggest a role in either maintaining open chromatin at the TSS, or in remodeling chromatin following DNA damage.

1.4 R-loops

1.4.1 Introduction to R-loops

R-loops are triple-strand structures consisting of an RNA:DNA hybrid and a displaced single strand of DNA. They are generally believed to form during transcription, when the nascent RNA strand invades the complementary upstream DNA duplex, forming a stable, thermodynamically favourable DNA:RNA duplex with the template strand (Shaw and Arya 2008).

R-loops were traditionally considered to be rare transcriptional by-products, with deleterious effects on genomic stability. They have, however, been recognised as essential intermediates in certain cellular events, including immunoglobulin class-switching (Yu et al. 2003; Chaudhuri and Alt 2004), and specific replication of bacteriophage (Kreuzer and Brister 2010), *E. coli* (Kogoma 1997), and mitochondrial DNA (Itoh and Tomizawa 1980; Baldacci et al. 1984; Pohjoismaki et al. 2010; Holt and Reyes 2012). There is evidence that R-loops function in transcription regulation, including maintenance of demethylation at CpG island promoters (Ginno et al. 2012; Ginno et al. 2013), and transcription termination at a subset of mammalian genes. It is still likely, however, that in many cases DNA:RNA hybrids are deleterious side-effects that may represent

failures to fully protect susceptible regions of nascent RNA through co-transcriptional processing. Mutants exhibiting increased R-loop formation are associated with hyper-recombination phenotypes and increased genomic instability. This may be due to ssDNA vulnerability or DNA replication fork interference (Huertas and Aguilera 2003; Li and Manley 2006; Gan et al. 2011; Aguilera and Garcia-Muse 2012).

1.4.2 Recognition of R-loops

The RNA:DNA hybrid is predicted to form a structure intermediate between the A and B forms of DNA (Roberts and Crothers 1992), with differences in major and minor groove conformation significant enough to be distinguishable from either of the standard forms. A number of dedicated factors with the ability to recognise this formation have been identified. These include a number of proteins involved in resolving R-loops and preventing their potentially hazardous accumulation in the cell. The RNase H family specifically recognise RNA:DNA hybrids and cleave the RNA strand (Cerritelli et al. 2003; Cerritelli and Crouch 2009). This has also proven useful in studying R-loops, as the susceptibility of a nucleotide structure or cellular process to RNase H but resistance to RNase A (which cleaves ssRNA) suggests the presence of R-loops. The unique structure of the hybrid has also allowed the raising of an antibody (S9.6) to specifically detect the structures, enabling techniques such as DRIP (DNA:RNA Immuno-Precipitation) and immunofluorescence imaging to be used to study the phenomenon (Boguslawski et al. 1986). Other detection techniques

include electron microscopy and mutation profiling using sodium bisulphite, which converts C to U only in ssDNA.

1.4.3 Features favouring hybrid formation

Although not dependent on a particular nucleotide sequence, R-loop formation is favoured by a number of sequence biases. Firstly, R-loops are more prevalent at regions of high GC content. However, not all GC-rich regions are equally favourable - the sequences that tend to form RNA:DNA duplexes are homopurine-homopyrimidine stretches, which have a predicted propensity to form structures such as G-quadruplexes (Huppert et al. 2008). These regions also tend to have most guanine nucleotides in the (displaced) non-coding strand (GC skew). This may be due to the formation of G-quadruplexes in the displaced ssDNA strand stabilising the triple-strand structure, but the increased stability of an RNA:DNA duplex with an rPu/dPy configuration as compared to rPy/dPu is also likely to be a contributing factor (Duquette et al. 2004). G-rich regions alone may not be sufficient - G clusters are known to be important in the initiation of R-loop formation (Roy and Lieber 2009), perhaps due to stability being critical during annealing of the first few nucleotides. R-loops have also been shown to be associated with the transcription of certain disease-associated trinucleotide repeat expansions (Reddy et al. 2011; Groh et al. 2014), namely CCG, CGG, CAG, CTG and GAA repeats, with the most thermodynamically stable having the greatest R-loop forming capacity.

Links have also been demonstrated between negative super-coiling and R-loop formation in *E. coli,* yeast and mammals (Drolet et al. 1994; Drolet 2006). Topoisomerase mutant strains show increased R-loop formation in a transcription-dependent manner, and this correlates with decreased transcription efficiency and replication impairment (Baaklini et al. 2004; El Hage et al. 2010). Similarly, temperature sensitive gyrase reactivation can impair RNA synthesis unless R-loop resolving enzymes are also over-expressed. The link appears reciprocal, with increased R-loop formation increasing hyper-negative super-coiling. Negative super-coiling appears to allow the formation of less thermodynamically stable hybrids, perhaps compensating for the energetically costly DNA remodelling and displacement. DNA nicks also favour *in vitro* R-loop formation, again perhaps altering the thermodynamic cost of formation (Roy et al. 2010).

A number of mutant strains have been isolated with increased R-loop accumulation. Many of these mutations have been shown to map to factors involved in the different stages of mRNA biogenesis. Proteins involved in mRNP packaging and processing are particularly highly represented, for example loss of the THO/TREX (Huertas and Aguilera 2003) complex or ASF/SF2 splicing factor activities (Li and Manley 2005). This has led to the suggestion that correct mRNP formation is crucial to prevent the re-annealing of the nascent mRNA to the template, by a physical barrier between the otherwise favourable base-pair interactions. It has even been suggested that prevention of R-loops is a major reason for transcription-translation coupling in bacteria (Gowrishankar and Harinarayanan 2004). However, at present the alternative explanation, that loss

of processing factors increases mRNA half-life, increasing steady-state levels and the likelihood of RNA invasion of the DNA duplex, cannot be fully discounted.

Little is known about protein factors that directly assist strand invasion, but recent evidence suggests a role for the RecA/Rad51 family of strand exchange factors characterised for their role in homologous recombination, in causing R-loop formation (see below).

1.4.4 R-loops and Transcriptional Regulation

There is now a growing body of evidence indicating R-loops form functionally in several areas of transcriptional regulation. Perhaps their best understood role is in the pause-type transcriptional termination in mammalian and yeast genes.

Transcription pause sites are GC rich and enriched for G repeats, favouring R-loop formation, and R-loops have been detected in the 3'UTR/pause regions of a number of genes (Skourti-Stathaki et al. 2011; Skourti-Stathaki et al. 2014; Wahba et al. 2016). The yeast RNA:DNA helicase Sen1 and the human orthologue SETX have been shown to resolve R-loops, which accumulate at these positions in genes on loss of the helicase (Mischo et al. 2011; Yuce and West 2013). Sen1 has been shown to interact with the Rpb1 subunit of RNAPII and to co-operate with Rat1 to cause efficient termination. Loss of SETX leads to transcriptional read-through in mammals, suggesting RNA:DNA hybrids must first be resolved before efficient termination of transcription can occur. Removal of R-loops by RNase HI over-expression also leads to transcriptional read-through, suggesting

R-loops may play a role in pausing/slowing the polymerase before being resolved, allowing Xrn2/Rat1 to cause termination (Skourti-Stathaki et al. 2011). However, the order of some events - whether R-loops cause pausing, or pausing increases the likelihood of hybrid formation - is still somewhat unclear. Recent work on the *β-actin* gene as a model example of pause-dependent termination has suggested a role for R-loops in establishing repressive heterochromatin marks at termination sites. R-loop formation over pause elements downstream of the pA site facilitates antisense transcription. This leads to dsRNA production that is processed by Dicer. Argonaut proteins (Ago1 and 2) and G9a/GLP methyltransferases are then recruited, leading to Histone H3 Lysine 9 (H3K9) dimethyl mark formation over this termination region and HP1γ recruitment. R-loops and H3K9 di-methylation combine to facilitate RNAPII pausing prior to termination (figure 1.7) (Skourti-Stathaki et al. 2014).

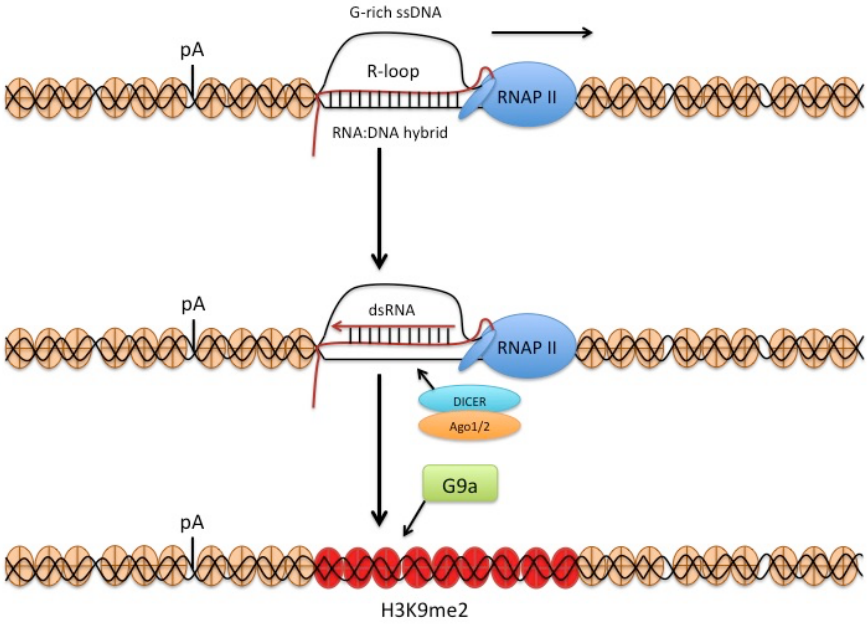


Figure 1.7 A proposed model for R-loop function in the pause-type termination mechanism. Emerging nascent RNA re-anneals to the DNA template, causing displacement of the G-rich ssDNA strand. Through some as yet unclear mechanism, this triggers dsRNA production. This acts as a substrate for Dicer and Argonaute (Ago1/2) proteins and leads to the recruitment of histone methyltransferases such as G9a, causing localized increases in H3 lysine 9 dimethylation, a repressive mark. This works synergistically with the G-rich sequence and R-loops to re-enforce pausing. Adapted from (Skourti-Stathaki et al. 2014)

1.4.5 Deleterious effects of R-loops

Many studies indicate that R-loops increase genomic instability (Gonzalez-Aguilera et al. 2008; Wahba et al. 2011; Chernikova and Brown 2012; Stirling et al. 2012), as suggested by a number of markers for loss of genomic integrity (such as the histone variant γH2AX and 53BP1). Indeed, an siRNA screen for factors involved in maintenance of genomic stability identified a large number of factors involved in transcription elongation, splicing, mRNP packaging, and export that likely relate to R-loops (Paulsen et al. 2009). Damage foci were

decreased in cells stably expressing RNase HI, again implicating R-loops in DNA damage. Many transcription-dependent damage foci and *in silico* predicted "R-loop forming sequences" map to proto-oncogenes, making a mechanistic understanding all the more important (Wongsurawat et al. 2012). Exactly how R-loops would cause DNA lesions is in many cases still unclear, although a number of explanations have been suggested. Firstly, the increased vulnerability of the displaced non-template strand to chemical modification and/or breakage, for example by AID (which can trigger a number of DNA repair pathways), may contribute to first single-strand and later, if unresolved, double-strand breaks.

Alternatively, retardation of the transcription machinery, either by R-loop formation or more directly through loss of RNA processing factors, may increase the likelihood of collisions with the replication fork (Prado and Aguilera 2005; Wellinger et al. 2006; Gottipati et al. 2008; Gan et al. 2011), which can cause DSBs through backtracking (and formation of a "chicken-foot" structure) (Mirkin and Mirkin 2007; Belotserkovskii et al. 2010). In support of this view, R-loop related DNA damage appears more common over long genes (Helmrich et al. 2011), and when R-loop repressing factors are lost, γH2AX appears more commonly at gene-rich regions. Collisions of the replication machinery with R-loops have been shown to directly impair replication fork progression in bacteria and eukaryotes, and this appears to lead to chromosomal rearrangement. Again, caution is needed in interpreting some of these results, as more direct effects of loss of mRNA processing factors on DNA damage checkpoints and responses have been suggested. G quadruplexes may also directly contribute to

transcription-related instability (Lopes et al. 2011). Hyper-recombination associated with these mutants has also yet to be shown to be S-phase specific.

As well as SETX, a number of other helicases have been found to resolve potentially deleterious R-loops (Sollier et al. 2014; Calo et al. 2015; Bhatia et al. 2017; Song et al. 2017). These include Aquarius, DDX19, 21 and 23, the latter of which is activated by phosphorylation by splicing-pathway kinases upon pausing of RNAPII (Sridhara et al. 2017).

1.4.6 Rad51 and R-loops

As mentioned above, one recent study has implemented the homologous recombination (HR) factors Rad51 and its activator Rad52 in catalysing strand invasion by RNA. Rad51 is known to function in the process of homologous recombination during both meiotic chromosomal crossover and DNA double-strand break repair. Rad51 polymerises on ssDNA as a right-handed helical filament. Bound ssDNA is extended, as compared to the B-form of DNA, with Rad51 units binding the phosphate backbone with bases exposed in triplet clusters (Pellegrini et al. 2002; Xu et al. 2017). This is thought to facilitate the search for complementary bases and catalyses the broken ssDNA strand invasion of the double-stranded homologous region (Sung and Klein 2006). DNA polymerases can then extend the broken strand in an error-free process using the homologous chromosome as a template. Rad52 facilitates the loading of Rad51 onto ssDNA in yeast (New et al. 1998; Shinohara and Ogawa 1998) and later annealing of the second ssDNA end (Sugiyama et al. 1998). In mammals,

BRCA2 fulfils much of this role, although Rad52 can still facilitate Rad51 loading in the absence of BRCA2 (Feng et al. 2011; Liu and Heyer 2011).

The Koshland laboratory showed that these proteins are necessary to cause lesions in a number of RNA processing mutants associated with R-loop formation (Wahba et al. 2013). Using a yeast artificial chromosome model, they were also able to show that Rad51 can catalyse *trans* R-loop (forming away from the initial transcription point, e.g. on homologous chromosomes) as well as *cis* R-loop formation and DNA damage. This work follows on from *in vitro* studies suggesting the bacterial Rad51 homologue, RecA, can form R-loops using mature RNA in the absence of transcription (Hong et al. 1995). Such evidence implies a broader view of R-loop formation is needed, and that more RNA species than previously considered are capable of causing genomic damage. For example the RNA:DNA hybrid formed by CRISPR/Cas9 is in fact a *trans* R-loop. The Rad51 findings also suggest the scenario that DNA damage repair factors such as Rad51 may be capable of causing DNA damage if misregulated/over-expressed.

1.5 DNA damage

1.5.1 Ultraviolet radiation

Ultraviolet (UV) radiation produces bulky adducts – cyclobutane pyrimidine dimers (CPDs) and 6-4 photoproducts – which if untreated can lead to DNA and RNA polymerase stalling, replication errors and a number of types of cancer, most famously melanoma (Kuraoka et al. 2003; Mei Kwei et al. 2004).

1.5.2 Nucleotide excision repair

The only repair pathway in placental mammals that can repair these bulky lesions is the Nucleotide excision repair pathway (NER). Here, DNA damage, specifically helix distortions and bulky adducts, are recognized by a number of proteins. The transcription factor TFIIH (Yokoi et al. 2000; Volker et al. 2001) and a number of Xeroderma Pigmentosum (XP) factors are then recruited to the site of damage, unwind DNA around the lesion, and make single-stranded cuts either side of the adduct (Tapias et al. 2004; Compe and Egly 2012). When the lesion strand is removed, a ssDNA section of 22-30 nucleotides is left (Marteijn et al. 2014). The complementary strand is then repaired by DNA polymerases δ, ε and κ, depending on the cell cycle stage of the cell (Moser et al. 2007; Ogi et al. 2010). A number of DNA ligases then seal nicks.

NER is divided into two main sub-pathways, which can be distinguished by the method by which they sense DNA damage (figure 1.8). In global genomic repair (GG-NER), a number of "damage sensors" recognize distortions. These include the DNA-damage binding proteins (DDB1/2) (Chu and Chang 1988; Wakasugi et al. 2002; Scrima et al. 2008) and the XPC-Rad23B complex (Sugasawa et al. 1998). These then recruit repair proteins including CUL4a (Groisman et al. 2003). In transcription-coupled repair (TC-NER), RNA polymerase stalls at lesions during gene transcription, and this recruits Cockayne Syndrome proteins (CSA/B), leading to backtracking of the polymerase (Sigurdsson et al. 2010) and recruitment of repair proteins (Fousteri et al. 2006; Gaillard and Aguilera 2013; Weems et al. 2017). How exactly stalled polymerase is initially recognized is still

unclear. The downstream repair pathways after recognition are very similar in both pathways (some proteins such as DDB2 are specific to GG-NER) (Marteijn et al. 2014). In the event of TC-NER failure to remove adducts, stalled RNA polymerase may also be removed by ubiquitinylation and degradation (Woudstra et al. 2002; Li et al. 2007; Wilson et al. 2013).

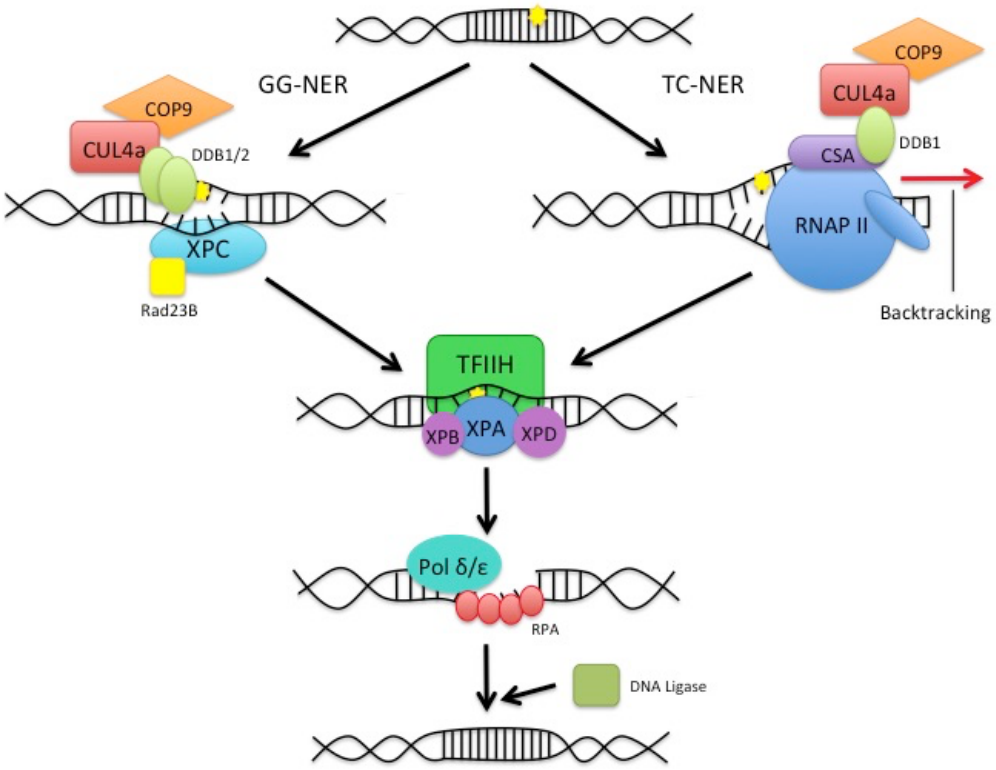


Figure 1.8 Schematic of the Nucleotide Excision Repair, showing the differences in sensing damage in global and transcription coupled pathways. Bulky DNA lesions are sensed by either DDB1/2 and XPC/Rad23B (Global) or stalled RNAPII (transcription coupled). CSA causes back tracking of RNAPII, allowing access to the lesion for DNA repair factors. The downstream repair pathway involves TFIIH and XP protein recruitment, excision of the single strand around the lesion, binding of RPA to protect the remaining ssDNA, and new DNA synthesis by DNA polymerases and ligases. It is similar in both GG- and TC- pathways. Reviewed in (Marteijn et al. 2014).

It has been shown that adducts on the strand transcribed by RNA polymerase are repaired more quickly than those on the other strand, or in transcriptionally

silent regions (Gaillard and Aguilera 2013; Hu et al. 2015). TC-NER is also better at recognizing CPDs than GG-NER, and may also recognize lesions otherwise repaired by Base Excision Repair (Stevnsner et al. 2008). This highlights the importance of the TC-NER pathway in rapidly dealing with harmful lesions. Stalled RNA polymerase has also been shown to be able to trigger DNA-damage responses even in the absence of damage itself, and may constitute a major general signal of DNA damage (Derheimer et al. 2007; Lindsey-Boltz and Sancar 2007).

A second role for some NER factors and DNA damage signal proteins has been suggested, where XP proteins are recruited to RNAPI and II promoters to maintain chromatin conformations suitable for transcription initiation (Le May et al. 2010; Bunch et al. 2015).

1.5.3 UV and transcription

In addition to RNAPII stalling at lesion sites, UV damage can also cause a global repression in gene expression, in undamaged genes as well (Mayne and Lehmann 1982; Rockx et al. 2000; Proietti-De-Santis et al. 2006). This is particularly marked in CSB mutant cells. This global effect is specific to bulky adduct damage such as produced by UV, and is not seen in other forms of DNA damage such as double-strand breaks (DSBs).

UV irradiation triggers hyper-phosphorylation of RNAPII or at least a reduction in the hypo-phosphorylated form. This suggests a repression of the initiation

stage of transcription (Rockx et al. 2000). H4 acetylation (thought to be a mark of open chromatin) is also deficient around housekeeping gene promoters in CSB mutant cells post UV treatment. This is one aspect of a general increase in chromatin dynamics and recycling post UV irradiation (Dinant et al. 2013; Hauer et al. 2017).

GRO-seq data suggest UV irradiation affects the speed of RNAPII elongation, and therefore the choice/co-ordination of RNA processing events, for example splicing and alternative pA site selection (Munoz et al. 2009; Andrade-Lima et al. 2015; Munoz et al. 2017; Williamson et al. 2017). Alternative last exon selection often leads to a switch to production of shorter isoforms after irradiation. Significant shutdown in transcription and reduction of elongation speed to as little as 5% WT levels is seen within 25 minutes of irradiation, with full recovery after 24 hours.

1.5.4 γH2AX as a marker of DNA damage

Phosphorylated H2AX is a well-known marker of DNA damage (Seo et al. 2012; Sharma et al. 2012). H2AX can be phosphorylated at two positions, S139 and Y142. S139 is phosphorylated (γH2AX) by the ATM kinase at the site of damage (particularly double strand breaks but also at sites of NER) (Takahashi and Ohnishi 2005; Marti et al. 2006). It acts as a signal recruiting repair factors to damage foci (Paull et al. 2000). γH2AX phosphorylation can spread into large domains, but propagation has been found to be repressed in sites of active transcription (Iacovoni et al. 2010). Y142 is constitutively phosphorylated by

BAZ1b (WSTF), and dephosphorylated upon damage by the EYA phosphatases. Loss of either kinase or phosphatase has been shown to increase cell death upon damage, and Y142 phosphorylation is believed to act in the pathway deciding between repair and apoptosis in specific cell types, depending on levels of damage (Cook et al. 2009). There is evidence of crosstalk between the pathways downstream of the two phosphorylation sites – phospho-Ser139 damage foci are decreased in Y142A mutants and BAZ1b knockdowns (Xiao et al. 2009; Brown et al. 2012).

1.6 Aims

In this thesis I set out to better understand the interconnectedness of transcriptional regulation with both DNA damage repair pathways and chromatin configuration. The following chapters will investigate the hypothesis that Rad51 functions in mammalian R-loop production during transcription, as well as the functions of the Y1P and T4P marks in regulation of stages of the transcription cycle, and follow up on data suggesting links between the CTD, UV damage, and the nucleotide excision repair pathway. They will also describe modifications made to the mNET-seq protocol to aid in these purposes.

Chapter 2

Materials and Methods

2. Materials and Methods

2.1 Tissue culture

2.1.1 Growth and maintenance

All cell lines were grown at 37°C with 5% CO_2.

HeLa and U2OS – DMEM, 10% FBS, (optional), 100 units/ml Penicillin/Streptomycin. Cells passaged at 80-90% confluency.

eHAP – IMDM, 10% FBS, 100 units/ml Penicillin/Streptomycin. Cells passaged at 60-70% confluency to avoid producing diploids.

2.1.2 Plasmid Transfections and RNAi

2.1.2.1 HeLa

Transfection of HeLa cells with GFP-RNase H1 plasmid was carried out as described previously (Skourti-Stathaki et al. 2011), using Lipofectamine 2000 (ThermoFisher Scientific). For $1x10^6$ cells, 20 μg GFP-RNase H1 vector was transfected and cells were harvested 32 hr after transfection.

siRNA knockdown of Rad51 was carried out as described previously (Skourti-Stathaki et al. 2011). Cells at 80% confluency in 24 well plates were placed in media without Penicillin/Streptomycin. In one tube, 3 μl Lipofectamine RNAiMAX (ThermoFisher Scientific) was incubated with 12 μl Opti-MEM for

7 min. In a separate tube, 3 µl of 20 mM siRNA duplex/buffer was mixed with 47 µl Opti-MEM. These mixtures were then combined and incubated, RT, 25 min. Then, Opti-MEM was added to a final volume of 100 µl and this mixture was added to a well. After overnight incubation, cells were split onto 6 well plates. The next day, a second transfection as above was performed. At this point cells could be split onto 10 cm^2 plates if needed. siRNA treated cells were harvested 24 hr after the second transfection.

siRNA knockdown of CPSF73, BAZ1b, CUL4a, and SNF2h were carried out as in (Nojima et al. 2015). For $1x10^6$ cells, 30 µl RNAiMAX and 100 µM siRNA duplex in 2 ml Opti-MEM was incubated for 25 min, then added to cell media, and cells were harvested after 72 hr.

2.1.2.2 eHAP

For eHAP cell transfection, a 1:4 ratio µg DNA : µg TransIT-LT1 transfection reagent was used. For $2x10^5$ cells, 0.5 µg Cas9:gRNA plasmid in 50 µl OptiMEM was added to cells in 500 µl fresh IMDM media. 24 hr after transfection, media was replaced with IMDM containing 1 µg/ml puromycin for selection of transient transfection. 24-48 hr later, puromycin media was removed to allow cells to recover for approximately 3 days. Cells were then diluted to an average of 0.5 cells per well for clonal selection.

2.1.2.3 Clonal selection

Colonies were lysed with Squish buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 25 mM NaCl, 200 µg/ml Proteinase K, 30 min, 65°C, followed by inactivation at 95°C, 2 min) and 1 µl aliquots used in a Taq screening PCR using deletion check primers to test for loss of target region. The PCR products were analysed by 1% agarose gel electrophoresis, followed by ethidium bromide staining.

The area surrounding the target region in positive colonies was then amplified for sequencing by Phusion PCR. Sequencing was performed by Source Bioscience.

2.1.3 BO2 inhibitor treatment

Small molecule inhibitor BO2 (Sigma Aldrich) was dissolved in cell media to a final concentration of 10 µM and cells were incubated for 2 hr, as has been shown to inhibit ionizing radiation (4 & 5 Gy) induced Rad51 foci formation (Lodhia et al. 2015) (F. Esashi, personal communication).

2.1.4 UV irradiation

Cells were irradiated (20 J/m^2, 254 nm) using a Spectrolinker XL-1000, when 60-90% confluent. For most experiments, unless otherwise stated, cells were harvested 2 hr post UV treatment.

2.2 Isolation of chromatin, ChrRNA, nucleoplasmic and total RNA.

Total RNA was harvested by washing plates with PBS and then adding 1 ml TRIzol reagent (Invitrogen) and then isolated using manufacturer's instructions. Chromatin, chromatin-RNA and nucleoplasm RNA were isolated as described in (Nojima et al. 2015; Nojima et al. 2016). Cells were first washed with PBS and then pelleted, 500 g, 5 min. Cells were resuspended in 4 ml ice-cold HLB+N buffer (10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 2.5 mM $MgCl_2$, 0.5% NP-40) and incubated on ice for 5 min. This sample was then underlaid with 1 ml HLB+NS (HLB+N with 10% sucrose). Nuclei were then pelleted, 300 g, 5 min, and supernatant discarded. Nuclear pellets were then resuspended in 125 µl NUN1 buffer (20 mM Tris-HCl, pH 7.9, 75 mM NaCl, 0.5 mM EDTA, 50% glycerol) by pipetting and transferred to new tubes. 1.2 ml ice-cold NUN2 buffer (20 mM HEPES-KOH, pH 7.6, 0.2 mM EDTA, 200 mM NaCl, 7.5 mM $MgCl_2$, 1% NP-40, 1 M Urea + protease/phosphatase inhibitors) was added and samples were kept on ice with intermittent mixing by vortex for 15 min. Samples were then centrifuged, 17,000 g, 10 min, and the supernatant (nucleoplasm) and pellet (chromatin) separated.

RNA can be isolated from nuclear or chromatin pellets by the addition of 200 µl HSB buffer (10 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 50 mM NaCl), 3 µl TURBO DNase (2 U/L, ThermoFisher Scientific), and incubation with shaking at 37°C, 10 min, 1400 rpm (Eppendorf Thermomixer C). 200 µl Proteinase K/SDS solution (0.75 mg/ml Proteinase K, 0.4% SDS) was then added and samples incubated, 37°C, 1400 rpm, 10 min (Eppendorf Thermomixer C). RNA was then isolated

using acidic phenol:chloroform extraction, followed by TRIzol extraction to more fully remove any DNA contaminants.

RNA libraries were prepared using the Ribo-Zero™ H/M/R Kit (Illumina) to remove rRNA, and then the NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina®, and NEBNext® Multiplex Oligos.

2.3 PCR

2.3.1 RT-PCR (cDNA synthesis)

cDNA was made by reverse transcribing RNA using Superscript III Reverse Transcriptase (Invitrogen). 0.5-1 µg RNA was added to a 25 µl reaction containing 0.4 µM gene specific primers, or 75 mM random hexamers, 0.4 µM dNTPs, 40 µU RiboLock RNase Inhibitor (ThermoFisher Scientific), 5 µM DTT and 200 U enzyme. This reaction mixture was then incubated at 55˚C for 1 hr.

2.3.2 Standard PCR

Template was amplified using either Taq or Phusion High Fidelity DNA polymerases (NEB) according to manufacturers protocols. Primer annealing temperatures were set approximately 5˚C below primer melting temperatures.

2.3.3 qPCR

qPCR reactions were carried out using SYBRgreen qPCR mix (Qiagen) and a Corbett Rotorgene 3000 and software. 2 µl of DNA was used with 0.5 µM primers and 7.5 µl SYBRgreen mix in a 15 µl reaction. Reactions were performed in

triplicate. Genomic DNA of known concentrations was used to produce a standard curve for calculating concentrations.

2.4 Chromatin Immunoprecipitation

ChIP and DRIP experiments were carried out as described in (Skourti-Stathaki et al. 2011) with modifications.

For ChIP, $5x10^6$ cells were fixed by adding 1% formaldehyde to cell medium, and incubation, 37˚C, 15 min with mild shaking. Cross-linking was quenched by the addition of 125 mM glycine and incubation at 37˚C, 10 min. Cells were then washed using cold PBS and collected by scraping. Cells were lysed using 300 µl Cell Lysis Buffer (0.5% NP-40, 5 mM PIPES, pH 8.0, 85 mM KCl + protease inhibitors) on ice, 10 min. Nuclei were pelleted, 6000 g, and then resuspended in 400 µl Nuclear Lysis Buffer (25 mM Tris-HCl, pH 8.0, 5 mM EDTA, 1% SDS + protease inhibitors) for 10 min, on ice. Samples were sonicated in 1.5 ml tubes using a Bioruptor (Diagenode), medium power, 30 s pulses at 30 s intervals, 25 min. Samples were then centrifuged, 17,000 g, 10 min, to remove cell debris, and supernatant was added to 2.6 ml IP dilution buffer (16.7 mM Tris-HCl pH 8.0, 1.2 mM EDTA, 167 mM NaCl, 1.1% Triton X-100, 0.01% SDS + protease inhibitors). Samples were pre-cleared using 30 µl Dynabeads (ThermoFisher Scientific) (A/G or a 1:1 ratio depending on antibody) with rotation, 4˚C, 1 h. Beads were then removed and a 10% aliquot of sample kept for use as Input. Remaining sample was split equally into IP and NO IP fractions, and 5 µg of antibody was then added to IP fractions and incubated with rotation, 4˚C, overnight. 30 µl Dynabeads was then added and incubated with rotation, 4˚C, 1.5 hr. Beads were

then collected using a magnetic rack and washed with Wash Buffers A (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS), B (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS), C (10 mM Tris-HCl pH 8.0, 250 mM LiCl, 1 mM EDTA, 1% NP-40, 1% sodium deoxycholate) and three times with TE buffer. DNA-protein was eluted by 2x incubation with 250 µl Elution Buffer (100 mM $NaHCO_3$, 1% SDS). Cross-linking was reversed by the addition of 0.3 M NaCl and 30 µg RNase A, and incubation for 4-12 h, 65˚C. Samples were then treated with 20 µg Proteinase K with 40 mM Tris-HCl pH 6.5, 10 mM EDTA, 45˚C, 2 h. DNA was then purified using PCR clean up columns (Qiagen). Input, IP and NO fractions were then analysed by qPCR. NO signal was subtracted from IP and this divided by Input. Values are shown as % Input.

Where specified, a second cross-linker (EGS, ethylene glycol bis(succinimidyl succinate), a cross-linker with a 12-atom spacer arm)) was used: cells were first washed with PBS and then re-covered in PBS. EGS dissolved in DMSO to a concentration of 15 mM was then added to a final concentration of 1.5 mM. Cells were then incubated, room temperature, 30 min, mild shaking. Formaldehyde was then added and quenched as usual.

2.5 DNA:RNA Immunoprecipitation

DRIP was performed as in the ChIP protocol, but with some modifications. No cross-linking step was used, and after the addition of Nuclear Lysis buffer, 30 µg Proteinase K was added, and samples incubated at 55˚C, 3-5 h. DNA:RNA was isolated by ethanol precipitation, and resuspended in 400 µl IP dilution buffer.

The ChIP protocol was then followed from the sonication step, however without the de-crosslinking step.

2.6 Protein extraction

2.6.1 SDS buffer

Cell, nuclear and chromatin pellets were resuspended in 2x SDS loading buffer (100 µM Tris-HCl, pH 6.8, 4% SDS, 20% glycerol, 0.2% BPB, 200 µM DTT). Nucleoplasm supernatant was diluted 1:1 with the same buffer. Samples were boiled, 95˚C, 10 min (chromatin samples were boiled for 30 min - 1 hr).

2.6.2 Acid histone extraction

$5x10^6$ cells were pelleted at 500 g, 5 min. Cells were washed twice with PBS and resuspended in 1 ml TEB buffer (PBS, 0.5% Triton-X, protease/phosphatase inhibitors) and left on ice 10 min. Samples were spun, 500 g, 10 min, and pellets washed in 500 µl TEB, and spun again. Pellets were resuspended in 250 µl 0.2 M HCl, and rotated, 4^oC, 3-12 hr. Samples were spun at 10,000 g and supernatant collected.

2.7 Western Blot analysis

Equal volumes were loaded onto protein gels appropriate for size separation and RNAPII dimer resolution (4-12% Bis-Tris and 3-8% Tris-Acetate (Invitrogen), 4-15% Tris-Glycine (Bio-Rad) with appropriate running buffers (MOPS and MES, Tris-Acetate, Tris-Glycine buffers respectively). Proteins were transferred to

PVDF membrane using a wet transfer method, Tris-glycine transfer buffer. Membrane transfer was confirmed by Ponceau S staining.

For ECL detection, membranes were first blocked with 5% milk in PBS with 0.1% Tween (PBST), 1 hr, RT, shaking. Primary antibodies were diluted (1:1000 unless otherwise stated) in PBS with 2% milk and membranes were incubated with the dilution, 1.5 hr, RT, shaking. Membranes were then washed, 3x, PBS, 7 min, RT, shaking. Membranes were then incubated in horseradish peroxidase-coupled secondary antibody dilutions in PBS with 2% milk, 1 hr, RT, shaking. Membranes were then washed again 3x, PBS, 7 min, RT, shaking. Membranes were then air-dried and proteins visualised using ECL (Pierce).

For LI-COR detection, membrane blocking and antibody dilution both utilised 3% BSA in PBST. Fluorescent secondary antibodies (680 and 800 nm) and subsequent washes were incubated in the dark. Proteins were visualised using an Odyssey CLx and Image Studio software (LI-COR Biosciences).

2.8 MNase digestion and RNAPII/chromatin precipitation

Chromatin pellets were first washed in 100 µl MNase buffer (NEB) and centrifuged, 17,000 g, 30 s. Supernatant was removed and 100 µl MNase reaction mixture (1x MNase buffer, 1x BSA solution (NEB), 40 U/µl MNase) was added to each pellet. Samples were incubated with shaking, 37˚C, 2 min, 1400 rpm (Eppendorf Thermomixer C). Once the pellet was digested, the MNase reaction was stopped by the addition of EGTA, 25 mM final concentration. Samples were

then centrifuged, 17,000 g, 10 min and supernatant collected. When followed by pulldown, 10 µl aliquots were taken as Inputs.

Antibody conjugated beads were prepared as follows: 50 µl Protein A/G or sheep anti-mouse IgG Dynabeads (depending on antibody affinity) per $2x10^7$ cells were washed twice with ice-cold NET2 buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.05% NP-40). Beads were resuspended in 50 µl ice-cold NET2 buffer and a range of 5-20 µg antibody was added depending on antibody. Beads were then incubated, 4˚C, 1 hr – overnight. Beads were then washed 3x with ice-cold NET2 buffer to remove unconjugated antibody.

Antibody-conjugated beads were then combined with MNase solubilized chromatin and diluted 1:10 with NET2. Samples were incubated with slow rotation, 4˚C, 1 hr. Beads were then washed 8x with NET2 buffer.

For mass spectrometry protocols and western blot confirmation IPs no BSA was used in the MNase reaction mixture and the last two NET2 washes post-IP were performed without NP-40. Samples were then resuspended in 8 M Urea, 100 µM TEAB and processed by standard trypsin digestion/FASP protocol.

2.9 Mass spectrometry

2.9.1 FASP

Samples for mass spectrometry analysis were transferred to 10K filter columns and incubated in 200 µl Lysis Buffer (8 M Urea, 100 µM TEAB pH 8.0), 30 min,

RT, gentle shaking. Samples were centrifuged, 17,000 g, 30 min. Columns were washed with 200 µl Lysis buffer, 17,000 g, 30 min. 200 µl Reduction buffer (8 M Urea, 100 µM TEAB pH 8.0, 20 nM TCEP) was added to columns, with gentle pipetting to resuspend beads. Samples were incubated, 30 min, RT, gentle shaking, followed by centrifugation, 17,000 g, 30 min. 200 µl Alkylation buffer (8 M Urea, 100 µM TEAB pH 8.0, 50 mM Chloroiodoacetamide) was added to columns, which were incubated, 30 min, RT, in dark. Samples were centrifuged, 17,000 g, 30 min, and then washed twice with Wash Buffer (1 M Urea, 50 mM TEAB pH 8.0), 17,000 g, 30 min, to produce conditions suitable for trypsin. Columns were then transferred to new tubes and then incubated with 100 µl Trypsin Buffer (1 M Urea, 50 mM TEAB pH 8.0, 0.01 µg/µl trypsin), 37˚C, 30 min, gentle shaking. A further 200 µl Trypsin buffer was then added and samples incubated 37˚C, gentle shaking, overnight. Samples were then centrifuged, 17,000 g, 30 min, and flowthrough collected. Column membranes were washed with 200 µl 0.1% trifluoroacetic acid, 17,000 g, 30 min, and this flowthrough combined with the previous. Membrane was quickly washed (to avoid destruction of column) with 200 µl 50% acetonitrile in 0.1% TFA, and centrifuged, 17,000 g, 30 min, and flowthrough combined with the two previous. Samples were dried using a SpeedVac, 37˚C, >2 hr.

