TERMINATION OF RNA POLYMERASE II TRANSCRIPTION

BY THE 5’-3’ EXONUCLEASE XRN2

by

MICHAEL ANDRES CORTAZAR OSORIO

B.S., Universidad del Valle – Colombia, 2011

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Molecular Biology Program

2018
This thesis for the Doctor of Philosophy degree by
Michael Andrés Cortázar Osorio
has been approved for the
Molecular Biology Program
by

Mair Churchill, Chair
Richard Davis
Jay Hesselberth
Thomas Blumenthal
James Goodrich
David Bentley, Advisor

Date: Aug 17, 2018
Cortázar Osorio, Michael Andrés (Ph.D., Molecular Biology)

Termination of RNA polymerase II transcription by the 5’-3’ exonuclease Xrn2

Thesis directed by Professor David L. Bentley

**ABSTRACT**

Termination of transcription occurs when RNA polymerase (pol) II dissociates from the DNA template and releases a newly-made mRNA molecule. Interestingly, an active debate fueled by conflicting reports over the last three decades is still open on which of the two main models of termination of RNA polymerase II transcription does in fact operate at 3’ ends of genes. The torpedo model indicates that the 5’-3’ exonuclease Xrn2 targets the nascent transcript for degradation after cleavage at the polyA site and chases pol II for termination. In contrast, the allosteric model asserts that transcription through the polyA signal induces a conformational change of the elongation complex and converts it into a termination-competent complex. In this thesis, I propose a unified allosteric-torpedo mechanism.

Consistent with a polyA site-dependent conformational change of the elongation complex, I found that pol II transitions at the polyA site into a mode of slow transcription elongation that is accompanied by loss of Spt5 phosphorylation in the elongation complex. Inhibition of polyA signal recognition by expression of the flu virus NS1A protein or CRISPR mediated mutation of a consensus polyA signal further revealed that binding of 3’-end processing factors to the polyA signal on the RNA transcript is required for both deceleration of pol II and loss of Spt5 phosphorylation downstream of polyA sites. Consistent with the torpedo model, I report that the 5’ PO₄ RNA end generated at the polyA cleavage site accumulates to high levels when the exonuclease activity of Xrn2 is inhibited. The primary product of polyA site cleavage is therefore directly degraded by Xrn2. Furthermore, inhibition of Xrn2 resulted in accumulation of
polymerases downstream of polyA sites that fail to dissociate from the DNA template after a switch to slow elongation. I propose an allosteric-torpedo model in which polyA site-dependent transcriptional deceleration by 3’-end processing factors permits the Xrn2 torpedo to catch pol II and dismantle the elongation complex.

Additionally, I found that premature termination of transcription is a common phenomenon on human promoters and that Xrn2 participates in premature termination of a fraction of polymerases that can elongate into gene bodies without licensing by pTEF-b. Consistent with a role of Xrn2 in premature termination, I found that Xrn2 is recruited to gene promoters and migrates to 3’ ends of genes associated with the elongation complex. Finally, I investigated R-Loop structures genome-wide and provide evidence that mammalian cells possess a protective mechanism against accumulation of R-Loops on pol II transcribed genes to prevent their potential threat to genomic instability.

The form and content of this abstract are approved. I recommend its publication.

Approved: David L. Bentley
GRAPHICAL ABSTRACT

TERMINATION OF RNA POLYMERASE II TRANSCRIPTION

BY THE 5'-3' EXONUCLEASE XRN2
# TABLE OF CONTENTS

## CHAPTER

### I. INTRODUCTION

- 1.1 Transcription by the Different Eukaryotic RNA Polymerases ................................................. 1
- 1.2 Transcription by RNA Polymerase II .......................................................................................... 6
- 1.3 Promoter-Proximal Pausing of Pol II .......................................................................................... 6
- 1.4 The Elongation Factor Spt5 ................................................................................................. 13
- 1.5 mRNA 3’-end Processing ......................................................................................................... 15
- 1.6 The 5’-3’ Exonuclease Xrn2 .................................................................................................... 18
- 1.7 Termination of Pol II Transcription .......................................................................................... 21
- 1.8 The Transcriptional Response to Heat Shock .............................................................................. 29
- 1.9 R-Loop Structures During Transcription by RNA pol II............................................................ 31

### II. MATERIALS AND METHODS

- 2.1 Chromatin Immunoprecipitation Coupled to Sequencing (ChIP-seq) .......................... 37
  - 2.1A Chromatin Immunoprecipitation (ChIP) ...................................................................... 37
  - 2.1B ChIP-seq and Mapping ................................................................................................. 38
- 2.2 Antibodies .............................................................................................................................. 39
- 2.3 Bromouridine Sequencing (Bruseq) ....................................................................................... 39
  - 2.3A Isolation of Nascent Transcripts ...................................................................... 39
  - 2.3B Bruseq and Mapping ...................................................................................................... 40
2.4 5’-PO₄ Bromouridine Sequencing (5’-Bruseq) .......................................................... 41
2.5 DNA-RNA Immunoprecipitation Coupled to Sequencing (DRIP-seq) ................. 44
2.6 Engineering of pcDNA5-NS1A Inducible Flp-in TREX Cells ................................. 44
2.7 CRISPR-Cas9 Genome Editing of the PSMG1 polyA Signal in Hap1 Cells .......... 45
2.8 Doxycycline-Inducible shRNA Expression of Xrn2 and TTF2 in HEK293 Cells .... 48
2.9 Cell Lines Growth Conditions .............................................................................. 51
2.10 Real-Time quantitative (q) PCR ........................................................................... 51

III. INTEGRATION OF XRN2 INTO THE TRANSCRIPTION CYCLE OF RNA POLYMERASE II ........................................................................................................... 54

3.1 Introduction .............................................................................................................. 54
3.2 Results ...................................................................................................................... 58

3.2A Recruitment of Xrn2 Around Promoters does not Require Synthesis of a pre-
mRNA Transcript ....................................................................................................... 58
3.2B Xrn2 Associates with Promoter-Proximally Paused Elongation Complexes ... 60
3.2C Xrn2 Accompanies the Elongation Complex from Promoters to 3’ Ends of Genes ......................................................................................................................... 63
3.2D Xrn2 Degrades Nascent Transcripts Downstream and Upstream of polyA Sites ................................................................................................................................. 72
3.2E 5’-Bruseq Detects 5’-Monophosphorylated Ends of Nascent Transcripts ....... 77
3.2F Xrn2 Recognizes the 5’ PO₄ End of Nascent Transcripts at polyA Sites for Degradation ...................................................................................................................... 83
3.2G Xrn2 Degrades the Nascent Transcript Downstream of the pre-rRNA 3’ End 87
3.2H Xrn2 Targets the 5’ End of snRNAs .................................................................... 90
7.3 Global Resolution of R-Loops in Response to Inhibition of DNA Topoisomerase I 239

REFERENCES...........................................................................................................259

APPENDIX

A. CHAPTER III EXTENDED FIGURE ............................................................................. 270

B. CHAPTER IV EXTENDED FIGURE ............................................................................. 271

C. DOUBLE STABLE EXPRESSION OF shRNAs AGAINST XRN2 AND TTF2
   SUGGESTS A ROLE OF TTF2 IN TERMINATION OF POL II TRANSCRIPTION......272

D. GENE LISTS ............................................................................................................. 285
LIST OF FIGURES

1-1. The transcription cycle of RNA polymerase II ............................................................ 7
1-2. The torpedo model of RNA polymerase II transcription termination ......................... 12
1-3. The allosteric model of RNA polymerase II transcription termination ....................... 25
1-4. Activation of the heat shock response by HSF1 ........................................................ 30
1-5. The R-Loop structure ............................................................................................... 32
2-1. Overview of the 5’-Bruseq procedure ....................................................................... 42
2-2. Expression of the flu virus NS1A protein causes cell death within 72 hours of induction... 46
2-3. Schematic diagram of the CRISPR-Cas9 genome editing strategy ............................... 47
2-4. Sequence alignment reveals mutation of the PSMG1 polyA signal .............................. 49
2-5. BamH I digestion confirms successful CRISPR-Cas9 editing around the polyA site of the PSMG1 gene in Hap1 cells .................................................................................. 50
2-6. Inducible knockdown of Xrn2 and TTF2 in HEK293 cells ......................................... 53
3-1. Xrn2 is recruited to the transcription start site of the RPL23 gene .............................. 59
3-2. Xrn2 is recruited to the transcription start site of gene promoters genome-wide .......... 61
3-3. Xrn2 is recruited to the transcription start site of gene promoters genome-wide – Biological replicate ..................................................................................................................... 62
3-4. Camptothecin causes pile-up of polymerases on the body region of the GAPDH gene and reduces pol II occupancy around its promoter ..................................................... 64
3-5. Pol II and Xrn2 accumulate on the body region of the EIF1 and ACTB genes after inhibition of transcription elongation with camptothecin ............................................. 66
3-6. Camptothecin inhibits transcription elongation genome-wide .................................... 67
3-7. Camptothecin induces global accumulation of Xrn2 on gene bodies ................. 68
3-8. Xrn2 accompanies the elongation complex from the promoter region to the 3’ end of the NEAT1 gene ............................................................................................... 69
3-9. Xrn2 accompanies the elongation complex from the promoter region to the 3’ end of the EIF1 gene.

3-10. Xrn2 accompanies the elongation complex from promoters to 3’ ends of genes genome-wide.

3-11. Xrn2 targets the nascent transcript for degradation downstream of the polyA site of the EIF1, CYR61 and JUN genes.

3-12. Xrn2 targets the nascent transcript for degradation downstream of the 3’ end of histone genes.

3-13. Xrn2 targets nascent transcripts for degradation around promoters and 3’ ends of genes.

3-14. 5’-Bruseq detects 5’ PO₄ ends of tRNAs and the 5’ PO₄ product of endonucleolytic cleavage by RNase Z at their 3’ end.

3-15. Xrn2 targets 5’ PO₄s ends of nascent transcripts at the 5’ and 3’ end of the pre-rRNA transcription unit.

3-16. Xrn2 targets 5’ PO₄s ends of nascent transcripts at the 5’ and 3’ end of the pre-rRNA transcription unit – Biological Replicate.

3-17. Xrn2 targets 5’ PO₄s of nascent transcripts generated via different pathways.

3-18. Xrn2 trims the 5’ leaders of intronic snoRNAs SNORD15A and SNORD11 after intron debranching.

3-19. Xrn2 targets 5’ PO₄s ends of nascent transcripts at polyA sites for degradation.

3-20. Xrn2 targets 5’ PO₄s ends of nascent transcripts at polyA sites for degradation – Biological Replicate.

3-21. Inhibition of Xrn2 differentially affects 5’ PO₄s of nascent transcripts at histones, snoRNA, snRNA and miRNA genes.

3-22. Inhibition of Xrn2 affects 5’ PO₄s of nascent transcripts at histones and snoRNA genes – Biological Replicate.

3-23. Xrn2 targets nascent transcripts at the 5’ and 3’ end of the pre-rRNA transcription unit.

3-24. Xrn2 targets nascent transcripts at the 5’ and 3’ end of the pre-rRNA transcription unit – Biological Replicate.

3-25. Model for integration of Xrn2 into the transcription process and termination of pol II transcription.
4-1. Inhibition of Xrn2 generates a pol II termination defect on the ACTB gene ..................... 107
4-2. Xrn2 is required for release of pol II from the DNA template at the 3’ end of the ACTB gene ............................................................................................................................... 108
4-3. Xrn2 is required for release of pol II from the DNA template at the 3’ end of the EIF1 gene ....................................................................................................................................... 109
4-4. Xrn2 is required for efficient release of pol II from the DNA template downstream of polyA sites genome-wide .......................................................................................................... 111
4-5. Xrn2 is required for efficient release of pol II from the DNA template downstream of polyA sites genome-wide – Biological Replicate ........................................................................... 112
4-6. Inhibition of Xrn2 causes a stronger pile-up of polymerases at the 3’ end of the ACTB gene ....................................................................................................................................... 113
4-7. Inhibition of Xrn2 causes a stronger pile-up of polymerases at the 3’ end of the EIF1 gene ....................................................................................................................................... 114
4-8. Estimated rates of transcription elongation in Xrn2WT and Xrn2MT cells ...................... 115
4-9. Pol II slows down transcription elongation downstream of polyA sites ..................... 118
4-10. Inhibition of Xrn2 generates stronger accumulation of polymerases downstream of polyA sites – Biological replicate .......................................................................................................... 121
4-11. Recognition of the polyA signal is required for slowdown of pol II downstream of the ACTB and EIF1 genes ........................................................................................................... 122
4-12. Recognition of the polyA signal is required for slowdown of pol II downstream of polyA sites ........................................................................................................................................... 123
4-13. The polyA signal at the 3’ end of the PSMG1 gene is required for early build-up of pol II density at its 3’ end .................................................................................................................. 124
4-14. Spt5 accumulates at the 3’ end of the ACTB and MYC genes in a hypo-phosphorylated state ........................................................................................................................................... 128
4-15. Spt5 accumulates at 3’ ends of genes in a hypo-phosphorylated state ..................... 129
4-16. Spt5 phosphorylation increases upstream and downstream of the transcription start site of genes ........................................................................................................................................... 130
4-17. Elongation complexes accumulate at the 3’ end of the ACTB gene with hypo-phosphorylated Spt5 .............................. 131

4-18. Elongation complexes accumulate at 3’ ends of genes with hypo-phosphorylated Spt5 . 132

4-19. Recognition of the polyA signal is required for loss of Spt5 phosphorylation downstream of the polyA site of the ACTB gene ............................................................ 135

4-20. Recognition of the polyA signal is required for loss of Spt5 phosphorylation downstream of the polyA sites .......................................................... 136

4-21. The polyA signal at the 3’ end of the PSMG1 gene is required for early build-up of pol II and total Spt5 at its 3’ end ................................................................................ 137

4-22. The allosteric-torpedo model of RNA pol II transcription termination ................. 145

5-1. Xrn2 targets pol II around promoter regions ............................................................. 154

5-2. Xrn2 targets pol II within the body region and downstream of the RPL23 and ACTB genes ............................................................................................................. 155

5-3. Xrn2 targets pol II on histone genes ........................................................................ 156

5-4. Xrn2 targets pol II throughout the length and downstream of human genes ........ 157

5-5. Xrn2 targets pol II throughout the length and downstream of human genes – Biological replicate ......................................................................................... 158

5-6. Xrn2WT and Xrn2MT cell lines display similar genomic mitochondrial DNA content.... 161

5-7. Xrn2 targets nascent transcripts at the promoter region of the CYR61 and GAPDH genes ........................................................................................................ 162

5-8. Xrn2 limits sense and antisense transcription from human promoters ........................ 163

5-9. Promoter-proximally paused polymerases are highly turned over ......................... 164

5-10. Xrn2 limits escape of promoter-proximally paused polymerases into the body region of the GAPDH gene ................................................................................. 165

5-11. Xrn2 limits escape of promoter-proximally paused polymerases into the body region of the XIST gene .................................................................................. 166

5-12. Xrn2 targets promoter-proximally paused polymerases ........................................ 167
5-13. Xrn2 limits escape of promoter-proximally paused polymerases into the body region of the ACTB gene (N. M12) ................................................................................................... 171

5-14. Xrn2 limits escape of promoter-proximally paused polymerases into the body region of the GAPDH gene (N. M12) ................................................................................................ 172

5-15. Xrn2 limits escape of promoter-proximally paused polymerases into the histone gene HIST4H4 (N. M12) ........................................................................................................... 173

5-16. Xrn2 targets promoter-proximally paused polymerases (N. M12) .................................. 174

5-17. Xrn2 targets polymerases creeping into the body region of the ACTB gene for premature termination of transcription ........................................................................................... 175

5-18. Xrn2 targets polymerases creeping into the body region of the GAPDH gene for premature termination of transcription ........................................................................................... 176

5-19. Xrn2 targets polymerases creeping from promoter regions of histone genes for premature termination .................................................................................................................... 177

5-20. Xrn2 targets polymerases creeping from promoter regions into gene bodies genome-wide ............................................................................................................................. 178

5-21. NELFa density does not migrate with creeping polymerases into the body region of the RPL23 gene .................................................................................................................. 180

5-22. NELFa density does not migrate with creeping polymerases into the body region of the GAPDH gene ................................................................................................................ 181

5-23. Global NELFa density does not migrate with creeping polymerases into gene bodies .... 182

5-24. The HSP90AA1 gene displays a termination defect after inhibition of Xrn2 ............... 185

5-25. Inhibition of Xrn2 generates stronger accumulation of pol II within the body region of the HSP90AA1 gene after heat shock .................................................................................. 186

5-26. Inhibition of Xrn2 generates stronger accumulation of pol II within the body region of the DNAJB1 gene after heat shock .................................................................................. 187

5-27. Inhibition of Xrn2 generates stronger accumulation of pol II within the body region of the HSPH1 gene after heat shock .................................................................................. 188

5-28. Promoter-proximal pausing of pol II is maintained under heat shock conditions in Xrn2WT and Xrn2MT cells ........................................................................................................... 189
5-29. Inhibition of Xrn2 causes stronger accumulation of pol II within the body region of heat shock genes upon heat shock ................................................................. 190

5-30. Xrn2 is required for upregulation of nascent transcript synthesis on heat shock genes ... 191

5-31. Heat shock induces upregulation of nascent transcript synthesis on the DNAJB1 gene .. 192

5-32. Heat shock induces upregulation of nascent transcript synthesis on the HSP90AA1 gene .................................................................................................................. 193

5-33. Heat shock induces upregulation of nascent transcript synthesis on the HSPH1 gene .... 194

5-34. Heat shock causes downregulation of nascent transcript synthesis on the CYR61 gene . 195

5-35. Heat shock causes downregulation of nascent transcript synthesis on the EIF1 gene..... 196

5-36. Xrn2 is required for proper upregulation of nascent transcript synthesis on heat shock genes upon heat shock ................................................................................................. 197

5-37. Inhibition of Xrn2 does not prevent global repression of nascent transcript synthesis .... 198

5-38. Model of premature termination of transcription by Xrn2 ........................................... 205

6-1. Distribution profile of R-Loops on the GAPDH gene ...................................................... 210

6-2. Global R-Loop distribution on most active genes ......................................................... 211

6-3. Comparison of DRIP-seq distribution profiles with previously reported data.............. 212

6-4. Upstream R-Loop formation coincides with divergent nascent transcript signal from the TSS ............................................................................................................................... 213

6-5. Inhibition of Xrn2 increases R-Loop formation within termination zones relative to gene bodies ................................................................. 215

6-6. Meta-gene plot reveals global increase in R-Loop formation downstream of the polyA site after inhibition of Xrn2 .......................................................... 216

6-7. Camptothecin treatment induces R-Loop formation on the rDNA repeating unit ....... 217

6-8. Camptothecin treatment eliminates R-loops on the GAPDH gene .............................. 218

6-9. Camptothecin treatment eliminates R-Loop formation in both Xrn2WT and Xrn2MT cells on the RPL13 gene .................................................................................................................. 219
6-10. Camptothecin treatment eliminates R-Loop formation in both Xrn2WT and Xrn2MT cells on the GAPDH gene .................................................................................................................................220

6-11. Camptothecin treatment generates loss of R-Loops genome-wide ........................................221

6-12. Camptothecin treatment does not affect the global distribution profile of R-Loops in the mitochondrial genome ..............................................................................................................224

6-13. Camptothecin treatment increases nascent transcript synthesis at least 1kb into the body region of the GAPDH gene from its promoter ..................................................................................225

A-1. Xrn2 and pol II ChIP-seq densities on 33 to 78 kb long genes -Related to figure 3-10 .... 241

B-1. Recruitment of Spt5 to promoter regions requires RNA synthesis by RNA pol II .............. 243

C-1. Double knockdown of Xrn2 and TTF2 reveals a mild termination defect on the RPL23 and HSP90AA1 genes ...............................................................................................................273

C-2. Double knockdown of Xrn2 and TTF2 reveals a mild termination defect genome-wide .... 274

C-3. Pol II ChIP-seq distribution profiles after inducible knockdown of Xrn2 or TTF2 ............ 279

C-4. Exposure to CdSO$_4$ induces transcriptional activation of the MT2A gene ..................... 280

C-5. Exposure to CdSO$_4$ generates a strong termination defect on the HSP90AA1 gene after double knockdown of Xrn2 and TTF2 .....................................................................................281

C-6. Exposure to CdSO$_4$ generates a strong termination defect on the DNAJB1 gene after double knockdown of Xrn2 and TTF2 ..................................................................................282

C-7. Exposure to CdSO$_4$ does not generate detectable changes in the distribution profile of pol II genome-wide ......................................................................................................................283

C-8. Double knockdown of Xrn2 and TTF2 increases the expression levels of heat shock protein genes .................................................................................................................................284
CHAPTER I

INTRODUCTION

1.1 Transcription by the Different Eukaryotic RNA Polymerases

Expression of the genetic information in eukaryotic cells is a complex process that requires three distinct RNA polymerases. These enzymes transfer the information contained in the DNA into an RNA molecule. This process, namely transcription, is tightly regulated and its misregulation can lead to human disease (Lee and Young 2013). The most fundamental aspects of this process are common to the three different RNA polymerases. Transcription factors (TFs) are required for RNA polymerases to bind promoter regions of genes and be assembled into a pre-initiation complex. During assembly of the transcription machinery, melting of the DNA duplex allows polymerases to load a single DNA template strand into their active site and initiate incorporation of nucleotides dictated by the DNA template sequence. All polymerases require several additional factors that are critical for successful completion of the different steps of the transcription process as will be discussed below. Eventually, RNA polymerases arrive at the end of genes and terminate transcription via specialized mechanisms in a controlled manner.

Eukaryotic RNA polymerases are structurally and functionally related (Cramer et al. 2008) but transcribe different group of genes. RNA polymerase (pol) I is responsible for transcription of ribosomal (r) DNA repeats distributed over five chromosomes in humans (~400 copies) and concentrated in one chromosome in yeast (~200 copies) (Henderson, Warburton, and Atwood 1972; McStay and Grummt 2008; Petes 1979). RNA pol I synthesizes a precursor RNA (35S in yeast and 45S in humans) that is processed into mature 18S, 5.8S and 28S (25S in yeast) rRNAs. Pol III transcribes a group of non-coding RNAs including transfer RNAs (tRNAs), the 5S rRNA, and the U6 spliceosomal snRNA. The activities of pol I and III dominate in the nucleus which
combined account for more than 80% of total RNA synthesis in a yeast cell (Warner 1999). While RNA pol I and III transcribe a limited group of genes, RNA polymerase II transcribes a great variety of protein-coding genes and numerous noncoding RNAs such as small nucleolar RNAs (snoRNAs), cryptic unstable transcripts (CUTs), stable uncharacterized transcripts (SUTs) and microRNA (miRNA) precursors.

Interestingly, each RNA polymerase is not only recruited to specific genes but is also concentrated at discrete and different locations in the nuclear compartment. Transcription by RNA pol I is localized at discrete sites called nucleoli where rRNA transcripts are produced, processed and assembled into ribosomes (Shaw and Jordan 1995). Confocal microscopy and electron microscopy of HeLa cells revealed that transcription by RNA pol III similarly takes place at about 2000 sites that are different and separated from those of RNA pol II transcription, approximately 8000 per nucleus (Pombo et al. 1999).

The catalytic core of the three RNA polymerases is composed by 10 conserved subunits but additional polymerase-specific factors are required for regulation of gene-specific transcription (Vannini and Cramer 2012). All polymerases require the TATA box-binding protein (TBP) (Cormack and Struhl 1992), which directly binds the DNA upstream of the TSS (Rhee and Pugh 2012) and induces a 90º DNA bend (Nikolov et al. 1995). In addition, functionally and structurally related factors are found in all three initiation complexes. Initiation by RNA pol II also requires TFIIB, TFIIF and TFIIE; factors that appear to have counterparts in initiation complexes of RNA pol I and RNA pol III (Vannini and Cramer 2012). TFIIB is able to bridge pol II and TBP (Kostrewa et al. 2009). TFIIF is required for recruitment of pol II (Flores et al. 1991), start site selection (Killeen, Coulombe, and Greenblatt 1992) initiation and elongation (Tan, Conaway, and Conaway 1995). TFIIE additionally stabilizes and activates the preinitiation complex via
interactions with GTFs and recruits TFIIH to pol II (Ohkuma 1997). In contrast to pol I and pol III, which most likely use binding energy to melt the DNA template like bacterial RNA polymerase (Feklistov and Darst 2011), pol II requires TFIIH to open the DNA template. TFIIH contains XPB, a helicase with ATPase activity needed for promoter opening and promoter escape (Lin, Choi, and Gralla 2005). Nevertheless, a recent study argued that the helicase activity of XPB is required in the pre-initiation complex to overcome a block imposed by XPB itself and that its presence in the complex is not absolutely required for transcription (Alekseev et al. 2017). In that report it was shown that while inhibition of the helicase activity of XPB blocks transcription initiation, depletion of XPB does not seem to cause major effects in human osteosarcoma U2OS cells. In Chapter III I further show that in human cells inhibition of the helicase activity of XPB with the drug triptolide (Titov et al. 2011) causes accumulation of pre-initiation complexes at the transcription start site (TSS) of genes and that these complexes cannot translocate to downstream positions (Fig. 3-2A).

Once RNA polymerases transcribe a certain length of DNA, elongation complexes clear the promoter region and transition into a mode of productive elongation characterized by high elongation rates and processivity. Previous analysis of yeast pol III indicated an average rate of 22 nucleotides per second at 20ºC that was similar to the elongation rate for pol II in Drosophila (Matsuzaki, Kassavetis, and Geiduschek 1994). Using Global run-on sequencing (GRO-seq), a high-resolution assay for determining the location and orientation of actively transcribing Pol II genome-wide (Core, Waterfall, and Lis 2008), the elongation rate of pol II was calculated in human MCF-7 breast cancer cells after activation of transcription with 17β-estradiol (E2) (Danko et al. 2013). Interestingly, in that report elongation rates of pol II varied across both genes and cell types and induction conditions up to four-fold. For example, elongation rates on 140 long genes ranged between 0.37 and 3.57 kb/min, with a median rate of 2.1 kb/min. On the other hand, a recent study
using fluorescence microscopy observed fairly constant rates of Pol II elongation in the early Drosophila embryo with less than 25% variation (2.4–3.0 kb/min) for different activators and promoters using reporter genes (Fukaya, Lim, and Levine 2017). In Chapter IV, I additionally show results of the calculated global elongation rate of pol II in HEK293 cells (2.0 kb/min) using chromatin immunoprecipitation of pol II followed by sequencing (ChIP-seq) (Fig. 4-8). Pausing sites, however, can greatly affect transcription rates (Mayer, Landry, and Churchman 2017). In the case of pol II, promoter-proximal pausing of the elongation complex has emerged as a widespread regulatory mechanism also involved in signal-responsive pathways (Adelman and Lis 2012). A section dedicated to function and regulation of promoter-proximal pausing of pol II is presented below.

It is important that polymerases be released from the DNA template at 3’ ends of genes in a timely manner to avoid interference with downstream transcription units (Shearwin, Callen, and Egan 2005). To terminate transcription the three transcription machineries utilize different mechanisms. Termination of RNA pol I and III occurs at discrete sites closely downstream of 3’ ends of genes. Termination of pol I transcription occurs near a strong terminator element that is recognized by Reb1 in yeast (Lang et al. 1994) and by TTF-I in mammalian cells (Grummt et al. 1985). Binding of these factors to the terminator element in the right orientation is thought to induce pausing of pol I. A transcript-release element consisting of a stretch of T’s upstream of the Reb1 and TTF-1 binding sites is thought to further destabilize the paused elongation complex (Grummt et al. 1985). It has been reported that a release factor is also required in mammalian cells (Mason, Sander, and Grummt 1997). Similar to pol I, pol III terminates at a defined region. Pol III requires a T-rich consensus for termination (Bogenhagen and Brown 1981) frequently consisting of only four T’s in mammals (Braglia, Percudani, and Dieci 2005). The pol III-specific subunits
C37 and C53 are sufficient to recognize the stretch of T’s in the absence of auxiliary factors (Landrieux et al. 2006). It was proposed that recognition of the T-rich terminator region by C37/C53 induces a slowdown of pol III and release of the nascent transcript. This intrinsic ability of pol III to recognize a terminator element is not a common feature with pol I and pol II, which need auxiliary factors for termination.

In contrast, termination of RNA pol II does not occur at discrete sites. Termination of RNA pol II is coupled to 3’-end formation of the nascent transcript, a process in which the RNA 3’ end receives a tail of (A)n nucleotides (polyA) (Colgan and Manley 1997). Both, RNA 3’-end processing and termination require a polyA signal consisting of the hexamer AAUAAA sequence (further introduced below) (Connelly and Manley 1988b; Whitelaw and Proudfoot 1986). Termination of transcription downstream of the beta-globin (major) gene in nuclei from induced mouse erythroleukemia cells takes place within the region between 700 to 2000 bases downstream from the polyA site, but not at particularly favored sites (Citron et al. 1984). Genome-wide analysis of pol II positions around genes have also shown that termination of pol II transcription downstream of polyA sites occurs over a broad window of several kilobases downstream of the cleavage site (Fong et al. 2015b; Core, Waterfall, and Lis 2008), observed also in this study to occur over a window of more than 6 kb from polyA sites (Fig. 4-9). The emerging view is that termination of pol II transcription is a remarkably complex process that requires many different factors including those involved in 3’-end formation of the nascent transcript (Proudfoot 2016). Termination of pol II transcription is the main subject of discussion in this thesis and will be further introduced below.
1.2 Transcription by RNA Polymerase II

In order to produce a complete and functional mRNA molecule, pol II must undergo multiple successful and complex steps requiring an army of auxiliary factors. The transcription process, which can be regulated at almost every step, can be divided into four main stages: recruitment, initiation, elongation and termination (Fig 1-1). In this thesis, I present results from investigation of two important steps during transcription: promoter-proximal pausing of pol II and termination of pol II transcription. Results presented in this thesis provide additional description of the molecular events leading to termination of pol II transcription at 3’ ends of genes and have also uncovered that premature termination of promoter-proximally paused polymerases is a common phenomenon on human genes.

1.3 Promoter-Proximal Pausing of Pol II

Traditionally, it was considered that recruitment of RNA polymerases to promoters was a limiting step in the control of gene expression. Nevertheless, analysis of the uninduced Drosophila heat shock protein (HSP) genes revealed that transcriptionally engaged polymerases accumulate immediately downstream of the HSP promoters associated with short transcripts ranging in length from 20 to 60 nucleotides (nts) (Gilmour and Lis 1986; Rougvie and Lis 1988; Giardina, Perez, and Lis 1992; Rasmussen and Lis 1993). This population of polymerases that accumulate at promoter-proximal regions were then referred to as “paused” polymerases (Rougvie and Lis 1990).

Additional work performed in mammalian cells demonstrated that transcriptionally engaged polymerases also accumulate downstream of human promoters, including that of c-myc and Fos, adding strength to the idea of promoter-proximal pausing of pol II (Strobl and Eick 1992; Plet, Eick, and Blanchard 1995; Krumm et al. 1992). In contrast to these observations, studies in yeast revealed that recruitment of pol II to promoters was a major mode of gene regulation and no
The transcription process can be divided into four main steps: recruitment of RNA polymerase II (pol II), initiation, elongation and termination. Recruitment of pol II is accomplished by general transcription factors GTFs (red circles), which position pol II at the transcription start site (TSS). After association with GTFs and the Mediator complex (blue circles) pol II becomes hyper-phosphorylated on its C-terminal domain (yellow tail) and synthesizes the first phosphodiester bond (initiation). Early after initiation, pol II associates with elongation and RNA processing factors (orange circles) and form the elongation complex (EC). At about 30 to 70 nucleotides (nts) DSIF binds both the EC and the nascent transcript and cooperates with NELF to induce promoter-proximal pausing of pol II. At 3’ ends of genes, endonucleolytic cleavage at the cleavage and polyadenylation site (CPS) defines the 3’ end of the mRNA molecule (blue) which becomes polyadenylated and generates an uncapped 5’-monophosphorylated (PO₄) nascent transcript. Termination of transcription occurs when the EC ceases transcription and disassembles from the DNA template in a controlled manner by termination factors allowing for recycling of pol II for subsequent rounds of transcription.
evidence for promoter-proximal pausing of pol II was found (Stargell and Struhl 1996; Ptashne and Gann 1997). However, with the recent advances in sequencing technology it is clear that pausing of pol II, after transcription of the first 30 to 70 nucleotides, is a major limiting step in multicellular organisms. For instance, this bottle neck has been visualized by GRO-seq (Core, Waterfall, and Lis 2008) or PRO-seq (Kwak et al. 2013) in which maps of pol II positions demonstrate global accumulation of pol II densities downstream of the transcription start site (TSS). Additionally, these methods have also shown that paused polymerases are competent to undergo elongation, given that paused polymerases can resume elongation in vitro after treatment with the detergent sarkosyl (Core, Waterfall, and Lis 2008).

Two complexes cooperate to induce pausing at the promoter-proximal region: the negative elongation factor NELF and the DRB (5,6-dichloro-1-β-D-ribofuranosylbenzimidazole) sensitivity-inducing factor (DSIF), composed by Spt5 and Spt4 (Wu et al. 2003; Yamaguchi, Takagi, et al. 1999). In the current view, binding of DSIF to both pol II and the RNA molecule, once it is extruded from the active site, allows for recruitment of NELF and repression of transcription elongation (Fig. 1-1) (Missra and Gilmour 2010). In order for pol II to transition into productive elongation it requires the action of the positive transcription elongation factor b (pTEF-b), which phosphorylates DSIF and Pol II (Peterlin and Price 2006; Renner et al. 2001). The multiple phosphorylation events that take place at this transition are thought to cause release of NELF from the elongation complex and escape of pol II into transcription elongation (Yamada et al. 2006; Bourgeois et al. 2002b; Fitz, Neumann, and Pavri 2018).

It is thought that promoter-proximal pausing of pol II could have multiple roles during transcription. Pausing of pol II might increase the residence time around promoters to increase the “window of opportunity” for recruitment of processing factors to the carboxy-terminal domain
(CTD) of the largest subunit of pol II, Rpb1. The CTD is a flexible extension of pol II and consists of tandem heptad repeats (26 in yeast, 52 in humans) with the amino acid consensus sequence Tyr$_1$-Ser$_2$-Pro$_3$-Thr$_4$-Ser$_5$-Pro$_6$-Ser$_7$ (Corden 1990). These repeats are essential for life and are dynamically phosphorylated during the transcription cycle (Komarnitsky, Cho, and Buratowski 2000; Mayer, Heidemann, et al. 2012; Chapman et al. 2007; Zehring et al. 1988). Paused polymerases are highly phosphorylated on Ser$_5$ within the heptad repeats of the CTD which stimulates the 5’ capping enzyme (Ghosh, Shuman, and Lima 2011). Recruitment of the 5’ capping enzyme might additionally be facilitated via its interactions with the pausing-inducing factor DSIF (Mandal et al. 2004; Moore and Proudfoot 2009). Additionally, it has been proposed that pausing might allow integration of multiple regulatory signals. While some activators of transcription such as Sp1 (Blau et al. 1996) and GAGA factor (Lee et al. 2008) appear to stimulate transcription via recruitment of pol II at the promoter, others appear to stimulate recruitment of pTEF-b and pause release (Rahl et al. 2010). Furthermore, some activators of transcription can facilitate both recruitment and pause release of pol II (Barboric et al. 2001; Blau et al. 1996). This suggests that a concerted action or “integration of regulatory signals”, depending on sequence context, may be facilitated by pausing of pol II. On the other hand, pausing of pol II might serve to facilitate a permissive promoter architecture. Interestingly, it has been observed that genes with promoter-proximal pol II pausing persist in a nucleosome-deprived state that is dependent on the presence of the paused pol II (Gilchrist et al. 2010; Gilchrist et al. 2008). This has led to the notion that pausing of pol II might also serve as a barrier against nucleosome formation and repressive chromatin structure by occupying the DNA template around the promoter-proximal region.

Importantly, an alternative fate of paused polymerases might be premature termination of transcription. This idea has recently gained substantial support from analyses revealing a short
residence time of pol II in the paused state (Steurer et al. 2018). While previous studies using ChIP-seq (Henriques et al. 2013; Shao and Zeitlinger 2017), Gro-Seq (Jonkers, Kwak, and Lis 2014), or methyltransferase footprinting assays (Krebs et al. 2017) addressed pol II kinetics and provided measurements in the order of minutes, these studies commonly used the drug triptolide to inhibit transcription initiation. Since triptolide targets the ATPase XPB helicase and prevents melting of the DNA template (Titov et al. 2011), the rationale involved the assumption that triptolide prevented recruitment of new polymerases at the transcription start site. However, my results in chapter III demonstrate that triptolide treatment does not prevent recruitment of the preinitiation complex (fig. 3-1A, 3-2A) and allows for a peak of pol II at the TSS that could interfere with measurements at low resolution. Furthermore, studies showing that this drug not only affects pol II levels (Wang et al. 2011; Manzo et al. 2012) but also has a slow mode of action (Nilson et al. 2017), opens the possibility of overestimation of the residence time of pol II in the pause state from experiments using triptolide. In fact, fluorescence microscopy and photobleaching studies on GFP-RPB1, the largest subunit of pol II, indicated that pol II transcription pauses for less than a minute (~ 42 s) (Steurer et al. 2018). Thus, an alternative view points to a dynamic population of paused polymerases controlled by release from the DNA template and continuous reinitiation.

In accord with this view, at least three potential premature termination mechanisms have been proposed to operate on paused polymerases. All of these agree with the generation of a 5’-monophosphorylated transcript that is attacked by the 5’-3’ exonuclease Xrn2. As further introduced below, the exonuclease Xrn2 contributes in multiple aspects or RNA biology but is also an important factor involved in termination of pol II transcription. The role of Xrn2 in termination of pol II transcription is the main focus of this study and results are discussed in the following Chapters. According to the torpedo model, Xrn2 recognizes the 5’-monophosphorylated end of the
nascent transcript produced by RNA cleavage at 3’ ends of genes and degrades it 5’ to 3’ in a chase for pol II. When Xrn2 catches up with the elongation complex it is thought to trigger release of pol II from the DNA template by a yet unknown mechanism, further introduced below (Fig. 1-2) (Connelly and Manley 1988b; West, Gromak, and Proudfoot 2004; Gromak, West, and Proudfoot 2006). The first example of a premature termination mechanism that involves generation of a 5’-monophosphorylated end that is attacked by Xrn2 was proposed from investigation of the HIV promoter, a well-known system from which the effects of DSIF, NELF and pTEF-b have been investigated. The endoribonuclease normally involved in miRNA maturation, Microprocessor (Han et al. 2006), was shown to bind the RNA stem loop structure generated around the HIV promoter after initiation of transcription, known as the transactivation response (TAR) element, and trigger cleavage to allow termination by Xrn2, SETX and Rrp6 (Wagschal et al. 2012). Interestingly, this mechanism appeared to extend to a subset of human endogenous genes, consistent with HITS-CLIP data demonstrating interactions of Microprocessor with many non-miRNA cellular RNAs (Macias et al. 2012). Alternatively, the presence of decapping factors around promoters by ChIP-seq and a reduction in the pausing index of pol II genome-wide upon depletion of Dcp1a, or Xrn2, also postulated co-transcriptional decapping as a possible route to premature termination (Brannan et al. 2012). A similar role was proposed for the human decapping factor DXO on transcripts that fail to acquire the correct cap structure (Jiao et al. 2013). And finally, a recent study reported premature termination from usage of cryptic polyA signals associated with the +1 nucleosome for a large class of mammalian promoters (Chiu et al. 2018). Although in this case, polymerases in front of a nucleosomal barrier, rather than promoter-proximally paused polymerases in the same vicinity, were proposed to be the target of a premature termination checkpoint controlled by the U1 snRNP.
Figure 1-2. The torpedo model of RNA polymerase II transcription termination

According to the torpedo model of transcription termination, cleavage of the nascent transcript at the cleavage and polyadenylation site (CPS) generates a 5’-monophosphorylated (PO₄) RNA end that becomes a target of the 5’-3’ RNA exonuclease Xrn2. Cleavage of the nascent transcript occurs at about 17-25 nucleotides (nts) downstream of the polyA (pA) signal (AAUAAA). Degradation of the nascent transcript is proposed to allow Xrn2 to chase pol II. Eventually, Xrn2 reaches the elongation complex and somehow induce dissociation of pol II from the DNA template. The RNA 5’ cap structure is represented as a red circle and the polyA signal as a short red stretch in the RNA molecule (blue). The purple arrow indicates direction of transcription and the green arrow marks an upstream transcription start site (TSS).
Nevertheless, it has not been demonstrated that premature termination is a global phenomenon on human promoters. Results presented in Chapter V, however, provide strong evidence that promoter-proximally paused polymerases are highly turned over by premature termination of transcription genome-wide. Furthermore, in that chapter I identified a population of polymerases that can migrate into gene bodies without licensing by pTEF-b and show that this population of polymerases is targeted for premature termination of transcription by Xrn2.

1.4 The Elongation Factor Spt5

Interestingly, Spt5 (NusG in bacteria) is a pol II-associated factor that is conserved in all domains of life (Werner 2007) and can positively and negatively impact transcription elongation. As described above Spt5 is part of the DSIF complex (Spt4/Spt5) and binds to both the nascent transcript and pol II and cooperates with NELF to induce promoter-proximal pausing of a reconstituted elongation complex \textit{in vitro} (Missra and Gilmour 2010; Yamaguchi, Takagi, et al. 1999). Importantly, phosphorylation of the C-terminal domain of human Spt5 by pTEF-b has been proposed to switch Spt5 from a pausing factor into a positive elongation factor (Yamada et al. 2006). This was proposed based on the observation that phosphorylation of Spt5 by pTEF-b was not required for its repressive activity on elongation, but it was critical for Spt5-mediated processive transcription elongation \textit{in vitro} (Yamada et al. 2006). Additionally, chromatin immunoprecipitation analysis has shown that drosophila Spt5 is found across the body region of genes \textit{in vivo} (Andrulis et al. 2000; Aida et al. 2006), leading to the notion that Spt5 positively impacts elongation throughout the length of genes. A recent study in yeast \textit{Schizosaccharomyces pombe} showed that Spt5 phosphorylation levels increase downstream of the TSS and drop immediately downstream of 3’ ends of genes (Parua et al. 2018). In that study, it was reported that Dis2, an isoform of protein phosphatase 1 (PP1), can dephosphorylate Spt5 \textit{in vitro} and that Dis2
is required to prevent transcription beyond normal termination zones. Additionally, in the same report, it was proposed that a transition of pol II from elongation to termination is regulated by opposing activities of Cdk9 and Dis2 towards their common substrate Spt5. Nevertheless, it is still unclear how changes in Spt5 phosphorylation impact transcription elongation within gene bodies or at 3’ ends of genes after Spt5 phosphorylation drops. Similarly, in Chapter IV I show that levels of Spt5 phosphorylation rapidly increase downstream of human promoters over the first ~2 kb within gene bodies (Fig. 4-16) and that Spt5 phosphorylation levels drop downstream of 3’ ends of genes (Fig. 4-20). An interesting speculation that I make in Chapter IV, based on additional interesting results, is that loss of Spt5 phosphorylation at 3’ ends of genes might revert back the repressive activity of Spt5 on transcription elongation to promote slowdown of the elongation complex before termination of pol II transcription.

Signals triggering such transcriptional repression might come from the nascent transcript. Binding of Spt5 with the nascent RNA was found to sharply increase immediately downstream of the polyA site in yeast by PAR-CLIP analysis (Baejen, Andreani, Torkler, Battaglia, Schwalb, Lidschreiber, Maier, Boltendahl, Rus, Esslinger, Soding, et al. 2017), pointing to the possibility that Spt5 might “sense” passage of pol II over the polyA site. A potential role of Spt5 in RNA 3’-end processing and termination is further suggested by the fact that its C-terminal domain contributes to recruitment of the 3’ RNA cleavage factor I at 3’-ends of genes in yeast (Mayer, Schreieck, et al. 2012).

Spt5 contains a conserved NGN (NusG N-terminal) domain, which binds Spt4, and different KOW (Kyprides-OnzonisWoese) domains connected by flexible C-terminal repeat regions (CTRs) (Meyer et al. 2015; Werner 2012). According to structural studies Spt5 “clamps” both the DNA and the RNA exiting from pol II while making contacts with pol II. Spt4 and Spt5
NGN and KOW1 domains encircle the exiting DNA, and Spt5 KOW4 and KOW5 domains encircle the exiting RNA transcript (Ehara et al. 2017; Bernecky, Plitzko, and Cramer 2017). Therefore, the multiple contacts of Spt5 with DNA and RNA, as well as simultaneous binding of Spt5 with pol II and the non-template DNA strand within the paused transcription bubble (Yakhnin et al., 2016) is consistent with the possibility that this highly conserved factor may directly induce a specific pol II conformational change conducive to pausing.

1.5 mRNA 3’-end Processing

Transcripts originated from protein-coding genes receive a poly(A)n tail by cleavage and processing at their 3’ ends which is essential for their function and stability (Fig. 1-2) (Eckmann, Rammelt, and Wahle 2011). For cleavage and polyadenylation (polyA) of the nascent transcript to occur an important motif needs to be transcribed, AAUAAA, initially revealed by comparison of nucleotide sequences upstream of the polyA sites of several mRNAs. Such mRNAs included the α- and β-globulin mRNAs of rabbit and human, the immunoglobulin light chain mRNA of mouse (MOPC 21) and the ovalbumin mRNA of chicken (Proudfoot and Brownlee 1976). Since then, the canonical A[A/U]UAAA hexamer sequence has been found to be highly conserved, being AUUAAA the most frequent variant (Beaudoing et al. 2000). Mutations at any other positions in the hexamer strongly inhibit 3’-end processing (Sheets, Ogg, and Wickens 1990; Wickens and Stephenson 1984; Wilusz, Pettine, and Shenk 1989). In addition to the polyA signal, a downstream sequence element (DSE) corresponding to a poorly conserved and diffuse sequence of Us and/or G/U is normally located 10-30 nucleotides downstream of the polyA site. In the case of U-rich elements, these have been described to be of short length ~5 nucleotides (Chou, Chen, and Wilusz 1994; Gil and Proudfoot 1987), and the GU-rich found downstream of polyA sites in a survey of 70 mammalian genes found the consensus YGUGUUYY (Y = pyrimidine) (McLauchlan et al. 1997).
1985). Some genes can have both a U-rich and a G/U-rich element that can function synergistically (Gil and Proudfoot 1987). In humans, the cleavage event occurs between the polyA signal and the DSE at two frequent dinucleotides: CA and UA (Li and Du 2013).

The polyA signal and the DSE are recognized on the nascent transcript by direct binding of the 3’-end processing machinery. Four important multi-subunit factors are part of the core of 3’-end processing machinery. These are the cleavage and polyadenylation specificity factor (CPSF), the cleavage stimulation factor (CstF), and mammalian cleavage factors I and II (CFIm and CFIIm). Recognition of the polyA signal is achieved by cooperation of CPSF binding to the polyA signal and Cstf binding to the DSE (Gilmartin and Nevins 1991). Mutual cooperative interactions of CPSF, CstF, mammalian cleavage factors and the polyA polymerase (PAP), which is responsible for adding adenosine residues at the 3’ end of RNAs (Gunderson et al. 1994), catalyze endonucleolytic cleavage and polyadenylation of the RNA 3’ end (Gilmartin and Nevins 1991; Murthy and Manley 1995; Ruegsegger, Beyer, and Keller 1996; Wilusz et al. 1990). CstF is composed of CstF77, CstF64, and CstF50, of which CstF64 makes direct contacts with the G/U-rich element sequence and also interacts with CstF77, which in turn interacts with CstF50 (Takagaki and Manley 2000). CPSF contains CPSF160, CPSF100, CPSF73, CPSF30 and Fip1 (Kaufmann et al. 2004). While CPSF30 is responsible for direct binding to the polyA signal (Chan, Huppertz, et al. 2014), CPSF73 is the endonuclease that catalyzes the cleavage reaction (Mandel et al. 2006). A complex network of protein-protein interactions stabilizes the entire 3’-end complex and juxtaposes CPSF73 with the cleavage site. The factor symplekin can interact with several polyadenylation factors in this network and is thought to function as an important scaffold during 3’-end processing (Takagaki and Manley 2000; Ghazy et al. 2009; Kyburz et al. 2003). The two-
step reaction consisting of endonucleolytic cleavage and polyadenylation takes under 10 seconds for the SV40 early polyA site to complete in vivo (Chao et al. 1999).

During infection by the influenza A virus, the viral NS1A protein inhibits polyA signal recognition by CPSF30 and is responsible for inhibiting production of functional cellular mRNAs (Chen, Li, and Krug 1999; Nemeroff et al. 1998; Shimizu et al. 1999), including the production of antiviral interferon IFN-β mRNAs (Noah, Twu, and Krug 2003). X-ray crystal structures show that the C-terminal domain of NS1A binds two zinc finger domains of CPSF30 (Das et al. 2008). NS1A binding to CPSF30 in vitro prevents CPSF binding to the RNA transcript and inhibits 3’ end cleavage and polyadenylation of pre-mRNAs (Nemeroff et al. 1998). NS1A expression in vivo similarly leads to accumulation of uncleaved host pre-mRNAs that remain in the nucleus (Nemeroff et al. 1998). In Chapter IV, I additionally show that expression of NS1A in mammalian cells leads to a transcription termination defect characterized by increased pol II transcription farther away from 3’ ends of genes, consistent with a previous report showing that NS1A expression in mammalian COS cells results in increased synthesis of nascent transcripts downstream of the SV40 late polyA signal (Nag, Narsinh, and Martinson 2007). In Chapter IV, I utilize expression of NS1A to inhibit binding of CPSF30 to the nascent transcript and show that polyA signal recognition is required for both preventing read-through transcription of pol II farther away from polyA sites and for loss of Spt5 phosphorylation downstream of polyA sites. Coupling of 3’-end processing with the termination process has been observed for a long time (Birse et al. 1998; Connelly and Manley 1988b; Yonaha and Proudfoot 2000). How 3’-end processing factors may contribute in termination of pol II transcription is introduced below. In Chapter IV, I propose that 3’-end processing factors cooperate with the 5’-3’ exonuclease Xrn2 by slowing down the elongation complex to facilitate Xrn2-dependent termination of pol II transcription.
1.6 The 5’-3’ Exonuclease Xrn2

Xrn2 is an evolutionarily conserved exonuclease in eukaryotes (Sugano et al. 1994; Shobuike et al. 1995; Zhang et al. 1999; Zakrzewska-Placzek et al. 2010). It was originally identified in yeast *Saccharomyces cerevisiae* as an essential gene required for export of polyadenylated transcripts after synthesis (Amberg, Goldstein, and Cole 1992). This protein was named Rat1 for ribonucleic acid trafficking protein 1; although Xrn2/Rat1 is also known as HKE1, DHM1, TAP1.