2.9.2 Processing and analysis

Samples were run on a Q Exactive Mass Spectrometer (ThermoFisher). The CPFP server was used for protein identification. Spectral Index scores were used for relative quantitation.

2.10 NetMNase-seq

2.10.1 Protocol

For chromatin isolation, digested chromatin or bead-bound polymerase/histones were treated with 100 µg/ml RNase A, 30 min, 37°C, 1400 rpm (Eppendorf Thermomixer C) followed by equal volume Proteinase K/SDS solution, 20 min, 37°C, 1400 rpm (Eppendorf Thermomixer C), and DNA-phenol extraction. Samples were then size separated on 6% TBE acrylamide gels (280 V, 20 min), visualised by EtBr, and bands of wanted size cut and purified using RNA/DNA elution buffer (1 M NaOAc, 1 mM EDTA), 37°C, 12 hr followed by column filtration with Whatman GF/D glass filters. DNA too large for library preparation was fragmented using the NEBNext® dsDNA Fragmentase kit.

DNA libraries were prepared using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® and NEBNext® Multiplex Oligos. Library quality and concentration was checked using 2100 Bioanalyzer (Agilent) with DNA High sensitivity chips. Deep sequencing was conducted by the WTCHG in Oxford using an Illumina HiSeq 4000, paired end, 75 or 150 bp reads.

2.10.2 Human genome annotation

Library bioinformatic analysis was performed by Dr Michael Tellier.

Non-overlapping transcription unit annotations from GENCODE v19 and a custom Python script were used to keep only non-overlapping units superior to 2 kb within a region 2.5 kb upstream of the TSS and downstream of the TES.

Annotations of different types of divergent promoter, for example PROMPT-mRNA, eRNA-eRNA and mRNA-mRNA were taken from (Chen et al. 2016). Divergent mRNA-mRNA and mRNA-lincRNA pairs were defined as transcription in opposing directions from independent promoters within 3 kb of each other.

2.10.3 Data processing

Reads were trimmed with Cutadapt v. 1.9.1 (Martin 2011) in paired-end mode to remove adaptor sequences. Sequences were then mapped to the human hg19 reference sequence with Bowtie2 v. 2.2.5. (Langmead and Salzberg 2012). Unmapped reads were removed with SAMtools v. 1.3.1. (Li et al. 2009). PCR duplicates were removed with Picard. Bam files were sorted, indexed and counted using SAMtools. Trackhubs in the UCSC browser were created by creating bedGraph files normalised to 100 million reads using the Bedtools v. 2.26.0 (Quinlan and Hall 2010) bedcov tool. These were used to create Bigwig files using the bedGraphtoBigWig tool.

For metaprofiles, RPKM normalised bigwig files were created using the deepTools2 v. 2.2.4 (Ramirez et al. 2016) bamCoverage tool. The metaprofile matrices were calculated with deepTools2 computeMatrix tool and obtained with the plotProfile tool. IP efficiency was calculated as IP read density/Input read density. Read density (FPKM) as IP - Input. The Regions of Interest were defined as: eRNAs: region -2 to +2 kb around the TSS; PROMPTs: TSS to +3 kb; lincRNAs: TSS to pA site; protein-coding genes: TSS to pA site. Metaprofiles were created with GraphPad Prism 6.

In order to identify potential details lost in averaging of all genes, *k*-means clustering was used to separate mRNA genes into five groups based on expression levels. This was determined by chromatin RNA-seq and mNET-seq data. Groups 4 and 5, containing unexpressed or very lowly expressed genes, were not included in metagene analysis.

2.11 Immunofluorescence

Cells were grown on coverslips in 6 well plates. Cells were first washed 2x in PBS. Cells were fixed in 4% PFA-PBS, 20 min, RT. Cells were then washed with PBS, RT, 5 min. Cells were permeabilised using PBS containing 0.5% Triton X-100, RT, 20 min. Cells were blocked in 3% BSA-PBS, RT, 20 min. Cells were incubated with primary antibodies diluted in 3% BSA-PBS, RT, 1 hr. Samples were then 4x washed using 0.05% Tween–PBS, RT, 5 min. Samples were then incubated in secondary antibodies diluted in 3% BSA-PBS, RT, in dark, 30 min - 1 hr. Cells were washed, 3x, 0.05% Tween–PBS, RT, dark, 5 min. Cells were then kept for 10 min in PBS, RT, dark, before being mounted on slides in 7.5 µl Vectashield with DAPI (Vector Laboratories), dried for 1 hr in the dark, and sealed with nail polish. Antibodies were used at the following concentrations: γH2AX – 1/500, Y1P CTD – 1/500, secondaries 1/250.

Samples were viewed using an Olympus FV1000 confocal microscope. Colocalisation correlation was assessed using FIJI (ImageJ) Colocalisation Coloc2 plugin, as described in (Costes et al. 2004), on selected nuclei. Spatial resolution: pixel = 207nm, PSF = 0.9 microns. DAPI staining was used to define nuclear threshold.

2.12 Antibodies

Protein	Antibody	Company	Specific protocols
RNAPII			
PolII (unphosphorylated CTD)	8WG16, MMS-126R	BioLegend	WB, ChIP
PolII, RBP1 subunit amino terminus	H-224, sc-9001	Santa Cruz Biotechnology	WB, ChIP
PolII, RBP1 subunit, amino terminus	N20, sc899	Santa Cruz Biotechnology	WB
PolII CTD	MABI0601	MAB Institute	WB, ChIP, IP
PolII CTD Y1P	clone 3D12, MABE350	Millipore	WB, ChIP, IP, IF
PolII CTD T4P	clone 6D7, 61361	active motif	WB, ChIP, IP
Histones			
H3	ab1791	abcam	
H3K4me3	ab8580	abcam	
H3K9me2	ab1220	abcam	
H3K9me3	ab8898	abcam	
H3K36me3	ab9050	abcam	
H4	ab17036	abcam	
HP1α	ab77256	abcam	WB, ChIP
HP1α	clone 15.19s2, #05-689	Millipore	WB, ChIP
HP1γ	ab11174	abcam	
γH2AX (pS139)	[EP854(2)Y] ab81299	abcam	WB
γH2AX (pS139)	[9F3] ab26350	abcam	WB
γH2AX (pS139)	ab2893	abcam	WB, IF
γH2AX (pS139)	clone JBW301, #05-636	Millipore	WB, IF, ChIP
γH2AX (pY142)	ab94602	abcam	
Transcription factors			
CPSF68	A301-358-M	Bethyl Laboratories	
CPSF100	A301-582A	Bethyl Laboratories	
SPT6	C-6, sc-393938	Santa Cruz Biotechnology	
PAF1	A300-172A	Bethyl Laboratories	
LEO1	A300-175A	Bethyl Laboratories	
CTR9	A301-395A	Bethyl Laboratories	
CDC73	A300-170A	Bethyl Laboratories	

DNA damage			
Rad51	D4B10	Cell Signalling Technology	WB
Rad51	7946	Gift from F. Esashi	ChIP
phospho-T13/S14 Rad51	5544	Gift from F. Esashi	ChIP
Rad51C	2H11/6	Novus Biologicals	
Rad51C	ab55728	abcam	
CSB	ab96089	abcam	
CSB	NB100-61070	Novus Biologicals	
WSTF/BAZ1b	ab50850	abcam	ChIP
WSTF/BAZ1b	A300-446A	Bethyl Laboratories	WB
SNF2H	ab3749	abcam	
CUL4A	#2699	Cell Signalling Technology	
DDB1	A300-462A	Bethyl Laboratories	
PHIP	A302-054A	Bethyl Laboratories	
PALB2	A301-246A	Bethyl Laboratories	
R-loops			
S9.6	N/a	Produced in lab from mouse hybridoma cell line, gift from Richard Gibbons	DRIP
S9.6	ENH001	Kerafast	
Senataxin	A301-105A	Bethyl Laboratories	
Loading controls			
Tubulin	T5168	Sigma Aldrich	
Other			
KMT1C/G9a	ab40542	abcam	
ABL1	ab69810	abcam	
ABL1	ab15130	abcam	
ABL2	ab54696	abcam	
FLAG	M2 clone, A2066	Sigma Aldrich	
CRISPR-cas9	[7A9-3A3] ab191468	abcam	
Secondary			
goat anti-rat IgG (H+L) FITC	A110-105F	Bethyl Laboratories	IF
goat anti-rabbit IgG (H+L) Alexa Fluor 594	A11012	ThermoFisher Scientific	IF

2.13 Primers

Name	**Sequence (5' – 3')**
PKM	
Promoter F	CGGTCCTAACACAATGAATGG
Promoter R	CCTCGGAGACTGAGCTGAAC
Intron 1 F	TCATTGAAACTGCGTTCTGG
Intron 1 R	GCGACCAGAGCAGAGAGG
Intron 2 F	GGAGCAGATGCTTTTTGGTC
Intron 2 R	TACTGACCCTGGCTCCAATC
Intron 4 F	TCTGGCTGGGTAGAGAGCAT
Intron 4 R	GGCCTCAAACACAAACCCTA
Intron 9 F	ACCACACCCGGCTAATTTTT
Intron 9 R	GGGCAAGCCATTTAACCTCT
Intron 10 F	GGCCATGATGTTCTGTTCCT
Intron 10 R	GTGAAGACCAGGAGGCCATA
Exon 12 F	AGAAACAGCCAAAGGGGACT
Exon 12 R	TGCTGCTGTAACTTGGAGGA
Intron 12 F	TAAAGTGGCTGCCTTTCTGG
Intron 12 R	TTGCCCAAAGACACATTCAA
Exon 13 F	ATAACCAGCCTCTTGCTCCA
Exon 13 R	ACCAGTCCTTCACCAAGTGC
Exon 14 F	CACCTTGTGGTGTGAGGAAG
Exon 14 R	GGTCAGCACAATGACCACAT
Pause 1 F	CCACTTTGGTCTTCCTGCAT
Pause 1 R	CCTTTGGCACTACAGCATCA
Pause 2 F	CGCAAGACTAGGACGACAGA
Pause 2 R	CTTCCCCAGCTTCTCTCCTT
Pause 3 F	TGTCTTGGCTGAGTCTGCTG
Pause 3 R	GCAATTAGGACTGGGAGTCG
Pause 4 F	GATGTCTTCACCCCATGTCC
Pause 4 R	AGGTTGCAGTGAGCCAAGAT
Pause 5 F	TTACAGGTGTGAGCCACTGC
Pause 5 R	CATGGGAGAGCTTCCAAGAG
β-actin	
prom F	GAGGGGAGAGGGGGTAAA
prom R	AGCCATAAAAGGCAACTTTCG
in1 F	CGGGGTCTTTGTCTGAGC
in1 R	CAGTTAGCGCCCAAAGGAC
in3 F	TAACACTGGCTCGTGTGACAA
in3 R	AAGTGCAAAGAACACGGCTAA
5' pause F	TTACCCAGAGTGCAGGTGTG
5' pause R	CCCCAATAAGCAGGAACAGA
pause F	GGGACTATTTGGGGGTGTCT
pause R	TCCCATAGGTGAAGGCAAAG
UTR F	CCTCACTTCCAGACTGACAGC
UTR R	TGGGCCACTTAATCATTCAAC
UTR2 F	CAGTGGTGTGGTGTGATCTTG
UTR2 R	GGCAAAACCCTGTATCTGTGA

YWHAZ	
Promoter F	CAGCCTTGTGGTCCACTGTA
Promoter R	CTCCCCATTGATTCTTCCAA
5' F	CGCAAGTCCTGATCACAATG
5' R	CCAGTCACTCCCTCTGGAAG
Gene Body 1 F	ACAAGCGTAGAGGAGCTGGA
Gene Body 1 R	GATGACGTCAAACGCTGCTA
Gene Body 2 F	GGACAAGCGTAGAGGAGCTG
Gene Body 2 R	GATGACGTCAAACGCTGCTA
Body intron 1 F	AAGATGGCAGCCAAAAATTG
Body intron 1 R	CCTCAGGCCATGCATCTATT
Body intron 2 F	CCTTCCAGGTTCACACCATT
Body intron 2 R	CGGATCACGAGGTTAGGAGA
Last intron 1 F	TTGACAGGAGGAGCACACAG
Last intron 1 R	GAGGGGAAGGGGAAAATACA
Last intron 1 small R	TATCTCTGGCCTGGAAAGGA
Last intron 2 F	GGCCCTGATCTACTGTCTGC
Last intron 2 R	GCCAAGCTGTTCCTGAACTC
Last intron 2 small R	TTGCAATATGCGCTCTTCAT
Last exon 1 F	GATGATGGGAGGCAGTGAGT
Last exon 1 R	CCAAGGTCTCCATTGCACTT
Last exon 2 F	GAGACTGGGCACTTGGAAAA
Last exon 2 R	GCAAGGTTGGACATCCTGTT
3'1 F	GATCAGCGGCGACATTAGAT
3'1 R	CACTTAAATGGGAGCCCTGA
3'2 F	TGCAGCTGATGTCTGTAGGG
3'2 R	CCCAGCTACTTGGGAGATTG
3'3 F	CATCACCTCCAGATGGGACT
3'3 R	GAGGGTTGGTTTCAGGATGA
3'small F	TGCAGCTGATGTCTGTAGGG
3'small R	TTGCATTTTTCCCATCTGTG
CCND1	
Splice exon F	TCCATGCGGAAGATCGTCG
Splice intron R	CCCTAGAAACACCACGGCAA
Splice exon R	GTCGGTGTAGATGCACAGCT
polyA 1 F	GGTGCACGTGACTGGACTTA
polyA 1 R	CAGCTGACAAGTAGCACCCA
polyA 2 F	TGGGTGCTACTTGTCAGCTG
polyA 2 R	GAGAGGCTGGTGGAGACCTA
CCNB1	
First Exon F	TCTGCTGGGTGTAGGTCCTT
First Exon R	AGCAGGCAGCAGCTAAGAAG
Well back F	TATGGGCATGCAGGTTAGTG
Well back R	AGAAGGAGGAAAGTGCACCA
Intron 3 F	CTTGGAAGGCCCATACTGATT
Intron 3 R	GGAGTTGATGCTTGAGCTGTT
Penultimate intron F	GCTGAGGCAGGAGAATCACT
Penultimate intron R	TGACTTGCCCACCCTAAAAC
Last intron F	TGGCATTGTAAGATGCTGAAA
Last intron R	TCTGGAAGACAAAAAGATAGCAA

Last exon F	GGTTTACCTGGGGATCCAAT
Last exon R	TGCAGCAGATTCTCCATGAC
Last exon 2 F	AAGTATGCCACATCGAAGCA
Last exon 2 R	AAGGGGCCACAAGCTTTATT
1st gRNA F	TCATTTTGGTTTTGTGTTTTGG
1st gRNA R	TCTGTATTTATACAAGGCATGAAGGA
Between gRNAs F	TTTCAACATACATAAGCCTACCTG
Between gRNAs R	TGTGTATCTAGGCACAAGAAGAGG
3rd gRNA F	GCCTGCAGTATTCAATAATATGGT
3rd gRNA R	TGCTTAAGGACTGGGATATGG
CoTC 1 F	CCCTGTTCGTTCAGATTGCT
CoTC 1 R	AGGAGAATCGCTTGAACCTG
CoTC 2 F	CGCCCGGCTAATTTTGTATT
CoTC 2 R	AGTAAATTGGCCAGGTGTGG
CoTC 3 F	CCTCCCAGGTTCAAGTGATTC
CoTC 3 R	TGAGACCATCCTGGCTAACAC
Thr4/1 F	CGTGTTAGCCAGGATGGTCT
Thr4/1 R	AAGCAGGGTTTCAGTCATGG
Thr4/2 F	AGGGAGTCTCGTTCTGTTGC
Thr4/2 R	ACGTGGTGAAACCCTGTCTC
Thr4/3 F	CTGGCCTCAAGTGACTCTCC
Thr4/3 R	AGTTTGCAGTGAGCCGAGAT
T2b F	GCCCTTCAATAACGTTGGAA
T2b R	CTGGGCGACAGAGAGACATT
T2c	TGTTGCATGCCCTTCAATAA
T2d	CTGGGCGACAGAGAGACATT
Aft1 F	TTCGACCTTTGCAGTGAGTG
Aft1 R	GCTGCTCACTCTTTGGGTCT
Aft2 F	GCAGTGCAGACCCAAAGAGT
Aft2 R	ATGGACCAATCAGCAGGAAG
Aft3 F	AGACCAGGAGCTCAAGACCA
Aft3 R	ACAGGGTCTCACTCCCACAC
Aft4 F	CCTTTTCCTGTGTGGCAGAT
Aft4 R	GTCCCAAGAGCCAGGTTACA
Aft5 F	TTTGACTGAGGGGCATAAGG
Aft5 R	TGCCTTTTTAACATGCCACA
Melt Assay Primers	
HRM F	CCTTAGGCTCTTTCTAAATCCCAC
HRM R	TGGGGCACATGACTATTTATCTG
HRM2 F	GTTAGTTTACTGCTTCCTTAGGC
HRM2 R	TCTAGATGGCATACCCTCCTGC
HRM3 F	GTGGTCCCATAAGATTATGATGCC
HRM3 R	GTTCAAACATCAGTGTACTCACAC
Deletion Check Primers	
DelCheck F1	CCACACCTGGCCAATTTACT
DelCheck F2	ACCACACCTGGCCAATTTAC
DelCheck R1	CCTGGCTAACACGGTGAAAT
DelCheck R2	CCACAGTTTGAGCACAGCAT
DelCheck R3	GGGTCTGCACTGCCTTTATG

GAPDH	
Exon1 F	TGAGCAGTCCGGTGTCACTA
Exon1 R	GCTTGGATGAAACAGGAGGA
Exon2 F	GGGTCTTTGCAGTCGTATGG
Exon2 R	ACCTCCTGTTTCTGGGGACT
Intron3 F	CCCACACACATGCACTTACC
Intron3 R	CCCACCCCTTCTCTAAGTCC
Last Exon F	CAACGAATTTGGCTACAGCA
Last Exon R	AGGGGTCTACATGGCAACTG
3' F	AACAAGGCCTTTTCCTCTCC
3' R	AGGCTGAGCTCCACTAACCA
WT T4 F	AGCCCCTAAGGTCTTCAAGC
WT T4 R	TTCCCCTAAGGACGATGATG
Border F	TCCAGCCTAGGCAACAGAGT
Border R	TCTTGGAAAAGCCTGGAGAA
KD1 F	TCTCCAGGCTTTTCCAAGAA
KD1 R	GCTCTTGCTCTCCTGGTCAC
KD2 F	GCATGGAGCTTCACCTTCTC
KD2 R	CACCTTGTCCCTGCTATGGT
KD3 F	GACAACAGGGATGGAGGAAA
KD3 R	GTGTCTGGAGCTAGGCTTGG
KD4 F	CGCAGTGAAAGCAGGAGACT
KD4 R	AGGGTGGGTTTGGAGAGAGT
Aft1 F	GAGGCGGAGGTTGTAGTGAG
Aft1 R	TTGTGCATTTGACCCTGTGT
Aft2 F	TTAACAGGTGGCGTTCACAG
Aft2 R	TCCAGGAAAAGCAGAGGAAA
Aft3 F	CCAGACAGCAGAGATGGTCA
Aft3 R	GGCTGGTGGCACTGTATTTT
Aft4 F	TCCTGAGGTCAGCAAGTCCT
Aft4 R	AGGTGGAGCTCAGGCTGTAA
Aft5 F	CGGGTCAGTGGAAAAACTGT
Aft5 R	CCTGGGTGACAGAGCAAGAT
Aft6 F	CCATCCAAAAGAGGGAGACA
Aft6 R	AAACTGTCCTGCTCCTGCAT
Aft7 F	CTGTCCCAAGGATCCACACT
Aft7 R	TATCCAAATTGCCAGCATCA
Aft8 F	TCGTAGAGACGGGGTTTCAC
Aft8 R	TAGAGGCAAGTGGAGCTGGT
Aft9 F	GGGAGGGGAAGAGAGAGAGA
Aft9 R	CTGAAGACCTCAGGCTCTGG
Deletion check and sequencing primers	
GAPDH_1_HRM_F	AGGTCAGGAGTTTGAGACCAG
GAPDH_1_HRM_F2	GCCAACATGGTGAAACTCCGTC
GAPDH_1_HRM_R2	CTGTTGCCTAGGCTGGAGTG
GAPDH_1_HRM_R	GGAGTCTCATACTCTGTTGCC
GAPDH_2_HRM_F	ACACCGCAACAAGAAAGTG
GAPDH_2_HRM_F2	AATCACAAAAATACTGCAGATAGAC
GAPDH_2_HRM_R2	TGGTGAAACCCCATCTCTAC
GAPDH_2_HRM_R	TTCAAGGCAAGCCTGGCCAAC
hGAPDHnew HRM_F	TCCCGGGCCTCACTTTGCACG

hGAPDHnew_HRM_R	AGTTTCACCATGTTGGCCAG
hGAPDHnew_HRM_F2	AGAATCCATTTGAATGTCAGCTC
hGAPDHnew_HRM_R2	GATGTGCTTCCTCCGTGTCTG
hGAPDHnew_HRM_F3	GGCTCAGTTGCTTTGGGCAG
hGAPDHnew_HRM_R3	AGGCTGGTCTCAAACTCCTG
hGAPDHnew_HRM_F4	TCAGCTCAACACAGCCTCC
hGAPDHnew_HRM_R4	AGGGTCTCTGTTTAAGTATCAGC
5S rRNA	
5S F	GGCCATACCACCCTGAACGC
5S R	CAGCACCCGGTATTCCCAGG
5Sv2 F	CTAGAGCACTTGGGGAGCAC
5Sv2 R	CAGTCTCCCATCCAAGCACT
5Sv2 R2	TATTCCCAGTCTCCCATCCA
7SL rRNA	
7SL F	GTGTCCGCACTAAGTTCGGCATCAATATGG
7SL R	TATTCACAGGCGCGATCCCACTACTGATC
7SLv2 F	GGAGTTCTGGGCTGTAGTGC
7SLv2 R	ATCAGCACGGGAGTTTTGAC
7SLv2 R2	CTGATCAGCACGGGAGTTTT
U6	
U6 F	CTCGCTTCGGCAGCACA
U6 R	AACGCTTCACGAATTTGCGT
cen9	
Cen9 F	CCTATGATGAAAAAGGTAATATCTTC
Cen9 R	CTGAAAGCGCTTAAAACGTCC

5S and 7SL Up, Promoter and Down primers are from (Sadeghifar et al. 2015).

2.14 siRNA duplex sequences

Rad51:

Sense: 5' G.C.C.A.G.A.U.C.G.U.A.U.A.C.A.A.U.U.A.U.U 3'

Antisense: 5' U.A.A.U.U.G.U.A.U.A.C.G.A.U.C.U.G.G.C.U.U 3'

SMARTpool siRNA against human CPSF73, BAZ1b, CUL4A and SNF2h were purchased from ThermoFisher Scientific.

CPSF3: SMARTpool L-006365-01-0020

BAZ1b: SMARTpool L-006901-00-0005

CUL4A: SMARTpool L-012610-00-0005

SNF2h: SMARTpool L-011478-00-0005

siLuc: Control siRNA HA08751859

2.15 Plasmids and gRNAs

The GFP-RNase H1 plasmid (Cerritelli et al. 2003) was a gift from R J Crouch.

gRNAs were designed with the help of Andrew Bassett (Genome Engineering Oxford), using the Zhang laboratory online generator (http://crispr.mit.edu). Dr Bassett then cloned and tested constructs as follows. gRNAs were cloned into the sgRNA-HA-Cas9-2A-Puro vector (pX-Cas9-Puro) cassette. gRNA cutting efficiency was tested in HeLa, using a High Resolution Melt assay (HRMA).

Sequences:

CCNB1:

Upstream:

1: 5’-ACTTGCAACCTCCGCCTCCCAGG-3’

Downstream

1: 5’-CGAAATTTGTTCCTTCCGGTGGG-3’

2: 5’-TCCATGCTGTGCTCAAACTGTGG-3’

GAPDH:

Upstream:

1: 5’-CAGGCGTGCGCGCTCTACCGAGG-3’

2: 5’-GGTTCACAATGCCTCGGTATAGG-3’

Downstream

1: 5’-AAAGCCAAGGCCCTAGAGTCAGG-3’

Chapter 3

R-loops and Rad51

3. R-loops and Rad51

3.1 Introduction

Pause-type termination has only been characterised in detail for a small group of genes (*β-actin,* and to a lesser extent *ENSA* and *GEMIN7)*, which were selected for the same criteria (coincidence of RNAPII CTD Serine 2 phosphorylation, HP1γ ChIP, and DRIP-seq signals) (Skourti-Stathaki et al. 2014). Repeating the experiments carried out in the study of *β-actin* termination on another gene, chosen for different reasons, would give a better idea of the prevalence of this subset of termination. If it generated clear evidence of pause-type termination, this would provide a potential model gene to study more novel aspects of the system.

The previously discovered connection of Rad51 with R-loops in yeast (Wahba et al. 2013) suggests a similar role may be played by the mammalian homologue. This could be determined by extending the analysis of the termination regions of pause-type genes to look for the presence of Rad51, and whether its removal has an effect either on R-loop prevalence, or level of downstream DNA damage due to loss of repair factors. Comparison of R-loops at termination regions and those elsewhere in the genome, e.g. near promoters, could indicate whether Rad51 is necessary for formation of R-loops at different locations or with different biological roles (e.g. association with CpG islands or transcription termination).

3.2 PKM as a model gene for pause-type termination

Pyruvate kinase (*PKM*) fulfils a number of criteria suggesting it may terminate using a pause-type mechanism: the 3' flanking region is significantly G rich, and contains numerous GC clusters. Recent genome-wide low resolution DRIP-seq (Ginno et al. 2012; Ginno et al. 2013) suggests R-loops may be present at both 5' and 3' ends of *PKM* (figure 3.1). Bioinformatic analysis also suggested *PKM* as a negative control for CoTC-type termination. It differs from *β-actin* in being less highly expressed, much larger in size, and in being alternatively spliced, making it an interesting test of the relevance of the *β-actin* model.

Firstly, Chromatin Immunoprecipitation (ChIP) was used to establish the RNAPII transcriptional profile (RNAPII occupancy) over *PKM*, as this would give a good indication as to whether pausing occurs at the 3' end (figure 3.2). Higher RNAPII occupancy at the promoter indicates promoter proximal pausing, and the significant accumulation of RNAPII downstream of the pA site suggests 3' pausing. RNAPII terminates approximately 1.5 kb from the pA signal.

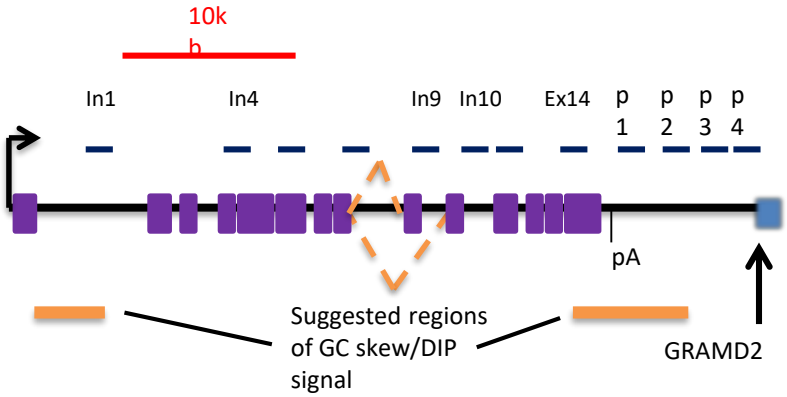


FIGURE 3.1 Schematic of pyruvate kinase (*PKM*). Intron 1, 4, 9, 10, Exon 14 and Pause 1, 2, 3, and 4 primer sets are marked as blue bars above the gene. Orange dotted line shows the alternatively spliced exons and orange bars the suggested R-loop regions according to (Ginno et al., 2013; Ginno et al., 2012). Red bar denotes scale in kilobase pairs (the 3' flanking region has been expanded 5x). *GRAMD2* is the next downstream gene.

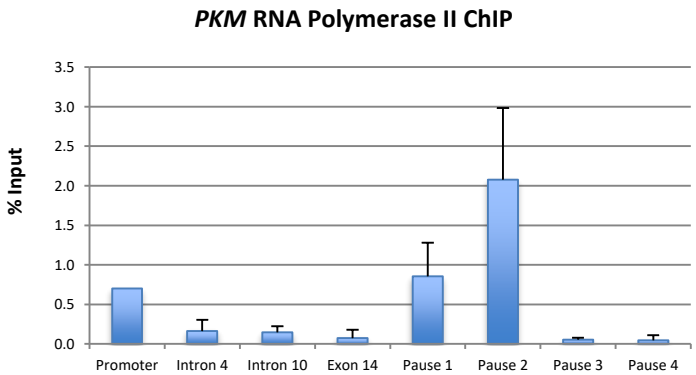


Figure 3.2 ChIP analysis of RNA Polymerase II (H-224) profile across *PKM*. Error bars for all graphs represent standard deviation over four repeats unless otherwise stated.

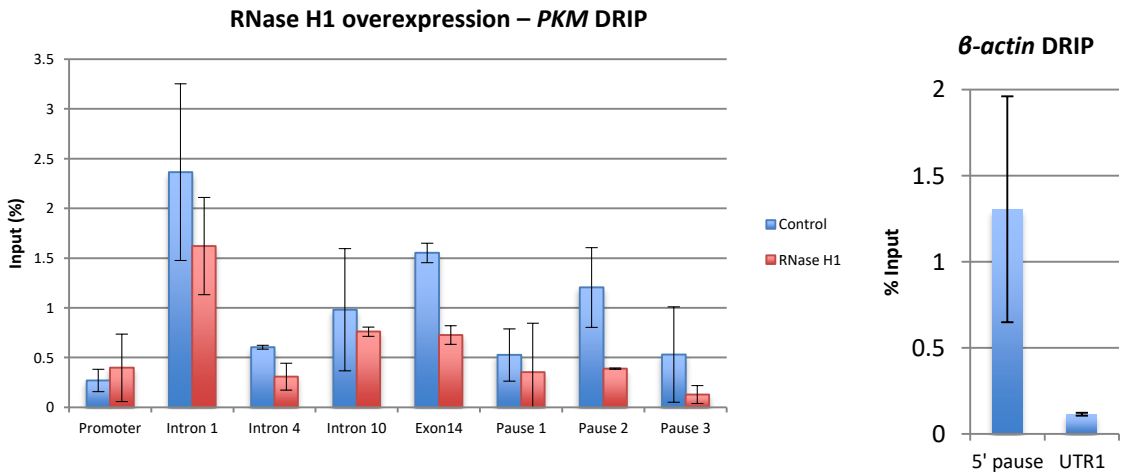


Figure 3.3 DRIP analysis over *PKM* with and without RNase H1 over-expression. Error bars represent standard deviation over 6 repeats.
Positive and negative controls for DRIP analysis, using *β-actin* primers.

To detect R-loops over the 3' region, DNA-RNA immunoprecipitation (DRIP) analysis was performed, as described previously (Skourti-Stathaki et al. 2011; Skourti-Stathaki et al. 2014). This technique detects R-loops using the S9.6 antibody that specifically recognises RNA/DNA hybrids. This was then coupled with RNase H1 overexpression, to test whether the observed peaks of signal across the gene were specifically RNA-DNA hybrids. This analysis suggested R-loops were present either side of the pA site, in the regions of the exon 14 and pause 2 primer sets, supporting the hypothesis that *PKM* transcription terminated via a pause-type mechanism (figure 3.3). Potential R-loop signal was also observed at the 5' end of the gene, although error bars remained too large to confirm, despite multiple repeats (n=6).

To investigate whether the later stages of the proposed termination model (Skourti-Stathaki et al. 2011; Skourti-Stathaki et al. 2014) were also indicated in *PKM*, RT-qPCR was performed on total RNA to look for the presence of antisense RNA at the termination region. Higher than background levels were found downstream of the pA site, as well as over intron 1 (figure 3.4), perhaps associated with promoter-upstream transcripts (PROMPTS) and/or the potential 5' R-loop. This was also seen for *β-actin.* The presence of 3' antisense RNA correlates with the before mentioned pause-type termination model, where ssDNA is used as template to generate antisense RNA.

Chromatin environment was then analysed by H3K9 di- and tri-methylation ChIP (figures 3.5 & 3.6). These modifications are considered repressive chromatin marks and are enriched in heterochromatic regions. H3K9 di-methylation was

observed in the R-loop region in *β-actin* (Skourti-Stathaki et al. 2014). High di-methylation signal was detected over the alternatively spliced exons (9 & 10) in *PKM*, as previously suggested to lead to RNAPII pausing and alternative exon inclusion in the human fibronectin 1 gene (Allo et al. 2009). However, no significant accumulation of either mark was observed in the vicinity of the pA or pause/R-loop regions. H3K9 di-methylation accumulation was observed further downstream (primer sets pause 4/5), however this area is over 2kb downstream of the PAS, and well beyond any observed R-loop signal. Enriched H3K9me3 signal was observed over the intron 1 region but is low over the 3' region. H3K9 tri-methylation over intron 1 may perhaps function in repressing transcription initiation from the transcription bubble of paused RNAPII or from a 5'-associated R-loop.

HP1γ, a marker for relatively transient heterochromatin, was shown to be present in higher levels over termination regions compared to the gene body (figure 3.7), suggesting heterochromatin is formed over the pause region. How HP1γ recruitment occurs without the presence of the di-/tri-methyl marks requires further investigation, and this finding differs from that observed in the pause region of *β-actin* (and more recently in the *ENSA* and *GEMIN7* genes) (Skourti-Stathaki et al. 2014).

These results suggest that *PKM* transcription terminates via a pause-type mechanism, involving R-loop formation in the RNAPII pause region. However, results also lead to the tentative conclusion that the R-loop/pause mechanism

involved in *PKM* termination may not feature all stages of heterochromatin formation as discovered in *β-actin* termination.

While further experiments could be performed to fully characterise *PKM* in the same way as *β-actin* it was felt that little novel information would be generated, and time would better be spent on using *PKM* and *β-actin* to investigate as yet untested hypotheses.

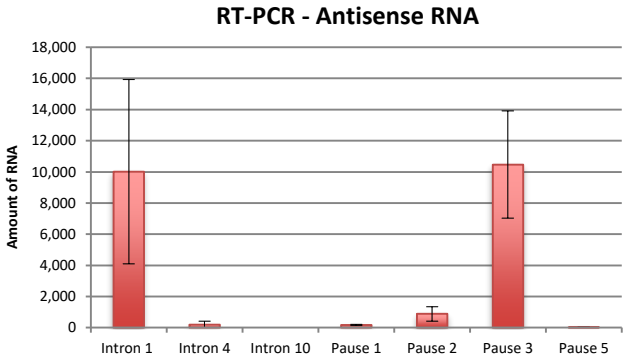


Figure 3.4 RT-qPCR analysis using forward primers specific for each gene region shows significant levels of antisense RNA in pause regions of *PKM*.

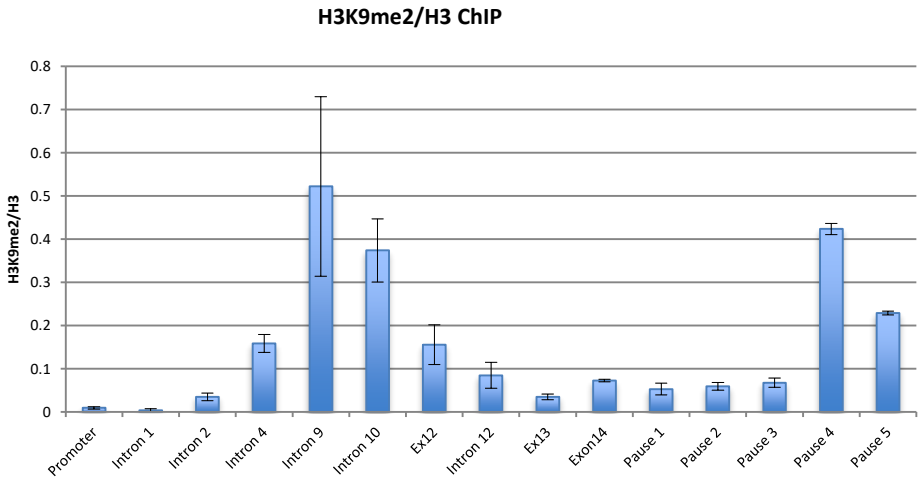


Figure 3.5 H3K9 di-methylation ChIP analysis across the *PKM* gene.
H3K9me2 levels normalised to H3. See Materials and Methods for antibodies used.

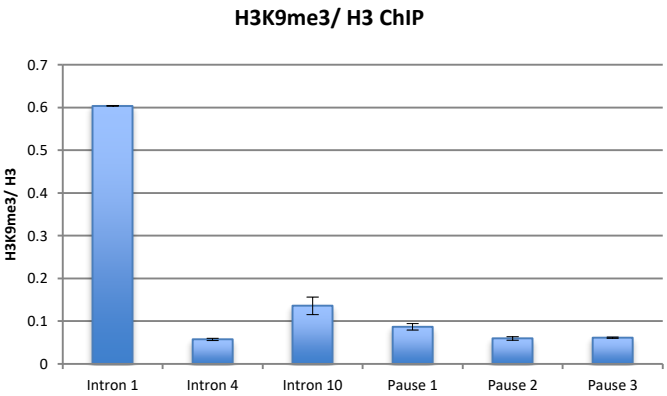


Figure 3.6 H3K9 tri-methylation ChIP analysis show no significant enrichment over pause regions. H3K9me3 levels normalised to H3.

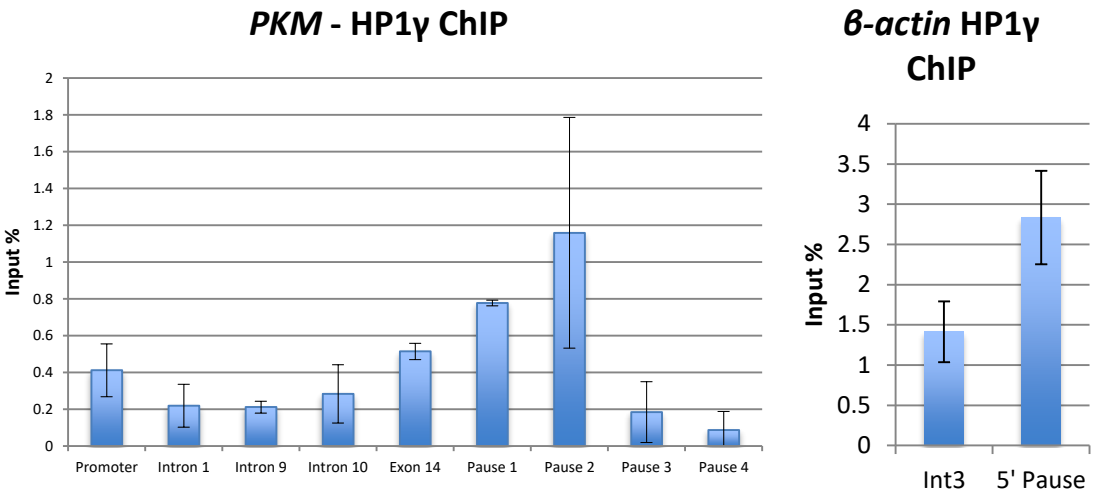


Figure 3.7 HP1γ ChIP across *PKM.* Signal for HP1γ increases over the pause region. Positive and negative controls for HP1γ, using *β-actin* primers.