Xrn2 localizes within the nuclear compartment (Johnson 1997) and contains regions homologous with the cytoplasmic 5’-3’ exonuclease Xrn1 (Aldrich et al. 1993). Analogous to Xrn1, Xrn2 processively degrades single-stranded RNAs containing a 5’-monophosphate by degrading “nucleotide-by-nucleotide” in the 5’ to 3’ direction with no sequence specificity (Stevens and Poole 1995). Interestingly, both Xrn1 and Xrn2 can complement each other for degradation of RNA transcripts when they become mislocalized to the nuclear or cytoplasmic compartment, respectively. In yeast, a mutation of Rat1 in its nuclear localization signal (NLS) led to mislocalization of Rat1 in the cytoplasm and could complement defects of Xrn1 mutants in cytoplasmic RNA turnover (Johnson 1997). Conversely, it was also shown that localization of Xrn1 in the nuclear compartment by fusion of a NLS to Xrn1 could restore the phenotypic defect of Rat1 mutant cells (Johnson 1997).

Xrn2 is an essential exonuclease involved in many aspects of cell biology via degradation of nascent transcripts. Depletion of Xrn2 causes growth arrest in yeast *S. cerevisiae* (Amberg, Goldstein, and Cole 1992) and slow growth and sterility in *C. elegans* (Chatterjee and Großhans 2009; Frand, Russel, and Ruvkun 2005). Significantly, Xrn2 has been well-known for its participation in termination of pol II transcription. After transcription of the polyA signal, RNA
cleavage by 3’-end processing factors generates a 5’ phosphate that, until now, had been envisioned to be recognized by Xrn2 for degradation. Degradation of the nascent transcript has been proposed to direct Xrn2 towards pol II as a “torpedo” for termination of transcription (Fig. 1-2) (Connelly and Manley 1988b; West, Gromak, and Proudfoot 2004; Pearson and Moore 2013). In this thesis, I present evidence confirming that this is in fact the case and that Xrn2 loads onto nascent transcripts for 5’-3’ degradation after cleavage at polyA sites in human genes (Chapter III). The evidence suggesting that Xrn2 participates in termination of transcription originally came from studies in yeast. Depletion of Rat1 lead to termination defects in which pol II transcribes farther downstream of polyA sites (Kim et al. 2004). Later, additional experiments by the Bentley lab and others showed that the exonuclease activity of Xrn2 is required for proper termination of transcription in yeast and mammalian cells genome-wide (further discussed below) (Baejen, Andreani, Torkler, Battaglia, Schwalb, Lidschreiber, Maier, Boltendahl, Rus, Esslinger, Söding, et al. 2017; West, Gromak, and Proudfoot 2004; Fong et al. 2015b). The role of Xrn2 in termination of pol II transcription is furthered discussed in the next section.

A role in premature termination of transcription by Xrn2 after co-transcriptional decapping was also proposed as a potential mechanism to control transcription from promoter regions (Brannan et al. 2012). Interestingly, human promoters commonly generate anti-sense transcription upstream from protein-coding genes, although its role is still unclear (Andersson et al. 2015). The Bentley lab proposed that a role of Xrn2 in premature termination in human promoters might additionally serve to limit antisense transcription from human promoters and contribute to promoter-directionality (Brannan et al. 2012). Consistent with a role of Xrn2 in premature termination, Xrn2 was found to be highly enriched around promoters in this study. In accord with those observations, Xrn2 affects a population of small RNA fragments associated with promoter
regions. Human promoters generate small transcription start site associated (TSSa) RNAs (16-30nts) thought to be derived from pol II pausing events that additionally protects them from exonucleolytic degradation while being bound by pol II (Seila et al. 2008). Importantly, the 5’ ends of TSSa RNAs do not match with their annotated sequences for the first nucleotides but are extended and mapped to the TSS after depletion of Xrn2, suggesting a role of Xrn2 in degradation of TSSa transcripts (Valen et al. 2011). Furthermore, depletion of both Xrn1 and Xrn2 led to a smaller peak of TSSa RNA transcripts and to an increased fraction of TSSa RNAs that were longer than 30 nucleotides (Valen et al. 2011). An unresolved question was whether Xrn2 can induce premature termination of transcription around promoters. In this thesis, I provide evidence suggesting that indeed Xrn2 functions in global premature termination of promoter-proximally paused polymerases in the human genome (Chapter V).

Xrn2 is additionally involved in processing and degradation of a variety of RNAs. During processing of ribosomal rRNAs in yeast and mammals, Xrn2 is involved in maturation of the 5.8s and the 25S (28S in humans) rRNAs, as well as some spacer fragments excised from the pre-rRNA (Petfalski et al. 1998; Wang and Pestov 2011). Likewise, Xrn2 has been reported to be required for biogenesis of snoRNAs in yeast after cleavage by Rnt1 (Petfalski et al. 1998; Chanfreau et al. 1998; Qu et al. 1999) or after excision of snoRNAs from an intron of the BEL1 transcript by 5’-3’ trimming of 5’ leaders (Petfalski et al. 1998). Interestingly, Xrn2 can control miRNAs levels and degrade transfected siRNAs (Wei et al. 2012; Chatterjee and Grohsans 2009). Finally, Xrn2 and the nuclear exosome additionally participate in degradation of pre-mRNAs in the nuclear compartment to control mRNA levels (Das, Butler, and Sherman 2003; Bousquet-Antonelli, Presutti, and Tollervey 2000).
Quality control mechanisms involved in elimination of aberrant transcripts also appear to commonly harness the exonuclease activity of Xrn2. Defective pre-rRNA transcripts can be eliminated by Xrn2 in an apparent coordination with the nuclear exosome in yeast (Fang, Phillips, and Butler 2005). Hypo-modified tRNAs, such as those lacking proper methylation required for tRNA function, can be degraded by Xrn2 and Xrn1 (Chernyakov et al. 2008). Importantly, Xrn2 additionally targets nascent pol II transcripts that fail to acquire a proper cap structure in yeast (Jiao et al. 2010; Jimeno-Gonzalez et al. 2010). Furthermore, failure to splice the pre-mRNA, process its 3’-end or terminate transcription can also lead to degradation by Xrn2 (Bousquet-Antonelli, Presutti, and Tollervey 2000; Davidson, Kerr, and West 2012). Therefore, the 5’-3’ exonuclease Xrn2 has multiple functions that are relevant for RNA biology. This can easily explain why Xrn2 is essential for cell growth in yeast, although it is still unclear if only one or multiple functions of Xrn2 are responsible its essentiality. A role of Xrn2 in control of cellular homeostasis in response to heat shock will also be proposed in this thesis in Chapter V.

1.7 Termination of Pol II Transcription

Termination of transcription occurs when pol II ceases synthesis of the nascent transcript and releases the DNA template. Termination is a controlled and coordinated process that must occur in a timely manner to prevent interference with downstream genes and synthesis of potentially harmful intergenic transcripts (Greger and Proudfoot 1998). In addition, termination might serve to facilitate transcription initiation by increasing the pool of available polymerases for additional rounds of transcription initiation. This idea has been suggested by the observation that some active yeast genes appear to have a “gene-loop” conformation in which the termination region can be physically linked to its promoter (O'Sullivan et al. 2004). Most of what is known about termination comes from investigation of three different classes of pol II transcribed genes:
protein-coding genes, histone genes and those that produce non-coding small nuclear snRNAs and small nucleolar snoRNAs. Elements required for processing of 3’ ends of pre-mRNAs, including histones pre-mRNAs and small non-coding RNA genes are required for termination (Chodchoy, Pandey, and Marzluff 1991; Carneiro et al. 2008; Morlando et al. 2002; Gu and Marzluff 1996; Connelly and Manley 1988b; Buratowski 2005). This linkage to RNA 3’-end processing has made analysis of the termination process more challenging and, in the case of protein-coding genes, controversial as discussed below.

Two major pathways of pol II transcription termination have been the focus of active investigation: 1. Termination of pol II transcription at 3’ ends of polyA signal-containing genes and 2. termination of snRNA and snoRNAs genes in yeast which are not polyadenylated (Kim et al. 2006). The Nrd1 protein is an essential RNA-binding factor that participates in termination of these small non-coding RNAs and some mRNAs in yeast in association with the helicase Sen1 (Steinmetz and Brow 1996; Arigo et al. 2006). Nrd1 additionally interacts with another essential protein Nab3 and with the CTD of pol II via its C-terminal interacting domain (CID) (Meinhart and Cramer 2004; Conrad et al. 2000; Yuryev et al. 1996). These three proteins form the Nrd1-Nab3-Sen1 complex which localizes to promoter regions of snoRNAs, and mutations in either Sen1 or Nrd1 cause strong termination defects on these genes (Kim et al. 2006; Nedea et al. 2003). Termination regions at snoRNAs contain multiple sequence motifs that are recognized by Nrd1 (GUAA/G) and Nab3 (UCUU) on the RNA transcript (Steinmetz and Brow 1998; Carroll et al. 2004) and their mutation also leads to termination defects (Carroll et al. 2004). Additional experiments have shown that Sen1 binds the CTD of pol II and mutations that affect Sen1 helicase activity similarly result in read-through transcription by pol II at 3’ ends of snoRNA genes (Kim et al. 2006; Steinmetz et al. 2006). It is thought that the helicase activity of Sen1 may disrupt the
RNA-DNA hybrid in the elongation complex leading to termination. It is worth pointing out here that in Chapter IV, I speculate that dephosphorylation of the elongation factor Spt5 is important for slowdown of polymerases downstream of polyA sites and termination of pol II, which might be dephosphorylated by protein phosphatase I (PP1) based on a recent study in yeast (Parua et al. 2018). Interestingly, the catalytic subunit of yeast PP1, Glc7, is a regulator of snoRNA termination, interacts with Sen1 and was shown to dephosphorylate Sen1 \textit{in vitro} (Nedea et al. 2008), indicating that dephosphorylation activity might be important for termination at both small non-coding RNAs and polyA site-dependent genes.

Nrd1 binds preferentially to the CTD of pol II phosphorylated on Ser5 (Vasiljeva et al. 2008), a CTD phosphorylation mark highly enriched around promoters that decreases progressively within gene bodies (Komarnitsky, Cho, and Buratowski 2000; Heidemann et al. 2013), which explains the observed localization of Nrd1-Nab3-sen1 complex to promoter regions and why this pathway preferentially functions on short transcription units (Steinmetz et al. 2006). Additionally, Nrd1 associates with the 3′-5′ exonuclease complex, exosome, and the exosome-activating complex TRAMP to stimulate the exosome’s ability to trim snoRNA 3′ ends (Vasiljeva and Buratowski 2006). This function of Nrd1 couples 3′-end processing of snoRNAs with termination of pol II transcription. Coupling of 3′-end processing and termination has also been observed during termination of transcription of polyA site-dependent genes.

For a long time, it has been observed that the polyA signal, required for 3′-end processing, is also a cis element required for termination of pol II transcription (Connelly and Manley 1988b; Logan, Falck, et al. 1987; Whitelaw and Proudfoot 1986). From that observation, two models dependent on a polyA signal were proposed: the allosteric model and the torpedo model. In the allosteric model, transcription of the polyA signal leads to a conformational change of the
elongation complex by dissociation of an antitermination/elongation factor or by association of a termination factor (Fig. 1-3) (Logan, Falck, et al. 1987). In the torpedo model, cleavage at the polyA site was proposed to create an entry point for a 5’-3’ RNA exonuclease that targets the uncapped 5’ end on the growing transcript for degradation. According to this model, “catching up” of the exonuclease with pol II is proposed to somehow induce release of pol II from the DNA template (Fig. 1-2) (Connelly and Manley 1988b). After extensive investigations, Xrn2 has emerged as the “torpedo” exonuclease proposed to induce termination at 3’ end of genes (Fong et al. 2015b; Nojima et al. 2013; West, Gromak, and Proudfoot 2004; Kim et al. 2004). An intense debate over the last three decades for and against these models point to a more intertwined mechanism of transcription termination. I have aimed to contribute to the resolution of such debate in Chapters IV and VII.

Initial evidence for the torpedo model comes from studies in yeast where mutation of Rat1 generates a substantial termination defect downstream of polyA sites (Kim et al. 2004). Additionally, self-cleavage of the nascent transcript at a co-transcriptional cleavage (CoTC) site downstream of the human beta-globin gene (Teixeira et al. 2004), was shown to act as a precursor to termination by presenting an uncapped 5’ RNA end that is recognized by Xrn2 (West, Gromak, and Proudfoot 2004). In that report, it was shown that depletion of Xrn2 results in reduced termination efficiency downstream of the CoTC cleavage site and stabilization of the downstream nascent transcript on co-transfected plasmids in HeLa cells. This result added strength to the idea that Xrn2 functions as a torpedo that degrades the nascent transcript and chases pol II for termination. Nevertheless, a global requirement of Xrn2 in termination of pol II transcription was challenged by seemingly contradictory observations. In vitro assays have shown that disassembly of polymerases can occur in the absence of cleavage or Xrn2 exonuclease activity
Figure 1-3. The allosteric model of RNA polymerase II transcription termination
According to the allosteric model, transcription of pol II through the polyA signal leads to a conformational change of the elongation complex (represented by a change in color of the elongation complex from orange to dark red) that converts it into a termination-competent elongation complex. This is proposed to occur either by (A) release of a positive elongation factor “E” from the elongation complex, or (B) association of a termination factor “T” with the elongation complex.
Similarly, partial depletion of Xrn2 in human cells does not seem to cause a termination defect (Brannan et al. 2012; Nojima et al. 2015a), suggesting that transcription termination does not require cleavage and “torpedo” action by Xrn2. Recent investigations have reported contrasting results in human cells however. On one hand, depletion of CPSF73 does in fact generate a strong termination defect that cannot be rescued by overexpression of a cleavage-deficient CPSF73 mutant (Eaton et al. 2018). On the other hand, the use of an exonucleotically-dead mutant (MT) of Xrn2 (D235A) demonstrated that degradation of the nascent transcript is in fact required for proper termination of pol II transcription on most human genes (Fong et al. 2015a). In that report, it was shown that expression of Xrn2MT results in stronger pol II occupancy downstream of polyA sites and increased transcription farther downstream of 3’ ends of genes. In agreement with kinetic competition as described by the torpedo model, it was also shown that mutations that affect the elongation rate of pol II lead to early termination of slow polymerases, and delayed termination of fast polymerases downstream of the polyA signal. Therefore, it is still unclear to what extent cleavage-independent mechanisms can support termination of transcription in vivo.

Alternative mechanisms that might involve a conformational change of the elongation complex conducive to pausing and termination, according to the allosteric model, have been proposed. Notably, the termination factor Pcf11, a core component of the cleavage and polyadenylation machinery, is able to release pol II in vitro independent of transcript cleavage or degradation (Zhang, Fu, and Gilmour 2005; Zhang and Gilmour 2006). It was proposed that simultaneous binding of Pcf11 to the nascent transcript and the C-terminal domain (CTD) of pol II might induce a conformational change capable of evicting the elongation complex, a mechanism that appears common with the termination factor Nrd1 during termination of polyA-independent
genes in yeast (Arigo et al. 2006). Additionally, it has been reported that binding of cleavage and polyadenylation factors might lead to a conformational change of the elongation complex that can be blocked by the transcriptional inhibitor alpha-amanitin before termination \textit{in vitro} (Zhang, Rigo, and Martinson 2015). Nevertheless, it is still not clear if the elongation complex suffers a polyA site-dependent conformational change capable of inducing disassembly from the DNA template \textit{in vivo}. It seems plausible that at least pausing of a highly processive elongation complex might be a prerequisite for termination. In prokaryotic systems, pausing of RNA polymerases is a common feature of the elongation process (Mustaev, Roberts, and Gottesman 2017) that might facilitate opening of the clamp, and render the elongation complex susceptible to termination (Sekine et al. 2015; Weixlbaumer et al. 2013). In eukaryotic systems, the common observation that pol II accumulates downstream of transcription units by ChIP-seq (Lian et al. 2008; Fong et al. 2015b), Gro-seq (Core et al. 2012) or mNET-Seq (Nojima et al. 2015b) is also consistent with pol II pausing downstream of the polyA signal. However, it is still not known if pol II pausing is required for termination and which mechanism would be responsible for it.

Several reports suggest that cleavage and polyadenylation factors may contribute to pol II pausing at $3'$-ends of genes. Binding of the CPSF factor to both the AAUAAA hexamer sequence and the body of pol II induces pausing in a reporter containing the SV40 late polyA signal (Nag, Narsinh, and Martinson 2007). Further support for a role of $3'$-end processing factors in limiting elongation of pol II upon binding to the polyA signal was shown in the same study, overexpression of the influenza virus protein NS1A, which targets CPSF30 and blocks CPSF binding to the polyA signal (Twu et al. 2006; Nemeroff et al. 1998), resulted in increased pol II transcription farther downstream of the SV40 late polyA signal. Thus, CPSF30 bound to the elongation complex might be responsible for recognition of the AAUAAA motif once it is extruded from the pol II active site.
to tether the polyA signal to the elongation complex and somehow trigger a slowdown of pol II at 3’ ends of genes. Such a mechanism is further suggested by the observation that CPSF associates with the transcription initiation factor TFIID in mammalian cells and yeast (Dantonel et al. 1997; Sanders et al. 2002), and that it is transferred from the transcription factor TFIID to the elongating polymerase in an in vitro mammalian system where it is thought to scan the RNA, awaiting the nascent polyA signal (Dantonel et al. 1997).

A role of 3’-processing factors in slowing down the elongation complex upon polyA signal recognition would explain the still controversial observation that depletion of processing factors generally cause stronger termination defects than depletion of Xrn2/Rat1 (Nojima et al. 2015a; Baejen, Andreani, Torkler, Battaglia, Schwalb, Lidschreiber, Maier, Boltendahl, Rus, Esslinger, Soding, et al. 2017; Eaton et al. 2018). If 3’-processing factors induce slowdown of pol II and facilitate catching-up of Xrn2 with the elongation complex downstream of the polyA signal, inhibition of these factors would cause pol II to read-through termination regions at high elongation rates farther downstream of gene ends. In principle, pausing of pol II may facilitate termination by shortening the path of Xrn2 to catch up with the elongation complex. In fact, specific pausing sequences downstream of transcription units contribute to termination of pol II (Enriquez-Harris et al. 1991; Ashfield et al. 1994; Yonaha and Proudfoot 1999; Plant et al. 2005; Gromak, West, and Proudfoot 2006). Nevertheless, it is still not known to what extent pol II slows down past the polyA signal, and whether or not pausing is required for termination. In Chapter IV I provide strong evidence that pol II transitions at the polyA site into a mode of slow transcription elongation consistent with a polyA site-dependent conformational change of the elongation complex before it becomes a target of Xrn2. In that chapter I propose a unified allosteric-torpedo model of RNA polymerase II transcription termination.
1.8 The Transcriptional Response to Heat Shock

The transcriptional heat shock response (HSR) is critical for cellular homeostasis and survival of all organisms. Under conditions of elevated temperature, essential proteins are easily unfolded and lose their function (Dubois, Hovanessian, and Bensaude 1991). Cells are properly equipped to deal with such deleterious events. It is thought that cells trigger activation of the heat shock response as a consequence of an imbalance of protein homeostasis in general rather than by sensing heat per se (Richter, Haslbeck, and Buchner 2010). Hence, a variety of stresses that induce protein unfolding including oxidative stress and heavy metals can trigger activation of the heat shock response (Courgeon, Maisonhaute, and Best-Belpomme 1984; Heikkila et al. 1982). To cope with the induced proteotoxic stress, heat shock protein (HSP) genes are expressed after exposure to heat (Lindquist 1986). Most of these factors are molecular chaperones that facilitate folding of proteins and are critical for survival (Lindquist 1986). The heat shock proteins that behave as molecular chaperones belong to five conserved families: HSP100s, HSP90s, HSP70s, HSP60s and small heat shock proteins (sHSPs) (Jakob et al. 1999). In general, molecular chaperones interact promiscuously with unfolded proteins due to the common feature of exposed hydrophobic amino acids by damaged proteins (Bukau, Hesterkamp, and Luirink 1996; Sharma et al. 2008; Muller et al. 2002).

The “master regulator” of the HSR, heat shock factor 1 (HSF1), is well-known for its role in triggering rapid expression of HSP genes (Parker and Topol 1984). HSF1 is constitutively expressed as a monomer and upon heat shock associates as a trimer and binds to inverted repeats of nGAAAn pentamers known as the heat shock element (HSE) on the promoter of the HSP70 gene (Fig. 1-4) (Westwood, Clos, and Wu 1991). Subsequently, HSF1 initiates assembly of the transcription machinery and can induce recruitment of pTET-b to alleviate promoter-proximal pol
Figure 1-4. Activation of the heat shock response by HSF1
Upon heat stress, cellular proteins unfold (purple). As a consequence, molecular chaperons in complex with HSF1 (blue) no longer form a complex, which induces trimerization of HSF1. The HSF1 trimer shuttles to the nuclear compartment where it binds to the DNA Heat Shock Element (HSE) around promoters of a heat shock genes. HSF1 binding to the HSE triggers assembly of the transcription machinery and activation of transcription on heat shock genes.
II pausing (Boehm et al. 2003; Hieda et al. 2005; Lis et al. 2000; Akerfelt, Morimoto, and Sistonen 2010). Interestingly, the inactive HSF1 monomer is kept in a complex with components of the HSP90 system. According to the widely accepted chaperone titration model, the presence of increasing numbers of unfolded proteins upon heat shock releases HSF1 from these chaperone complexes, as chaperones are required to bind unfolded proteins (Voellmy and Boellmann 2007).

A recent report additionally provided evidence that transcriptional activation of the HSR can occur independent of HSF1, although the mechanism is still unclear (Mahat, Salamanca, et al. 2016). In that report, analysis of PRO-seq density, which maps pol II positions on genes at nucleotide resolution (Mahat, Kwak, et al. 2016), indicated that HSF1 induces transcription upregulation by increasing release of promoter-proximally paused polymerases into productive elongation. Furthermore, it was shown that only hundreds of genes are upregulated while thousands are downregulated in mammalian cells in response to heat. Transcriptional repression was found to be HSF1 independent and proposed to also be achieved by limiting pol II escape from the promoter-proximal pause. In Chapter V I further show that the exonuclease activity of Xrn2 is important for proper upregulation of the heat shock response. In that Chapter, I present evidence that Xrn2 targets polymerases that have not matured into an optimal elongation complex at the promoter-proximal region for premature termination. This function of Xrn2 is proposed to prevent those polymerases from transitioning into a suboptimal elongation phase that can limit upregulation of nascent transcript synthesis.

1.9 R-Loop Structures During Transcription by RNA pol II

R-Loop structures consist of an RNA-DNA hybrid and a non-template displaced DNA strand (Fig. 1-5) (Thomas, White, and Davis 1976). Initial studies showed that these structures occur in bacteria and are a consequence of the transcription process (Drolet et al. 1995). Since
**Figure 1-5. The R-Loop structure**
During synthesis of an RNA transcript by the transcription complex (TC), the nascent transcript (blue) can anneal with the DNA template strand forming an RNA-DNA hybrid with a displaced DNA strand. The DNA duplex is shown in green and the 5’-cap of the nascent transcript is shown as a red circle. The bent arrow represents the position of an upstream transcription start site.
then, it has become clear that R-Loops are frequently generated during transcription and are present across species from bacteria to mammals (Skourti-Stathaki and Proudfoot 2014). Crystallography data showing that DNA and RNA molecules exit pol II through different channels (Westover, Bushnell, and Kornberg 2004) have supported the “thread back” model in which R-Loops form after extrusion of the nascent transcript from the active site of pol II. Although, it has also been proposed that RNA molecules in trans could also generate R-Loops in yeast (Wahba, Gore, and Koshland 2013) and humans (Nadel et al. 2015).

R-Loops are known to have important functions under physiological and pathological conditions. These structures promote class-switch recombination in activated B cells (Yu et al. 2003), they can positively influence gene expression by blocking repressive modifications on the DNA and histones (Aguilera and Garcia-Muse 2012; Skourti-Stathaki and Proudfoot 2014; Powell et al. 2013; Grunseich et al. 2018; Ginno et al. 2012) and can also contribute to transcription termination potentially by facilitating pol II pausing downstream from the polyA signal (Skourti-Stathaki et al., 2014; Skourti-Stathaki et al., 2011; Lionel A. Sanz 2016). In addition, formation of R-Loop structures on ribosomal DNA (rDNA) is conserved from yeast to humans (Nadel et al. 2015; Ginno et al. 2012; Chan, Aristizabal, et al. 2014) but their function remains unknown. Nevertheless, mounting evidence suggests that uncontrolled levels of R-loop formation induce genomic instability and disease (Groh and Gromak 2014). Their toxicity seems to originate from the increased sensitivity of the displaced DNA strand to damage (Li and Manley 2006; Aguilera and Garcia-Muse 2012) and/or conflicts with the replication fork leading to double strand breaks (Aguilera 2002; Gan et al. 2011; Houlard et al. 2011).

Different enzymes can target R-Loops in mammalian cells. These include members of the RNase H family that can degrade the RNA moiety in the RNA-DNA hybrid (Cerritelli and Crouch
2009), or helicases such as SETX and DHX9 that can unwind the RNA/DNA hybrid (Skourtis-Stathaki, Proudfoot, and Gromak 2011; Cristini et al. 2018). Additionally, recent evidence suggests that splicing factors play a critical role in limiting R-Loop formation and disease (Sorrells et al. 2018; Chen et al. 2018; Li and Manley 2005). Multiple high-risk cancer mutations in splicing factors, known to cause dissimilar splicing patterns, commonly augment R-Loop formation (Chen et al. 2018). Interestingly, a recent report arguing that a free 5’ RNA end is necessary to induce R-Loop formation around promoters (Chen et al. 2017) raises the possibility that 5’-3’ RNA exonucleolytic digestion could also play a role in elimination of R-Loops. If free 5’ RNA ends are/or become monophosphorylated, the nuclear 5’-3’ RNA exonuclease Xrn2 might contribute to elimination of R-Loops via recognition of the free 5’ end followed by 5’ to 3’ degradation of the nascent transcript. Xrn2 is a transcription termination factor that can efficiently degrade the RNA moiety in a RNA-DNA hybrid after phosphorylation by Cdk9 in vitro (Sanso et al. 2016).

Notably, in addition to specific sequences or DNA topology (Ginno et al. 2012; Duquette et al. 2004), negative superhelicity behind pol II is a major facilitator of R-Loop formation (Roy et al. 2010; Drolet, Bi, and Liu 1994). This is explained by the fact that negative superhelicity leads to a more open DNA structure that might facilitate annealing of RNA with the DNA template. To prevent this, human cells also possess topoisomerases that relax supercoiled DNA (Wang 2002). Topoisomerase I, for instance, creates a single strand break on supercoiled DNA and covalently binds the nicked 3’ end allowing controlled rotation around the intact strand. After the DNA has relaxed, topoisomerase I re-ligates the nicked strand and releases DNA (Koster et al. 2005). Interestingly, a highly effective anticancer agent called camptothecin (CPT) targets the active site of topoisomerase I and affects both its rate of uncoiling (Koster et al. 2007) and the re-ligation step leading to accumulation of R-Loops (Marinello et al. 2013). The efficacy of CPT in cancer
treatment depends on its ability to induce R-Loop formation, double strand breaks and apoptosis (Pommier 2006). Thus, maintaining R-Loop structures under control is critical for cell integrity.

Analysis using a monoclonal antibody (S9.6) specific for RNA/DNA hybrids (Boguslawski et al. 1986), and more recently catalytically dead RNase H mutant (Chen et al. 2017; Ginno et al. 2012), have revealed that R-Loops occur frequently in the yeast genome (Chan, Aristizabal, et al. 2014; El Hage et al. 2014; Wahba et al. 2016) and are also a common feature of human promoters (Ginno et al. 2012; Nadel et al. 2015; Sanz et al. 2016). It is thought that a major role of these structures around promoters is to promote transcription by blocking DNA methylation. An interesting unsolved question, however, is whether pol II transcribed genes are able to survey and control R-Loop formation to prevent their potential negative effects.

Intriguingly, CPT causes accumulation of R-Loops in nucleoli as evidenced by fluorescent microscopy staining using the S9.6 antibody (Marinello et al. 2013), but whether R-Loops around promoters are similarly affected has not been addressed. Accumulation of R-Loops around promoters upon treatment of cells with CPT has been assumed to occur based on observations that CPT treatment not only results in increased recruitment and escape of pol II from the promoter-proximal pause, but also in enhanced antisense transcription and formation of more open chromatin (Baranello et al. 2010; Capranico, Marinello, and Baranello 2010; Bertozzi et al. 2011; Ljungman and Hanawalt 1996). Nevertheless, transcription by pol II might be more resistant to accumulation of R-Loops in response to CPT treatment. A recent study showed that R-Loop formation transiently increases on a few genes upon addition of CPT to HCT116 cells but R-Loop levels at those sites rapidly decrease within the first 10 minutes of treatment (Marinello et al. 2016). In Chapter VI, I show evidence that indeed CPT treatment does not lead to accumulation of R-Loops on pol II transcribed genes. Surprisingly, I found that CPT-mediated inhibition of
topoisomerase I eliminates R-Loops from pol II transcribed genes and I further propose that mammalian cells effectively eliminate R-Loops throughout the length of pol II transcribed genes in CPT via resolution of the RNA-DNA hybrid.
CHAPTER II
MATERIALS AND METHODS

2.1 Chromatin Immunoprecipitation Coupled to Sequencing (ChIP-seq)

2.1A Chromatin Immunoprecipitation (ChIP)

ChIPs were performed as previously described (Kim et al. 2011). Cells at 70-90% confluency were crosslinked by directly adding 1% (v/v, final) formaldehyde for 7 minutes and quenched with 2.5 M glycine for 5 minutes. For ChIPs including the M12 spike-in, 2.0x10^6 mouse M12 cells were added to human cultured cells in a 150 mm dish at ~70-90% confluency right before addition of formaldehyde. Cells were harvested by centrifugation, washed twice with 1X PBS and lysed in RIPA buffer (150 mM NaCl, 1% Nonidet P-40, 0.5% deoxycholate, 0.1% SDS, 50 mM Tris pH 8.0, 5 mM EDTA). Chromatin was fragmented into ~200 bps fragments using Bioruptor Pico (Diagenode) at 4ºC (4 cycles of 10 minutes at high power, with intervals of 30 seconds on and 30 seconds off). Lysates were pre-cleared by centrifugation at maximum speed for 10 minutes and further cleared by incubation with 30 µl of unblocked Protein A-Sepharose 4B beads (Invitrogen) for 1 hour at 4ºC on nutator. For ChIPs, 30 µl of Protein A-Sepharose 4B beads (Invitrogen) in RIPA buffer were blocked with *E. coli* tRNA (10 mg/ml) and BSA (10 mg/ml) before incubation with the selected antibody at room temperature for 30 minutes. Cell lysates (~2 mgs total protein) were mixed with antibody-containing Protein-A-Sepharose 4B beads and incubated at 4ºC for at least 4 hours on nutator.

ChIPs were washed twice with RIPA buffer, four times with high-salt buffer (500 mM LiCl, 1% (v/v) Nonidet P-40, 1% (w/v) deoxycholic acid, 100 mM Tris-HCL pH 8.5), twice with RIPA buffer and twice with 1X TE buffer (10 total washes). For reverse crosslinking, washed beads in 100 µl TE buffer were added 200 µl of 1.5X Talianidis Elution buffer (70 mM Tris-HCl
pH 8.0, 1 mM EDTA pH 8.0, 1.5% (w/v) SDS), 12 µl 5 M NaCl and incubated at 65ºC overnight. Samples were treated with Proteinase K at 42ºC for 30 minutes and DNA was purified with phenol/chloroform extraction and ethanol precipitation.

2.1B ChIP-seq and Mapping

ChIP-seq libraries were prepared as previously described (Brannan et al. 2012). Sequencing libraries of input DNA fragments (2 µg) from cell lysates containing mouse M12 cells were additionally prepared. DNA fragments were end repaired at 20ºC for 30 minutes with T4 polynucleotide kinase, T4 DNA polymerase and DNA polymerase I (Klenow) in 1X ligase buffer containing ATP and 10 mM dNTPs. DNA purifications after each enzymatic reaction were made with AMPure XP SPRI beads (Beckman Coulter Genomics). A-base addition to DNA 3’ ends was performed at 37ºC for 30 minutes with DNA polymerase I (klenow, exo-minus) in 1X NEB buffer and 10 mM dATP. Ligation of the TruSeq universal adapter (TUA) was performed at 15ºC overnight with DNA ligase in 1X DNA ligase buffer and 1 mM TUA. Size selection was made by binding DNA fragments longer than 400 bps with SPRI beads solution (0.65 volumes) and transferring smaller fragments in solution into a new tube. Smaller fragments were purified by adding additional SPRI beads solution (0.35 volumes). TruSeq indexed primers were added by PCR enrichment with Phusion DNA polymerase in 1X Phusion buffer, 10 mM dNTPs, 0.6 µM Truseq indexed primer and 0.6µM PE forward primer (universal) [98ºC:30 seconds, 16x(98ºC:10 seconds, 65ºC:30 seconds, 72ºC:30 seconds), 72ºC:5 minutes].

ChIP-seq libraries were sequenced on the Illumina Hi-Seq platforms. Single-end reads were trimmed to 50 bases and mapped to the hg19 UCSC human genome (Feb, 2009). For input libraries and ChIP-seq libraries made from M12 spike-in lysates, reads mapping uniquely to the mm10 UCSC mouse genome (Dec, 2011) and hg19 UCSC human genome (Feb, 2009) were kept
for analysis. Mapping was done using Bowtie version 2.3.2 (Langmead et al. 2009). bigWig files were generated using 50 bp bins and 200 bp windows assuming a 180 bp fragment size shifting effect. For visualization with the UCSC genome browser, libraries were normalized by total number of mapped reads (RPBM: reads, per bin per million mapped reads). M12 spike-in normalization was performed as previously described (Hu et al. 2015). After normalization by total number of mapped reads, libraries were further normalized to the occupancy ratio (OR) value: OR = (Input_m \times IP_h)/(Input_h \times IP_m). Input_m = total mouse mapped reads in the input library, Input_h = total human mapped reads in the input library, IP_m = total mouse mapped reads in the IP library, IP_h = total human mapped reads in the IP library.

2.2 Antibodies

Rabbit anti-pan CTD (Schroeder et al. 2000), rabbit anti-Xrn2 (Brannan et al. 2012), mouse monoclonal anti-RNA-DNA hybrids (S9.6) (Boguslawski et al. 1986), rabbit anti-Total Spt5 (Glover-Cutter et al. 2008), rabbit anti-PO₄ Spt5 (Glover-Cutter, doctoral thesis: “Integration of Pol II Transcription”, UC Denver, 2008) have all been previously described.

2.3 Bromouridine Sequencing (Bruseq)

2.3A Isolation of Nascent Transcripts

Bruseq has been previously described (Paulsen et al. 2013). Nascent transcripts were pulse-labelled by addition of 2 mM 5-bromouridine (Invitrogen) to cell cultures and incubation for 30 minutes. Total RNA was extracted using commercial RNA mini-prep kits from Zymo Research. Samples were DNased on the column. For RNA immunoprecipitation (IP), 500 µl of cell medium from cells secreting anti-BrdU B44 antibody (Gratzner 1982) were mixed with 30 µL Protein A or Protein G Dynabeads (Invitrogen) and incubated for 1 hour at room temperature. For IPs with Protein G Dynabeads, Rat anti-Mouse IgG (Becton Dickinson) was used to immobilize the B44
antibody on the beads. Beads were washed three times with wash buffer (0.2% (v/v) triton X-100, 1X DEPC_PBS). In all steps including RNA, buffers were prepared with PBS or autoclaved water previously treated with 0.1% (v/v) diethyl pyrocarbonate (DEPC). For IP of bromo-labelled nascent transcripts, 50 µg of total RNA were incubated at 65°C for 3 minutes and put back on ice. Anti-BrdU-Dynabeads were added to the RNA sample, wash buffer was added to bring the mixture to 500 µl and incubated 1 hour at room temperature on nutator. IPs were washed five times with wash buffer. Elution of bromo-labelled RNA was done by incubation of samples at 95°C for 3 minutes in 20 µL DEPC-H₂O (~150 ng total RNA was usually obtained).

2.3B Bruseq and Mapping

For preparation of Bruseq libraries the TruSeq Stranded mRNA Sample Prep LS Protocol Kit from Illumina was used. The first steps in this protocol are designed for isolation of mRNA from total RNA. These steps were skipped and samples were treated as indicated in the protocol starting at the RNA fragmentation step. For this, sample volumes were adjusted to maintain final concentrations as shown below.

12 µl Bromo-labelled RNA were mixed with 19 µl of Fragment, Prime, Finish Mix (FPF) buffer and incubated as indicated in step “Incubate RFP” to fragment RNA: “RNA Elution 2 - Frag - Prime (94°C for 8 minutes, 4°C hold)”. 10.5 µl of First Strand Synthesis Act D Mix (FSA) and 1.5 µl RT SuperScript II were added to fragmented RNA and incubated as indicated in step “Incubate 1 CDP” for cDNA synthesis: “Synthesize 1st Strand Program” (25°C:10 minutes, 42°C:15 minutes, 70°C:15 minutes). For second DNA strand synthesis, 30 µl of Second Strand Marking Master Mix (SMM) were added and incubated as indicated in “Incubate 2 CDP”: 16°C for 1 hour. DNA fragments were purified using 130 µl of SPRI beads solution. From this point, samples were treated as indicated in the protocol starting at step “Purify CDP”.

40
Bruseq libraries were sequenced on the Illumina Hi-Seq platforms. Single-end reads were trimmed to 50 bases and mapped to the hg19 UCSC human genome (Feb, 2009). Mapping was done using Bowtie version 2.3.2 (Langmead et al. 2009). Duplicate mapped reads were excluded from Bam files with Picard tool MarkDuplicates (http://broadinstitute.github.io/picard/). Bedtools genomecov was used to create bedGraph files reporting coverage over windows of variable size. Libraries were normalized by thousand mitochondrial mapped reads for visualization with the UCSC genome browser and metagene plots using R.

2.4 5′-PO₄ Bromouridine Sequencing (5′-Bruseq)

A graphical description of the 5′-Bruseq method utilized for experiments presented in Chapter III is shown in figure 2-1. Nascent transcripts were isolated as described for Bruseq. After elution of Bromo-labelled RNA from Dynabeads in the last step, and after each enzymatic reaction below, the RNA was purified following the “Reaction Clean-up” protocol with the Direct-zol™ RNA MicroPrep kit from Zymo Research. 5′ RNA adenylation was made at 37°C for 4 hours with Mth RNA ligase in 1X 5′ DNA adenylation buffer, 0.05 mM ATP and 1 µl RiboLock (20 µl total reaction). A DNA adaptor “5pBlocked DNA-(rN)₄ RNA” (Box66-A6) containing 4 random ribonucleotides at its 3′ end and blocked at its 5′ end with the amino-modifier “5′ Amino Modifier C6” from IDT (5pDNA-RNA₄) was ligated to the 5′ pre-adenylated RNA end. Different enzymes were assayed for ligation (T4 RNA ligase 1, T4 RNA Ligase 2-truncated K227Q (NEB), 5′ AppDNA/RNA Ligase (NEB) and CircLigase RNA ligase). T4 RNA Ligase 1 and T4 RNA Ligase 2-truncated K227Q (NEB) were more efficient at ligating the 5pDNA-RNA₄ adaptor to a DppH treated, in vitro transcribed RNA, further preadenylated at its 5′ end before incubation with T4
Figure 2-1. Overview of the 5′-Bruseq procedure

Nascent transcripts are pulse-labelled by incubating cells with 5-bromouridine. Bromouridine-labelled nascent transcripts are isolated from total RNA via immunoprecipitation using a specific BrdU antibody (B44). In order to map 5’-monophosphorylated (PO₄) ends in the human genome, nascent transcripts are 5’ adenylated with Mth RNA ligase and ligated at their 5’ App RNA end with a 5’ adaptor using T4 Rnl2tr. The 5’-adaptor contains four random ribonucleotides at its 3’ end that serve as four nucleotides long molecular identifiers for suppression of PCR duplicates. Finally, RNA fragmentation and cDNA synthesis using random priming are carried out followed by amplification of the DNA fragments and high throughput sequencing.
RNA Ligase 2-truncated K227Q (NEB). Best yields of 5’-Bruseq library preparations were obtained with T4 RNA ligase 2-truncated K227Q and it was used for the experiments presented in Chapter I. Ligation reaction were performed with 0.15 µM 5pDNA-RNAN4 adaptor and T4 Rnl2tr RNA ligase in 1X T4 RNA ligase buffer with 12% PEG 8000 at room temperature for 1 hour. Ligated RNA was fragmented by incubation at 95°C for 3 minutes in 1X SuperScript II RT, 5X First-Strand Buffer (Invitrogen) containing 1 µl of 10 µM “BrUseq-cDNA-FirstStrand Reverse primer” (Box65-A1) for cDNA synthesis. Priming to the RNA was made with a random nonamer sequence at the 3’ end of the BrUseq-cDNA-FirstStrand Reverse primer and reverse transcription was made with SuperScript II RT (Invitrogen). The 5pDNA-RNAN4 adaptor and the BrUseq-cDNA-FirstStrand Reverse primer contain priming sequences for PCR amplification with the Multiplexing Index Read Sequencing Primer (MIRSP) and the TruSeq Universal Adapter (TUA) for sequencing on the Illumina Hi-Seq platforms. RNase H treatment was performed, and cDNA fragments were purified with SPRI beads. PCR amplification was done with Phusion DNA polymerase in 1X Phusion buffer, 10 mM dNTPs, 0.6 µM Truseq indexed primer Rv and 0.6 µM PE Fw (universal) [98°C:30 seconds, (98°C:10 secods, 65°C:30 seconds, 72°C:30 seconds)x16, 72°C:5 minutes].

5’-Bruseq libraries were sequenced on the Illumina Hi-Seq platforms. Adapter sequences in single-end sequenced reads were trimmed with CutAdapt. Duplicate reads with the same sequence for the first four bases added during ligation with the 5pDNA-RNAN4 adaptor were discarded and a single read was kept for mapping. Reads were further trimmed to 50 bases and mapped to the hg19 UCSC human genome (Feb, 2009). Mapping was done using Bowtie version 2.3.2 (Langmead et al. 2009). bigWig files with coverage of ligation events at nucleotide resolution were created for visualization with the UCSC genome browser. Libraries were normalized by total
number of mapped reads (RPBM: reads, per bin per million mapped reads). For metagene plots, total 5’-Bruseq coverage was normalized by thousand mitochondrial mapped reads.

2.5 DNA-RNA Immunoprecipitation Coupled to Sequencing (DRIP-seq)

DRIP-seq has been previously described (Ginno et al. 2012). Cells were washed once with 1X PBS and resuspended in lysis buffer (50 mM Tris-HCL pH 8.0, 0.7 mM EDTA, 1% w/v SDS, 200 µg/ml Proteinase K). Lysates were incubated at 37°C overnight and extracted with phenol/chloroform. Genomic DNA was precipitated with isopropanol and digested with a cocktail of restriction enzymes: BsrGI, EcoRI, HindIII, SspI and XbaI. Digested fragments were purified with phenol/chloroform extraction and precipitated with ethanol. DNA was resuspended in IP DRIP-seq buffer (10 mM sodium phosphate pH 7.0, 140 mM NaCl, 0.05% (v/v) Triton X-100) and RNA-DNA hybrids were immunoprecipitated using the S9.6 antibody (Boguslawski et al. 1986) and 30 µl Protein A Sepharose 4B beads (Invitrogen). Binding was done at 4°C overnight on nutator. IPs were washed six times with IP DRIP-seq buffer and RNA-DNA hybrids were eluted by adding 1.5 volumes of 1.5X Talianidis Elution buffer (70 mM Tris-Cl pH 8.0, 1 mM EDTA pH 8.0, 1.5% (w/v) SDS) and incubation at 65°C for 15 minutes. Samples were treated with Proteinase K at 42°C, purified with phenol/chloroform extraction and precipitated with ethanol. RNA-DNA hybrids were resuspended in 10 mM Tris-HCl, 1 mM EDTA and sonicated at high power for 15 minutes at high power with intervals of 30 seconds on 30 seconds off. Sonicated fragments were used as input in the regular protocol for preparation of ChIP-seq libraries to obtain DRIP-seq libraries. DRIP-seq libraries were mapped and analyzed as made for ChIP-seq libraries.

2.6 Engineering of pcDNA5-NS1A Inducible Flp-in TREX Cells

The NS1A cDNA sequence was PCR amplified with primers “Hind III 3XHA NS1A Fw” and “NS1A Not I Rv” which added three in-frame tandem HA tag sequences at the N-terminal of
the NS1A protein. The PCR amplicon was digested with HindIII and Not I restriction enzymes and cloned into the pcDNA5 FRT/TO vector. Site-specific integration of the pcDNA5 NS1A expression plasmid into HEK293-Flp-in T-REX cells (Invitrogen) was mediated by Flp recombinase. Flp-in TREX cells were co-transfected with the pcDNA5-NS1A plasmid (3 µg) and the pOG44 vector (3 µg) for expression of Flp-Recombinase. Lipofectamine 2000 was utilized for transfection. After 48 hours of transfection cells were selected with hygromycin B 200 µg/ml. NS1A expression was induced with doxycycline (3.0 µg/ml) and total protein was extracted and resolved by SDS-PAGE. Resolved proteins were transferred to a PVDF membrane with wet-transfer apparatus (BioRad) at 0.4 amps for 45 min. The PVDF membrane was blocked in 5% non-fat dry milk in PBST for 30 minutes. The blocked membrane was cut and separately probed with either rabbit-Xrn2 or mouse monoclonal anti-HA (12CA5) in 2% non-fat dry milk in PBST overnight. Membranes were washed with PBST and incubated with either rabbit anti-mouse Ig-HRP or swine anti-rabbit Ig-HRP in 5% non-fat dry milk in PBST for 30 minutes. Membranes were visualized by horseradish peroxidase (HRP) enhanced chemiluminescence (Thermo Fisher). Expression of NS1A was confirmed (Fig. 2-2).

2.7 CRISPR-Cas9 Genome Editing of the PSMG1 polyA Signal in Hap1 Cells

A schematic diagram of the CRISPR-Cas9 genome editing strategy is shown in figure 2-3. Oligonucleotides were designed according to the PSMG1 gene sequence around the polyA signal, synthesized (Integrated DNA Technologies), annealed and cloned into BbsI sites directly downstream of the Pol III U6 promoter in pX330-U6-Chimeric_BB-CBh-hSpCas9 vector (Feng Zhang, Addgene plasmid #42230). After integration of the annealed primers, the pX330-U6-Chimeric_BB-CBh-hSpCas9 vector directs expression of a chimeric small guide and
Figure 2-2. Expression of the flu virus NS1A protein causes cell death within 72 hours of induction

(A) Western blot showing induction of the recombinant HA-NS1A protein after incubation of HEK293 Flp-in TREX pcDNA5-NS1A cells with 3 μg/ml doxycycline (Dox) for 4 and 24 hours. Xrn2 was used as a loading control. Parent HEK293 Flp-in TREX cells were included as negative control (B) Contrast microscope images of cell cultures before and after incubation with doxycycline for 72 hrs. Note impaired growth and spherical morphology in the presence of doxycycline after 72 hrs.
Figure 2-3. Schematic diagram of the CRISPR-Cas9 genome editing strategy

Shown at the top is the PSMG1 gene target sequence for double strand break by the CRISPR-Cas9 system in Hap1 cells. The polyA signal subjected to sequence editing is shown in red in the coding strand, next to the PAM (NGG) sequence required for cleavage by the Cas9 endonuclease. The gRNA sequence is shown in green and the dotted line marks the expected site for DNA double-strand break. A donor ssDNA oligonucleotide (template strand sequence) was co-transfected with the CRISPR-Cas9 system for homology-directed repair. The donor oligonucleotide was designed to replace the polyA signal with the BamHI restriction sequence and includes an additional single nucleotide mutation in the PAM sequence to prevent multiple rounds of Cas9 cleavage (intended edits in orange). The outcome of the CRISPR-Cas9 strategy is shown at the bottom wherein bases in purple correspond to final edits.
transactivating RNA (sgRNA) transcript under the control of the human Pol III U6 promoter, as well as a human codon-optimized S. pyogenes Cas9 nuclease transcript driven by a hybrid chicken beta-actin (Cbh) promoter (Cong et al. 2013). 2 µg of PSMG1-pX330-U6-Chimeric_BB-CBh-hSpCas9 plasmid were co-transfected into Hap1 cells with 2 µg of a reporter plasmid expressing GFP-tagged histone (“pEGFP3-1”) and a single stranded DNA donor of 107 nts for homologous recombination (Fig. 2-3) (800 pmol). The design of the donor ssDNA oligo was made following as a guide (Chen et al. 2011) and (Davis and Maizels 2014). Transfection of both plasmids and ssDNA donor into Hap1 cells was done with Lipofectamine 2000 in a single well of a 6-well plate at ~70% confluency. The donor fragment contained the BamHI restriction sequence in place of the AATAAA PSMG1 hexamer sequence. GFP-positive cells were isolated by Fluorescence-Activated Cell Sorting (FACS) and seeded at low density for colony formation. Individual CRISPR-modified cell colonies were identified by PCR amplification of the PSMG1 target region and screening of PCR amplicons sensitive to BamHI digestion by gel electrophoresis. Successfully modified CRISPR cells were confirmed by sanger sequencing (Fig. 2-4) and yield a PCR amplicon from around the polyA signal of the PSMG1 gene that is sensitive to BamHI digestion (Fig. 2-5).