3.3 Rad51 is present at pause regions

To establish whether Rad51 was present at R-loop regions such as those found at pause-type terminators, ChIP analysis was performed using a number of different Rad51 antibodies to establish which was most suitable for ChIP. While Rad51 ChIP and ChIP-seq have been performed before (Aymard et al. 2014; Yata et al. 2014), these experiments are usually performed on DNA treated with damaging agents, and this meant that defining a positive control region to test the antibodies in undamaged cells was difficult. However, after unsuccessful attempts with commercially available antibodies (data not shown), a potential signal was detected over the *PKM* R-loop region (pause 2) using a polyclonal Rad51 antibody (7946) (figure 3.8). This signal remained unreproducible despite multiple repeats, therefore a new ChIP was attempted using an antibody raised against a diphosphorylated form of Rad51 (5544) (Yata et al. 2012). This antibody is believed to have higher affinity for Rad51 (Fumiko Esashi, personal communication), and may therefore produce higher IP to background ratios. This produced a more significant peak at the expected R-loop region (pause 1 & 2) of *PKM* as compared to the gene body (figure 3.9, left panel). Rad51 recruitment over *β-actin* was also analysed (figure 3.9, right panel) and was enriched over the 5' pause region, previously shown to be R-loop associated (Skourti-Stathaki et al. 2011; Skourti-Stathaki et al. 2014).

ChIP analysis with a second batch of the antibody again produced a significant peak over the termination element of *PKM* (figure 3.10). However this experiment also revealed a significant peak over intron 1, also a potential R-loop

forming region. This may suggest that Rad51 is involved in R-loop formation at 5' ends as well as at termination events. ChIP profiles across Cyclin B1 (*CCNB1*), which features a CoTC-type terminator, showed no increase in signal over the body (1) or termination region (3'2, 3'3), indicating the Rad51 signal is specific to pause-type genes.

Rad51 ChIP and R-loop DRIP profiles were next produced using U2OS cells instead of HeLa, in order to more closely corroborate existing ChIP-seq data (Yata et al. 2014) and so that existing protocols for the Rad51 inhibitor (BO2) could be used to determine the effects of Rad51 inhibition on R-loop profiles. U2OS Rad51 ChIP and DRIP profiles over *PKM* and *β-actin* were similar to HeLa (figures 3.10 (right panel) & 3.11).

3.4 Rad51 does not effect R-loop formation

Both siRNA knockdown of Rad51 and the BO2 inhibitor were used to study the effects of protein depletion/inhibition on R-loop dynamics (figures 3.11-13). The BO2 inhibitor targets the DNA binding site of Rad51, preventing filament formation (Lodhia et al. 2015), and could provide cleaner results than protein depletion, and the drug treatment would also require a much shorter time period for application.

However, neither the BO2 inhibitor nor Rad51 siRNA knockdown had a significant effect on R-loop signals. This indicates that, at least in the case of hybrids forming in pause-type termination, Rad51 does not promote R-loop

formation, but may instead simply function downstream of this potentially hazardous species formation as a DNA repair factor. Since this function of Rad51 is well characterised, it was decided that further investigations into its role here were unwarranted.

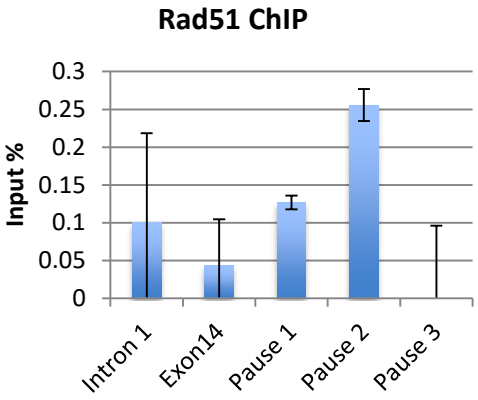
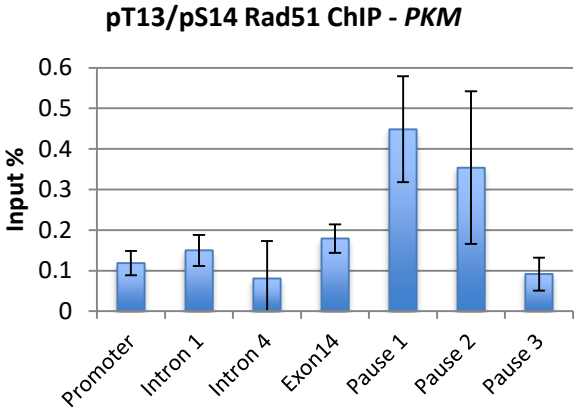


Figure 3.8 Rad51 ChIP (7946) over *PKM*. A potential enrichment is seen over pause 2.



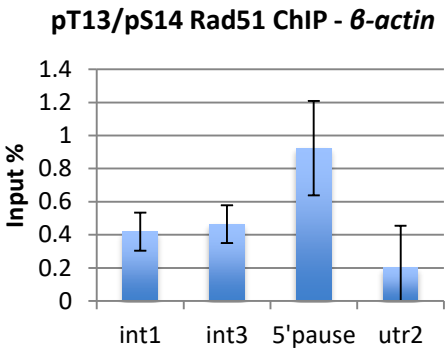
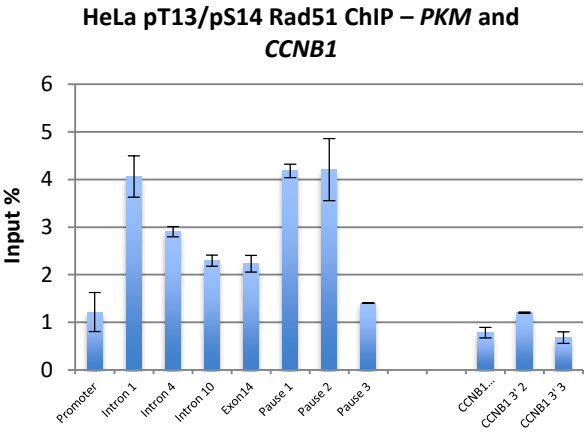


Figure 3.9 pT13/pS14 Rad51 ChIP (Ab 5544) over *PKM* and *β-actin*. Clearer signal indicates a peak of Rad51 over the pause region of both genes.



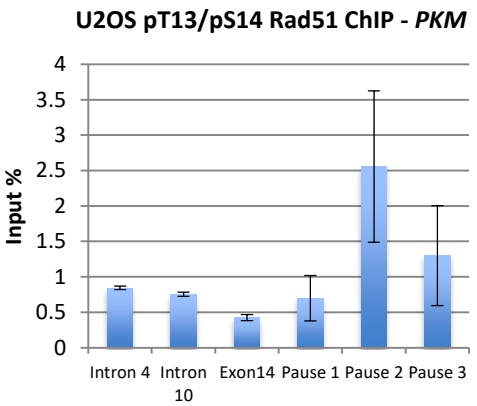


Figure 3.10 Repeated pT13/pS14 Rad51 ChIP using new antibody batch, and both HeLa and U2OS cells.

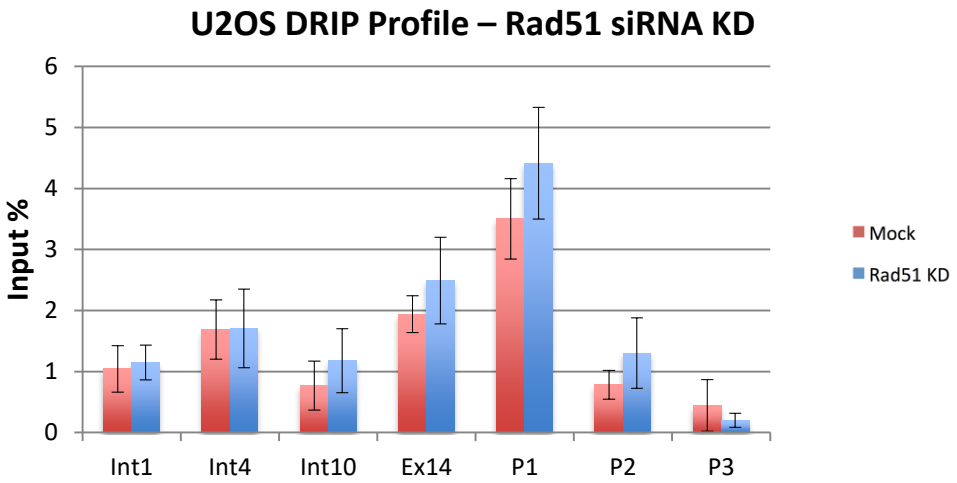


Figure 3.11 U2OS DRIP over *PKM* with and without Rad51 siRNA knockdown indicates no significant difference in R-loop signal was observed on Rad51 knockdown.

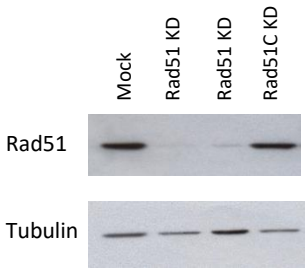


Figure 3.12 siRNA knockdowns. Whole HeLa cell extracts levels analysed by Western Blot after treatment with Rad51 and Rad51C specific siRNAs showing efficient knockdown of Rad51. See methods for antibodies used. Molecular weights: Rad51 – 37 kDa, Tubulin – 49.9 kDa.

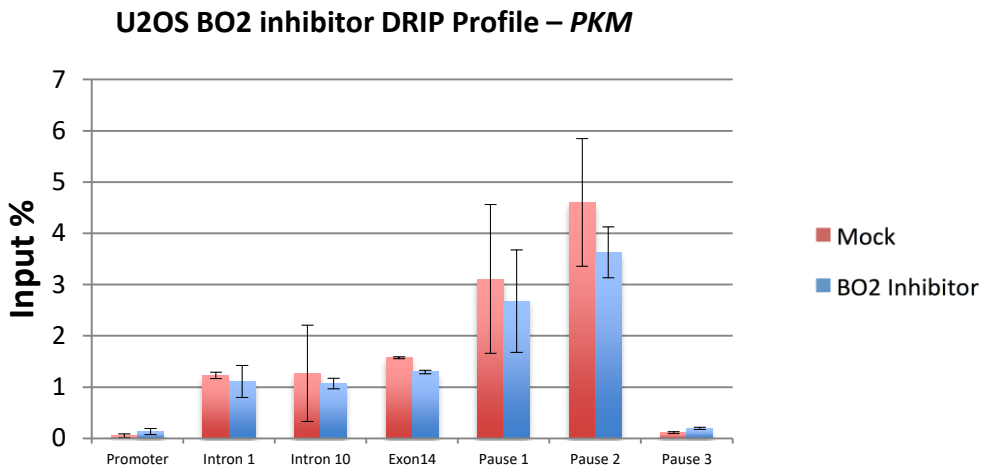


Figure 3.13 Rad51 DRIP profile over *PKM* with and without a 2-hour treatment with the Rad51 BO2 inhibitor.

3.5 Discussion

RNAPII, HP1γ ChIP, DRIP analyses and antisense RNA detection results (figures 3.2-4 & 7) obtained on *PKM* indicate that this gene could be utilised to further validate the mechanism of pause-type termination, but also study different aspects of R-loop biology, such as Rad51 involvement. These results also appear to show a good example of H3K9 di-methylation over alternative exons. Although most results appear to corroborate those seen for *β-actin*, the lack of H3K9 di-methylation in pause regions suggests some aspects of the *β-actin* model may not be universal. Perhaps in genes with lower expression (e.g. *PKM*), R-loop formation is enough to trigger pausing and termination, whereas in highly expressed genes the formation of heterochromatin may be needed to work in synergy with R-loops to ensure tight transcriptional regulation. To test this hypothesis, RNAPII ChIP (to measure pausing) and Br-UTP Nuclear Run-On or mNET-seq approaches (measuring nascent transcription levels) could be performed in RNase H1 over-expressing conditions to study whether R-loops are sufficient for pausing and termination. The dynamics of the HP1γ signal on RNase H1 and BIX (H3K9 methylase inhibitor) treatment might clarify the mechanism by which HP1γ is recruited despite the lack of H3K9 di-methylation.

The results obtained on Rad51 recruitment and its potential role on R-loop formation (figures 3.8-13) indicate that whilst Rad51 does appear to be present at pause-type termination regions, and not at CoTC-type, its presence does not affect the level of R-loop formation, and Rad51 may be present as a consequence

of R-loop induced DNA damage leading to repair factor recruitment, as seen with BRCA1 (Hatchi et al. 2015).

Although significant reduction in Rad51 foci formation was seen on treatment with this BO2 regimen in a number of cell lines including U2OS (Lodhia et al. 2015) (F. Esashi, personal communication), there remains the possibility that some active Rad51 was present after both BO2 and siRNA treatments. Higher inhibitor doses should therefore be trialled to fully rule out the possibility of Rad51 involvement in R-loop production.

A number of further experiments can be performed: a) study whether Rad51 recruitment is triggered by R-loop formation (perhaps due to DNA damage) - this could be tested by RNase H1 over-expression to remove R-loops; b) investigate whether γH2A.X signal increases in these R-loop regions. A further avenue is to investigate whether the Rad51 paralogues in mammals function in R-loop formation, since their role is ill defined. ChIP analysis with the two batches of pT13/pS14 Rad51 antibody differed in indicating a significant peak of Rad51 over intron 1 of *PKM*. This should be further investigated to test whether Rad51 is necessary for the formation of different types of R-loops. Intron 1 of *β-actin* should also be analysed using the new antibody batch. The clearer signal achieved using the pT13/pS14 antibody is interesting. It may be that a higher affinity antibody is needed to generate clear ChIP signals for an enzyme such as Rad51. Alternatively, only the diphosphorylated form of Rad51 is specifically enriched in R-loop regions.

Chapter 4

Development of NetMNase-seq

4. Development of NetMNase-seq

4.1 Introduction

A modification of mNET-seq protocol has already been mentioned for mass spectrometry. By gently isolating chromatin and bound polymerase without crosslinking, one can not only analyse the nascent RNA as done in mNET-seq, but also by pulling down transcribing RNAPII one can use mass spectrometry to detect protein complexes bound to different isoforms of the CTD. However this is not the only potential modification that can be made to the early stages of mNET-seq.

I aimed to create an MNase-seq like assay (NetMNase-seq) to analyse DNA protected from MNase digestion by histones, polymerase and/or other bound factors but performed under the same chromatin isolation conditions as mNET-seq (figures 4.1 & 4.2). This would allow results from the two assays to be directly compared and DNA and RNA even isolated from the same cells. Addition of an IP stage as in mNET-seq could also produce a native ChIP-seq like technique for histones and other bound factors. Below is a description of the development of this NetMNase-seq technique. For more information on buffers and conditions used in the final developed protocol, see methods.

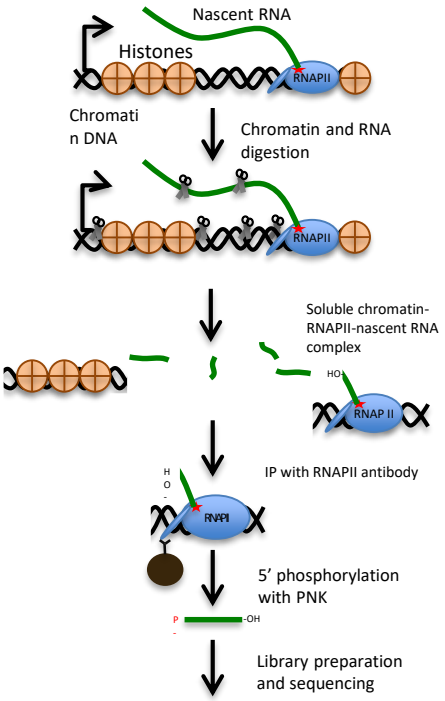


Figure 4.1 Diagram showing major steps in mNET-seq protocol. Cells are gently lysed to collect chromatin with elongating polymerase still bound. This chromatin fraction is then solubilized by an MNase digestion (scissors). RNA within the active site of the polymerase is protected from digestion. Polymerase is then pulled down with an IP step (antibody conjugated beads) and the RNA is then extracted and processed for sequencing.

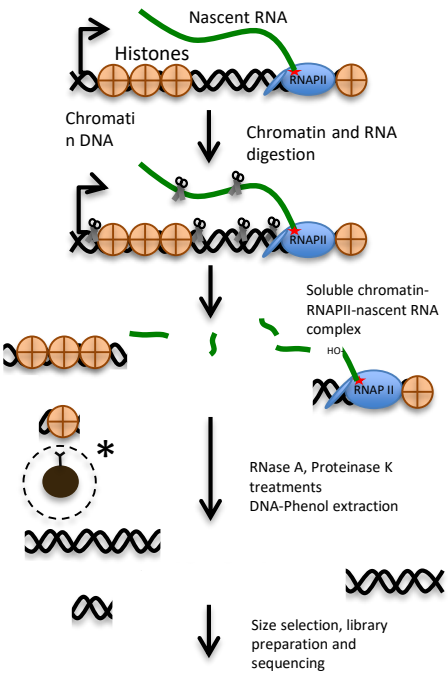


Figure 4.2 Similar schematic for aims in developing NetMNase-seq. During similar chromatin isolation and digestion to mNET-seq, histones protect DNA fragments which can be isolated following RNase and Proteinase K treatment. An IP stage could be added at the same point in the protocol to that in mNET-seq (*).

4.2 Establishment of chromatin isolation and digestion

In setting up this protocol, the early chromatin isolation conditions of mNET-seq were followed. Following MNase treatment, samples were treated with Proteinase K and then DNA-phenol extraction. Firstly it was necessary to establish the number of cells needed to collect a DNA library of sufficient size and coverage, and to titrate the MNase digestion in order to produce fragments of the expected mononucleosome size. Published values necessary for mNET-seq were taken as a starting point, and $4x10^7$ cells was found to be more than sufficient per condition ($5x10^6$ and lower have since been found to be adequate). The digestion produces a pattern familiar from MNase protection assays – a ladder corresponding to mono, di, tri and higher nucleosome configurations. An MNase titration (figure 4.3(i)) determined that a 40 U/µl MNase solution incubated for 2 minutes, was appropriate in order to produce a clear mononucleosome band without potentially over-digesting and removing/breaking down all higher fractions into mononucleosomes. Over digestion might mask regions of higher structure, or completely digest partially susceptible fragments.

Following nucleic acid isolation, sample aliquots were treated with RNase or DNase to confirm correct identification of the nucleic acid fragments as DNA, not RNA (figure 4.3(ii)). To make certain of this, an RNase treatment stage before Proteinase K digestion was added to all future method repeats, as shown in figure 4.2.

4.3 Size selection and library preparation

In order to isolate fragments of interest, gel purification of mononucleosome fragments was employed (figure 4.3(iii)). Excising all fragments between 75 and 200 bp was decided upon in order to take in the entire mononucleosome band and so the method can be consistently repeated, cutting at the visible marker points. Cutting at 75 bp avoids smaller fragments, for example corresponding to DNA protected in active site of polymerases (20-35 bp (Tornaletti et al. 1999)).

Mononucleosome fragments were then processed for paired end next generation sequencing. This allowed both ends of fragments to be read to identify the whole size and position of the mononucleosomes.

4.4 First NetMNase-seq analysis

Data processing from these first libraries produced profiles as seen in figure 4.4. Peaks corresponding to the approximate size protected by nucleosomes (140-200 bp) were seen across the genome, with particular sites showing low signal. One class of these was found in the proximity of gene transcription start sites (TSS). This can be seen in the metagene profile for mRNA genes (figure 4.5), in comparison to the mNET-seq profile. The peaks here correspond to nucleosome positions, with the peaks just after the TSS being most well defined due to regular nucleosome spacing around the TSS, and further into the gene nucleosome spacing becomes noisier as it varies from gene to gene. This repeat of a result found via a number of methods gives a good indication that the protocol is functioning correctly. The low signal corresponding to nucleosome

"free" or unprotected regions around the TSS is also clearly visible. The metagene profile for mNET-seq of Total RNAPII shows a peak at the TSS indicating paused polymerase before it enters the elongation stage of transcription. When this profile is overlaid on the NetMNase-seq profile, this peak appears to be situated just before the first regular nucleosome. This might suggest that polymerase is loaded at the TSS in an open chromatin region (susceptible to MNase digestion) and that it pauses before licensing for elongation just before the first regular nucleosome in the gene body. This agrees with a number of studies illustrating the importance of nucleosome spacing for transcription initiation (Dreos et al. 2016) and as barriers to transcription (Bondarenko et al. 2006).

These results indicate that the lysis and wash conditions used for mNET-seq (including the 1 M urea chromatin isolating NUN2 buffer) are appropriate for isolating chromatin bound histones as well as polymerase.

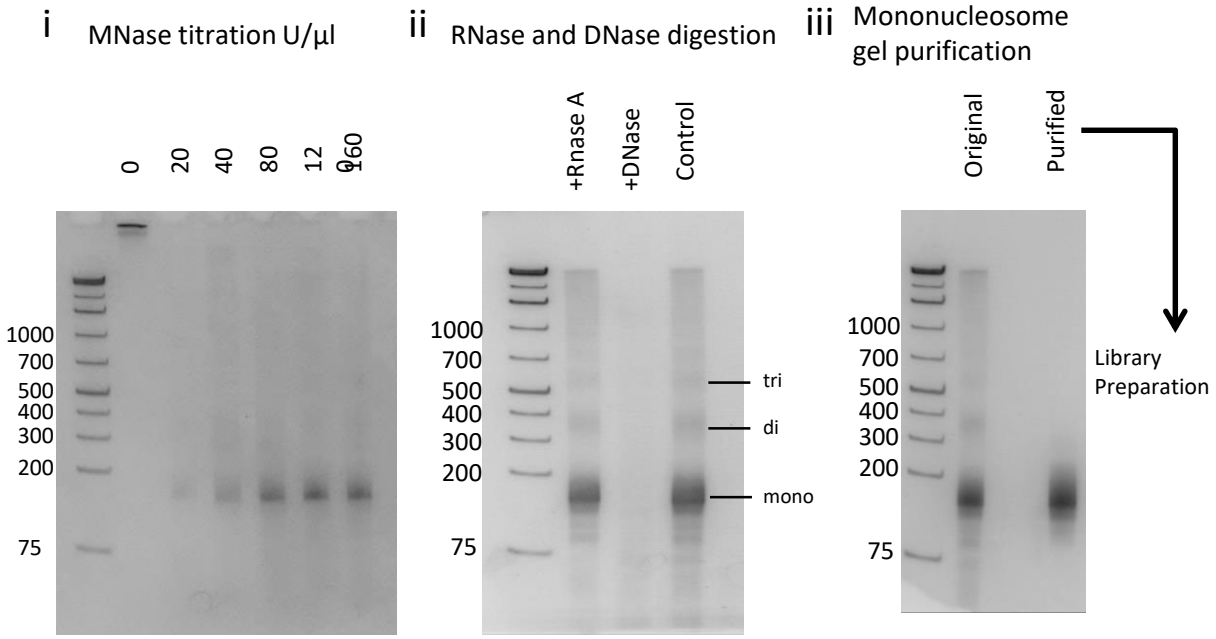


Figure 4.3 DNA digestion pattern confirmation. i) Gel visualisation of MNase digestion titration, showing predominantly mononucleosome (~140-200 bp) fraction as well as larger fragments. ii) Confirmation of DNA nature of fragments by RNase A and DNase (TURBO DNase) treatments, showing loss in signal on digestion with the latter only. Bands corresponding to di-, tri-, and higher nucleosome configurations are marked. iii) Mononucleosome fragment purification. DNA fragments between 75 and 200 bp were cut from the gel and purified (see methods). Purified fragment size was then confirmed by gel visualisation.

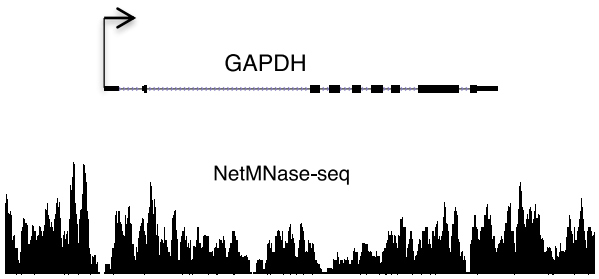


Figure 4.4 Visualisation of library generated from mononucleosome fragments. Genome browser visualisation of the nucleosome pattern observed over example (*GAPDH*) gene.

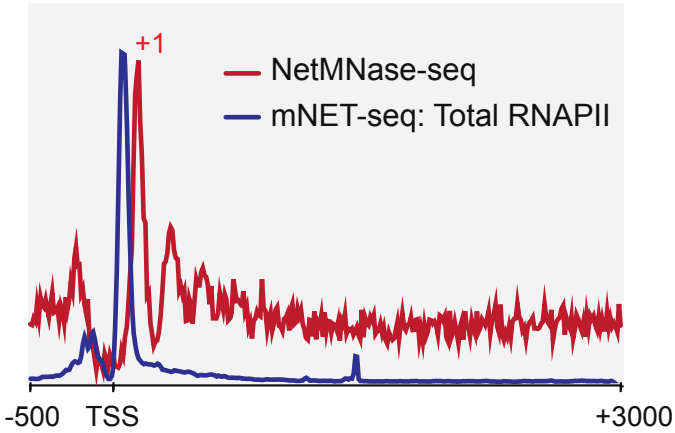


Figure 4.5 Metagene comparison of mNET-seq to NetMNase-seq. The metagene profile for NetMNase-seq across mRNA genes is in red. The metagene profile for mNET-seq of total RNAPII (data generated by Takayuki Nojima) is in blue. All metagene profiles were produced by Michael Tellier.

4.5 Addition of immunoprecipitation stage

While the above protocol gives a certain amount of information on nucleosome configuration, it would also be useful to be able to analyse only the DNA protected by specific proteins, such as histone variants' phosphorylation states. Additionally it might be possible to isolate mononucleosomes in close proximity to RNAPII isoforms by immunoprecipitation of RNAPII using antibodies specific to different post-translational modifications of the CTD. To test this, solubilised chromatin was incubated and washed in the same IP buffers and conditions as mNET-seq with antibody conjugated beads. Results for RNAPII IP are shown in figures 4.6-11. Histone modifications are discussed later (Chapter 6, figures 6.13 - 6.18).

DNA from input solubilised chromatin and Total RNAPII IP are shown in figure 4.6. The DNA co-precipitated with RNAPII was predominantly of a high molecular weight. One potential explanation for this is that predominantly undigested chromatin bound to many RNAPII molecules per DNA moiety will be more efficiently precipitated. Heterochromatic regions protected from digestion may also prevent any DNA bound to RNAPII from being digested to only ~30 nucleotides and so being lost from the gel. However, a certain proportion of DNA corresponding to mononucleosome and dinucleosome sizes could also be seen.

To test whether different fragment sizes might give different profiles, and to see which was most appropriate for further analysis, gel fragments corresponding to mono and di nucleosomes and larger fragments were excised for both input and RNAPII IP. Larger fragments did not elute successfully, so as a substitute whole

Input and IP samples were treated with a Fragmentase enzyme to produce DNA of a length suitable for sequencing.

4.6 Sequencing results for RNAPII immunoprecipitation

Sequencing results showed RNAPII precipitated mononucleosome fractions produced profiles over gene bodies (figures 4.7-9) that bear similarities to RNAPII ChIP-seq. This is encouraging since this protocol bears similarities to a native ChIP. Total RNAPII profiles show a prominent 5' peak as seen through other sequencing methods. However, unlike a ChIP-seq protocol, peaks here correspond to nucleosomes. This suggests that at least some proportion of transcribing RNAPII is closely associated with nucleosomes – close enough to prevent digestion of the DNA linking the polymerase and nucleosome. Two prevalent patterns were observed across genes. As examples, *CD9* shows high signal at the TSS and lower signal across the rest of the gene body. *MYC* shows high signal throughout the gene, perhaps corresponding to level of expression.

Metagene profiles for Input DNA (figure 4.8, left panel) show mononucleosome fractions produce the most highly defined, regularly spaced first nucleosomes. Fragmented total DNA lacks the same resolution. This may indicate that the first regularly spaced nucleosomes tend to be mononucleosomes, rather than being packed into higher configurations and therefore the linker DNA between is easily digested. Dinucleosome DNA shows the second nucleosome peak to be higher in read density than the first. This could be due to more reads in the di- fraction containing the second nucleosomal sequence, combined with either nucleosome

one or three. Dinucleosome read density is higher across the genome due to the increased read size, as read density is calculated from fragment length.

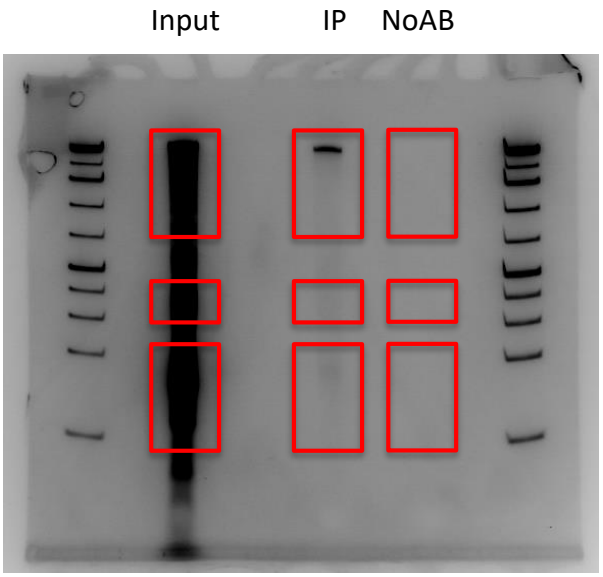


Figure 4.6 Test of immunoprecipitation addition to NetMNase-seq and library preparation from different fragment sizes. Gel visualisation of Input digestion (10%) and fragments precipitated with Total RNAPII antibody (MABI0601) and no antibody control (NoAB). The red highlighted regions, corresponding to mono, dinucleosome, and larger fragments were cut from the gel.

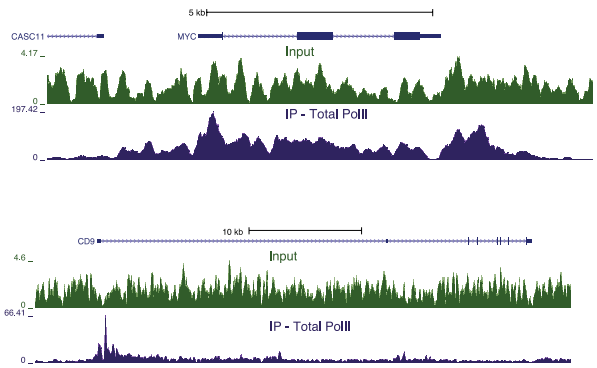


Figure 4.7 Visualisation of RNAPII IP results. Two examples genes (*MYC* and *CD9*) as representatives of two prevalent patterns observed on mononucleosome sized fragments extracted following RNAPII immunoprecipitation.

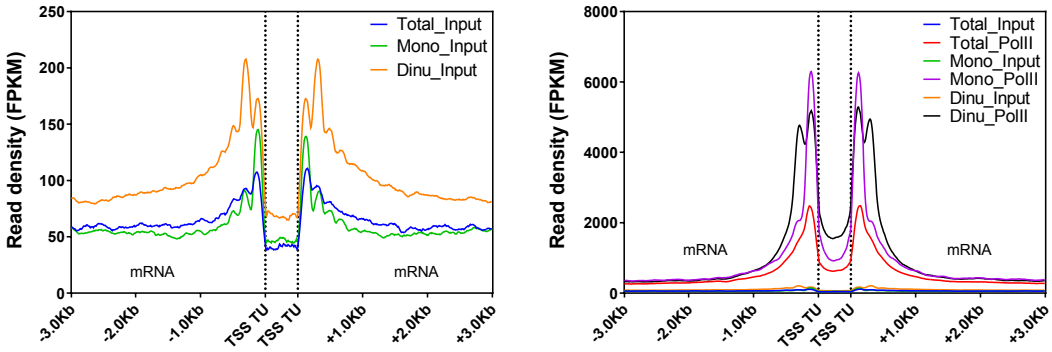


Figure 4.8 Metagene analysis of divergent mRNA genes. i) Comparison of libraries produced from mono, di and total Input fractions. ii) Comparison of RNAPII co-precipitated libraries (mono, di and total).

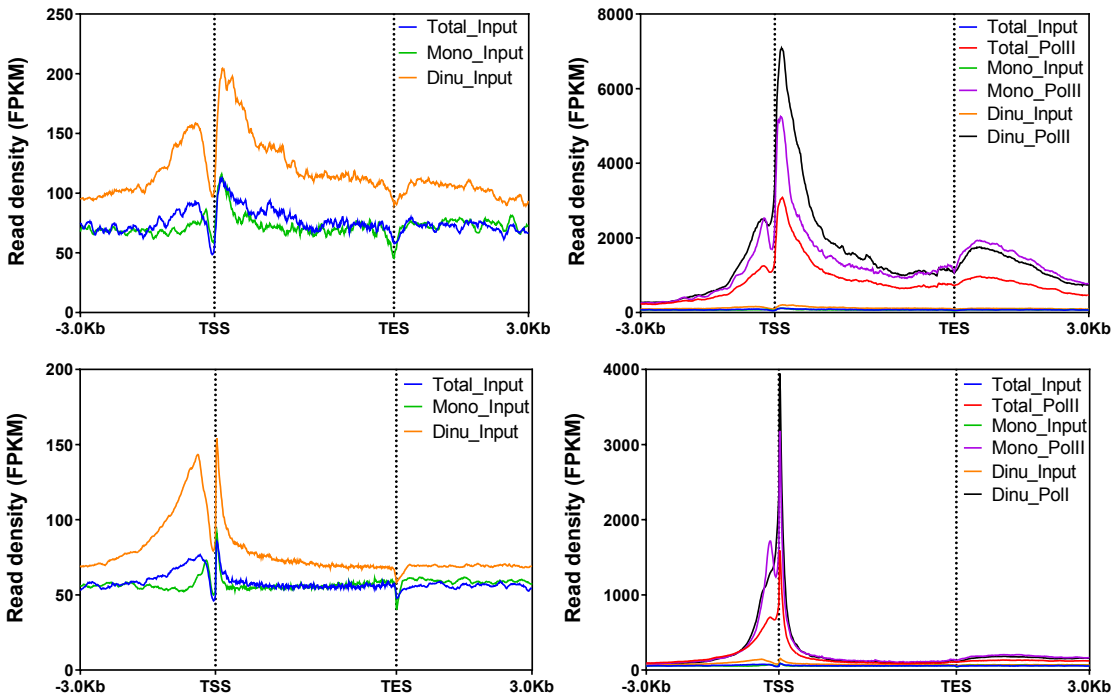


Figure 4.9 Metagene analysis of Input and IP signal across mRNA gene bodies. Genes were sorted into three groups depending on level of expression. Graphs represent the highest (top) and lowest (below) expression groups.

Again, the metagene profile of RNAPII IP for divergent mRNA indicates the mono- and di- nucleosome fractions produce a more highly defined peak at the gene TSSs than the total precipitated DNA (figure 4.8, right panel). This might suggest that the nucleosomes at active transcription start sites, closely associated with RNAPII, are regularly spaced, and linker DNA between nucleosomes is easily digested suggesting low levels of compaction. The "doublet" seen in the dinucleosome fraction profile is expected as the predominant fraction precipitated with RNAPII would likely be the first two nucleosomes bound to paused polymerase sitting before the first regular nucleosome.

Analysis of PROMPTs (figure 4.10) shows the regular spacing of nucleosomes and association with RNAPII also occurs for other species of RNAPII transcription. The overall lower signal for both regular nucleosomes and associated polymerase may be partly due to lower transcription levels, but also potentially denser packaging of chromatin, perhaps to prevent aberrant transcription. Both nucleosome and RNAPII patterns are also visible in eRNA metagene profiles (figure 4.11); however both are much less well defined than for mRNA, suggesting both less regular nucleosome spacing than at promoters, and lower RNAPII accumulation.

These results suggest that for most future NetMNase-seq experiments, sequencing the mononucleosome fraction only will be sufficient due to the high resolution and low background for individual nucleosomes and other factors.

Further NetMNase-seq experiments performed using different RNAPII antibodies to specific CTD phosphorylation states (S5P) have since shown to have a similar profile distribution to total RNAPII (Takayuki Nojima, personal communication). This is unlike mNET-seq, where sequencing data from different RNAPII isoforms can be easily distinguished, but similar to ChIP-seq results for the same antibodies. This is perhaps unsurprising as NetMNase-seq including an IP stage bears many similarities to native ChIP-seq.

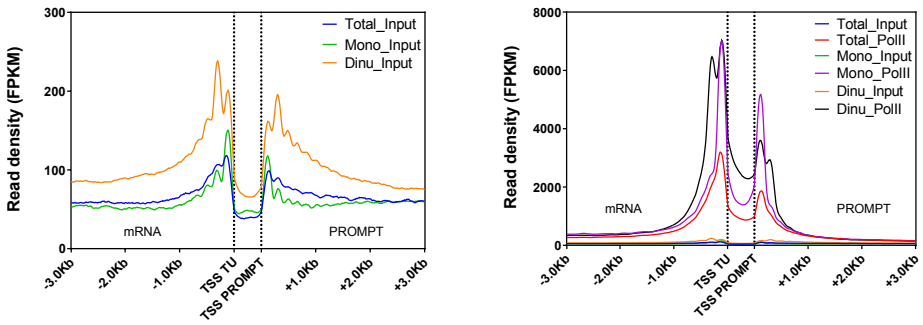


Figure 4.10 Metagene analyses showing PROMPT-mRNA pairs. PROMPTs show a similar overall pattern to mRNA but with an overall lower signal level.

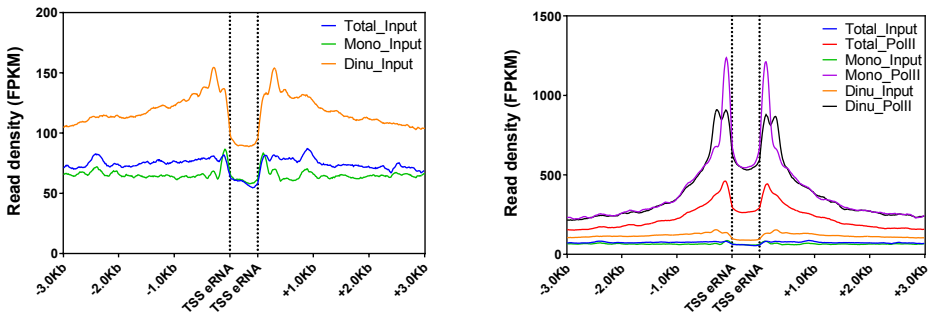


Figure 4.11 Metagene analyses showing histone arrangement around eRNAs. Regular spacing of the first and second nucleosome around enhancers producing eRNA is detectable.

4.7 Discussion

NetMNase-seq was developed to allow direct comparison of mNET-seq RNA sequencing of nascent transcription with nucleosome occupancy and compaction, under the same isolation conditions for both techniques, and even using the same cells. This is important, as MNase-seq techniques have been shown to vary considerably in identified nucleosome positioning depending on isolation conditions, cell type, cell cycle stage, and library preparation and analysis approaches (Flores et al. 2014). The protocol can function as an MNase-seq or native ChIP-seq like protocol depending on the addition of the immunoprecipitation step.

One advantage NetMNase-seq appears to have over ChIP-seq, for certain antibodies, is in resolution. In chapter 6, figure 6.15 shows examples of H3K4me3 histone signal over promoters. Here NetMNase-seq is able to resolve to a greater degree than the corresponding ChIP-seq the individual nucleosomes with this modification surrounding the TSS. This sensitivity may also be exploited to allow for this protocol to be carried out using fewer cells. Trials in our and other laboratories are currently working to minimize the number of cells required for mNET-seq, and similar refinements can be made to NetMNase-seq. In its current state, the NetMNase-seq protocol requires approximately 7-9 PCR cycles to amplify DNA precipitated from $5x10^6$ cells using histone antibodies. Firstly, trials with lower DNA amounts have shown that satisfactory libraries can be generated using up to at least 18 cycles. Secondly, antibody concentrations to extract the maximum DNA per cell can be further optimized.

Column purification instead of gel extraction would also minimize DNA loss in size purification; actual visualisation of the mononucleosome fraction is no longer necessary with every experiment. All these protocol modifications could reduce the number of cells necessary for sequencing to near single-cell levels.

The protocol does suffer from a number of limitations however. DNA co-immunoprecipitation is only possible for proteins sufficiently strongly associated with DNA to remain bound to chromatin under the 1 M urea wash conditions used in mNET-seq. This buffer could be modified, although direct comparison with mNET-seq results would be less definite.

NetMNase-seq with RNAPII IP may have limited functionality. The results obtained from this experiment are essentially a measure of both RNAPII and nucleosome states, and therefore could only provide information on the fraction of RNAPII closely associated with nucleosomes. This may provide too many independent, confounding factors to deliver much useful information for wider questions. However, it does produce the interesting result that some fraction at least of transcribing RNAPII is closely associated with *in situ* nucleosomes. This would indicate RNAPII transcribes genes with nucleosomes still in place, and they are recycled/remade quickly around it.

Cutting only 0-75 bp areas of the DNA gel could be trialed, to see if any DNA is present in RNAPII IP fractions. This protocol would essentially be an RNAPII ChIP-seq. There may however be problems with loss of unique sequence alignment below 20 nucleotides.

In future experiments using histone IP (see figures 6.13-18) mono- and di-nucleosome fractions after IP could be analyzed separately to distinguish marks associated with single nucleosomes or higher order structures. A similar, recently devised technique has been used to differentiate certain modifications of single or di-nucleosomes at promoters (Lorzadeh et al. 2016)

It would also be advisable to fully titrate the effects of the time of digestion on both Input and IP fraction distributions. Personal experience has shown that extent of MNase digestion of chromatin pellets can vary depending on MNase batch, extent of agitation of the pellet by vortex and any remnant lysis buffers. It is therefore necessary to strictly standardize the procedure so results do not vary between experimenters. A time titration of digestion would provide an idea of how different sequencing profiles can appear depending on extent of digestion (Mieczkowski et al. 2016). For example, a histone modification associated with heterochromatin may predominantly precipitate with long, protected DNA fragments and so be largely absent from mononucleosome fractions used for sequencing. However, over-digestion of the chromatin pellet may cause this DNA to be fragmented to mononucleosome sizes and so change the sequenced profiles. Under-digestion has also been shown to reduce read density and coverage (Flores et al. 2014). In this sort of situation, a digestion time course may actually provide a useful scale of how compacted/protected is a region of DNA that different modifications are situated in.

Several other techniques generate comparable data to NetMNase-seq. While chemical mapping of nucleosomes has some advantages over MNase digestion

analysis (Voong et al. 2016; Voong et al. 2017), it requires introduction of transgenic histone variants, limiting its efficacy in mammalian cells. The technique also essentially maps *in vitro* nucleosome positioning, making comparison with RNAPII and nascent RNA mapping less applicable. It also does not guarantee equal distribution of modified histones throughout the genome, and requires larger cell numbers per library. Comparison of the results of both techniques should however help to identify "fragile" nucleosome positioning at the TSS.

ATAC-seq provides different information and is in some ways the reverse of MNase-seq approaches. Reads represent both DNA accessibility around nucleosome positions and exposed open DNA stretches (areas available to transposase insertion). MNase-seq is at present better at defining the beginning and end of DNA sequences around a nucleosome, and can more readily distinguish between histones and other factors bound to DNA. ATAC-seq is also not so readily able to define nucleosome-depleted regions, for example at the TSS, without in-depth bioinformatic pattern analysis (Meyer and Liu 2014; Zentner and Henikoff 2014). Also the transposase reaction is performed on isolated nuclei meaning results are not so readily comparable to RNAPII/RNA mapping techniques with different isolation conditions. The current protocol for ATAC-seq does not allow for an immunoprecipitation step. This technique could also be used in complement with NetMNase-seq, as different information is obtained. Both techniques have different inherent sequence cleavage preferences and this may introduce read biases.

Chapter 5

CTD Threonine 4

5. CTD Threonine 4

5.1 Introduction

mNET-seq results for T4P show very little signal across the body of mRNA genes. However, a peak in signal is observed downstream of the TES, in the same region as the drop in signal seen in ChrRNA caused by transcription termination (Schlackow et al. 2017) (figure 5.1). This suggests T4P is a signal marker triggering termination or that it is a mark of chromatin-bound polymerase post termination. mNET-seq results from our laboratory from cells post CPSF73 siRNA knockdown show that the T4P signal diminishes in the region observed in wild type cells, and often increases further downstream. ChrRNA signal post siCPSF73 treatment increases downstream of the termination site, but still diminishes at a further 3' point, often in the region of the siCPSF73-induced T4P signal increase (Takayuki Nojima, personal communication). This may indicate some form of secondary termination event, and that in the absence of cleavage and poly-adenylation stimulated by CPSF73, some factor causes a "failsafe" termination event further downstream. This could contribute to preventing deleterious transcription read-through. It is unclear whether the CPSF73-independent termination event is caused by some specific sequence or protein factor, or whether secondary termination is triggered by a positioning effect downstream of the WT poly-adenylation site. The areas of these "failsafe" T4P signals have little sequence or size similarity, varying from 20 to >2000 nucleotides.

5.2 Factors associated with T4P

A potential method to investigate protein factors associated with the T4P mark is the mass spectrometry throughput adaptation of mNET-seq. Any potential factors particularly associated with the T4P mark compared to any other modification could then be depleted by siRNA transfection and the effects on T4P mark positioning could then be studied with mNET-seq and ChrRNA-seq/RT-qPCR. This could also be done with a double knockdown of CPSF73 and the factors of interest. This could establish if any protein factors are either responsible for the phosphorylation of threonine-4 post cleavage, or whether T4P recruits factors that would trigger termination.

However, mass spectrometry analysis of my T4P CTD pull-downs identified very few co-immunoprecipitated factors – none of which were unique to T4P pull-downs alone. Loss of some factors as compared to other CTD pull-down datasets was perhaps to be expected - lack of splicing component association with T4P is compatible with yeast data (Harlen et al. 2016). However, association of some termination factors with T4P had appeared likely. It was speculated that any bound mRNA processing factors might be lost from polymerase very rapidly after cleavage and poly-adenylation. However, comparison of MS results from siLuc and siCPSF73 transfected cells showed no increase in factors precipitating with T4P (see appendix 1.1). This may suggest that threonine 4 is only significantly phosphorylated post removal of most processing factors, or that the phosphorylation of this mark acts as a steric repellent of factors that would otherwise bind to the CTD.

It is possible that the salt or urea concentrations used in isolating chromatin disrupt protein factor interactions with T4P, if these interaction affinities are of a weaker magnitude to those with other CTD modification. Alternatively, it was speculated that the presence of antibody and/or bead proteins in MS samples might mask low levels of interacting factors. Different elution protocols using urea washes or other elution buffers to separate RNAPII and complexes from antibody and beads prior to MS processing were trialled but these were found to also elute significant levels of antibody (data not shown). In any case the MS equipment is capable of distinguishing peptides from many proteins present at very low levels despite high levels of contaminants such as antibody chains or keratin, meaning factors should have been identified if they were present.

5.3 CRISPR-Cas9 employed to generate mutant strains lacking "failsafe" region.

Since identification of protein factors linked to T4P proved unfruitful, an approach to investigate the sequence and positioning effect of the "failsafe" region was devised. This involved the use of CRISPR-Cas9 to delete the "failsafe" region from a selection of model genes, and then check for further termination defects on CPSF73 depletion. If the deleted region itself was inherent to the termination process, a more significant termination defect and read-through into downstream genes might be observed, as seen upon herpes and influenza virus infections (Nemeroff et al. 1998; Rutkowski et al. 2015), osmotic stress (Vilborg et al. 2015), and SETD2 inactivation (Grosso et al. 2015). If the effect is purely positional, any aberrant transcription might terminate a similar distance

downstream as seen in the wild type strain under siCPSF73 conditions, despite the absence of the precise sequence.

Firstly, a selection was made of two genes that provided good targets for CRISPR-Cas9 editing. These genes needed to fulfil a number of criteria. Firstly they needed to be good examples of the effect of siCPSF73 transfection causing diminution of the original mark and increase in T4P in a downstream region, with a corresponding termination defect observable in ChrRNA. Preferably they should be significantly expressed in the cell line chosen. The gene and potential regions downstream of wild type termination sites should not overlap with other genes, at least genes with any mRNA expression in that cell line. The "failsafe" regions should provide a good selection of targets for CRISPR and be of a reasonable size that CRISPR deletion occurs with high enough frequency. Too small a region may increase the chance of non-homologous end-joining removing part of or the entire WT termination region as well. Two choices with slightly different characteristics may also help to determine internal variation.

GAPDH and *Cyclin B1* (*CCNB1*) were chosen. *GAPDH* provides one of the clearest examples of the effect described above: a clear section of T4P signal that shifts to a non-overlapping region downstream on CPSF73 depletion. *CCNB1* provides a subtler example where the WT signal region is still partially visible in the siCPSF73 library, and the level of signal change in the "failsafe" region is more a matter of degree. The target regions are highlighted in figure 5.1. *GAPDH* is believed to terminated via a pause-type mechanism, whereas *CCNB1* was identified as having a CoTC type termination element (Nojima et al. 2013). The

selection of these two genes might therefore allow investigation of the similarities and differences in a secondary failsafe termination.

Next a suitable cell line was chosen. This line needed to be suitable for deletions of all alleles. To maximize the chance of this happening, a haploid or at most diploid cell line was preferable. Examples considered include eHAP-1, RPE-1, and HEK293. The HeLa ploidy state makes it particularly unsuitable. Preferably the cell line should also be suitable for nuclear and chromatin isolation as used in ChrRNA-seq and mNET-seq analysis. Tests were carried out to establish in which of these lines it was possible to separate chromatin and nucleoplasmic fractions. HEK were found to be unsuitable. The eHAP cell line passed fractionation tests and has only one copy of each gene, and so should be the easiest in which to ascertain total loss of the target regions.

gRNAs were designed to target the 5' and 3' ends of the highlighted regions. These were selected to take into account as much of the region as possible, so not leaving part of the "failsafe" intact, without overlapping with the WT termination region. In *CCNB1* the beginning of the "failsafe" region was defined as the first peaks to consistently show up-regulation in the siCPSF73 libraries. Multiple gRNAs for the same target up- or downstream of the target were transfected in tandem to maximize chance of cuts (figure 5.2).

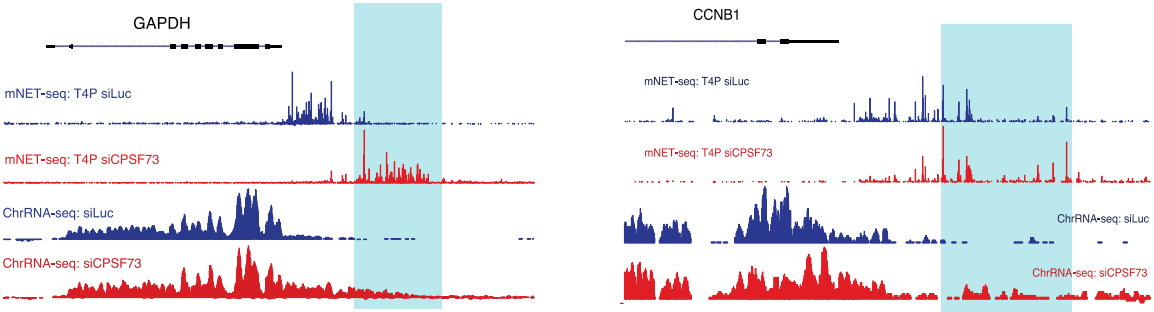


Figure 5.1 mNET-seq and ChrRNA-seq data shows T4P is a potential termination mark. Data produced by Takayuki Nojima. The region of the potential "failsafe" termination region is highlighted in blue, for the two selected target genes, *GAPDH* and *CCNB1*.

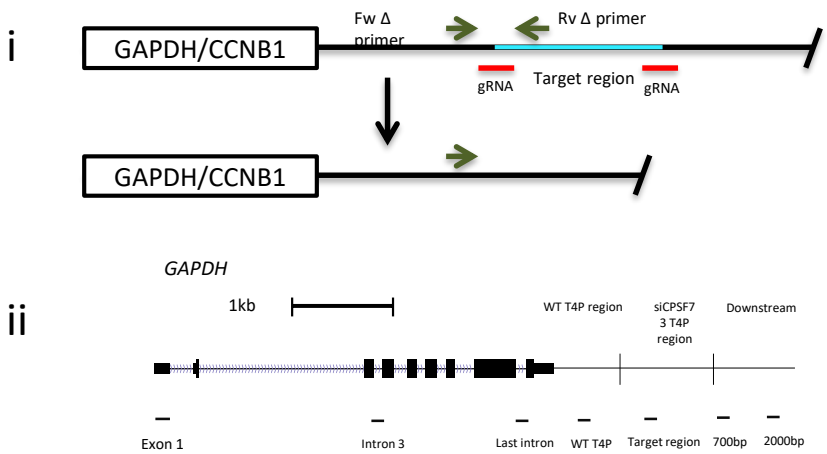


Figure 5.2 i) Schematic of gRNA and strain deletion check primer locations around the target locations. gRNAs (red) were designed to cut at the borders of the target "failsafe" region (blue), with deletion check primers (green) designed to amplify a region only if the deletion has not occurred. ii) Schematic of the *GAPDH* gene showing positions of primers used for ChIP, RT- and qPCR. Primers over Exon 1, Intron 3, the last intron, wild type termination region, target "failsafe" region, and 700 bp and 2000 bp downstream of the target region are marked.

To screen for potential clones, PCR primers overlapping with the target region, one outside, one in, were employed (figures 5.2 & 5.3). A positive result would appear as a loss of a band. This can then be checked by qPCR and sequencing. Despite acceptable levels of transfection judging by GFP control transfections and Cas9 expression levels tested by western blot (data not shown), I found a very low rate of generation of positive colonies, with only one out of a hundred colonies for *GAPDH* actually having the deletion, and none for *CCNB1*. It was also noted on testing potential gRNAs for cutting facility using an HRMA assay that an unusually low number of gRNAs with good *in silico* predictions cut well *in vivo*. This could perhaps suggest something about this region makes gRNA binding, Cas9 cleavage or even non-homologous end-joining difficult. If this were chromatin compaction or a similar effect, it might help to explain a failsafe termination, with nucleosomes forming a barrier to further transcription.

Nevertheless, deletion of target region was achieved for *GAPDH* (figures 5.3 & 5.4). This was confirmed by sequencing (appendix 2).

5.4 "Failsafe" termination appears to be positionally determined

Chromatin immunoprecipitation was used to examine the effects of deletion of the "failsafe" region on termination, upon transfection of CPSF73 or luciferase control siRNA (figure 5.5). RNAPII signal in both WT and deletion strains is increased in the WT termination region upon siCPSF73 transfection. The deletion strain self-evidently lacks ChIP signal within the deleted region. 700 bp downstream of the "failsafe" region, the siCPSF73 termination defect is

significantly greater in the strain lacking the targeted region compared to the equivalent WT strain. This is not the case in the siLuc conditions, as the standard termination region is still in effect. However, in the deletion and siCPSF73 condition, the increased signal diminishes 2000bp downstream, suggesting a later termination is still in effect in the deletion strain.

RT-qPCR analysis of chromatin RNA under the same conditions as the ChIP experiment shows similar results (figure 5.6). CPSF73 depletion causes an increase in RNA in and downstream of the WT termination region, and this effect diminishes by the 700bp downstream primer set. In combination with the deletion of the failsafe region, this effect is extended. However by 2000bp downstream differences in RNA level between all four conditions are minimal, suggesting that “failsafe” termination is a positional effect rather than due specifically to the sequence of the deleted region.

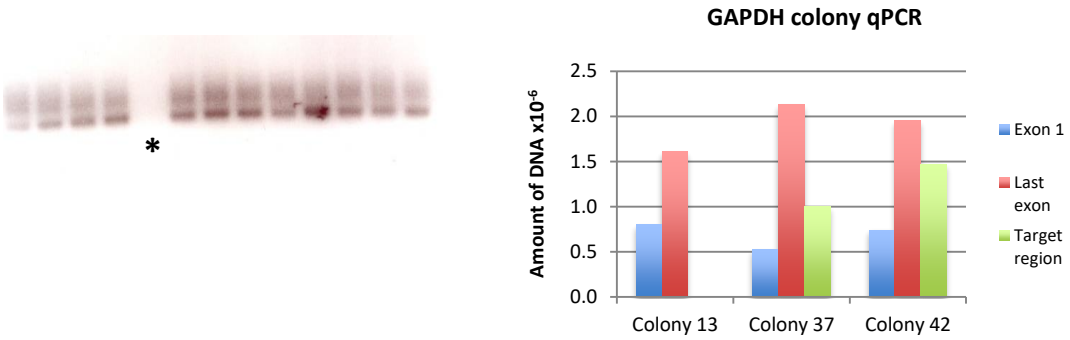


Figure 5.3 Colony PCR evidence of target deletion in one strain. Loss of the band (*) in a colony PCR using the deletion check primers indicated deletion of the target region. qPCR using primer pairs flanking and internal to the region also suggested deletion had occurred (colony 13).

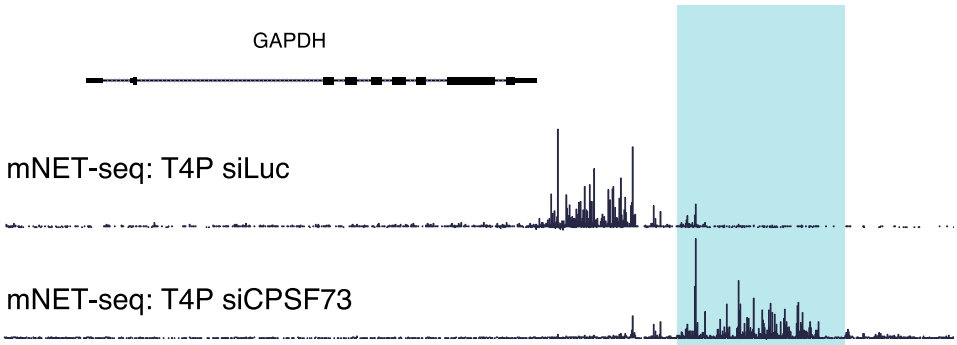


Figure 5.4 Schematic showing exact region deleted in the successful strain.

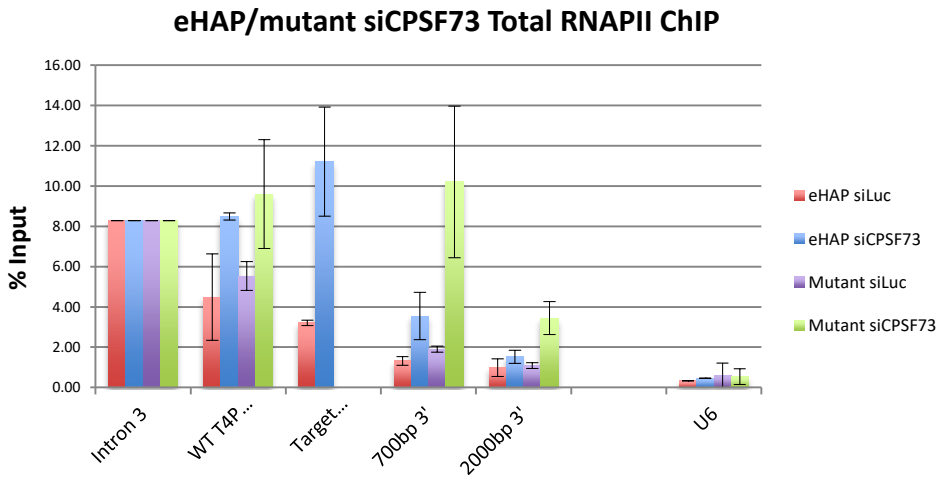


Figure 5.5 Total RNAPII (MABI0601) ChIP in wildtype and deletion strain cells, following siLuc or siCPSF73 transfection. Error bars from two repeats.

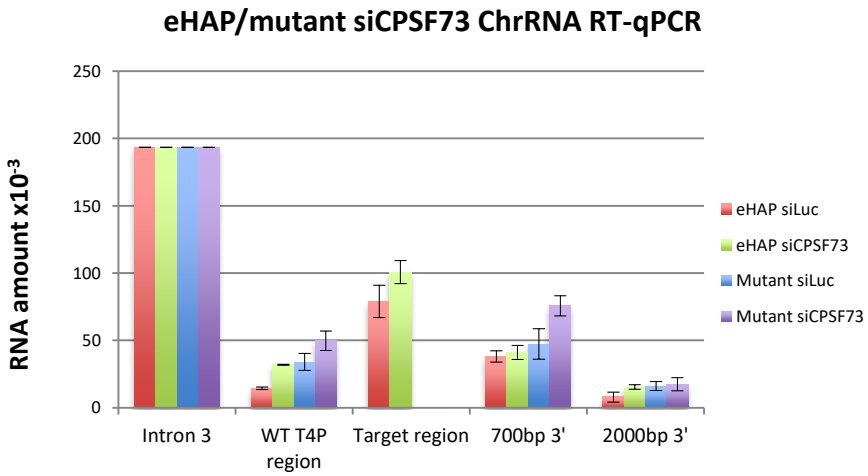


Figure 5.6 RT-qPCR on CPSF73 depletion in WT and deletion strains.
Error bars from two repeats.

5.5 CPSF73 depletion alters nucleosome coverage downstream of the pA site

In order to further explore potential causes for a secondary termination mechanism, NetMNase-seq was used to investigate nucleosome configuration downstream of the pA site. It was reasoned that by comparing wild type and CPSF73 depleted cells, it might be possible to discover changes in nucleosome profile triggered by a primary termination failure.

A dip in read density at mRNA TES sites has been seen before for WT mononucleosome Input libraries (figures 4.9, 5.7). A similar drop in signal at the TES has been seen for mNET-seq signal, where it may be caused by rapid cleavage and/or increased polymerase speed at the pA site. This trough appears to be a genuine region of low read density, and not an artefact of defining the TES, as it is not seen at exon boundaries or other features when they are used to define boundaries in metagene profiles. This suggests there is a nucleosome depleted or fragile nucleosome region at the pA site (Mavrich et al. 2008). The cause of this is unclear. Speculative reasons for maintenance of this region include DNA sequence bias disfavouring nucleosome formation, RNAPII pausing and occupancy at the 3' end, and perhaps termination factor recruitment, in a similar way to transcription and activator binding at promoters and enhancers.

Downstream of the TSS detecting regular nucleosome structures becomes difficult due to the variability between genes in length and nucleosome spacing. However, the maintenance of an NDR at the TES may allow the detection of regular nucleosome peaks around it.

Comparing siLuc and siCPSF73 samples, an increase in mononucleosome signal is evident after the TES upon depletion of CPSF73 (figures 5.7, 5.9). This would indicate that cleavage failure and/or aberrant transcription read-through somehow triggers chromatin condensation. The nucleosome-depleted region at the TES seen in WT cells is not significantly altered by siCPSF73 treatment.

Also observable is depletion in signal of the first few regularly spaced nucleosomes at the TSS following siCPSF73 treatment, but not the -1 positioned nucleosome (figure 5.8). This may indicated a feedback loop, where termination defects alter transcription initiation and the maintenance of 5' nucleosome positioning. This TSS depletion is not seen in the peak associated with the TSS of PROMPTs, suggesting it is specific to mRNA genes (figure 5.10).

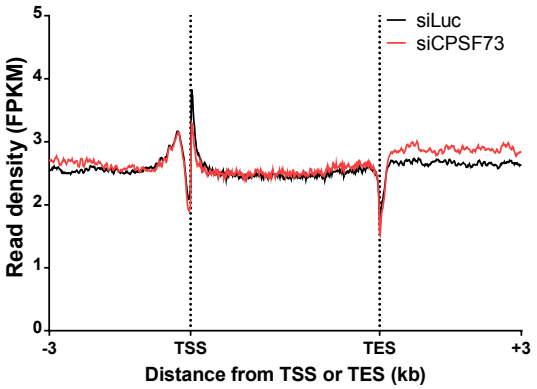


Figure 5.7 Metagene profile of Input NetMNase-seq signal comparing siLuc and siCPSF73 treated cells. Metagene profiles by Michael Tellier.

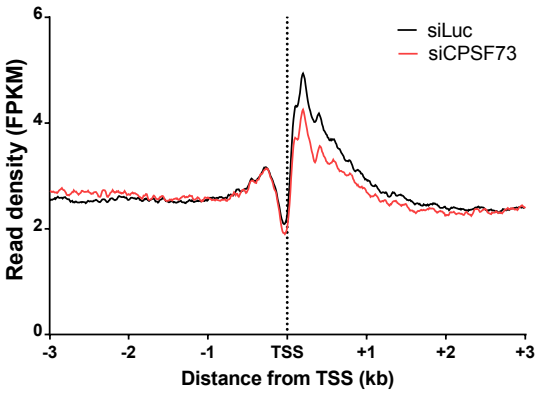


Figure 5.8 Magnification of TSS region from figure 5.7.

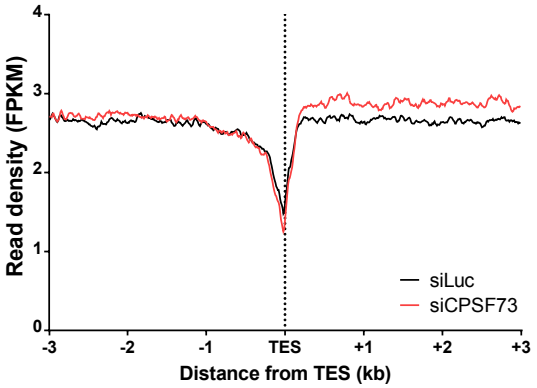


Figure 5.9 Magnification of the TES region from figure 5.7. The increase in signal upon depletion of the endonuclease is specific to 3' of the TES (pA site).

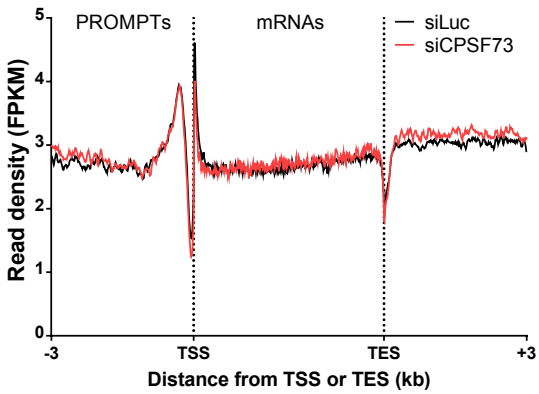


Figure 5.10 Metagene representation of divergent mRNA-PROMPT pairs. The nucleosome signal at +1 position at PROMPT TSS is unaffected by CPSF73 depletion.

5.6 Discussion

RNAPII ChIP and RT-qPCR results for *GAPDH*, comparing the WT and deletion strains, indicate that removing the region associated with the secondary T4P signal peaks does not prevent termination further downstream. This, coupled with the lack of sequence or size similarities between this region in different gene 3' UTRs suggests that the CPSF73-independent termination mechanism occurs within a given distance of the pA site, and is not determined by sequence.

The apparent difficulty in gRNA-Cas9 cutting in this region may suggest some sort of physical barrier. This could be to either protein-DNA interaction, or DNA unwinding, allowing the gRNA:DNA hybrid to form. Perhaps this is due to condensed chromatin. However, gRNAs have been used before to cut constitutive heterochromatin, and clearly nucleosomes do not provide a prodigious barrier to Cas9. No very marked change in signal in this region is seen in WT NetMNase-seq data, although this generally will only provide mononucleosomal information, or at least not information on the overall physical/conformational packaging of chromatin.

Failure to produce a *CCNB1* mutant is regrettable, as it would have given a second example with which to test if the "failsafe" termination is positional. *CCNB1* is also thought to terminate using a CoTC-type mechanism. This gene may therefore have shown some differences in termination upon region deletion, compared to the pause-type termination mechanism of *GAPDH*.

NetMNase-seq shows a significant increase in nucleosome signal in the 3' flanking region following CPSF73 depletion. This may suggest a downstream mechanism triggered by either loss of the endonuclease or the absence of the cleavage reaction itself. Perhaps factors sponsoring heterochromatin formation are recruited. Alternatively, if the CPA complex and other factors remain bound due to the lack of cleavage, nucleosome-remodelling enzymes may remain bound to RNAPII and chromatin further downstream than in WT conditions, triggering chromatin compaction. Clearly the NDR at the TES is maintained even in the absence of cleavage.

The loss of TSS-associated nucleosome signal upon CPSF73 knockdown may indicate feedback from termination to initiation, perhaps preferentially on the same gene (Andersen et al. 2013). Perhaps efficient termination is necessary for maintenance of regularly spaced nucleosomes at TSS positions.

The exact function of the T4P mark downstream of the pA site is still as yet uncertain. It may recruit termination factors, or sterically hinder binding of any proteins post-termination. The lack of peptide identification for CPA factors in the T4P MS experiments does not necessarily mean this mark is not associated with cleavage and poly-adenylation complexes. Since no mass spectrometry experiments with total polymerase or any individual phosphorylation have identified CPA factors, it seems likely that these complexes are either transiently or weakly bound, compared to the spliceosome for example. Further co-IP experiments could therefore be tried, using the MNase digestion on nuclear

extract rather than chromatin to avoid 1 M urea buffers which may disrupt CTD interactions.

It may also be useful to delete WT termination regions (as indicated by T4P peaks in untreated cells). This may have an effect on distance of termination (or at least T4 phosphorylation) downstream of the pA site. However, given the pA signal would still be intact it is quite likely no changes would be seen.

Given the significant effect seen in both mNET-seq and NetMNase-seq of CPSF73 depletion, it would be interesting to test the effects of depletion of other CPA factors on T4P signal and chromatin organisation. For example, CstF64 and XRN2 depletion have both been shown to produce termination defects, and these may alter T4P accumulation in different ways. Different viruses and cell stresses interfere with host transcription termination in different manners, causing different levels of aberrant transcriptional read-through, and following the effect these mechanism have on T4P profiles may be instructive (Nemeroff et al. 1998; Rutkowski et al. 2015; Vilborg et al. 2015).

Additionally, repeats of the NetMNase-seq data are necessary before more definite conclusions can be drawn. It would also be interesting to investigate whether particular histone isoforms are associated with the observed 3' signal increase – are transcription-repressive or heterochromatin-promoting post-translational modifications upregulated in the 3'UTR following loss of termination factors?

It should be noted that an increase in mononucleosome signal post-TES is a slightly ambiguous result that needs further investigating. It may suggest a simple increase in the number of nucleosomes positioned in regions 3' of termination sites, leading to increased MNase protection and therefore increased read density. However, it may also be caused by a shift from higher order chromatin to mononucleosomes, leading to increase in DNA sequences from 3' regions being isolated from mononucleosome gel bands. Investigating individual histone modification presence in this region would help to settle this issue. Alternatively an MNase digestion time course could help to resolve how easily different genomic regions are digested.

Chapter 6

CTD Tyrosine 1

6. CTD Tyrosine 1

6.1 Introduction

Initial mass spectrometry results generated by our laboratory have revealed potential roles for Tyrosine 1 phosphorylation in chromatin remodelling and DNA repair (Takayuki Nojima, personal communication). Y1P co-immunoprecipitates with a number of proteins involved in chromatin remodelling and DNA damage repair, most notably a number of factors involved in the repair of DNA lesions caused by UV damage. The main factors identified that are associated with these roles are as follows.

BAZ1b/WSTF is an atypical tyrosine kinase and has been found to phosphorylate the C-terminal residue of histone H2AX (Y142) (Xiao et al. 2009). It acts as a chromatin remodeller and as part of the WICH complex to reconfigure irregular chromatin into regular nucleosomal arrays (Bozhenok et al. 2002; Poot et al. 2005; Barnett and Krebs 2011). Through this process it can act as a transcriptional regulator, and has been shown to regulate the levels of RNA Polymerase I and III transcripts (Cavellan et al. 2006; Percipalle and Farrants 2006; Percipalle et al. 2006; Vintermist et al. 2011). It is also recruited by PCNA to DNA replication foci (Poot et al. 2004).

SNF2H/SMARCA5 is the other half of the WICH complex, and is a helicase with nucleosome remodelling activity (Sun et al. 2016). It is also part of a nucleolar remodelling complex (NoRC) recruiting histone-modifying enzymes to ribosomal

genes (Santoro et al. 2002). Genome-wide studies find SNF2h accumulates at promoters and remodels nucleosomes adjacent to transcription factor binding sites (Morris et al. 2014; Wiechens et al. 2016). Members of the SNF protein family have been found to remodel nucleosomes around sites of NER repair (Hara and Sancar 2002; Zhao et al. 2009; Watanabe et al. 2017).

DNA damage-binding protein 1 (DDB1) forms a complex with DDB2 to bind photo-dimers and other UV damage products (Sugasawa et al. 2005). It also acts as an adaptor molecule for CUL4A (He et al. 2006; Higa et al. 2006) (see below) leading to ubiquitinylation of histones at sites of damage (Kapetanaki et al. 2006; Guerrero-Santoro et al. 2008). It also recruits other remodelling factors to UV damage sites, for example Ino80 (Jiang et al. 2010; Fei et al. 2011).

CUL4A is a core component of a number of E3 ubiquitin-ligase complexes. These can contain either DDB1 (Hu et al. 2004) or DDB2. Different adaptor molecules direct the complex to different sites, and can cause ubiquitinylation of XPC (inhibiting adduct removal (Liu et al. 2009)) and histones H3 and H4 in nucleosome remodelling (Wang et al. 2006). It is required for efficient DNA-replication coupled and independent nucleosome assembly (Han et al. 2013).

PHIP/DCAF14 is one of a number of DDB1 and CUL4A-associated factors (DCAFs) shown to act as substrate receptors (Lee and Zhou 2007). It may have an anti-apoptotic role in the damage recognition pathway (De Semir et al. 2012).

The association of these proteins with Y1P raises many questions. Since Y1P is particularly associated with the TSS (and eRNAs) (Descostes et al. 2014), could chromatin remodelling factors like BAZ1b and SNF2h be involved in the maintenance of nucleosome depletion or regular spacing of nucleosomes around the promoter? They might therefore affect expression levels as they have been found to do for RNAPI and III genes. Chromatin remodelling to facilitate transcription might then lead to increased DNA damage (Perera et al. 2016). NER factors may therefore be recruited to the same genomic locations as Y1P due to increased damage associated with open chromatin. There is also evidence that NER can be impaired by the binding of transcription factors to promoters (Sabarinathan et al. 2016). It may be necessary for increased repair factor recruitment to the TSS to deal with persistent damage.

However, the reverse could also be possible. Since other members of the SNF family have been found to remodel chromatin around sites of NER, perhaps stalled polymerase (perhaps causing phosphorylation at Y1 residues) could recruit both NER factors and chromatin remodelers. Could promoter proximal pausing recruit similar factors to stalled polymerase?

6.2 Repeat mass spectrometry identifies the same proteins associated specifically with Y1P polymerase.

Firstly I performed a repeat Y1P pull-down for mass spectrometry under the mNET-seq isolation conditions in order to confirm the initial mass spectrometry experiment (table 6.1 & appendix 1.2). The above listed proteins were again

identified as co-immunoprecipitating with Y1P CTD. They were not present in repeated T4P pull-downs I performed (chapter 5 & appendix 1.1), and were not significantly enriched in S5P or S2P samples (Takayuki Nojima, personal communication).

	SPECTRAL COUNTS		SPECTRAL INDEX (MIC SIn)	
Protein	No Ab	Y1P	No Ab	Y1P
SPT6		119.95		8.73E-06
DDB1		54.27		6.51E-06
mRNA turnover protein 4 homolog		16.00		5.05E-06
PHIP		67.50		4.52E-06
Serine/threonine-protein phosphatase PGAM5		11.95		4.44E-06
Regulation of nuclear pre-mRNA domain-containing protein 1A		13.95		3.76E-06
Nuclear cap-binding protein subunit 1		41.00		3.50E-06
RNA-binding protein Raly		8.91		3.48E-06
Cap-specific mRNA methyltransferase 1		35.00		3.20E-06
CUL4a		26.00		2.79E-06
Nuclear cap-binding protein subunit 2		6.95		2.19E-06
SNF2h/SMARCA5		30.00		2.08E-06
DHX15		28.00		1.95E-06
ATP-dependent RNA helicase DDX39A		8.88		1.59E-06
BAZ1b		40.00		1.59E-06
Heterochromatin protein 1-binding protein 3		13.76		1.43E-06

Table 6.1 List of top Y1P associated proteins as identified by mass spectrometry (ranked by Spectral Index and compared to no antibody control)**.** Histones and RNAPII subunits, and proteins present in the No Ab sample have been removed for clarity. DNA repair factors and chromatin remodellers are highlighted. For full list see appendix 1.2.

6.3 Western blot analysis shows no indication that BAZ1b is a CTD kinase

Since BAZ1b is a tyrosine kinase, an siRNA knockdown of BAZ1b was performed to test whether Y1P levels were affected (figure 6.1). This would indicate that BAZ1b was a kinase for Y1. No discernable change in Y1 phosphorylation could be seen on knockdown, but given that for other CTD residues seven or more

kinases have been found for one residue, this may be due to redundancy masking any effect. An *in vitro* kinase assay would better establish whether BAZ1b can phosphorylate CTD residues. Unfortunately the lack of readily available recombinant kinase domains for BAZ1b made this test beyond the scope of this project.

Somewhat surprisingly HP1α (a heterochromatic marker) levels appear to increase markedly on BAZ1b depletion, here shown from total cell lysate. This has also been shown for chromatin bound HP1α (Poot et al. 2004).

6.4 UV damage

Since MS experiments suggested an association between Y1P and NER factors, and NER is the main repair pathway for UV damage, it seemed apposite to test the effects of UV irradiation on Y1P levels, and transcriptional profiles.
Firstly, a choice needed to be made for UV treatment regimen. Both UV energy levels and incubation time after UV irradiation vary significantly in the literature. 20 Jm^{-2} was chosen as an acceptable medium dosage. Time courses were then performed to establish damage levels after exposure. γH2AX was used as a proxy measure of DNA damage. This indicated that significant levels of damage response were seen after 1-2 hours, and that levels of S139 phosphorylation did not change significantly between 4 and 24 hours (figure 6.2). Selecting 2 hours as a main time-point would therefore be adequate. Previous studies have shown that 2 hours is past the main transcriptional shutdown post-UV, but before full recovery (~18 hours). I therefore followed similar reasoning to (Boeing et al.

2016) that two hours incubation could provide information on protein factors involved in both early and late stages of the UV response.

A second time-course was performed to analyse CTD residue phosphorylation post-UV (figure 6.3). Western blot analysis showed that UV irradiation causes an increase in the ratio of hyper- to hypo- phosphorylated forms of the CTD after just one hour, as indicated by Total RNAPII doublet ratios. This happens in both the chromatin and nucleoplasm fractions. This time-course may also indicate some decrease in Y1P and S2P levels on chromatin post-UV treatment, and a corresponding small increase in nucleoplasm levels of these phosphorylations. This might suggest increased turnover of RNAPII with these modifications from the chromatin-bound fraction into the nucleoplasm following UV treatment. This correlates with the known ubiquitinylation and removal of some stalled RNAPII for degradation following UV damage.