2.8 Doxycycline-Inducible shRNA Expression of Xrn2 and TTF2 in HEK293 Cells

For inducible knockdown of Xrn2 or TTF2 proteins, double strand DNA fragments encoding shRNA cassettes (Xrn2, 5’-CCAGTGTATTCTAGATCATCTA-3’; TTF2, 5’-ACCTGACAATGATTGCGCTCAT-3’) with pINDUCER10 sequence homology arms were designed following the guidelines a previously described (Fellmann et al. 2011). Double strand DNA fragments were cloned into the EcoRI-XhoI digested pINDUCER10 vector (Meerbrey et al. 2011) with In-Fusion cloning technology (Takara Bio Inc.). High titer lentivirus was produced by transfecting 293FT cells with Lipofectamine 2000 3 µg each pPACKH1 SBI packaging vectors
Figure 2-4. Sequence alignment reveals mutation of the PSMG1 polyA signal
PCR amplicons from the 3’ end of the PSMG1 gene in Hap1 PSMG1 pA-WT and PSMG1 pA-MT cells were sequenced by Sanger Sequencing and aligned with the Clustal Omega Multiple Sequence Alignment tool (https://www.ebi.ac.uk/Tools/msa/clustalo/). Shown is the output alignment. Note that the physiological polyA signal (shown in red) is mutated in PSMG1 pA-MT cells, which contain a BamH I restriction site (shown in green).
Figure 2-5. BamH I digestion confirms successful CRISPR-Cas9 editing around the polyA site of the PSMG1 gene in Hap1 cells
Genomic DNA was extracted from parent and CRISPR-modified PSMG1-pA-MT BamHI Hap1 cells. The PSMG1 polyA target region was amplified by PCR using specific primers and subjected to digestion with BamHI restriction enzyme. Final products were resolved by agarose gel electrophoresis. Note efficient digestion of the DNA amplicon from PSMG1-pA-MT cells. Gel shift of the PCR amplicon from the PSMG1-pA-MT cells is explained by additional nucleotide sequence insertions during editing of the PSMG1 gene.
(Open Biosystems), and 2 µg lentiviral pINDUCER10-Xrn2 or pINDUCER10-TTF2. Virus containing supernatant was added along with polybrene to HEK293 cells at ~50% confluence plated the previous day. Cells were selected in 2 µg/ml puromycin 48 hours after infection. Doxycycline-inducible knockdown of Xrn2 and TTF2 was confirmed by western blot (Fig. 2-6).

2.9 Cell Lines Growth Conditions

HEK293 shRNA, Flp-in T-REX pcDNA5 HA-Xrn2WT and HEK293 shRNA, Flp-in T-REX pcDNA5 HA-Xrn2MT (D235A) cells, previously described (Fong et al. 2015b), were maintained in DMEM medium supplemented with 10% fetal bovine serum, 1% pen/strep, 2 µg/ml puromycin and 200 µg/ml hygromycin B at 37°C and 5% CO². HEK293 Flp-in T-REX pcDNA5 HA-NS1A cells were maintained in DMEM medium supplemented with 10% fetal bovine serum, 1% pen/strep and 200 µg/ml hygromycin B at 37°C and 5% CO². Hap1 cells were maintained in IMDM medium supplemented with 10% fetal bovine serum and 1% pen/strep at 37°C and 5% CO². HEK293 pINDUCER10 Xrn2 and HEK293 pINDUCER10 TTF2 cell lines were maintained in DMEM medium supplemented with 10% fetal bovine serum, 1% pen/strep and 2 µg/ml puromycin at 37°C and 5% CO².

2.10 Real-Time quantitative (q) PCR

10 µl PCR reactions were performed in 384 well plates with iTaq™ Universal SYBR Green Supermix (Biorad) using the LC-480 instrument (Roche Applied Science). Activation was done at 95°C for 30 seconds followed by 50 cycles of 5 seconds at 95°C and 30 seconds at 60°C. Primer sets were designed using Primer-BLAST at NCBI (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). Primer melting temperatures were between 57°C and 63°C, product sizes between 70 and 120 base pairs. Primer pairs with efficiencies between 1.8 and 2.2 and error less than 0.01 were
included for analysis. For analysis of relative gene expression levels the $2^{(-\Delta\Delta C_t)}$ method was utilized (Livak and Schmittgen 2001).
Figure 2-6. Inducible knockdown of Xrn2 and TTF2 in HEK293 cells
Western blot showing gradual depletion of Xrn2 and TTF2 in HEK293 cells containing the pINDUCER10 lentiviral system. Cells were treated with doxycycline (Dox) (3 $\mu$g/mL) for induction of shRNAs against Xrn2 (shXrn2) and TTF2 (shTTF2) and cell lysates were prepared for analysis at the indicated times. As negative control induction of a scrambled (Scr) shRNA was included. Actin was used as a loading control.
CHAPTER III
INTEGRATION OF XRN2 INTO THE TRANSCRIPTION CYCLE
OF RNA POLYMERASE II

3.1 Introduction

RNA polymerase II (pol II) must overcome multiple barriers throughout transcription in order to produce a full-length mRNA molecule. These barriers are alleviated by different factors known to be strategic targets of potent inhibitors of transcription. First, the TFIIH-associated helicase XPB harbors an ATPase activity that is thought to contribute to DNA melting in the pre-initiation complex, although its presence in the pre-initiation complex might not be absolutely required for transcription (Lin, Choi, and Gralla 2005; Alekseev et al. 2017). This helicase is the target of a potent inhibitor of transcription initiation called triptolide (Titov et al. 2011). The next major barrier is a strong pause encountered after transcription of the first 30 to 70 nucleotides imposed by two factors: the negative elongation factor NELF and the DRB (5,6-dichloro-1-β-D-ribofuranosylbenzimidazole) sensitivity-inducing factor (DSIF), composed by two subunits called Spt5 and Spt4 (Yamaguchi, Wada, et al. 1999; Wu et al. 2003). DSIF was initially identified genetically in a suppressor screen in \emph{S. cerevisiae} (Winston et al. 1984) and later found to be involved in transcription elongation (Hartzog et al. 1998). DSIF was the factor conferring sensitivity of pol II to the drug DRB (Wada, Takagi, Yamaguchi, Ferdous, et al. 1998), a nucleoside analog that inhibits the kinase activity of the positive transcription elongation factor b (pTEF-b) (Mancebo et al. 1997). It turns out that DSIF and NELF induce a strong pause of the elongation complex that can be alleviated by the kinase activity of pTEF-b (Fujinaga et al. 2004; Wada, Takagi, Yamaguchi, Watanabe, et al. 1998). It is thought that pTEF-b phosphorylates pol
II and Spt5, causing release of NELF and pol II escape from the pause (Kim and Sharp 2001; Yamaguchi, Wada, et al. 1999).

Once pol II embarks on elongation of the whole gene body, it faces potential road blocks such as nucleosomes, DNA-binding proteins or accumulating DNA superhelicity. To efficiently traverse gene bodies, pol II is not only loaded with chromatin remodeling factors (Kulaeva et al. 2013; Workman 2006) but is also supported by topoisomerases that relax supercoiled DNA (Wang 2002). Topoisomerase I, for instance, creates a single strand break on supercoiled DNA and covalently binds the nicked 3’ end allowing controlled rotation around the intact strand. After the DNA has relaxed, topoisomerase I re-ligates the nicked strand and releases DNA (Koster et al. 2005). A widely used anticancer drug, CPT, is also able to inhibit transcription elongation by targeting the active site of topoisomerase I, blocking the final ligation step and transcription elongation (Li et al. 2017; Hertzberg, Caranfa, and Hecht 1989; Capranico et al. 2007; Ljungman and Hanawalt 1996).

When pol II arrives at 3’ ends of genes transcription must terminate. Failure in termination of transcription can lead to transcription interference with downstream genes or production of potentially harmful intergenic transcripts (Shearwin, Callen, and Egan 2005; Greger and Proudfoot 1998). According to the torpedo model, cleavage downstream of the polyA signal allows the 5’-3’ exonuclease Xrn2/Rat1 (human/yeast) to degrade the nascent transcript and chase pol II (Fig. 1-2). It is proposed that catching-up of Xrn2 with the elongation complex induces disassembly by a yet unknown mechanism (Connelly and Manley 1988a; West, Gromak, and Proudfoot 2004). Consistent with this model, it has been shown that anti-Xrn2 ChIP-seq density accumulates at both 3’ and 5’ ends of genes (Brannan et al. 2012). Xrn2-independent mechanisms have also been suggested. At least in Caenorhabditis elegans, two distinct transcription termination modes
dictated by promoters have been proposed (Miki, Carl, and Grosshans 2017). In that study, it was observed that although a subset of genes requires Xrn2, termination of transcription on other genes that additionally recruit Xrn2 to their 3’ ends is not affected in Xrn2-deficient *C. elegans* strains. It was also reported that promoters instruct the choice of termination mode and that Xrn2-independent termination additionally requires a compatible region downstream from the 3’ end cleavage site.

Consistent with the idea that alternative Xrn2-independent termination mechanisms may operate at 3’ ends of genes, depletion of Xrn2 in some studies have not uncovered a noticeable termination defect (Brannan et al. 2012; Nojima et al. 2015b). Nevertheless, the Bentley lab recently reported that inhibition of the exonuclease activity of Xrn2 does lead to accumulation of polymerases downstream of polyA sites genome-wide consistent with a termination defect (Fong et al. 2015b). According to the torpedo model, Xrn2 chases pol II via degradation of the nascent transcript and induces termination upon catching up with the elongation complex. Thus, in this model, the rate of extension of the nascent transcript in the 3’ direction by pol II is in competition with the rate of 5’-3’ exonucleolytic digestion by Xrn2 and must be overcome. In agreement with a kinetic competition, in the latter report it was shown that a slow pol II terminates earlier while a fast pol II terminates later.

Interestingly, in the study by Brannan, Kim et al. 2012 depletion of Xrn2 did not generate a termination defect at 3’ ends of genes, but a strong change in the distribution profile of pol II within gene bodies was revealed after depletion of either Xrn2 or decapping factors. This change in the distribution profile of pol II was characterized by an increase in pol II density within gene bodies relative to pol II density around promoters and it was interpreted as an increase in the flux
of pol II from promoters. It was proposed that Xrn2 functions in global premature termination of nascent transcripts that are decapped co-transcriptionally.

In agreement with Brannan, Kim et al. 2012, the study by Nojima, Gomes et al. 2015 showed that depletion of Xrn2 with siRNAs in mammalian cells resulted in increased levels of promoter-associated RNA 3’ ends thought to originate by protection of pol II at its active site bound to the DNA template. Additionally, Xrn2 was proposed to limit transcription on the HIV promoter after endonucleolytic cleavage by Microprocessor at the Transactivator Response Element (TAR) (Wagschal et al. 2012). Altogether, these observations indicated that Xrn2 might be involved in premature termination of transcription. In Chapter V, I further provide strong evidence that Xrn2 targets promoter-proximally paused polymerases for termination and that its exonuclease activity is required for such function.

How Xrn2 is recruited to promoters is an intriguing question that remains unsolved. Search for interacting proteins using mass spectrometry analysis of Xrn2 immunoprecipitates from HeLa nuclear extract revealed splicing and cleavage-polyadenylation factors but not pol II (Brannan et al. 2012), suggesting that Xrn2 does not directly contact the elongation complex. In that approach RNase treatment of nuclear extract was performed before immunoprecipitation, which eliminated potential RNA-mediated tethering of Xrn2 to the elongation complex. One system in which Xrn2 appears to be recruited to the promoter region via interactions with the pre-mRNA appears to be the HIV promoter. Recruitment of Xrn2 to the HIV promoter was shown to be facilitated by the endonuclease Microprocessor, which directly binds the RNA stem-loop known as the trans-activator response element (TAR) (Wagschal et al. 2012). On the other hand, Xrn2 co-immunoprecipitates with decapping factors and the termination factor TTF2, both shown to
localize at 5’ ends of genes (Brannan et al. 2012). Thus, multiple anchoring points seem to link Xrn2 to promoters.

Significantly, treatment of cell with DRB, which inhibits pol II release from the promoter proximal pause, restricted accumulation of both pol II and Xrn2 to the 5’ end of the GAPDH gene by ChIP-qPCR (Brannan et al. 2012). This result suggests an intimate association between Xrn2 and pol II around promoters. Nevertheless, it is still unknown whether Xrn2 associates with the elongation complex and, if so, at which step during transcription. Here, I utilized the transcriptional inhibitors mentioned above to prevent progression of pol II at different stages during transcription and evaluate recruitment of Xrn2. The results presented here demonstrate that Xrn2 is indeed recruited to promoter regions early after transcription initiation and becomes intimately associated with the elongation complex until 3’ ends of genes. Furthermore, inhibition of the exonuclease activity of Xrn2 led to genome-wide stabilization of 5’-monophosphorylated nascent transcripts generated by cleavage at polyA sites. This result additionally confirms that once Xrn2 reaches the end of human genes, it targets the nascent transcript for degradation and chases pol II as indicated by the torpedo model of transcription termination.

3.2 Results

3.2A Recruitment of Xrn2 Around Promoters does not Require Synthesis of a pre-mRNA Transcript

The ChIP-seq distribution profile of Xrn2 normally shows strong densities at both the promoter and termination regions as shown on the RPL23 gene (fig. 3-1B). Nevertheless, the resolution offered by ChIP-seq on steady state pol II precludes identification of the precise site of recruitment during transcription. To resolve this, I adopted a strategy consisting in the use of a variety of inhibitors to block transcription at different steps and ask if the distribution profile of Xrn2 follows the effects observed for pol II, suggestive of an association between Xrn2 and pol II.
Figure 3-1. Xrn2 is recruited to the transcription start site of the RPL23 gene

HEK293 cells were treated with different transcription inhibitors: triptolide (10 µM, 10 min), DRB (100 µM, 1 hr) and flavopiridol (300 nM, 1 hr). Dimethyl sulfoxide (DMS) was included as vehicle control. (A) pol II ChIP-seq density on the RPL23 gene. (B) Xrn2 ChIP-seq density on the RPL23 gene. Note persistence of Xrn2 signal around the promoter region of the RPL23 gene after inhibition of transcription initiation with triptolide or after inhibition of pol II escape from the promoter-proximal pause with DRB or Flavopiridol (indicated by red arrows). A biological replicate of Xrn2 ChIP-seq density under triptolide conditions is shown at the genome-wide level in figure 3-3. UCSC genome browser screenshots. Blue arrow indicates direction of transcription.
First, HEK293 cells were treated with triptolide for 10 minutes to inhibit transcription initiation. As expected, pol II density was rapidly lost within the body region of the RPL23 gene but could still be detected at its 3’ end (fig. 3-1A). Interestingly, triptolide treatment still allowed for the presence of both pol II and Xrn2 ChIP-seq peaks at the 5’ end of the RPL23 gene (fig. 3-1). To visualize effects of triptolide treatment on pol II distribution at the genome-wide level, I selected genes with the strongest shift in pol II density from the promoter-proximal region to the transcription start site (TSS) after triptolide treatment for 10 minutes. Genes with the greatest increase in the ratio of pol II density around the TSS (-50 nts to +50 nts from the TSS) over pol II density around the promoter-proximal region (+50 nts to +200 nts from the TSS) upon triptolide treatment were selected for analysis (n=3000). As observed on the RPL23 gene, this group of genes shows that pol II density is still present around promoters under triptolide conditions. Interestingly, pol II density shifts from the promoter-proximal region to the TSS upon treatment with triptolide (Fig. 3-2A). This result is interpreted as an accumulation of inhibited pre-initiation complexes at the TSS accompanied with clearance of polymerases that have already initiated transcription from promoter regions. Importantly, Xrn2 density persisted around promoters after treatment with triptolide (Fig 3-2B). A biological replicate of Xrn2 accumulation around promoters under triptolide conditions is shown in Figure 3-3. Thus, localization of Xrn2 around promoters does not require that a nascent transcript be made by pol II. Such observation suggests that recruitment of Xrn2 to promoter regions does not require interactions with the pre-mRNA.

3.2B Xrn2 Associates with Promoter-Proximally Paused Elongation Complexes

Next, I inhibited the transition of promoter-proximally paused polymerases into elongation by treating cells with the elongation inhibitor DRB (1 hour), which targets pTEF-b. Human promoters are intrinsically bidirectional, which means that antisense transcription is frequently
Figure 3-2. Xrn2 is recruited to the transcription start site of gene promoters genome-wide
HEK293 cells were treated with triptolide (10 μM, 10 min), DRB (100 μM, 1 hr) or flavopiridol (300 nM, 1 hr). Dimethyl sulfoxide (DMSO) was included as vehicle control. Genes with the strongest shift in pol II density toward the TSS after triptolide treatment (n=3000) were aligned at the transcription start site (TSS) (see text). (A) pol II ChIP-seq density before and after inhibition of transcription initiation with triptolide. (B) Xrn2 ChIP-seq density before and after inhibition of transcription initiation with triptolide normalized relative to signal at -1 kb upstream of the TSS. (C) pol II ChIP-seq signal before and after inhibition of pol II escape from the promoter-proximal pause with DRB. (D) Xrn2 ChIP-seq density in the presence of the different transcription inhibitors normalized relative to signal at -1 kb upstream of the TSS. Note strong accumulation of Xrn2 ChIP-seq density at the pol II pausing site under DRB and Flavopiridol conditions (indicated by red arrows). The pol II promoter-proximal pausing site is marked by the grey dotted line. A biological replicate of Xrn2 ChIP-seq density under triptolide conditions is shown in figure 3-3. Rel. Fr. (Relative Frequency of the Mean), Mean (Reads per bin per million mapped reads).
Figure 3-3. Xrn2 is recruited to the transcription start site of gene promoters genome-wide – Biological replicate
HEK293 cells were treated with triptolide (10 µM, 10 min) to inhibit transcription initiation. Dimethyl sulfoxide (DMSO) was included as vehicle control. Shown is the mean distribution profile of Xrn2 ChIP-seq density normalized relative to signal at -1 kb upstream of the TSS. Biological replicate of figure 3-2 (n=3000 genes). Mean (Reads per bin per million mapped reads).
generated upstream from the promoter in opposite direction to sense transcription (Duttke et al. 2015). While transcription in the antisense direction is not stable, pol II similarly accumulates with pausing factors at the promoter-proximal region upstream of the transcription start site as evidenced in mouse Embryonic Stem Cells (mESC) (Flynn et al. 2011). Consistent with pausing of pol II in both directions, pol II density displayed a bimodal peak around promoters that was strengthened by inhibition of pol II escape from the pause with DRB genome-wide (Fig. 3-2C), also seen on the promoter of the RPL23 gene (Fig. 3-1A). Importantly, Xrn2 and pol II density were lost at the 3’ end of the RPL23 gene, which demonstrates that recruitment of Xrn2 at the 3’ end of the RPL23 gene requires the presence of an elongation complex at that site. Additionally, accumulation of Xrn2 around promoters under DRB conditions was strongest at the pol II pausing site on thousands of genes (indicated by red arrows) (Fig. 3-2D). Strong accumulation of Xrn2 density at the pol II pausing site was reproduced by inhibition of pTEF-b with an independent drug, flavopiridol (Fig. 3-1B, 3-2D) (Chao and Price 2001). I interpret this result to suggest that Xrn2 associates with promoter-proximally paused elongation complexes.

3.2C Xrn2 Accompanies the Elongation Complex from Promoters to 3’ Ends of Genes

The fact that Xrn2 is recruited to promoter regions prompted the idea that Xrn2 might reach 3’ ends of genes via association with the elongation complex from promoter regions. To investigate this possibility, I performed two independent experiments. First, I tested the prediction that an increase in pol II density within gene bodies would be accompanied by accumulation of Xrn2. For this, I used CPT to inhibit topoisomerase I and slow down elongation complexes within gene bodies (Collins, Weber, and Levens 2001; Ljungman and Hanawalt 1996). As expected, CPT treatment generated an increase in pol II occupancy within the body region of the GAPDH gene observed by anti-pol II ChIP-qPCR (Fig. 3-4). Additionally, pol II occupancy was reduced around
Figure 3-4. Camptothecin causes pile-up of polymerases on the body region of the GAPDH gene and reduces pol II occupancy around its promoter

HEK293 cells were treated with camptothecin (CPT) (10 μM, 30 min) to inhibit transcription elongation. Dimethyl sulfoxide (DMSO) was included as vehicle control. (A) Anti-pol II ChIP-qPCR signal using primer pairs targeting different positions on the GAPDH gene. Note increase pol II occupancy within the gene body at positions 1334, 2436 and 3420. (B) Map of the GAPDH gene with positions of primer pairs used for anti-pol II ChIP-qPCR analysis. The bent arrow marks the direction and position of the transcription start site. The downward arrow indicates the position of the polyadenylation site (pA). Error bars correspond to the standard deviation of technical replicates from a single experiment (n = 3).
the GAPDH promoter, consistent with data suggesting that accumulated negative superhelicity of the DNA template caused by CPT facilitates escape of paused pol II around promoters (Fig. 3-4) (Capranico et al. 2007). In agreement with the idea that Xrn2 associates with the elongation complex early during transcription, CPT generated accumulation of both pol II and Xrn2 density within the body region of thousands of genes (Fig. 3-5 – 3-7).

To test for continuous association of Xrn2 with the elongation complex throughout transcription, I turned to an alternative strategy. Paused polymerases were released into elongation after 1 hour of DRB treatment to generate a synchronous wave of pol II from promoters. ChIP-seq analysis was performed at time points (T) 0, 10 and 20 minutes after removal of DRB from cultured cells. Significantly, a front wave of pol II tracked closely with a wave of Xrn2 after release of paused polymerases (Fig. 3-8 – 3-10). On the NEAT1 and ACTB genes, pol II and Xrn2 density was restricted to the promoter region at T₀, but both pol II and Xrn2 migrated into the body region at T₁₀ with similar distribution profiles (Fig. 3-8, 3-9). At T₂₀, both pol II and Xrn2 reached the 3’ end of these genes. A global wave of pol II and Xrn2 tracking together could be seen at the genome-wide level (Fig. 3-10). Taken together, these data demonstrate that Xrn2 is recruited to promoter regions and associates with the elongation complex throughout the length of genes. I interpret these results to suggest that Xrn2 migrates with elongation complexes from promoter regions to 3’ ends of genes. This migration of Xrn2 molecules from promoters can occur either by permanent association of Xrn2 with the elongation complex from promoters to 3’ ends of genes, or by continuous events of binding and release of Xrn2 molecules on the elongation complex during elongation.
Figure 3-5. Pol II and Xrn2 accumulate on the body region of the EIF1 and ACTB genes after inhibition of transcription elongation with camptothecin
HEK293 cells were treated with the camptothecin (CPT) (10 µM for 30 min) to inhibit transcription elongation. Dimethyl sulfoxide (DMSO) was included as vehicle control. (A) Pol II and Xrn2 ChIP-seq density on the ACTB gene. (B) Pol II and Xrn2 ChIP-seq density on the EIF1 gene. Note accumulation of both pol II and Xrn2 density within the body regions of the ACTB and EIF1 genes under CPT conditions when compared to DMSO control (indicated by red arrows). UCSC genome browser screenshots. Blue arrow indicates the direction of transcription.
Figure 3-6. Camptothecin inhibits transcription elongation genome-wide
HEK293 cells were treated with the camptothecin (CPT) (10 µM for 30 min) to inhibit transcription elongation. Dimethyl sulfoxide (DMSO) was included as vehicle control. The relative distribution profile of anti-pol II ChIP-seq density is shown on most active genes (Top 25% with highest pol II density within the body region in normal conditions). Genes were aligned at the transcription start site (TSS) and the cleavage and polyadenylation site (CPS). Additionally, the body region between 0.5 kb downstream of the TSS and 0.5kb upstream of the CPS was normalized to gene size. Note global accumulation of pol II density within gene bodies relative to distal positions under CPT treatment when compared to DMSO control (indicated by a red arrow).
**Figure 3-7. Camptothecin induces global accumulation of Xrn2 on gene bodies**

HEK293 cells were treated with the camptothecin (CPT) (10 μM for 30 min) to inhibit transcription elongation. Dimethyl sulfoxide (DMSO) was included as vehicle control. The relative distribution profile of anti-Xrn2 ChIP-seq density is shown on most active genes (Top 25% with highest pol II density within the body region in normal conditions). Genes were aligned at the transcription start site (TSS) and the cleavage and polyadenylation site (CPS). Additionally, the body region between 0.5 kb downstream of the TSS and 0.5kb upstream of the CPS was normalized to gene size. Note global accumulation of Xrn2 density within gene bodies relative to distal positions under CPT treatment when compared to DMSO control (indicated by red arrows).
Figure 3-8. Xrn2 accompanies the elongation complex from the promoter region to the 3’ end of the NEAT1 gene

HEK293 cells were treated with DRB (100 µM, 2.5 hrs) to inhibit escape of pol II from the promoter-proximal pause. Polymerases were released into elongation by DRB wash and ChIP-seq was performed at time points (T) T₀, T₁₀ and T₂₀ minutes after DRB wash. Shown are the distribution profiles of pol II and Xrn2 ChIP-seq densities on the NEAT1 gene. Note that the front wave of pol II is accompanied by a wave of Xrn2 from the promoter to the 3’ end of the NEAT1 gene (illustrated by red arrows). UCSC genome browser screen shot. Blue arrow indicates direction of transcription.
Figure 3-9. Xrn2 accompanies the elongation complex from the promoter region to the 3’ end of the EIF1 gene

HEK293 cells were treated with DRB (100 µM, 2.5 hrs) to inhibit escape of pol II from the promoter-proximal pause. Polymerases were released into elongation by DRB wash and ChIP-seq was performed at time points (T) T₀, T₁₀ and T₂₀ minutes after DRB wash. Shown are the distribution profiles of pol II and Xrn2 ChIP-seq densities on the EIF1 gene. Note that the front wave of pol II is accompanied by a wave of Xrn2 from the promoter to the 3’ end of the EIF1 gene (illustrated by red arrows). UCSC genome browser screen shot. Blue arrow indicates direction of transcription.
Figure 3-10. Xrn2 accompanies the elongation complex from promoters to 3’ ends of genes genome-wide
HEK293 cells were treated with DRB (100 µM, 2.5 hrs) to inhibit escape of pol II from the promoter-proximal pause. Polymerases were released into elongation by DRB wash and ChIP-seq was performed at time points (T) T₀, T₁₀ and T₂₀ minutes after DRB wash. Shown are the mean distribution profiles of anti-pol II and anti-Xrn2 ChIP-seq densities normalized to signal at -1 kb upstream of the TSS. Short genes (1 kb to 6.5 kb) were aligned at the transcription start site (TSS). A wave of Xrn2 could not be detected on longer genes due to low signal to noise ratio for Xrn2 ChIP-seq libraries. A meta-plot on long genes (33 to 78 kb long) is shown in Appendix A (Fig. A-1).
3.2D Xrn2 Degrades Nascent Transcripts Downstream and Upstream of polyA Sites

The results above illustrate clearly that Xrn2 arrives to the end of genes associated with the elongation complex. According to the torpedo model, Xrn2 must recognize the 5’-monophosphorylated end of the nascent transcript for degradation after cleavage at the polyA site to chase pol II and induce termination. Nevertheless, it has not yet been demonstrated that Xrn2 targets the nascent transcript for degradation at the polyA site to chase pol II in vivo. I set out to confirm this essential aspect of the torpedo model. For this, I analyzed levels of nascent transcripts, as well as levels of 5’-monophosphorylated ends (5’ PO₄s) of nascent transcripts at polyA sites after inhibition of the exonuclease activity of Xrn2 genome-wide. For analysis of nascent transcripts I employed Bromouridine Sequencing (Bruseq) (Paulsen et al. 2014). In this method nascent transcripts are pulsed-labelled with bromouridine and isolated via immunoprecipitation with an anti-BrdU (B44) specific antibody (Gratzner 1982). Subsequently, nascent transcripts are processed into sequencing libraries for mapping to the genome. Note that Bruseq probes for active transcription since nascent transcripts must be extended to become labelled with bromouridine.

A previous report by the Bentley lab showed that expression of dominant negative, D235A mutant (MT) Xrn2MT causes a pol II termination defect under stable expression of an shRNA targeting the endogenous Xrn2 (Fong et al. 2015b). In that report it was shown that pol II transcribes farther downstream of polyA sites after expression of Xrn2MT when compared to control cells expressing wild-type (WT) Xrn2WT (also shown in Chapter IV, Fig. 4-1). Here I utilized the same cell lines and approach to inhibit the exonuclease activity of Xrn2 and assess levels of nascent transcripts (and 5’ PO₄s described in the section below). Localization of Xrn2 is
restricted to the nuclear compartment (Krzyszton et al. 2012), therefore mitochondrial signal is expected to serve as an ideal internal control.

Bruseq coverage was normalized to mitochondrial reads and signal from Xrn2WT and Xrn2MT cells were overlaid on the same track. As predicted by the torpedo model, inhibition of the exonuclease activity of Xrn2 resulted in an increase in the levels of nascent transcripts downstream of the polyA sites of the CYR61 and JUN genes (Fig. 3-11). Note that nascent transcript signal continues farther away from the polyA site in Xrn2MT cells relative to control Xrn2WT cells (indicated by red arrows) (Fig. 3-11). Two independent demonstrations of this effect are shown. This result strongly suggests that Xrn2 targets the nascent transcript for degradation at 3’ ends of genes consistent with the idea that Xrn2 chases pol II for termination. A global increase in nascent transcript signal was also consistently detected at 3’ ends of genes (Fig. 3-13). Nevertheless, inhibition of Xrn2 generated a modest increase in nascent transcript signal. It is possible that inhibition of Xrn2 might be compensated by additional termination mechanisms. At least in *C. elegans* two distinct transcription termination modes dictated by promoters have been proposed (Miki, Carl, and Grosshans 2017).

Interestingly, inhibition of Xrn2 similarly caused an increase in nascent transcript signal downstream of the 3’ end of histone genes (Fig. 3-12). This result is in contrast with a previous report arguing that Xrn2 plays no significant role in termination of histone genes using conditional depletion of Xrn2 (Eaton et al. 2018), but it is consistent with the observation that pol II terminates later at these sites in these Xrn2MT cells (Fong et al. 2015b). Surprisingly, inhibition of Xrn2 also caused elevation of nascent transcript signal within gene body regions (Fig. 3-13), suggesting that Xrn2 targets nascent transcripts for premature termination. Further support for this interpretation is the fact that Xrn2 was found to associate with the elongation complex early during transcription,
Figure 3-11. Xrn2 targets the nascent transcript for degradation downstream of the polyA site of the EIF1, CYR61 and JUN genes

Nascent transcripts were isolated from HEK293 Flp-in TREX cells expressing dominant negative, mutant (MT) (D235A) Xrn2MT or wild-type (WT) control Xrn2WT. Shown are overlaid Bruseq profiles of Xrn2WT and Xrn2MT cells on the EIF1, CYR61 and JUN genes. The overlap of Bruseq signals is shown in black. Bruseq coverage was normalized to thousand mitochondrial mapped reads. The blue dotted line marks the position of the polyA site. Note strong increase in nascent transcripts downstream of the polyA site in Xrn2MT cells indicated by red arrows. Also note that stronger red signal (from Xrn2MT cells) is more strongly observed than blue signal (from Xrn2WT cells) above the overlap upstream of the polyA site. Two biological replicates are shown. USCS genome browser screen shots. Blue arrow indicates the direction of transcription.
Figure 3-12. Xrn2 targets the nascent transcript for degradation downstream of the 3’ end of histone genes

Nascent transcripts were isolated from HEK293 Flp-in TREX cells expressing dominant negative, mutant (MT) (D235A) Xrn2MT or wild-type (WT) control Xrn2WT. Shown is the Bruseq coverage normalized to thousand mitochondrial mapped reads on HIST1H3B, HIST1H4E and HIST1H1C histone genes. Two biological replicates are shown. The blue dotted line marks the position of the polyA site. Note increased signal of nascent transcripts downstream of the polyA site after inhibiton of Xrn2 (indicated by red arrows). USCS genome browser screen shots. Blue arrow indicates the direction of transcription.
HEK293 Flp-in TREX cells were treated with doxycycline for 24 hrs to induce expression of wild-type (WT) Xrn2WT or exonuclease-dead, mutant (MT) (D235A) Xrn2MT. Most active genes (top 25% with highest pol II ChIP-seq density in normal conditions) were aligned at the transcription start site (TSS) and the cleavage and polyadenylation site (CPS). Additionally, the body region between the TSS and CPS was normalized to gene length. Shown is the distribution profile of Bruseq signal (nascent transcript coverage) for two biological replicates (Run 1 and Run 2) normalized to mitochondrial reads. Note a slight, consistent increase in nascent transcript signal from the TSS up to +3 kb downstream of the CPS after inhibition of Xrn2. Also, note a small increase in nascent transcript signal in the antisense direction upon inhibition of Xrn2. N.M.: mean of total coverage per bin, further normalized per thousand mitochondrial reads.
as shown above. Indeed, strong evidence that Xrn2 functions in global premature termination of transcription around promoters is presented in Chapter V. In conclusion, I interpret the global increase in Bruseq signal upon inhibition of Xrn2 to suggest stabilization of nascent transcripts that are targets of Xrn2 for transcription termination.

3.2E 5′-Bruseq Detects 5′-Monophosphorylated Ends of Nascent Transcripts

To map 5′ PO₄s I developed 5′ phosphate bromouridine sequencing (5′-Bruseq) (Fig. 2-1). In this approach 5′ PO₄s of bromouridine-labelled nascent transcripts are detected by ligation to an adaptor using RNA ligase. High throughput sequencing and mapping of sequence tags can then inform on the identity of the ligation events. Nascent transcripts with a cap structure or 5′ ends other than a 5′ PO₄ are not competent for ligation in this method.

Preliminary inspection of 5′-Bruseq signal confirmed detection of bona fide 5′-PO₄ RNA ends. The 5′ ends of mature tRNAs are generated as a result of RNase P cleavage (Kirsebom 2007) and their 3′ ends are defined by RNase Z cleavage, which generates a 5′ monophosphate on the trailer transcript (Spath, Canino, and Marchfelder 2007). Both 5′ PO₄ ends can be readily detected by this assay (Fig. 3-14). Alignment of tRNAs at their 3′ end revealed two strong 5′-Bruseq peaks (indicated by red arrows), one at their 3′ end and one 75 nts upstream of their 3′ end consistent with their short length. Similarly, the mature 5.8S, 18S and 28S ribosomal RNAs are processed from a single primary transcript by the concerted action of a variety of nucleases (Fernandez-Pevida, Kressler, and de la Cruz 2015). Interestingly, mapping of 5′-Bruseq libraries to full human rDNA repeat (GenBank accession no. U13369) revealed signal of 5′ PO₄ ends at the annotated start site of each of the 5.8S, 18S and 28S rRNA species (Fig. 3-15). A biological replicate of 5′-Bruseq signal on the rDNA repeating unit is shown in figure 3-16. Furthermore, the 5′ PO₄ ends of miRNAs, snRNAs and snoRNAs are also strongly detected (Fig. 3-17D, 3-18, 3-21B-D). In
Figure 3-14. 5′-Bruseq detects 5′ PO₄ ends of tRNAs and the 5′ PO₄ product of endonucleolytic cleavage by RNase Z at their 3′ end

Nascent transcripts were isolated from HEK293 Flp-in TREX cells expressing dominant negative, mutant (MT) (D235A) Xrn2MT or wild-type (WT) control Xrn2WT. 5′ PO₄ ends of nascent transcripts were probed with 5′-Bruseq. (A) 5′-Bruseq signal on hg19 TRNA Leu: chr14:21093529-21093610. (B) 5′-Bruseq signal on hg19 TRNA Gly: chr1:16872434-16872504. (C) Mean distribution profile of sense 5′-Bruseq signal on tRNAs aligned at their 3′ end. Red arrows point to peaks of 5′ PO₄s ends generated by RNase P and RNase Z cleavage of nascent tRNA transcripts. N.M.: mean of total number of reads per bin (40 bins, 5nts each), further normalized per million mitochondrial reads (N.M.). RPNM: Reads per nucleotide per thousand mitochondrial mapped reads.
Figure 3-15. Xrn2 targets 5’ PO₄s ends of nascent transcripts at the 5’ and 3’ end of the pre-rRNA transcription unit

Nascent transcripts were isolated from HEK293 Flp-in TREX cells expressing dominant negative, mutant (MT) (D235A) Xrn2MT or wild-type (WT) control Xrn2WT. 5’ PO₄ ends of nascent transcripts were mapped to a full human rDNA repeat (GenBank accession no. U13369). Shown is the distribution profile of sense 5’-Bruseq signal on the rDNA repeating unit. Note increase in 5’ PO₄s at the beginning of the rDNA unit (indicated by red arrow) relative to a downstream peak of 5’ PO₄s (indicated by blue arrows). Also, note increase in 5 PO₄s closely downstream of the 28S rRNA in Xrn2MT cells (indicated by green arrow). 5’-Bruseq coverage was plotted at nucleotide resolution and normalized to million mapped reads. A biological replicate of 5’-Bruseq signal on the rDNA unit is shown in figure 3-16. No substantial 5’-Bruseq signal was detected in the antisense direction. RPNM: Reads per nucleotide per million.
Figure 3-16. Xrn2 targets 5’ PO4s ends of nascent transcripts at the 5’ and 3’ end of the pre-rRNA transcription unit – Biological Replicate
Nascent transcripts were isolated from HEK293 Flp-in TREX cells expressing dominant negative, mutant (MT) (D235A) Xrn2MT or wild-type (WT) control Xrn2WT. 5’ PO4 ends of nascent transcripts were mapped to a full human rDNA repeat (GenBank accession no. U13369). Shown is the distribution profile of sense 5’-Bruseq signal on the rDNA repeating unit. Note increase in 5’ PO4s at the beginning of the rDNA unit (indicated by red arrows) relative to a downstream peak of 5’ PO4s (indicated by blue arrows). Also, note increase in 5 PO4s closely downstream of the 28S rRNA in Xrn2MT cells (indicated by a green arrow). 5’-Bruseq coverage was plotted at nucleotide resolution and normalized to million mapped reads. RPNM: Reads per nucleotide per million.
Figure 3-17. Xrn2 targets 5’ PO₄s of nascent transcripts generated via different pathways

HEK293 Flp-in TREX cells were treated with doxycycline for 24 hrs to induce expression of wild-type (WT) Xrn2WT or exonuclease-dead, mutant (MT) (D235A) Xrn2MT. (A) 5’-Bruseq signal on the EIF1 gene. (B) 5’-Bruseq signal on the NEAT1 gene. (C) 5’-Bruseq signal on the histone gene HIST2H4A. (D) 5’-Bruseq signal around the intronic snoRNA SNORD4B. Note accumulation of 5’ PO₄ ends that are generated via different mechanism after inhibition of Xrn2 (indicated by red arrows). RNA cleavage by CPSF73 generates a 5’ PO₄ end on the nascent transcript at the polyA site of the EIF1 gene. RNA cleavage by RNase P produces a 5’ PO₄ end on the nascent transcript at the 3’ end of the NEAT1 gene. polyA-independent cleavage by CPSF-73 generates a 5’ PO₄ end on the nascent transcript at 3’ ends of histone genes. Finally, 5’ PO₄ ends at the intron start site most likely correspond to intron debranching events.

RPNM: Reads per nucleotide per million mapped reads. UCSC genome browser screenshots. Blue arrow indicates direction of transcription.
Figure 3-18. Xrn2 trims the 5’ leaders of intronic snoRNAs SNORD15A and SNORD11 after intron debranching

HEK293 Flp-in TREX cells were treated with doxycycline for 24 hrs to induce expression of wild-type (WT) Xrn2WT or exonuclease-dead, mutant (MT) (D235A) Xrn2MT. (A) 5’-Bruseq signal on SNORD15A. (B) 5’-Bruseq signal on SNORD11. Note accumulation of 5’ PO₄s at the intron start site (indicated by red arrows) accompanied by a reduction in 5’ PO₄s at the start site of the snoRNA (indicated by blue arrows). RPNM: Reads per nucleotide per million mapped reads. UCSC genome browser screenshots. Blue arrow indicates direction of transcription.
summary, this method can be used for accurate detection of 5’-monophosphorylated ends of nascent transcripts.

3.2F Xrn2 Recognizes the 5’ PO₄ End of Nascent Transcripts at polyA Sites for Degradation

To determine if Xrn2 targets the 5’ end of nascent transcripts after cleavage, I assessed the levels of these species at the polyA site of genes in cells expressing dominant negative, exonuclease-dead Xrn2MT (D235A) and wild-type (WT) control Xrn2 (Fong et al. 2015b). Remarkably, in accord with the prediction of the torpedo model, inhibition of Xrn2 led to stabilization of nascent transcripts downstream of polyA sites as evidenced on the EIF1, NEAT1 and HIST2H4A genes (Fig. 3-17A-C). Additionally, 5’ PO₄ ends were detected at the 5’ end of intron 2 of the EIF1 gene. Furthermore, this signal increased in intensity in both the first and second introns in Xrn2MT cells (Fig. 3-17A). These 5’ PO₄ species most likely correspond to intron-debranching events that occur co-transcriptionally or rapidly after excision of the intron lariat by pre-mRNA splicing and, therefore, are detected by 5’-Bruseq. My observation suggests that, following debranching, introns are degraded by Xrn2.

To visualize effects at the genome-wide level 5’-Bruseq signal from the mitochondrial genome was used for normalization. Expression of the mitochondrial genome is highly complex and includes maturation of 5’ and 3’ ends of transcripts that can generate 5’ PO₄ RNA ends (Binder and Kuhn 2015). Additionally, 5’ PO₄s mapping to the start site of introns were excluded to eliminate visualization of intron-debranching events. For analysis, all RefSeq genes with detectable signal were aligned to the TSS and the cleavage and polyadenylation site (CPS) at 3 ends of genes. Strikingly, 5’-Bruseq signal revealed that 5’ PO₄s are present within gene bodies indicating that 5’ uncapped nascent transcripts are common during transcription (Fig. 3-19A). Note that, as expected, a peak of 5’ PO₄s can be detected at the CPS generated by endonucleolytic
cleavage by CPSF73 (indicated by a red arrow). However, in Xrn2MT cells 5’-Bruseq signal is sharply elevated downstream and upstream of the TSS in the sense and antisense direction respectively (Fig. 3-19A). A biological replicate of this observation is shown in figure 3-20. This result suggests that nascent transcripts are frequently decapped or cleaved by endonucleolytic cleavage during transcription. This is a relevant result supporting the idea that nascent transcripts can be exposed for 5’ to 3’ degradation by Xrn2 in a premature termination mechanism proposed in Chapter V.

Significantly, inhibition of Xrn2 resulted in a strong and sharp accumulation of 5’ PO₄₅s at the CPS demonstrating that inhibition of the exonuclease activity of Xrn2 stabilizes transcripts at the ends of genes with a 5’ end that maps precisely at polyA sites (Fig 3-19, 3-20). This result demonstrates that Xrn2 loads onto the nascent transcript after cleavage at the polyA site provides a 5’ PO₄ that can be recognized by Xrn2 for degradation. A biological replicate of this result is shown in figure 3-20. This result is consistent with the torpedo model in which Xrn2 is proposed to chase the elongation complex via degradation of the nascent transcript at 3’ ends of genes. Additionally, in one experiment inhibition of Xrn2 led to an increase in 5’-Bruseq signal at the polyA site of genes in the antisense direction (Fig. 3-19C), but this effect was not reproduced (Fig. 3-20C). Intriguingly, inhibition of Xrn2 reduced the levels of 5’ PO₄ ends detected in the sense and antisense direction from the promoter region (Fig. 3-19, 3-20). This is a counterintuitive result. Inhibition of the exonuclease activity of Xrn2 is expected to stabilize 5’ PO₄₅s that are degraded by this factor; especially since nascent transcript levels were found to be increased after inhibition of Xrn2 on gene bodies (fig. 3-13). This unexpected result is further discussed below with respect to a proposed role of Xrn2 in premature termination of transcription presented in Chapter V.
Figure 3-19. Xrn2 targets 5’ PO4s ends of nascent transcripts at polyA sites for degradation

HEK293 Flp-in TREX cells were treated with doxycycline for 24 hrs to induce expression of wild-type (WT) Xrn2WT or exonuclease-dead, mutant (MT) (D235A) Xrn2MT. All RefSeq genes with at least one single sequencing read within each of the three gene regions (upstream of the TSS, downstream of the CPS and the body region (TSS to CPS)) were aligned at the transcription start site (TSS) and the cleavage and polyadenylation site (CPS). Additionally, the body region between the TSS and CPS was normalized to gene length. (A) 5’-Bruseq in control Xrn2WT cells. (B) Sense 5’-Bruseq signal in control and Xrn2MT cells. Note sharp increase in 5’ PO4 RNA ends at the CPS after inhibition of Xrn2. (C) Antisense 5’-Bruseq signal from Xrn2WT and Xrn2MT cells is shown overlaid with sense signal from control cells (Xrn2WT) for comparison. A biological replicate of 5’-Bruseq signal is shown in figure 3-20. N.M.: mean of total number of reads per bin, further normalized per thousand mitochondrial reads.
Figure 3-20. Xrn2 targets 5’ PO4s ends of nascent transcripts at polyA sites for degradation – Biological Replicate
HEK293 Flp-in TREX cells were treated with doxycycline for 24 hrs to induce expression of wild-type (WT) Xrn2WT or exonuclease-dead, mutant (MT) (D235A) Xrn2MT. Most active genes (top 25% with highest pol II ChIP-seq density in normal conditions) were aligned at the transcription start site (TSS) and the cleavage and polyadenylation site (CPS). Additionally, the body region between the TSS and CPS was normalized to gene length. (A) 5’-Bruseq in control Xrn2WT cells. (B) Sense 5’-Bruseq signal in control and Xrn2MT cells. Note sharp increase in 5’ PO4 RNA ends at the CPS after inhibition of Xrn2. (C) Antisense 5’-Bruseq signal from Xrn2WT and Xrn2MT cells is shown overlaid with sense signal from control cells (Xrn2WT) for comparison. Signal is shown for a reduced number of genes (n=1585) due to lower coverage of 5’-Bruseq reads over coding genes for this biological replicate relative to Run 1 (Fig. 3-19). N.M.: mean of total number of reads per bin, further normalized per thousand mitochondrial reads.
Additionally, nascent transcripts downstream of the 3’ cleavage site of genes whose 3’ ends are generated via mechanisms that do not produce a polyA tail were also stabilized when Xrn2 activity was suppressed. This was the case for Histone whose 3’ end formation depends on a RNA stem-loop structure at their 3’ ends (Fig. 3-17C, 3-21A) (Dominski et al. 1999). A biological replicate of 5’-Bruseq signal on histone genes is shown in Figure 3-22A. The NEAT1 gene for which RNase P cleavage defines its 3’ end (Sunwoo et al. 2009), also displayed an increase in 5’ PO₄s at the 3’ end in Xrn2MT cells (Fig. 3-17B). Surprisingly, inhibition of Xrn2 resulted in a reduction in 5’ PO₄ ends at the beginning of small nucleolar RNAs (snoRNAs) (Fig. 3-17D, 3-18, 3-21B). Interestingly, visualization of 5’-Bruseq signal on the intron-contained snoRNA, SNORD4B, revealed that inhibition of Xrn2 resulted in a reduction in 5’ PO₄s at the beginning of the SNORD4B but also led to an increase of 5’ PO₄s at the 5’ end of the associated intron (Fig. 3-17D, 3-18). This 5’ shift of 5’-Bruseq density is indicative of a reduction in 5’ to 3’ trimming of the 5’ leaders of snoRNAs by Xrn2 after intron debranching in Xrn2MT cells. That interpretation is consistent with a previous report indicating a role of Xrn2 in processing of intronic snoRNAs in yeast (Petfalski et al. 1998). A biological replicate showing a reduction of 5’-PO₄s at the beginning of snoRNAs after inhibition of Xrn2 is shown in Figure 3-22B.

3.2G Xrn2 Degrades the Nascent Transcript Downstream of the pre-rRNA 3’ End

Previous results have argued that Xrn2 contributes to processing and degradation of aberrant precursor ribosomal RNA (pre-rRNA) fragments in mammals (Wang and Pestov 2011) and in termination of RNA pol I transcription in a torpedo-like mechanism in yeast (El Hage et al. 2008; Kawauchi et al. 2008). Depletion of Xrn2 in mouse cells led to accumulation of short pre-rRNA fragments originating from cleavage events around the TSS of pol I transcription (Wang and Pestov 2011). In agreement with that observation, 5’-Bruseq signal on
Figure 3-21. Inhibition of Xrn2 differentially affects 5' PO₄s of nascent transcripts at histones, snoRNA, snRNA and miRNA genes
HEK293 Flp-in TREX cells were treated with doxycycline for 24 hrs to induce expression of wild-type (WT) Xrn2WT or exonuclease-dead, mutant (MT) (D235A) Xrn2MT. (A) Sense 5’-Bruseq signal on histone genes aligned at their 3’-end. Note increase in 5’ PO₄ ends at the 3’ end of histone genes after inhibition of Xrn2. (B) Sense 5’-Bruseq signal on snoRNAs aligned at their 5’ end. Note a global reduction in the levels 5’ PO₄ ends at the start site of snoRNAs after inhibition of Xrn2. (C) Sense 5’ Bruseq signal on snRNAs aligned at their 5’ end. (D) Sense 5’ Bruseq signal on miRNAs aligned at their 5’ end. A biological replicate of 5’-Bruseq signal on Histone and snoRNAs is shown in figure 3-22. N.M.: mean of total number of reads per bin, further normalized per thousand mitochondrial reads.
Figure 3-22. Inhibition of Xrn2 affects 5’ PO_4s of nascent transcripts at histones and snoRNA genes – Biological Replicate

HEK293 Flp-in TREX cells were treated with doxycycline for 24 hrs to induce expression of wild-type (WT) Xrn2WT or exonuclease-dead, mutant (MT) (D235A) Xrn2MT. (A) Sense 5’-Bruseq signal on histone genes aligned at their 3’-end. Note increase in 5’ PO_4 ends at the 3’ end of histone genes after inhibition of Xrn2. (B) Sense 5’-Bruseq signal on snoRNAs aligned at their 5’ end. Note a global reduction in the levels 5’ PO_4 ends at the start site of snoRNAs after inhibition of Xrn2. N.M.: mean of total number of reads per bin, further normalized per thousand mitochondrial reads.
the ribosomal DNA repeating unit (rDNA) revealed that inhibition of Xrn2 results in accumulation of 5’ PO₄ ends at the beginning of the pre-rRNA transcription unit (Fig. 3-15). Additionally, inhibition of Xrn2 stabilized 5’ PO₄s closely downstream of the 28S rRNA, consistent with the idea that Xrn2 degrades the nascent transcript after cleavage at the 3’ end of the pre-rRNA for termination of pol I transcription. A biological replicate of these effects is shown in figure 3-16. Furthermore, analysis of Bruseq signal showed that nascent transcript signal is additionally increased downstream of the stabilized 5’ PO₄ peak in Xrn2MT cells (indicated in the green box) (Figure 3-23). A biological replicate is shown in figure 3-24. Taken together, these results strongly suggest that Xrn2 targets the nascent transcript for degradation after cleavage of the pre-rRNA transcript at its 3’ end and strengthen the idea that Xrn2 contributes to termination of pol I transcription.