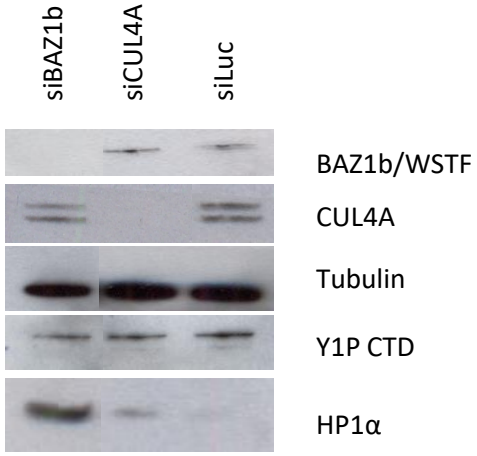


Figure 6.1 siBAZ1b and siCUL4A knockdown western blot. Samples prepared from whole cell extracts. Approximate molecular weights (kDa): BAZ1b – 170, CUL4A - 88, Tubulin 1-alpha - 50, PolII CTD/POLR2A - 220, HP1α – 23.

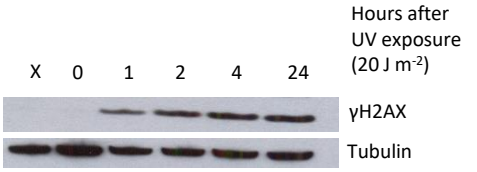


Figure 6.2 UV irradiation time-course. Cells were irradiated with 20 Jm^{-2} UV and harvested after 0, 1, 2, 4 and 24 hours post irradiation. X = no irradiation. Approximate molecular weights (kDa): (γ)H2AX – 15, H4 – 11.

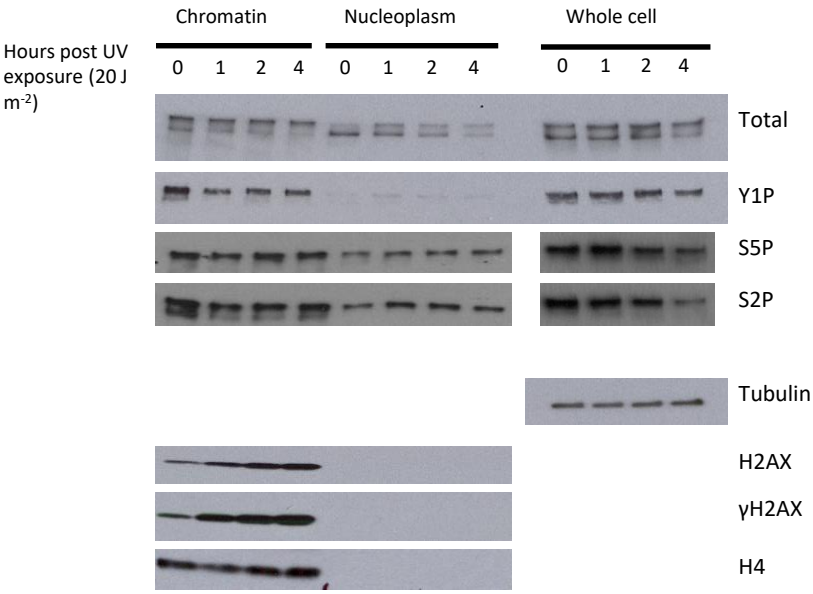


Figure 6.3 CTD phosphorylation post-UV irradiation time-course. Chromatin, nucleoplasm and whole cell extract were harvested from irradiated cells after 0, 1, 2 and 4 hours.

6.5 UV and damage foci do not co-localise

To investigate whether Y1P is associated with sites of DNA damage, confocal immunofluorescence was performed to examine whether tyrosine 1 phosphorylation foci localised to sites of UV induced damage (figures 6.4 & 6.5). As expected, the number and intensity of γH2AX foci increased significantly two hours post UV irradiation. Any change in Y1P signal level was too small to detect. In neither condition was there a significant co-localisation of Y1P and γH2AX signal, (Pearson's correlation coefficient –UV = -0.16, +UV = -0.13, 10-16 nuclei per sample). This suggests there is no strong association of CTD phosphorylated at tyrosine 1 and sites of DNA damage.

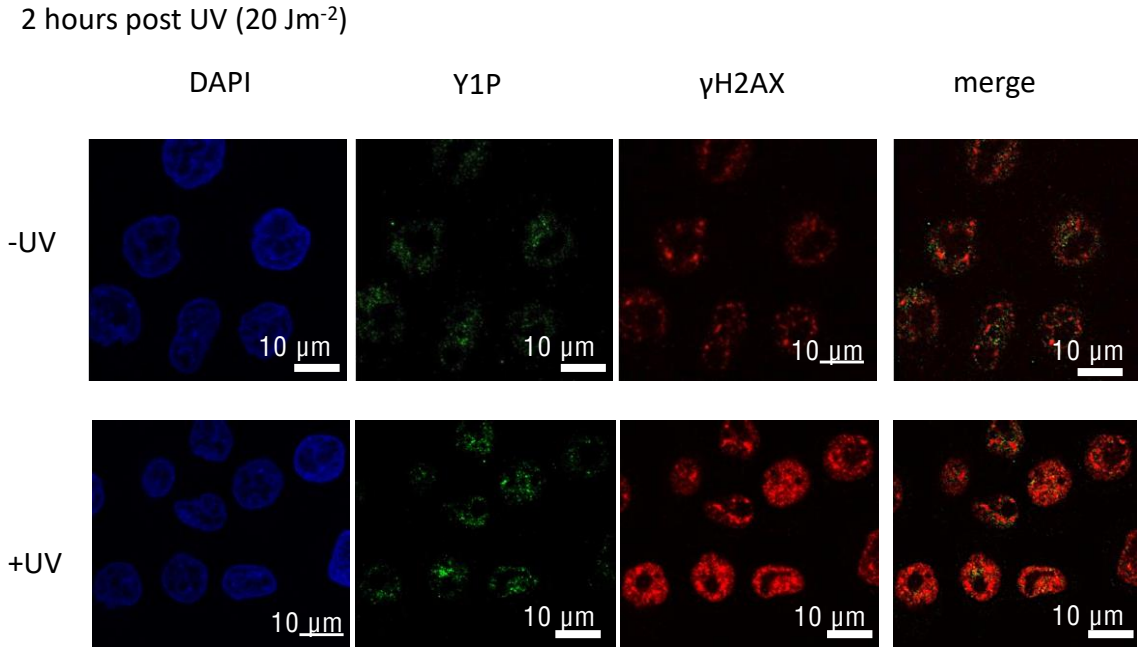


Figure 6.4 Immunofluorescence confocal imaging of Y1P CTD modifications upon UV damage. Fixed cells were probed for Y1P (green) and γH2AX (red) as a damage marker. γH2AX signal increased markedly on ultraviolet irradiation. Any change in Y1P signal after irradiation was smaller than variation between cells in the same population.

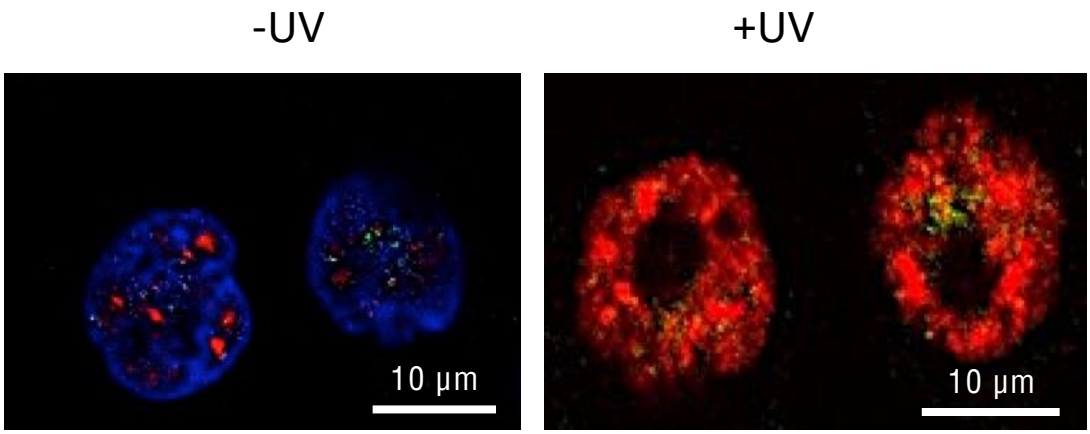


Figure 6.5 Expanded view of Y1P and γH2AX overlay.

6.6 UV damage causes a global elongation defect

To investigate how UV radiation might be affecting transcription and Y1P phosphorylation, Y1P and Total RNAPII mNET-seq after UV irradiation were performed by Dr Takayuki Nojima. Figure 6.6 shows the Total and Y1P signal profiles over the *YWHAZ* gene as a particularly marked example of the effect noted. A loss of the RNAPII signal in the 3' region of the gene is evident for both Y1P and Total, two hours post irradiation. No subsequent signal increase is seen further downstream as in the siCPSF73 T4P data. This effect is also shown in this metagene profile (figure 6.7), showing a significant drop in Y1P signal after the transcription end site. A small decrease in signal at the TSS is also noticeable. This indicates that UV irradiation is inducing an elongation defect. This effect is not simply a direct effect of photo adducts causing blockage in particular genes, as the effect is global and reproducible, and not therefore due to random placement of adducts. This agrees with the conclusions of the Svejstrup laboratory (Williamson et al. 2017), who also identified a global elongation defect using GRO-seq. Our data do show some differences however – the mNET-seq signal over gene bodies appears similar in UV and control samples, with most loss in signal near the TES. This differs from analysis of very long genes indicating a loss of transcription across the body of the gene. It should also be noted that the loss of signal is more evident in Y1P profiles than Total, indicating a specific effect acting on this RNAPII isoform.

RNA-seq experiments on chromatin-bound and nuclear RNA were also performed (figure 6.8). The profile over many genes does not appear to change

significantly on UV irradiation. However, certain gene profiles (e.g. *TAF9*, *CSTF3*, *CCNC, p53, KIAA0824*) show an increase in RNA level in 5' regions and a loss in 3' regions, indicative of an elongation defect. The loss of signal in a number of these genes occurs around known alternative poly-adenylation sites (Tian et al. 2007). This might suggest UV elongation inhibition leads to a switch to the short isoform of these factors. Longer genes may also show a more significant elongation defect.

The elongation defect is also seen globally in the metagene analysis of ChrRNA-seq, which shows increased signal at the TSS followed by decreased signal towards the 3' end, consistent with an elongation defect causing 5' build up of RNAPII. This indicates that there is a subtle but significant effect on elongation across a wide number of genes, not just the more extreme switches as seen in those genes mentioned above.

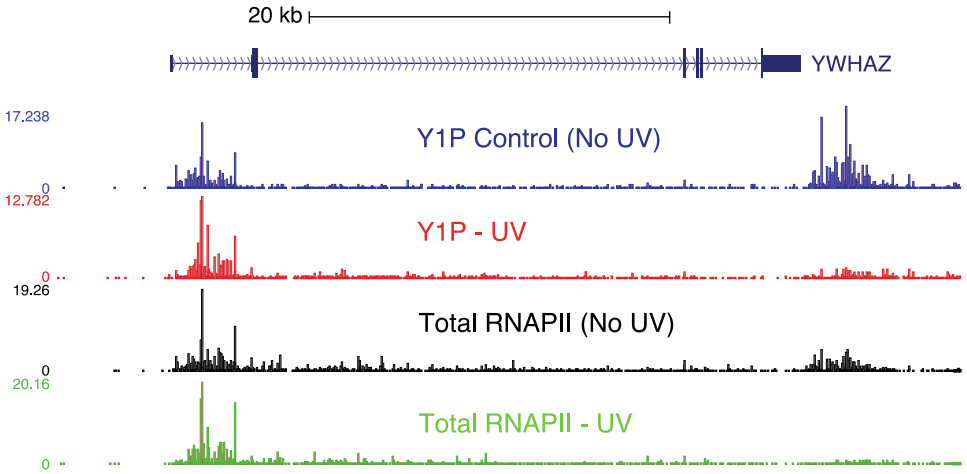


Figure 6.6 *YWHAZ* as an example of effects of UV irradiation on Y1P and Total RNAPII mNET-seq profiles. UV irradiation: (20 Jm^{-2}). Data generated by Takayuki Nojima

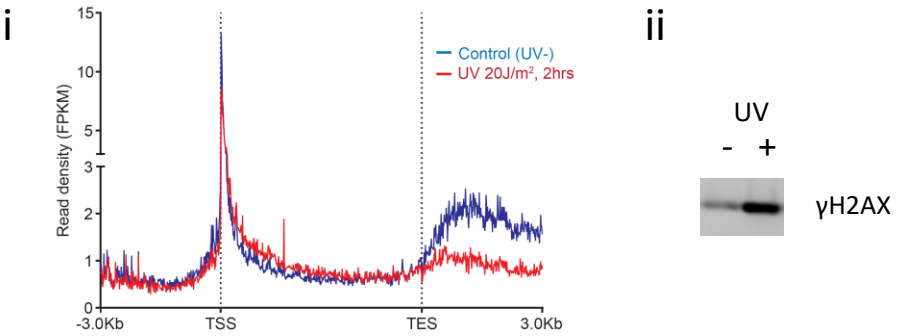


Figure 6.7 Metagene analysis of mRNA gene Y1P mNET-seq signal. i) Post irradiation, a significant decrease in signal is seen after the TES of mRNA genes. ii) Western blot confirmation of damage caused by UV irradiation in cells harvested for mNET-seq. γH2AX used as damage proxy.

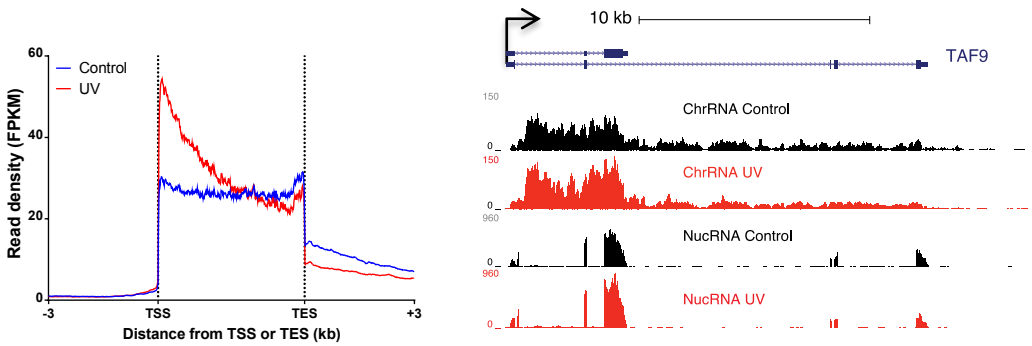


Figure 6.8 Chromatin and nuclear RNA sequencing library data. Metagene analysis of ChrRNA-seq over mRNA genes (left). *TAF9* (right) as single gene example showing a marked shift to 5' signal upon UV irradiation, consistent with expression of a short isoform. Data generated by Takayuki Nojima.

6.7 UV damage increases the association of the PAF complex with RNAPII

I next performed a further mass spectrometry analysis, of proteins associated with Total RNAPII, comparing no and 2 hour UV treatment. This was done to determine whether the level of previously identified NER factors in association with RNAPII increased after UV treatment. The experiment could also identify further factors binding to RNAPII upon DNA damage.

This approach revealed a significant increase in a number of proteins (appendix 1.3). Perhaps unsurprisingly these included the UV repair proteins identified in the original Y1P mass spectrometry (which showed a small, ~2 fold increase), as well as >25-fold increase in ERCC6 and 8 (CSA/B), two other proteins involved in coupling the Nucleotide Excision Repair pathway to RNAPII (Groisman et al. 2006; Boeing et al. 2016). Increase in peptide count for these factors acts as a positive control for the technique, indicating the MS isolation protocol can identify increases in UV repair factors upon UV damage.

But somewhat more interestingly most members of the PAF1 complex (PAF1C) were also markedly increased/present only in the +UV sample. This is in agreement with a previous proteomic analysis of RNAPII interactions (Boeing et al. 2016). PAF1C is composed of PAF1, CTR9, LEO1, CDC73, SKI8 (Zhu et al. 2005). All but SKI8 was significantly upregulated (or only present) in the UV irradiated MS sample.

PAF1C appears to have many roles in transcription, often seemingly contradictory. PAF1C is associated with S2P, S5P and unphosphorylated CTD

(Moniaux et al. 2006; Yu et al. 2015). It associates with elongation factors in HeLa (Kim et al. 2010; Chen et al. 2017) and yeast (Squazzo et al. 2002; Hampsey and Reinberg 2003), and knockdown of PAF1 increased promoter proximal RNAPII, and an increase in chromatin-bound S5P signal. However some reports suggest it is a negative regulator of promoter escape (Chen et al. 2015). Its role seems to depend on cellular context and component phosphorylation state (Takahashi et al. 2011). It is believed to regulate a number of histone modifications; causing H2B ubiquitinylation, and methylation of H3K4me3. Additionally it appears to regulate 3' end formation, and has been found in association with cleavage and poly(A) factors (Nagaike et al. 2011). PAF1 knockdown leads to reduction of S2P but not total RNAPII profiles at the 3' end (Mueller et al. 2004; Yu et al. 2015). Members of the complex have also been shown to accumulate in the termination region, where the loss of the mNET-seq signal is seen post-UV treatment (Yu et al. 2015; Yang et al. 2016).

A number of DNA and RNA helicases were also upregulated in the +UV sample. These include DDX54, which has been shown to regulate expression of ionizing radiation-induced genes (Milek et al. 2017), and DDX21, which is involved in RNAPI and II elongation control and R-loop resolution.

Unfortunately, although follow-up repeats and +/- UV Y1P pulldown MS experiments identified the same proteins as being upregulated on UV irradiation, they did not provide comparable overall levels of co-immunoprecipitation. Statistical comparison of the mass spectrometry data was therefore untenable.

Therefore in order to corroborate the above results, western blot confirmation of co-IP was attempted (figure 6.9). Detecting most of the proteins of interest proved difficult in the IP fraction. Perhaps this is due to low levels detectable only by sensitive mass spectrometry/low affinity antibodies/masking by bead smears. Low salt NET2 (100mM) washes were also trialed for the IP, to test whether less stringent washes would prevent loss of proteins associated with RNAPII. No greater co-IP ratio was seen.

However, chromatin bound proteins were detectable in the Input fraction, and did show an increase in band intensity for CTR9, PAF1, CDC73, and CSB upon UV irradiation. An increase in RNAPII associated CDC73 was also detectable in the IP fraction.

SPT6 appeared to be downregulated on UV damage in some MS experiments. This is in agreement with data from the Svejstrup laboratory suggesting it is degraded by the proteasome upon UV damage. SPT6 is another elongation factor and histone chaperone, and promotes the methylation of Histone 3 Lysine 36 (H3K36me3), a mark deposited intragenically in the wake of active transcription. The level of this protein associating with chromatin was also therefore probed by immunoblot, and found to diminish upon UV treatment.

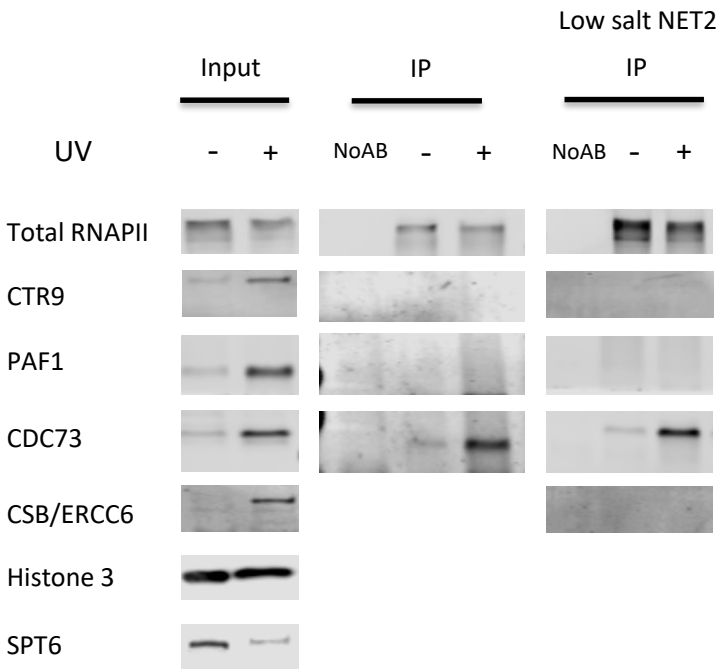


Figure 6.9 Western blot confirmation of UV irradiation mass spectrometry. Total RNAPII antibody (MABI0601) used for IP. Histone H3 loading control for Input chromatin fraction. Equal levels of RNAPII for IP. Low salt NET2 (100mM) was used to test effects of low stringency washes on co-IP. UV treatment: 20 Jm^{-2}. Approximate molecular weights (kDa): RNAPII/POLR2A - 220, Spt6 – 200, CTR9 – 135, PAF1 - 60, CDC73 - 60, CSB - 170, H3 - 15.

6.8 BAZ1b association with 3' ends does not alter upon UV damage

Chromatin immunoprecipitation was used to investigate whether BAZ1b levels at 3' ends were altered by UV treatment (figures 6.10 & 11). Conditions for BAZ1b ChIP were tested comparing siBAZ1b and siLuc treated samples to show the specificity of BAZ1b signal. UV- and non- treated cells were then tested for 3' signal levels in the *YWHAZ* gene, which showed the particularly marked loss of Y1P mark on UV treatment. No significant change was seen following UV irradiation, suggesting BAZ1b levels in terminator regions were unaffected and cannot account for the observed Y1P defect. This does not rule out changes in kinase activity or accumulation elsewhere from affecting gene expression, perhaps through chromatin remodelling.

6.9 PAF1C components are lost from chromatin at the 3' ends of genes

To investigate the effect of UV treatment on the profile of the PAF complex over the gene body, I performed Chromatin Immunoprecipitation over the *YWHAZ* gene (figures 6.10 & 12). Both PAF1 and LEO1 show a decrease in signal over the 3' and termination region (primers 5 and 6) upon UV irradiation. CDC73 may also show a decrease in signal at the 3' end of *YWHAZ* upon UV irradiation, although large variation between repeats makes this uncertain. This was somewhat surprising given that the PAF complex appears to increase in the mass spectrometry RNAPII pull down on UV treatment, and it is still not certain where on chromatin this increase is occurring.

However, this loss at the 3' end may help to explain the elongation defect apparent in mNET-seq and RNA-seq data, since PAF1C may function as a positive regulator of elongation.

It was anticipated that these results could be corroborated by generating ChIP-seq libraries to produce a genome-wide view of changes in PAF1C signal upon UV damage. Unfortunately, attempts to produce libraries with sufficient read density and distribution proved unsuccessful in the time frame allotted.

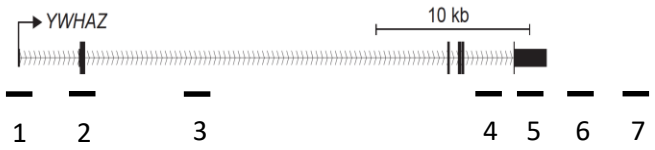


Figure 6.10 Schematic of the *YWHAZ* gene showing positions of primers used for ChIP. Primers over promoter (1), gene body exon (2), gene body intron (3) the last intron (4), last exon (5), and 700 bp (6) and 1500 bp (7) downstream of the target region are marked.

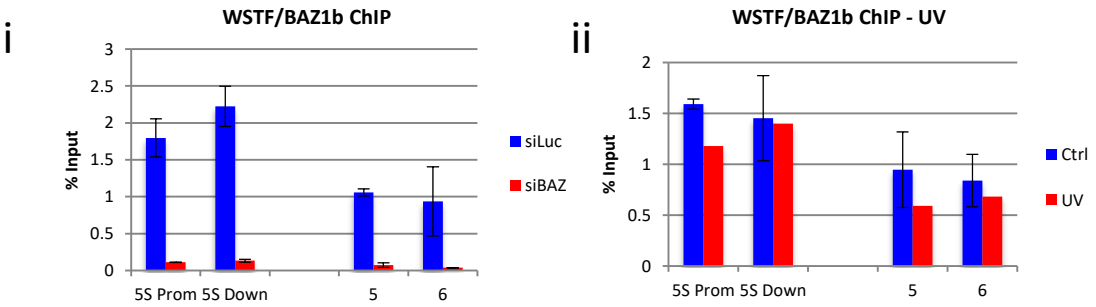


Figure 6.11 BAZ1b Chromatin immunoprecipitation experiments. i) BAZ1b ChIP establishment using primers previously shown to detect high levels of BAZ1b in rRNA (5S Prom/Down from Sadeghifar et al., 2015), and in the 3' end of *YWHAZ*. siBAZ1b knockdown compared to siLuc control. ii) BAZ1b ChIP investigating effects of UV damage. 2 hour incubation +/- 20 Jm^{-2}. Error bars from two repeats.

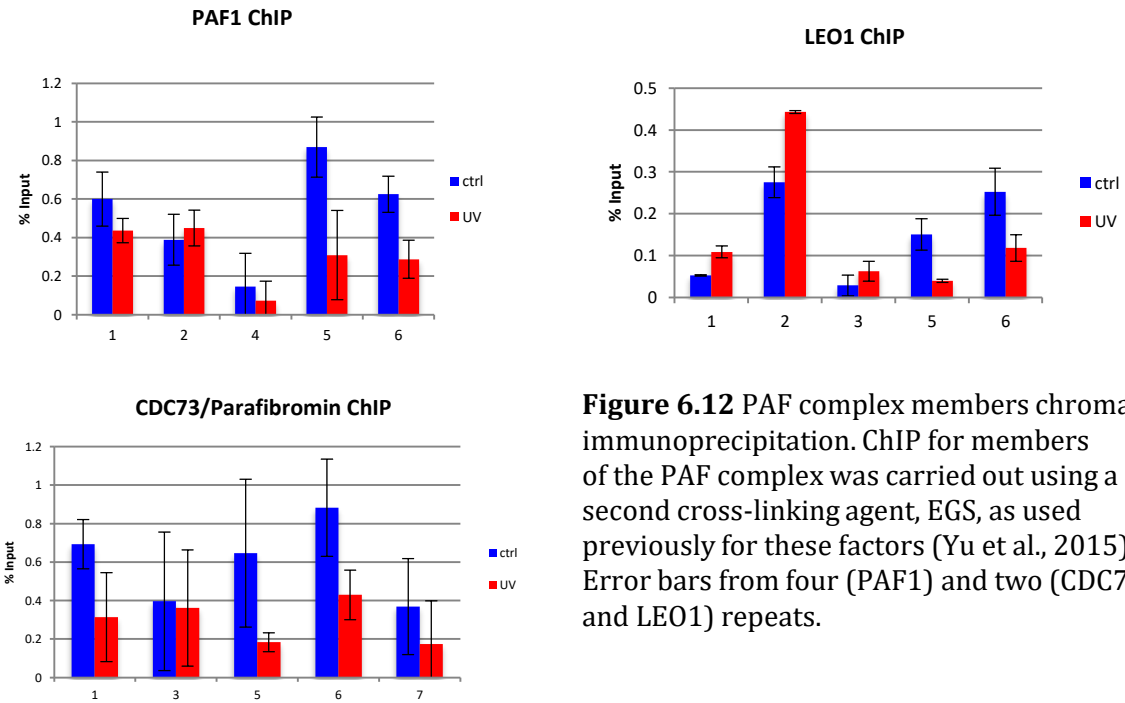


Figure 6.12 PAF complex members chromatin immunoprecipitation. ChIP for members of the PAF complex was carried out using a second cross-linking agent, EGS, as used previously for these factors (Yu et al., 2015). Error bars from four (PAF1) and two (CDC73 and LEO1) repeats.

6.10 UV irradiation causes a loss of nucleosomes and specific histone marks

Next I employed the NetMNase-seq technique after UV treatment to test whether UV irradiation causes alterations in nucleosome occupancy or post-translational modification (figure 6.13-18). Comparing Input mononucleosome fractions, depletion in nucleosome signal across genes is evident post UV (figure 6.16). This is particularly marked at the 5' end. This depletion might suggest increased nucleosome recycling rates over genes upon UV irradiation. A similar effect has been shown before for yeast (Hauer et al. 2017).

An immunoprecipitation stage was also used in this experiment to specifically identify DNA associated with particular histone methylation marks, suggestive of active transcription and repressive chromatin to test whether UV changes these profiles as well (figure 6.13). Three marks were chosen. H3K36me3 is generally associated with active RNAPII elongation. H3K4me3 is associated primarily with promoter regions of actively transcribed genes. H3K9me3 is believed to be a silencing mark and is found to accumulate over heterochromatin, constitutively silenced genes and ncRNA transcription units. It was hypothesised these three methyl marks would provide a range of profiles to investigate the effects of UV irradiation on nucleosome profiles.

Signal for the H3K36me3 mark accumulated in non-irradiated cells over gene bodies, H3K4me3 either side of gene promoters, and H3K9me3 over ncRNA and heterochromatin (figure 6.14). These correspond well to equivalent ChIP-seq

data for the same marks (H3K4me3 comparison shown in figure 6.15). This indicates the technique is working correctly.

Metagene profiles for active mRNA genes show UV irradiation causes a depletion in H3K36me3 across the gene body, and depletion in H3K4me3 at promoters (figure 6.17). This suggests a loss of histone marks promoting or being "written" downstream of active transcription upon UV damage. H3K9me3 signal is similar across mRNA genes for the two conditions, perhaps showing a minimal reduction in signal at the TSS and in the gene body. The prevalence of this mark is low over mRNA however. These effects are in addition to the general loss in nucleosome signal seen in figure 6.16, as IP efficiency is calculated relative to the Input nucleosome signal. This means the proportional loss of these marks is greater than the general nucleosome depletion.

A similar reduction in H3K36me3 and H3K4me3 marks is seen over eRNA transcription units (figure 6.18). Both H3K36me3 and H3K4me3 IPs show a distinct loss of signal over and to either side of eRNA transcription start sites. H3K9me3 signal appears to be much less effected. This may indicate that UV irradiation causes a general reduction in histone marks favouring active transcription and RNA processing, assisting in a genome-wide reduction in expression level.

6.11 BAZ1b and SNF2h depletion causes nucleosome remodelling

The NetMNase-seq technique was also used to investigate the effects of knockdown of some of the chromatin remodelling factors identified in the original Y1P mass spectrometry data. Input mononucleosome reads for siBAZ1b show a reduction in signal over the body of mRNA genes, compared to luciferase control. A smaller reduction in signal is also evident upon SNF2h depletion (figure 6.19). This signal reduction is greatest close to the TSS, in the regularly spaced nucleosome region (figure 6.20). A loss of signal from the +1, +2, and +3 nucleosome positions is evident.

Interestingly, siSNF2h treatment also produces an increase in nucleosome signal upstream of mRNA TSSs. This is not seen in BAZ1b siRNA treated cells. This SNF2h specific effect is also seen when only focusing on PROMPTs (data not shown). BAZ1b and SNF2h knockdowns (figure 6.22) appear to have different effects around eRNAs as well (figure 6.21). Both knockdowns have little effect on the signal of the first two regularly spaced nucleosomes either side of eRNA transcription start sites, but then differ in their effects further from the TSS. BAZ1b depletion causes a loss of nucleosome signal compared to siLuc control. SNF2h knockdown on the other hand causes an increase in mononucleosome signal either side of eRNA promoters.

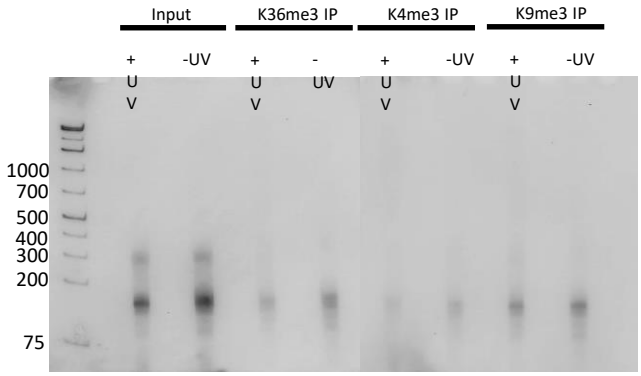


Figure 6.13 Gel visualisation of DNA isolated from histone immunoprecipitation under ultraviolet damage conditions. Following UV irradiation and NetMNase-seq chromatin digestion, histone-DNA complexes were isolated by precipitation with antibodies specific to different methylation marks – H3K36me3, H3K4me3 and H3K9me3. Mononucleosomes fractions were then purified.

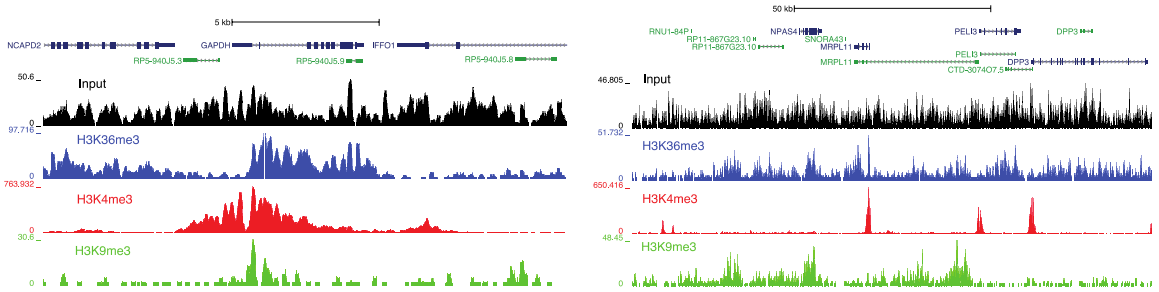


Figure 6.14 Visualisation of read coverage for DNA associated with different histone marks. *GAPDH* as an example gene and a wider view (right).

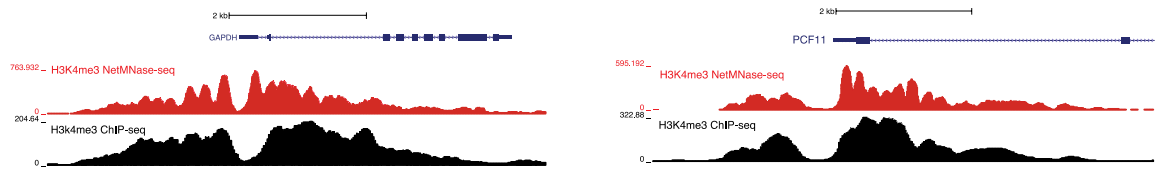


Figure 6.15 Comparison of NetMNase-seq with ChIP-seq for the H3K4me3 mark. *GAPDH* and *PCF11* as example genes.

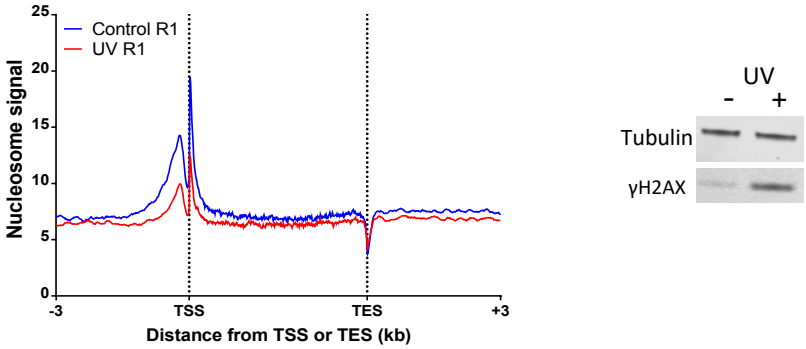


Figure 6.16 Metagene analysis of Input mononucleosome signal +/- UV damage. One repeat of three shown. Metagene profiles by Michael Tellier.

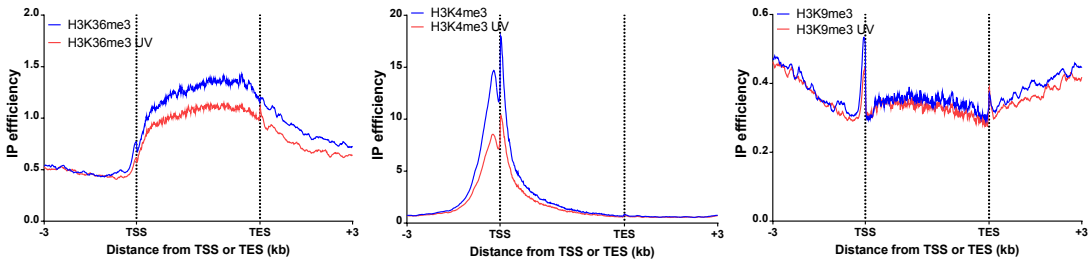


Figure 6.17 Metagene analysis of histone methylation state over mRNA genes. H3K36me3, H3K4me3 and H3K9me3 IPs relative to Input.

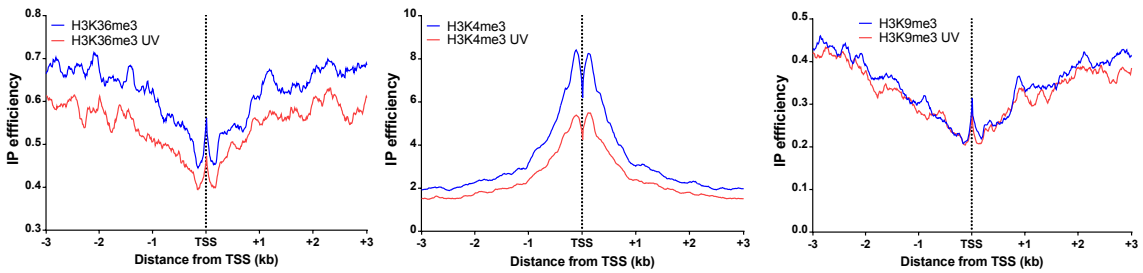


Figure 6.18 Metagene analysis of histone methylation around eRNAs.

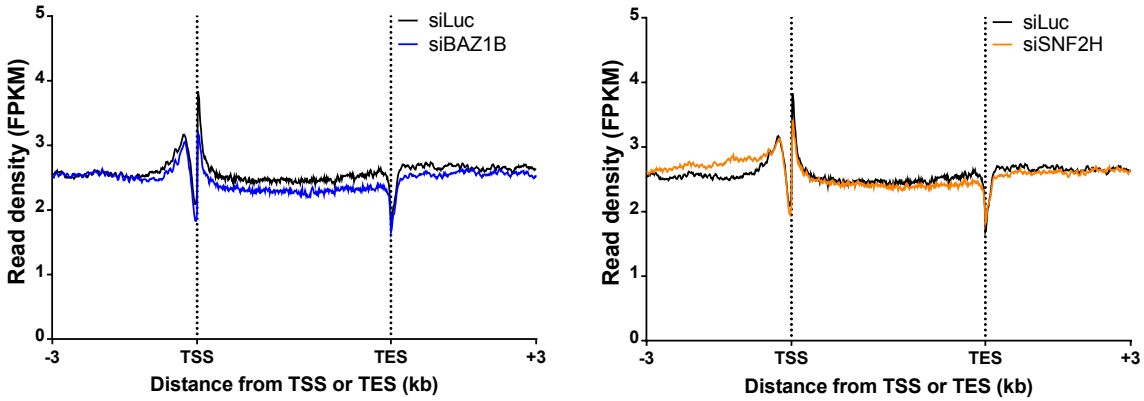


Figure 6.19 Metagene analysis of Input histone occupancy across mRNA transcription units upon BAZ1b and SNF2h depletion.