3.2H Xrn2 Targets the 5’ End of snRNAs

Lastly, 5’-PO₄ ends of tRNAs (n = 405) and miRNAs (n=614) appeared unaffected by inhibition of Xrn2 (Fig. 3-14C – 3-21D). Interestingly, inhibition of Xrn2 generated a slight increase in 5’ PO₄s at the start site of snRNAs (n = 530) (Fig. 3-21C). This is an interesting result given that snRNAs, similar to mRNAs, receive a cap structure at their 5’ ends (Gu et al. 2012). snRNAs are decapped by Dcp2 and degraded 5’-3’ by Xrn1 or from their 3’ end by 3’-5’ degradation by the exosome in yeast (Shukla and Parker 2014). Therefore, stabilization of 5’ PO₄s at the start site of snRNAs after inhibition of Xrn2 suggests that this exonuclease additionally participates in turnover of decapped snRNAs by targeting them for 5’-3’ exonucleolytic degradation.
Figure 3-23. Xrn2 targets nascent transcripts at the 5’ and 3’ end of the pre-rRNA transcription unit

Nascent transcripts were isolated from HEK293 Flp-in TREX cells expressing dominant negative, mutant (MT) (D235A) Xrn2MT or wild-type (WT) control Xrn2WT. Bruseq signal of nascent transcripts was mapped to a full human rDNA repeat (GenBank accession no. U13369). Shown are the distribution profiles of Bruseq signal of Xrn2WT and Xrn2MT cells overlaid on the rDNA repeating unit. Sense Bruseq signal is shown on the positive “y” axis and antisense Bruseq signal is shown on the negative “y” axis. Note a 5’ shift in nascent transcript signal toward the transcription start site of the pre-rRNA relative to distal positions on the rDNA repeating unit in Xrn2MT cells when compared to Xrn2WT cells. 5’-Bruseq signal has been included above Bruseq profiles to indicate the position of the stabilized 5’ PO₄ end of the nascent transcript downstream of the pre-rRNA transcription unit (green box). Note increased signal of nascent transcripts downstream of the stabilized 5’ PO₄s in Xrn2MT cells (indicated by green arrows). A biological replicate of 5’ Bruseq signal on the rDNA repeating unit is shown in figure 3-24. CPBM: coverage per bin per million mapped reads. RPNM: Reads per nucleotide per million.
Figure 3-24. Xrn2 targets nascent transcripts at the 5’ and 3’ end of the pre-rRNA transcription unit – Biological Replicate

Nascent transcripts were isolated from HEK293 Flp-in TREX cells expressing dominant negative, mutant (MT) (D235A) Xrn2MT or wild-type (WT) control Xrn2WT. Bruseq signal of nascent transcripts was mapped to a full human rDNA repeat (GenBank accession no. U13369). Shown are the distribution profiles of Bruseq signal of Xrn2WT and Xrn2MT cells overlaid on the rDNA repeating unit. Sense Bruseq signal is shown on the positive “y” axis and antisense Bruseq signal is shown on the negative “y” axis. Note a 5’ shift in nascent transcript signal toward the transcription start site of the pre-rRNA relative to distal positions on the rDNA repeating unit in Xrn2MT cells when compared to Xrn2WT cells. 5’-Bruseq signal has been included above Bruseq profiles to indicate the position of the stabilized 5’ PO₄ end of the nascent transcript downstream of the pre-rRNA transcription unit (green box). Note increased signal of nascent transcripts downstream of the stabilized 5’ PO₄s in Xrn2MT cells (indicated by green arrows). CPBM: coverage per bin per million mapped reads.

RPNM: Reads per nucleotide per million.
In summary, 5'-'Bruseq has allowed detection of 5'-monophosphorylated ends of nascent transcripts, identified the different targets of Xrn2 and confirmed the role of Xrn2 in degradation of nascent transcripts at the polyA site as anticipated by the torpedo model.

3.3 Discussion

The results presented in this chapter addressed two important aspects of the termination factor Xrn2 relevant for termination of RNA pol II: (1) they showed that Xrn2 is recruited at the transcription start site during transcription and (2) they confirmed the role of Xrn2 in degradation of nascent transcripts at the polyA site as proposed by the torpedo model. Previous studies have reported the presence of Xrn2 at 3’ ends of genes consistent with its role in termination, but information on when Xrn2 is recruited during transcription was lacking. The data above, however, established that Xrn2 is recruited at the transcription start site of genes. Importantly, Xrn2 accumulated to high densities at the promoter of the NEAT1, EIF1 and RPL13 genes under conditions in which no detectable signal of Xrn2 was present at their 3’ ends (Fig. 3-1B – 3-8, 3-9). Such result rules out the possibility that signal of Xrn2 around promoters is simply caused by a potential “gene-loop” conformation in which the termination region can be physically linked to its promoter (O'Sullivan et al. 2004).

Inhibition of the helicase activity of XPB with triptolide has been assumed to prevent recruitment of pol II to promoters in previous studies (Henriques et al. 2013; Shao and Zeitlinger 2017; Jonkers, Kwak, and Lis 2014). The data presented here, however, demonstrates that triptolide treatment still allows for a peak of pol II at the TSS (Fig 3-1A, 3-2A). This observation indicates that recruitment of pol II can still occur after inhibition of the XPB helicase with triptolide. A recent report argued that the helicase activity of XPB is required in the pre-initiation complex to overcome a block imposed by XPB itself (Alekseev et al. 2017). If that is true,
inhibition of the helicase activity of XPB with triptolide may inhibit transcription initiation by stabilizing the XPB-mediated block in the pre-initiation complex.

I have found that Xrn2 maintains association with the elongation complex until it reaches the 3’ ends of genes (Fig 3-8 - 3-10). This finding is particularly relevant given the implications of having a termination factor present throughout the elongation phase of pol II. Premature termination becomes a feasible event. Elongation complexes destined for termination might be marked through generation of a 5’-monophosphorylated end that may be easily detected by Xrn2 localized in its vicinity. This is in line with results in Chapter V and several reports suggesting a role of Xrn2 in premature termination of transcription (Brannan et al. 2012; Wagschal et al. 2012; Nojima et al. 2015b).

Several reports have supported a role of Xrn2 at 3’ ends of genes as described in the torpedo model of transcription termination (Fong et al. 2015b; West, Gromak, and Proudfoot 2004; Gromak, West, and Proudfoot 2006). Nonetheless, it has not been shown that Xrn2 targets the 5’ end generated by cleavage at the polyA site in vivo. Here I detected 5’ PO4 ends of nascent transcripts using 5’-Bruseq to confirm this essential aspect of the model. Remarkably, inhibition of the exonuclease activity of Xrn2 generated accumulation of RNA 5’ PO4s at the cleavage site of polyA-dependent genes as predicted by the torpedo model (Fig. 3-17A, 3-19). The stabilization of these 5’ PO4s after inhibition of Xrn2 strengthens the premise that degradation of the nascent transcript is an essential function of Xrn2 during termination of RNA pol II. Interestingly, 5’-Bruseq signal is sharply elevated downstream and upstream of the TSS in the sense and antisense direction respectively in control Xrn2WT cells (Fig. 3-19A, 3-20A), suggesting that nascent transcripts are frequently decapped or cleaved by endonucleolytic cleavage during transcription. This is a relevant result supporting the idea that nascent transcripts can be exposed
for 5’ to 3’ degradation by Xrn2 in a premature termination mechanism proposed in Chapter V. Nevertheless, inhibition of Xrn2 led to a reduction in 5’-PO4s within body regions (Fig. 3-19, 3-20). This is a counterintuitive result, especially since nascent transcript levels were found to be increased after inhibition of Xrn2 on gene bodies (Fig. 3-13). Furthermore, strong evidence was found in Chapter V suggesting that Xrn2 functions in premature termination of transcription around promoters and that inhibition of its exonuclease activity recovers a fraction of polymerases within gene bodies (Fig. 5-4, 5-5). Therefore, there is no clear explanation for the observation that while both nascent transcripts and polymerases are recovered on gene bodies, 5’ PO4s ends are reduced on these sites upon inhibition of Xrn2. One possibility that would explain such a result is that failure to prematurely terminate elongation complexes by Xrn2, or inhibition of the exonuclease activity of Xrn2, can signal to decapping or cleavage factors to reduce production of 5’ PO4s ends that will not be effectively removed. Although this scenario seems unlikely, detection of interactions of Xrn2 with decapping factors by mass spectrometry analysis, suggests that at least physical communication between Xrn2 and decapping factors does exist (Brannan et al. 2012).

A role of Xrn2 in termination of pol I transcription was also suggested by the observation that 5’ PO4s and nascent transcripts accumulated at the end of the pre-rRNA transcription unit after inhibition of Xrn2. This result is consistent with torpedo action by Xrn2 at the 3’ end of the pre-rRNA transcription unit (Fig. 2-23, 3-24). Interestingly, inhibition of Xrn2 also led to accumulation of nascent transcripts within the pre-rRNA transcription unit suggesting that Xrn2 might additionally function in premature termination of RNA pol I transcription. This is consistent with previous reports arguing that Xrn2 targets aberrant pre-rRNA transcripts for degradation (Fang, Phillips, and Butler 2005; Wang and Pestov 2011).
The 5’-3’ exonuclease activity of Rat1, the yeast homolog of Xrn2, is proposed to trim the 5’ extensions after Rnt1 cleavage to define the 5’ ends of snR190 and U14 (Chanfreau et al. 1998). Furthermore, it has been shown that snoRNAs originating from introns contained in protein-coding genes require intron debranching and processing by Rat1 to define the 5’ ends of snoRNAs U18 and U24 (Petfalski et al. 1998). In agreement with these observations, inhibition of the exonuclease activity of Xrn2 reduces the levels of 5’ PO₄ of snoRNAs (Fig. 3-17D, 3-18, 3-21B, 3-22B), while increasing the signal of the 5’ PO₄ at the start site of the associated intron (Fig. 3-17D, 3-18). This is consistent with a reduction in trimming of 5’-extensions of snoRNAs from the 5’-end of introns after debranching. Thus, Xrn2 seems to be involved in processing of snoRNAs by trimming 5’-end extensions after endonucleolytic cleavage or after intron debranching.

Interestingly, a role of Xrn2 in termination of histone genes remains controversial. While depletion of Xrn2 did not reveal a significant effect on nascent transcript signal at 3’ ends of histone genes (Eaton et al. 2018), inhibition of the exonuclease activity of Xrn2 resulted in a modest increase in pol II density farther downstream (Fong et al. 2015b). While polyA-dependent genes produce polyadenylated transcripts at their 3’-ends, histone genes do not. The 3’ ends of these transcripts are processed via a different mechanism. A specialized processing machinery recognizes two unique features present in histone pre-mRNAs, a stem-loop and a purine-rich histone downstream element (HDE) (Dominski et al. 1999). A difference in processing machinery at 3’ ends might explain a potential alternative mechanism of transcription termination independent of Xrn2. Nevertheless, results obtained above clearly demonstrate that Xrn2 targets the nascent transcript generated by cleavage at 3’ ends of histone genes for degradation (Fig. 3-17C, 3-21A, 3-22A). This result suggests that Xrn2 participates in termination of transcription on histone genes. Interestingly, the processing machinery responsible for histone 3’-end formation
contains a number of factors involved in cleavage/polyadenylation (Dominski, Yang, and Marzluff 2005), including the endonuclease CPSF73 responsible for cleavage at polyA-dependent genes. In summary, CPSF73 endonucleolytic cleavage and 5'-3' Xrn2 exonucleolytic digestion of the nascent transcript at the 3’ end of histone genes are also common to polyA-dependent genes. Such similarities with polyA-dependent genes further supports a role of Xrn2 in termination of transcription of histone genes. An alternative role of Xrn2 at the end of histone genes may additionally be release of the U7 small nuclear RNP (snRNP) bound to the nascent transcript downstream of the cleavage site. It has been shown that after cleavage, the downstream product is rapidly degraded, resulting in release of the U7 snRNP complex, which can be used in another round of processing (Walther et al. 1998).

Taken together results in this chapter allowed me to identify how Xrn2 integrates with the elongation complex and confirmed an unverified prediction of the torpedo model. As illustrated in figure 3-25, these data suggest a model in which Xrn2 integrates with the transcription process at the transcription start site by association with the pre-initiation complex and migrates associated with the elongation complex until it reaches the 3’ ends of genes. After cleavage at the polyA site, Xrn2 loads onto the 5’-monophosphorylated end and chases pol II by degrading the nascent RNA. When it reaches the polymerase it contributes to dislodge it from the DNA template.
Figure 3-25. Model for integration of Xrn2 into the transcription process and termination of pol II transcription
Xrn2 binds the pre-initiation complex (PIC) at the transcription start site and migrates associated with the elongation complex until it reaches the 3’ ends of genes. After cleavage at the polyA site, Xrn2 loads onto the 5’-monophosphorylated end and degrades the nascent transcript it in a chase for pol II.
CHAPTER IV

THE ALLOSTERIC-TORPEDO MODEL OF RNA POL II TRANSCRIPTION TERMINATION

4.1 Introduction

Termination of transcription occurs when RNA polymerase ceases synthesis of the nascent transcript and dissociates from the DNA template. Failure in termination of transcription can result in production of potentially harmful intergenic transcripts and transcription interference with downstream genes (Shearwin, Callen, and Egan 2005; Greger and Proudfoot 1998). Therefore, timely and efficient termination is crucial to properly define transcription units and prevent alteration of gene expression patterns. RNA pol II transcription termination has long been known to be coupled to 3′-end processing of the pre-mRNA. A polyadenylation (polyA) signal consisting of the hexamer sequence AAUAAA is required for both cleavage and polyadenylation of the pre-mRNA transcript immediately downstream of the polyA signal, as well as for termination of pol II transcription (Connelly and Manley 1988b; Buratowski 2005; Birse et al. 1998). The cleavage and polyA site (CPS) will also be referred here simply as the polyA site. From this observation, two main models have aimed to describe the mechanism of RNA pol II transcription termination downstream of a polyA signal on protein-coding genes: the torpedo model and the allosteric model. According to the torpedo model, cleavage downstream of the polyA signal allows the 5′-3′ exonuclease Xrn2/Rat1 (human/yeast) (introduced in Chapter I) to degrade the nascent transcript and chase pol II for termination (Fig. 1-2) (Connelly and Manley 1988a; Kim et al. 2004; West, Gromak, and Proudfoot 2004). It is thought that catching-up of Xrn2 with pol II induces disassembly of the elongation complex by a yet unknown mechanism. The allosteric model, in contrasts, conveys that the elongation complex undergoes a conformational change due to loss of
an antitermination factor, or gain of a termination factor, leading to pausing and termination (Logan, Falck-Pedersen, et al. 1987; Calvo and Manley 2001; Greenblatt et al. 1998; Connelly and Manley 1988b; Whitelaw and Proudfoot 1986).

An intense debate over the last three decades for and against these models point to a more intertwined mechanism of transcription termination. Initial evidence for the torpedo model comes from studies in yeast where depletion of Rat1 generates a substantial termination defect downstream of polyA sites (Kim et al. 2004). Additionally, self-cleavage of the nascent transcript at a co-transcriptional cleavage (CoTC) site downstream of the human beta-globin gene (Teixeira et al. 2004), was shown to act as a precursor to termination by presenting an uncapped 5’ RNA end that is recognized by Xrn2 (West, Gromak, and Proudfoot 2004). In that report, it was shown that depletion of Xrn2 results in reduced termination efficiency downstream of the CoTC cleavage site and stabilization of the downstream nascent transcript on co-transfected plasmids in HeLa cells. This result added strength to the idea that Xrn2 functions as a torpedo that degrades the nascent transcript and chases pol II for termination. Nevertheless, a global requirement of Xrn2 in termination of pol II was challenged by seemingly contradictory observations and conflicting reports. In vitro assays have shown that disassembly of polymerases can occur in the absence of cleavage or Xrn2 exonuclease activity (Zhang, Fu, and Gilmour 2005; Pearson and Moore 2013; Zhang, Rigo, and Martinson 2015). Similarly, partial depletion of Xrn2 in human cells does not seem to cause a termination defect (Brannan et al. 2012; Nojima et al. 2015a), suggesting that transcription termination does not require “torpedo” action by Xrn2. Recent investigations have reported contrasting results in human cells however. On one hand, depletion of CPSF73 does in fact generate a strong termination defect that cannot be rescued by overexpression of a cleavage-deficient CPSF73 mutant (Eaton et al. 2018). On the other hand, the use of an exonucleotically-
dead mutant (MT) of Xrn2 (D235A) demonstrated that degradation of the nascent transcript is in fact required for proper termination of pol II transcription on most human genes (Fong et al. 2015a). In that report, it was shown that expression of Xrn2MT results in stronger pol II occupancy downstream of polyA sites and increased transcription farther downstream of 3’ ends of genes. In agreement with a kinetic competition as described by the torpedo model, it was also shown that mutations that affect the elongation rate of pol II lead to early termination of slow polymerases, and delayed termination of fast polymerases downstream of the polyA signal. Therefore, it is still unclear to what extent cleavage-independent mechanisms can support termination of transcription in vivo.

Alternative mechanisms that might involve a conformational change of the elongation complex conducive to pausing and termination, according to the allosteric model, have been proposed. Notably, the termination factor Pcf11, a core component of the cleavage and polyadenylation machinery, is able to release pol II in vitro independent of transcript cleavage or degradation (Zhang, Fu, and Gilmour 2005; Zhang and Gilmour 2006). It was proposed that simultaneous binding of Pcf11 to the nascent transcript and the C-terminal domain (CTD) of pol II might induce a conformational change capable of evicting the elongation complex, a mechanism that appears common with the termination factor Nrd1 during termination of polyA-independent genes in yeast (introduced in Chapter I) (Arigo et al. 2006). Additionally, it has been reported that binding of cleavage and polyadenylation factors might lead to a conformational change of the elongation complex that can be blocked by the transcriptional inhibitor alpha-amanitin before termination in vitro (Zhang, Rigo, and Martinson 2015). Nevertheless, it is still not clear if the elongation complex suffers a polyA site-dependent conformational change capable of inducing disassembly from the DNA template in vivo. It seems plausible that at least pausing of a highly
processive elongation complex might be a prerequisite for termination. In prokaryotic systems, pausing of RNA polymerases is a common feature of the elongation process (Mustaev, Roberts, and Gottesman 2017) that might facilitate opening of the clamp, and render the elongation complex susceptible to termination (Sekine et al. 2015; Weixlbaumer et al. 2013). In eukaryotic systems, the conserved observation that pol II accumulates downstream of transcription units by ChIP-seq (Lian et al. 2008; Fong et al. 2015b), Gro-seq (Core et al. 2012) or mNET-Seq (Nojima et al. 2015b) is also consistent with pol II pausing downstream of the polyA signal. However, it is still not known if pol II pausing is required for termination and what mechanism would be responsible for it.

Termination of pol II transcription might also involve a slowdown in the rate of transcription elongation before pol II can dissociate from the DNA template. Interestingly, Spt5 is a pol II-associated factor that is conserved in all domains of life (Werner 2007) and can positively and negatively impact transcription elongation. Binding of Spt5 to both the nascent transcript and pol II has been shown to induce promoter-proximal pausing of a reconstituted elongation complex in vitro in cooperation with the negative elongation factor NELF (Missra and Gilmour 2010; Yamaguchi, Takagi, et al. 1999). Promoter-proximal pausing of pol II is introduced in Chapter I. In order for a paused pol II to transition into productive elongation it requires the action of the positive transcription elongation factor b (pTEF-b), which phosphorylates Spt5 and Pol II (Peterlin and Price 2006; Renner et al. 2001). Importantly, phosphorylation of the C-terminal domain of human Spt5 by pTEF-b has been proposed to switch Spt5 from a pausing factor into a positive elongation factor (Yamada et al. 2006). This was proposed based on the observation that phosphorylation of Spt5 by pTEF-b was not required for its repressive activity on elongation, but it was critical for Spt5-mediated processive transcription elongation in vitro (Yamada et al. 2006).
Additionally, chromatin immunoprecipitation analysis has shown that *Drosophila* Spt5 is found across the body region of genes *in vivo* (Andrulis et al. 2000; Aida et al. 2006), leading to the notion that Spt5 positively impacts elongation throughout the length of genes. However, it is still not known if phosphorylation levels of Spt5 change during transcription elongation or how changes in Spt5 phosphorylation levels impact transcription elongation by pol II. A recent study in yeast *Schizosaccharomyces pombe* suggested that Dis2, an isoform of protein phosphatase 1 (PP1), can dephosphorylate Spt5 *in vitro* and that Dis2 is required to prevent transcription beyond normal termination zones (Parua et al. 2018). In that report, it was proposed that a transition of pol II from elongation to termination is regulated by opposing activities of Cdk9 and Dis2 towards their common substrate Spt5. An interesting possibility could be that loss of Spt5 phosphorylation at 3’ ends of genes might revert back the repressive activity of Spt5 on transcription elongation to promote slowdown of the elongation complex and termination of pol II. Signals triggering such suppression might come from the nascent transcript. Binding of Spt5 with the nascent RNA was found to sharply increase by PAR-CLIP immediately downstream of the polyA site in yeast (Baejen, Andreani, Torkler, Battaglia, Schwalb, Lidschreiber, Maier, Boltendahl, Rus, Esslinger, Soding, et al. 2017), pointing to the possibility that Spt5 might “sense” passage of pol II over the polyA site. A potential role of Spt5 in RNA 3’-end processing and termination is further suggested by the fact that its C-terminal domain contributes to recruitment of the 3’ RNA cleavage factor I at 3’-ends of genes in yeast (Mayer, Schreieck, et al. 2012).

Additionally, several reports suggest that cleavage and polyadenylation factors may contribute to pol II pausing at 3’-ends of genes. Binding of the CPSF factor to both the AAUAAA hexamer sequence and the body of pol II induces pausing in a reporter containing the SV40 late polyA signal (Nag, Narsinh, and Martinson 2007). Further support for a role of 3’-end processing
factors in limiting elongation of pol II upon binding to the polyA signal was shown in the same study. Overexpression of the influenza virus protein NS1A, which targets CPSF30 and blocks CPSF binding to the polyA signal (Twu et al. 2006; Nemeroff et al. 1998), resulted in increased pol II transcription farther downstream of the SV40 late polyA signal. Thus, CPSF30 bound to the elongation complex might be responsible for recognition of the AAUAAA motif once it is extruded from the pol II active site to tether the polyA signal to the elongation complex and somehow trigger a slowdown of pol II at 3’ ends of genes. Such a mechanism is further suggested by the observation that CPSF associates with the transcription initiation factor TFIID in mammalian cells and yeast (Dantonel et al. 1997; Sanders et al. 2002), and that it is transferred from the transcription factor TFIID to the elongating polymerase in an in vitro mammalian system where it is thought to scan the RNA, awaiting the nascent polyA signal (Dantonel et al. 1997).

A role of 3’-processing factors in slowing down the elongation complex upon polyA signal recognition would explain the still controversial observation that depletion of processing factors generally cause stronger termination defects than depletion of Xrn2/Rat1 (Nojima et al. 2015a; Baejen, Andreani, Torkler, Battaglia, Schwalb, Lidschreiber, Maier, Boltendahl, Rus, Esslinger, Soding, et al. 2017; Eaton et al. 2018). If 3’-processing factors are required to slow down pol II and facilitate catching-up of Xrn2 downstream of the polyA signal, inhibition of these factors may cause pol II to “overlook” the signal to slow down at the ends of genes and continue transcription farther downstream of gene ends at high elongation rates. In principle, pausing of pol II may facilitate termination by shortening the path of Xrn2 to catch up with the elongation complex. In fact, specific pausing sequences downstream of transcription units contribute to termination of pol II (Enriquez-Harris et al. 1991; Ashfield et al. 1994; Yonaha and Proudfoot 1999; Plant et al. 2005;
Gromak, West, and Proudfoot 2006). Nevertheless, it is still not known to what extent pol II slows down past the polyA signal, and whether or not pausing is required for termination.

Here I demonstrate that elongation complexes transition at the polyA site into a strong slowdown and that Xrn2 is required for efficient release of pol II from the DNA template. I have directly detected poorly elongating complexes that fail to be removed from chromatin after inhibition of the exonuclease activity of Xrn2. Furthermore, I have utilized CRISPR-mediated mutation of a consensus polyA signal or overexpression of the flu virus NS1A protein to demonstrate that polyA signal recognition is responsible for both loss of Spt5 phosphorylation and slowdown of pol II downstream of polyA sites. I propose a unified allostERIC-torpedo model of RNA polymerase II transcription termination.

4.2 Results

4.2A Xrn2 is Required for Release of pol II from the DNA Template Downstream of polyA Sites

As mentioned above, a previous report by the Bentley lab provided support for the torpedo model by showing that inhibition of the exonuclease activity of Xrn2, via expression of dominant negative D235A mutant (MT) Xrn2MT causes a global pol II termination defect downstream of polyA sites (Fong et al. 2015b). It was shown that expression of Xrn2MT in HEK293 cells, under stable shRNA-mediated knockdown of the endogenous Xrn2, resulted in augmented pol II occupancy and increased pol II transcription farther downstream of 3’ ends of genes when compared to cells expressing wild-type (WT) Xrn2. Here, using the same approach, I decided to test the torpedo model further by investigating participation of Xrn2 in dissociation of pol II from the DNA template downstream of polyA sites. For this, the approach consisted of inhibiting transcription initiation and evaluating clearance of the last wave of polymerases immediately downstream of polyA sites after inhibition of the exonuclease activity in Xrn2MT cells.
First, expression of Xrn2MT resulted in increased pol II occupancy downstream of the polyA site of the ACTB gene relative to pol II occupancy around its promoter (Fig 4-1), reproducing the termination defect reported by Fong et al., 2015. Having recapitulated the termination defect, I incorporated inhibition of transcription initiation to evaluate clearance of polymerases at 3’ ends of genes. For this experiment, I incubated cells with triptolide. This drug targets the ATPase XPB helicase in the TFIIH transcription initiation factor and inhibits transcription initiation (Titov et al. 2011). As a result, polymerases that have already initiated transcription continue through the cycle and terminate transcription. As expected, pol II ChIP-seq signals on the ACTB and EIF1 genes show progressive clearance of pol II density over time that is completely lost after 15 minutes of treatment (Fig. 4-2, 4-3), indicated by red arrows. Strikingly, Xrn2MT cells exhibit a dramatic delay in clearance of pol II density downstream of the polyA site on these genes (Fig. 4-2, 4-3) (indicated by red arrows). It can be observed that while Xrn2WT cells have completely cleared polymerases downstream of the polyA site of the ACTB and EIF1 genes at time point 15 min, Xrn2MT cells do not and exhibit accumulation of pol II until 20 minutes, or even up to 30 minutes in the case of the ACTB gene (Fig. 4-2, 4-3). Two biological replicates of this observation are shown. This result strongly suggests that Xrn2 is required for release of pol II from the DNA template in a timely manner downstream of polyA sites. Significantly, the accumulated density of pol II in Xrn2MT cells does not seem to translocate farther downstream from the polyA site over time (Fig. 4-2, 4-3), suggesting that polymerases do not efficiently transcribe past the polyA site and “wait” for Xrn2 at sites of termination. This idea is further supported by results following this section, below.

To further investigate global failure in pol II clearance at 3’ ends of genes in Xrn2MT cells, ChIP-seq signal was analyzed on genes separated from downstream transcription units by at least
Figure 4-1. Inhibition of Xrn2 generates a pol II termination defect on the ACTB gene
HEK293 Flp-in TREX cells were treated with doxycycline for 24 hrs to induce expression of wild-type (WT) Xrn2WT or exonuclease-dead, mutant (MT) (D235A) Xrn2MT. Shown is the distribution profile of anti-pol II ChIP-seq signal on the ACTB gene. The red dotted line marks the position of the cleavage and polyadenylation site (CPS). Note a termination defect characterized by elevated pol II density downstream of the ACTB gene and increased transcription farther away from the CPS relative to control cells expressing Xrn2WT. RPBM: reads per bin per million. UCSC genome browser screen shot. Blue arrow indicates the direction of transcription.
Figure 4-2. Xrn2 is required for release of pol II from the DNA template at the 3’ end of the ACTB gene

HEK293 Flp-in TREX cells were treated with doxycycline for 24 hrs to induce expression of wild-type (WT) Xrn2WT or exonuclease-dead, mutant (MT) (D235A) Xrn2MT. Cell cultures were treated with 10 uM triptolide (Trip) to inhibit transcription initiation and anti-pol II ChIP-seq libraries were performed after 5, 10 and 15 min of triptolide treatment. Shown is pol II ChIP-seq signal on the ACTB gene. The red dotted line marks the position of the cleavage and polyadenylation site (CPS). Note a delay in the removal of polymerases downstream of the CPS in Xrn2MT cells relative to control Xrn2WT cells (indicated by red arrows). Control sample corresponds to cells in normal conditions. Two biological replicates are shown. RPBM: reads per bin per million. UCSC genome browser screen shots. Blue arrow indicates the direction of transcription.
Figure 4-3. Xrn2 is required for release of pol II from the DNA template at the 3’ end of the EIF1 gene

HEK293 Flp-in TREX cells were treated with doxycycline for 24 hrs to induce expression of wild-type (WT) Xrn2WT or exonuclease-dead, mutant (MT) (D235A) Xrn2MT. Cells were treated with 10uM triptolide (Trip) to inhibit transcription initiation and anti-pol II ChIP-seq was performed at the indicated time point after addition of triptolide. Shown is pol II ChIP-seq signal on the EIF1 gene. The red dotted line marks the position of the cleavage and polyadenylation site (CPS). Note a delay in the removal of polymerases downstream of the CPS in Xrn2MT cells relative to control Xrn2WT cells (indicated by red arrows). Control sample corresponds to cells in normal conditions. Two biological replicates are shown. RPBM: reads per bin per million. UCSC genome browser screen shots. Blue arrow indicates the direction of transcription.
+12 kb from their polyA site. To visualize effects on pol II occupancy at 3’ ends of genes from the initial minutes of triptolide treatment, genes 6 kb long and shorter were selected for meta-gene analysis (n=1621). As observed on the ACTB and EIF1 genes, this group of genes reveal global failure of pol II dissociation from the DNA template downstream of polyA sites (Fig. 4-4, 4-5). Two biological replicates are shown. It can be observed that while pol II occupancy drops dramatically after 10 minutes of triptolide treatment at the 3’ ends of these genes in Xrn2WT cells, pol II occupancy in Xrn2MT cells is still strong at this time point and is consistently higher at later time points downstream of the polyA site (Fig. 4-4, 4-5) (indicated by double-headed arrows). Therefore, this result suggests that Xrn2 is required for efficient dissociation of pol II downstream of polyA sites.

4.2B Pol II Decelerates Transcription Elongation Downstream of polyA Sites

The observation that accumulated elongation complexes in Xrn2MT cells do not appear to efficiently translocate away from the polyA site (Fig. 4-2, 4-3), suggested that elongation complexes slow down at 3’ ends of genes before they become targets of Xrn2. To test that idea, I turned to visualize the front wave of pol II generated upon synchronous release of paused polymerases from promoter regions. In this strategy I calculated the global elongation rate of pol II by analyzing movement of the front wave of pol II ChIP-seq density over time within gene bodies and tested for a potential slowdown in transcription elongation past the polyA site. For this experiment, I used DRB treatment (2.5 hrs, 100 µM), which blocks pTEF-b and prevents escape of pol II from the promoter-proximal pausing site (Mancebo et al. 1997).

As expected, release of paused polymerases generated a wave of pol II ChIP-seq density moving into gene bodies at time points (T) 5, 10 and 15 minutes that can be seen on individual genes (Fig. 4-6, 4-7) and on genes that are 50 kb and longer (Fig. 4-8). The global elongation rate
Figure 4-4. Xrn2 is required for efficient release of pol II from the DNA template downstream of polyA sites genome-wide
HEK293 Flp-in TREX cells were treated with doxycycline for 24 hrs to induce expression of wild-type (WT) Xrn2WT or exonuclease-dead, mutant (MT) (D235A) Xrn2MT. Cells were treated with 10 uM triptolide (Trip) to inhibit transcription initiation and anti-pol II ChIP-seq was performed after 5, 10 and 15 min of triptolide treatment. Shown is the pol II ChIP-seq distribution profile on short genes (6 kb long and shorter) separated from downstream transcription units by at least +12 kbs from their polyA site. The red dotted line marks the position of the cleavage and polyadenylation site (CPS). Note a global delay in the removal of polymerases downstream of the CPS in Xrn2MT cells relative to control Xrn2WT cells (illustrated by longer double-headed arrows in Xrn2MT cells for each time point). The overlay of time points from Xrn2WT and Xrn2MT cells is shown at the bottom. The mean distribution profile of pol II ChIP-seq density was normalized to signal at +12 kbs from the CPS.
Figure 4-5. Xrn2 is required for efficient release of pol II from the DNA template downstream of polyA sites genome-wide – Biological Replicate

HEK293 Flp-in TREX cells were treated with doxycycline for 24 hrs to induce expression of wild-type (WT) Xrn2WT or exonuclease-dead, mutant (MT) (D235A) Xrn2MT. Cell cultures were treated with 10uM triptolide to inhibit transcription initiation and anti-pol II ChIP-seq libraries were performed after 5, 10, 20 and 30 min of triptolide (Trip) treatment. Shown is the pol II ChIP-seq distribution profile on short genes (6 kb long and shorter), separated from downstream transcription units by at least +12 kbs from their polyA site. The red dotted line marks the position of the cleavage and polyadenylation site (CPS). Note a global delay in the removal of polymerases downstream of the CPS in Xrn2MT cells relative to control Xrn2WT (illustrated by longer double-headed arrows in Xrn2MT cells for each time point). The mean distribution profile of pol II ChIP-seq density was normalized to signal at - 1 kb from the CPS.
Figure 4-6. Inhibition of Xrn2 causes a stronger pile-up of polymerases at the 3’ end of the ACTB gene

HEK293 Flp-in TREX cells were treated with doxycycline for 24 hrs to induce expression of wild-type (WT) Xrn2WT or exonuclease-dead, mutant (MT) (D235A) Xrn2MT. Cells were treated with DRB (100 µM, 2.5 hrs) to inhibit pause release of pol II from promoters and clear polymerases from gene bodies. Paused polymerases were allowed to resume elongation by removal of DRB. Shown is the distribution profile of pol II ChIP-seq signal at time point (T) T₀, T₅, T₁₀ and T₁₅ minutes on the ACTB gene after DRB wash. Note that pol II arrives at the 3’ end of the ACTB gene after 5 minutes (T₅ min) of DRB wash and piles up immediately downstream of the cleavage and polyadenylation site (marked by the red dotted line). Also note that inhibition of Xrn2 results in stronger pol II pile-up downstream of the CPS (intensity indicated by a red bar) relative to pol II density at the promoter region (intensity indicated by a grey bar). The double-headed arrows illustrate the intensity of pol II pile-up at the 3’ end of the ACTB gene relative to the intensity of the pol II peak at the promoter region.
**Figure 4-7. Inhibition of Xrn2 causes a stronger pile-up of polymerases at the 3’ end of the EIF1 gene**

HEK293 Flp-in TREX cells were treated with doxycycline for 24 hrs to induce expression of wild-type (WT) Xrn2WT or exonuclease-dead, mutant (MT) (D235A) Xrn2MT. Cells were treated with DRB (100 µM, 2.5 hrs) to inhibit pause release of pol II from promoters and clear polymerases from gene bodies. Paused polymerases were allowed to resume elongation by removal of DRB. Shown is the distribution profile of pol II ChIP-seq signal at time point (T) T₀, T₅, T₁₀ and T₁₅ minutes on the EIF1 gene after DRB wash. Note that pol II piles up immediately downstream of the cleavage and polyadenylation site (marked by the red dotted line). Also note that inhibition of Xrn2 results in stronger pol II pile-up downstream of the CPS (intensity indicated by a red bar) relative to pol II density at the promoter region (intensity indicated by a grey bar). The double-headed arrows illustrate the intensity of pol II pile-up at the 3’ end of the EIF1 gene relative to the intensity of the pol II peak at the promoter region.
Xrn2WT and Xrn2MT cells were treated with DRB (100 µM, 2.5 hrs) to inhibit pause release of pol II from promoters and clear polymerases from gene bodies. Paused polymerases were allowed to resume elongation by removal of DRB. (A) Wave of pol II from the transcription start site (TSS) in Xrn2WT cells with an estimated elongation rate of 2.0 kb/min (B) Wave of pol II from the TSS in Xrn2MT cells with an estimated elongation rate of 2.3 kb/min (C) Overlay of the pol II waves in Xrn2WT and Xrn2MT cells. Shown is the pol II ChIP-seq signal at time point (T) T₀, T₅, T₁₀ and T₁₅ minutes on long genes (50 kb and longer) after DRB wash. The global elongation rate of pol II was estimated by calculating the observed distance traveled by pol II between time points T₅ and T₁₀, as well as between T₁₀ and T₁₅ (indicated by double-headed arrows). The average of these two values was calculated. Note that the elongation rates between Xrn2WT and Xrn2MT cells is not substantially different but pol II density is consistently shifted toward the 3’ of genes in the three time points relative to Xrn2WT cells (~3.7 kb farther downstream at T₁₀ min). This effect can be explained by results from Chapter V. to occur due to release of “creeping” polymerases located farther downstream of promoter regions in Xrn2MT cells under DRB conditions (creeping polymerases are proposed to accumulate downstream of promoters in Xrn2MT cells due to inhibition of Xrn2-mediated premature termination of those polymerases). Rel. Fr: relative frequency of the mean further normalized at signal at position +55 kb from the TSS.
of pol II within the body region of these long genes was inferred by direct observation of the
distance traveled by the front of the wave of pol II between time points T₅ and T₁₀ min, as well as
between T₁₀ and T₁₅ (indicated by double-headed arrows in figure 4-8) and the mean of these two
values was calculated. The calculated elongation rate of pol II within gene bodies was 2.0 kb/min
in Xrn2WT cells and 2.3 kb/min in Xrn2MT cells. The elongation rate between Xrn2WT and
Xrn2MT cells does not appear to be substantially different. Nevertheless, it can be noted that in
each time point after DRB wash out pol II density in Xrn2MT cells is consistently shifted toward
the 3’ of genes relative to Xrn2WT cells (Fig. 4-8C) (~3.7kb farther downstream at T₁₀ min). This
effect can be explained from results in Chapter V. Inhibition of the exonuclease activity of Xrn2
is proposed in Chapter V to inhibit premature termination of “creeping” polymerases that can
slowly migrate to downstream positions from promoter regions under DRB conditions. It is
possible that DRB wash out allows for release of accumulated creeping polymerases with a head
start farther downstream of promoters in Xrn2MT cells, resulting in a wavefront of pol II that is
consistently shifted 3’ relative to that observed in Xrn2WT cells. However, only a single
experiment has been performed for this analysis.

Interestingly, analysis of the wavefront of pol II on genes that are separated from
downstream transcription units by at least +12 kb from their polyA site, revealed that pol II travels
a shorter distance past the polyA site relative to that within body regions over the same window of
time, between T₁₀ and T₁₅ minutes (Fig. 4-9A, B). On short genes, the front wave of pol II can
already be seen crossing the polyA site after 5 minutes from DRB wash out (Fig. 4-9A, B).
Nevertheless, from T₁₀ to T₁₅ minutes, pol II migrated downstream of the polyA site at a rate of
0.58 kb/min in Xrn2WT cells and 0.54 kb/min in Xrn2MT cells (Fig. 4-9A, B), suggesting that
polymerases slow down the rate of transcription elongation past the polyA site up to ~75% before
termination of transcription by Xrn2. This is a relevant result suggesting that the common accumulation of pol II densities at 3’ ends of genes genome-wide (Lian et al. 2008; Fong et al. 2015b; Core et al. 2012; Nojima et al. 2015a), is due to increased residence time of pol II and not simply due to an increase in accessibility of pol II for immunoprecipitation at these sites. Furthermore, this result adds support to the idea that the elongation complex undergoes a polyA site-dependent conformational change that might lead to a slowdown of the elongation complex at 3’ ends of genes.

Notably, inhibition of Xrn2 did not affect the transition of pol II into slow elongation downstream of polyA sites (Fig. 4-9), indicating that degradation of the nascent transcript by Xrn2 is not involved in deceleration of the elongation complex. Nevertheless, consistent with a role of Xrn2 in pol II release from the DNA template, Xrn2MT cells did exhibit a stronger build-up of elongation complexes downstream of the polyA site at each time point after DRB wash out as evidenced on the ACTB and EIF1 genes (Fig. 4-6, 4-7) and at the genome-wide level (Fig. 4-9C) (indicated by red arrows). A biological replicate of stronger pol II build-up downstream of polyA sites at T15 in Xrn2MT cells is shown in figure 4-10. This result further supports the idea that Xrn2 is required for polymerases to be removed from the DNA template and prevent accumulation of pol II at 3’ ends of genes. In conclusion, the results above strongly suggest that pol II transitions at the polyA site into a mode of slow transcription elongation before it becomes target of Xrn2 for termination.

4.2C Slowdown of pol II Downstream of polyA Sites Requires Recognition of the polyA Signal

The interesting observation that pol II slows down the rate of elongation past the polyA site points to a potential role of the polyA signal in inducing deceleration of the elongation complex. To test this idea, I utilized two independent strategies to inhibit polyA signal recognition
Figure 4-9. Pol II slows down transcription elongation downstream of polyA sites
Xrn2WT and Xrn2MT cells were treated with DRB (100 µM, 2.5 hrs) to inhibit pause release of
pol II from promoters and clear polymerases from gene bodies. Paused polymerases were allowed
to resume elongation by removal of DRB. (A) Wave of pol II downstream of the cleavage and
polyadenylation site (CPS) in Xrn2WT cells with an estimated elongation rate of 0.58 kb/min (B)
Wave of pol II downstream of the CPS in Xrn2MT cells with an estimated elongation rate of 0.54
kb/min. (C) Overlay of time points for Xrn2WT and Xrn2MT cells. Shown is pol II ChIP-seq
signal downstream of the CPS of short genes (6kb and shorter) separated from downstream genes
by at least +12 kb from their polyA sites. Time point (T) T0, T5, T10 and T15 minutes after DRB
wash are shown. The global elongation rate of pol II downstream of the CPS was estimated by
calculating the observed distance traveled by pol II between time points T10 and T15, (indicated by
a double-headed arrow). Note that the elongation rates between Xrn2WT and Xrn2MT cells
downstream of the CPS are not substantially different but pol II occupancy is stronger in Xrn2MT
cells in each time point (illustrated by red arrows in panel C). A biological replicate of increased
pol II density downstream of the CPS in Xrn2MT cells at T15 is shown in figure 4-10. Plotted are
mean distribution profiles of pol II normalized to signal at position -1kb from the TSS.
and ask the question of whether accumulation of pol II is reduced at 3’ ends of genes. The first strategy involved overexpression of the flu virus NS1A protein which competes with the polyA signal for binding to CPSF30 (Twu et al. 2006). Consistent with inhibition of binding of 3’-end processing factors to the polyA signal, overexpression of NS1A caused major cell death after 72 hours from its induction (Fig. 2-2).

Interestingly, expression of NS1A for 24 hours eliminated the pol II build-up immediately downstream of the polyA site of the ACTB and EIF1 genes and led to increased pol II occupancy farther away from the 3’ end of these genes indicative of a termination defect (Fig. 4-11), indicated by blue arrows. This result suggests that recognition of the polyA signal by 3’-end processing factors is involved in decelerating the elongation complex at 3’ ends of genes and, therefore, in generation of the normal 3’ peak of pol II downstream of polyA sites. This shift in pol II occupancy toward more 3’ distal positions upon inhibition of polyA signal recognition can be seen at the genome-wide level on thousands of genes in two independent experiments (Fig. 4-12). Note that inhibition of CPSF30 binding to the polyA signal by expression of NS1A, in contrast to inhibition of Xrn2 (Fig. 4-1), causes loss of the 3’ peak of pol II density downstream of the ACTB gene (Fig. 4-11A). This difference indicates that Xrn2 and 3’-end processing factors limit transcription farther away from polyA sites via different mechanisms. Taking into account the observation above that the exonuclease activity of Xrn2 is not required for deceleration of the elongation complex (Fig. 4-9) but for release of pol II from the DNA template downstream of the polyA site, I interpret the data obtained up to this point to indicate that 3’-end processing factors facilitate termination of pol II by slowing down the elongation complex and shortening the path of Xrn2 to chase pol II.

I next asked whether direct mutation of a polyA consensus sequence recapitulates loss of the 3’ peak of pol II. To answer this question, I employed the CRISPR/Cas9 system to mutate the
polyA signal at the 3’ end of the PSMG1 gene as graphically illustrated in Figure 2-3, (Doudna and Charpentier 2014). To simplify gene editing and analysis, I used Hap1 cells that are mostly haploid. On the other hand, the PSMG1 gene was selected for editing on the basis that it contains a single reported polyA signal and that it is not classified as an essential gene (Shalem et al. 2014). The humanized *Streptococcus pyogenes* Cas9 (hSpCas9) endonuclease was expressed in cells from the pX330-U6-Chimeric_BB-CBh-hSpCas9 vector that additionally directs expression of a guide RNA (gRNA) for specific hSpCas9 DNA cleavage at the polyA signal of the PSMG1 gene in this experiment (Ansai and Kinoshita 2014). This plasmid was co-transfected with a single stranded ssDNA donor oligo for homology-directed recombination at the DNA double strand break. In place of the consensus AAUAAA sequence the ssDNA donor contained a BamH1 restriction site to mutate the PSMG1 polyA signal. As expected, the clonal population of successfully edited Hap1 cells, PSMG1 pA-MT cells, yield a PCR amplicon from the 3’ end of the PSMG1 gene that is sensitive to BamH I digestion (Fig 2-5). Mutation of the physiological PSMG1 polyA signal was further confirmed by sequencing (Fig. 2-4).

Interestingly, parent Hap1 cells exhibit the characteristic 3’-peak of pol II density downstream of the wild-type PSMG1 polyA site (Fig. 4-13). However, PSMG1-pAMT cells do not display such early accumulation of pol II density downstream of the mutated polyA site. Instead, accumulation of pol II takes place ~6 kb farther downstream relative to control parent cells (Fig. 4-13), indicated by a blue arrow. Two biological replicates of that observation are shown. Elimination of the pol II build-up immediately downstream of the mutated polyA signal indicates that usage of a functional polyA signal is responsible for both triggering slowdown of pol II at 3’ ends of genes and limiting transcription farther downstream of the PSMG1 gene.
Figure 4-10. Inhibition of Xrn2 generates stronger accumulation of polymerases downstream of polyA sites – Biological replicate

Xrn2WT and Xrn2MT cells were treated with DRB (100 µM, 2.5 hrs) to inhibit pause release of pol II from promoters and clear polymerases from gene bodies. Paused polymerases were allowed to resume elongation by removal of DRB. (A) Pol II ChIP-seq signal on the ACTB gene in DRB conditions (DRB T₀) and 15 minutes after removal of DRB (T₁₅min). Note increased accumulation of pol II downstream of the polyA site (marked by the red dotted line) relative to signal around the transcription start site (TSS) in Xrn2MT cells when compared to control Xrn2WT cells (the short double-headed arrow indicates a smaller 3’ peak of pol II relative to the peak of pol II at the TSS in Xrn2WT cells. (B) Metaplot showing accumulated pol II density around promoters after inhibition of pol II escape from the pause with DRB (genes at least 33kb and longer). (C). Pol II density downstream of the cleavage and polyadenylation site (CPS) 15 minutes (T₁₅min) after DRB-wash. Genes correspond to 6kb long and shorter, separated from downstream transcription units by at least +12 kbs from their polyA sites. Note stronger accumulation of pol II downstream of the CPS in Xrn2MT cells when compared to Xrn2WT cells (indicated by a red arrow). Biological replicate of the experiment shown in figure 4-9.
Figure 4-11. Recognition of the polyA signal is required for slowdown of pol II downstream of the ACTB and EIF1 genes

Flp-in TREX cells were incubated with doxycycline for 24hrs to induce expression of NS1A and inhibit polyA signal recognition. Cells not treated with doxycycline (No-Dox) and NS1A-expressing cells were used for analysis. (A) Pol II ChIP-seq on the ACTB gene. (B) Pol II ChIP-seq on the EIF1 gene. Note that in the presence of the NS1A protein the peak of pol II density immediately downstream of the cleavage and polyadenylation site (marked by the red dotted line) is dramatically reduced and pol II density continues farther downstream relative to the No-Dox condition. UCSC genome browser screen shot. Blue arrow indicates the direction of transcription.
Figure 4-12. Recognition of the polyA signal is required for slowdown of pol II downstream of polyA sites

Flp-in TREX cells were incubated with doxycycline for 24hrs to induce expression of NS1A and inhibit polyA signal recognition. Cells not treated with doxycycline (No-Dox) and NS1A-expressing cells were used for analysis. (A) Pol II ChIP-seq on most active genes (top 25% based on pol II density within gene bodies in normal conditions) aligned at the cleavage and polyadenylation site (CPS) (marked by the red dotted line). (B) Biological replicate. Pol II ChIP-seq on all RefSeq genes. Note that pol II occupancy immediately downstream of the CPS is dramatically reduced and increases at distal positions farther downstream after inhibition of polyA signal recognition by expression of NS1A.
The CRISPR-Cas9 system was used to mutate the polyA signal of the PSMG1 gene in Hap1 cells (mostly haploid). A clone of successfully modified PSMG1 polyA mutant (pA-MT) cells was selected for analysis (see Chapter II). (A) Pol II ChIP-seq density on the PSMG1 gene. Downstream of the mutated polyA signal (marked by the blue dotted line), 14 polyA consensus hexamer sequences of AAUAAA (pA) could be detected before accumulation of pol II density. 7 of those are spread over a region of ~6 kb (marked by the red dotted line) and 7 clustered at the beginning of the pol II build-up (between the two red upward arrows). Two biological replicates are shown. Note a delay in pol II build-up downstream of the PSMG1 gene in Hap1 pA-MT cells compared to parent Hap1 cells (indicated by blue arrows). (B) The location and DNA sequence of the first pA and last pA in the cluster at the beginning of the pol II build-up are shown. UCSC genome browser screen shot. Blue arrow indicates direction of transcription.
Importantly, inspection of the nucleotide sequence spanning the novel pol II pausing zone revealed the presence of multiple canonical AAUAAA sequences that might serve as potential cryptic polyA signals. 7 of those were found to be spread over ~6 kb downstream of the PSMG1 gene (marked by a red dotted line) (Fig. 4-13). Immediately upstream of the novel peak of pol II in PSMG1-pAMT cells, 7 additional AAUAAA sequences were clustered within a short window of ~1kb (Fig. 4-13) (between the two upward arrows). The location and surrounding sequence for the first and last hexamer sequence in this cluster is shown in Figure 4-13B. The fact that mutation of the physiological polyA signal of the PSMG1 gene eliminates accumulation of pol II at that site but results in a novel peak farther downstream, where multiple sequences might serve as cryptic polyA signals, suggests that usage of a polyA signal might be sufficient to induce slowdown of the elongation complex. Taken together the results above suggest that, in addition to direct cleavage and polyadenylation of the pre-mRNA, binding of 3’-end processing factors to the polyA signal on the nascent transcript is responsible for inducing deceleration of transcription elongation at 3’ ends of genes.