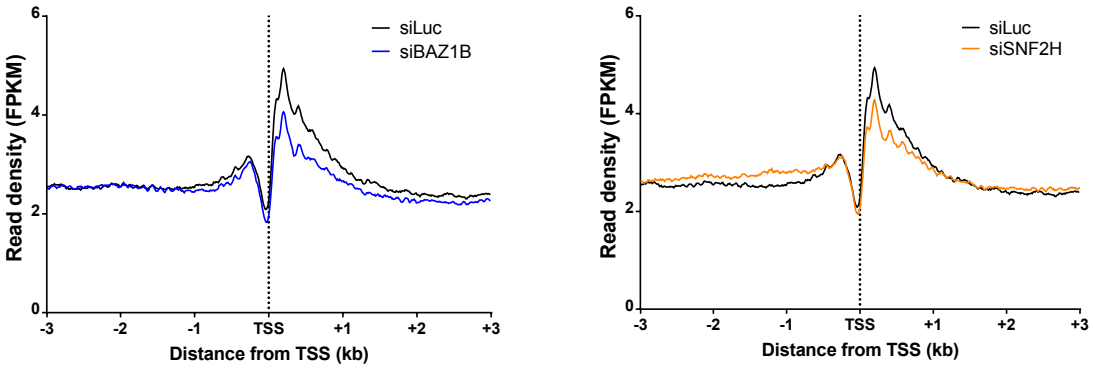


Figure 6.20 Magnification of mRNA TSS regions from figure 6.19.

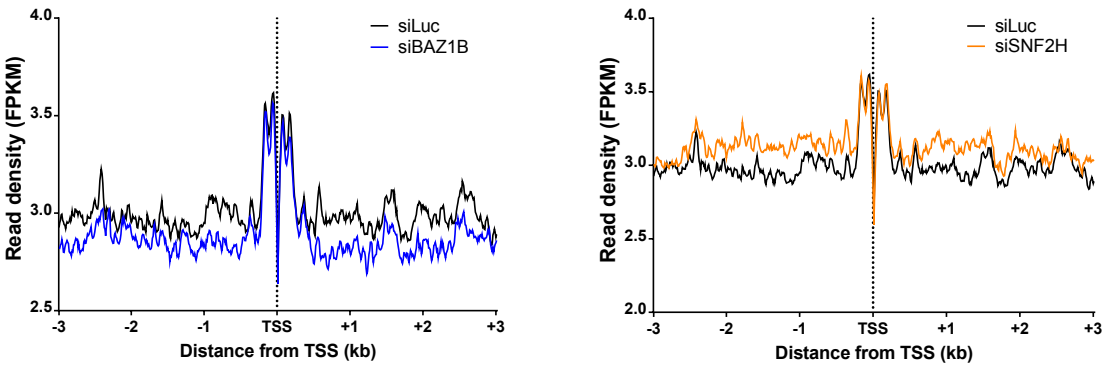


Figure 6.21 eRNA gene mononucleosome signal is altered in opposing directions by knockdown of BAZ1b and SNF2h.

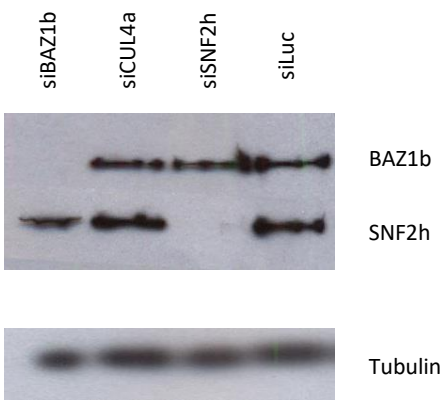


Figure 6.22 Western blot confirmation of BAZ1b and SNF2h siRNA knockdown. Confirmation for samples used in NetMNase-seq.
Approximate molecular weights (kDa): BAZ1b – 170, SNF2h – 120,
Tubulin 1-alpha – 50.

6.12 Discussion

Mass spectrometry results have suggested a link between tyrosine 1 phosphorylation and DNA repair, with a number of NER pathway factors specifically interacting with the Y1P moiety.

However a direct link between these factors associating with Y1P, the transcription cycle and DNA damage still remains elusive. Western blot analysis suggests an increase in hyper-phosphorylation of CTDs with at least some Y1 phosphorylation. However, immunofluorescence suggests no direct cause and effect between damage and tyrosine phosphorylation, at least at major damage foci. Perhaps both marks only correlate with promoters rather than associate at novel damage foci. Novel γH2AX foci may predominantly represent DSBs and not sites of NER.

The possibility that BAZ1b could potentially phosphorylate CTD tyrosine residues was thought worthy of investigation. So far, only c-Abl has been established as a kinase for this residue, and only *in vitro.* Like other CTD residues, it is quite likely there are a number of kinases for tyrosine-1, and BAZ1b could have provided a link between DNA damage and the phosphorylation of this residue. However, this potential degeneracy among kinases makes showing a loss of phospho-signal *in vivo* difficult. BAZ1b should not be fully ruled out as a CTD kinase until *in vitro* kinase assays are performed. Although the phospho-sites of H2AX and CTD do not share many sequence

similarities, they are both found in relatively unstructured C-terminal tails brought into close proximity to BAZ1b on chromatin.

However these investigations did lead to interesting results investigating UV irradiation. ChrRNA-seq results following irradiation corroborate what has previously been shown for the effect of UV irradiation on transcription – an accumulation of RNAPII and nascent RNA at the 5' end of genes, and an elongation defect. This is generally a small effect at the 2-hour post-UV time window, with little difference between conditions for most genes, suggesting transcription is mostly recovered for short genes in this time frame. More analysis of extremely long genes might show a greater effect. As identified above, there is also a more pronounced elongation defect evident for some shorter genes, including transcription factors such as *TAF9*, *CSTF3* and *KIAA0824*. UV treatment may favour production of a short isoform of these genes, although this would have to be confirmed with techniques such as 3' RACE. Perhaps UV irradiation triggers an alteration in transcription dynamics through alternative expression of these factors.

mNET-seq results appear to differ slightly from previously published studies. Little loss in signal is seen over the gene body following UV irradiation, with the major effect seen in both Total and Y1P RNAPII associated reads being a loss 3' of the pA signal. However, a similar small reduction in RNAPII peak at the TSS is seen using both mNET-seq and previous techniques. Differences could be due to a range of reasons. Previous reports have relied on chromatin RNA or GRO-seq run-on techniques (Rockx et al. 2000; Andrade-Lima et al. 2015; Williamson et al.

2017). mNET-seq differs from run-ons in that it does not rely on following transcription post the isolation of nuclei and introduction of foreign nucleotides to detect nascent RNA. UV treatments before or after administration of the elongation inhibitor DRB show different effects in GRO-seq, and the use of inhibitors and Br-UTP treatment may make the system more artificial. Previous analysis on steady state RNA levels instead of DRB treated cells has generally focused on only very long genes. mNET-seq metagene analysis on all expressed mRNA genes rather than just long mRNA may therefore mask an effect on gene body signal.

Other techniques also have no connection to different polymerase isoforms (CTD modifications). (Similarly, previous MS analyses have precipitated FLAG-tagged polymerase rather than pulling down individual CTD modifications (Boeing et al. 2016).) It is possible other modifications (S2P) would show an mNET-seq signal decrease over the gene body. In future mNET-seq experiments, spiking of RNA samples should also be used to better standardise signal levels.

Studies of chromatin conformations have not been included in published multiomic approaches to studying UV transcription effects (Boeing et al. 2016). NetMNase-seq may therefore provide a way of linking methods studying nascent RNA like mNET-seq, and proteomic analysis of chromatin, to mononucleosome mapping and positioning of histone modifications upon irradiation.

The most prominent effect of BAZ1b and SNF2h knockdowns on nucleosome architecture is seen around promoters. This is perhaps not surprising as SNF2h

has a known role in regularly spacing nucleosomes at this location, and the WICH complex has been found to accumulate at RNAPI and III promoters. Loss of nucleosome signal over the gene body could suggest a role in the recycling of histone proteins around RNAPII.

SNF2h has a markedly different effect on nucleosome signal upstream of promoters and in eRNA transcription units. This suggests an SNF2h specific function separate from its BAZ1b partner. Increased nucleosome density at PROMPTs and eRNAs on knockdown indicates a role in maintaining open chromatin on ncRNA genes. However, as noted in chapter 5, the meaning of an increase in Input mononucleosome read signal is still somewhat ambiguous.

BAZ1b depletion shows relatively small loss of mononucleosome signal over gene bodies and around eRNAs. Western blot analysis indicated a major increase in HP1α (a heterochromatic marker) as a result of BAZ1b depletion. This would indicate the increase in HP1α does not take place over transcription units, and it is more likely the HP1α increase occurs in regions of constitutive heterochromatin, areas of which the WICH complex is known to assist in maintaining (Culver-Cochran and Chadwick 2013) or around areas of newly replicated DNA (Poot et al. 2004). HP1α is also found to bind at sites of DNA damage (Bartova et al. 2017). Its expression may therefore be upregulated following loss of a nucleosome remodelling factor involved in DNA repair. Alternatively, HP1α remodelling of chromatin may not alter accessibility to MNase, and so not significantly alter MNase-seq profiles.

There is a discrepancy between the MS increase in PAF1C association with polymerase on UV irradiation, and the apparent drop in signal associated with the *YWHAZ* 3' end. The failure of ChIP-seq is regrettable as it may have shown in what genomic location the increase in PAF1C accumulation on chromatin occurred. However, if UV treatment inhibits PAF accumulation at gene 3' ends, it may no longer promote RNAPII elongation over these regions, leading to the observed elongation defect. However, it is also possible that these factors travel with the polymerase along the gene, and therefore fail to accumulate at the 3' end if UV damage prevents elongation through other means.

Current work in our laboratory shows that depletion of SPT6 causes premature termination, indicated by a decrease in chromatin RNA reads towards the 3' end of genes, and an increase in the potential termination mark T4P within gene bodies. More recently, the NetMNase-seq technique has been used to pull down DNA associated with the H3K36me3 mark, a marker for active transcription, and here a loss over the body of the gene is seen in SPT6 KD cells. SPT6 knockdown also produces a decrease in PAF1 signal across gene bodies and 3' ends (Takayuki Nojima, personal communication).

So a potential model to explain these results could be as follows – under normal circumstances, SPT6 and the PAF complex promote polymerase elongation. SPT6 promotes active transcription and H3K36 methylation. However, under UV irradiation, SPT6 is degraded, leading to repressed elongation and loss of the methyl mark.

It should be noted that the mass spectrometry methods employed are only semi-quantitative, and in order to make more certain judgments on the presence/absence or fold increase in the level of a certain protein between samples, more quantitative measures should be employed. These include SILAC (Stable isotope labeling of amino acids in cell culture), which would allow for a more definite comparison of two conditions grown in heavy and light media. This approach was however beyond the scope of the current report. The proteasome inhibitor MG132 could also be used to prevent loss of RNAPII and other factors upon UV irradiation, preventing loss of a proportion of stalled polymerase and associated factors.

Most experiments were only carried out comparing 2 hours incubation after UV irradiation and untreated cells. Time courses for a number of experiments would provide information at shorter time points (immediate response to DNA damage) and recovery stage (up to 24 hr). This could give information on the timing of elongation and shutdown of transcription (mNET- and ChrRNA- seq), the sequence of protein factor recruitment/association with RNAPII and chromatin (MS and ChIP) and nucleosome remodelling (NetMNase-seq). This would obviously require a very significant time investment. It would also be interesting to look at other CTD modifications with mNET-seq. For example, T4P signal. UV damage is believed to mostly trigger an elongation defect, but testing for increased T4P signal over the gene body could reveal whether any premature termination also occurs.

Ongoing work in our laboratory is investigating the effect of R-loops on *de novo* transcription initiation. Since Y1P peaks at the TSS, is associated with a number of RNA:DNA helicases, and R-loops are known to be sources of DNA damage, it would be interesting to test for a correlation between R-loops and Y1P, and whether RNase-H treatment alters Y1P mNET-seq profiles.

Given the significantly different changes to nucleosome profiles upon siRNA knockdown of different chromatin remodeling factors and termination factors (CPSF73, chapter 5) it would be interesting to use NetMNase-seq (both Input and IP) to test the effects of knockdown of factors like CUL4A and the PAF complex, since these have also been linked to chromatin remodeling and histone methylation. Likewise, NetMNase-seq IP of different histone marks following BAZ1b and SNF2h knockdown might help to dissect what effects these factors are causing. Since preliminary data suggest BAZ1b depletion may have an effect on alternative splicing only after UV damage (Gwendal Dujardin, personal communication), it may also be informative to investigate knockdown effects on nucleosome architecture in conjunction with UV treatment. It is also important to produce RNA-seq data for these factor knockdowns, to correlate chromatin remodeling to gene expression.

Chapter 7

General Discussion

7. General Discussion

7.1 Summary

This thesis has addressed a number of separate issues that each focus on the transcription cycle. A connecting theme is that chromatin remodelling is used to reinforce and facilitate different mechanisms, including transcription termination and DNA repair.

Pause-type termination has been shown to involve a synergistic relationship between DNA sequence, R-loops and heterochromatin (Skourti-Stathaki et al. 2014). G-rich DNA sequences with G-clusters in the non-coding strand facilitate R-loop formation, and RNAPII pausing. This leads to localized antisense RNA production and recruitment of DICER and Argonaute proteins, as well as histone methylases, causing H3K9 dimethylation. However, the cause and effect relationship of some of these events is still as yet unclear. Perhaps a pioneer round of transcription and R-loop induced pausing causes the introduction of H3K9me2 marks over the pause region. This then further slows polymerase in subsequent rounds of transcription to promote termination soon after the pA site.

The findings studying *PKM*, that G-rich regions with corresponding R-loops and antisense RNA production can exist without associated downstream heterochromatic marks, suggests that chromatin condensation is either not necessary at all genes, or is established through different mechanisms. Perhaps only very highly expressed genes like *β-actin* require both R-loops and

heterochromatin to slow polymerase. However, although H3K9 dimethylation was not found to increase in the pause region of *PKM*, HP1γ levels did peak over the R-loop regions, suggesting some mechanism of heterochromatin formation still occurs.

Rad51 is best known as a homologous recombination factor that polymerizes on ssDNA and facilitates strand invasion into homologous dsDNA during DNA repair (Godin et al. 2016). Previous work in yeast suggested that Rad51 can facilitate RNA invasion of double stranded DNA to produce R-loops (Wahba et al. 2013). While Rad51 was indeed found to be present on DNA in the vicinity of R-loops in pause-type regions, and not at the 3' end of CoTC-type genes, no effect on R-loops was seen following Rad51 inhibition or depletion. This would suggest that Rad51 in these regions is simply acting in the canonical DNA-repair pathway, mechanistically downstream of lesion-inducing R-loops. Another DNA repair factor, BRCA1 has recently been shown to function in just such a pathway (Hatchi et al. 2015).

In view of this lack of Rad51 effect on R-loop formation, further research in this area seemed non-productive. I therefore focused my efforts on other projects. Consequently my research shifted towards studying the roles of two CTD modifications, as suggested by mNET-seq and mass spectrometry data, and the development of the NetMNase-seq technique to aid these studies.

NetMNase-seq was set up to allow close comparison of mNET-seq nascent RNA reads, showing genome-wide RNAPII transcription activity, and nucleosome

architecture. Both techniques use the same, relatively gentle isolation conditions to extract nucleic acid protected from digestion. Comparison of data from the two techniques shows a close relationship between RNAPII promoter-proximal pausing and regularly spaced nucleosomes at the transcription start site. NetMNase-seq itself can also determine RNAPII positional information relative to nucleosomes, and the position of histone modifications, through immunoprecipitation. In some cases this can be done with higher resolution than with histone ChIP-seq. NetMNase-seq can also resolve changes in DNA protection, probably equating to nucleosome prevalence, at different positions on gene bodies following perturbations such as protein factor depletion or UV damage.

mNET-seq data has shown that RNAPII CTD threonine-4 phosphorylation is associated with mRNA 3' ends and may act as a mark of termination. Whether T4P recruits termination factors or T4 is phosphorylated after the release of pre-mRNA or RNA processing complexes is still as yet unclear. CPSF73 depletion shows that loss of the endonuclease causing cleavage at the pA site causes a termination defect, but that that defect is small, and that termination occurs at a relatively small distance downstream. A corresponding shift in T4P signal is seen. This differs from the massive disruption to termination caused by some virus infections, such as herpes simplex and influenza (Nemeroff et al. 1998; Rutkowski et al. 2015). Deletion of the DNA region associated with the secondary termination (as defined by T4P signal peaks) indicates that this CPSF73-independent, potential "failsafe" termination event is not governed by sequence elements but probably occurs at a certain position downstream of the wild type

termination region. CPSF73 depletion has also been shown to increase the mononucleosome signal specifically downstream of pA sites. This suggests that nucleosome remodelling may be partially responsible for preventing aberrant transcriptional run-on from causing massive read-through on loss of cleavage. CPSF73 depletion does not alter the position of TES NDRs, suggesting these are maintained by other factors, for example DNA sequence and other termination factor recruitment – perhaps triggering RNAPII pausing or directly remodelling chromatin. CPSF73 depletion also causes a small but significant loss of nucleosome signal at the TSS, suggesting feedback from transcription termination to regulation of the spacing of nucleosomes next to the depleted region around the promoter.

mNET-seq analysis of the RNAPII CTD Tyrosine-1 phosphorylation signal suggests this modification is mostly present at gene TSS and TES regions. Another modification to the mNET-seq protocol to allow processing by mass spectrometry has indicated that Y1P is associated with DNA repair and nucleosome remodelling factors. Specifically factors involved in nucleotide excision repair, suggesting a link between transcription regulation and UV damage repair. Further mass spectrometry experiments upon UV damage also indicated that association of members of the PAF complex with the CTD is increased upon UV damage, whereas SPT6 association decreases. PAF complex distribution across genes also appears to be altered upon UV irradiation, with a loss of ChIP signal at the 3' end of the *YWHAZ* gene. This region was also shown to specifically lose Y1P signal in mNET-seq analysis following exposure to UV. H3K36me3 and H3K4me3 marks are also specifically lost over the gene body

and promoter respectively on UV damage. Altogether, my data suggest that UV irradiation alters recruitment of SPT6 and PAF1C, with consequent changes to histone recycling and methylation, and a global RNAPII elongation defect.

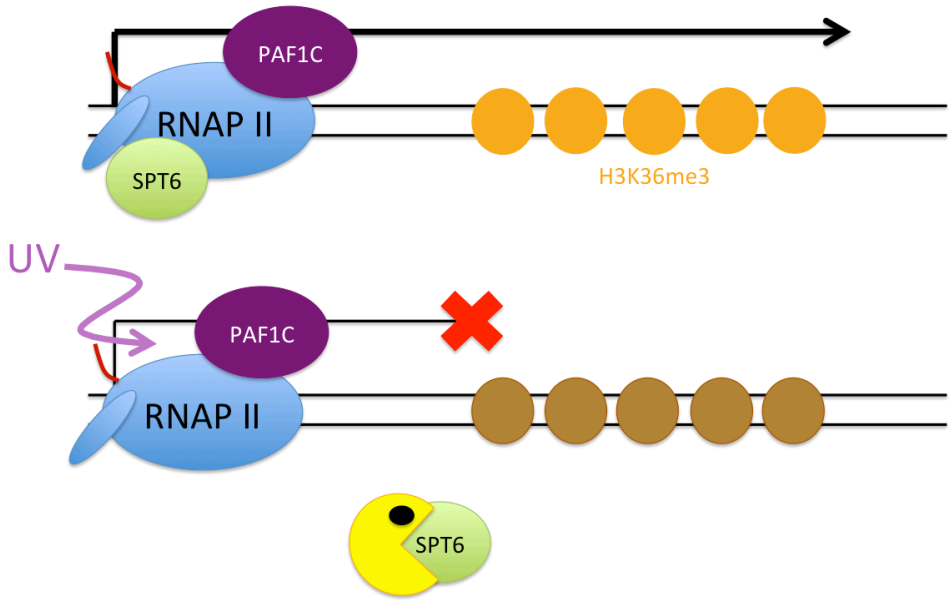


Figure 7.1 A potential model for the effect of UV irradiation on transcription elongation. Under normal conditions, SPT6 and PAF1C bind to RNAPII and promote transcriptional elongation, and the introduction of histone marks that correlate with active elongation, such as H3K36me3. However, upon UV irradiation, SPT6 is degraded by the proteasome, leading to a loss of these histone marks across the gene body, and an elongation defect. PAF1C may still bind to RNAPII and be lost at the 3' end due to the RNAPII elongation defect, or it may be actively displaced in these regions.

Despite the intriguing mass spectrometry link between Y1P and NER factors, the exact nature of the connection between DNA damage and Y1P still remains elusive.

7.2 Future Experiments

NetMNase-seq could easily be further modified and used to investigate both topics touched upon in this thesis, and other interests of our group. For example, the effects of R-loops on both heterochromatin formation in pause regions, and transcription initiation, could be investigated by RNase HI overexpression followed by NetMNase-seq. This could test for both changes in mononucleosome read density and localization of specific histone marks.

As mentioned in chapters 4 and 5, further refinements to NetMNase-seq are both necessary and potentially informative. A time-course of MNase digestion could be used to better understand the nature of sequences extracted from mononucleosome gel bands and the extent of compaction of regions associated with different histone marks.

It would be interesting to attempt pull down with antibodies for DNA binding factors other than RNAPII and histones. Similarly, IP with antibodies specific for ubiquitin may give information on the level of RNAPII or histone ubiquitinylation and removal around DNA damage sites. More comparison of mono- and di-nucleosome fractions could further probe the compaction state of histone modifications in different areas of the gene. Extending immunoprecipitation steps to include more histone marks can always be added to all the NetMNase-seq experiments. For example, the prevalence of H3K4me1 marks at enhancers, H3K9me2 marks in pause regions and H3K27 acetylation as a mark of active transcription at promoters. γH2AX and HP1α/γ mapping could also give

information on sites sensitive to DNA damage and regions of heterochromatin, respectively.

Deletion of other gene "failsafe" regions would be useful to compare termination mechanisms and the role of T4P. Multiple deletions in one strain could even justify genome-wide sequencing approaches to analyse termination behaviour. However, given the results for *GAPDH* indicate a simple positioning mechanism for failsafe termination, only a few other gene examples may be necessary to confirm this.

The nucleosome-depleted region at protein-coding gene TESs appears to be maintained in the same location even in absence of cleavage by CPSF73. This may be due to other factors in the CPA complex recognising the region of cleavage, particularly CPSF30 and WDR33 still being present and determining binding to the AAUAAA RNA consensus sequence. It would be interesting to determine whether this NDR is still maintained in the case of influenza infection that disrupts the CPSF30 interaction (Kuo and Krug 2009; Chan et al. 2014). Depletion of these other CPA factors may also help determine this. Bioinformatic analysis of the area around the TES should also be extended to see if any regular nucleosomes arrangement can be detected.

Mass spectrometry RNAPII precipitations could be performed with less stringent isolation and wash buffers to attempt co-immunoprecipitation of cleavage factors and to test which isoforms they are associated with. It may be difficult

however to isolate only chromatin bound, nascent RNA-transcribing RNAPII under such conditions.

Generating ChIP-seq libraries for PAF1C members upon UV damage is necessary to confirm the *YWHAZ* ChIP results, to test how prevalent the 3' loss is genome-wide and to understand where the increase of PAF1C associated with RNAPII CTD is occurring. Similar tests would be useful to understand the significant increase in HP1α upon BAZ1b depletion.

The effects of other forms of DNA damage on Y1P genomic distribution could also be investigated. A number of techniques could be employed to generate double strand breaks at multiple, known locations. The AsiSI enzyme targets cleavage sites on average >1 Mbp (although not all are cut due to inaccessibility) (Iacovoni et al. 2010). gRNAs used to direct Cas9 (variants of which can cut either one or both strands at a target site) could fulfill a similar role. These could be used to resolve what CTD isoforms are recruited to lesions and what factors are recruited at different time points. Confocal immunofluorescence of Y1P with factors like BAZ1b and CUL4A may also be useful to answer these questions, and generate a better picture of the roles of Y1P and its associated factors in transcriptional regulation, DNA damage and chromatin arrangement.

Overall my thesis has opened up a number of different experimental avenues that I hope will be addressed beyond the research described herein.

References

Adelman K, Lis JT. 2012. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature reviews Genetics* **13**: 720-731.
Aguilera A, Garcia-Muse T. 2012. R loops: from transcription byproducts to threats to genome stability. *Mol Cell* **46**: 115-124.
Allo M, Buggiano V, Fededa JP, Petrillo E, Schor I, de la Mata M, Agirre E, Plass M, Eyras E, Elela SA et al. 2009. Control of alternative splicing through siRNA-mediated transcriptional gene silencing. *Nature structural & molecular biology* **16**: 717-724.
Ambros V. 2004. The functions of animal microRNAs. *Nature* **431**: 350-355.
Andersen PK, Jensen TH, Lykke-Andersen S. 2013. Making ends meet: coordination between RNA 3'-end processing and transcription initiation. *Wiley interdisciplinary reviews RNA* **4**: 233-246.
Andrade-Lima LC, Veloso A, Paulsen MT, Menck CF, Ljungman M. 2015. DNA repair and recovery of RNA synthesis following exposure to ultraviolet light are delayed in long genes. *Nucleic acids research* **43**: 2744-2756.
Aymard F, Bugler B, Schmidt CK, Guillou E, Caron P, Briois S, Iacovoni JS, Daburon V, Miller KM, Jackson SP et al. 2014. Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. *Nature structural & molecular biology* **21**: 366-374.
Baaklini I, Hraiky C, Rallu F, Tse-Dinh YC, Drolet M. 2004. RNase HI overproduction is required for efficient full-length RNA synthesis in the absence of topoisomerase I in Escherichia coli. *Molecular microbiology* **54**: 198-211.
Baldacci G, Cherif-Zahar B, Bernardi G. 1984. The initiation of DNA replication in the mitochondrial genome of yeast. *The EMBO journal* **3**: 2115-2120.
Barnett C, Krebs JE. 2011. WSTF does it all: a multifunctional protein in transcription, repair, and replication. *Biochemistry and cell biology = Biochimie et biologie cellulaire* **89**: 12-23.
Bartova E, Malyskova B, Komurkova D, Legartova S, Suchankova J, Krejci J, Kozubek S. 2017. Function of heterochromatin protein 1 during DNA repair. *Protoplasma* **254**: 1233-1240.
Baskaran R, Chiang GG, Wang JY. 1996. Identification of a binding site in c-Ab1 tyrosine kinase for the C-terminal repeated domain of RNA polymerase II. *Molecular and cellular biology* **16**: 3361-3369.
Baskaran R, Dahmus ME, Wang JY. 1993. Tyrosine phosphorylation of mammalian RNA polymerase II carboxyl-terminal domain. *Proceedings of the National Academy of Sciences of the United States of America* **90**: 11167-11171.
Baskaran R, Escobar SR, Wang JY. 1999. Nuclear c-Abl is a COOH-terminal repeated domain (CTD)-tyrosine (CTD)-tyrosine kinase-specific for the mammalian RNA polymerase II: possible role in transcription elongation. *Cell growth & differentiation : the molecular biology journal of the American Association for Cancer Research* **10**: 387-396.
Belotserkovskii BP, Liu R, Tornaletti S, Krasilnikova MM, Mirkin SM, Hanawalt PC. 2010. Mechanisms and implications of transcription blockage

by guanine-rich DNA sequences. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 12816-12821.
Bhatia V, Herrera-Moyano E, Aguilera A, Gomez-Gonzalez B. 2017. The Role of Replication-Associated Repair Factors on R-Loops. *Genes* **8**.
Boeing S, Williamson L, Encheva V, Gori I, Saunders RE, Instrell R, Aygun O, Rodriguez-Martinez M, Weems JC, Kelly GP et al. 2016. Multiomic Analysis of the UV-Induced DNA Damage Response. *Cell reports*.
Boguslawski SJ, Smith DE, Michalak MA, Mickelson KE, Yehle CO, Patterson WL, Carrico RJ. 1986. Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. *Journal of immunological methods* **89**: 123-130.
Bondarenko VA, Steele LM, Ujvari A, Gaykalova DA, Kulaeva OI, Polikanov YS, Luse DS, Studitsky VM. 2006. Nucleosomes can form a polar barrier to transcript elongation by RNA polymerase II. *Mol Cell* **24**: 469-479.
Bozhenok L, Wade PA, Varga-Weisz P. 2002. WSTF-ISWI chromatin remodeling complex targets heterochromatic replication foci. *The EMBO journal* **21**: 2231-2241.
Brown JA, Eykelenboom JK, Lowndes NF. 2012. Co-mutation of histone H2AX S139A with Y142A rescues Y142A-induced ionising radiation sensitivity. *FEBS open bio* **2**: 313-317.
Bunch H, Lawney BP, Lin YF, Asaithamby A, Murshid A, Wang YE, Chen BP, Calderwood SK. 2015. Transcriptional elongation requires DNA break-induced signalling. *Nature communications* **6**: 10191.
Buratowski S. 2009. Progression through the RNA polymerase II CTD cycle. *Mol Cell* **36**: 541-546.
Burugula BB, Jeronimo C, Pathak R, Jones JW, Robert F, Govind CK. 2014. Histone deacetylases and phosphorylated polymerase II C-terminal domain recruit Spt6 for cotranscriptional histone reassembly. *Molecular and cellular biology* **34**: 4115-4129.
Calo E, Flynn RA, Martin L, Spitale RC, Chang HY, Wysocka J. 2015. RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. *Nature* **518**: 249-253.
Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, Shia WJ, Anderson S, Yates J, Washburn MP et al. 2005. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123**: 581-592.
Cavellan E, Asp P, Percipalle P, Farrants AK. 2006. The WSTF-SNF2h chromatin remodeling complex interacts with several nuclear proteins in transcription. *The Journal of biological chemistry* **281**: 16264-16271.
Cerritelli SM, Crouch RJ. 2009. Ribonuclease H: the enzymes in eukaryotes. *The FEBS journal* **276**: 1494-1505.
Cerritelli SM, Frolova EG, Feng C, Grinberg A, Love PE, Crouch RJ. 2003. Failure to produce mitochondrial DNA results in embryonic lethality in Rnaseh1 null mice. *Mol Cell* **11**: 807-815.

Chan S, Choi EA, Shi Y. 2011. Pre-mRNA 3'-end processing complex assembly and function. *Wiley interdisciplinary reviews RNA* **2**: 321-335.
Chan SL, Huppertz I, Yao C, Weng L, Moresco JJ, Yates JR, 3rd, Ule J, Manley JL, Shi Y. 2014. CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing. *Genes & development* **28**: 2370-2380.
Chaudhuri J, Alt FW. 2004. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nature reviews Immunology* **4**: 541-552.
Chen FX, Woodfin AR, Gardini A, Rickels RA, Marshall SA, Smith ER, Shiekhattar R, Shilatifard A. 2015. PAF1, a Molecular Regulator of Promoter-Proximal Pausing by RNA Polymerase II. *Cell* **162**: 1003-1015.
Chen FX, Xie P, Collings CK, Cao K, Aoi Y, Marshall SA, Rendleman EJ, Ugarenko M, Ozark PA, Zhang A et al. 2017. PAF1 regulation of promoter-proximal pause release via enhancer activation. *Science*.
Chen Y, Pai AA, Herudek J, Lubas M, Meola N, Jarvelin AI, Andersson R, Pelechano V, Steinmetz LM, Jensen TH et al. 2016. Principles for RNA metabolism and alternative transcription initiation within closely spaced promoters. *Nature genetics* **48**: 984-994.
Chernikova SB, Brown JM. 2012. R-loops and genomic instability in Bre1 (RNF20/40)-deficient cells. *Cell cycle (Georgetown, Tex)* **11**: 2980-2984.
Cho EJ, Takagi T, Moore CR, Buratowski S. 1997. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes & development* **11**: 3319-3326.
Chu G, Chang E. 1988. Xeroderma pigmentosum group E cells lack a nuclear factor that binds to damaged DNA. *Science* **242**: 564-567.
Churchman LS, Weissman JS. 2012. Native elongating transcript sequencing (NET-seq). *Current protocols in molecular biology* **Chapter 4**: Unit 4.14.11-17.
Colgan DF, Manley JL. 1997. Mechanism and regulation of mRNA polyadenylation. *Genes & development* **11**: 2755-2766.
Compe E, Egly JM. 2012. TFIIH: when transcription met DNA repair. *Nature reviews Molecular cell biology* **13**: 343-354.
Cook PJ, Ju BG, Telese F, Wang X, Glass CK, Rosenfeld MG. 2009. Tyrosine dephosphorylation of H2AX modulates apoptosis and survival decisions. *Nature* **458**: 591-596.
Corden JL. 1990. Tails of RNA polymerase II. *Trends in biochemical sciences* **15**: 383-387.
-. 2013. RNA polymerase II C-terminal domain: Tethering transcription to transcript and template. *Chemical reviews* **113**: 8423-8455.
Costes SV, Daelemans D, Cho EH, Dobbin Z, Pavlakis G, Lockett S. 2004. Automatic and quantitative measurement of protein-protein colocalization in live cells. *Biophysical journal* **86**: 3993-4003.
Cramer P, Bushnell DA, Kornberg RD. 2001. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**: 1863-1876.

Culver-Cochran AE, Chadwick BP. 2013. Loss of WSTF results in spontaneous fluctuations of heterochromatin formation and resolution, combined with substantial changes to gene expression. *BMC genomics* **14**: 740.

Das C, Tyler JK. 2013. Histone exchange and histone modifications during transcription and aging. *Biochimica et biophysica acta* **1819**: 332-342.

Davidson L, Muniz L, West S. 2014. 3' end formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II CTD are reciprocally coupled in human cells. *Genes & development* **28**: 342-356.

Davis R, Shi Y. 2014. The polyadenylation code: a unified model for the regulation of mRNA alternative polyadenylation. *Journal of Zhejiang University Science B* **15**: 429-437.

De Semir D, Nosrati M, Bezrookove V, Dar AA, Federman S, Bienvenu G, Venna S, Rangel J, Climent J, Meyer Tamguney TM et al. 2012. Pleckstrin homology domain-interacting protein (PHIP) as a marker and mediator of melanoma metastasis. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 7067-7072.

Derheimer FA, O'Hagan HM, Krueger HM, Hanasoge S, Paulsen MT, Ljungman M. 2007. RPA and ATR link transcriptional stress to p53. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 12778-12783.

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG et al. 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome research* **22**: 1775-1789.

Descostes N, Heidemann M, Spinelli L, Schuller R, Maqbool MA, Fenouil R, Koch F, Innocenti C, Gut M, Gut I et al. 2014. Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. *eLife* **3**: e02105.

Dinant C, Ampatziadis-Michailidis G, Lans H, Tresini M, Lagarou A, Grosbart M, Theil AF, van Cappellen WA, Kimura H, Bartek J et al. 2013. Enhanced chromatin dynamics by FACT promotes transcriptional restart after UV-induced DNA damage. *Mol Cell* **51**: 469-479.

Dreos R, Ambrosini G, Bucher P. 2016. Influence of Rotational Nucleosome Positioning on Transcription Start Site Selection in Animal Promoters. *PLoS computational biology* **12**: e1005144.

Drolet M. 2006. Growth inhibition mediated by excess negative supercoiling: the interplay between transcription elongation, R-loop formation and DNA topology. *Molecular microbiology* **59**: 723-730.

Drolet M, Bi X, Liu LF. 1994. Hypernegative supercoiling of the DNA template during transcription elongation in vitro. *The Journal of biological chemistry* **269**: 2068-2074.

Duquette ML, Handa P, Vincent JA, Taylor AF, Maizels N. 2004. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel

structures containing G4 DNA. *Genes & development* **18**: 1618-1629.
Egloff S, Murphy S. 2008. Cracking the RNA polymerase II CTD code. *Trends in genetics : TIG* **24**: 280-288.
Eick D, Geyer M. 2013. The RNA polymerase II carboxy-terminal domain (CTD) code. *Chemical reviews* **113**: 8456-8490.
El Hage A, French SL, Beyer AL, Tollervey D. 2010. Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes & development* **24**: 1546-1558.
Fei J, Kaczmarek N, Luch A, Glas A, Carell T, Naegeli H. 2011. Regulation of nucleotide excision repair by UV-DDB: prioritization of damage recognition to internucleosomal DNA. *PLoS biology* **9**: e1001183.
Feng Z, Scott SP, Bussen W, Sharma GG, Guo G, Pandita TK, Powell SN. 2011. Rad52 inactivation is synthetically lethal with BRCA2 deficiency. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 686-691.
Filipowicz W, Pogacic V. 2002. Biogenesis of small nucleolar ribonucleoproteins. *Current opinion in cell biology* **14**: 319-327.
Flores O, Deniz O, Soler-Lopez M, Orozco M. 2014. Fuzziness and noise in nucleosomal architecture. *Nucleic acids research* **42**: 4934-4946.
Fong N, Brannan K, Erickson B, Kim H, Cortazar MA, Sheridan RM, Nguyen T, Karp S, Bentley DL. 2015. Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. *Mol Cell* **60**: 256-267.
Fousteri M, Vermeulen W, van Zeeland AA, Mullenders LH. 2006. Cockayne syndrome A and B proteins differentially regulate recruitment of chromatin remodeling and repair factors to stalled RNA polymerase II in vivo. *Mol Cell* **23**: 471-482.
Gaillard H, Aguilera A. 2013. Transcription coupled repair at the interface between transcription elongation and mRNP biogenesis. *Biochimica et biophysica acta* **1829**: 141-150.
Gan W, Guan Z, Liu J, Gui T, Shen K, Manley JL, Li X. 2011. R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes & development* **25**: 2041-2056.
Gardini A. 2017. Global Run-On Sequencing (GRO-Seq). *Methods in molecular biology (Clifton, NJ)* **1468**: 111-120.
Ginno PA, Lim YW, Lott PL, Korf I, Chedin F. 2013. GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination. *Genome research* **23**: 1590-1600.
Ginno PA, Lott PL, Christensen HC, Korf I, Chedin F. 2012. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell* **45**: 814-825.
Godin SK, Sullivan MR, Bernstein KA. 2016. Novel insights into RAD51 activity and regulation during homologous recombination and DNA replication. *Biochemistry and cell biology = Biochimie et biologie cellulaire* **94**: 407-418.
Gonzalez-Aguilera C, Tous C, Gomez-Gonzalez B, Huertas P, Luna R, Aguilera A. 2008. The THP1-SAC3-SUS1-CDC31 complex works in

transcription elongation-mRNA export preventing RNA-mediated genome instability. *Molecular biology of the cell* **19**: 4310-4318.
Gottipati P, Cassel TN, Savolainen L, Helleday T. 2008. Transcription-associated recombination is dependent on replication in Mammalian cells. *Molecular and cellular biology* **28**: 154-164.
Gowrishankar J, Harinarayanan R. 2004. Why is transcription coupled to translation in bacteria? *Molecular microbiology* **54**: 598-603.
Groh M, Lufino MM, Wade-Martins R, Gromak N. 2014. R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. *PLoS genetics* **10**: e1004318.
Groisman R, Kuraoka I, Chevallier O, Gaye N, Magnaldo T, Tanaka K, Kisselev AF, Harel-Bellan A, Nakatani Y. 2006. CSA-dependent degradation of CSB by the ubiquitin-proteasome pathway establishes a link between complementation factors of the Cockayne syndrome. *Genes & development* **20**: 1429-1434.
Groisman R, Polanowska J, Kuraoka I, Sawada J, Saijo M, Drapkin R, Kisselev AF, Tanaka K, Nakatani Y. 2003. The ubiquitin ligase activity in the DDB2 and CSA complexes is differentially regulated by the COP9 signalosome in response to DNA damage. *Cell* **113**: 357-367.
Gromak N, West S, Proudfoot NJ. 2006. Pause sites promote transcriptional termination of mammalian RNA polymerase II. *Molecular and cellular biology* **26**: 3986-3996.
Grosso AR, Leite AP, Carvalho S, Matos MR, Martins FB, Vitor AC, Desterro JM, Carmo-Fonseca M, de Almeida SF. 2015. Pervasive transcription read-through promotes aberrant expression of oncogenes and RNA chimeras in renal carcinoma. *eLife* **4**.
Guerrero-Santoro J, Kapetanaki MG, Hsieh CL, Gorbachinsky I, Levine AS, Rapic-Otrin V. 2008. The cullin 4B-based UV-damaged DNA-binding protein ligase binds to UV-damaged chromatin and ubiquitinates histone H2A. *Cancer research* **68**: 5014-5022.
Guiro J, Murphy S. 2017. Regulation of expression of human RNA polymerase II-transcribed snRNA genes. *Open biology* **7**.
Hahn S. 2004. Structure and mechanism of the RNA polymerase II transcription machinery. *Nature structural & molecular biology* **11**: 394-403.
Hampsey M, Reinberg D. 2003. Tails of intrigue: phosphorylation of RNA polymerase II mediates histone methylation. *Cell* **113**: 429-432.
Han J, Zhang H, Zhang H, Wang Z, Zhou H, Zhang Z. 2013. A Cul4 E3 ubiquitin ligase regulates histone hand-off during nucleosome assembly. *Cell* **155**: 817-829.
Hara R, Sancar A. 2002. The SWI/SNF chromatin-remodeling factor stimulates repair by human excision nuclease in the mononucleosome core particle. *Molecular and cellular biology* **22**: 6779-6787.
Hardy JG, Norbury CJ. 2016. Cleavage factor Im (CFIm) as a regulator of alternative polyadenylation. *Biochemical Society transactions* **44**: 1051-1057.
Harlen KM, Trotta KL, Smith EE, Mosaheb MM, Fuchs SM, Churchman LS. 2016. Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. *Cell reports* **15**: 2147-2158.