4.2D Spt5 Transitions Into a Hypo-Phosphorylated State in the Elongation Complex Downstream of polyA Sites

A polyA site-dependent deceleration of transcription elongation at 3’ ends of genes strongly suggests that elongation complexes undergo a conformational change as proposed in the allosteric model of transcription termination. To investigate a potential conformational change of the elongation complex I examined the pol II-associated factor Spt5, known to participate in repression of transcription elongation by inducing pol II pausing at promoter regions (Missra and Gilmour 2010; Yamaguchi, Takagi, et al. 1999) and thought to switch into a positive elongation factor that contributes to processivity of the elongation complex over gene bodies when
phosphorylated by Cdk9 (Yamada et al. 2006). It is possible that a conformational change of the elongation complex involves a switch of Spt5 back into a hypo-phosphorylated state that might return its repressive function on transcription elongation.

To investigate this possibility, I first analyzed the distribution profile of Spt5 and its phosphorylation levels throughout transcription units by ChIP-seq. Importantly, human cultured cells were spiked in with mouse M12 cells to control for immunoprecipitation efficiency. Normalization of ChIP-seq libraries using a spiked-in internal control allows for assessment of the absolute levels of ChIP-seq signal between control and test samples (Hu et al. 2015). Notably, the distribution profile of total Spt5 closely follows the distribution profile of pol II on the ACTB and MYC genes, consistent with its direct association with pol II (Fig. 4-14). In contrast, using an antibody that preferentially binds the phosphorylated form of Spt5 (PO$_{4}$-Spt5, provided by B. Erickson) generates a distribution profile that differs from that of total Spt5 at the end of the ACTB and MYC genes (Fig. 4-14). In particular, PO$_{4}$-Spt5 signal drops rapidly past the polyA site, while total Spt5 increases.

Meta-gene plots on thousands of genes clearly show comparable distribution profiles of total and PO$_{4}$-Spt5 around promoters and gene bodies with strong accumulation of both signals around promoter regions (Fig. 4-15A). This result suggests that Spt5 becomes highly phosphorylated around promoter regions, consistent with the model of pol II pause release in which phosphorylation of Spt5 and pol II by Cdk9 at the promoter-proximal region is required for a transition of pol II from a paused state into productive elongation (Wada, Takagi, Yamaguchi, Watanabe, et al. 1998; Yamada et al. 2006). Furthermore, normalization of ChIP-seq signals to the M12 spike-in and additional calculation of the log$_{2}$ value for the ratio of PO$_{4}$-Spt5 over total Spt5 revealed that Spt5 phosphorylation levels are lowest around the TSS but sharply increase at distal
positions from the TSS and reaches a maximum level at ~+2 kb into gene bodies (Fig. 4-16). Two biological replicates of this observation are shown. At 3’ ends of genes, however, only total Spt5 and pol II accumulate downstream of polyA sites with a similar distribution profile, but PO$_4$-Spt5 does not (Fig. 4-15B, C). This observation suggests that Spt5 remains bound to the elongation complex at 3’ ends of genes, but it loses its phosphorylation levels downstream of polyA sites.

To further investigate an association of Spt5 in a hypo-phosphorylated state with the elongation complex past the polyA site, I decided to visualize elongation complexes that accumulate at 3’ ends of genes in Xrn2MT cells upon inhibition of transcription initiation. If Spt5 transitions into a hypo-phosphorylated state at 3’ ends of genes, and remains bound to the elongation complex, accumulation of polymerases is expected to result in a build-up of total Spt5 but not PO$_4$-Spt5 build-up downstream of the polyA site. As expected, triptolide treatment (10 µM, 20 min) led to loss of all three factors in Xrn2WT cells downstream of the polyA site of the ACTB gene (Fig. 4-17) (indicated by red arrows). However, Xrn2MT cells accumulated both pol II and total Spt5 downstream of the polyA site of the ACTB gene, and this effect was not accompanied by accumulation of PO$_4$-Spt5. Strong accumulation of total Spt5 downstream of the polyA site, with no accumulation of PO$_4$-Spt5, was observed on thousands of genes in two biological replicates (Fig 4-18). This result further suggests that Spt5 remains bound to the elongation complex but transitions into a hypo-phosphorylated state downstream of polyA sites.

Interestingly, previous experiments have shown that Spt5 requires an RNA transcript longer than 18 nucleotides for recruitment to a reconstituted pol II elongation complex in vitro (Missra and Gilmour 2010). Consistent with that observation, inhibition of transcription initiation with triptolide in the experiment above resulted in loss of total Spt5 and PO$_4$-Spt5 at the promoter region of the ACTB gene, in spite of the fact that pol II can still be detected at this site (Fig. 4-17).
Figure 4-14. Spt5 accumulates at the 3’ end of the ACTB and MYC genes in a hypo-phosphorylated state

HEK293 cells were crosslinked and cell lysates prepared for immunoprecipitation of pol II, total Spt5 and phosphorylated Spt5 (PO₄-Spt5). Shown are the Pol II, total and PO₄-Spt5 ChIP-seq profiles on the ACTB (A) and MYC gene (B). Note that pol II and total Spt5 track together with similar profiles and accumulate at 3’ ends of genes. However, Spt5 phosphorylation levels drop downstream of the polyA site (marked by the red dotted line). UCSC genome browser screen shot. Blue arrow indicates direction of transcription.
Figure 4-15. Spt5 accumulates at 3’ ends of genes in a hypo-phosphorylated state
HEK293 cells were crosslinked and cell lysates prepared for immunoprecipitation of pol II, total Spt5 and phosphorylated Spt5 (PO$_4$-Spt5). (A) Total Spt5 and PO$_4$-Spt5 ChIP-seq signal on most active genes (Top 25% with highest pol II density within their body region). Genes were aligned at the transcription start site (TSS) and the cleavage and polyadenylation site (CPS). The variable body region (0.5kb downstream of the TSS and 0.5kb upstream of the CPS) was normalized according to gene size. (B) Pol II, total Spt5 and PO$_4$-Spt5 ChIP-seq signal on all RefSeq genes aligned at the CPS. Note accumulation of pol II and total Spt5, but no enrichment of PO$_4$-Spt5 downstream of the CPS (C) Biological replicate of total Spt5 and PO$_4$-Spt5 ChIP-seq signal at 3’ ends of most active genes. Mean ChIP-seq signals are shown normalized to signal at position - 1kb in panel A and B and to signal at position - 0.5kb in panel C.
Figure 4-16. Spt5 phosphorylation increases upstream and downstream of the transcription start site of genes
Shown is the distribution profile of Spt5 phosphorylation levels. PO₄-Spt5 ChIP-seq signal was normalized to total Spt5 ChIP-seq signal by calculating the Log₂ ratio of (PO₄-Spt5/total Spt5). Note that Spt5 phosphorylation levels increase upstream and downstream of the TSS and plateaus at ~2 kb within the first 10 kb of gene bodies. Normalized to M12 spike-in (N. M12). All RefSeq genes were aligned at the transcription start site (TSS) (n=19568). Two biological replicates are shown.
Figure 4-17. Elongation complexes accumulate at the 3’ end of the ACTB gene with hypo-phosphorylated Spt5

HEK293 Flp-in TREX cells were treated with doxycycline for 24 hrs to induce expression of wild-type (WT) Xrn2WT or exonuclease-dead, mutant (MT) (D235A) Xrn2MT. Cells were treated with triptolide (10 µM, 20 min) to inhibit transcription initiation and visualize elongation complexes that fail to be released from the DNA template in a timely manner downstream of the polyA site (marked by the red dotted line) of the ACTB gene in Xrn2MT cells. Shown are pol II, total and phosphorylated (PO₄-Spt5) ChIP-seq signals on the ACTB gene under triptolide conditions for Xrn2WT and Xrn2MT cells. Note accumulation of pol II and total Spt5 downstream of the polyA site in Xrn2MT cells (indicated by red arrows), which is not accompanied by signal of PO₄-Spt5. ChIP-seq signal was normalized (N.) to an M12 mouse spike-in (N. M12). UCSC genome browser screen shot. Blue arrow indicates direction of transcription.
Figure 4-18. Elongation complexes accumulate at 3’ ends of genes with hypo-phosphorylated Spt5

HEK293 Flp-in TREX cells were treated with doxycycline for 24 hrs to induce expression of wild-type (WT) Xrn2WT or exonuclease-dead, mutant (MT) (D235A) Xrn2MT. Cells were treated with triptolide (10 µM, 15 min) to inhibit transcription initiation and visualize elongation complexes that fail to be released from the DNA template in a timely manner downstream of the cleavage and polyadenylation site (CPS) (marked by the red dotted line) in Xrn2MT cells. Shown is ChIP-seq signal for the indicated factors in Xrn2MT cells relative to signal in Xrn2WT cells after 15 min of treatment with triptolide. ChIP-seq signal was first normalized to a mouse M12 spike-in and the log2(Xrn2MT) – log2(Xrn2WT) value was calculated and plotted. ChIP-seq density was further normalized to signal at -1 kb upstream of the CPS. Note that Xrn2MT cells accumulate pol II and total Spt5 downstream of the CPS relative to Xrn2WT cells, but PO4-Spt5 does not accumulate. Two independent experiments are shown. Run1 is shown on all RefSeq genes and Run2 on most active genes (Top 25% with highest pol II density within their body region).
Loss of total Spt5 around promoters occurred on thousands of genes upon treatment with triptolide (Appendix B, Fig. B-1). Therefore, it can be concluded that recruitment of Spt5 to promoter regions in human cells requires synthesis of a pre-mRNA transcript by pol II. This conclusion is further supported by structures of Spt5 in the elongation complex (Bernecky, Plitzko et al. 2017, Ehara, Yokoyama et al. 2017).

4.2E Recognition of the polyA Signal is Required for Loss of Spt5 Phosphorylation Downstream of polyA Sites

Results above demonstrate that both deceleration of pol II and loss of Spt5 phosphorylation coincide at 3’-ends of genes. Thus, it is possible that recognition of the polyA signal is also involved in loss of Spt5 phosphorylation. To investigate this possibility, first I decided to test the prediction that inhibition of polyA signal recognition would prevent loss of Spt5 phosphorylation at 3’ ends of genes. For this experiment, I turned back to the NS1A inducible system. As predicted, NS1A expression led to increased total Spt5 and PO4-Spt5 signal farther downstream of transcription units as shown on the ACTB gene (Fig. 4-19) (indicated by red arrows). Furthermore, normalization of PO4-Spt5 to total Spt5 signal revealed that NS1A expression blocks the normal drop in phosphorylation levels downstream of polyA sites (Fig. 4-20). Thus, inhibition of polyA signal recognition not only prevents accumulation of pol II, but also leads to an inability of Spt5 to transition into a hypo-phosphorylated state at the polyA site.

To further investigate if loss of Spt5 phosphorylation at 3’ ends of genes depends on recognition of a functional polyA signal, I examined total Spt5 and PO4-Spt5 densities around the polyA site of the PSMG1 gene in PSMG1-pAMT cells. First, consistent with an association of Spt5 with slow moving elongation complexes at 3’ ends of genes, it could be observed that the novel peak of pol II downstream of the mutant polyA site is accompanied by a strong peak of total
Spt5, with a similar profile to that of pol II (Fig. 4-21) (indicated by red arrows). Interestingly, mutation of the physiological polyA signal at the 3’ end of the PSMG1 gene also led to a shift of PO₄-Spt5 density toward 3’ distal positions indicative of increased levels of PO₄-Spt5 farther downstream of the mutant polyA site (Fig 4-21), illustrated by a grey dotted line. However, a single experiment of this observation has been made. Although the low signal to noise ratio precludes visualization of a clear profile of PO₄-Spt5, the shift of PO₄-Spt5 signal toward more 3’ downstream positions suggests that a functional polyA signal is required to limit Spt5 phosphorylation levels past the polyA site. Taking into account the recent observations in yeast *S. pombe*, suggesting that Dis2, an isoform of protein phosphatase 1 (PP1), is able to dephosphorylate Spt5 *in vitro* and that it is required to prevent transcription beyond the normal termination zone (Parua et al. 2018), the results above point to polyA site-dependent dephosphorylation of Spt5 as a potential mechanism contributing to deceleration of the elongation complex at 3’ ends of genes. I speculate that recognition of the polyA signal by 3’-end processing factors triggers dephosphorylation of Spt5 reverting back its repressive activity on transcription elongation, which in turn leads to deceleration of the elongation complex downstream of polyA sites.

**4.3 Discussion**

**4.3A Pol II Decelerates Transcription Elongation Downstream of polyA Sites and Becomes Target of Xrn2**

According to the torpedo model an elongating pol II is chased by Xrn2 through degradation of the nascent transcript after cleavage at the polyA site. This scenario implies a kinetic competition in which elongation by pol II is overcome by processive exonuclease digestion. Nevertheless, results in this study have revealed that the major effect of inhibiting Xrn2 on pol II occupancy, is accumulation of pol II density immediately downstream of the polyA site not
Figure 4-19. Recognition of the polyA signal is required for loss of Spt5 phosphorylation downstream of the polyA site of the ACTB gene

Flp-in TREX cells were treated with doxycycline for 24 hours to induce expression of NS1A and inhibit polyA signal recognition. Shown is the distribution profile of pol II, total Spt5 and phosphorylated Spt5 (PO₄-Spt5) before and after induction of NS1A. Note increased occupancy of pol II, total Spt5 and PO₄-Spt5 farther downstream of the polyA site (marked by the red dotted line) after expression of NS1A (indicated by red arrows). Normalized to M12 spike-in (N. M12). UCSC genome browser screen shot. Blue arrow indicates direction of transcription.
Figure 4-20. Recognition of the polyA signal is required for loss of Spt5 phosphorylation downstream of the polyA sites
Flp-in TREX cells were treated with doxycycline for 24 hours to induce expression of NS1A and inhibit polyA signal recognition. To visualize levels of Spt5 phosphorylation downstream of cleavage and polyadenylation sites (CPS), PO4-Spt5 signal was normalized to total Spt5. First, ChIP-seq signals were normalized (N.) to a mouse M12 spike-in and then the log2(PO4-Spt5) – log2(Total Spt5) value was plotted for control (Ctrl), no doxycycline (No-Dox), and for NS1A expression conditions. Two biological replicates are shown on genes separated from other transcription units at least 5 kb from the CPS. Signal was normalized at the CPS in both experiments. Note the drop in Spt5 phosphorylation levels downstream of the CPS in control cells that is inhibited upon expression of NS1A (illustrated by red arrows).
Figure 4-21. The polyA signal at the 3’ end of the PSMG1 gene is required for early build-up of pol II and total Spt5 at its 3’ end

Shown is the distribution profile of Pol II, total Spt5 and phosphorylated Spt5 (PO$_4$-Spt5) on the PSMG1 gene for parent Hap1 cells and PSMG1 polyA mutant (pA-MT) cells. Note a delay in pol II build-up downstream of the PSMG1 gene in PSMG1 pA-MT cells relative to control Hap1 cells. A biological replicate of this effect is shown in figure 4-13. Also note that build-up of Spt5 downstream of the PSMG1 gene is also delayed and tracks with pol II density (indicated by red arrows). Although PO$_4$-Spt5 has a low signal to noise ratio, PO$_4$-Spt5 occupancy additionally appears to shift toward positions farther downstream of the PSMG1 gene (illustrated by a change in the profile of the grey dotted line indicating levels of PO$_4$-Spt5). UCSC genome browser screen shot. Blue arrow indicates direction of transcription.
increased transcription father downstream of termination zones (Fig. 4-1, 4-6, 4-7, 4-9, 4-10). This suggests that pol II does not efficiently elongate past the polyA site and becomes an “easy” target for Xrn2.

Investigation of the elongation rate of pol II downstream of the polyA site is particularly challenging since pol II occupancy within termination zones is dynamically maintained by ongoing pol II disassembly and continuous repopulation of arriving polymerases from multiple rounds of transcription. Here, the global elongation rate of pol II was calculated within body regions and downstream of polyA sites by release of promoter-proximally paused polymerases from a DRB block before and after inhibition of Xrn2 (Fig 4-8, 4-9). The calculated, global elongation rate of pol II within gene bodies was 2.0 kb/min and 2.3 kb/min in Xrn2WT and Xrn2MT cells respectively. These rates of transcription elongation are consistent with previous reported elongation rates of pol II which range from 1.1kb/min to 3.0kb/min (Ardehali et al. 2009; O'Brien and Lis 1993; Danko et al. 2013; Fuchs et al. 2014; Jonkers, Kwak, and Lis 2014). Importantly, elongation by pol II downstream of polyA sites was found to be dramatically reduced in both Xrn2WT and Xrn2MT cells to 0.58 kb/min and 0.54 kb/min respectively (Fig. 4-9). This drop in the rate of transcription elongation, of up to ~75% of that of pol II within gene bodies, suggests that the usual peak of pol II at 3’ ends of genes in yeast and mammalian cells is generated by increased residence time of pol II at these sites and is not a consequence of increased pol II accessibility for immunoprecipitation. Interestingly, the exonuclease activity of Xrn2 was not required for deceleration of the elongation complex at 3’ ends of genes (Fig. 4-9), suggesting that torpedo action by Xrn2 is not needed for slowdown of elongation complexes beyond the polyA site.
In particular, previous studies have argued for a role of Xrn2 in termination in yeast and mammalian cells based on the observation that pol II or nascent transcripts accumulate farther downstream of polyA sites after depletion of Xrn2 (Baejen, Andreani, Torkler, Battaglia, Schwalb, Lidschreiber, Maier, Boltendahl, Rus, Esslinger, Söding, et al. 2017; West, Gromak, and Proudfoot 2004; Gromak, West, and Proudfoot 2006; Eaton et al. 2018) or inhibition of its exonuclease activity (Fong et al. 2015b; Kim et al. 2004). In this investigation, however, I additionally show that Xrn2 is required for efficient release of pol II from the DNA template by direct observation of stabilized elongation complexes that fail to be removed downstream of polyA sites in Xrn2MT cells (Fig. 4-2 – 4-5). Importantly, stabilized elongation complexes do not appear to efficiently translocate on the DNA template over time (Fig. 4-2 – 4-5), suggesting that pol II transitions at the polyA site into a mode of slow transcription elongation before it becomes a target of Xrn2. Stronger support for a role of Xrn2 in release of pol II at 3’ ends of genes comes from the observation that arrival of polymerases downstream of polyA sites results in pol II build-up that is consistently stronger in Xrn2MT cells at different time points after removal of the block in elongation by DRB (Fig. 4-6, 4-7, 4-9C, 4-10).

4.3B Recognition of the polyA Signal by 3’-End Processing Factors Instructs Slowdown of the Elongation Complex

The role of the polyA signal in termination of pol II transcription has been well established (Whitelaw and Proudfoot 1986; Connelly and Manley 1988b; Logan, Falck, et al. 1987). At least in part, polyA-signal directed cleavage of the nascent transcript has explained its role in termination by providing an entry point for the exonuclease Xrn2 (Connelly and Manley 1988a). However, my results strongly suggest that an additional function of the polyA signal is to instruct a slowdown to the elongation complex at 3’ ends of genes. Overexpression of the viral NS1A
protein, which blocks direct binding of CPSF30 to the polyA signal (Twu et al. 2006; Nemeroff et al. 1998), prevented accumulation of pol II density at 3’ ends of genes and shifted pol II occupancy to downstream distal positions (Fig. 4-11, 4-12). Likewise, mutation of the human PSMG1 polyA signal by CRISPR delayed accumulation of pol II density approximately 6 kilobases until a region occupied by multiple AAUAAA hexamer sequences that might serve as cryptic polyA signals for RNA 3’-end processing (Fig. 4-13). The observation that accumulation of pol II shifted from the mutated polyA signal to a downstream region occupied by a cluster of multiple sequences that might serve as cryptic polyA sites (Fig. 4-13) further indicates that usage of a polyA signal might be sufficient to induce slowdown of the elongation complex.

A role of 3’-end processing factors in inducing deceleration of pol II downstream of the polyA site can easily explain the observation that knockdown of 3’-end processing factors generates a more profound termination defect in which pol II transcribes farther away from polyA sites than inhibition of Xrn2 (Nojima et al. 2015a; Baejen, Andreani, Torkler, Battaglia, Schwalb, Lidschreiber, Maier, Boltendahl, Rus, Esslinger, Soding, et al. 2017; Eaton et al. 2018). According to my results, the exonuclease activity of Xrn2 is involved in dissociation of polymerases from the DNA template after elongation complexes have already slowed down at 3’ ends of genes (Fig. 4-1 – 4-5, 4-9). In that scenario, inhibition of 3’-end processing factors is expected to cause a different and opposite termination defect with respect to that caused by inhibition of Xrn2, which operates at a downstream step in the termination process. While inhibition of polyA signal recognition eliminates the “brakes” of the elongation complex and causes pol II to transcribe farther away from 3’ ends of genes, inhibition of Xrn2 prevents DNA release by elongation complexes that have been efficiently instructed by 3’-end processing factors to slow down leading to a pile-up of pol II immediately downstream of polyA sites. Thus, accumulation of poorly elongating polymerases
downstream of polyA sites after inhibition of Xrn2 explains the strong densities of pol II at 3’ ends of genes in Xrn2MT cells and the modest increase in transcription beyond normal termination zones (Fig. 4-1).

An additional interesting observation after release of paused polymerases into elongation by DRB wash out, was that arrival of polymerases downstream of the polyA site of the ACTB and EIF1 genes generated a gradual build-up of pol II density over a broad range of several kilobases (Fig. 4-6, 4-7), rather than over a more discrete region localized at the polyA site. (Fig. 4-5 – 4-7). In agreement with a previous report (Steven J. (Kim and Martinson 2003), this appears to suggest a stochastic component in the mechanism triggering slowdown of pol II. While some polymerases might efficiently slow down immediately downstream of the polyA site, others might be able to reach several kilobases before recognition of the polyA signal triggers the switch into an elongation arrest. It is possible that binding of 3’-end processing factors to the polyA signal does not always occur immediately after extrusion of the polyA signal. For instance, recruitment of the cleavage factor CstF, which is part of the 3’-end processing machinery (introduced in Chapter I), can be rate-limiting for cleavage and polyadenylation in B cells and changes in cellular CstF levels can alter the choice between two potential polyA sites at the 3’ end of the pre-mRNA (Takagaki et al. 1996). A stochastic component in recognition of polyA signals would be consistent with the observation that over 50% of genes display alternative polyA site usage (Tian et al. 2005). A relevant additional purpose of slowing down pol II after commitment to process at a polyA signal, might be to ensure processing at the chosen polyA site by preventing elongation farther down and synthesis of potential downstream competing signals.
Notably, phosphorylation of Spt5 by Cdk9 is thought to switch Spt5 from a pol II pausing factor at the promoter-proximal region into a positive elongation factor (Yamada et al. 2006) that can contribute to processivity of pol II elongation downstream of the HIV-1 promoter by stabilizing the elongation complex on the DNA template (Ping and Rana 2001; Bourgeois et al. 2002b). Consistent with a positive role of Spt5 phosphorylation during elongation, I found that levels of Spt5 phosphorylation rapidly increase within gene bodies downstream of promoter regions (Fig. 4-16). Interestingly, similar to the transition of pol II into productive elongation characterized by concomitant phosphorylation of Spt5, the transition into a slowdown at the polyA site is also accompanied by a reversal of Spt5 into a hypo-phosphorylated state that is dependent on the polyA signal (Fig. 4-14, 4-15, 4-17 – 4-21). Thus, this observation allows for the speculation that recognition of the polyA signal might behave as a switch capable of reverting Spt5 into a hypo-phosphorylated state that is required to slowdown elongation. In agreement with that speculation, it has been recently proposed from studies in yeast S. pombe that Dis2, an isoform of protein phosphatase 1 (PP1), is able to dephosphorylate Spt5 in vitro and that it is required to prevent transcription beyond the normal termination zone at 3’ ends of genes (Parua et al. 2018). Additionally, deletion of the C-terminal domain of Spt5, a domain target of Cdk9 for phosphorylation (Yamada et al. 2006), led to increased synthesis of nascent transcripts downstream of polyA sites in budding yeast, suggesting that the C-terminal domain of Spt5 is required for proper termination of pol II (Baejen, Andreani, Torkler, Battaglia, Schwalb, Lidschreiber, Maier, Boltendahl, Rus, Esslinger, Söding, et al. 2017). It is possible that loss of Spt5 phosphorylation activates the repressive activity of Spt5 on transcription elongation, leading...
to slowdown of pol II in a mechanism similar to that for promoter-proximal pausing of pol II by Spt5. Previous reports have shown that the repressive function of Spt5 on transcription elongation at the promoter region does not always require NELF (Ainbinder et al. 2004; Baugh, Demodena, and Sternberg 2009), pointing to the possibility that interaction of Spt5 in a hypo-phosphorylated state with the elongation complex is sufficient to repress its translocation on the DNA template. Indeed, binding of Spt5 to the nascent transcript right after extrusion from the active site at the promoter region is thought to allow Spt5 to trigger pausing of pol II (Missra and Gilmour 2010). Significantly, a sharp increase in PAR-CLIP signal of Spt5 binding to the nascent transcript immediately downstream of the polyA site has been reported in budding yeast (Baejen, Andreani, Torkler, Battaglia, Schwalb, Lidschreiber, Maier, Boltendahl, Rus, Esslinger, Söding, et al. 2017). This observation further indicates that simultaneous binding of hypho-phosphorylated Spt5 to pol II and the nascent transcript, as it is thought to occur during pausing of pol II around promoters, might be a common conformation of the elongation complex with a repressive effect on elongation during transcription. The multiple interactions of Spt5 with pol II (Martinez-Rucobo et al. 2011; Klein et al. 2011; Sevostyanova and Artsimovitch 2010; Ehara et al. 2017; Bernecky, Plitzko, and Cramer 2017), or simultaneous binding to pol II and the non-template DNA strand within the paused transcription bubble (Yakhnin, Murakami, and Babitzke 2016) is consistent with the possibility that this highly conserved factor may directly induce a specific pol II conformational change conducive to pausing.

Taken together, these data suggest that binding of cleavage-polyadenylation factors to the polyA signal trigger a transition of the elongation complex into a mode of slow elongation rate. This is consistent with the allosteric model in which loss of an antitermination factor or gain of a termination factor leads to pausing and termination. Nevertheless, I have found that slowdown of
pol II does not render the elongation complex unstable and it remains bound to chromatin if it is not removed by Xrn2. Thus, I propose an allosteric-torpedo model of pol II transcription termination in mammalian cells (Fig. 4-22). In this model, transcription of pol II through a functional polyA site allows for binding of 3’-end processing factors to the nascent transcript on the polyA signal, which triggers a conformational change of the elongation complex involving dephosphorylation of Spt5 and deceleration of transcription elongation. This slowdown renders the elongation complex susceptible to termination by Xrn2, at least in part by facilitating catching-up of this exonuclease with pol II.
Figure 4-22. The allosteric-torpedo model of RNA pol II transcription termination

The transcription elongation complex travels along gene bodies associated with Xrn2 and phosphorylated Spt5. After transcription of the polyadenylation (polyA) signal, 3’-end processing factors bind the polyA signal on the RNA and trigger both dephosphorylation of the elongation factor Spt5 and deceleration of the elongation complex to facilitate chase of pol II by Xrn2. Cleavage of the nascent transcript at the cleavage and polyadenylation site (CPS) allows access of Xrn2 to degrade the uncapped, 5’-monophosphorylated transcript and chase pol II for termination.
CHAPTER V
PREMATURE TERMINATION AND CONTROL OF RNA POLYMERASE II
TRANSCRIPTION BY XRN2

5.1 Introduction

Traditionally, it was considered that recruitment of RNA polymerases to promoters was a limiting step in the control of gene expression. Nevertheless, analysis of the uninduced Drosophila heat shock protein (HSP) genes revealed that transcriptionally engaged polymerases accumulate immediately downstream of the HSP promoters associated with short transcripts ranging in length from 20 to 60 nucleotides (nts) (Gilmour and Lis 1986; Rougvie and Lis 1988; Giardina, Perez, and Lis 1992; Rasmussen and Lis 1993). This population of polymerases that accumulate at promoter-proximal regions were then referred to as “paused” polymerases (Rougvie and Lis 1990).

Additional work performed in mammalian cells demonstrated that transcriptionally engaged pol II also accumulates downstream of human promoters, including that of c-myc and Fos, adding strength to the idea of promoter-proximal pausing of pol II (Strobl and Eick 1992; Plet, Eick, and Blanchard 1995; Krumm et al. 1992; Bentley and Groudine 1986). In contrast to these observations, studies in yeast revealed that recruitment of pol II to promoters was a major mode of gene regulation and no evidence for promoter-proximal pausing of pol II was found (Stargell and Struhl 1996; Ptashne and Gann 1997). However, with the recent advances in sequencing technology it is clear that pausing of pol II after transcription of the first 30 to 70 nucleotides is a major limiting step in multicellular organisms. For instance, this bottleneck has been visualized by GRO-seq (Core, Waterfall, and Lis 2008), PRO-seq (Kwak et al. 2013) and ChIP-seq (Zeitlinger et al. 2007; Muse et al. 2007) in which maps of pol II positions demonstrate global accumulation of pol II densities downstream of the transcription start site (TSS).
Additionally, these methods have shown that paused polymerases are competent to undergo elongation, given that paused polymerases can resume elongation in vitro after treatment with the detergent sarkosyl (Core, Waterfall, and Lis 2008; Rougvie and Lis 1988). Two complexes cooperate to induce pausing at the promoter-proximal region: the negative elongation factor NELF and the DRB (5,6-dichloro-1-β-D-ribofuranosylbenzimidazole) sensitivity-inducing factor (DSIF), composed of Spt5 and Spt4 (Wu et al. 2003; Yamaguchi, Takagi, et al. 1999). In the current view, binding of DSIF to both pol II and the RNA molecule, once it is extruded from the active site, allows for recruitment of NELF and repression of transcription elongation (Missra and Gilmour 2010). In order for pol II to transition into productive elongation it requires the action of the positive transcription elongation factor b (pTEF-b), which phosphorylates DSIF and Pol II (Peterlin and Price 2006; Renner et al. 2001). The multiple phosphorylation events that take place at this transition not only alleviate pausing through release of NELF, but also switches Spt5 into an elongation factor (Yamada et al. 2006; Bourgeois et al. 2002b; Fitz, Neumann, and Pavri 2018).

Control of pol II release from the promoter-proximal pause has been shown to be highly important for regulation of gene expression (Core and Lis 2008; Muse et al. 2007; Nechaev and Adelman 2011; Rahl et al. 2010). Especially, rapid transcriptional upregulation in response to environmental and developmental cues appears to rely on immediate release of pol II into productive elongation (Zeitlinger et al. 2007; Gilmour and Lis 1986; Levine 2011). A clear example of this is the heat shock response (introduced in Chapter I). Heat shock genes maintain a strong peak of paused pol II around promoters in normal repressed conditions (Gilmour and Lis 1986; Rougvie and Lis 1988; Mahat, Kwak, et al. 2016). However, the proteotoxic environment generated by heat stress triggers a rapid and robust induction of these molecular chaperones (Lindquist 1986; Lindquist and Craig 1988; Mahat, Kwak, et al. 2016). This has been shown to be...
mediated by increased recruitment of pTEF-b (Boehm et al. 2003) and alleviation of the promoter-proximal pause (Hieda et al. 2005; Lis et al. 2000). Notably, pol II density increases within the body region after heat shock but remains greater at the promoter on these genes. This has indicated that even under strong transcriptional upregulation pausing remains a rate-limiting step (Boehm et al. 2003; Giardina, Perez, and Lis 1992). Interestingly, heat shock also generates global transcriptional repression achieved by preventing transition of paused polymerases into elongation (Mahat, Kwak, et al. 2016). Thus, control of pol II escape from the pause can greatly modulate gene expression levels. It appears that pausing of pol II provides a state of readiness and/or, as additionally proposed, may favor an accessible promoter architecture to facilitate future upregulation of gene expression (Gilchrist et al. 2008; Gilchrist et al. 2010; Adelman and Lis 2012).

Importantly, an alternative fate of paused polymerases might be premature termination of transcription. This idea has recently gained substantial support from analyses revealing a short residence time of pol II in the paused state (Steurer et al. 2018). While previous studies using ChIP-seq (Henriques et al. 2013; Shao and Zeitlinger 2017), Gro-Seq (Jonkers, Kwak, and Lis 2014), or methyltransferase footprinting assays (Krebs et al. 2017) addressed pol II kinetics and provided measurements in the order of minutes, these studies commonly used the drug triptolide to inhibit transcription initiation. Since triptolide targets the ATPase XPB helicase and prevents melting of the DNA template (Titov et al. 2011), the rationale involved the assumption that triptolide prevented recruitment of new polymerases at the transcription start site. However, my results in chapter III demonstrate that triptolide treatment does not prevent recruitment of the preinitiation complex (Fig. 3-1A, 3-2A) and allows for a peak of pol II at the TSS that could interfere with measurements at low resolution. Furthermore, studies showing that this drug not only affects pol
II levels (Wang et al. 2011; Manzo et al. 2012) but also has a slow mode of action (Nilson et al. 2017), opens the possibility of overestimation of the residence time of pol II in the pause state using triptolide. In fact, fluorescence microscopy and photobleaching studies on GFP-RPB1, the largest subunit of pol II, recently indicated that pol II pauses for less than a minute (~ 42 s) (Steurer et al. 2018). Thus, an alternative view points to a dynamic population of paused polymerases controlled by release from the DNA template and continuous reinitiation.

In accord with this view, at least three potential premature termination mechanisms have been proposed to operate on paused polymerases. All of these agree on generation of a 5’-monophosphorylated transcript that is attacked by the 5’-3’ exonuclease Xrn2. As described in the torpedo model, Xrn2 displaces pol II via degradation of the nascent transcript after cleavage at the polyA site (Connelly and Manley 1988b; West, Gromak, and Proudfoot 2004; Gromak, West, and Proudfoot 2006). The first example takes place on the HIV promoter, a well-known system from which the effects of DSIF, NELF and pTEF-b has been investigated. The endoribonuclease normally involved in miRNA maturation, Microprocessor (Han et al. 2006), was shown to bind the RNA stem loop structure known as the transactivation response (TAR) element and trigger cleavage to allow termination by Xrn2, SETX and Rrp6 (Wagschal et al. 2012). Interestingly, this mechanism appeared to extend to a subset of human endogenous genes, consistent with HITS-CLIP data demonstrating interactions of Microprocessor with many non-miRNA cellular RNAs (Macias et al. 2012). Alternatively, the presence of decapping factors around promoters by ChIP-seq and a reduction in the pausing index of pol II genome-wide upon depletion of Dcp1a, or Xrn2, also postulated decapping as a possible route to premature termination (Brannan et al. 2012). A similar role was proposed for the human decapping factor DXO on transcripts that fail to acquire the correct cap structure (Jiao et al. 2013). And finally, a recent study reported premature
termination from usage of cryptic polyA signals associated with the +1 nucleosome for a large class of mammalian promoters (Chiu et al. 2018). Although in this case, polymerases in front of a nucleosomal barrier, rather than promoter-proximally paused polymerases in the same vicinity, were proposed to be the target of a premature termination checkpoint controlled by the U1 snRNP.

The commonality of Xrn2 in these mechanisms agrees with the multiple proposed roles of Xrn2 in elimination of aberrant transcripts when either capping (Jimeno-Gonzalez et al. 2010), splicing or 3’ end processing is impaired co-transcriptionally (Davidson, Kerr, and West 2012). Interestingly, Rat1, the homolog of Xrn2 in yeast, was reported to induce premature termination of polymerases bearing a rpb1-N488D mutation that causes defects in capping and transcription elongation (Jimeno-Gonzalez et al. 2010). In that report, the role of Xrn2 in degradation of nascent transcripts, co-transcriptionally, was suggested by the observation that this mutation in pol II generated increased recruitment of Xrn2 to gene bodies, as evidenced by ChIP. Additionally, inactivation of Rat1 corrected the observed elongation defect of the mutant pol II and stabilized a population of uncapped nascent transcripts. Further support for a role of Xrn2 in targeting transcripts with processing defects co-transcriptionally came from analysis of beta-globin pre-mRNAs with mutations to inhibit splicing and 3’ end processing generated from constructs in HeLa cells (Davidson, Kerr, and West 2012). Depletion of Xrn2 with siRNAs led to accumulation of defective beta-globin nascent transcripts upstream of the polyA site as evidenced by a nuclear run on (NRO) assay.

In addition, I have found that Xrn2 is recruited early after initiation and accumulates with paused polymerases genome-wide (Chapter III). Therefore, a role of Xrn2 in targeting paused polymerases for premature termination is further supported by the observation that Xrn2 localizes around promoters on human genes. Nevertheless, it has not been demonstrated that premature
termination is a global phenomenon near human promoters. In this chapter, I provide evidence that Xrn2 functions in premature termination of polymerases genome-wide. In particular, Xrn2 targets promoter-proximally paused polymerases that can transition into an apparent suboptimal elongation phase for premature termination. A suboptimal elongation phase of polymerases that are targets of Xrn2 was supported by the observation that inhibition of Xrn2 resulted in stronger accumulation of polymerases within the body region of HSP genes but impaired upregulation of nascent transcript synthesis on these genes in response to heat shock. This discrepancy between increased pol II occupancy and reduced transcript synthesis on a DNA template fits a model in which polymerases that are not displaced by Xrn2 accumulate on HSP genes and do not optimally translocate on the DNA template limiting transit of fully competent complexes positioned upstream. This effect can be viewed as a “traffic jam”. I propose a checkpoint mechanism in which Xrn2 targets for premature termination paused elongation complexes that are not set to transition into productive elongation phase at the promoter-proximal region.

5.2 Results

5.2A Xrn2 Targets Pol II for Premature Termination of Transcription Genome-Wide

To determine whether Xrn2 functions in premature termination of transcription, I inhibited its exonuclease activity and investigated pol II and nascent transcript signal within gene bodies. For this experiment, the mutant (MT) dominant-negative (D235A) Xrn2MT was overexpressed and compared with control isogenic Flp-in TREX cells expressing wild-type (WT) Xrn2WT. These cell lines have been introduced and utilized in Chapter III and IV to demonstrate participation of Xrn2 in termination of pol II transcription downstream of polyA sites, as well as in a previous report by the Bentley lab (Fong et al. 2015b). Preliminary inspection of individual genes did not reveal major changes in pol II occupancy but meta-plots on most active genes
consistently displayed a relative increase in pol II density around promoters after inhibition of Xrn2 (Fig 5-1). Two biological replicates of that observation are shown.

To further evaluate absolute levels of pol II throughout the length of genes, human cultured cells were spiked in with mouse M12 cells right before preparation of cell lysates to control for immunoprecipitation efficiency (Hu et al. 2015). Interestingly, visualization of pol II ChIP-seq density after normalization to M12 mouse spike-in revealed a robust increase in the absolute levels of pol II on the RPL23 and ACTB genes after inhibition of Xrn2 (Fig. 5-2). It can be observed that pol II occupancy not only increases downstream of these genes, consistent with a termination defect (indicated by red arrows), but it is also elevated upstream around their promoter region. Importantly, a role of Xrn2 in termination of transcription at 3’ ends of histone genes remains controversial. While expression of Xrn2MT in these cells has previously revealed a slight increase in pol II density farther away from 3’ ends of histone genes relative to control Xrn2WT cells (Fong et al. 2015b), others have found no evidence of a termination defect at these sites after depletion of Xrn2 (Eaton et al. 2018). Significantly, I found that absolute levels of pol II occupancy are strongly increased within the body region and farther downstream of histone genes in Xrn2MT cells (Fig. 5-3) (indicated by red arrows). I interpret this result to suggest that Xrn2 functions in termination of transcription at 3’ ends of histone genes. This interpretation is further supported by the observation that inhibition of Xrn2 resulted in stabilization of nascent transcripts immediately downstream of 3’ ends of histone genes (Chapter III, Fig 3-12) and accumulation of 5’ PO₄s RNA ends that map precisely to the cleavage site at the 3’ end of histone genes (Chapter III, Fig. 3-17C, 3-21A).

Consistent with meta-plots showing a relative increase of pol II density around promoters (Fig. 5-1), normalization to the M12 spike-in control showed a global increase in pol II occupancy
around promoter regions and gene bodies after inhibition of Xrn2 (Fig. 5-4). A biological replicate of this result is shown in figure 5-5. I interpret this increase in pol II density as a recovery of elongation complexes that would otherwise be prematurely terminated by Xrn2 in control conditions. Further support to this interpretation is provided by results in Chapter III showing that inhibition of Xrn2 also causes an increase in the levels of nascent transcripts throughout the length of genes genome-wide (Fig. 3-11, 3-13). Note that pol II accumulates at 3’ ends of genes in Xrn2MT cells (Fig. 5-4, 5-5), in agreement with results in Chapter IV demonstrating that Xrn2 is required for proper dissociation of pol II from the DNA template downstream of polyA sites (Fig. 4-2 – 4-5).

5.2B Xrn2 Targets Nascent Transcripts for Premature Termination

If polymerases destined to prematurely terminate transcription by Xrn2 are transcriptionally engaged, then their recovery would be accompanied by an increase in nascent transcript synthesis. To test this possibility, I utilized Bromouridine Sequencing (Bruseq) for analysis of nascent transcripts upon inhibition of Xrn2 (Paulsen et al. 2014). In this method, nascent transcripts are pulse-labeled by incubating cells with 5-bromouridine (30 min, 2mM) and isolated via immunoprecipitation. As shown in Chapter III (Fig. 3-11), Bruseq libraries from nascent transcripts inform on transcription directionality and confirmed a termination defect in Xrn2MT cells. Importantly, Bruseq coverage over introns is as strong as coverage over exons (Fig. 5-7), which are only found in mature transcripts, confirming that signal from this method comes exclusively from nascent transcripts.

For assessment of the levels of nascent transcripts I utilized total mitochondrial coverage for normalization. Localization of Xrn2 is restricted to the nuclear compartment (Krzyszton et al. 2012), and no evidence of variability in mitochondria copy number was detected between Xrn2MT
Figure 5-1. Xrn2 targets pol II around promoter regions
HEK293 Flp-in TREX cells were treated with doxycycline for 24 hrs to induce expression of wild-type (WT) Xrn2WT or exonuclease-dead, mutant (MT) (D235A) Xrn2MT. Shown is pol II ChIP-seq density on most active genes (top 25% with highest pol II occupancy within their body region in normal conditions). Note higher pol II density around the TSS relative to distal positions in Xrn2MT cells when compared to control Xrn2WT cells. Mean pol II ChIP-seq density was normalized to signal at -0.5 kb from the transcription start site (TSS).
Figure 5-2. Xrn2 targets pol II within the body region and downstream of the RPL23 and ACTB genes
HEK293 Flp-in TREX cells were treated with doxycycline for 24 hrs to induce expression of wild-type (WT) Xrn2WT or exonuclease-dead, mutant (MT) (D235A) Xrn2MT. pol II ChIP-seq signal was normalized to M12 mouse spike-in (N. M12). Note augmented pol II occupancy throughout the length and downstream of both the RPL23 and the ACTB genes. Signal downstream of the polyA site (marked by the blue dotted line) is strongly increased consistent with a termination defect in Xrn2MT cells (indicated by red arrows). UCSC genome browser screen shots. Blue arrow indicates the direction of transcription.
Figure 5-3. Xrn2 targets pol II on histone genes

HEK293 Flp-in TREX cells were treated with doxycycline for 24 hrs to induce expression of wild-type (WT) Xrn2WT or exonuclease-dead, mutant (MT) (D235A) Xrn2MT. pol II ChIP-seq signal was normalized to M12 mouse spike-in (N. M12). Note augmented pol II occupancy throughout the length and downstream of histone genes (indicated by red arrows). UCSC genome browser screen shots. Blue arrow indicates the direction of transcription.
Figure 5-4. Xrn2 targets pol II throughout the length and downstream of human genes
HEK293 Flp-in TREX cells were treated with doxycycline for 24 hrs to induce expression of wild-type (WT) Xrn2WT or exonuclease-dead, mutant (MT) (D235A) Xrn2MT. Anti-pol II ChIP-seq signal was normalized to M12 mouse spike-in (N. M12) and plotted on most active genes (top 25% with highest pol II occupancy within their body region in normal conditions). Genes were aligned to the transcription start site (TSS) and the cleavage and polyadenylation site (CPS). The body region was defined as +0.5 kb downstream of the TSS and -0.5 kb upstream of the CPS and normalized to gene length. Note higher pol II density around promoters and downstream of the cleavage and polyadenylation site (CPS) in Xrn2MT cells relative to control Xrn2MT cells.
Figure 5-5. Xrn2 targets pol II throughout the length and downstream of human genes – Biological replicate

HEK293 Flp-in TREX cells were treated with doxycycline for 24 hrs to induce expression of wild-type (WT) Xrn2WT or exonuclease-dead, mutant (MT) (D235A) Xrn2MT. Anti-pol II ChIP-seq signal was normalized to M12 mouse spike-in (N. M12) and plotted on most active genes (top 25% with highest pol II occupancy within their body region in normal conditions). Genes were aligned to the transcription start site (TSS) and the cleavage and polyadenylation site (CPS). The body region was defined as +0.5 kb downstream of the TSS and -0.5 kb upstream of the CPS and normalized to gene length. Note higher pol II density around promoters and downstream of the cleavage and polyadenylation site (CPS) in Xrn2MT cells relative to control Xrn2MT cells.
and Xrn2WT cells (Fig. 5-6). Remarkably, inhibition of Xrn2 leads to a consistent increase in nascent transcript synthesis around the promoter region in the sense direction on the CYR61 and GAPDH genes in two independent experiments (Fig. 5-7). This effect can also be seen at the genome wide level (Fig. 5-8). Interestingly, nascent transcripts in the antisense direction are also slightly elevated in Xrn2MT cells relative to control Xrn2WT cells. Therefore, inhibition of Xrn2 not only results in a recovery of a population of polymerases around promoters but also leads to a recovery of nascent transcripts at those sites. It is worth pointing out that Bruseq signal can only be generated by incorporation of 5-bromouridine into nascent transcripts by active transcription. This means that increased Bruseq signal in Xrn2MT cells is a result of increased transcriptional activity and not simply increased stability of the nascent transcript on the DNA template. Taken together, these results strongly suggest that Xrn2 targets polymerases for premature termination of transcription around promoters and that polymerases targeted by Xrn2 are transcriptionally engaged and can elongate nascent transcripts.

5.2C Promoter-Proximally Paused Polymerases are Highly Turned Over by Premature Termination of Transcription

Recovery of polymerases and nascent transcripts from promoter-proximal regions suggested a potential role of Xrn2 in premature termination of paused polymerases. To evaluate this possibility, I inhibited release of paused polymerases by incubating cells with the elongation inhibitor DRB. This leads to accumulation of paused polymerases at the promoter region that would have two potential alternatives: 1. Remain stably bound at the promoter region if they can only transition into elongation, or 2. Dissociate from the DNA template by active premature termination. To distinguish between these two options, transcription initiation must be additionally inhibited so that multiple rounds of transcription initiation do not repopulate the
promoter-proximal pause and turnover of paused polymerases can be assessed. For this, the peak of paused pol II was monitored in the presence of DRB (1 hour) immediately after addition of the initiation inhibitor triptolide (5 and 10 min). Consistent with slow action of triptolide (Nilson et al. 2017), substantial changes in pol II distribution only occurred after 5 minutes of triptolide treatment (Fig. 5-9). Nevertheless, after 10 minutes of triptolide treatment the distribution of pol II transitioned from a dual peak into a single peak centered in close proximity at the TSS (Fig. 5-9). It can be observed that the peaks of pol II associated with promoter-proximal pausing sites upstream and downstream of the TSS under DRB are strongly reduced after addition of triptolide (indicated by red arrows). The loss of these pol II peaks is most easily explained by active turnover of promoter-proximally paused polymerases since a possible transition into elongation is blocked by DRB. A shift of pol II density toward the TSS can be additionally explained by accumulation of pre-initiation complexes at the TSS under triptolide treatment demonstrated in Chapter III (Fig. 3-2A). I interpret this result to suggest that paused polymerases are actively turned over around human promoters by premature termination of transcription.

5.2D Xrn2 Limits Escape of Promoter-Proximally Paused Polymerases into Gene Bodies

Surprisingly, inhibition of Xrn2 did not impair loss of the two main pol II peaks around the TSS in the experiment above (Fig. 5-9), indicating that it is not absolutely required for global turnover of polymerases at the promoter-proximal pausing site. Nonetheless, pol II turnover at the pausing site after addition of triptolide uncovered pol II density downstream of promoter regions that is increased in Xrn2MT cells as evidenced on the GAPDH and XIST genes (Fig 5-10, 5-11), as well as at the genome-wide level (Fig. 5-12) (indicated by red arrows). This result suggests that, alternatively, Xrn2 targets a fraction of paused polymerases that escape global turnover and “leak” into gene bodies. Note that DRB treatment before addition of triptolide was performed for an hour
Figure 5-6. Xrn2WT and Xrn2MT cell lines display similar genomic mitochondrial DNA content
Genomic DNA was extracted from cell lines in control conditions for assessment of mitochondrial content in Xrn2WT and Xrn2MT cells. Relative Real-Time qPCR was performed relative to two GAPDH primer pairs. Signal corresponds to a single experiment with three technical replicates. Note that these cell lines (originated from the same parent cell line) display comparable relative mitochondrial content.
Figure 5-7. Xrn2 targets nascent transcripts at the promoter region of the CYR61 and GAPDH genes
HEK293 Flp-in TREX cells were treated with doxycycline for 24 hrs to induce expression of wild-type (WT) Xrn2WT or exonuclease-dead, mutant (MT) (D235A) Xrn2MT. Bruseq coverage was normalized to thousand mitochondrial mapped reads (N. M.). Bruseq signals from Xrn2WT and Xrn2MT cells were overlaid on the same track. The overlap of Bruseq signals is shown in black. Sense transcription is shown on the positive “y” axis and antisense transcription on the negative “y” axis. Note the increase in nascent transcripts on the sense direction from the promoter of the CYR61 and GAPDH genes in Xrn2MT cells relative to control Xrn2WT cells (stronger red signal above the overlap). Two biological replicates are shown. UCSC genome browser screen shots. Blue arrow indicates the direction of transcription.
Figure 5-8. Xrn2 limits sense and antisense transcription from human promoters
HEK293 Flp-in TREX cells were treated with doxycycline for 24 hrs to induce expression of wild-type (WT) Xrn2WT or exonuclease-dead, mutant (MT) (D235A) Xrn2MT. Bruseq coverage was normalized to thousand mitochondrial mapped reads (N. M.) and plotted on most active genes (top 25% with highest pol II density within their body region in normal conditions). Sense and antisense transcription are shown for both Xrn2WT and Xrn2MT cells. Note consistent increase of nascent transcripts downstream of the transcription start site (TSS) on the sense direction after inhibition of Xrn2. Also, nascent transcripts in the antisense direction from the TSS slightly increase after inhibition of Xrn2. Two biological replicates are shown.
Figure 5-9. Promoter-proximally paused polymerases are highly turned over
Flp-in TREX cells were treated with doxycycline for 24hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead, mutant (D235A) Xrn2MT. Cells were treated with DRB (100 μM, 1 hr) to inhibit the transition of paused polymerases into transcription elongation. Triptolide (10 μM) was added to DRB-treated cells to further inhibit transcription initiation and pol II ChIP-seq was performed at time point T₀ (DRB), T₅ min and T₁₀ min after addition of triptolide (Trip). Shown is pol II ChIP-seq density normalized at the transcription start site (TSS) on RefSeq genes. Note how the dual peak of pol II around the TSS reduces its intensity and transitions to a single peak centered in close proximity to the TSS after addition of triptolide. No substantial differences were observed between Xrn2WT (Top) and Xrn2MT cells (Bottom).
Flp-in TREX cells were treated with doxycycline for 24 hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead, mutant (D235A) Xrn2MT. Cells were treated with DRB (100 µM, 1 hr) to inhibit the transition of paused polymerases into transcription elongation. Triptolide (10 µM) was added to DRB-treated cells to further inhibit transcription initiation and pol II ChIP-seq was performed at time point T₀ (DRB), T₅ min and T₁₀ min after addition of triptolide (Trip). Shown is the overlay of pol II ChIP-seq density from Xrn2WT and Xrn2MT cells on the GAPDH gene. The overlap of ChIP-seq signals is shown in black. Note that inhibition of Xrn2 increases the population of polymerases migrating into the body region of the GAPDH gene from its promoter (indicated by the red arrow). UCSC genome-browser screen shot. Blue arrow indicates the direction of transcription.
Figure 5-11. Xrn2 limits escape of promoter-proximally paused polymerases into the body region of the XIST gene

Flp-in TREX cells were treated with doxycycline for 24 hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead, mutant (D235A) Xrn2MT. Cells were treated with DRB (100 µM, 1 hr) to inhibit the transition of paused polymerases into transcription elongation. Triptolide (10 µM) was added to DRB-treated cells to further inhibit transcription initiation and pol II ChIP-seq was performed at time point T₀ (DRB), T₅ min and T₁₀ min after addition of triptolide (Trip). Shown is the overlay of pol II ChIP-seq density from Xrn2WT and Xrn2MT cells on the XIST gene. The overlap of ChIP-seq signals is shown in black. Note that inhibition of Xrn2 increases the population of polymerases migrating into the body region of the XIST gene from its promoter (indicated by a red arrow). UCSC genome-browser screen shot. Blue arrow indicates the direction of transcription.
Figure 5-12. Xrn2 targets promoter-proximally paused polymerases
Flp-in TREX cells were treated with doxycycline for 24 hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead, mutant (D235A) Xrn2MT. Cells were treated with DRB (100 µM, 1 hr) to inhibit the transition of paused polymerases into transcription elongation. Triptolide (10 µM) was added to DRB-treated cells to further inhibit transcription initiation and pol II ChIP-seq was performed at time point T₀ (DRB), T₅ min and T₁₀ min after addition of triptolide (Trip). Mean pol II ChIP-seq signal was plotted on RefSeq genes and normalized to signal at -0.25kb from the TSS. Note the increase in pol II occupancy downstream of the transcription start site (TSS) in Xrn2MT cells after inhibition of transcription initiation when compared to control Xrn2WT cells (indicated by red arrows).
which would allow plenty of time for the last wave of polymerases on gene bodies to move far away from promoters and even terminate transcription on most genes. Therefore, the density of pol II downstream of promoters in this experiment most likely corresponds to polymerases that escape the promoter-proximal pausing site and translocate into gene bodies.

To further investigate absolute levels of pol II occupancy in this experiment, a biological replicate was performed spiking mouse M12 cells to human cultured cells. Consistent with results above, addition of triptolide to DRB-treated polymerases generated a dramatic drop of the absolute levels of pol II around promoters in both Xrn2WT and Xrn2MT cells (Fig. 5-13 – 5-16). Note that pol II occupancy around promoters drops ~70% after 5 minutes and ~95% after 10 minutes of triptolide treatment on individual genes (Fig. 5-13 – 5-15) and at the genome-wide level (Fig. 5-16). Notably, Xrn2MT cells lost the bulk population of polymerases around promoters, as in Xrn2WT cells, but a substantial fraction of polymerases remained accumulated around and downstream of the TSS when compared to control Xrn2WT cells (Fig. 5-14 – 5-16) (indicated by red arrows). This result further suggests that Xrn2 limits escape of paused polymerases from promoter regions into gene bodies by premature termination of transcription.

5.2E Xrn2 Targets Creeping Polymerases for Premature Termination of Transcription

Given that triptolide treatment still allows for recruitment of pol II to promoters in the experiment above, it is possible that the broad peak of pol II around the TSS (generated by the limited resolution of the ChIP-seq method) might preclude visualization of a wave of polymerases migrating away from promoters after they escape turnover. In order to only visualize this potential fraction of polymerases and test whether they translocate into gene bodies, I decided to perform the same experiment, described above, but utilize a different strategy to additionally prevent recruitment of pol II to the TSS. In yeast and mammalian cells, treatment with NaCl (0.5M) causes
dissociation of pol II from promoters, but not at downstream regions (Proft and Struhl 2004; Wang et al. 2005). Based on the observation that in vitro pre-initiation complexes dissociate from the DNA template at 0.25M KCl but they become resistant to 0.6M KCl after elongation of the first 10 bases (Cai and Luse 1987), it is possible that in mammalian cells high-salt treatment might preferentially destabilize elongation complexes that have incorporated a few nucleotides but not those downstream at the promoter-proximal pausing site; which are proposed to be turned over by active premature termination in this study. Therefore, the strategy involved treatment of cells with DRB and then visualize turnover of polymerases on the DNA template after inhibition of both transcription initiation and recruitment of pol II to the TSS by addition of NaCl (0.35M, total 0.5M). The following results correspond to a single experiment. No biological replicates have been performed.

As predicted, addition of NaCl to DRB-treated polymerases revealed a population of polymerases that escape turnover at the promoter-proximal region and migrate into gene bodies as evidenced on the ACTB and GAPDH genes (Fig. 5-17, 5-18). Pol II densities from Xrn2WT and Xrn2MT cells were overlaid on the same track for each time point after addition of NaCl (T₀, T₂, T₅ and T₁₀ min). Note migration of pol II away from the TSS, illustrated by green arrows pointing to the highest point of the overlap of pol II ChIP-seq densities from both cell lines (Fig. 5-17, 5-18). Interestingly, migration of this wave on the DNA template is very slow. As observed on the ACTB and GAPDH genes, the position of the pol II peak indicated by the green arrows does not move more than ~1 kb between 2 to 10 minutes after addition of NaCl (Fig. 5-17, 5-18). This observation indicates that the elongation rate of this population of polymerases is less than ~0.2 kb per minute, which is dramatically slower than to the global elongation rate of pol II in normal conditions calculated in Chapter IV (~2.0 kb/min, Fig. 4-8). This result is particularly significant
since it is generally accepted that in order for paused polymerases to escape promoter regions, the kinase activity of pTEF-b is required (Peterlin and Price 2006; Renner et al. 2001; Core et al. 2012), and in these experiments it is inhibited by DRB. Therefore, a fraction of polymerases can slowly migrate into gene bodies without licensing by pTEF-b. This phenomenon of polymerases slowly moving into gene bodies has been previously observed in mammalian cells under oxidative stress caused with hydrogen peroxide (Nilson et al. 2017). In that report, polymerases slowly migrating into gene bodies under oxidative stress were referred to as “creeping” polymerases and were proposed to originate under conditions of reduced premature termination of transcription caused by hydrogen peroxide. Although slow-elongating polymerases under DRB conditions in this study are not necessarily the same as those detected under oxidative stress, I will refer to this population of polymerases as “creeping” polymerases.

Remarkably, inhibition of Xrn2 resulted in stronger accumulation of creeping polymerases within gene bodies as shown on individual genes (Fig. 5-17, 5-18) and at the genome wide-level (Fig. 5-20) (indicated by red arrows). I interpret this result to suggest that Xrn2 limits migration of creeping polymerases into gene bodies by premature termination of transcription. Importantly, strong accumulation of polymerases also took place downstream of histone genes at 10 minutes after addition of NaCl in Xrn2MT cells (Fig. 5-19) (indicated by red arrows). This observation adds further support for a role of Xrn2 in termination of transcription on histone genes.

5.2F Creeping Polymerases are not Associated with Detectable Signal of NELF

According to the current model, phosphorylation of pol II, DSIF and the negative elongation factor NELF by pTEF-b causes release of NELF from the elongation complex and escape of pol II from the promoter-proximal pausing site (Peterlin and Price 2006; Renner et al. 2001; Yamaguchi, Takagi, et al. 1999). Therefore, it was intriguing to determine if creeping
Figure 5-13. Xrn2 limits escape of promoter-proximally paused polymerases into the body region of the ACTB gene (N. M12)

Flp-in TREX cells were treated with doxycycline for 24 hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead, mutant (D235A) Xrn2MT. Cells were treated with DRB (100 µM, 1 hr) to inhibit the transition of paused polymerases into transcription elongation. Triptolide (10 µM) was added to DRB-treated cells to further inhibit transcription initiation and pol II ChIP-seq was performed at time point T₀ (DRB), T₅ min and T₁₀ min after addition of triptolide (Trip). Shown is the overlay of pol II ChIP-seq density from Xrn2WT and Xrn2MT cells on the ACTB gene after normalization to M12 mouse spike-in (N. M12). The overlap of ChIP-seq signals is shown in black. Note that inhibition of Xrn2 increases the population of polymerases migrating into the body region of the ACTB gene from its promoter (indicated by red arrows). UCSC genome-browser screen shot. Blue arrow indicates the direction of transcription.
Figure 5-14. Xrn2 limits escape of promoter-proximally paused polymerases into the body region of the GAPDH gene (N. M12)

Flp-in TREX cells were treated with doxycycline for 24 hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead, mutant (D235A) Xrn2MT. Cells were treated with DRB (100 μM, 1 hr) to inhibit the transition of paused polymerases into transcription elongation. Triptolide (10 μM) was added to DRB-treated cells to further inhibit transcription initiation and pol II ChIP-seq was performed at time point T0 (DRB), T5 min and T10 min after addition of triptolide (Trip). Shown is the overlay of pol II ChIP-seq density from Xrn2WT and Xrn2MT cells on the GAPDH gene after normalization to M12 mouse spike-in (N. M12). The overlap of ChIP-seq signals is shown in black. Note that inhibition of Xrn2 increases the population of polymerases migrating into the body region of the GAPDH gene from its promoter (indicated by red arrows). UCSC genome-browser screen shot. Blue arrow indicates the direction of transcription.
Figure 5-15. Xrn2 limits escape of promoter-proximally paused polymerases into the histone gene HIST4H4 (N. M12)
Flp-in TREX cells were treated with doxycycline for 24 hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead, mutant (D235A) Xrn2MT. Cells were treated with DRB (100 µM, 1 hr) to inhibit the transition of paused polymerases into transcription elongation. Triptolide (10 µM) was added to DRB-treated cells to further inhibit transcription initiation and pol II ChIP-seq was performed at time point T₀ (DRB), T₅ min and T₁₀ min after addition of triptolide (Trip). Shown is the overlay of pol II ChIP-seq density from Xrn2WT and Xrn2MT cells on the HIST4H4 gene after normalization to M12 mouse spike-in (N. M12). The overlap of ChIP-seq signals is shown in black. Note that inhibition of Xrn2 increases the population of polymerases migrating from the HIST4H4 gene (indicated by red arrows). UCSC genome-browser screen shot. Blue arrow indicates the direction of transcription.
Figure 5-16. Xrn2 targets promoter-proximally paused polymerases (N. M12)
Flp-in TREX cells were treated with doxycycline for 24 hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead, mutant (D235A) Xrn2MT. Cells were treated with DRB (100 µM, 1 hr) to inhibit the transition of paused polymerases into transcription elongation. Triptolide (10 µM) was added to DRB-treated cells to further inhibit transcription initiation and pol II ChIP-seq was performed at time point T_0 (DRB), T_5 min and T_10 min after addition of triptolide. Mean pol II ChIP-seq signal was normalized to mouse M12 spike-in (N. M12) and plotted on RefSeq genes. Note the values in the “y” axis showing strong loss of pol II density in both cells after addition of triptolide. Also note that Xrn2MT cells exhibit stronger pol II density around the transcription start site (TSS) and downstream of the TSS after inhibition of transcription initiation when compared to control Xrn2WT cells (indicated by red arrows).
Figure 5-17. Xrn2 targets polymerases creeping into the body region of the ACTB gene for premature termination of transcription

Flp-in TREX cells were treated with doxycycline for 24 hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead, mutant (D235A) Xrn2MT. Cells were treated with DRB (100 µM, 1 hr) to inhibit the transition of paused polymerases into transcription elongation. NaCl (500 mM total) was added to DRB-treated cells to further inhibit transcription initiation and pol II ChIP-seq was performed at time point T₀ (DRB), T₂ min, T₅ min and T₁₀ min after addition NaCl. Shown is the overlay of pol II ChIP-seq density from Xrn2WT and Xrn2MT cells on the ACTB gene. The overlap of ChIP-seq signals is shown in black. Note migration of creeping polymerases into the gene body indicated by the position of green arrows. Green arrows indicate the position of the point of maximum intensity within the overlap distribution. Also, note that inhibition of Xrn2 increases the population of creeping polymerases migrating from the promoter region (indicated by the red arrow). UCSC genome browser screen shot. Blue arrow indicates the direction of transcription.
Figure 5-18. Xrn2 targets polymerases creeping into the body region of the GAPDH gene for premature termination of transcription

Flp-in TREX cells were treated with doxycycline for 24 hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead, mutant (D235A) Xrn2MT. Cells were treated with DRB (100 µM, 1 hr) to inhibit the transition of paused polymerases into transcription elongation. NaCl (500 mM total) was added to DRB-treated cells to further inhibit transcription initiation and pol II ChIP-seq was performed at time point T₀ (DRB), T₂ min, T₅ min and T₁₀ min after addition NaCl. Shown is the overlay of pol II ChIP-seq density from Xrn2WT and Xrn2MT cells on the GAPDH gene. The overlap of ChIP-seq signals is shown in black. Note migration of creeping polymerases into the gene body indicated by the position of green arrows. Green arrows indicate the position of the point of maximum intensity within the overlap distribution. Also, note that inhibition of Xrn2 increases the population of creeping polymerases migrating from the promoter region (indicated by red arrows). UCSC genome browser screen shot. Blue arrow indicates the direction of transcription.
Figure 5-19. Xrn2 targets polymerases creeping from promoter regions of histone genes for premature termination

Flp-in TREX cells were treated with doxycycline for 24 hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead, mutant (D235A) Xrn2MT. Cells were treated with DRB (100 μM, 1 hr) to inhibit the transition of paused polymerases into transcription elongation. NaCl (500 mM total) was added to DRB-treated cells to further inhibit transcription initiation and pol II ChIP-seq was performed at time point T₀ (DRB), T₂ min, T₅ min and T₁₀ min after addition NaCl. Shown is the overlay of pol II ChIP-seq density from Xrn2WT and Xrn2MT cells on histone genes. The overlap of ChIP-seq signals is shown in black. Note that inhibition of Xrn2 increases the population of creeping polymerases migrating from the promoter region of histone genes (indicated by red arrows). UCSC genome browser screen shot. Blue arrow indicates the direction of transcription.
Figure 5-20. Xrn2 targets polymerases creeping from promoter regions into gene bodies genome-wide
Flp-in TREX cells were treated with doxycycline for 24 hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead, mutant (D235A) Xrn2MT. Cells were treated with DRB (100 µM, 1 hr) to inhibit the transition of paused polymerases into transcription elongation. NaCl (500 mM total) was added to DRB-treated cells to further inhibit transcription initiation and pol II ChIP-seq was performed at time point T₀ (DRB), T₂ min, T₅ min and T₁₀ min after addition NaCl. pol II ChIP-seq signal was plotted on RefSeq genes aligned at the transcription start site (TSS). Note the increased density of creeping polymerases migrating into gene bodies after inhibition of Xrn2 (indicated by red arrows).
polymerases detected under DRB conditions, in which pTEF-b is inhibited, were still bound by NELF within gene bodies. To investigate this possibility, I visualized the ChIP-seq distribution profile of NELFa, a subunit of NELF, after treatment of Xrn2MT cells with DRB and high-salt as performed above. In this experiment mouse M12 cells were spiked-in to human cells to assess absolute levels of NELFa. If creeping polymerases are bound by NELF, addition of NaCl would be expected to generate a wave of NELFa similar to that of creeping polymerases. Results from a single experiment are shown below. No biological replicates have been performed.

Interestingly, no wave of NELFa was detected. NELFa occupancy under DRB conditions was strong and limited to the promoter of the RPL23 and GAPDH genes (Fig. 5-21, 5-22), consistent with an association between NELF and promoter-proximally paused elongation complexes. Nevertheless, addition of NaCl to DRB-treated cells caused loss of NELFa from these promoters without noticeable migration of NELFa density into the body region of the RPL23 and GAPDH genes (Fig. 5-21, 5-22). Analysis of NELFa at a genome-wide scale showed that the relative distribution of this factor on human genes does not shift toward gene bodies while its levels drop after addition of high-salt (Fig. 5-23). In conclusion, the lack of detectable signal of NELFa migrating into gene bodies might indicate that creeping polymerases are not bound by NELF. If that is in fact the case, it may explain their capability to escape the promoter-proximal region. An important intriguing question from this result is whether creeping polymerases correspond to paused polymerases that are released into gene bodies upon loss of NELF or if, alternatively, creeping polymerases do not undergo promoter-proximal pausing and traverse gene bodies from the TSS without repression by NELF. The implication of the latter possibility would be that promoter-proximal pausing of pol II is not always an obligatory step during transcription on paused genes.
Figure 5-21. NELFa density does not migrate with creeping polymerases into the body region of the RPL23 gene

Flp-in TREX cells were treated with doxycycline for 24 hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead, mutant (D235A) Xrn2MT. Cells were treated with DRB (100 µM, 1 hr) to inhibit the transition of paused polymerases into transcription elongation. NaCl (500 mM total) was added to DRB-treated cells to further inhibit transcription initiation and pol II ChIP-seq was performed at time point T₀ (DRB), T₂ min, T₅ min and T₁₀ min after addition NaCl. Shown is NELFa ChIP-seq density normalized to mouse M12 spike-in (N. M12) on the RPL23 gene. UCSC genome browser screen shot. Blue arrow indicates the direction of transcription.
Figure 5-22. NELFa density does not migrate with creeping polymerases into the body region of the GAPDH gene

Flp-in TREX cells were treated with doxycycline for 24 hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead, mutant (D235A) Xrn2MT. Cells were treated with DRB (100 µM, 1 hr) to inhibit the transition of paused polymerases into transcription elongation. NaCl (500 mM total) was added to DRB-treated cells to further inhibit transcription initiation and pol II ChIP-seq was performed at time point T0 (DRB), T2 min, T5 min and T10 min after addition NaCl. Shown is NELFa ChIP-seq density normalized to mouse M12 spike-in (N. M12) on the GAPDH gene. UCSC genome browser screen shot. Blue arrow indicates the direction of transcription.
Figure 5-23. Global NELFa density does not migrate with creeping polymerases into gene bodies

Flp-in TREX cells were treated with doxycycline for 24 hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead, mutant (D235A) Xrn2MT. Cells were treated with DRB (100 µM, 1 hr) to inhibit the transition of paused polymerases into transcription elongation. NaCl (500 mM total) was added to DRB-treated cells to further inhibit transcription initiation and pol II ChIP-seq was performed at time point $T_0$ (DRB), $T_2$ min, $T_5$ min and $T_{10}$ min after addition NaCl. Shown is NELFa ChIP-seq density on RefSeq genes aligned at the transcription start site (TSS).
Interestingly, I additionally found evidence that polymerases that are targets of Xrn2 for premature termination do not efficiently elongate within gene bodies. In order to investigate whether premature termination of transcription around human promoters is important for proper gene expression, I decided to evaluate the transcriptional response of heat shock genes after heat shock (introduced in Chapter I) in Xrn2MT cells and in control Xrn2WT cells. Results from a single experiment are shown below. No biological replicates have been performed.

First, visualization of pol II density confirmed a termination defect at 3’ ends of heat shock genes suggesting that Xrn2 also functions in termination of transcription at these sites (Fig. 5-24). As expected, incubation of cells at 42°C for 30 minutes greatly increased pol II occupancy within the body region of this group of genes (Fig. 5-25 – 5-27). Remarkably, inhibition of Xrn2 resulted in stronger accumulation of pol II within the body region of HSP genes in response to heat shock, which can be easily appreciated by comparison of the pol II body density relative to the intensity of the pol II peak at the promoter-proximal region on the HSP40, HSP90 and HSPH1 genes (Fig. 5-25 – 5-27). The intensity of the promoter-proximal peak is illustrated by a red line. Note that inhibition of Xrn2 leads to stronger pol II peaks within the body region of HSP genes that more robustly exceed or reach the intensity of the promoter-proximal peak of pol II.

To observe global effects on pol II distribution, I searched for genes that become transcriptionally upregulated as well as those that become downregulated, upon heat shock. To do this, and additionally analyze nascent transcript signal during heat shock, I pulse-labeled nascent transcripts in Xrn2WT and Xrn2MT cells for 30 minutes in normal conditions, or simultaneously with incubation at 42°C, and performed Bruseq. Bruseq signal normalized to mitochondrial reads clearly detected upregulation of nascent transcript synthesis as shown on the HSP40, HSP90 and
HSPH1 genes (Fig. 5-31 – 5-33), as well as downregulation on the CYR61 and EIF1 genes (Fig. 5-34 – 5-35). Genes were classified as heat shock inducible genes on the basis that they present at least a two-fold increase in Bruseq coverage within their body region upon heat shock and also rank in the top 10% of genes with highest Bruseq coverage under heat shock conditions (n=155) (Appendix D). Repressed genes were selected on the basis that they rank within the top 50% of genes with highest Bruseq coverage within their body region in normal conditions but reduce it at least two-fold after heat shock (n=1319) (Appendix D). Meta-plots on heat shock inducible genes showed that pol II density increases in the body region of this group of genes in both Xrn2WT and Xrn2MT cells (Fig. 5-28). On the other hand, all genes (n=19548) appear to have a slight reduction in pol II occupancy around promoters and 3’ ends (Fig. 5-28), consistent with global repression of transcription during heat shock (Mahat, Kwak, et al. 2016). Also, the group of heat shock repressed genes lost their 3’ peak of pol II density after heat shock (Fig. 5-28) (indicated by red arrows) in agreement with Bruseq data indicating strong repression of nascent transcript synthesis on these genes (Fig. 5-37).

Importantly, stronger accumulation of polymerases within the body region of heat shock inducible genes in Xrn2MT cells was a global effect (Fig. 5-29). Plotting of the fold increase in pol II density upon heat shock shows that Xrn2MT cells present a much greater accumulation of polymerases immediately downstream of the TSS (indicated by a blue arrow). Strikingly, stronger accumulation of polymerases within the body region of HSP genes did not correlate with increased output of nascent transcript synthesis relative to control Xrn2WT cells. In fact, the opposite effect was observed. First, nascent transcripts were labeled with 5-bromouridine in normal conditions or at 42°C (to simultaneously induce the heat shock response) and were analyzed by reverse transcriptase (RT) qPCR before performing Bruseq libraries. Analysis of nascent transcripts from
Figure 5-24. The HSP90AA1 gene displays a termination defect after inhibition of Xrn2
Flp-in TREX cells were treated with doxycycline for 24 hrs for expression of wild-type Xrn2
(Xrn2WT) or exonuclease-dead, mutant (D235A) Xrn2MT. pol II ChIP-seq signal is shown on the
HSP90AA1 (HSP90) gene. Note increased pol II occupancy immediately downstream of the
polyA site (blue dotted line) when compared to control Xrn2WT cells. UCSC genome browser
screen shot. Blue arrow indicates direction of transcription.
Figure 5-25. Inhibition of Xrn2 generates stronger accumulation of pol II within the body region of the HSP90AA1 gene after heat shock

Flp-in TREX cells were treated with doxycycline for 24 hrs for expression of wild-type Xrn2 (Xrn2WT) or dominant-negative (D235A) mutant (Xrn2MT). Heat shock was performed by sinking cultured cells into a 42°C water bath for 30min. pol II ChIP-seq signal is shown on the HSP90AA1 (HSP90) gene. A red line illustrating the intensity of the promoter-proximal peak has been added. Note increased pol II density within the gene body relative to the intensity of the promoter-proximal peak upon heat shock after inhibition of Xrn2 (indicated by the red arrow). UCSC genome browser screen shot. Blue arrow indicates the direction of transcription.
Figure 5-26. Inhibition of Xrn2 generates stronger accumulation of pol II within the body region of the DNAJB1 gene after heat shock
Flp-in TREX cells were treated with doxycycline for 24 hrs for expression of wild-type Xrn2 (Xrn2WT) or dominant-negative (D235A) mutant (Xrn2MT). Heat shock was performed by sinking cultured cells into a 42°C water bath for 30min. pol II ChIP-seq signal is shown on the DNAJB1 (HSP40) gene. A red line illustrating the intensity of the promoter-proximal peak has been added. Note increased pol II density within the gene body relative to the intensity of the promoter-proximal peak upon heat shock after inhibition of Xrn2 (indicated by red arrows). UCSC genome browser screen shot. Blue arrow indicates the direction of transcription.
Figure 5-27. Inhibition of Xrn2 generates stronger accumulation of pol II within the body region of the HSPH1 gene after heat shock
Flp-in TREX cells were treated with doxycycline for 24 hrs for expression of wild-type Xrn2 (Xrn2WT) or dominant-negative (D235A) mutant (Xrn2MT). Heat shock was performed by sinking cultured cells into a 42°C water bath for 30min. pol II ChIP-seq signal is shown on the HSPH1 gene. A red line illustrating the intensity of the promoter-proximal peak has been added. Note increased pol II density within the gene body relative to the intensity of the promoter-proximal peak upon heat shock after inhibition of Xrn2 (illustrated by a double-headed arrow). UCSC genome browser screen shot. Blue arrow indicates the direction of transcription.
Figure 5-28. Promoter-proximal pausing of pol II is maintained under heat shock conditions in Xrn2WT and Xrn2MT cells

HEK293 Flp-in TREX cells were treated with doxycycline for 24hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead (D235A) Xrn2MT. Heat shock was performed by sinking cultured cells into a 42ºC water bath for 30min. Mean pol II ChIP-seq signal normalized to signal at -1.5 kb from the TSS. Three different groups of genes are shown. Top: all RefSeq genes; Middle: induced genes based on strongest upregulation of nascent transcript signal within their gene body, Bottom: repressed genes based on strongest downregulation of nascent transcript signal within their gene body. Note that pol II density increases within the body region of heat shock (HS) induced genes, but the peak of pol II around the transcription start site (TSS) is maintained under heat shock. Also, note loss of the 3’ peak of pol II downstream of the cleavage and polyadenylation site (CPS) on repressed genes (indicated by red arrows). Mean pol II ChIP-seq density was normalized to signal at -1.5 kb from the TSS.
Figure 5-29. Inhibition of Xrn2 causes stronger accumulation of pol II within the body region of heat shock genes upon heat shock
Flp-in TREX cells were treated with doxycycline for 24 hrs for expression of wild-type Xrn2 (Xrn2WT) or dominant-negative (D235A) mutant (Xrn2MT). Heat shock was performed by sinking cultured cells into a 42°C water bath for 30min. pol II ChIP-seq signal at 42°C was normalized to signal at 37°C in Xrn2WT and Xrn2MT cells by calculating the Log₂ value of the ratio (pol II ChIP-seq 42°C/pol II ChIP-seq 37°C). The fold increase in pol II density is shown on heat shock induced genes. Genes were aligned at the transcription start site (TSS) and the cleavage and polyadenylation site (CPS). The body region was defined as +0.5 kb downstream of the TSS and -0.5 kb upstream of the CPS and normalized to gene length. Note stronger accumulation of pol II immediately downstream of the TSS upon heat shock after inhibition of Xrn2 (indicated by the blue arrow).
Figure 5-30. Xrn2 is required for upregulation of nascent transcript synthesis on heat shock genes

Flp-in TREX cells were treated with doxycycline for 24hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead, mutant (D235A) Xrn2MT. Cells were then incubated at 42°C in the presence of 5-bromouridine (2mM) to label nascent transcripts during heat shock. Nascent transcripts were isolated via immunoprecipitation and analyzed by Reverse Transcription qPCR. Absolute quantification with primers targeting the 5’-end of the HSP40 and HSP70 cDNA from a single experiment is shown. Error bars correspond to the standard error of technical replicates (n=3). Note impaired upregulation of the HSP40 and HSP70 nascent transcripts in Xrn2MT cells (green bars) relative to control Xrn2WT cells (blue bars) after heat shock.
Figure 5-31. Heat shock induces upregulation of nascent transcript synthesis on the DNAJB1 gene

Flp-in TREX cells were treated with doxycycline for 24hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead, mutant (D235A) Xrn2MT. Cells were then incubated at 42°C in the presence of 5-bromouridine (2mM) to label nascent transcripts during heat shock. Shown is the overlay of Bruseq signal in normal conditions and after heat shock for Xrn2WT and Xrn2MT cells normalized to mitochondrial reads (N. M.) on the DNAJB1 gene. Sense transcription is shown on the positive “y” axis and antisense transcription in the negative “y” axis. The overlap of Bruseq signal in non-induced conditions with Bruseq signal under heat shock conditions is shown as the darker distribution profile.
Figure 5-32. Heat shock induces upregulation of nascent transcript synthesis on the HSP90AA1 gene

Flp-in TREX cells were treated with doxycycline for 24hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead, mutant (D235A) Xrn2MT. Cells were then incubated at 42°C in the presence of 5-bromouridine (2mM) to label nascent transcripts during heat shock. Shown is the overlay of Bruseq signal in normal conditions and after heat shock for Xrn2WT and Xrn2MT cells normalized to mitochondrial reads (N. M.) on the HSP90AA1 gene. Sense transcription is shown on the positive “y” axis and antisense transcription in the negative “y” axis. The overlap of Bruseq signal in non-induced conditions with Bruseq signal under heat shock conditions is shown as the darker distribution profile.
Figure 5-33. Heat shock induces upregulation of nascent transcript synthesis on the HSPH1 gene
Flp-in TREX cells were treated with doxycycline for 24hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead, mutant (D235A) Xrn2MT. Cells were then incubated at 42°C in the presence of 5-bromouridine (2mM) to label nascent transcripts during heat shock. Shown is the overlay of Bruseq signal in normal conditions and after heat shock for Xrn2WT and Xrn2MT cells normalized to mitochondrial reads (N. M.) on the HSPH1 gene. Sense transcription is shown on the positive “y” axis and antisense transcription in the negative “y” axis. The overlap of Bruseq signal in non-induced conditions with Bruseq signal under heat shock conditions is shown as the darker distribution profile.
Figure 5-34. Heat shock causes downregulation of nascent transcript synthesis on the CYR61 gene

Flp-in TREX cells were treated with doxycycline for 24hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead, mutant (D235A) Xrn2MT. Cells were then incubated at 42°C in the presence of 5-bromouridine (2mM) to label nascent transcripts during heat shock. Shown is the overlay of Bruseq signal in normal conditions and after heat shock for Xrn2WT and Xrn2MT cells normalized to mitochondrial reads (N. M.) on the CYR61 gene. Sense transcription is shown on the positive “y” axis and antisense transcription in the negative “y” axis.
Figure 5-35. Heat shock causes downregulation of nascent transcript synthesis on the EIF1 gene
Flp-in TREX cells were treated with doxycycline for 24hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead, mutant (D235A) Xrn2MT. Cells were then incubated at 42°C in the presence of 5-bromouridine (2mM) to label nascent transcripts during heat shock. Shown is the overlay of Bruseq signal in normal conditions and after heat shock for Xrn2WT and Xrn2MT cells normalized to mitochondrial reads (N. M.) on the EIF1 gene. Sense transcription is shown on the positive “y” axis and antisense transcription in the negative “y” axis.
Figure 5-36. Xrn2 is required for proper upregulation of nascent transcript synthesis on heat shock genes upon heat shock
Flp-in TREX cells were treated with doxycycline for 24 hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead mutant (D235A) (Xrn2MT). Cells were then incubated at 42°C in the presence of 5-bromouridine (2mM) to label nascent transcripts during heat shock. Bruseq signal normalized to mitochondrial reads (N.M.) is shown for genes most highly upregulated during heat shock in control Xrn2WT cells. These genes were selected on the basis that they present at least two-fold increase in Bruseq coverage within their body region upon heat shock and also rank in the top 10% of genes with highest Bruseq coverage under heat shock conditions. Note that in normal conditions heat shock genes display higher levels of nascent transcripts in Xrn2MT cells. However, upon heat shock, upregulation of nascent transcript synthesis is impaired in Xrn2MT cells relative to control Xrn2WT cells.
Figure 5-37. Inhibition of Xrn2 does not prevent global repression of nascent transcript synthesis
Flp-in TREX cells were treated with doxycycline for 24 hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead mutant (D235A) (Xrn2MT). Cells were then incubated at 42°C in the presence of 5-bromouridine (2mM) to label nascent transcripts during heat shock. Bruseq signal normalized to mitochondrial reads (N.M.) is shown for genes repressed during heat shock in control Xrn2WT cells. These genes were selected on the basis that they rank within the top 50% of genes with highest Bruseq coverage within their body region in control conditions but reduce it at least two-fold after heat shock. Note that in normal conditions this group of genes display higher levels of nascent transcripts in Xrn2MT cells. However, upon heat shock, downregulation of nascent transcript synthesis occurs to the same extent on these genes in both cell lines.
the HSP40 and HSP70 genes revealed that Xrn2MT cells fail to efficiently upregulate nascent transcript synthesis on these genes in response to heat shock when compared to control Xrn2WT cells (Fig. 5-30). Impaired upregulation of nascent transcript synthesis in Xrn2MT cells was a global effect on heat shock inducible genes (Fig. 5-36). In spite of the fact that inhibition of Xrn2 results in increased transcription of this group of genes in normal conditions, Bruseq signal in Xrn2MT cells was not upregulated to the same extent for these genes as in Xrn2WT cells in response to heat shock (Fig. 5-36) (red arrow indicates amplification of Bruseq signals). This discrepancy between increased pol II occupancy and reduced nascent transcript synthesis strongly suggests that recovered polymerases do not efficiently elongate and that rapid upregulation of the heat shock response is therefore limited. I interpret this result to suggest that polymerases that are not removed by Xrn2 accumulate on the body region of HSP genes and do not optimally translocate on the DNA template limiting transit of fully competent complexes positioned upstream; causing a “traffic jam”. Interestingly, the exonuclease activity of Xrn2 was not required for a drop in nascent transcript signal in the group of heat shock repressed genes (Fig. 5-37), in spite of the fact that in normal conditions these genes present higher levels of nascent transcripts upon inhibition of Xrn2. Therefore, this observation suggests that downregulation of nascent transcript signal on these genes in response to heat shock does not occur via increased Xrn2 exonuclease-dependent premature termination of transcription.

5.3 Discussion

In Chapters III and IV, I expanded the current knowledge on the molecular events leading to termination of RNA pol II at 3’ ends of genes and showed that Xrn2 participates in removal of polymerases from the DNA template during termination of pol II. Here, I report a previously unknown function of Xrn2 on human promoters and propose an Xrn2-mediated premature
termination checkpoint that targets paused polymerases not properly equipped for productive elongation.

Premature termination of transcription has been previously demonstrated in bacterial systems (Nudler and Gottesman 2002) in yeast (Jiao et al. 2010; Gudipati et al. 2008; Lykke-Andersen and Jensen 2007; Steinmetz et al. 2001) and on viral promoters in eukaryotes (Wagschal et al. 2012; Hay, Skolnik-David, and Aloni 1982). Nevertheless, to date premature termination has not been established as a global phenomenon on human genes. The Bentley lab proposed a mechanism in which regulated decapping provides entrance of Xrn2 early during transcription. The idea came from the observation that depletion of Xrn2, or decapping factors, led to a global increase in pol II density within gene bodies and a reduction in the promoter-proximal peak of pol II suggestive of an increased flux of pol II into gene bodies (Brannan et al. 2012). Knockdown of Xrn2 and its potential effects on the heat shock response were additionally investigated in this study, shown in Appendix C. Interestingly, inhibition of the exonuclease activity in this investigation and in a previous report (Fong et al. 2015b), did not recapitulate the shift in the distribution profile of pol II characterized by increased pol II density within gene bodies relative to the promoter-proximal region after depletion of Xrn2 (Brannan et al. 2012). Although not completely clear why Xrn2MT cells do not display the change in the distribution profile of pol II caused by depletion of Xrn2, but rather an increase in absolute levels of pol II throughout the length of genes (Fig. 5-2 – 5-5), both results support the model in which Xrn2 functions in premature termination of transcription. A scenario that would reconcile this discrepancy would be a potential contribution of Xrn2 in regulation of the promoter-proximal pause. If a structural component of Xrn2 promotes pausing of pol II around promoters, depletion of Xrn2 might cause a reduction in pausing and therefore a drop in the promoter-proximal peak of pol II in addition to
a reduction in Xrn2 exonuclease-dependent premature termination. In the case of inhibition of the exonuclease activity of Xrn2, the structural component of Xrn2 is still present maintaining the promoter-proximal peak of pol II, although levels of pol II occupancy increase throughout genes due to reduced Xrn2 exonuclease-dependent premature termination of transcription. Further investigation is needed to determine if this is in fact the case.

Importantly, normalization of ChIP-seq libraries allowed me to uncover increased levels of pol II throughout the length of gene bodies and promoter regions after inhibition of Xrn2 (Fig. 5-2 – 5-5) which is not easily observed by relative frequency distributions of pol II (Fig. 4-1) (Fong et al. 2015b). This result positively correlated with increased levels of nascent transcripts from the 5’ ends to the 3’ ends of genes as evidenced by Bruseq (Fig. 3-13, 5-8), strongly suggesting that Xrn2 functions in premature termination of transcription. Importantly, the results obtained in this investigation suggest that the fraction of polymerases that are targets of Xrn2 for premature termination are transcriptionally engaged and elongate nascent transcripts based on the fact that inhibition of Xrn2 also led to an increase in the population of bromo-labeled nascent transcripts from promoter regions, which requires active transcription (Fig. 5-8). This does not rule out the possibility that Xrn2 also targets polymerases that are stalled or arrested on the DNA template. Interestingly, Xrn2 also limits divergent transcription from promoters (Fig 5-8), although this was not a strong effect, indicating that it might also contribute to establish promoter directionality. A role of Xrn2 around promoters is further supported by recruitment of Xrn2 to promoter regions (Chapter III).

The above observations raised the question of how stable paused polymerases are around promoters. Interestingly, I found that maintenance of a strong peak of paused pol II under conditions in which elongation is blocked requires continuous initiation events (Fig. 5-9).
Therefore, paused polymerases rapidly turnover at the pausing site, in spite of the fact that they cannot transition into the elongation phase. This result suggests that indeed premature termination is a global phenomenon on human promoters, consistent with a recent study claiming short-lived pausing events (Steurer et al. 2018). Surprisingly, inhibition of Xrn2 did not prevent global turnover (Fig. 5-9). Instead, Xrn2 was required to limit slow and progressive migration of paused polymerases into gene bodies (Fig. 5-17 – 5-20). Importantly, it is widely thought that pol II requires phosphorylation by pTEF-b to be released from the promoter-proximal pause and transition into productive elongation (Peterlin and Price 2006; Renner et al. 2001). Nevertheless, in this study I found that a fraction of polymerases, referred to as creeping polymerases, are able to slowly migrate into gene bodies from promoter regions in the presence of DRB, an inhibitor of the kinase activity of pTEF-b (Fig. 5-17 – 5-20) (Mancebo et al. 1997). Interestingly, Xrn2 limits transcription by this population of polymerases (Fig. 5-10 – 5-20). Therefore, a fraction of polymerases can “leak” into gene bodies without licensing by pTEF-b where they become targets of Xrn2 for premature termination. Such a role of Xrn2 might ensure that complexes that fail to undergo proper maturation at the pausing site do not transition into the elongation phase of pol II. For instance, failure to recruit RNA processing factors or acquire a proper cap structure, which coincides with pausing (Rasmussen and Lis 1993), might lead to premature termination. In line with this idea Rat1, the homolog of Xrn2 in yeast, prevents synthesis of full length uncapped transcripts in mutants defective for mRNA capping (Jiao et al. 2010; Jimeno-Gonzalez et al. 2010).

Additionally, I found evidence that polymerases that are not displaced by Xrn2 from the DNA template do not efficiently elongate within gene bodies. This conclusion was supported by the observation that inhibition of Xrn2 led to stronger accumulation of polymerases within the body region of HSP genes upon their activation by heat shock (Fig. 5-25 – 5-27, 5-29), but it was
accompanied with impaired upregulation of HSP pre-mRNAs (Fig. 5-30, 5-36). This discrepancy fits a model in which polymerases that are not displaced by Xrn2 accumulate on HSP genes and do not optimally translocate on the DNA template limiting transit of fully competent complexes positioned upstream; causing a “traffic jam”. It is possible that in normal conditions, low frequency of initiation rates would not lead to potential collisions by fast approaching polymerases with polymerases that do not optimally elongate farther downstream. In contrast, high initiation rates under heat shock, which can support upregulation of transcript synthesis up to 1000-fold (Kobayashi et al. 2014), a suboptimal elongation rate by a fraction of polymerases may limit transit of all polymerases upstream. The fact that the promoter-proximal peak of pol II is not eliminated during heat shock (Fig. 5-25 – 5-28), further points to the importance of promoter-proximal pausing of pol II as a potential checkpoint that ensures maturation of elongation complexes for proper transition into productive elongation; even when rapid upregulation of gene expression is needed. It is possible that failure of elongation complexes to fully mature at the promoter-proximal pausing site, such as by failing to recruit critical elongation or processing factors, might trigger premature termination by Xrn2. This role of Xrn2 might prevent release of these elongation complexes into a suboptimal elongation phase that might negatively impact gene expression or rapid upregulation of nascent transcript synthesis as shown on heat shock genes (Fig. 5-36).

Further evidence that polymerases targets of Xrn2 are not properly associated with factors important for transcription was the result that upon inhibition of Xrn2, the increased population of creeping polymerases were not accompanied by detectable signal of a subunit of NELF (NELFa) under inhibition of pTEF-b with DRB, which would be expected to be bound to the elongation complex under inhibition of pTEF-b (Fig. 5-21 – 5-23). According to the current model pTEF-b is required for release of NELF from the elongation complex around promoters (Peterlin and Price
2006; Renner et al. 2001). That observation might reflect a failure in the system decoding pausing signals around promoters by pol II and explain loss of the paused state for these polymerases. If creeping polymerases are in fact not bound by NELF, it may also indicate that NELF can eventually release paused polymerases without the positive action by pTEF-b. Or more intriguingly, it would open the possibility that creeping polymerases are not repressed by NELF and correspond to a fraction of polymerases for which promoter-proximal pausing is not an obligatory step during transcription.

Overall, multiple lines of evidence suggest a model (Fig. 5-38) in which Xrn2 targets for premature termination paused polymerases that fail to mature into an optimal elongation complex that can efficiently transcribe within gene bodies. The importance of Xrn2 in clearance of these suboptimal elongation complexes is underlined by the fact that their accumulation within the body region of HSP genes is accompanied by impaired induction of the heat shock response, an essential mechanism of cells required to overcome the proteotoxic environment generated by heat.
After transcription initiation, pol II pauses at the promoter-proximal region and the elongation complex matures into a fully elongation-competent complex for productive elongation. Failure to mature into a fully elongation-competent complex produces a 5’ PO₄ RNA end by either decapping factors, endonucleolytic cleavage, or pyrophosphohydrolase activity on a 5’ triphosphorylated end of the nascent transcript. Xrn2 in the elongation complex recognizes the generated 5’ PO₄ RNA end and degrades the nascent transcript in a chase for pol II that results in premature termination of transcription upon catching-up with the elongation complex.
CHAPTER VI

GLOBAL RESOLUTION OF R-LOOP STRUCTURES IN RESPONSE TO INHIBITION
OF DNA TOPOISOMERASE I

6.1 Introduction

During transcription, the newly synthesized nascent transcript can anneal with the DNA template strand and form what is called an R-Loop structure (Fig. 1-5). Thus, the R-Loop structure consist of an RNA-DNA hybrid and a non-template displaced DNA strand (Thomas, White, and Davis 1976). R-Loops are known to have important functions under physiological and pathological conditions. These structures promote class-switch recombination in activated B cells (Yu et al. 2003), they can positively influence gene expression by blocking repressive modifications on the DNA and histones (Aguilera and Garcia-Muse 2012; Skourt-Stathaki and Proudfoot 2014; Powell et al. 2013; Grunseich et al. 2018; Ginno et al. 2012) and can also contribute to transcription termination potentially by facilitating pol II pausing downstream from the polyA signal (Skourt-Stathaki et al., 2014; Skourt-Stathaki et al., 2011; Lionel A. Sanz 2016). In addition, formation of R-Loop structures on ribosomal DNA (rDNA) is conserved from yeast to humans (Nadel et al. 2015; Ginno et al. 2012; Chan, Aristizabal, et al. 2014) but their function remains unknown. Nevertheless, mounting evidence suggests that uncontrolled levels of R-loop formation induce genomic instability and disease (Groh and Gromak 2014). Their toxicity seems to originate from the increased sensitivity of the displaced DNA strand to damage (Li and Manley 2006; Aguilera and Garcia-Muse 2012) and/or conflicts with the replication fork leading to double strand breaks (Aguilera 2002; Gan et al. 2011; Houlard et al. 2011).

Different enzymes can target R-Loops in mammalian cells. These include members of the RNase H family that can degrade the RNA moiety in the RNA-DNA hybrid (Cerritelli and Crouch
207), or helicases such as SETX and DHX9 that can unwind the RNA/DNA hybrid (Skourtis-Stathaki, Proudfoot, and Gromak 2011; Cristini et al. 2018). Additionally, recent evidence suggests that splicing factors play a critical role in limiting R-Loop formation and disease (Sorrells et al. 2018; Chen et al. 2018; Li and Manley 2005). Multiple high-risk cancer mutations in splicing factors, known to cause dissimilar splicing patterns, commonly augment R-Loop formation (Chen et al. 2018). Interestingly, a recent report arguing that a free 5′ RNA end is necessary to induce R-Loop formation around promoters (Chen et al. 2017) raises the possibility that 5′-3′ RNA exonucleolytic digestion could also play a role in elimination of R-Loops. If free 5′ RNA ends are/or become monophosphorylated, the nuclear 5′-3′ RNA exonuclease Xrn2 might contribute to elimination of R-Loops via recognition of the free 5′ end followed by 5′ to 3′ degradation of the nascent transcript. Xrn2 is a transcription termination factor that can efficiently degrade the RNA moiety in a RNA-DNA hybrid after phosphorylation by Cdk9 in vitro (Sanso et al. 2016).

Notably, in addition to specific sequences or DNA topology (Ginno et al. 2012; Duquette et al. 2004), negative superhelicity behind pol II is a major facilitator of R-Loop formation (Roy et al. 2010; Drolet, Bi, and Liu 1994). This is explained by the fact that negative superhelicity leads to a more open DNA structure that might facilitate annealing of RNA with the DNA template. To prevent this, human cells also possess topoisomerases that relax supercoiled DNA (Wang 2002). Topoisomerase I, for instance, creates a single strand break on supercoiled DNA and covalently binds the nicked 3′ end allowing controlled rotation around the intact strand. After the DNA has relaxed, topoisomerase I re-ligates the nicked strand and releases DNA (Koster et al. 2005). Interestingly, a highly effective anticancer agent called camptothecin (CPT) targets the active site of topoisomerase I and affects both its rate of uncoiling (Koster et al. 2007) and the re-ligation step leading to accumulation of R-Loops (Marinello et al. 2013). The efficacy of CPT in cancer
treatment depends on its ability to induce R-Loop formation, double strand breaks and apoptosis (Pommier 2006). Thus, maintaining R-Loop structures under control is critical for cell integrity.

Analysis using a monoclonal antibody (S9.6) specific for RNA/DNA hybrids (Boguslawski et al. 1986), and more recently a catalytically dead RNase H mutant (Chen et al. 2017; Ginno et al. 2012), have revealed that R-Loops occur frequently in the yeast genome (Chan, Aristizabal, et al. 2014; El Hage et al. 2014; Wahba et al. 2016) and are also a common feature of human promoters (Ginno et al. 2012; Nadel et al. 2015; Sanz et al. 2016). It is thought that a major role of these structures around promoters is to promote transcription by blocking DNA methylation. An interesting unsolved question, however, is whether pol II transcribed genes are able to survey and control R-Loop formation to prevent their potential negative effects.

Intriguingly, CPT causes accumulation of R-Loops in nucleoli as evidenced by fluorescent microscopy staining using the S9.6 antibody (Marinello et al. 2013), but whether R-Loops around promoters are similarly affected has not been addressed. Accumulation of R-Loops around promoters upon treatment of cells with CPT has been assumed to occur based on observations that CPT treatment not only results in increased recruitment and escape of pol II from the promoter-proximal pause, but also in enhanced antisense transcription and formation of more open chromatin (Baranello et al. 2010; Capranico, Marinello, and Baranello 2010; Bertozzi et al. 2011; Ljungman and Hanawalt 1996). Nevertheless, pol II transcribed genes might be less prone to accumulation of R-Loops in CPT. A recent study showed that R-Loop formation transiently increases on a few genes upon addition of CPT to HCT116 cells but R-Loop levels at those sites rapidly decrease within the first 10 minutes of treatment (Marinello et al. 2016).

Here I investigated R-Loop formation in mammalian cells and provide evidence that pol II transcribed genes effectively eliminate R-Loops from promoters and gene bodies upon treatment
with CPT. Consistent with previous observations, CPT treatment resulted in accumulation of R-Loops in the rDNA repeating unit but induced complete loss of R-Loops around promoters genome-wide. Furthermore, analysis of nascent transcripts on the GAPH gene suggests that loss of R-Loops does not occur via degradation of the nascent transcript, but rather through resolution of the RNA-DNA hybrid. The evidence reported here strongly suggests that CPT-mediated inhibition of topoisomerase I results in activation of a protective mechanism capable of targeting R-Loops for resolution on pol II transcribed genes.