Hatchi E, Skourti-Stathaki K, Ventz S, Pinello L, Yen A, Kamieniarz-Gdula K, Dimitrov S, Pathania S, McKinney KM, Eaton ML et al. 2015. BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. *Mol Cell* **57**: 636-647.

Hauer MH, Seeber A, Singh V, Thierry R, Sack R, Amitai A, Kryzhanovska M, Eglinger J, Holcman D, Owen-Hughes T et al. 2017. Histone degradation in response to DNA damage enhances chromatin dynamics and recombination rates. *Nature structural & molecular biology* **24**: 99-107.

He YJ, McCall CM, Hu J, Zeng Y, Xiong Y. 2006. DDB1 functions as a linker to recruit receptor WD40 proteins to CUL4-ROC1 ubiquitin ligases. *Genes & development* **20**: 2949-2954.

Helmrich A, Ballarino M, Tora L. 2011. Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol Cell* **44**: 966-977.

Higa LA, Wu M, Ye T, Kobayashi R, Sun H, Zhang H. 2006. CUL4-DDB1 ubiquitin ligase interacts with multiple WD40-repeat proteins and regulates histone methylation. *Nature cell biology* **8**: 1277-1283.

Hintermair C, Heidemann M, Koch F, Descostes N, Gut M, Gut I, Fenouil R, Ferrier P, Flatley A, Kremmer E et al. 2012. Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. *The EMBO journal* **31**: 2784-2797.

Hintermair C, Voss K, Forne I, Heidemann M, Flatley A, Kremmer E, Imhof A, Eick D. 2016. Specific threonine-4 phosphorylation and function of RNA polymerase II CTD during M phase progression. *Scientific reports* **6**: 27401.

Holt IJ, Reyes A. 2012. Human mitochondrial DNA replication. *Cold Spring Harbor perspectives in biology* **4**.

Hong X, Cadwell GW, Kogoma T. 1995. Escherichia coli RecG and RecA proteins in R-loop formation. *The EMBO journal* **14**: 2385-2392.

Hsin JP, Li W, Hoque M, Tian B, Manley JL. 2014a. RNAP II CTD tyrosine 1 performs diverse functions in vertebrate cells. *eLife* **3**: e02112.

Hsin JP, Sheth A, Manley JL. 2011. RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing. *Science* **334**: 683-686.

Hsin JP, Xiang K, Manley JL. 2014b. Function and control of RNA polymerase II C-terminal domain phosphorylation in vertebrate transcription and RNA processing. *Molecular and cellular biology* **34**: 2488-2498.

Hu J, Adar S, Selby CP, Lieb JD, Sancar A. 2015. Genome-wide analysis of human global and transcription-coupled excision repair of UV damage at single-nucleotide resolution. *Genes & development* **29**: 948-960.

Hu J, McCall CM, Ohta T, Xiong Y. 2004. Targeted ubiquitination of CDT1 by the DDB1-CUL4A-ROC1 ligase in response to DNA damage. *Nature cell biology* **6**: 1003-1009.

Huertas P, Aguilera A. 2003. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol Cell* **12**: 711-721.

Huppert JL, Bugaut A, Kumari S, Balasubramanian S. 2008. G-quadruplexes: the beginning and end of UTRs. *Nucleic acids research* **36**: 6260-6268.
Iacovoni JS, Caron P, Lassadi I, Nicolas E, Massip L, Trouche D, Legube G. 2010. High-resolution profiling of gammaH2AX around DNA double strand breaks in the mammalian genome. *The EMBO journal* **29**: 1446-1457.
Itoh T, Tomizawa J. 1980. Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonuclease H. *Proceedings of the National Academy of Sciences of the United States of America* **77**: 2450-2454.
Jensen TH, Jacquier A, Libri D. 2013. Dealing with pervasive transcription. *Mol Cell* **52**: 473-484.
Jenuwein T, Allis CD. 2001. Translating the histone code. *Science* **293**: 1074-1080.
Jiang Y, Wang X, Bao S, Guo R, Johnson DG, Shen X, Li L. 2010. INO80 chromatin remodeling complex promotes the removal of UV lesions by the nucleotide excision repair pathway. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 17274-17279.
Jonkers I, Lis JT. 2015. Getting up to speed with transcription elongation by RNA polymerase II. *Nature reviews Molecular cell biology* **16**: 167-177.
Joshi AA, Struhl K. 2005. Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. *Mol Cell* **20**: 971-978.
Kapetanaki MG, Guerrero-Santoro J, Bisi DC, Hsieh CL, Rapic-Otrin V, Levine AS. 2006. The DDB1-CUL4ADDB2 ubiquitin ligase is deficient in xeroderma pigmentosum group E and targets histone H2A at UV-damaged DNA sites. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 2588-2593.
Kim J, Guermah M, Roeder RG. 2010. The human PAF1 complex acts in chromatin transcription elongation both independently and cooperatively with SII/TFIIS. *Cell* **140**: 491-503.
Kim SM, Dubey DD, Huberman JA. 2003. Early-replicating heterochromatin. *Genes & development* **17**: 330-335.
Kogoma T. 1997. Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiology and molecular biology reviews : MMBR* **61**: 212-238.
Komarnitsky P, Cho EJ, Buratowski S. 2000. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes & development* **14**: 2452-2460.
Kowalczyk MS, Hughes JR, Garrick D, Lynch MD, Sharpe JA, Sloane-Stanley JA, McGowan SJ, De Gobbi M, Hosseini M, Vernimmen D et al. 2012. Intragenic enhancers act as alternative promoters. *Mol Cell* **45**: 447-458.
Kreuzer KN, Brister JR. 2010. Initiation of bacteriophage T4 DNA replication and replication fork dynamics: a review in the Virology Journal series on bacteriophage T4 and its relatives. *Virology journal* **7**: 358.
Krishnamurthy S, He X, Reyes-Reyes M, Moore C, Hampsey M. 2004. Ssu72 Is an RNA polymerase II CTD phosphatase. *Mol Cell* **14**: 387-394.

Kuo RL, Krug RM. 2009. Influenza a virus polymerase is an integral component of the CPSF30-NS1A protein complex in infected cells. *Journal of virology* **83**: 1611-1616.
Kuraoka I, Endou M, Yamaguchi Y, Wada T, Handa H, Tanaka K. 2003. Effects of endogenous DNA base lesions on transcription elongation by mammalian RNA polymerase II. Implications for transcription-coupled DNA repair and transcriptional mutagenesis. *The Journal of biological chemistry* **278**: 7294-7299.
Laitem C, Zaborowska J, Isa NF, Kufs J, Dienstbier M, Murphy S. 2015. CDK9 inhibitors define elongation checkpoints at both ends of RNA polymerase II-transcribed genes. *Nature structural & molecular biology* **22**: 396-403.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357-359.
Le May N, Egly JM, Coin F. 2010. True lies: the double life of the nucleotide excision repair factors in transcription and DNA repair. *Journal of nucleic acids* **2010**.
Lee J, Zhou P. 2007. DCAFs, the missing link of the CUL4-DDB1 ubiquitin ligase. *Mol Cell* **26**: 775-780.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.
Li H, Zhang Z, Wang B, Zhang J, Zhao Y, Jin Y. 2007. Wwp2-mediated ubiquitination of the RNA polymerase II large subunit in mouse embryonic pluripotent stem cells. *Molecular and cellular biology* **27**: 5296-5305.
Li X, Manley JL. 2005. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* **122**: 365-378.
-. 2006. Cotranscriptional processes and their influence on genome stability. *Genes & development* **20**: 1838-1847.
Licatalosi DD, Geiger G, Minet M, Schroeder S, Cilli K, McNeil JB, Bentley DL. 2002. Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. *Mol Cell* **9**: 1101-1111.
Lindsey-Boltz LA, Sancar A. 2007. RNA polymerase: the most specific damage recognition protein in cellular responses to DNA damage? *Proceedings of the National Academy of Sciences of the United States of America* **104**: 13213-13214.
Liu J, Heyer WD. 2011. Who's who in human recombination: BRCA2 and RAD52. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 441-442.
Liu L, Lee S, Zhang J, Peters SB, Hannah J, Zhang Y, Yin Y, Koff A, Ma L, Zhou P. 2009. CUL4A abrogation augments DNA damage response and protection against skin carcinogenesis. *Mol Cell* **34**: 451-460.
Lodhia KA, Gao S, Aleksic T, Esashi F, Macaulay VM. 2015. Suppression of homologous recombination sensitizes human tumor cells to IGF-1R inhibition. *International journal of cancer* **136**: 2961-2966.

Lopes J, Piazza A, Bermejo R, Kriegsman B, Colosio A, Teulade-Fichou MP, Foiani M, Nicolas A. 2011. G-quadruplex-induced instability during leading-strand replication. *The EMBO journal* **30**: 4033-4046.
Lorzadeh A, Bilenky M, Hammond C, Knapp DJ, Li L, Miller PH, Carles A, Heravi-Moussavi A, Gakkhar S, Moksa M et al. 2016. Nucleosome Density ChIP-Seq Identifies Distinct Chromatin Modification Signatures Associated with MNase Accessibility. *Cell reports* **17**: 2112-2124.
Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**: 251-260.
Luse DS. 2014. The RNA polymerase II preinitiation complex. Through what pathway is the complex assembled? *Transcription* **5**: e27050.
Macfarlane LA, Murphy PR. 2010. MicroRNA: Biogenesis, Function and Role in Cancer. *Current genomics* **11**: 537-561.
Mahat DB, Kwak H, Booth GT, Jonkers IH, Danko CG, Patel RK, Waters CT, Munson K, Core LJ, Lis JT. 2016. Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nature protocols* **11**: 1455-1476.
Marteijn JA, Lans H, Vermeulen W, Hoeijmakers JH. 2014. Understanding nucleotide excision repair and its roles in cancer and ageing. *Nature reviews Molecular cell biology* **15**: 465-481.
Marti TM, Hefner E, Feeney L, Natale V, Cleaver JE. 2006. H2AX phosphorylation within the G1 phase after UV irradiation depends on nucleotide excision repair and not DNA double-strand breaks. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 9891-9896.
Martin M. 2011. Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. *EMBnetjournal* **17**.
Marzluff WF, Wagner EJ, Duronio RJ. 2008. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nature reviews Genetics* **9**: 843-854.
Mavrich TN, Ioshikhes IP, Venters BJ, Jiang C, Tomsho LP, Qi J, Schuster SC, Albert I, Pugh BF. 2008. A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome research* **18**: 1073-1083.
Mayer A, Heidemann M, Lidschreiber M, Schreieck A, Sun M, Hintermair C, Kremmer E, Eick D, Cramer P. 2012. CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science* **336**: 1723-1725.
Mayne LV, Lehmann AR. 1982. Failure of RNA synthesis to recover after UV irradiation: an early defect in cells from individuals with Cockayne's syndrome and xeroderma pigmentosum. *Cancer research* **42**: 1473-1478.
Mei Kwei JS, Kuraoka I, Horibata K, Ubukata M, Kobatake E, Iwai S, Handa H, Tanaka K. 2004. Blockage of RNA polymerase II at a cyclobutane pyrimidine dimer and 6-4 photoproduct. *Biochemical and biophysical research communications* **320**: 1133-1138.
Meinhart A, Cramer P. 2004. Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature* **430**: 223-226.

Meinhart A, Kamenski T, Hoeppner S, Baumli S, Cramer P. 2005. A structural perspective of CTD function. *Genes & development* **19**: 1401-1415.
Meyer CA, Liu XS. 2014. Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. *Nature reviews Genetics* **15**: 709-721.
Mieczkowski J, Cook A, Bowman SK, Mueller B, Alver BH, Kundu S, Deaton AM, Urban JA, Larschan E, Park PJ et al. 2016. MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. *Nature communications* **7**: 11485.
Milek M, Imami K, Mukherjee N, Bortoli F, Zinnall U, Hazapis O, Trahan C, Oeffinger M, Heyd F, Ohler U et al. 2017. DDX54 regulates transcriptome dynamics during DNA damage response. *Genome research* **27**: 1344-1359.
Mirkin EV, Mirkin SM. 2007. Replication fork stalling at natural impediments. *Microbiology and molecular biology reviews : MMBR* **71**: 13-35.
Mischo HE, Gomez-Gonzalez B, Grzechnik P, Rondon AG, Wei W, Steinmetz L, Aguilera A, Proudfoot NJ. 2011. Yeast Sen1 helicase protects the genome from transcription-associated instability. *Mol Cell* **41**: 21-32.
Moniaux N, Nemos C, Schmied BM, Chauhan SC, Deb S, Morikane K, Choudhury A, Vanlith M, Sutherlin M, Sikela JM et al. 2006. The human homologue of the RNA polymerase II-associated factor 1 (hPaf1), localized on the 19q13 amplicon, is associated with tumorigenesis. *Oncogene* **25**: 3247-3257.
Mora A, Sandve GK, Gabrielsen OS, Eskeland R. 2016. In the loop: promoter-enhancer interactions and bioinformatics. *Briefings in bioinformatics* **17**: 980-995.
Morris SA, Baek S, Sung MH, John S, Wiench M, Johnson TA, Schiltz RL, Hager GL. 2014. Overlapping chromatin-remodeling systems collaborate genome wide at dynamic chromatin transitions. *Nature structural & molecular biology* **21**: 73-81.
Moser J, Kool H, Giakzidis I, Caldecott K, Mullenders LH, Fousteri MI. 2007. Sealing of chromosomal DNA nicks during nucleotide excision repair requires XRCC1 and DNA ligase III alpha in a cell-cycle-specific manner. *Mol Cell* **27**: 311-323.
Mueller CL, Porter SE, Hoffman MG, Jaehning JA. 2004. The Paf1 complex has functions independent of actively transcribing RNA polymerase II. *Mol Cell* **14**: 447-456.
Munoz MJ, Nieto Moreno N, Giono LE, Cambindo Botto AE, Dujardin G, Bastianello G, Lavore S, Torres-Mendez A, Menck CFM, Blencowe BJ et al. 2017. Major Roles for Pyrimidine Dimers, Nucleotide Excision Repair, and ATR in the Alternative Splicing Response to UV Irradiation. *Cell reports* **18**: 2868-2879.
Munoz MJ, Perez Santangelo MS, Paronetto MP, de la Mata M, Pelisch F, Boireau S, Glover-Cutter K, Ben-Dov C, Blaustein M, Lozano JJ et al. 2009. DNA damage regulates alternative splicing through inhibition of RNA polymerase II elongation. *Cell* **137**: 708-720.

Nagaike T, Logan C, Hotta I, Rozenblatt-Rosen O, Meyerson M, Manley JL. 2011. Transcriptional activators enhance polyadenylation of mRNA precursors. *Mol Cell* **41**: 409-418.
Nemeroff ME, Barabino SM, Li Y, Keller W, Krug RM. 1998. Influenza virus NS1 protein interacts with the cellular 30 kDa subunit of CPSF and inhibits 3'end formation of cellular pre-mRNAs. *Mol Cell* **1**: 991-1000.
New JH, Sugiyama T, Zaitseva E, Kowalczykowski SC. 1998. Rad52 protein stimulates DNA strand exchange by Rad51 and replication protein A. *Nature* **391**: 407-410.
Nojima T, Dienstbier M, Murphy S, Proudfoot NJ, Dye MJ. 2013. Definition of RNA polymerase II CoTC terminator elements in the human genome. *Cell reports* **3**: 1080-1092.
Nojima T, Gomes T, Carmo-Fonseca M, Proudfoot NJ. 2016. Mammalian NET-seq analysis defines nascent RNA profiles and associated RNA processing genome-wide. *Nature protocols* **11**: 413-428.
Nojima T, Gomes T, Grosso AR, Kimura H, Dye MJ, Dhir S, Carmo-Fonseca M, Proudfoot NJ. 2015. Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* **161**: 526-540.
Ogi T, Limsirichaikul S, Overmeer RM, Volker M, Takenaka K, Cloney R, Nakazawa Y, Niimi A, Miki Y, Jaspers NG et al. 2010. Three DNA polymerases, recruited by different mechanisms, carry out NER repair synthesis in human cells. *Mol Cell* **37**: 714-727.
Paule MR, White RJ. 2000. Survey and summary: transcription by RNA polymerases I and III. *Nucleic acids research* **28**: 1283-1298.
Paull TT, Rogakou EP, Yamazaki V, Kirchgessner CU, Gellert M, Bonner WM. 2000. A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. *Current biology : CB* **10**: 886-895.
Paulsen RD, Soni DV, Wollman R, Hahn AT, Yee MC, Guan A, Hesley JA, Miller SC, Cromwell EF, Solow-Cordero DE et al. 2009. A genome-wide siRNA screen reveals diverse cellular processes and pathways that mediate genome stability. *Mol Cell* **35**: 228-239.
Pellegrini L, Yu DS, Lo T, Anand S, Lee M, Blundell TL, Venkitaraman AR. 2002. Insights into DNA recombination from the structure of a RAD51-BRCA2 complex. *Nature* **420**: 287-293.
Percipalle P, Farrants AK. 2006. Chromatin remodelling and transcription: be-WICHed by nuclear myosin 1. *Current opinion in cell biology* **18**: 267-274.
Percipalle P, Fomproix N, Cavellan E, Voit R, Reimer G, Kruger T, Thyberg J, Scheer U, Grummt I, Farrants AK. 2006. The chromatin remodelling complex WSTF-SNF2h interacts with nuclear myosin 1 and has a role in RNA polymerase I transcription. *EMBO reports* **7**: 525-530.
Perera D, Poulos RC, Shah A, Beck D, Pimanda JE, Wong JW. 2016. Differential DNA repair underlies mutation hotspots at active promoters in cancer genomes. *Nature* **532**: 259-263.

Peterlin BM, Price DH. 2006. Controlling the elongation phase of transcription with P-TEFb. *Mol Cell* **23**: 297-305.
Phatnani HP, Greenleaf AL. 2006. Phosphorylation and functions of the RNA polymerase II CTD. *Genes & development* **20**: 2922-2936.
Pohjoismaki JL, Holmes JB, Wood SR, Yang MY, Yasukawa T, Reyes A, Bailey LJ, Cluett TJ, Goffart S, Willcox S et al. 2010. Mammalian mitochondrial DNA replication intermediates are essentially duplex but contain extensive tracts of RNA/DNA hybrid. *Journal of molecular biology* **397**: 1144-1155.
Poot RA, Bozhenok L, van den Berg DL, Hawkes N, Varga-Weisz PD. 2005. Chromatin remodeling by WSTF-ISWI at the replication site: opening a window of opportunity for epigenetic inheritance? *Cell cycle (Georgetown, Tex)* **4**: 543-546.
Poot RA, Bozhenok L, van den Berg DL, Steffensen S, Ferreira F, Grimaldi M, Gilbert N, Ferreira J, Varga-Weisz PD. 2004. The Williams syndrome transcription factor interacts with PCNA to target chromatin remodelling by ISWI to replication foci. *Nature cell biology* **6**: 1236-1244.
Prado F, Aguilera A. 2005. Impairment of replication fork progression mediates RNA polII transcription-associated recombination. *The EMBO journal* **24**: 1267-1276.
Proietti-De-Santis L, Drane P, Egly JM. 2006. Cockayne syndrome B protein regulates the transcriptional program after UV irradiation. *The EMBO journal* **25**: 1915-1923.
Proudfoot NJ. 2016. Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science* **352**: aad9926.
Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.
Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dundar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**: W160-165.
Rau M, Ohlmann T, Morley SJ, Pain VM. 1996. A reevaluation of the cap-binding protein, eIF4E, as a rate-limiting factor for initiation of translation in reticulocyte lysate. *The Journal of biological chemistry* **271**: 8983-8990.
Reddy K, Tam M, Bowater RP, Barber M, Tomlinson M, Nichol Edamura K, Wang YH, Pearson CE. 2011. Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. *Nucleic acids research* **39**: 1749-1762.
Reyes-Reyes M, Hampsey M. 2007. Role for the Ssu72 C-terminal domain phosphatase in RNA polymerase II transcription elongation. *Molecular and cellular biology* **27**: 926-936.
Roberts RW, Crothers DM. 1992. Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. *Science* **258**: 1463-1466.
Robinson PJ, Fairall L, Huynh VA, Rhodes D. 2006. EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure. *Proceedings of the National*

Academy of Sciences of the United States of America **103**: 6506-6511.

Rockx DA, Mason R, van Hoffen A, Barton MC, Citterio E, Bregman DB, van Zeeland AA, Vrieling H, Mullenders LH. 2000. UV-induced inhibition of transcription involves repression of transcription initiation and phosphorylation of RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 10503-10508.

Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A. 2004. Identification of mammalian microRNA host genes and transcription units. *Genome research* **14**: 1902-1910.

Rosonina E, Yurko N, Li W, Hoque M, Tian B, Manley JL. 2014. Threonine-4 of the budding yeast RNAP II CTD couples transcription with Htz1-mediated chromatin remodeling. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 11924-11931.

Roy D, Lieber MR. 2009. G clustering is important for the initiation of transcription-induced R-loops in vitro, whereas high G density without clustering is sufficient thereafter. *Molecular and cellular biology* **29**: 3124-3133.

Roy D, Zhang Z, Lu Z, Hsieh CL, Lieber MR. 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. *Molecular and cellular biology* **30**: 146-159.

Rutkowski AJ, Erhard F, L'Hernault A, Bonfert T, Schilhabel M, Crump C, Rosenstiel P, Efstathiou S, Zimmer R, Friedel CC et al. 2015. Widespread disruption of host transcription termination in HSV-1 infection. *Nature communications* **6**: 7126.

Sabarinathan R, Mularoni L, Deu-Pons J, Gonzalez-Perez A, Lopez-Bigas N. 2016. Nucleotide excision repair is impaired by binding of transcription factors to DNA. *Nature* **532**: 264-267.

Sadeghifar F, Bohm S, Vintermist A, Ostlund Farrants AK. 2015. The B-WICH chromatin-remodelling complex regulates RNA polymerase III transcription by promoting Max-dependent c-Myc binding. *Nucleic acids research* **43**: 4477-4490.

Santoro R, Li J, Grummt I. 2002. The nucleolar remodeling complex NoRC mediates heterochromatin formation and silencing of ribosomal gene transcription. *Nature genetics* **32**: 393-396.

Schlackow M, Nojima T, Gomes T, Dhir A, Carmo-Fonseca M, Proudfoot NJ. 2017. Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. *Mol Cell* **65**: 25-38.

Schonemann L, Kuhn U, Martin G, Schafer P, Gruber AR, Keller W, Zavolan M, Wahle E. 2014. Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. *Genes & development* **28**: 2381-2393.

Schreieck A, Easter AD, Etzold S, Wiederhold K, Lidschreiber M, Cramer P, Passmore LA. 2014. RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7. *Nature structural & molecular biology* **21**: 175-179.

Schuller R, Forne I, Straub T, Schreieck A, Texier Y, Shah N, Decker TM, Cramer P, Imhof A, Eick D. 2016. Heptad-Specific Phosphorylation of RNA Polymerase II CTD. *Mol Cell* **61**: 305-314.

Scott MS, Ono M. 2011. From snoRNA to miRNA: Dual function regulatory non-coding RNAs. *Biochimie* **93**: 1987-1992.

Scrima A, Konickova R, Czyzewski BK, Kawasaki Y, Jeffrey PD, Groisman R, Nakatani Y, Iwai S, Pavletich NP, Thoma NH. 2008. Structural basis of UV DNA-damage recognition by the DDB1-DDB2 complex. *Cell* **135**: 1213-1223.

Seo J, Kim SC, Lee HS, Kim JK, Shon HJ, Salleh NL, Desai KV, Lee JH, Kang ES, Kim JS et al. 2012. Genome-wide profiles of H2AX and gamma-H2AX differentiate endogenous and exogenous DNA damage hotspots in human cells. *Nucleic acids research* **40**: 5965-5974.

Shandilya J, Roberts SG. 2012. The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. *Biochimica et biophysica acta* **1819**: 391-400.

Sharma A, Singh K, Almasan A. 2012. Histone H2AX phosphorylation: a marker for DNA damage. *Methods in molecular biology (Clifton, NJ)* **920**: 613-626.

Shaw NN, Arya DP. 2008. Recognition of the unique structure of DNA:RNA hybrids. *Biochimie* **90**: 1026-1039.

Shinohara A, Ogawa T. 1998. Stimulation by Rad52 of yeast Rad51-mediated recombination. *Nature* **391**: 404-407.

Sigurdsson S, Dirac-Svejstrup AB, Svejstrup JQ. 2010. Evidence that transcript cleavage is essential for RNA polymerase II transcription and cell viability. *Mol Cell* **38**: 202-210.

Sims RJ, 3rd, Rojas LA, Beck DB, Bonasio R, Schuller R, Drury WJ, 3rd, Eick D, Reinberg D. 2011. The C-terminal domain of RNA polymerase II is modified by site-specific methylation. *Science* **332**: 99-103.

Skourti-Stathaki K, Kamieniarz-Gdula K, Proudfoot NJ. 2014. R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature*.

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

Sogaard TM, Svejstrup JQ. 2007. Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. *The Journal of biological chemistry* **282**: 14113-14120.

Sollier J, Stork CT, Garcia-Rubio ML, Paulsen RD, Aguilera A, Cimprich KA. 2014. Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol Cell* **56**: 777-785.

Song C, Hotz-Wagenblatt A, Voit R, Grummt I. 2017. SIRT7 and the DEAD-box helicase DDX21 cooperate to resolve genomic R loops and safeguard genome stability. *Genes & development*.

Squazzo SL, Costa PJ, Lindstrom DL, Kumer KE, Simic R, Jennings JL, Link AJ, Arndt KM, Hartzog GA. 2002. The Paf1 complex physically and functionally associates with transcription elongation factors in vivo. *The EMBO journal* **21**: 1764-1774.

Sridhara SC, Carvalho S, Grosso AR, Gallego-Paez LM, Carmo-Fonseca M, de Almeida SF. 2017. Transcription Dynamics Prevent RNA-Mediated Genomic Instability through SRPK2-Dependent DDX23 Phosphorylation. *Cell reports* **18**: 334-343.

St Laurent G, Wahlestedt C, Kapranov P. 2015. The Landscape of long noncoding RNA classification. *Trends in genetics : TIG* **31**: 239-251.

Stevnsner T, Muftuoglu M, Aamann MD, Bohr VA. 2008. The role of Cockayne Syndrome group B (CSB) protein in base excision repair and aging. *Mechanisms of ageing and development* **129**: 441-448.

Stiller JW, McConaughy BL, Hall BD. 2000. Evolutionary complementation for polymerase II CTD function. *Yeast (Chichester, England)* **16**: 57-64.

Stirling PC, Chan YA, Minaker SW, Aristizabal MJ, Barrett I, Sipahimalani P, Kobor MS, Hieter P. 2012. R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. *Genes & development* **26**: 163-175.

Sugasawa K, Ng JM, Masutani C, Iwai S, van der Spek PJ, Eker AP, Hanaoka F, Bootsma D, Hoeijmakers JH. 1998. Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. *Mol Cell* **2**: 223-232.

Sugasawa K, Okuda Y, Saijo M, Nishi R, Matsuda N, Chu G, Mori T, Iwai S, Tanaka K, Tanaka K et al. 2005. UV-induced ubiquitylation of XPC protein mediated by UV-DDB-ubiquitin ligase complex. *Cell* **121**: 387-400.

Sugiyama T, New JH, Kowalczykowski SC. 1998. DNA annealing by RAD52 protein is stimulated by specific interaction with the complex of replication protein A and single-stranded DNA. *Proceedings of the National Academy of Sciences of the United States of America* **95**: 6049-6054.

Suh H, Ficarro SB, Kang UB, Chun Y, Marto JA, Buratowski S. 2016. Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II. *Mol Cell* **61**: 297-304.

Sun H, Martin JA, Werner CT, Wang ZJ, Damez-Werno DM, Scobie KN, Shao NY, Dias C, Rabkin J, Koo JW et al. 2016. BAZ1B in Nucleus Accumbens Regulates Reward-Related Behaviors in Response to Distinct Emotional Stimuli. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **36**: 3954-3961.

Sung P, Klein H. 2006. Mechanism of homologous recombination: mediators and helicases take on regulatory functions. *Nature reviews Molecular cell biology* **7**: 739-750.

Takahashi A, Ohnishi T. 2005. Does gammaH2AX foci formation depend on the presence of DNA double strand breaks? *Cancer letters* **229**: 171-179.

Takahashi A, Tsutsumi R, Kikuchi I, Obuse C, Saito Y, Seidi A, Karisch R, Fernandez M, Cho T, Ohnishi N et al. 2011. SHP2 tyrosine phosphatase converts parafibromin/Cdc73 from a tumor suppressor to an oncogenic driver. *Mol Cell* **43**: 45-56.

Tapias A, Auriol J, Forget D, Enzlin JH, Scharer OD, Coin F, Coulombe B, Egly JM. 2004. Ordered conformational changes in damaged DNA induced by nucleotide excision repair factors. *The Journal of biological chemistry* **279**: 19074-19083.

Tian B, Manley JL. 2013. Alternative cleavage and polyadenylation: the long and short of it. *Trends in biochemical sciences* **38**: 312-320.
Tian B, Pan Z, Lee JY. 2007. Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. *Genome research* **17**: 156-165.
Tornaletti S, Reines D, Hanawalt PC. 1999. Structural characterization of RNA polymerase II complexes arrested by a cyclobutane pyrimidine dimer in the transcribed strand of template DNA. *The Journal of biological chemistry* **274**: 24124-24130.
Tycowski KT, Steitz JA. 2001. Non-coding snoRNA host genes in Drosophila: expression strategies for modification guide snoRNAs. *European journal of cell biology* **80**: 119-125.
Vilborg A, Passarelli MC, Yario TA, Tycowski KT, Steitz JA. 2015. Widespread Inducible Transcription Downstream of Human Genes. *Mol Cell* **59**: 449-461.
Vintermist A, Bohm S, Sadeghifar F, Louvet E, Mansen A, Percipalle P, Ostlund Farrants AK. 2011. The chromatin remodelling complex B-WICH changes the chromatin structure and recruits histone acetyl-transferases to active rRNA genes. *PloS one* **6**: e19184.
Volker M, Mone MJ, Karmakar P, van Hoffen A, Schul W, Vermeulen W, Hoeijmakers JH, van Driel R, van Zeeland AA, Mullenders LH. 2001. Sequential assembly of the nucleotide excision repair factors in vivo. *Mol Cell* **8**: 213-224.
Voong LN, Xi L, Sebeson AC, Xiong B, Wang JP, Wang X. 2016. Insights into Nucleosome Organization in Mouse Embryonic Stem Cells through Chemical Mapping. *Cell* **167**: 1555-1570.e1515.
Voong LN, Xi L, Wang JP, Wang X. 2017. Genome-wide Mapping of the Nucleosome Landscape by Micrococcal Nuclease and Chemical Mapping. *Trends in genetics : TIG* **33**: 495-507.
Wagner EJ, Carpenter PB. 2012. Understanding the language of Lys36 methylation at histone H3. *Nature reviews Molecular cell biology* **13**: 115-126.
Wahba L, Amon JD, Koshland D, Vuica-Ross M. 2011. RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Mol Cell* **44**: 978-988.
Wahba L, Costantino L, Tan FJ, Zimmer A, Koshland D. 2016. S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes & development* **30**: 1327-1338.
Wahba L, Gore SK, Koshland D. 2013. The homologous recombination machinery modulates the formation of RNA-DNA hybrids and associated chromosome instability. *eLife* **2**: e00505.
Wakasugi M, Kawashima A, Morioka H, Linn S, Sancar A, Mori T, Nikaido O, Matsunaga T. 2002. DDB accumulates at DNA damage sites immediately after UV irradiation and directly stimulates nucleotide excision repair. *The Journal of biological chemistry* **277**: 1637-1640.
Wang H, Zhai L, Xu J, Joo HY, Jackson S, Erdjument-Bromage H, Tempst P, Xiong Y, Zhang Y. 2006. Histone H3 and H4 ubiquitylation by the CUL4-

DDB-ROC1 ubiquitin ligase facilitates cellular response to DNA damage. *Mol Cell* **22**: 383-394.

Watanabe R, Kanno SI, Mohammadi Roushandeh A, Ui A, Yasui A. 2017. Nucleosome remodelling, DNA repair and transcriptional regulation build negative feedback loops in cancer and cellular ageing. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* **372**.

Weems JC, Slaughter BD, Unruh JR, Boeing S, Hall SM, McLaird MB, Yasukawa T, Aso T, Svejstrup JQ, Conaway JW et al. 2017. Cockayne syndrome B protein regulates recruitment of the Elongin A ubiquitin ligase to sites of DNA damage. *The Journal of biological chemistry* **292**: 6431-6437.

Wellinger RE, Prado F, Aguilera A. 2006. Replication fork progression is impaired by transcription in hyperrecombinant yeast cells lacking a functional THO complex. *Molecular and cellular biology* **26**: 3327-3334.

Westover KD, Bushnell DA, Kornberg RD. 2004. Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science* **303**: 1014-1016.

White E, Kamieniarz-Gdula K, Dye MJ, Proudfoot NJ. 2013. AT-rich sequence elements promote nascent transcript cleavage leading to RNA polymerase II termination. *Nucleic acids research* **41**: 1797-1806.

Wiechens N, Singh V, Gkikopoulos T, Schofield P, Rocha S, Owen-Hughes T. 2016. The Chromatin Remodelling Enzymes SNF2H and SNF2L Position Nucleosomes adjacent to CTCF and Other Transcription Factors. *PLoS genetics* **12**: e1005940.

Williamson L, Saponaro M, Boeing S, East P, Mitter R, Kantidakis T, Kelly GP, Lobley A, Walker J, Spencer-Dene B et al. 2017. UV Irradiation Induces a Non-coding RNA that Functionally Opposes the Protein Encoded by the Same Gene. *Cell* **168**: 843-855.e813.

Wilson MD, Harreman M, Svejstrup JQ. 2013. Ubiquitylation and degradation of elongating RNA polymerase II: the last resort. *Biochimica et biophysica acta* **1829**: 151-157.

Wongsurawat T, Jenjaroenpun P, Kwoh CK, Kuznetsov V. 2012. Quantitative model of R-loop forming structures reveals a novel level of RNA-DNA interactome complexity. *Nucleic acids research* **40**: e16.

Woudstra EC, Gilbert C, Fellows J, Jansen L, Brouwer J, Erdjument-Bromage H, Tempst P, Svejstrup JQ. 2002. A Rad26-Def1 complex coordinates repair and RNA pol II proteolysis in response to DNA damage. *Nature* **415**: 929-933.

Wu C, McGeehan JE, Travers A. 2016. A metastable structure for the compact 30-nm chromatin fibre. *FEBS letters* **590**: 935-942.

Xiao A, Li H, Shechter D, Ahn SH, Fabrizio LA, Erdjument-Bromage H, Ishibe-Murakami S, Wang B, Tempst P, Hofmann K et al. 2009. WSTF regulates the function of H2A.X via a novel tyrosine kinase activity. *Nature* **457**: 57-62.

Xiao T, Hall H, Kizer KO, Shibata Y, Hall MC, Borchers CH, Strahl BD. 2003. Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. *Genes & development* **17**: 654-663.

Xu J, Zhao L, Xu Y, Zhao W, Sung P, Wang HW. 2017. Cryo-EM structures of human RAD51 recombinase filaments during catalysis of DNA-strand exchange. *Nature structural & molecular biology* **24**: 40-46.
Yang Y, Li W, Hoque M, Hou L, Shen S, Tian B, Dynlacht BD. 2016. PAF Complex Plays Novel Subunit-Specific Roles in Alternative Cleavage and Polyadenylation. *PLoS genetics* **12**: e1005794.
Yata K, Bleuyard JY, Nakato R, Ralf C, Katou Y, Schwab RA, Niedzwiedz W, Shirahige K, Esashi F. 2014. BRCA2 coordinates the activities of cell-cycle kinases to promote genome stability. *Cell reports* **7**: 1547-1559.
Yata K, Lloyd J, Maslen S, Bleuyard JY, Skehel M, Smerdon SJ, Esashi F. 2012. Plk1 and CK2 act in concert to regulate Rad51 during DNA double strand break repair. *Mol Cell* **45**: 371-383.
Yates LA, Norbury CJ, Gilbert RJ. 2013. The long and short of microRNA. *Cell* **153**: 516-519.
Yokoi M, Masutani C, Maekawa T, Sugasawa K, Ohkuma Y, Hanaoka F. 2000. The xeroderma pigmentosum group C protein complex XPC-HR23B plays an important role in the recruitment of transcription factor IIH to damaged DNA. *The Journal of biological chemistry* **275**: 9870-9875.
Yu K, Chedin F, Hsieh CL, Wilson TE, Lieber MR. 2003. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nature immunology* **4**: 442-451.
Yu M, Yang W, Ni T, Tang Z, Nakadai T, Zhu J, Roeder RG. 2015. RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. *Science* **350**: 1383-1386.
Yuce O, West SC. 2013. Senataxin, defective in the neurodegenerative disorder ataxia with oculomotor apraxia 2, lies at the interface of transcription and the DNA damage response. *Molecular and cellular biology* **33**: 406-417.
Yurko NM, Manley JL. 2017. The RNA polymerase II CTD "orphan" residues: Emerging insights into the functions of Tyr-1, Thr-4 and Ser-7. *Transcription*: 0.
Zaborowska J, Egloff S, Murphy S. 2016. The pol II CTD: new twists in the tail. *Nature structural & molecular biology* **23**: 771-777.
Zentner GE, Henikoff S. 2014. High-resolution digital profiling of the epigenome. *Nature reviews Genetics* **15**: 814-827.
Zhang H, Rigo F, Martinson HG. 2015. Poly(A) Signal-Dependent Transcription Termination Occurs through a Conformational Change Mechanism that Does Not Require Cleavage at the Poly(A) Site. *Mol Cell* **59**: 437-448.
Zhang M, Wang XJ, Chen X, Bowman ME, Luo Y, Noel JP, Ellington AD, Etzkorn FA, Zhang Y. 2012. Structural and kinetic analysis of prolyl-isomerization/phosphorylation cross-talk in the CTD code. *ACS chemical biology* **7**: 1462-1470.
Zhao Q, Wang QE, Ray A, Wani G, Han C, Milum K, Wani AA. 2009. Modulation of nucleotide excision repair by mammalian SWI/SNF chromatin-remodeling complex. *The Journal of biological chemistry* **284**: 30424-30432.