6.2 Results

6.2A R-loop Structures are a Common Feature of Human Promoters and Termination Regions

To investigate formation of R-Loop structures during transcription by RNA pol II, DNA-RNA immunoprecipitation coupled to sequencing (DRIP-seq) was performed (Ginno et al. 2012). In this approach DNA-RNA hybrids are immunoprecipitated using the S9.6 antibody and DNA fragments are identified by high throughput sequencing. As negative control, samples were treated with RNase H before the immunoprecipitation step. For comparison, previously reported data by the Chèdin’s laboratory (Sanz et al. 2016) was downloaded and included for analysis. Importantly, treatment of samples with RNase H before immunoprecipitation eliminated DRIP-seq signal (Fig 6-1, 6-2) confirming specificity of the S9.6 antibody.

I found that the distribution of R-Loops in HEK293 cells shares features with that from Ntera2 cells at the promoter region (Fig. 6-3) (Sanz et al. 2016). In particular, both data sets agree in a bimodal peak of DRIP-seq density around promoters with its valley located in close proximity to the transcription start site (TSS) (Fig. 6-3). On the GAPDH and RPS15 genes it can also be observed that R-Loop formation upstream of the TSS is accompanied by Bruseq signal of divergent transcription (Fig 6-4) (indicated by red arrows). This result is consistent with the
Figure 6-1. Distribution profile of R-Loops on the GAPDH gene
Genomic DNA was extracted from HEK293 cells and DRIP-seq libraries were performed. As negative control, genomic DNA was treated with RNase H before immunoprecipitation of RNA-DNA hybrids. Shown is the DRIP-seq distribution profile on the GAPDH gene for two independent experiments. In addition, DRIP-seq data previously published was downloaded and mapped to the human genome for comparison (Sanz et al. 2016). UCSC genome browser screen shot. Blue arrow indicates the direction of transcription.
Figure 6-2. Global R-Loop distribution on most active genes

DRIP-seq signal from HEK293 cells was plotted on the top 25% of most active genes. As negative control, genomic DNA was treated with RNase H before immunoprecipitation. Shown is the mean distribution profile of R-Loops normalized to signal at -1.5 kb upstream from the transcription start site (TSS). Note that signal is particularly low at the TSS but peaks at its boundaries (indicated by red arrows). The position of the cleavage and polyadenylation site (CPS) is marked by a red dotted line.
Figure 6-3. Comparison of DRIP-seq distribution profiles with previously reported data

DRIP-seq signal from HEK293 cells was plotted on the top 25% of most active genes. As negative control, genomic DNA was treated with RNase H before immunoprecipitation. Shown is the mean distribution profile of R-Loops normalized to signal at -1.5 kb upstream from the transcription start site (TSS). The position of the cleavage and polyadenylation site (CPS) is marked by a red dotted line. DRIP-seq signal previously published by the Chédin Lab has been included for comparison. Note that DRIP-seq signal accumulates downstream of the CPS is HEK293 cells.
Figure 6-4. Upstream R-Loop formation coincides with divergent nascent transcript signal from the TSS

Genomic DNA was extracted from HEK293 cells and DRIP-seq libraries were performed. Shown is the DRIP-seq distribution profile on the GAPDH and RPS15 genes with their respective Bruseq signal of nascent transcripts. Note that presence of R-Loop structures coincides with divergent transcription from the GAPDH and RPS15 promoters (indicated by red arrows). UCSC genome browser screen shot. Blue arrow indicates the direction of transcription.
observation that antisense transcripts can form R-Loops around human promoters (Marinello et al. 2013).

In contrast to DRIP-seq data in Ntera2 cells, my data appears to detect global R-Loop formation within termination zones as reported in an independent study at 3’ ends of mammalian genes (Sanz et al. 2016) (Fig. 6-3) (indicated by a red arrow). In fact, expression of dominant negative (D235A) Xrn2MT, which causes increased transcription at 3’ ends of genes (Chapter IV) (Fong et al. 2015b), results in increased R-Loop formation farther downstream of termination zones (Fig. 6-5, 6-6). These data suggest that R-Loops are a common feature of human promoters and termination regions.

6.2B Camptothecin Treatment Eliminates R-loops from pol II Transcribed Genes

The topoisomerase I inhibitor CPT leads to increased R-Loop formation during transcription by RNA pol I due to increased DNA negative superhelicity (Marinello et al. 2013). To investigate if CPT treatment leads to accumulation of R-Loops during transcription by RNA pol II, Xrn2WT and Xrn2MT cells were treated with CPT. Consistent with a block to elongation, treatment with CPT (10µM, 30min) led to global accumulation of pol II density within gene bodies (Fig. 3-4 – 3-6). To visualize R-Loop signal during synthesis of the primary ribosomal transcript by RNA pol I, DRIP-seq libraries were mapped to a custom-built hg19 assembly containing a rDNA repeating unit. In agreement with previous observations, treatment with CPT resulted in stronger R-Loop signal on the rRNA transcription unit, especially towards its 5’ end where the primary transcript coding for the 18S, 5.8S and 28S ribosomal RNAs is located (Fig. 6-7).

Strikingly, treatment with CPT generates a complete loss of R-Loop structures on pol II transcribed genes. This can be observed on the RPL13 and GAPDH genes (Fig. 6-8 – 6-10) and at the genome-wide level (Fig. 6-11). No substantial differences were detected between Xrn2WT and
Figure 6-5. Inhibition of Xrn2 increases R-Loop formation within termination zones relative to gene bodies
Flp-in TREX cells were treated with doxycycline for 24hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead mutant (D235A) Xrn2 (Xrn2MT). The DRIP-seq distribution profile reveals increased R-Loop density downstream of the RPL13 and DNAJB1 (HSP40) genes relative to signal within their body region when compared to control cells. UCSC genome browser screen shot. Blue arrow indicates the direction of transcription.
Figure 6-6. Meta-gene plot reveals global increase in R-Loop formation downstream of the polyA site after inhibition of Xrn2
Flp-in TREX cells were treated with doxycycline for 24hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead mutant (D235A) Xrn2 (Xrn2MT). Shown is DRIP-seq density on the top 25% of most active genes. Note that inhibition of Xrn2 results in increased R-Loop formation downstream of the cleavage and polyadenylation site (CPS) in Xrn2MT cells. Mean DRIP-seq density was normalized to signal at -1.5 kb from the transcription start site (TSS).
Figure 6-7. Camptothecin treatment induces R-Loop formation on the rDNA repeating unit
Flp-in TREX cells were treated with doxycycline for 24 hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead mutant (D235A) Xrn2 (Xrn2MT) and followed by treatment with the topoisomerase I inhibitor, camptothecin (CPT) (10 µM, 30min). The sequence of the rDNA repeating unit was introduced in the initial telomeric region of chromosome 1 (chr1) and DRIP-seq reads were mapped to the rDNA unit in the context of the hg19 reference genome. Shown is the DRIP-seq distribution profile on the human ribosomal repeating unit. Note that CPT treatment generates an increase in R-Loop signal toward the 5’ end of the rDNA unit where the pre-rDNA transcription unit is located. UCSC genome browser screen shot
Figure 6-8. Camptothecin treatment eliminates R-loops on the GAPDH gene
HEK293 cells were treated with the topoisomerase I inhibitor, camptothecin (CPT) (10 μM, 30 min), and RNA-DNA hybrids were immunoprecipitated using the S9.6 antibody. As negative control an input sample was pre-treated with RNase H before immunoprecipitation of RNA-DNA hybrids. (A) DRIP-qPCR on the GAPDH gene using primer pairs targeting different positions. (B) Map of the GAPDH gene with positions of primer pairs used for DRIP-qPCR analysis. The bent arrow indicates the position of the transcription start site. The red downward arrow indicates the position of the polyadenylation site (pA). Error bars correspond to the standard error of technical replicates in single experiment (n=3).
Figure 6-9. Camptothecin treatment eliminates R-Loop formation in both Xrn2WT and Xrn2MT cells on the RPL13 gene

Flp-in TREX cells were treated with doxycycline for 24hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead mutant (D235A) Xrn2 (Xrn2MT) and followed by treatment with the topoisomerase I inhibitor, camptothecin (CPT) (10 µM, 30 min). Shown is the distribution profile of R-Loops on the RPL13 gene. Note that CPT treatment causes complete loss of R-Loops in both cell lines. UCSC genome browser screen shot. Blue arrow indicates direction of transcription.
Figure 6-10. Camptothecin treatment eliminates R-Loop formation in both Xrn2WT and Xrn2MT cells on the GAPDH gene

Flp-in TREX cells were treated with doxycycline for 24hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead mutant (D235A) Xrn2 (Xrn2MT) and followed by treatment with the topoisomerase I inhibitor, camptothecin (CPT) (10 µM, 30 min). Shown is the distribution profile of R-Loops on the GAPDH gene. Note that CPT treatment causes complete loss of R-Loops in both cell lines. UCSC genome browser screen shot. Blue arrow indicates direction of transcription.
Figure 6-11. Camptothecin treatment generates loss of R-Loops genome-wide
Flp-in TREX cells were treated with doxycycline for 24hrs for expression of wild-type Xrn2 (Xrn2WT) or exonuclease-dead (D235A) Xrn2 (Xrn2MT) and followed by treatment with the topoisomerase inhibitor camptothecin (CPT) (10 μM, 30 min). Shown is DRIP-seq density on the top 25% of most active genes aligned at the transcription start site (TSS) and the cleavage and polyadenylation site (CPS). Note that CPT treatment results in a dramatic loss of R-Loop structures throughout the length of genes in both cell lines. Mean DRIP-seq density was normalized to signal at -1.5 kb from the TSS.
Xrn2MT cells. Interestingly, no changes in R-Loop formation were observed on the mitochondrial genome during the course of CPT treatment (Fig. 6-12) showing that loss of R-Loop structures is unique to chromosomal, pol II transcribed genes. Importantly, the signal detected by the S9.6 antibody cannot be explained by detection of the RNA-DNA hybrid in the active site of pol II. In the presence of CPT pol II strongly accumulates within gene bodies (Fig. 3-4 – 3-6), yet the S9.6 antibody does not detect RNA-DNA hybrids within gene bodies (Fig 6-8 – 6-11). In summary, these data show that camptothecin treatment induces accumulation of R-Loops specifically on the rDNA repeating unit but generates a global loss of R-Loop structures on pol II transcribed genes.

Loss of R-Loop structures during CPT treatment might occur via two possible mechanisms: (1) Digestion of the RNA moiety in the RNA-DNA hybrid, which could be accomplished by RNase H activity (Cerritelli and Crouch 2009), or (2) Resolution of the RNA-DNA hybrid by a helicase such as SETX (Skourti-Stathaki, Proudfoot, and Gromak 2011). To elucidate which of these two options might be taking place, levels of nascent transcripts were analyzed before and during incubation of cells with CPT. If R-Loops are eliminated by degradation of the RNA moiety, a potential decrease in nascent transcript signal would be expected to occur under CPT. If alternatively, R-Loops are resolved by a helicase that does not degrade the nascent transcript in the RNA-DNA hybrid, levels of nascent transcripts are not expected to decrease. To elucidate which possibility explains elimination of R-Loops, I labelled nascent transcripts by incubating cells with 5-bromouridine in normal conditions or simultaneously with incubation with CPT. Interestingly, plotting of the fold change in nascent transcript signal detected by RT-qPCR revealed that nascent transcripts remain mostly unchanged on the GAPDH gene (Fig. 6-13). In fact, consistent with data suggesting that CPT treatment leads to increased escape of paused polymerases from the promoter region (Marinello et al. 2016), a modest increase in nascent
transcript signal was detected from the TSS up to ~1kb downstream on the GAPDH gene in both cell lines. Additionally, Xrn2MT cells did not show a noticeable alteration in the fold-change of nascent transcript levels on the GAPDH gene when compared to control Xrn2WT cells (Fig. 6-13), further suggesting that CPT-mediated elimination of R-Loops does not occur via degradation of nascent transcripts by Xrn2. Taken together, these results suggest that CPT treatment leads to elimination of R-Loop structures on pol II transcribed genes via resolution of the DNA-RNA hybrid, rather than digestion of the nascent transcript.

6.3 Discussion

While R-Loops appear to positively influence gene expression around promoters by blocking repressive modifications on the DNA and histones (Aguilera and Garcia-Muse 2012; Skourti-Stathaki and Proudfoot 2014; Powell et al. 2013; Grunseich et al. 2018; Ginno et al. 2012), they are also a potential threat to genome integrity. This dual property of R-Loops inevitably suggests that cells must have also evolved mechanisms to keep the potential harmful effects of R-Loops under control. In effect, the data reported here suggest that cells can respond to increased levels of R-Loop formation and eliminate these structures via resolution of the RNA-DNA hybrid.

In this study I used CPT to inhibit topoisomerase I and showed that in agreement with a previous report (Marinello et al. 2013) it induces accumulation of R-Loops on the ribosomal DNA repeating unit (Fig. 6-7). Strikingly, this was not the case on protein coding genes (Fig. 6-8 – 6-11). While R-Loops can be strongly detected in normal conditions around promoters, treatment with CPT led to a complete loss of R-Loops on protein coding genes similar to the effect of pre-treating input samples with RNase H as a negative control (Fig. 6-8 – 6-11). This unexpected result strongly suggests that R-Loops are actively eliminated in response to inhibition of topoisomerase I with CPT. Loss of R-Loop signal cannot be explained by repression of transcription since CPT
Figure 6-12. Camptothecin treatment does not affect the global distribution profile of R-Loops in the mitochondrial genome

HEK293 Flp-in TREX cells were treated with doxycycline for 24hrs for expression of wild-type (WT) Xrn2 (Xrn2WT) or exonuclease-dead, mutant (D235A) Xrn2 (Xrn2MT). Cells were treated with the topoisomerase I inhibitor, camptothecin (CPT) (10 µM, 30 min). As negative control an input sample was pre-treated with RNase H before immunoprecipitation of RNA-DNA hybrids. Shown is DRIP-seq density on the mitochondrial genome.
Figure 6-13. Camptothecin treatment increases nascent transcript synthesis at least 1kb into the body region of the GAPDH gene from its promoter
To assess levels of nascent transcript synthesis during camptothecin (CPT) treatment, nascent transcripts were labelled with 5-bromouridine in normal conditions or simultaneous with CPT treatment (10 µM, 30 min). Nascent transcripts were isolated via immunoprecipitation as for Bruseq but analyzed by reverse transcriptase (RT) qPCR. HEK293 Flp-in TREX cells expressing wild-type (WT) Xrn2WT or dominant negative, mutant (MT) (D235A) Xrn2MT were used for analysis. (A) RT-qPCR on the GAPDH gene using primer pairs targeting different positions as indicated. (B) Map of the GAPDH gene with positions of primer pairs used for RT-qPCR analysis. The bent arrow indicates the position of the transcription start site. The red downward arrow indicates the position of the polyadenylation site (pA).
treatment leads to accumulation of pol II within gene bodies (Fig. 3-4 – 3-6). In addition, nascent transcript signal was not reduced under CPT treatment (Fig. 6-13), consistent with results showing that CPT slows down transcription elongation but does not completely halt the transcription process (Veloso et al. 2013). In fact, synthesis of nascent transcripts increased at the 5’ end of the GAPDH gene (Fig. 6-13), in line with reports suggesting that CPT facilitates pol II recruitment and alleviates the promoter-proximal pause (Baranello et al. 2010; Capranico, Marinello, and Baranello 2010; Bertozzi et al. 2011; Ljungman and Hanawalt 1996). I conclude that CPT-mediated loss of R-Loops occurs by active resolution of the RNA-DNA hybrid, rather than by degradation of the RNA moiety.

The discovery that RNA-DNA hybrids are resolved genome-wide during CPT treatment is particularly relevant given that it shows that inhibition of topoisomerase I with CPT does not lead to global accumulation of R-Loops, contrary to the implication of initial results at the rDNA locus (Marinello et al. 2013). This notion has been supported by both immunofluorescence showing accumulation of R-Loops at nucleoli (Marinello et al. 2013) and detection of enhanced transcription around promoters upon treatment with CPT (Ljungman and Hanawalt 1996; Baranello et al. 2010); also shown here downstream of the GAPDH gene promoter (Fig. 6-13). In this study, however, I have found that CPT leads to accumulation of R-Loops specifically on the rRNA genes but not on pol II transcribed genes where R-Loops disappear under CPT. Interestingly, no changes in R-Loop formation were observed on the mitochondrial genome during the course of CPT treatment (Fig. 6-12). While this may indicate that mitochondrial topoisomerase I (Zhang et al. 2001), which is highly homologous to nuclear topoisomerase I, is not particularly sensitive to CPT, this result could be better explained by the limited permeability of mitochondria
to non-cationic molecules and/or because CPT is readily inactivated at alkaline pH (Dalla Rosa et al. 2014).

Even though R-Loop signal was analyzed after 30 min of CPT treatment and not immediately after addition of CPT to cells, a previous report looking at a small group of pol II promoters showed that R-Loops transiently accumulate after 2 minutes of treatment using the same CPT concentration (Marinello et al. 2016). Therefore, it remains possible that a sudden increase in R-Loop formation may trigger activation of a protective response capable of effectively resolving these structures. CPT prevents the final re-ligation step required to eliminate the covalent bond between topoisomerase I and DNA. As a consequence, the stabilized topoisomerase I can become a road block on the DNA template (Pommier 2006). It could also be envisioned that prolonged interaction of topoisomerase I with the DNA template, or collisions of the DNA replication fork with topoisomerase I bound on the DNA template (Capranico, Marinello, and Baranello 2010; Pommier 2006) might generate signals leading to activation of the proposed protective response. For example, prolonged binding of Topo1 on the DNA template might be recognized by a particular factor leading to activation of the response.

The differential effects of CPT on transcription by RNA pol I and pol II further suggest that factors associated with pol II might be involved in triggering resolution of R-Loops. Splicing factors are particularly attractive for such a role since inhibition of RNA processing or mutations in splicing factors can generate more R-Loops (Chen et al. 2018). For instance, sequestration of the RNA by the DNA template would prevent access of splicing factors and progression of the splicing reaction activating potential checkpoints (Chathoth et al. 2014; Brzyzek and Swiezewski 2015).
The results presented here were obtained in two independent cell lines engineered to express dominant negative Xrn2MT or wild-type Xrn2WT. In addition to representing replicates of two independent experiments, these cell lines further indicate that Xrn2 is not involved in elimination of R-Loops during CPT treatment (Fig. 6-11). This conclusion is also consistent with the observation that levels of nascent transcripts on the GAPDH gene are not reduced upon treatment of cells with CPT, suggesting that nascent transcripts are not degraded in CPT (Fig. 6-13). Nevertheless, inhibition of Xrn2 resulted in accumulation of R-Loops farther downstream within termination zones in normal conditions (Fig. 6-6). This is consistent with a failure in transcription termination after inhibition of Xrn2 shown in Chapter III, IV and (Fong et al. 2015b), and formation of R-Loops at 3’ ends of mammalian genes as previously reported (Sanz et al. 2016); although to a lesser extent than at promoter regions.

Taken together, it can be concluded that CPT-mediated inhibition of topoisomerase I results in accumulation of R-Loops at nucleoli but triggers global resolution of R-Loops at pol II transcription units. In addition, I propose that inhibition of topoisomerase I activates a rapid and potent response to protect cells from accumulating RNA-DNA hybrids, known to induce double DNA strand breaks and genomic instability.
CHAPTER VII

CONCLUSIONS

7.1 The unified Allosteric-Torpedo Model of RNA pol II Transcription Termination

For over three decades an extensive debate has been focused on two models aiming to explain termination of RNA polymerase II transcription (Libri 2015). In the torpedo model, RNA cleavage is proposed to expose a 5’ PO₄ RNA end at the polyA site that is attacked by the 5’-3’ exonuclease Xrn2 (Fig. 1-2). In this model Xrn2 degrades the nascent transcript up to the elongation complex and somehow induces pol II disassembly from the DNA template (Connelly and Manley 1988b). On the other hand, the allosteric model asserts that a polyA site-dependent conformational change, or loss of an anti-termination factor, transitions the elongation complex into a termination competent state (Fig. 1-3) (Greenblatt, Nodwell, and Mason 1993; Logan, Falck, et al. 1987). Importantly, results presented in this study suggest that the molecular events described in both models are in fact not mutually exclusive and appear to cooperate at 3’ ends of genes to afford proper termination of pol II transcription.

In support of the torpedo model I have shown for the first time that inhibition of the exonuclease activity of Xrn2, via expression of dominant negative (D235A) Xrn2MT, leads to a global accumulation of 5’ PO₄ RNA ends at polyA sites in mammalian cells (Fig. 3-17, 3-19, 3-20). This result is an essential prediction of the torpedo model not previously demonstrated in vivo or in vitro. In agreement with the idea that Xrn2 chases pol II for termination, Xrn2 loads onto the nascent transcript for degradation after cleavage at the polyA site. Interestingly, nascent 5’ PO₄ ends generated by cleavage events independent of a polyA signal (i.e. cleavage at the 3’ end of histone genes, tRNAs and the NEAT1 gene) were also stabilized after inhibition of Xrn2 (Fig. 3-14, 3-17, 3-21), demonstrating that different types of cleavage events at 3’ ends of genes

229
commonly provide an entry point for Xrn2. Thus, degradation of the nascent transcript at 3’ ends of genes is a widespread function of Xrn2 that does not necessarily require a polyA signal; as long as a 5’ PO₄ end is provided.

Mapping of nascent 5’ PO₄ RNA ends in Xrn2MT cells allowed me to confirm previous suggested roles of Xrn2. Consistent with reports arguing that Xrn2 participates in processing of the ribosomal precursor RNA (pre-rRNA) and degradation of aberrant pre-rRNA fragments (Wang and Pestov 2011; Amberg, Goldstein, and Cole 1992), inhibition of Xrn2 resulted in an increase of 5’ PO₄ ends near the transcription start site of the pre-rRNA (Fig. 3-15, 3-16). Additionally, it has been proposed that Xrn2 contributes to termination of RNA pol I transcription in a torpedo-like mechanism after cleavage by the Rnt1 endonuclease (El Hage et al. 2008; Kawauchi et al. 2008). In support of that idea, inhibition of Xrn2 generated accumulation of 5’ PO₄ ends and nascent transcripts downstream and in close proximity to the annotated 28S rRNA where termination of RNA pol I is thought to occur (Fig. 3-23, 3-24). Finally, analysis of 5’ PO₄ ends indicated that Xrn2 additionally trims the 5’ leaders of intronic snoRNAs after intron debranching (Fig. 3-17, 3-18). In the future, it will be interesting to identify the binding partners of Xrn2 that are responsible for its recruitment and coordination of specific functions at different sites in the genome. The available mass spectrometry database of interacting partners of Xrn2 (Brannan et al. 2012), along with ChIP-seq analysis of potential partners at specific loci may be instrumental to elucidate site-specific interactions of Xrn2 that are relevant for its function.

In particular, previous studies have argued for a role of Xrn2 in termination in yeast and mammalian cells based on the observation that pol II or nascent transcripts accumulate farther downstream of polyA sites after depletion of Xrn2 (Baejen, Andreani, Torkler, Battaglia, Schwalb, Lidschreiber, Maier, Boltendahl, Rus, Esslinger, Söding, et al. 2017; West, Gromak, and Proudfoot
2004; Gromak, West, and Proudfoot 2006; Eaton et al. 2018) or inhibition of its exonuclease activity (Fong et al. 2015b; Kim et al. 2004). In this investigation, however, I additionally show that Xrn2 is required for efficient release of pol II from the DNA template by direct observation of stabilized elongation complexes that fail to be removed downstream of polyA sites in Xrn2MT cells (Fig. 4-2 – 4-5). This was clearly detected by analysis of the last wave of elongation complexes after imposing a block to transcription initiation using triptolide. While control cells cleared pol II ChIP-seq density at 3’ ends of genes after treatment with triptolide, Xrn2MT cells accumulated polymerases downstream of the polyA site and failed to remove pol II in a timely manner (Fig. 4-2 – 4-5). Therefore “torpedo” action by Xrn2 after cleavage at the polyA site is essential for proper disassembly of elongation complexes from the DNA template. I will refer to this strategy of utilizing triptolide to assess removal of pol II at 3’ ends of genes as the “3’-end disassembly assay”.

How Xrn2 is able to dismantle the elongation complex from the DNA template remains to be elucidated. Previous reports have shown that degradation of the nascent transcript alone is not sufficient to induce termination of pol II in vivo and in vitro (Pearson and Moore 2013; Luo, Johnson, and Bentley 2006; Dengl and Cramer 2009). It is then likely that degradation of the nascent transcript serves as a conduit that guides Xrn2 to contact pol II or to recruit a helicase factor associated with Xrn2 for termination.

The possibility that Xrn2 contributes to termination of transcription by recruiting a helicase factor is particularly interesting. Termination factors in different biological systems commonly utilize ATP hydrolysis to disrupt the RNA-DNA hybrid in the pol II active site: Rho in bacteria (Nudler and Gottesman 2002), NPH-I in vaccinia virus (Deng and Shuman 1998) and Sen1 in yeast (Steinmetz et al. 2001; Porrua and Libri 2013). Thus, it seems highly possible that in order
for Xrn2 to promote template release by pol II, a helicase might also be required. The transcription termination factor 2, TTF2, is a DNA-dependent ATPase that might serve this function and that may be recruited by Xrn2. TTF2 has helicase activity toward RNA-DNA hybrids and is required to release the nascent transcript and pol II from the DNA template during mitosis (Jiang et al. 2004). Recruitment of TTF2 by Xrn2 is supported by the fact that this protein was identified amongst the most strongly enriched factors in a mass spectrometry analysis of Xrn2 immunoprecipitates (Brannan et al. 2012). It remains to be tested whether or not TTF2 is required for termination of interphase transcription. The 3’-end disassembly assay might be a suitable tool to test its role in termination. If the helicase activity of TTF2 is required for DNA release, inhibition and depletion of TTF2 might lead to accumulation of elongation complexes at 3’ ends of genes similar to inhibition of Xrn2 in this study. Alternatively, the RNA-DNA helicase SETX, known to contribute in termination of transcription at some specific genes (Skourti-Stathaki, Proudfoot, and Gromak 2011), could also be tested following the same approach.

My experiments strongly suggested that elongation complexes undergo a conformational change at the polyA site that triggers slowdown of polymerases before termination by Xrn2 (Chapter IV, 4.2B). Furthermore, binding of cleavage and polyadenylation factors to the polyA signal on the nascent transcript was found to be the event most likely communicating deceleration to the elongation complex (Chapter IV, 4.2C). This is consistent with a previous report showing that simultaneous binding of CPSF30 to the polyA signal and the body of pol II induces pol II pausing (Nag, Narsinh, and Martinson 2007). A role of CPSF30 in inducing deceleration of pol II upon recognition of the polyA signal might explain the observation that RNA 3’-end processing and termination are coupled. In principle, a reduction in the rate of nascent transcript synthesis after recognition of the polyA signal would favor Xrn2-mediated termination by shortening the
path of Xrn2 to “torpedo” the elongation complex. Additionally, the polyA site-dependent conformational change of the elongation complex, and not just deceleration of transcription elongation, might be a prerequisite for Xrn2 to displace pol II from the DNA template. It is possible that recognition of the polyA signal by the 3’-end processing machinery alters the composition of the elongation complex by exchange of factors that is needed for Xrn2-dependent termination of pol II. Such an additional role of the polyA signal, independent of polyA site cleavage, in Xrn2-mediated termination of pol II might explain why at the end of the human β-gene gene the polyA signal is still required for termination in spite of the fact that cleavage at the polyA site is not required (West, Gromak, and Proudfoot 2004; West, Proudfoot, and Dye 2008). In those reports, it was proposed that entrance of Xrn2 to the nascent transcript and termination at the end of the β-globin gene is satisfied by a self-cleaving (CoTC) RNA element located downstream of the polyA signal, and that Xrn2 is required for termination at that site. It is possible that the role of the polyA signal at the end of β-gene, in addition to direct 3’ end processing of the nascent transcript, is to induce a conformational change of the elongation complex that is required for Xrn2-mediated torpedo termination downstream of the proposed CoTC element.

An unsolved question is whether a polyA site-dependent conformational change of the elongation complex might still be sufficient to induce termination of pol II in the absence of polyA site cleavage. Xrn2MT cells showed a delay in removal of polymerases at 3’ ends of genes, but polymerases eventually cleared the DNA template (Fig. 4-2 – 4-5), suggesting that other factors might be able to release pol II at 3’ ends independent of the exonuclease activity of Xrn2. This result suggests that termination of transcription in the absence of exonucleolytic digestion of the nascent transcript may still occur, although Xrn2 might be required for efficient termination and prevent accumulation of polymerases on the DNA template at 3’ ends of genes. It is still formally
possible that cleavage and polyadenylation factors are able to destabilize the elongation complex via a conformational change that might lead to DNA release. At least in vitro, the cleavage and polyadenylation factor Pcf11 is able to dislodge pol II from the DNA template in the absence of RNA cleavage (Zhang, Fu, and Gilmour 2005). Interestingly, some specific mutations in Pcf11 affect termination of transcription but do not affect its function in cleavage at the polyA site (Sadowski et al. 2003), suggesting that the functions of Pcf11 in cleavage and termination are separable. It would be interesting to elucidate the mechanism by which these Pcf11 mutations, reported not to affect cleavage (Sadowski et al. 2003), affect termination of pol II. According to my results, deceleration of transcription elongation occurs before release of pol II from the DNA template downstream of the polyA signal (Fig. 4-9). Therefore, it is possible that Pcf11 contributes to either slowing down of pol II or release of pol II from the DNA template, or both. If Pcf11 is required for slowing down polymerases downstream of the polyA site during termination, those mutations in Pcf11 would cause pol II to transcribe farther downstream of polyA sites in the 3’-end disassembly assay, similar to the effect on pol II distribution at 3’ ends seen after expression of NS1A (Fig. 4-11, 4-12). If in contrast Pcf11 is required to remove polymerases at 3’ ends of genes, those mutations in Pcf11 would be expected to cause accumulation of pol II immediately downstream of polyA sites in the 3’-end disassembly assay, similar to accumulation of polymerases at 3’ ends of genes seen after inhibition of Xrn2 (Fig. 4-2 - 4-5).

A potential cooperation of Xrn2 with RNA 3’-end processing factors during termination is suggested by the observation that its homolog in yeast, Rat1, is important for the efficient recruitment of 3’-end processing factors and that inactivation of Pcf11 also affects recruitment of Rat1 (Luo, Johnson, and Bentley 2006). Functional cooperation between Xrn2 and 3’-end processing factors is further supported by the observation that coupling of Xrn2 to 3’ processing
factors was shown to stimulate exonuclease activity in nuclear extract (Kaneko et al. 2007). In this study, recruitment of Xrn2 to the cleavage and polyadenylation machinery was proposed to occur via interactions with p54nrb/PSF, an important factor involved in transcription, splicing and polyadenylation (Rosonina et al. 2005; Liang and Lutz 2006). In support of a concerted action between 3'-end processing factors and Xrn2, my data clearly demonstrates that the polyA signal is involved in slowing down elongation complexes and that Xrn2 targets poorly elongating polymerases downstream of the polyA site (Chapter IV). Thus, the allosteric and torpedo model instead of being mutually exclusive appear to complement each other to better explain termination of pol II transcription at the ends of genes. I propose a unified allosteric-torpedo model in which binding of cleavage and polyadenylation factors to the polyA signal causes a conformational change of the elongation complex that leads to deceleration and converts it into an easy target for the torpedo Xrn2 (Fig. 4-22).

Spt5 is a pol II-associated factor well-known for its role in inducing promoter-proximal pausing (Missra and Gilmour 2010). Additionally, phosphorylation of Spt5 around promoters is proposed to switch it into a positive elongation factor that enhances pol II processivity (Bourgeois et al. 2002a; Fitz, Neumann, and Pavri 2018). Consistent with this idea, Spt5 phosphorylation increased genome-wide from promoter regions into gene bodies and reached maximum levels at about 2 kb from transcription start sites (Fig. 4-16). Intriguingly, the 3’ transition of pol II into a mode of slow transcription elongation was accompanied by a transition of Spt5 into a hypo-phosphorylated state that also depended on recognition of the polyA signal. Notably, inhibition of polyA signal recognition via expression of the NS1A protein or mutation of a consensus polyA sequence inhibited the 3’ transition of both pol II and Spt5 at the polyA site (Chapter IV). Therefore, I speculate that loss of Spt5 phosphorylation might revert back its repressive activity.
on transcription elongation to induce deceleration of the elongation complex. Reports suggesting that Spt5 can induce pol II pausing in the absence of NELF (Ainbinder et al. 2004; Baugh, Demodena, and Sternberg 2009) and PAR-CLIP data showing that Spt5 regains binding to the nascent transcript immediately downstream of the polyA site in yeast (Baejen, Andreani, Torkler, Battaglia, Schwalb, Lidschreiber, Maier, Boltendahl, Rus, Esslinger, Söding, et al. 2017), further suggest a repressive behavior of Spt5 at 3’ ends of genes. It will be important to test if failure to drop Spt5 phosphorylation levels at the polyA site is sufficient to eliminate slowdown of the elongation complex at 3’ ends of genes. A recent publication by the Fisher Lab reported that Spt5 is dephosphorylated by protein phosphatase 1 (PP1) at 3’ ends of genes in yeast (Parua et al. 2018). Using okadeic acid to block PP1 (Swingle, Ni, and Honkanen 2007) might aide to inhibit Spt5 dephosphorylation in the proposed experiment.

7.2 Premature Termination of RNA pol II Transcription by Xrn2

Premature termination of transcription has been previously demonstrated in bacterial systems (Nudler and Gottesman 2002) in yeast (Jiao et al. 2010; Gudipati et al. 2008; Lykke-Andersen and Jensen 2007; Steinmetz et al. 2001) and at viral promoter-proximal regions (Hay, Skolnik-David, and Aloni 1982; Skolnik-David, Hay, and Aloni 1982; Wagschal et al. 2012). Nevertheless, to date premature termination has not been demonstrated for endogenous human genes. Results from this study, however, showed that promoter-proximally paused polymerases are not stable and are highly turned over genome-wide (Fig. 5-9). After imposing a block to transcription elongation with DRB, maintenance of the pol II promoter-proximal peak required continuous transcription initiation events. Inhibition of transcription initiation with triptolide or addition of 350 mM sodium chloride (500 mM total salt) led to a rapid loss of polymerases; within 2 minutes in the case of NaCl treatment (Fig. 5-17 – 5-19). Given that pre-initiation complexes
become resistant to 600 mM KCl after synthesis of the first 10 bases in vitro (Cai and Luse 1987), and that the pol II peak is also lost with triptolide treatment, loss of paused polymerases is best explained by an active premature termination mechanism.

Significantly, inhibition of Xrn2 recovered a fraction of polymerases and nascent transcripts throughout the length of thousands of genes in normal conditions (Fig. 3-13, 5-4, 5-5). This observation strongly suggests that Xrn2 additionally functions in premature termination of transcription. Interestingly, Xrn2 was not required for global turnover of polymerases, but rather for premature termination of a fraction of slow moving or “creeping” polymerases from the promoter region on thousands of genes (Fig. 5-17 – 5-20). This result has raised two important questions. Which factors, other than Xrn2, are involved in premature termination that can account for global loss of the bulk of polymerases at the promoter-proximal region? And why do only a fraction of polymerases creep from the promoter region into gene bodies? It would be interesting to test the role of previously proposed termination factors such as TTF2, Pcf11 or SETX in turnover of paused polymerases as performed for Xrn2 in this study. My results further revealed that creeping polymerases do not appear to be bound by the negative elongation factor NELF helping to explain why these polymerases are able to transcribe beyond the promoter-proximal region. This could indicate that creeping polymerases are not properly loaded with essential elongation factors, which might expose them to premature termination by Xrn2 in a fail-safe mechanism. Given that Xrn2 requires 5’ PO₄ to degrade the nascent transcript it is expected that either decapping or cleavage factors might cooperate to provide an entry point before termination. The decapping factor Dcp2 or the endonuclease activity of Microprocessor may fit the bill. Both factors have been proposed to provide entrance for Xrn2 early after transcription initiation (Brannan et al. 2012; Wagschal et al. 2012). Additionally, the integrator complex, known to
associate with the C-terminal domain of pol II and thought to carry endonuclease activity (Baillat et al. 2005), might also provide an entry point for Xrn2.

The relevance of Xrn2 during synthesis of nascent transcripts could be appreciated from preliminary results showing that the heat shock response is impaired after inhibition of Xrn2 (Chapter V, 5.2G). Significantly, upregulation of nascent transcript signal was not optimal in Xrn2MT cells in spite of the fact that the fold increase in pol II density within the body region of heat shock genes was stronger in Xrn2MT cells (Fig. 5-25 – 5-27, 5-29, 5-36). I interpreted this discrepancy as a pol II “traffic jam” on heat shock genes caused by inhibition of Xrn2 under conditions of increased initiation rates at the promoter. I speculate that elongation complexes that are marked for premature termination by Xrn2 are “suboptimal” complexes that do not efficiently travel on the DNA template. Under conditions of high transcription initiation events, such as during heat shock, “suboptimal” complexes may become a barrier to fast approaching polymerases in close proximity, limiting global upregulation of nascent transcript synthesis. A “traffic jam” could be further investigated by analysis of the last wave of pol II by ChIP-seq after induction of heat shock genes and subsequent inhibition of transcription initiation. Xrn2WT and Xrn2MT cells could be incubated at 42°C for 30 minutes to induce transcription activation on heat shock genes and then triptolide or NaCl could be added to visualize the last wave of pol II. If the traffic jam model is true, the last wave of pol from promoters moving to 3’ ends of genes would be expected to be delayed in Xrn2MT cells.

Taken together, results from this study suggest that Xrn2 targets a fraction of elongation complexes at the promoter-proximal region that are not properly equipped to transition into the elongation phase. It is possible that Xrn2 operates in a fail-safe mechanism when elongation complexes fail to receive a proper cap structure or does not associate with specific elongation
factors. It will be interesting to see how future studies will help to identify which polymerases are targeted by Xrn2 and whether the mechanism employed for eviction of pol II during premature termination is similar to or diverges from polyA-dependent termination at 3’ ends of genes.

7.3 Global Resolution of R-Loops in Response to Inhibition of DNA Topoisomerase I

R-Loop structures are RNA-DNA hybrids generated during transcription when the nascent transcript anneals with the DNA template (Drolet et al. 1995). These structures are a common feature of protein coding genes and are associated with hypo-methylated promoters (Sanz et al. 2016; Ginno et al. 2012) but are also known to be potential threats to genomic instability (Groh and Gromak 2014). Interestingly, it is not known if cells possess surveillance mechanisms that might control R-Loop formation to prevent their potential deleterious effects. Here, I have found evidence suggesting that mammalian cells can prevent accumulation of R-Loops on pol II transcribed genes after treatment with camptothecin (CPT), an inhibitor of topoisomerase I that leads to accumulation of R-Loops at nucleoli.

Consistent with fluorescence microscopy observations showing increased R-Loop formation at nucleoli under CPT treatment (Marinello et al. 2013), DRIP-seq data revealed accumulation of R-Loops on the ribosomal repeating unit after incubation of cells with CPT for 30 minutes (10 μM) (Fig. 6-7). The mitochondrial genome did not display any substantial effects (Fig. 6-12). Strikingly, DRIP-seq signal completely disappeared throughout the length of pol II transcribed genes suggestive of an active mechanism responsible for elimination of these structures under CPT conditions (Chapter VI, 6.2B). Analysis of pol II ChIP-seq density genome-wide (Fig. 3-6) and nascent transcript signal on the GAPDH gene (Fig. 6-13) further demonstrated that polymerases are not lost from gene bodies and that nascent transcripts are not degraded during
treatment with this drug. Therefore, active resolution of the RNA-DNA hybrid rather than degradation of the nascent transcript explains how these structures are eliminated.

It will be interesting to test if any of the different RNA-DNA helicases previously known to target R-Loop structures such as SETX and DHX9 (Skourtis-Stathaki, Proudfoot, and Gromak 2011; Cristini et al. 2018) are involved in resolution of RNA-DNA hybrids in response to CPT treatment. In addition, it would be important to investigate the initial response to CPT treatment to determine if R-Loops accumulate before being resolved. This would strengthen the idea that resolution of R-Loop structures occurs in response to conditions that trigger accumulation of R-Loops in the first place. Nevertheless, the striking complete elimination of R-Loops specifically on pol II transcribed genes appears to suggest the existence of a potent mechanism capable of eliminating R-Loop structures on protein-coding genes. Taken together, it can be concluded that CPT-mediated inhibition of topoisomerase I results in accumulation of R-Loops at nucleoli but triggers global resolution of R-Loops in mammalian cells. I propose that inhibition of topoisomerase I activates a rapid and potent response to protect cells from accumulating RNA-DNA hybrids, known to be potential threats to genomic instability.
REFERENCES


Andrulis, E. D., E. Guzman, P. Doring, J. Werner, and J. T. Lis. 2000. 'High-resolution localization of drosophila spt5 and spt6 at heat shock genes in vivo: roles in promoter proximal pausing and transcription elongation [In Process Citation]', Genes Dev, 14: 2635-49.


Dominski, Z., L X Zheng, R Sanchez, and W F Marzluff. 1999. 'Stem-loop binding protein facilitates 3'-end formation by stabilizing U7 snRNP binding to histone pre-mRNA.', *Molecular and Cellular Biology*, 19: 3561-70.


Fitz, J., T. Neumann, and R. Pavri. 2018. 'Regulation of RNA polymerase II processivity by Spt5 is restricted to a narrow window during elongation', *EMBO J*.


Greger, I. H., and N. J. Proudfoot. 1998. 'Poly(A) signals control both transcriptional termination and initiation between the tandem GAL10 and GAL7 genes of Saccharomyces cerevisiae', *EMBO J*, 17: 4771-9.


Johnson, A. W. 1997. 'Rat1p and Xrn1p are functionally interchangeable exoribonucleases that are restricted to and required in the nucleus and cytoplasm, respectively', *Mol Cell Biol*, 17: 6122-30.

Kaneko, S., O. Rozenblatt-Rosen, M. Meyerson, and J. L. Manley. 2007. 'The multifunctional protein p54nrb/PSF recruits the exonuclease XRN2 to facilitate pre-mRNA 3' processing and transcription termination', *Genes Dev*, 21: 1779-89.


Li, F., T. Jiang, Q. Li, and X. Ling. 2017. 'Camptothecin (CPT) and its derivatives are known to target topoisomerase I (Top1) as their mechanism of action: did we miss something in CPT analogue molecular targets for treating human disease such as cancer?', Am J Cancer Res, 7: 2350-94.


Liang, S., and C. S. Lutz. 2006. 'p54nrb is a component of the snRNP-free U1A (SF-A) complex that promotes pre-mRNA cleavage during polyadenylation', RNA, 12: 111-21.


Logan, J., E. Falck-Pedersen, J. E. Darnell, Jr., and T. Shenk. 1987. 'A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene', *Proc Natl Acad Sci U S A*, 84: 8306-10.

Luo, W., A. W. Johnson, and D. L. Bentley. 2006. 'The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model', *Genes Dev*, 20: 954-65.


Missra, Anamika, and David S Gilmour. 2010. 'Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the Drosophila RNA polymerase II transcription elongation complex', Proceedings of the National Academy of Sciences of the United States of America, 107: 11301-06.


Paulsen, Michelle T, Artur Veloso, Jayendra Prasad, Karan Bedi, Emily A Ljungman, Brian Magnuson, Thomas E Wilson, and Mats Ljungman. 2014. 'Use of Bru-Seq and BruChase-Seq for genome-wide assessment of the synthesis and stability of RNA', Methods (San Diego, Calif), 67: 45-54.


Shukla, S., and R. Parker. 2014. 'Quality control of assembly-defective U1 snRNAs by decapping and 5'-to-3' exonucleolytic digestion', *Proc Natl Acad Sci USA*, 111: E3277-86.


Whitelaw, E., and N. Proudfoot. 1986. 'Alpha-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3' end processing in the human alpha 2 globin gene', *EMBO J*, 5: 2915-22.


Zhang, Z., and D. S. Gilmour. 2006. 'Pcf11 is a termination factor in Drosophila that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript', *Mol Cell*, 21: 65-74.
HEK293 cells were treated with DRB (100 µM, 2.5 hrs) to inhibit escape of pol II from the promoter-proximal pause. Polymerases were released into elongation by DRB wash out and ChIP-seq was performed at time points (T) T0, T10 and T20 minutes after DRB wash out. Shown are the mean distribution profiles of anti-pol II and anti-Xrn2 ChIP-seq densities normalized to signal at -1 kb upstream of the TSS. Long genes (33 kb to 78 kb) were aligned at the transcription start site (TSS). No wave of Xrn2 signal is detected on long genes due to low signal to noise ratio of Xrn2 ChIP-seq libraries. A wave of pol II and Xrn2 can be observed on short genes in figure 3-10.
Figure B-1. Recruitment of Spt5 to promoter regions requires RNA synthesis by RNA pol II

HEK293 Flp-in cells were treated with triptolide (10 µM, 15 minutes) to inhibit transcription initiation. Shown is the mean distribution profile of anti-Spt5 ChIP-seq density for all RefSeq genes normalized relative to signal at -1 kb upstream of the TSS. Note complete loss of Spt5 signal around the transcription start site (TSS) after treatment with triptolide. Signals of Spt5 in control and triptolide conditions come from two independent experiments. Control condition corresponds to Spt5 ChIP-seq signal from HEK293 Flp-in TREX cells containing the pcDNA5-NS1A system in non-inducible conditions (no doxycycline), not relevant for this figure. Spt5 ChIP-seq signal in triptolide conditions is from HEK293 Flp-in TREX cells containing the pcDNA5-Xrn2WT system, after induction of Xrn2 with doxycycline for 24 hours, also not relevant for this figure.
APPENDIX C

DOUBLE STABLE EXPRESSION OF shRNAs AGAINST XRN2 AND TTF2 SUGGESTS A ROLE OF TTF2 IN TERMINATION OF POL II TRANSCRIPTION

The transcription termination factor 2 (TTF2) has been thought to cooperate with Xrn2 in termination of pol II transcription based on the observation that this protein is required for removal of RNA pol II from mitotic chromosomes during replication (Jiang et al. 2004) and can dislodged polymerases from the DNA template in vitro (Cheng et al. 2012; Liu, Xie, and Price 1998). Additionally, mass spectrometry analysis of Xrn2 immunoprecipitates (IP) from RNase-treated HeLa nuclear extract identified TTF2 among the most strongly enriched proteins in the Xrn2 IP (Brannan et al. 2012). Furthermore, depletion of Xrn2 led to an apparent increase in the flux of pol II from promoter regions that was augmented when supplemented with knockdown of TTF2 in HEK293 cells (Brannan et al. 2012). Nevertheless, in this latter report, double knockdown of Xrn2 and TTF2 did not cause a global termination defect of pol II at 3’ ends of genes. To further investigate a potential role of TTF2 in termination of pol II transcription, I utilized the previously published double knockdown of Xrn2 and TTF2 (shXrn2+shTTF2) cell line, and additionally created HEK293 cell lines that can be induced with doxycycline for individual expression of shRNAs against Xrn2 (shXrn2) or TTF2 (shTTF2) using the pINDUCER10 lentiviral system.

Unexpectedly, shXrn2+shTTF2 cells did not display a dramatic change in the distribution profile relative to the control shScr cells as previously published (Brannan et al. 2012), but the same effect of reduced intensity of the promoter-proximal peak of pol II relative to downstream regions was observed (Fig. C-1, C-2) (indicated by red arrows). This can be noticed specially at the genome-wide level (Fig. C-2). Surprisingly, in the present study, a slight increase in pol II density could be observed farther away from the 3’ ends of genes indicative of a termination defect
Figure C-1. Double knockdown of Xrn2 and TTF2 reveals a mild termination defect on the RPL23 and HSP90AA1 genes
Shown is pol II ChIP-seq density from HEK293 cells expressing two shRNAs targeting Xrn2 and TTF2 (shXrn2+shTTF2) or a scrambled shRNA (shScr) control. Note slight increase in pol II density downstream of the polyA site (marked by the red dotted line) of the RPL23 and HSP90AA1 genes in shXrn2+shTTF2 cells relative to control shScr cells (indicated by red arrows). UCSC genome browser screen shot. Blue arrow indicates direction of transcription.
Figure C-2. Double knockdown of Xrn2 and TTF2 reveals a mild termination defect genome-wide

Shown is pol II ChIP-seq density from HEK293 cells expressing two shRNAs targeting Xrn2 and TTF2 (shXrn2+shTTF2) or a scrambled shRNA (shScr) control. Most active genes (top 25% genes with highest pol II density within the body region in HEK293 cells in normal conditions) were aligned at the transcription start site (TSS) and the cleavage and polyadenylation site (CPS). The body region was defined as +0.5 kb downstream of the TSS and -0.5 kb upstream of the CPS and normalized to gene length. Note the increase in pol II density downstream of the CPS relative to signal around promoters in shXrn2+shTTF2 cells when compared to shScr cells.
after depletion of both Xrn2 and TTF2 (Fig. C-1, C-2) (indicated by blue arrows). While analysis of the protein levels of Xrn2 and TTF2 were not successful for this experiment, the fact that this cell line was previously reported to have a partial depletion of Xrn2 and TTF2, and that I detected a mild termination defect, encouraged me to continue investigation with these cell line.

To elucidate the contributions of Xrn2 and TTF2 in termination of pol II transcription at 3’ ends of genes I decided to analyze the distribution profile of pol II after individual depletion of Xrn2 and TTF2 using the inducible-shRNA cell lines. Xrn2 and TTF2 protein levels were effectively knockdown over time after incubation of cells with doxycycline (Fig. 2-6). Cells were treated with doxycycline (3 µg/mL) for 72 hours which causes strong depletion of both Xrn2 and TTF2 down to ~30% (Fig. 2-6). As a control, cells induced for expression of a scrambled shRNA were used in this experiment. Pol II ChIP-seq densities on individual genes or at the genome-wide levels did not reveal substantial changes in the distribution profile of pol II on genes after individual knockdown of Xrn2 or TTF2 (Fig. C-3), consistent with a previous study showing that knockdown of Xrn2 does not result in termination defects at 3’ ends of genes (Nojima et al. 2015b). Nevertheless, in a recent study, an auxin-inducible degron (AID) for rapid Xrn2 depletion in HCT116 cells did reveal a termination defect at 3’ ends of genes detected by mNET-seq (Eaton et al. 2018), opening the possibility that the effects of knocking down Xrn2 are rather short lived and that cells might be able to compensate for its loss via alternative mechanisms of transcription termination, as proposed to exist in Caenorhabditis elegans (Miki, Carl, and Grosshans 2017).