Zhu B, Mandal SS, Pham AD, Zheng Y, Erdjument-Bromage H, Batra SK, Tempst P, Reinberg D. 2005. The human PAF complex coordinates transcription with events downstream of RNA synthesis. *Genes & development* **19**: 1668-1673.

Appendices

Appendix 1. Mass spectrometry results - *1.1 Phospho-threonine 4 Mass Spectrometry Results*

RNAPII components

	SPECTRAL COUNTS			SPECTRAL INDEX (MIC SIn)		
Protein	**NoAb Control**	**T4P siCPSF73**	**T4P siLuc**	**NoAb Control**	**T4P siCPSF73**	**T4P siLuc**
Ig kappa chain V-II region Cum		40.00	40.00		5.36E-03	5.21E-03
Histone H4	6.00	10.00	12.00	2.75E-04	8.40E-05	5.19E-05
RNAPII RPB3		15.00	18.00		7.91E-05	7.31E-05
RNAPII RPB2		105.00	131.00		7.19E-05	8.19E-05
RNAPII RPB1		167.00	189.00		6.85E-05	7.00E-05
RNAPII RPB9		10.99	12.99		5.69E-05	6.85E-05
RNAPI, II and III RPABC1 subunit		16.00	23.00		5.64E-05	8.58E-05
RNAPII RPB11a		6.94	8.91		5.56E-05	3.84E-05
Histone H2B type 1-K	3.96	4.96	4.97	1.04E-04	5.43E-05	9.87E-06
RNAPII RPB4 subunit		11.95	15.95		4.40E-05	5.92E-05
Ig kappa chain C region OS	14.00	11.00	10.00	4.29E-04	3.59E-05	2.34E-05
RNAPII RPB7		11.98	16.98		3.40E-05	4.22E-05
RNAPI, II and III RPABC3 subunit		6.00	10.00		3.37E-05	3.69E-05
RNAPI, II and III RPABC2 subunit		3.00	3.00		2.98E-05	2.53E-05
RNAPI, II and III RPABC5 subunit		5.00	8.00		2.62E-05	3.65E-05
Histone H3.1	1.96	3.93	4.92	3.84E-05	2.18E-05	2.83E-05
Immunoglobulin lambda-like polypeptide 5	6.00	2.50	1.50	1.17E-04	1.54E-05	1.36E-06
Histone H2A type 1-B/E	1.98	2.98	1.98	5.98E-05	1.30E-05	8.03E-06
RNAPI, II and III RPABC4 subunit		2.00	3.00		1.29E-05	1.66E-05
Ig gamma-3 chain C	1.00	3.50	5.50	6.12E-06	9.60E-06	1.03E-05
Histone H1.2	4.99	3.99	2.98	5.07E-05	5.80E-06	2.57E-06
Spt6		1.00	3.00		8.62E-08	3.09E-07
Spt5			3.00			2.71E-07
Serine/arginine-rich splicing factor 7	1.50		1.00	1.20E-05		6.07E-07

1.2 Phospho-tyrosine 1 Mass spectrometry results

Group

Group NoAb control

Group Y1P RNAPII

RNAPII components

DNA repair and chromatin remodelling proteins

Other transcriptional regulators and RNA binding proteins

Results

	SPECTRAL COUNTS		SPECTRAL INDEX (MIC SIn)		RATIO
Protein	**NoAb**	**Y1P**	**NoAb**	**Y1P**	**Y1P/ctrl**
Histone H4	7.00	64.00	1.59E-04	4.42E-04	2.79
Histone H2A 2C		5.96		4.14E-04	Y1P Only
Histone H2B 1K	3.97	45.60	1.16E-04	2.07E-04	1.78
RNAPII RPB7		78.95		1.70E-04	Y1P Only
RNAPII RPB3		75.00		1.69E-04	Y1P Only
RNAPI, II and III RPABC1 subunit		47.00		1.68E-04	Y1P Only
RNAPII RPB4 subunit		52.83		1.46E-04	Y1P Only
RNAPII RPB1		536.00		1.32E-04	Y1P Only
RNAPII RPB9		18.98		1.30E-04	Y1P Only
RNAPII RPB2		346.00		1.06E-04	Y1P Only
RNAPII RPB11		21.79		9.17E-05	Y1P Only
Ribosome biogenesis protein BRX1 homolog	7.00	31.00	1.23E-04	6.24E-05	0.51
RNAPI, II and III RPABC3 subunit		45.36		5.30E-05	Y1P Only
RNAPI, II and III RPABC4 subunit		5.00		2.59E-05	Y1P Only
Histone macro-H2A.1		11.96		2.01E-05	Y1P Only
RNAPI, II and III RPABC5 subunit		13.00		2.01E-05	Y1P Only
Histone H3.1		6.93		1.98E-05	Y1P Only

Ig kappa chain C region	9.00	12.00	4.46E-04	1.91E-05		0.04
RNAPI, II and III RPABC2 subunit		10.95		1.73E-05	Y1P Only	
60S ribosomal protein L8	4.95	15.82	5.03E-05	1.66E-05		0.33
Ribosomal L1 domain-containing protein 1	6.93	44.41	1.50E-05	1.61E-05		1.08
60S ribosomal protein L18a	1.98	13.91	8.04E-06	1.61E-05		2.00
Ig lambda-2 chain C regions	3.97	2.98	7.47E-04	1.06E-05		0.01
Probable rRNA-processing protein EBP2	6.00	22.00	2.61E-05	1.02E-05		0.39
Histone H1.5	4.00	11.00	3.95E-05	9.62E-06		0.24
Ribosome production factor 2 homolog	2.00	19.00	1.18E-05	9.41E-06		0.80
Histone H2A.Z		5.94		9.38E-06	Y1P Only	
Ribosome biogenesis regulatory protein homolog	2.00	24.00	5.24E-06	9.24E-06		1.76
Heterogeneous nuclear ribonucleoproteins C1/C2	2.93	21.39	1.69E-05	8.95E-06		0.53
SPT6		119.95		8.73E-06	Y1P Only	
Histone H1x	2.00	7.00	1.31E-05	7.74E-06		0.59
Polyubiquitin-B	0.98	8.84	5.95E-06	7.36E-06		1.24
Isoform 2 of Nucleophosmin		12.93		7.06E-06	Y1P Only	
DDB1		54.27		6.51E-06	Y1P Only	
NEDD8		3.00		5.49E-06	Y1P Only	
Nucleolar GTP-binding protein 1		31.70		5.46E-06	Y1P Only	
mRNA turnover protein 4 homolog		16.00		5.05E-06	Y1P Only	
DDX54	1.99	33.79	2.18E-06	4.96E-06		2.28
PHIP		67.50		4.52E-06	Y1P Only	
Serine/threonine-protein phosphatase PGAM5		11.95		4.44E-06	Y1P Only	
Coiled-coil domain-containing protein 86		8.00		4.32E-06	Y1P Only	
Probable 28S rRNA methyltransferase	1.00	20.92	1.24E-06	3.98E-06		3.22
rRNA 2'-O-methyltransferase fibrillarin	0.99	14.76	5.12E-06	3.91E-06		0.76
Regulation of nuclear pre-mRNA domain-containing protein 1A		13.95		3.76E-06	Y1P Only	
Histone H2A type 2-B		1.99		3.59E-06	Y1P Only	
Nuclear cap-binding protein subunit 1		41.00		3.50E-06	Y1P Only	
Eukaryotic translation initiation factor 6	1.00	8.00	2.68E-06	3.49E-06		1.30

RNA-binding protein Raly		8.91		3.48E-06	Y1P Only	
Cap-specific mRNA methyltransferase 1		35.00		3.20E-06	Y1P Only	
Serine/arginine-rich splicing factor 7	2.00	5.50	4.81E-05	2.85E-06		0.06
Nucleolar protein 16		5.00		2.84E-06	Y1P Only	
CUL4a		26.00		2.79E-06	Y1P Only	
pre-rRNA processing protein FTSJ3		22.00		2.60E-06	Y1P Only	
Histone H1.2		2.00		2.43E-06	Y1P Only	
Guanine nucleotide-binding protein-like 3	1.00	16.00	1.03E-06	2.41E-06		2.35
Nucleolar GTP-binding protein 2		25.00		2.36E-06	Y1P Only	
Nuclear cap-binding protein subunit 2		6.95		2.19E-06	Y1P Only	
Pumilio domain-containing protein KIAA0020		17.99		2.19E-06	Y1P Only	
Glioma tumour suppressor candidate region gene 2 protein		8.96		2.14E-06	Y1P Only	
Small nuclear ribonucleoprotein Sm		4.97		2.13E-06	Y1P Only	
SNF2h/SMARCA5		30.00		2.08E-06	Y1P Only	
Chromatin target of PRMT1	1.00	5.00	7.17E-06	2.05E-06		0.29
MKI67 FHA domain-interacting nucleolar phosphoprotein		8.96		2.02E-06	Y1P Only	
DHX15		28.00		1.95E-06	Y1P Only	
Small nuclear ribonucleoprotein-associated protein B	1.00	6.00	2.23E-06	1.73E-06		0.77
DDX27	1.00	19.00	1.22E-06	1.72E-06		1.41
Histone H3		3.00		1.63E-06	Y1P Only	
ATP-dependent RNA helicase DDX39A		8.88		1.59E-06	Y1P Only	
BAZ1b		40.00		1.59E-06	Y1P Only	
Nucleolar protein 56		11.00		1.55E-06	Y1P Only	
Heterochromatin protein 1-binding protein 3		13.76		1.43E-06	Y1P Only	
Eukaryotic initiation factor 4A-III		8.00		1.29E-06	Y1P Only	
Histone-binding protein RBBP7		6.90		1.29E-06	Y1P Only	
RRP15-like protein		4.00		1.24E-06	Y1P Only	
Transcription termination factor 2		27.00		1.22E-06	Y1P Only	
Ribosome production factor 1		6.00		1.20E-06	Y1P Only	
Notchless protein homolog 1		8.99		1.17E-06	Y1P Only	

RNA-binding protein 34		9.94		1.16E-06	Y1P Only	
SPT5		17.99		1.11E-06	Y1P Only	
Vimentin	1.99	10.93	2.77E-06	1.08E-06		0.39
ATP-dependent RNA helicase DDX18		12.00		1.07E-06	Y1P Only	
Isoform 2 of Histone H1.0		2.00		1.00E-06	Y1P Only	
RNA-binding protein 28		12.97		9.71E-07	Y1P Only	
Nucleolin		15.00		9.59E-07	Y1P Only	
Nucleolar complex protein 3 homolog		12.00		9.16E-07	Y1P Only	
Ribosome biogenesis protein BOP1		13.95		9.15E-07	Y1P Only	
Regulation of nuclear pre-mRNA domain-containing protein 1B		6.00		9.13E-07	Y1P Only	
Heterogeneous nuclear ribonucleoprotein H		6.83		9.02E-07	Y1P Only	
Histone H2A type 1-B/E		1.00		9.02E-07	Y1P Only	
ATP-dependent RNA helicase A		28.00		8.80E-07	Y1P Only	
WD repeat-containing protein 18		6.98		8.57E-07	Y1P Only	
Probable ribosome biogenesis protein RLP24		3.00		8.14E-07	Y1P Only	
SCAF11		12.00		8.06E-07	Y1P Only	
Serine/arginine-rich splicing factor 10	1.00	3.00	9.54E-06	7.87E-07		0.08
Heterogeneous nuclear ribonucleoprotein F		4.00		7.49E-07	Y1P Only	
Nucleolar protein 12		2.00		7.39E-07	Y1P Only	
Myb-binding protein 1A		17.98		7.13E-07	Y1P Only	
Nuclear valosin-containing protein-like		12.74		6.70E-07	Y1P Only	
Periodic tryptophan protein 1 homolog		6.00		6.65E-07	Y1P Only	
Chromodomain-helicase-DNA-binding protein 4		21.92		6.63E-07	Y1P Only	
MAK16 homolog		5.00		6.62E-07	Y1P Only	
DDX21		12.00		6.52E-07	Y1P Only	
RNA-binding protein 8A		1.00		6.29E-07	Y1P Only	
Pescadillo homolog		6.91		6.22E-07	Y1P Only	
Chromatin modification-related protein MEAF6		3.00		5.97E-07	Y1P Only	
Splicing factor 3B subunit 2		10.99		5.79E-07	Y1P Only	
Probable ATP-dependent RNA helicase DDX56		5.98		5.54E-07	Y1P Only	

Surfeit locus protein 6		3.00		5.33E-07	Y1P Only	
H/ACA ribonucleoprotein complex subunit 4		3.97		4.55E-07	Y1P Only	
Regulation of nuclear pre-mRNA domain-containing protein 2		12.87		4.40E-07	Y1P Only	
Poly(rC)-binding protein 1		3.00		4.39E-07	Y1P Only	
116 kDa U5 small nuclear ribonucleoprotein component	1.98	10.96	2.33E-06	4.29E-07		0.18
Proline-, glutamic acid- and leucine-rich protein 1		11.87		4.21E-07	Y1P Only	
WD repeat-containing protein 74		3.00		3.78E-07	Y1P Only	
Metastasis-associated protein MTA2		8.99		3.54E-07	Y1P Only	
NHP2-like protein 1		2.00		3.44E-07	Y1P Only	
Chromobox protein homolog 3		1.98		3.40E-07	Y1P Only	
Ribosomal RNA processing protein 1		1.00		3.38E-07	Y1P Only	
Splicing factor 3B	1.00	12.00	1.36E-06	3.37E-07		0.25
Elongation factor 1-alpha 1		1.99		3.37E-07	Y1P Only	
Suppressor of SWI4 1 homolog		3.98		3.22E-07	Y1P Only	
RNA-binding protein with serine-rich domain 1		2.00		3.20E-07	Y1P Only	
KRR1 small subunit processome component homolog		3.00		3.14E-07	Y1P Only	
Nuclease-sensitive element-binding protein 1		1.98		3.13E-07	Y1P Only	
U5 small nuclear ribonucleoprotein 40 kDa protein		2.00		3.07E-07	Y1P Only	
Splicing factor 3B subunit 3	2.00	7.99	1.56E-06	3.06E-07		0.20
FACT complex subunit SPT16		4.00		2.84E-07	Y1P Only	
Transcription factor AP-2-alpha		2.00		2.81E-07	Y1P Only	
THO complex subunit 4		2.00		2.81E-07	Y1P Only	
RuvB-like 1		6.93		2.76E-07	Y1P Only	
U1 small nuclear ribonucleoprotein 70 kDa		3.96		2.51E-07	Y1P Only	
Protein LLP homolog		2.00		2.48E-07	Y1P Only	
U5 small nuclear ribonucleoprotein 200 kDa helicase		11.00		2.45E-07	Y1P Only	
Nucleolar protein 58		6.00		2.45E-07	Y1P Only	
Ribosome biogenesis protein WDR12		3.00		2.32E-07	Y1P Only	
Nucleolar complex protein 2		3.00		2.21E-07	Y1P Only	
Polynucleotide 5'-hydroxyl-kinase NOL9		4.00		2.19E-07	Y1P Only	

Ribosomal biogenesis protein LAS1L		5.94		2.14E-07	Y1P Only	
Zinc finger MYND domain-containing protein 11		2.97		2.08E-07	Y1P Only	
RuvB-like 2		2.99		2.02E-07	Y1P Only	
NF-kappa-B-repressing factor		4.00		1.99E-07	Y1P Only	
Pre-mRNA-processing-splicing factor 8	3.00	10.00	1.91E-06	1.81E-07		0.09
Probable ATP-dependent RNA helicase DDX10		3.00		1.73E-07	Y1P Only	
Actin-like protein 6A		3.00		1.64E-07	Y1P Only	
Heterogeneous nuclear ribonucleoprotein M		2.00		1.61E-07	Y1P Only	
Bromodomain and WD repeat-containing protein 1		8.97		1.54E-07	Y1P Only	
Heat shock 70 kDa protein 1A/1B		2.98		1.54E-07	Y1P Only	
Lysine-specific demethylase PHF2		3.00		1.52E-07	Y1P Only	
Matrin-3		3.97		1.40E-07	Y1P Only	
Peregrin		4.00		1.37E-07	Y1P Only	
FACT complex subunit SSRP1		3.00		1.32E-07	Y1P Only	
Bromodomain and WD repeat-containing protein 3		3.50		1.29E-07	Y1P Only	
RNA exonuclease 1 homolog		5.00		1.28E-07	Y1P Only	
Heterogeneous nuclear ribonucleoprotein U		2.00		1.28E-07	Y1P Only	
Transcription activator BRG1/SMARCA4		6.99		1.21E-07	Y1P Only	
ATPase family AAA domain-containing protein 2		5.00		1.18E-07	Y1P Only	
Histone deacetylase 2		1.00		1.11E-07	Y1P Only	
Poly [ADP-ribose] polymerase 2		1.00		1.07E-07	Y1P Only	
Cullin-4B		3.00		8.32E-08	Y1P Only	

1.3 Total RNAPII +/- UV irradiation mass spectrometry results

Group

Group 1. NoAb - negative control
Group 2. -UV IP
Group 3. +UV IP
Group 4. Input NoAb
Group 5. Input -UV
Group 6. Input +UV

RNAPII components
Proteins of interest

Results

	SPECTRAL COUNTS						SPECTRAL INDEX (MIC SIn)					
Name	**NoAb**	**-UV IP**	**+UV IP**	**Input NoAb**	**Input -UV**	**Input +UV**	**NoAb**	**-UV IP**	**+UV IP**	**Input NoAb**	**Input - UV**	**Input +UV**
Ig kappa chain V-IV region Len	1.00	31.00	28.00	1.00		1.00	8.99E-06	4.48E-04	6.41E-04	2.38E-07		1.42E-07
RNAPII RPB4 subunit	0.99	28.86	22.89	9.96	10.95	10.95	2.30E-05	3.58E-04	1.80E-04	4.09E-06	3.94E-06	3.58E-06
Histone H4	9.00	15.00	18.00	310.00	341.00	296.00	1.44E-04	6.02E-05	1.40E-04	2.76E-03	2.95E-03	2.72E-03
RNAPI, II and III RPABC1 subunit		52.00	41.00	16.00	15.00	15.00		1.33E-04	1.36E-04	7.07E-06	1.00E-05	7.22E-06
Histone H2B	4.88	12.72	22.48	262.83	309.85	273.65	2.08E-05	3.87E-05	1.24E-04	1.02E-03	1.01E-03	1.46E-03
RNAPI, II and III RPABC3 subunit		16.83	13.86	8.91	9.90	7.92		1.42E-04	1.21E-04	4.70E-06	4.84E-06	4.22E-06
RNAPII RPB3		35.00	33.00	17.00	18.00	16.00		1.20E-04	1.06E-04	4.30E-06	5.10E-06	5.77E-06
RNAPII RPB1		283.00	266.00	120.00	129.00	107.00		9.20E-05	9.86E-05	3.61E-06	3.39E-06	3.28E-06
RNAPII RPB7		30.97	25.97	10.98	14.98	9.98		1.02E-04	8.99E-05	4.47E-06	5.84E-06	3.65E-06
RNAPII RPB9		16.96	18.96	7.98	8.97	9.98		4.53E-04	8.61E-05	3.06E-06	3.27E-06	4.30E-06
RNAPII RPB2		181.23	165.37	58.43	63.37	54.47		1.13E-04	8.59E-05	3.25E-06	2.63E-06	3.21E-06

RNAPII RPB11		10.86	9.88	7.90	7.90	4.94		1.64E-04	7.67E-05	2.18E-06	1.69E-06	3.71E-06
RNAPI, II and III RPABC4 subunit		3.00	4.00	1.00	1.00	1.00		1.76E-05	2.88E-05	1.83E-06	4.89E-06	1.30E-06
RNAPI, II and III RPABC5 subunit		8.00	6.00	4.00	4.00	3.00		2.49E-05	2.74E-05	2.76E-06	2.17E-06	1.94E-06
Regulation of pre-mRNA protein 1A RPRD1A		12.91	12.91	9.94	12.92	8.95		2.13E-05	2.51E-05	2.36E-06	2.78E-06	1.67E-06
Histone H1.5	3.00	7.00	10.00	12.00	10.00	14.00	5.74E-06	5.30E-06	2.37E-05	5.42E-06	2.60E-06	8.89E-06
Serine/arginine-rich splicing factor 7	1.00	5.50	5.50	24.00	31.50	22.50	1.91E-06	1.15E-05	1.44E-05	2.33E-05	2.99E-05	1.80E-05
Regulation of pre-mRNA protein 1B RPRD1B		11.98	12.97	11.99	11.99	8.00		1.18E-05	1.34E-05	1.42E-06	1.53E-06	1.18E-06
RNAPI, II and III RPABC2 subunit		4.98	5.97	1.99	1.99	1.00		1.33E-05	1.34E-05	8.94E-07	1.98E-06	3.46E-07
Regulation of pre-mRNA protein 2 RPRD2	1.99	44.55	39.60	35.63	31.67	19.80	3.23E-07	9.05E-06	1.05E-05	9.17E-07	5.76E-07	3.83E-07
Ig kappa chain V-III region Ti	3.00	4.00	3.00				3.29E-05	7.84E-06	1.01E-05			
Spt6		64.96	59.98	46.99	44.99	28.99		7.78E-06	9.36E-06	9.48E-07	8.97E-07	5.29E-07
PAF1			20.87	0.99	1.99	8.97			8.85E-06	1.43E-08	3.37E-09	5.00E-07
Nucleolar protein 58			3.00	13.00	16.00	15.00			8.65E-06	1.02E-06	1.52E-06	1.55E-06
ERCC6		2.00	49.92			6.97		1.25E-07	8.00E-06			9.82E-08
Eukaryotic translation initiation factor 6	2.00	5.00	5.00	6.00	7.99	5.99	1.91E-06	5.00E-06	7.88E-06	1.21E-06	1.02E-06	1.67E-06
60S ribosomal protein L	1.99	9.93	13.92	16.92	15.92	13.93	1.70E-06	4.01E-06	7.49E-06	1.61E-06	1.90E-06	3.62E-06
Chromatin target of PRMT1 protein	1.00	7.00	8.00	13.00	11.00	9.00	1.44E-06	6.73E-06	7.17E-06	3.15E-06	2.08E-06	4.74E-06
60S ribosomal protein L23	0.99	4.96	5.96	5.96	5.96	6.95	6.27E-07	3.88E-06	7.09E-06	1.64E-06	1.73E-06	2.51E-06
Ribosome biogenesis regulatory protein homolog		5.00	10.00	20.00	20.00	17.00		1.39E-06	6.60E-06	3.68E-06	4.35E-06	4.46E-06
CTR9			30.00			14.00			6.51E-06			2.50E-07
BAZ1b	1.00	25.00	44.00	77.00	82.00	70.00	1.35E-07	2.50E-06	6.35E-06	2.60E-06	2.10E-06	2.68E-06
CDC73			18.00			14.00			5.88E-06			5.89E-07
DDB1		15.77	23.66	30.56	33.52	39.44		1.56E-06	5.52E-06	1.51E-06	1.10E-06	2.30E-06
SNF2h		24.99	33.99	80.98	79.98	62.98		3.60E-06	5.34E-06	3.73E-06	3.69E-06	3.47E-06
RNA-binding protein Raly	0.99	5.90	8.85	42.51	44.49	33.55	3.67E-07	3.39E-06	5.24E-06	1.30E-05	1.23E-05	1.33E-05
Ribosomal L1 domain-containing protein 1		6.91	12.84	30.65	29.67	25.70		1.46E-06	5.22E-06	4.07E-06	3.10E-06	5.68E-06
WDR61			9.87			3.96			5.17E-06			1.78E-07
60S ribosomal protein L6		7.97	6.97	9.97	7.98	10.97		4.81E-06	5.09E-06	1.98E-06	2.02E-06	3.24E-06
Serine/arginine repetitive matrix protein 2		43.83	40.83	58.81	62.77	47.84		4.62E-06	4.93E-06	1.41E-06	1.29E-06	1.27E-06

RNA-binding protein 8A		3.00	4.00	31.00	39.00	25.00		4.38E-06	4.84E-06	1.57E-05	1.70E-05	1.24E-05
Pre-mRNA-processing factor 19		11.99	8.98	32.94	36.93	23.95		6.23E-06	3.96E-06	5.59E-06	5.91E-06	3.63E-06
Splicing factor, arginine/serine-rich 19		15.50	16.50	13.50	13.00	13.00		2.15E-06	3.94E-06	4.34E-07	2.43E-07	3.68E-07
SCAF11	1.00	45.00	31.00	14.00	15.00	9.00	5.01E-08	6.33E-06	3.92E-06	2.46E-07	2.54E-07	1.19E-07
Initiation factor 4A-III		10.00	9.00	59.00	55.00	42.00		4.15E-06	3.58E-06	1.42E-05	1.41E-05	1.09E-05
Putative histone H2B type 2-C		1.96	1.96	26.58	30.60	21.64		1.43E-06	3.40E-06	3.54E-05	2.80E-05	3.59E-05
Glioma tumour suppressor GLTSCR2	1.98	2.98	7.97	11.96	6.96	9.95	1.71E-06	7.18E-07	3.27E-06	1.12E-06	4.99E-07	9.93E-07
Small nuclear ribonucleoprotein Sm D3	1.00	3.00	3.00	22.00	31.00	9.00	8.85E-08	3.40E-06	3.25E-06	1.66E-05	1.62E-05	4.93E-06
60S ribosomal protein L7a		4.99	7.99	15.97	15.97	11.97		2.41E-06	3.20E-06	1.57E-06	1.51E-06	1.73E-06
Serine/arginine-rich splicing factor 9		4.98	2.99	28.88	29.88	21.91		4.54E-06	3.10E-06	1.25E-05	9.89E-06	9.37E-06
Chromobox protein homolog 5	1.00	3.00	6.00	21.50	24.00	15.50	1.33E-06	1.50E-06	2.64E-06	8.44E-06	6.91E-06	4.28E-06
Histone H1.2	1.00	1.00	1.00	3.00	3.00	4.00	3.01E-06	2.11E-06	2.49E-06	1.05E-06	3.19E-07	1.08E-06
U5 small nuclear ribonucleoprotein 40		7.97	4.98	27.96	31.94	17.96		3.58E-06	2.35E-06	7.08E-06	6.21E-06	4.95E-06
ERCC8			5.96			0.99			2.34E-06			2.36E-08
Guanine nucleotide-binding protein-like 3		5.00	8.98	6.98	6.98	9.97		6.02E-07	2.31E-06	2.52E-07	3.15E-07	6.94E-07
LEO1			7.00	1.00		2.00			2.24E-06	8.89E-09		7.83E-08
Nucleolar GTP-binding protein 1		5.94	6.94	16.85	22.79	23.78		1.45E-06	2.20E-06	7.89E-07	9.97E-07	1.34E-06
Nucleolar protein 16			2.99	7.98	8.98	7.98			2.11E-06	1.51E-06	1.63E-06	2.32E-06
Heterogeneous nuclear ribonucleoprotein H	1.94	2.89	2.89	88.72	107.20	59.34	2.82E-06	1.28E-06	1.90E-06	1.11E-05	8.90E-06	5.81E-06
DDX54		2.00	8.00	15.50	20.50	19.50		2.59E-07	1.74E-06	4.58E-07	5.83E-07	9.16E-07
Small nuclear ribonucleoprotein Sm D2		2.98	1.99	35.76	32.77	19.87		1.33E-06	1.67E-06	3.59E-05	3.82E-05	2.17E-05
pre-rRNA processing protein FTSJ3		6.00	11.00	16.00	19.00	19.00		3.99E-07	1.63E-06	4.81E-07	4.36E-07	9.53E-07
Histone macro-H2A.1 H2AFY		1.97	3.96	60.44	76.30	61.47		7.36E-07	1.43E-06	2.55E-05	2.83E-05	3.59E-05
Thyroid hormone receptor-associated protein 3		7.00	7.00	23.00	20.00	15.00		1.07E-06	1.07E-06	6.96E-07	5.51E-07	4.27E-07
Pre-mRNA-processing-splicing factor 8	3.00	32.99	21.00	140.99	141.99	118.99	9.35E-08	1.54E-06	1.02E-06	2.98E-06	3.00E-06	3.27E-06
Pescadillo homolog			4.97	22.86	23.86	21.87			8.21E-07	1.37E-06	1.07E-06	2.06E-06
PHIP		15.00	10.00	18.99	21.00	13.99		8.10E-07	7.55E-07	3.41E-07	3.47E-07	2.97E-07
PHD and RING finger domain-containing protein 1		13.50	6.50	3.50	2.00			1.39E-06	7.47E-07	2.96E-08	2.24E-08	
Nucleolar complex protein 3 homolog			5.00	11.00	12.00	16.00			7.00E-07	4.18E-07	3.81E-07	8.61E-07

U5 small nuclear ribonucleoprotein 200 kDa	3.00	16.00	13.00	131.99	149.99	99.00	2.55E-07	8.07E-07	6.84E-07	3.89E-06	3.85E-06	2.92E-06
Apoptotic chromatin condensation inducer ACIN1		6.89	6.87	27.62	24.68	8.86		4.75E-07	6.81E-07	6.20E-07	4.96E-07	2.79E-07
Splicing factor 3B subunit 3		1.00	3.99	93.89	108.87	72.93		3.14E-08	6.09E-07	4.79E-06	6.57E-06	4.98E-06
Serine/arginine-rich splicing factor 1		0.99	0.99	60.39	61.38	47.52		4.76E-07	5.83E-07	4.27E-05	3.39E-05	3.56E-05
Nucleolar protein 56			4.98	16.95	15.95	15.95			4.67E-07	1.11E-06	9.89E-07	1.34E-06
Heterochromatin protein 1-binding protein 3		0.98	1.96	15.73	15.73	17.68		1.01E-07	4.32E-07	1.47E-06	1.30E-06	2.07E-06
Pumilio domain-containing protein KIAA0020		1.00	4.99	40.99	45.00	35.99		5.58E-08	4.25E-07	2.54E-06	2.71E-06	2.54E-06
116 kDa U5 small nuclear ribonucleoprotein	1.00	9.95	2.98	75.79	80.78	53.83	1.35E-07	1.95E-06	4.10E-07	5.59E-06	4.75E-06	4.30E-06
Isoform 2 of Serine/arginine-rich splicing factor 3		2.50	0.50	32.00	33.50	21.50		2.28E-06	3.92E-07	3.13E-05	3.71E-05	2.62E-05
DDX27			4.00	15.00	12.00	17.00			3.77E-07	4.37E-07	3.91E-07	6.77E-07
DDX21			4.00	26.00	27.00	27.00			3.43E-07	1.12E-06	1.03E-06	1.35E-06
Splicing factor 3B subunit 1		5.00	4.00	112.98	119.98	80.99		6.63E-07	3.01E-07	6.68E-06	6.15E-06	6.03E-06
RNA-binding protein 34			1.97	4.97	5.97	7.95			2.67E-07	3.40E-07	3.85E-07	5.55E-07
Splicing factor 3B subunit 2	0.98	7.89	2.97	80.28	93.17	66.42	2.60E-07	8.88E-07	2.40E-07	6.92E-06	6.84E-06	7.32E-06
RNA-binding protein 28			2.00	8.97	12.98	10.96			1.84E-07	1.98E-07	2.74E-07	2.94E-07
XRN2		4.00	2.00	24.00	26.00	18.00		9.58E-07	1.75E-07	1.22E-06	9.62E-07	8.01E-07
DHX15	1.00	6.00	1.00	41.00	50.00	32.00	5.79E-09	6.24E-07	1.46E-07	3.88E-06	3.97E-06	2.02E-06
Isoform 2 of PHD finger protein 3		13.90	7.94	2.95	1.97			4.71E-07	1.40E-07	9.15E-09	4.92E-09	
Transcription termination factor 2		4.00	2.00	1.00				4.10E-07	1.34E-07	6.56E-09		
Structural maintenance of chromosomes protein 3			3.00	82.00	92.00	77.00			1.14E-07	4.68E-06	4.73E-06	4.19E-06
Cap-specific mRNA methyltransferase 1		8.00			1.00			6.54E-07			6.42E-09	

Appendix 2. Sequencing results

TACCTAGTGGAGACAAAAGTACATACCTCGGTAGAGCGCGCACGCCTGTAACCCCAGCAC
TTTGGGAGGCCAAGGTGGGTGTATCACCTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAA
CATGGTGAAACTCCGTCTCTACTAAAATTACAAAAATCAGCCAGGCTTCATGGCACATGC
CTATAGTCCCAGCTACAGGCATGCTGAAGCAGGAGAATCGCTTGCACCCCGGAGGCAGAG
GCTGCAGTGAGCTGAGACCACACCACTGCACTCCAGCCTAGGCAACAGAGTATGAGACTC
CATCTCAAAAAAAAAAAAAGTACCTACCTCAGAGTTCAAACTAGTGAATATTAGGAAGTG
CTTGAGACAGTGACACCAAAGTGCACAATAAATACTCGCCAGTTTCATTATTATTAAAGA
ATCCATTTGAATGTCAGCTCAACACAGCCTCCTATACCGAGGCATTGTGAACCGCATCTC
CCCAGCTTCTCCAGGCTTTTCCAAGAATCAGGGACACTGTAGCCTGTTGGTCTCAGTGTA
TGACAGACACGGAGGAAGCACATCTTTAGCTGATACTTAAACAGAGACCCTGAGCGCACA
TACACCCGCGCACACATGCATGGAGCTTCACCTTCTCTGTCATTCTGCAGTGACCAGGAG
AGCAAGAGCTCCCACCTCCCTTCAAAACACTGTGCCCATCCCGGGCACTAAGGCCTCTTT
AAAGCACGGCACCTCCACGAGGGAGGGCCACAGCCACATACACTCCACCTGGCAGGTGGA
CAGCGTGAGCACGTGGACCATAGCAGGGACAAGGTGCCCCGGCCAGCCCCAACGCCCTCT
GCCGCTGACAGGGACAGAAGCCCTCTCCAGCTGCGTGTGCTGCAGAGGCCATGCGTAGCC
TCCAGCTGCATTCTATTCCACTCCAGTGCCTGGGCCAGTTAGCACCAGTGTGGAAGACAG
TGAGCTGGCTCCGGACAACAGGGATGGAGGAAAGGTCCCACATTCACATTCCTGATACGT
GGACAAGGTGAGGGGCCGCAATCGCTCTGGCAGCATTTTAAAGATGGGGAAGTAGCAGAC
ACCCACGCGTGAAGGCAGGAGAGCCCCAACTGTGGTGGAAATGGCCCCAGAATGGTAGGG
CCAAGCCTAGCTCCAGACACCCCAGAGCCCTGGAGAAGCCAAGACTGAGGGAGAAAGCCT
GAGGGAGGAGCGCCCCAGTCCCCAGGGACCGGCCTGGTGCAGAGCTGCAGCTGATGTTCC
CCTCTGTGCAGCCCCACCCTCTGCCTCGCTGAGCTCCCTGCTGCGAGGGCCTCGGGTGCA
AGGGGGAGGCAGGTCTCTATCTCATGGAGCTGTCAGATGAGACATCGCGATCGGAGTCCT
CAGCCTCGCTTGGCGGCGGCGGCGGGTCGCTAAGCGGGACCGCAGTGAAAGCAGGAGACT
TTCTAGAAAAAAACACCAGTTGTCAACCTTGGGGCAGGCAGGAATCCTGAAGACGGACGG
CACTCCTCCTCCTGCTGCCTCACCCTCTGGCAGCCCGTGAGAAGTACCGGAAGCGAGGGC
GGGGCCGCGGGATGGCGAGGGAGCGGCAGGGACTGAACTCTCTCCAAACCCACCCTGACA
GGGAAATGGGCCCCGCCTGTGTCTTGGGAACTCAGAGGCTGAGGTCAGGCATGATGGCTC
ACGCCTGTAATTCCAGCACTTTGGGAGGCAGAGGCGGGTGGATCACGACGTCAGGAGTTC
AAGGCAAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATGCAAAAATTAGCCA
GGTGTGGTGGCGGGCGCCTTGGAGGCTGAGGCAAGATAATCACTTGAACCTGGGAGGCGG
AGGTTGTAGTGAGCCAAAAAAAAAAAAAAAAAGAAATAGCTGAAGTCACAGTAGGAGAGAA
GCTGCTGAGCCTCCAGCACCCTGACTCTAGGGCCTTGGCTTTATGTCTATCTGCAGTATT
TTTGTGATTTTTAAAAATTCACTTTCTTGTTGCGGTGTAACTTACACAGGGTCAAATGCA
CAAATCATGGCCCTGACTTTAGATAAAAATCTGCCCCCACAACCCTTCTGTTCCTTGCCA
GTTTTTAAACTGCCTCTAACCAGGGGAACCACCAGAGCTGGTGGCCTTGGGAGGTTTCAG
CCCTCCCGTCATGAATGGACATAGCTCATCCAACTGCCAAGGGAGAGAGCTGTGGGTCTG
GGCCAGCCCCACCAGGTAACTCCCAAAGGGCAGCCCCACAGCAAGATGTGACCCAGTCAT

TACCTAGTGGAGACAAAAGTACATACCTCGGTAGAGCGCGCACGCCTGTAACCCCAGCAC
TTTGGGAGGCCAAGGTGGGTGTATCACCTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAA
CATGGTGAAACTCCGTCTCTACTAAAATTACAAAAATCAGCCAGGCTTCATGGCACATGC
CTATAGTCCCAGCTACAGGCATGCTGAAGCAGGAGAATCGCTTGCACCCCGGAGGCAGAG
GCTGCAGTGAGCTGAGACCACACCACTGCACTCCAGCCTAGGCAACAGAGTATGAGACTC
CATCTCAAAAAAAAAAAAAGTACCTACCTCAGAGTTCAAACTAGTGAATATTAGGAAGTG
CTTGAGACAGTGACACCAAAGTGCACAATAAATACTCGCCAGTTTCATTATTATTAAAGA
ATCCATTTGAATGTCAGCTCAACACAGCCTCCTATATCTAGGGCCTTGGCTTTATGTCTA
TCTGCAGTATTTTTGTGATTTTTAAAAATTCACTTTCTTGTTGCGGTGTAACTTACACAG
GGTCAAATGCACAAATCATGGCCCTGACTTTAGATAAAAATCTGCCCCCACAACCCTTCT
GTTCCTTGCCAGTTTTTAAACTGCCTCTAACCAGGGGAACCACCAGAGCTGGTGGCCTTG
GGAGGTTTCAGCCCTCCCGTCATGAATGGACATAGCTCATCCAACTGCCAAGGGAGAGAG
CTGTGGGTCTGGGCCAGCCCCACCAGGTAACTCCCAAAGGGCAGCCCCACAGCAAGATGT
GACCCAGTCAT

Appendix 2: Sequence of the *GAPDH* termination region (top) showing target deletion region (cyan) and deletion strain sequence results (green). Sequence of deletion mutant only (bottom), showing target region removed.