Given that individual depletion of Xrn2 and TTF2 did not reveal a termination defect at 3’ ends of genes, the observed termination defect in Figure C-2 in shXrn2+shTTF2 cells is better explained by an additive inhibition of the functions of Xrn2 and TTF2 in termination of
transcription. I interpret this result to suggest that both Xrn2 and TTF2 is involved in termination of pol II transcription at 3’ ends of genes.

To further investigate the requirement for Xrn2 and TTF2 in rapid upregulation of transcription, I turned to evaluate the response of heat shock genes upon exposure of cells to cadmium sulfate. Heavy metals induce oxidative stress and trigger expression of heat shock genes (Kim et al. 2014). To this end cells were treated with CdSO$_4$ at 50 µM for 4 hours. As expected, pol II ChIP-seq density increased on both the heavy-metal responsive metallothionein MT2A gene (Fig C-4) and on heat shock protein genes (Fig. C-5, C-6), although to a modest extent for the latter. Surprisingly, treatment of cells with CdSO$_4$ uncovered a strong termination defect on the HSP40 and HSP90 genes in shXrn2+shTTF2 cells when compared to control stable shScr cells (Fig. C-5, C-6) (indicated by red arrows). In spite of the strong termination defect detected on the HSP40 and HSP90 genes, this was not a global effect. In order to identify cadmium-responsive genes, I selected genes ranking within the top 5% of genes with highest pol II density within their body region after treatment with CdSO$_4$ (n=966). As expected, metallothionein genes as well as heat shock protein genes were identified in this gene list (See appendix D). Meta-plots of pol II distribution on this group of genes or heat shock inducible genes did not show a major change in the distribution profile of pol II, which agrees with the mild effects on pol II distribution on heat shock genes after heat stress shown in figure 5-28. Evidence that the group of HSP27, HSP40, HSP60, HSP70, HSP90 and HSP105 genes are in fact induced after exposure of cells to CdSO$_4$ was obtained by analysis of their mRNA levels by RT-qPCR (Fig. C-9). Nevertheless, the intriguing observation that CdSO$_4$ treatment generates a termination defect on the HSP40 and HSP90 genes in shXrn2+shTTF2 (Fig. C-5 – C-6) indicates that reduced protein levels of Xrn2 and TTF2 might not support an increased demand for termination after activation of gene
expression. An increased demand for termination at the end of the HSP40 gene is supported by the observation that its mRNA levels were upregulated 40-fold after exposure to CdSO\(_4\) in control shScr cells (Fig C-9).

Next, I investigated mRNA levels from HSP genes before and after exposure of cells to CdSO\(_4\). For this experiment, I included a new biological replicate of cells that were freshly transduced to recapitulate the double shXrn2+shTTF2 cell line (#2). No protein levels were analyzed to confirm knockdown of Xrn2 or TTF2 in this cell line. Consistent with a previous measurement of mRNA levels on the HSP90 gene (Brannan et al. 2012), both shXrn2+shTTF2 cells (#1 and #2) showed a trend of increased mRNA levels for the group of HSP27, HSP40, HSP60, HSP70, HSP90, HSP105, in normal conditions (not significant for all genes) (Fig. C-8). This result is consistent with the result that inhibition of the exonuclease activity of Xrn2 results in increased levels of nascent transcripts on heat shock inducible genes (Fig. 5-36). Exposure of cells to CdSO\(_4\) led to upregulation of mRNA levels from HSP genes, specially HSP40 and HSP70 in control shScr cells (Fig. C-9). Interestingly, both shXrn2+shTTF2 cells (#1 and #2) showed a trend of impaired upregulation of mRNA levels for the group of HSP27, HSP40, HSP60, HSP70, HSP90, HSP105, in normal conditions (not significant for all genes) (Fig. C-9). This result is also consistent with impaired upregulation of nascent transcript signal after inhibition of the exonuclease activity of Xrn2 (Fig. 5-36). Hence, taken together, these results suggest that Xrn2 and TTF2 are not only involved in termination of pol II transcription at 3’ ends of genes but are also required for upregulation of the heat shock response. Taken together, these results suggest that Xrn2 and TTF2 are important pol II termination factors that can impact gene expression and maintenance of cellular homeostasis. It remains to be elucidated if the defects observed in
upregulation of the heat shock response in the experiments above are also due to depletion of TTF2 or just Xrn2 or both.
Figure C-3. Pol II ChIP-seq distribution profiles after inducible knockdown of Xrn2 or TTF2
HEK293 cells were treated with doxycycline (48 hrs) to individually express an shRNA against
Xrn2 (shXrn2) or an shRNA against TTF2 (shTTF2). A scrambled shRNA (shScr) sequence was
used as control. (A) pol II ChIP-seq distribution after knockdown of Xrn2 or TTF2 on the GAPDH
and the ACTB genes. UCSC genome browser screen shots. (B) Meta-gene plots of pol II ChIP-
seq density on most active genes (top 25% genes with highest pol II density within the body region
in HEK293 cells in normal conditions). Genes were aligned at the transcription start site (TSS)
and the cleavage and polyadenylation site (CPS). The body region was defined as +0.5 kb
downstream of the TSS and -0.5 kb upstream of the CPS and normalized to gene length.
Figure C-4. Exposure to CdSO₄ induces transcriptional activation of the MT2A gene
HEK293 cells expressing two shRNAs targeting Xrn2 and TTF2 (shXrn2+shTTF2) or a scrambled shRNA (shScr) control were treated with CdSO₄ (50 μM for 4 hrs) to induce expression of heat shock genes. Shown is pol II ChIP-seq density on the heavy-metal responsive metallothionein MT2A gene. UCSC genome browser screen shot. Blue arrow indicates the direction of transcription.
Figure C-5. Exposure to CdSO₄ generates a strong termination defect on the HSP90AA1 gene after double knockdown of Xrn2 and TTF2

Stable shXrn2+shTTF2 or scrambled shRNA (shScr) control cells were treated with CdSO₄ (50 µM for 4 hours). Shown is pol II ChIP-seq density on the HSP90AA1 (HSP90) gene before and after exposure to CdSO₄. Note stronger accumulation of pol II density downstream of the polyA site (marked by the red dotted line) after exposure to CdSO₄ in shXrn2+shTTF2 cells when compared to control shScr cells (indicated by the red arrow). UCSC genome browser screenshot. Blue arrow indicates direction of transcription.
Figure C-6. Exposure to CdSO₄ generates a strong termination defect on the DNAJB1 gene after double knockdown of Xrn2 and TTF2

Stable shXrn2+shTTF2 or scrambled shRNA (shScr) control cells were treated with CdSO₄ (50 µM for 4 hours). Shown is pol II ChIP-seq density on the DNAJB1 (HSP40) gene before and after exposure to CdSO₄. Note stronger accumulation of pol II density downstream of the polyA site (marked by the red dotted line) after exposure to CdSO₄ in shXrn2+shTTF2 cells when compared to control shScr cells (indicated by the red arrow). UCSC genome browser screenshot. Blue arrow indicates direction of transcription.
Figure C-7. Exposure to CdSO₄ does not generate detectable changes in the distribution profile of pol II genome-wide

Shown is pol II ChIP-seq density from HEK293 cells expressing two shRNAs targeting Xrn2 and TTF2 (shXrn2+shTTF2) or a scrambled shRNA (shScr) control before and after exposure of cells to CdSO₄ (50 µM for 4 hours). **Top**: most active genes (top 25% genes with highest pol II density within the body region in HEK293 cells in normal conditions). **Middle**: CdSO₄-responsive genes (top 5% genes with highest pol II density within their body region under CdSO₄ conditions in control shScr cells). **Bottom**: Heat shock responsive genes. These genes were selected on the basis that they present at least two-fold increase in Bruseq coverage within their body region upon heat shock and also rank in the top 10% of genes with highest Bruseq coverage under heat shock conditions.
Figure C-8. Double knockdown of Xrn2 and TTF2 increases the expression levels of heat shock protein genes

Total RNA from stable shXrn2+shTTF2 and scrambled shRNA (shScr) control cells were extracted for analysis of mRNA expression levels of heat shock genes. Cell lines from two independent transductions to knock down both Xrn2 and TTF2 were included for analysis (shXrn2+shTTF2 #1 and shXrn2+shTTF2 #2). RT-qPCR quantification was made relative to GAPDH expression levels. Shown is the fold change in expression levels of HSP27, HSP40, HSP60, HSP70, HSP90 and HSP105 genes in shXrn2+shTTF2 cells relative to shScr cells. Note a trend suggestive of increased mRNA levels of heat shock genes in shXrn2+shTTF2 cells relative to shScr cells. Error bars correspond to the standard error of technical replicates of a single experiment (n=3).
# APPENDIX D

## GENE LISTS

### Heat Shock Induced Genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACRC</td>
<td>DNAJB6</td>
<td>HSP90AA1</td>
<td>MRPS6</td>
<td>SEPW1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACSL3</td>
<td>DUSP7</td>
<td>HSP90AB1</td>
<td>NCKAP5L</td>
<td>SERPINC1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGAP4</td>
<td>DVL2</td>
<td>HSPA4</td>
<td>NEK8</td>
<td>SERTD1B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGAP6</td>
<td>DYRK1B</td>
<td>HSPA4L</td>
<td>NF51</td>
<td>SFI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AHS1</td>
<td>E2F3</td>
<td>HSPD1</td>
<td>NPIPL3</td>
<td>SKI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AHS2</td>
<td>ECH1</td>
<td>HSPE1-MO4</td>
<td>NPTX1</td>
<td>SLC12A2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AKAP5</td>
<td>ELOVL5</td>
<td>HSPH1</td>
<td>NUDC</td>
<td>SLC25A10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALDH1A2</td>
<td>EN2</td>
<td>ID12-AS1</td>
<td>NXT2</td>
<td>SLC29A1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANP32A</td>
<td>ENGASE</td>
<td>IFFO1</td>
<td>PARD6G-AS1</td>
<td>SLC6A8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AOC2</td>
<td>EPAH</td>
<td>IFI35</td>
<td>PHKG2</td>
<td>SLC9A3R1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AOC3</td>
<td>EPPK1</td>
<td>INSIG1</td>
<td>PHLDA1</td>
<td>SNAPC5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APLP2</td>
<td>ERF</td>
<td>INTS6</td>
<td>PIK3R1</td>
<td>SNX3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAG3</td>
<td>EZH2</td>
<td>JAG1</td>
<td>PIK3R3</td>
<td>SOX9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BASP1</td>
<td>FAM46A</td>
<td>JMD6</td>
<td>PLAC8L1</td>
<td>SPRED3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCL2L11</td>
<td>FAM46C</td>
<td>KLF10</td>
<td>PLXNA3</td>
<td>SQLE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCL2L12</td>
<td>FAM72A</td>
<td>LAG3</td>
<td>PNLD1C</td>
<td>SQSTM1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C4orf32</td>
<td>FAM72D</td>
<td>LARP4B</td>
<td>PNPLA2</td>
<td>ST13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C7orf53</td>
<td>FASN</td>
<td>LDB1</td>
<td>POLG2</td>
<td>STIP1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CACYBP</td>
<td>FBR5</td>
<td>LDLR</td>
<td>POMT1</td>
<td>TAOK2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBX4</td>
<td>FKB5</td>
<td>LIN37</td>
<td>PPID</td>
<td>TCP1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCND1</td>
<td>FLNA</td>
<td>LINC00263</td>
<td>PPIEL</td>
<td>TECR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDKN1A</td>
<td>FOSB</td>
<td>LINC00641</td>
<td>PTGES3</td>
<td>TOB2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDKN2B</td>
<td>GDF11</td>
<td>LRF1</td>
<td>RBM38</td>
<td>TRMT61A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHORDC1</td>
<td>GFM2</td>
<td>LSS</td>
<td>RDH10</td>
<td>UBTF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHRNA5</td>
<td>GLS</td>
<td>LTB4R2</td>
<td>RRP12</td>
<td>USP1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLU</td>
<td>GNB3</td>
<td>MAFG</td>
<td>RUSC1</td>
<td>VARS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSK</td>
<td>GRHL1</td>
<td>MAP2K3</td>
<td>RYBP</td>
<td>WLS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEDD2</td>
<td>GU1</td>
<td>MGC27345</td>
<td>SCAMP3</td>
<td>ZBTB11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DGKE</td>
<td>HMGCS1</td>
<td>MIA</td>
<td>SCD</td>
<td>ZFAND2A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNAJA1</td>
<td>HNRNPU-AS1</td>
<td>MKNK2</td>
<td>SCRIB</td>
<td>ZNF324</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNAJB4</td>
<td>HS3ST3B1</td>
<td>MRPL18</td>
<td>SECISBP2</td>
<td>ZNF34</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Heat Shock Repressed Genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAED1</td>
<td>ADIPOR1</td>
<td>ALG10</td>
<td>AP1S2</td>
<td>ARRDC4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AARS2</td>
<td>ADPRM</td>
<td>ALG10B</td>
<td>API</td>
<td>ASAH2B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AASDHPP1</td>
<td>ADSL</td>
<td>ALG11</td>
<td>APOL2</td>
<td>ASB13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABCD4</td>
<td>AEN</td>
<td>ALG12</td>
<td>APOOL</td>
<td>ASB8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABHD10</td>
<td>AFG3L1P</td>
<td>ALG2</td>
<td>APPL2</td>
<td>ASH2L</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABHD13</td>
<td>AFG3L2</td>
<td>AMOT</td>
<td>ARAF</td>
<td>ASNS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACAAA2</td>
<td>AGA</td>
<td>AMZ2</td>
<td>ARF3</td>
<td>ASNDS1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACBD3</td>
<td>AIMP1</td>
<td>AMZ2P1</td>
<td>ARHGEF35</td>
<td>ASPSCR1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACOT2</td>
<td>AIMP2</td>
<td>ANGEL1</td>
<td>ARHGEF39</td>
<td>ASUN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACP2</td>
<td>AIP</td>
<td>ANGEL2</td>
<td>ARL1</td>
<td>ATF5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACTR10</td>
<td>AKIRIN1</td>
<td>ANKRD39</td>
<td>ARL10</td>
<td>ATG14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACTR5</td>
<td>AKR7A2</td>
<td>ANKRD49</td>
<td>ARL14EP</td>
<td>ATG4D</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACTR8</td>
<td>ALDH16A1</td>
<td>ANKRD5</td>
<td>ARL2BP</td>
<td>ATMN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADAMTS1</td>
<td>ALDH1B1</td>
<td>ANKZF1</td>
<td>ARL5B</td>
<td>ATP5A1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADAT2</td>
<td>ALDH5A1</td>
<td>ANP32B</td>
<td>ARM1</td>
<td>ATP5C1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADH5</td>
<td>ALG1</td>
<td>ANXA7</td>
<td>ARMC5</td>
<td>ATP5F1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gene</td>
<td>Chromosome</td>
<td>Description</td>
<td>Enzyme</td>
<td>Gene</td>
<td>Chromosome</td>
<td>Description</td>
<td>Enzyme</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>------------</td>
<td>-------------</td>
<td>--------</td>
<td>-------</td>
<td>------------</td>
<td>-------------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>ATP5G2</td>
<td>C19orf10</td>
<td>CD99P1</td>
<td>COX6C</td>
<td>DMWD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATP5G3</td>
<td>C19orf42</td>
<td>CDAN1</td>
<td>COX7B</td>
<td>DNAJA2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATP5H</td>
<td>C1D</td>
<td>CDC5L</td>
<td>CPT2</td>
<td>DNAJC16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATP5O</td>
<td>C1orf115</td>
<td>CDC7</td>
<td>CRBN</td>
<td>DNAJC18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATP5S</td>
<td>C1orf198</td>
<td>CDCA4</td>
<td>CREB2F</td>
<td>DNAJC19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATP5SL</td>
<td>C1orf220</td>
<td>CDCA7</td>
<td>CREG1</td>
<td>DNAJC27</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATP6AP2</td>
<td>C1orf50</td>
<td>CDC7L</td>
<td>CRNDE</td>
<td>DNASE1L1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATP6V1B2</td>
<td>C20orf111</td>
<td>CDK1</td>
<td>CRTAP</td>
<td>DNASE2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATPAF2</td>
<td>C21orf2</td>
<td>CDK2</td>
<td>CSDE1</td>
<td>DNEP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATRAID</td>
<td>C21orf67</td>
<td>CDKS5RAP3</td>
<td>C5R1</td>
<td>DP1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATRIP</td>
<td>C22orf13</td>
<td>CDKN2AIPNL</td>
<td>CSTF2</td>
<td>DPH1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AXN10</td>
<td>C22orf28</td>
<td>CDKN2C</td>
<td>CTBS</td>
<td>DPY30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATXN3</td>
<td>C2orf29</td>
<td>CDKN3</td>
<td>CTC1</td>
<td>DRG1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATXN7L2</td>
<td>C2orf47</td>
<td>CDX2</td>
<td>CTPS1</td>
<td>DSC3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUNIP</td>
<td>C2orf68</td>
<td>CEBPG</td>
<td>CTSL1</td>
<td>DTDW1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B2M</td>
<td>C3orf17</td>
<td>CENPV</td>
<td>CUEDC2</td>
<td>DTYMK</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B4GALT3</td>
<td>C3orf19</td>
<td>CEP68</td>
<td>CUL4B</td>
<td>DUS3L</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B9DI</td>
<td>C3orf33</td>
<td>CGRF1</td>
<td>CXorf26</td>
<td>DUS4L</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAK1</td>
<td>C5orf51</td>
<td>CHAMP1</td>
<td>CXorf56</td>
<td>DUSP12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BBS2</td>
<td>C6orf130</td>
<td>CHCHD7</td>
<td>CYCS</td>
<td>DUSP3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BBS5</td>
<td>C6orf162</td>
<td>CHD1L</td>
<td>CYP2U1</td>
<td>DYNLT1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCKDHA</td>
<td>C6orf57</td>
<td>CHMP1A</td>
<td>DAP</td>
<td>DYNLT3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCL7B</td>
<td>C6orf89</td>
<td>CHMP2B</td>
<td>DBR1</td>
<td>E2F7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCL7C</td>
<td>C7orf73</td>
<td>CHMP4C</td>
<td>DCAF11</td>
<td>ECHS1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCRP2</td>
<td>C8orf40</td>
<td>CHMP7</td>
<td>DCAF15</td>
<td>EC12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BLOC1S2</td>
<td>C9orf156</td>
<td>CHRFAM7A</td>
<td>DCAF4</td>
<td>EEF1A1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BLOC1S6</td>
<td>C9orf72</td>
<td>CHTOP</td>
<td>DCLRE1A</td>
<td>EFHA1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BLVRA</td>
<td>C9orf78</td>
<td>CIAPIN1</td>
<td>DCTN3</td>
<td>EI24</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMS1P1</td>
<td>CA2</td>
<td>CINP</td>
<td>DCTN4</td>
<td>EIF1AX</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMS1P2</td>
<td>CALCOCO2</td>
<td>CIRBP</td>
<td>DCTN6</td>
<td>EIF1A2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMS1P2</td>
<td>CAMLG</td>
<td>CKA2L</td>
<td>DD1</td>
<td>EIF2A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMS1P5</td>
<td>CASC3</td>
<td>CKA4</td>
<td>DHD2</td>
<td>EIF2B4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BNI1P1</td>
<td>CBX3</td>
<td>CLCC1</td>
<td>DDI2</td>
<td>EIF2C4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOLA3</td>
<td>CC2D1B</td>
<td>CLN5</td>
<td>DDO1</td>
<td>EIF2S2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BRAT1</td>
<td>CCBL2</td>
<td>CLPX</td>
<td>DDX19B</td>
<td>EIF2S3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BRMS1L</td>
<td>CCDC127</td>
<td>CMAS</td>
<td>DDX20</td>
<td>EIF3D</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BTF3</td>
<td>CCDC167</td>
<td>CDN2P</td>
<td>DDX47</td>
<td>EIF3F</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BTF3L4</td>
<td>CCDC23</td>
<td>CNHI</td>
<td>DD4</td>
<td>EIF4A2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BTG1</td>
<td>CCDC25</td>
<td>CNHI4</td>
<td>DD5</td>
<td>EFK4B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BUB3</td>
<td>CCDC28A</td>
<td>CNOT7</td>
<td>DDH5</td>
<td>EFK4BP1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BUD13</td>
<td>CCDC43</td>
<td>COG1</td>
<td>DDH5</td>
<td>EFK4G2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BZW1</td>
<td>CCDC59</td>
<td>COG8</td>
<td>DEDD</td>
<td>EFK5A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1orf57</td>
<td>CCDC9</td>
<td>COIL</td>
<td>DEM1</td>
<td>ELAC1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C12orf32</td>
<td>CCDC94</td>
<td>COMM2</td>
<td>DENR</td>
<td>ELP6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C12orf43</td>
<td>CCDC99</td>
<td>COMM6</td>
<td>DEPPDC1</td>
<td>EMC10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C12orf45</td>
<td>CCNA2</td>
<td>COMM8</td>
<td>DERL2</td>
<td>EMC2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C12orf76</td>
<td>CCNB1</td>
<td>COMM9</td>
<td>DHOHD</td>
<td>EMC4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C14orf126</td>
<td>CCNB1P1</td>
<td>COP53</td>
<td>DHPS</td>
<td>EMC7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C14orf166</td>
<td>CCNB2</td>
<td>COP54</td>
<td>DHR5</td>
<td>EME1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C14orf28</td>
<td>CCNDPBP1</td>
<td>COP58</td>
<td>DHR5-SAS1</td>
<td>EMG1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C15orf61</td>
<td>CCNG1</td>
<td>COQ10B</td>
<td>DIX16</td>
<td>ENO1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C16orf87</td>
<td>CCNJ</td>
<td>COQ2</td>
<td>DIX35</td>
<td>ENTPD4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C17orf75</td>
<td>CD47</td>
<td>COX14</td>
<td>DIXF</td>
<td>EPRS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C17orf79</td>
<td>CD99</td>
<td>COX20</td>
<td>DIS3L</td>
<td>ERAL1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C17orf85</td>
<td>CD99</td>
<td>COX411</td>
<td>DLD</td>
<td>ERCC1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C18orf21</td>
<td>CD99P1</td>
<td>COX5A</td>
<td>DMAP1</td>
<td>ERCC6L2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

286
<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Gene Symbol</th>
<th>Gene Symbol</th>
<th>Gene Symbol</th>
<th>Gene Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERH</td>
<td>FDXACB1</td>
<td>GRK6</td>
<td>ICT1</td>
<td>LGALS8</td>
</tr>
<tr>
<td>ERLC1</td>
<td>FEM1B</td>
<td>GRN</td>
<td>IDH3B</td>
<td>LGALSL</td>
</tr>
<tr>
<td>ERLN2</td>
<td>FEZ2</td>
<td>GRPEL1</td>
<td>IDH3G</td>
<td>LIG3</td>
</tr>
<tr>
<td>ERMAP</td>
<td>FEZF1-AS1</td>
<td>GRPEL2</td>
<td>IER3IP1</td>
<td>LIPT1</td>
</tr>
<tr>
<td>ESD</td>
<td>FH</td>
<td>GRWD1</td>
<td>IFIT5</td>
<td>LMAN1</td>
</tr>
<tr>
<td>ETAA1</td>
<td>FIZ1</td>
<td>GSTA4</td>
<td>IFNGR1</td>
<td>LRR1</td>
</tr>
<tr>
<td>ETFB</td>
<td>FKRP</td>
<td>GSTK1</td>
<td>IFT46</td>
<td>LRRC42</td>
</tr>
<tr>
<td>EXOC8</td>
<td>FLJ10661</td>
<td>GTF101</td>
<td>IKZF5</td>
<td>LRRC8E</td>
</tr>
<tr>
<td>EXOG</td>
<td>FLJ46906</td>
<td>GTF2B</td>
<td>IL12A</td>
<td>LSM12</td>
</tr>
<tr>
<td>EXOSC2</td>
<td>FLYWCH2</td>
<td>GTF3C6</td>
<td>ILF2</td>
<td>LSM3</td>
</tr>
<tr>
<td>EXOSC3</td>
<td>FN3KRP</td>
<td>GTPBP1</td>
<td>ILVBL</td>
<td>LSM6</td>
</tr>
<tr>
<td>EXOSC8</td>
<td>FNTA</td>
<td>GTPBP8</td>
<td>IMMT</td>
<td>LTV1</td>
</tr>
<tr>
<td>EXOSC9</td>
<td>FOPNL</td>
<td>GYGI</td>
<td>IMPACT</td>
<td>LYRM5</td>
</tr>
<tr>
<td>EZH1</td>
<td>FRG1</td>
<td>HAUS1</td>
<td>INO80C</td>
<td>LYSMD4</td>
</tr>
<tr>
<td>FAH2D2A</td>
<td>FRG1B</td>
<td>HAUS2</td>
<td>IRAK1BP1</td>
<td>MAD2L1</td>
</tr>
<tr>
<td>FAH2D2B</td>
<td>FRS3</td>
<td>HAUS4</td>
<td>IREB2</td>
<td>MAD2L1BP</td>
</tr>
<tr>
<td>FAH2DCP</td>
<td>FTSJD2</td>
<td>HAUS5</td>
<td>ISCA1</td>
<td>MAG2-AS3</td>
</tr>
<tr>
<td>FAIM</td>
<td>FUBP1</td>
<td>HAUS8</td>
<td>ISCU</td>
<td>MAGOHB</td>
</tr>
<tr>
<td>FAM101B</td>
<td>FUCAL2</td>
<td>HBPI</td>
<td>ISOC2</td>
<td>MAN2C1</td>
</tr>
<tr>
<td>FAM111A</td>
<td>FUNDC1</td>
<td>HBXIP</td>
<td>ITFG2</td>
<td>MAP7D3</td>
</tr>
<tr>
<td>FAM115A</td>
<td>FUT1</td>
<td>HCCS</td>
<td>ITM2B</td>
<td>MAPK1P1L</td>
</tr>
<tr>
<td>FAM156A</td>
<td>FUT4</td>
<td>HCFC1</td>
<td>ITPR1P2</td>
<td>MAPK3</td>
</tr>
<tr>
<td>FAM157A</td>
<td>FXR2</td>
<td>HCFC2</td>
<td>JAKAMP</td>
<td>MAPK6</td>
</tr>
<tr>
<td>FAM157B</td>
<td>FXY5D</td>
<td>HCG11</td>
<td>KANSL2</td>
<td>MARVELD3</td>
</tr>
<tr>
<td>FAM160A2</td>
<td>FZD3</td>
<td>HDDC2</td>
<td>KARS</td>
<td>MAT2B</td>
</tr>
<tr>
<td>FAM160B1</td>
<td>FZD6</td>
<td>HEATR6</td>
<td>KAT5</td>
<td>MBIP</td>
</tr>
<tr>
<td>FAM174A</td>
<td>G6PC3</td>
<td>HEATR8-</td>
<td>KAT7</td>
<td>MBLAC2</td>
</tr>
<tr>
<td>FAM188B1</td>
<td>GABARAP1L1</td>
<td>TTC4</td>
<td>KCNG1</td>
<td>MBTPS2</td>
</tr>
<tr>
<td>FAM200B</td>
<td>GALNT3</td>
<td>HEBP1</td>
<td>KCNJ8</td>
<td>MCM6</td>
</tr>
<tr>
<td>FAM210B</td>
<td>GARS</td>
<td>HEBP2</td>
<td>KDEL5C</td>
<td>MDH2</td>
</tr>
<tr>
<td>FAM217B</td>
<td>GATAD1</td>
<td>HEXA</td>
<td>KDM8</td>
<td>MDM2</td>
</tr>
<tr>
<td>FAM21A</td>
<td>GBA2</td>
<td>HEXB</td>
<td>KEAPI</td>
<td>MEF6</td>
</tr>
<tr>
<td>FAM21C</td>
<td>GBAS</td>
<td>HINT1</td>
<td>KIAA0141</td>
<td>MED1</td>
</tr>
<tr>
<td>FAM221A</td>
<td>GCA</td>
<td>HINT3</td>
<td>KIAA0195</td>
<td>MED19</td>
</tr>
<tr>
<td>FAM35B</td>
<td>GCBSH</td>
<td>HLA-DQB1</td>
<td>KIAA1143</td>
<td>MED21</td>
</tr>
<tr>
<td>FAM35B2</td>
<td>GDAP1</td>
<td>HMBG1</td>
<td>KIAA1147</td>
<td>MED28</td>
</tr>
<tr>
<td>FAM40A</td>
<td>GDPG1</td>
<td>HMBG3</td>
<td>KIAA1586</td>
<td>MED30</td>
</tr>
<tr>
<td>FAM45A</td>
<td>GEMIN2</td>
<td>HMCL</td>
<td>KIF22</td>
<td>MESC2</td>
</tr>
<tr>
<td>FAM63A</td>
<td>GEMIN5</td>
<td>HMGN4</td>
<td>KIF23</td>
<td>METAP2</td>
</tr>
<tr>
<td>FAM64A</td>
<td>GEMIN8</td>
<td>HMGXB4</td>
<td>KIF3A</td>
<td>METTL17</td>
</tr>
<tr>
<td>FAM76A</td>
<td>GGA3</td>
<td>HNRNPA1L2</td>
<td>KLC1</td>
<td>METTL21D</td>
</tr>
<tr>
<td>FAM86A</td>
<td>GGCT</td>
<td>HNRNPA3</td>
<td>KLHCDC2</td>
<td>MFF</td>
</tr>
<tr>
<td>FAM86B1</td>
<td>GINM1</td>
<td>HNRNPAB</td>
<td>KLHL12</td>
<td>MFS1</td>
</tr>
<tr>
<td>FAM86B2</td>
<td>GLTSCR2</td>
<td>HNRNPK</td>
<td>KLHL36</td>
<td>MFS1</td>
</tr>
<tr>
<td>FAM86DP</td>
<td>GMFB</td>
<td>HOXA9</td>
<td>KPN2</td>
<td>MIER2</td>
</tr>
<tr>
<td>FAM86FP</td>
<td>GNBL21</td>
<td>HPS1</td>
<td>KRR1</td>
<td>MINOS1</td>
</tr>
<tr>
<td>FAM98B</td>
<td>GNPD1A</td>
<td>HPS4</td>
<td>LACTB2</td>
<td>MKS1</td>
</tr>
<tr>
<td>FAN1</td>
<td>GNPNAT1</td>
<td>HRASLS</td>
<td>LANCL1</td>
<td>MLF11P</td>
</tr>
<tr>
<td>FAS</td>
<td>GOLT1B</td>
<td>HSDL1</td>
<td>LARP6</td>
<td>MLX</td>
</tr>
<tr>
<td>FASTKD2</td>
<td>GORASP2</td>
<td>HSPA13</td>
<td>LCP1</td>
<td>MLYCD</td>
</tr>
<tr>
<td>FBXL12</td>
<td>GOT1</td>
<td>HTATSF1</td>
<td>LDHA</td>
<td>MMD</td>
</tr>
<tr>
<td>F BX21</td>
<td>GOT2</td>
<td>HUS1</td>
<td>LDHB</td>
<td>MMGT1</td>
</tr>
<tr>
<td>FBOXO21</td>
<td>GPAM</td>
<td>HYLS1</td>
<td>LDOC1L</td>
<td>MMS19</td>
</tr>
<tr>
<td>FBOXO3</td>
<td>GPATCH3</td>
<td>IAH1</td>
<td>LEMD2</td>
<td>MNF1</td>
</tr>
<tr>
<td>FBOXO7</td>
<td>GPR180</td>
<td>IBA57</td>
<td>LEO1</td>
<td>MOB1A</td>
</tr>
<tr>
<td>FDX1L</td>
<td>GPX8</td>
<td>ICMT</td>
<td>LETMD1</td>
<td>MOCS2</td>
</tr>
<tr>
<td>Gene</td>
<td>Gene</td>
<td>Gene</td>
<td>Gene</td>
<td>Gene</td>
</tr>
<tr>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>MORN2</td>
<td>NDUFA6</td>
<td>P2RX4</td>
<td>PMPCB</td>
<td>RAB39B</td>
</tr>
<tr>
<td>MPI</td>
<td>NDUFA7</td>
<td>PABPC1</td>
<td>PMS2P1</td>
<td>RAB9A</td>
</tr>
<tr>
<td>MRPL10</td>
<td>NDUFA9</td>
<td>PABPC4</td>
<td>PMS2P3</td>
<td>RABGGTA</td>
</tr>
<tr>
<td>MRPL15</td>
<td>NDUFAF5</td>
<td>PAICS</td>
<td>PNN</td>
<td>RABGGTB</td>
</tr>
<tr>
<td>MRPL19</td>
<td>NDUFB8</td>
<td>PAIP1</td>
<td>POFUT1</td>
<td>RAD1</td>
</tr>
<tr>
<td>MRPL2</td>
<td>NDUFB9</td>
<td>PALB2</td>
<td>POLE2</td>
<td>RAD21</td>
</tr>
<tr>
<td>MRPL32</td>
<td>NDUFS2</td>
<td>PANK4</td>
<td>POLE4</td>
<td>RBBP7</td>
</tr>
<tr>
<td>MRPL35</td>
<td>NDUFS5</td>
<td>PARP2</td>
<td>POLM</td>
<td>RBRP9</td>
</tr>
<tr>
<td>MRPL39</td>
<td>NDUFV1</td>
<td>PBK</td>
<td>POLR2D</td>
<td>RBAF</td>
</tr>
<tr>
<td>MRPL4</td>
<td>NDUFV2</td>
<td>PCCB</td>
<td>POLR2G</td>
<td>RBL2</td>
</tr>
<tr>
<td>MRPL42</td>
<td>NDUFV3</td>
<td>PCIF1</td>
<td>POLR3D</td>
<td>RBM11</td>
</tr>
<tr>
<td>MRPL44</td>
<td>NECAP1</td>
<td>PCK2</td>
<td>PON2</td>
<td>RBM18</td>
</tr>
<tr>
<td>MRPL46</td>
<td>NEFL</td>
<td>PCMTD2</td>
<td>PPA1</td>
<td>RBM39</td>
</tr>
<tr>
<td>MRPL49</td>
<td>NEK3</td>
<td>PCNP</td>
<td>PPIA</td>
<td>RBM45</td>
</tr>
<tr>
<td>MRPL52</td>
<td>NENF</td>
<td>PDCD5</td>
<td>PPIL1</td>
<td>RBM7</td>
</tr>
<tr>
<td>MRPS10</td>
<td>NHLRC2</td>
<td>PDCL</td>
<td>PPIL3</td>
<td>RBM8A</td>
</tr>
<tr>
<td>MRPS14</td>
<td>NIPSnap3A</td>
<td>PDE12</td>
<td>PP1R8</td>
<td>RBMX</td>
</tr>
<tr>
<td>MRPS18C</td>
<td>NT2</td>
<td>PDCL3</td>
<td>PPT1</td>
<td>RBX1</td>
</tr>
<tr>
<td>MRPS22</td>
<td>NKAP</td>
<td>PDHAI</td>
<td>PPTC7</td>
<td>RC3H2</td>
</tr>
<tr>
<td>MRPS30</td>
<td>NKAPP1</td>
<td>PDHB</td>
<td>PQBP1</td>
<td>RCBT1</td>
</tr>
<tr>
<td>MRPS35</td>
<td>NLN</td>
<td>PDIK1L</td>
<td>PQLC2</td>
<td>RCHY1</td>
</tr>
<tr>
<td>MSH5</td>
<td>NMD3</td>
<td>PDK1</td>
<td>PRDM4</td>
<td>RCN2</td>
</tr>
<tr>
<td>MSL3</td>
<td>NOB1</td>
<td>PDPR</td>
<td>PRDX1</td>
<td>RELL1</td>
</tr>
<tr>
<td>MSX2</td>
<td>NOL11</td>
<td>PEBP1</td>
<td>PRDX3</td>
<td>RFC5</td>
</tr>
<tr>
<td>MTCP1NB</td>
<td>NOL9</td>
<td>PEX1</td>
<td>PRDX4</td>
<td>RFE5D</td>
</tr>
<tr>
<td>MTERFD1</td>
<td>NONO</td>
<td>PEX10</td>
<td>PRKRA</td>
<td>RFT1</td>
</tr>
<tr>
<td>MTFMT</td>
<td>NOP56</td>
<td>PEX11A</td>
<td>PRMT1</td>
<td>RFWD3</td>
</tr>
<tr>
<td>MTHFD2</td>
<td>NOP9</td>
<td>PEX11B</td>
<td>PROK2</td>
<td>RHBDD2</td>
</tr>
<tr>
<td>MTMR6</td>
<td>NPM1</td>
<td>PEX13</td>
<td>PRPS1</td>
<td>RINT1</td>
</tr>
<tr>
<td>MTPAP</td>
<td>NR1D2</td>
<td>PEX19</td>
<td>PRSS23</td>
<td>RIOK2</td>
</tr>
<tr>
<td>MTRF1L</td>
<td>NRBF2</td>
<td>PEX2</td>
<td>PSG1</td>
<td>RIPK2</td>
</tr>
<tr>
<td>MTX3</td>
<td>NRNS2</td>
<td>PEX26</td>
<td>PSG4</td>
<td>RLIM</td>
</tr>
<tr>
<td>MUT</td>
<td>NSFL1C</td>
<td>PFDN1</td>
<td>PSG7</td>
<td>RM11</td>
</tr>
<tr>
<td>MYADM</td>
<td>NSFP1</td>
<td>PFDN4</td>
<td>PSKH1</td>
<td>RM12</td>
</tr>
<tr>
<td>MYC</td>
<td>NSMCE4A</td>
<td>PFMN2</td>
<td>PSMB2</td>
<td>RNASEH1</td>
</tr>
<tr>
<td>MYEF2</td>
<td>NSUN3</td>
<td>PGAM1</td>
<td>PSMC3</td>
<td>RNF139</td>
</tr>
<tr>
<td>MYL12A</td>
<td>NSUN5</td>
<td>PGP3</td>
<td>PSMC6</td>
<td>RNF14</td>
</tr>
<tr>
<td>MYSM1</td>
<td>NTSC3L</td>
<td>PGM2</td>
<td>PSMD10</td>
<td>RNF141</td>
</tr>
<tr>
<td>MZT1</td>
<td>NUCKS1</td>
<td>PHAX</td>
<td>PSMD9</td>
<td>RNF146</td>
</tr>
<tr>
<td>N4BP1</td>
<td>NUCCD2</td>
<td>PHC2</td>
<td>PSMG1</td>
<td>RNF25</td>
</tr>
<tr>
<td>N6AMT1</td>
<td>NUDDT2</td>
<td>PHF19</td>
<td>PSTK</td>
<td>RNF6</td>
</tr>
<tr>
<td>NAA38</td>
<td>NUDDT15</td>
<td>PHF5A</td>
<td>PTAR1</td>
<td>RNF7</td>
</tr>
<tr>
<td>NAEL1</td>
<td>NUDDT16</td>
<td>PIK2B</td>
<td>PTD2</td>
<td>RNF7</td>
</tr>
<tr>
<td>NAGLU</td>
<td>NUDDT19</td>
<td>PIAG</td>
<td>PTGER4</td>
<td>RNT1</td>
</tr>
<tr>
<td>NANGD</td>
<td>NUDDT19</td>
<td>PIAG</td>
<td>PTGER4</td>
<td>RNT1</td>
</tr>
<tr>
<td>NAPN</td>
<td>NUFD2</td>
<td>PIGH</td>
<td>PTPM1</td>
<td>RNPEP</td>
</tr>
<tr>
<td>NARS</td>
<td>NUFI</td>
<td>PIGO</td>
<td>PTS</td>
<td>RP9</td>
</tr>
<tr>
<td>NBN</td>
<td>NUFI</td>
<td>PINK1</td>
<td>PTTG1</td>
<td>RP1F</td>
</tr>
<tr>
<td>NCL</td>
<td>NUFI</td>
<td>PIP4K2C</td>
<td>PWP1</td>
<td>RPRG</td>
</tr>
<tr>
<td>NCOA4</td>
<td>OCIAD2</td>
<td>PJA2</td>
<td>PYROX1D1</td>
<td>RPL14</td>
</tr>
<tr>
<td>NCOA5</td>
<td>OGFRL1</td>
<td>PKN1</td>
<td>QDPR</td>
<td>RPL15</td>
</tr>
<tr>
<td>NCSTN</td>
<td>OIP5-A51</td>
<td>PLAA</td>
<td>QRS1</td>
<td>RPL21</td>
</tr>
<tr>
<td>NDRG1</td>
<td>ORC6</td>
<td>PLD6</td>
<td>QSOX2</td>
<td>RPL22</td>
</tr>
<tr>
<td>NDRG2</td>
<td>OS9</td>
<td>PLK2</td>
<td>RAB11B</td>
<td>RPL22L1</td>
</tr>
<tr>
<td>NDUFA12</td>
<td>OSGIN2</td>
<td>PLK4</td>
<td>RAB22A</td>
<td>RPL26</td>
</tr>
<tr>
<td>NDUFA13</td>
<td>OTUD6B</td>
<td>PLSCR3</td>
<td>RAB2B</td>
<td>RPL26L1</td>
</tr>
<tr>
<td>NDUFA5</td>
<td>OXA1L</td>
<td>PM20D2</td>
<td>RAB33B</td>
<td>RPL27A</td>
</tr>
<tr>
<td>Gene</td>
<td>Description</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>--------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPL3</td>
<td>Ribosomal Protein L3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPL36A</td>
<td>Ribosomal Protein L36A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPL38</td>
<td>Ribosomal Protein L38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPL39</td>
<td>Ribosomal Protein L39</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPL4</td>
<td>Ribosomal Protein L4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPL5</td>
<td>Ribosomal Protein L5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPN1</td>
<td>Ribosomal Protein N1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPP38</td>
<td>Ribosomal Protein P38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPP40</td>
<td>Ribosomal Protein P40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPS15A</td>
<td>Ribosomal Protein S15A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPS19</td>
<td>Ribosomal Protein S19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPS23</td>
<td>Ribosomal Protein S23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPS3A</td>
<td>Ribosomal Protein S3A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPS5</td>
<td>Ribosomal Protein S5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPS9</td>
<td>Ribosomal Protein S9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPSA</td>
<td>Ribosomal Protein S4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPUSD4</td>
<td>Ribosomal Protein UPUSD4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RRM1</td>
<td>Ribosomal Binding Protein M1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RN3</td>
<td>Ribosomal Nucleic Acid</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNP3P1</td>
<td>Ribosomal Protein P3P1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RN3P2</td>
<td>Ribosomal Protein P3P2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RUNDC1</td>
<td>Ribosomal Nuclear Protein C1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RWDD1</td>
<td>Ribosomal WDD1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S100A13</td>
<td>S100 Family Member A13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS2</td>
<td>Sarcoma</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAYSD1</td>
<td>Say's Disease 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SBDs</td>
<td>Sarcoidosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCML1</td>
<td>Smooth Muscle Actin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCO1</td>
<td>Smooth Muscle Actin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCYL3</td>
<td>Smooth Muscle Actin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEC22C</td>
<td>SEC22C Protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEC61B</td>
<td>SEC61B Protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEC62</td>
<td>SEC62 Protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEH1L</td>
<td>SEH1L Protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SENP2</td>
<td>Senescence</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Top 5% of genes with highest pol II density within the gene body region in CdSO4 conditions

<table>
<thead>
<tr>
<th>AARS</th>
<th>ARHI1</th>
<th>BRIX1</th>
<th>C1orf56</th>
<th>CCT2</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABHD14A-</td>
<td>ARL6IP1</td>
<td>BRWD1-IT2</td>
<td>C1orf63</td>
<td>CCT3</td>
</tr>
<tr>
<td>ACY1</td>
<td>ARRD3C</td>
<td>BTF3</td>
<td>C3orf71</td>
<td>CCT4</td>
</tr>
<tr>
<td>ACADVL</td>
<td>ASH1L</td>
<td>BTF3L4</td>
<td>C5orf28</td>
<td>CCT5</td>
</tr>
<tr>
<td>ACAT2</td>
<td>ASH1L-AS1</td>
<td>BTG1</td>
<td>C5orf34</td>
<td>CCT7</td>
</tr>
<tr>
<td>ACHE</td>
<td>ATF4</td>
<td>BTG2</td>
<td>C5orf51</td>
<td>CCT8</td>
</tr>
<tr>
<td>ACPT</td>
<td>ATF5</td>
<td>BYSL</td>
<td>C6orf1</td>
<td>CD3EAP</td>
</tr>
<tr>
<td>ACTB</td>
<td>ATN1</td>
<td>C10orf2</td>
<td>C6orf120</td>
<td>CD68</td>
</tr>
<tr>
<td>ACTG1</td>
<td>ATP5B</td>
<td>C11orf83</td>
<td>C6orf48</td>
<td>CDC42E1</td>
</tr>
<tr>
<td>ACTN4</td>
<td>ATP5J2-</td>
<td>C12orf57</td>
<td>C7orf59</td>
<td>CDC42E1</td>
</tr>
<tr>
<td>ADAMTS1</td>
<td>PTD1</td>
<td>C12orf61</td>
<td>CACTBP</td>
<td>CDC3</td>
</tr>
<tr>
<td>ADCK3</td>
<td>ATP6V1G2-</td>
<td>C12orf65</td>
<td>C4M4</td>
<td>CDK1</td>
</tr>
<tr>
<td>ADM</td>
<td>DDX39B</td>
<td>C12orf73</td>
<td>CALM2</td>
<td>CDK16</td>
</tr>
<tr>
<td>AGXT2L2</td>
<td>ATXN2L</td>
<td>C15orf37</td>
<td>CALR</td>
<td>CDK4</td>
</tr>
<tr>
<td>AKIRIN2</td>
<td>AURKB</td>
<td>C15orf39</td>
<td>CAND2</td>
<td>CDKN1B</td>
</tr>
<tr>
<td>AKT1S1</td>
<td>AXL</td>
<td>C15orf44</td>
<td>CANX</td>
<td>CDT1</td>
</tr>
<tr>
<td>ALDH16A1</td>
<td>B3GALT4</td>
<td>C16orf59</td>
<td>CAPNS1</td>
<td>CEP95</td>
</tr>
<tr>
<td>ALDOA</td>
<td>B4GALT2</td>
<td>C16orf79</td>
<td>CAPRIN1</td>
<td>CETN3</td>
</tr>
<tr>
<td>AMD1</td>
<td>B4GALT7</td>
<td>C17orf76-AS1</td>
<td>CAPZA1</td>
<td>CFL1</td>
</tr>
<tr>
<td>AMH</td>
<td>BAG6</td>
<td>C18orf32</td>
<td>CBX3</td>
<td>CHAC1</td>
</tr>
<tr>
<td>APBB3</td>
<td>BAX</td>
<td>C19orf24</td>
<td>CBX5</td>
<td>CHD4</td>
</tr>
<tr>
<td>APEX1</td>
<td>BCDIN3D-AS1</td>
<td>C19orf35</td>
<td>CBX6</td>
<td>CHMP2A</td>
</tr>
<tr>
<td>APOA1BP</td>
<td>BCLAF1</td>
<td>C19orf48</td>
<td>CCDC106</td>
<td>CHTOP</td>
</tr>
<tr>
<td>APOM</td>
<td>BEST1</td>
<td>C19orf54</td>
<td>CCDC24</td>
<td>CIRBP</td>
</tr>
<tr>
<td>APRT</td>
<td>BEST4</td>
<td>C19orf55</td>
<td>CCDC47</td>
<td>CIRBP-AS1</td>
</tr>
<tr>
<td>ARF1</td>
<td>BOLA2</td>
<td>C1orf213</td>
<td>CCDC84</td>
<td>CITED2</td>
</tr>
<tr>
<td>ARHGDA</td>
<td>BOLA2B</td>
<td>C1orf31</td>
<td>CCDC97</td>
<td>CKAP2</td>
</tr>
<tr>
<td>ARID4B</td>
<td>BRD2</td>
<td>C1orf35</td>
<td>CCNG1</td>
<td>CKB</td>
</tr>
</tbody>
</table>

290
<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Gene Name</th>
<th>Gene Name</th>
<th>Gene Name</th>
<th>Gene Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIN37</td>
<td>MRPS26</td>
<td>PAQR4</td>
<td>PRPF38B</td>
<td>ROMO1</td>
</tr>
<tr>
<td>LINC00263</td>
<td>MRPS34</td>
<td>PCBP1</td>
<td>PRR3</td>
<td>RPL10</td>
</tr>
<tr>
<td>LINC00641</td>
<td>MRPS7</td>
<td>PCBP2</td>
<td>PRR7</td>
<td>RPL10A</td>
</tr>
<tr>
<td>LONP1</td>
<td>MRT04</td>
<td>PCF11</td>
<td>PRR7-AS1</td>
<td>RPL11</td>
</tr>
<tr>
<td>LRRC41</td>
<td>MT1X</td>
<td>PCIF1</td>
<td>PRRC2A</td>
<td>RPL12</td>
</tr>
<tr>
<td>LRSAM1</td>
<td>MTF2</td>
<td>PCNA</td>
<td>PRRG2</td>
<td>RPL13</td>
</tr>
<tr>
<td>LSM14A</td>
<td>MTIF2</td>
<td>PDAP1</td>
<td>PRRT2</td>
<td>RPL13A</td>
</tr>
<tr>
<td>LSM1D</td>
<td>MTRF1L</td>
<td>PDCD2L</td>
<td>PSAT1</td>
<td>RPL14</td>
</tr>
<tr>
<td>LSR</td>
<td>MYBBP1A</td>
<td>PDDC1</td>
<td>PSEnen</td>
<td>RPL15</td>
</tr>
<tr>
<td>LTC4S</td>
<td>MYL6</td>
<td>PFN1</td>
<td>PSG4</td>
<td>RPL18</td>
</tr>
<tr>
<td>LZTR1</td>
<td>MYL6B</td>
<td>PGAM1</td>
<td>PSMB4</td>
<td>RPL18A</td>
</tr>
<tr>
<td>MALAT1</td>
<td>MYLPF</td>
<td>PGP</td>
<td>PSMC4</td>
<td>RPL19</td>
</tr>
<tr>
<td>MAN2C1</td>
<td>MZF1</td>
<td>PHB2</td>
<td>PSMC5</td>
<td>RPL22L1</td>
</tr>
<tr>
<td>MAP3K12</td>
<td>NAA25</td>
<td>PHGDH</td>
<td>PSMD8</td>
<td>RPL23</td>
</tr>
<tr>
<td>MARCKS</td>
<td>NACA</td>
<td>PHIP</td>
<td>PTCH2</td>
<td>RPL23A</td>
</tr>
<tr>
<td>MARCKSL1</td>
<td>NASP</td>
<td>PHKG2</td>
<td>PTGES3</td>
<td>RPL24</td>
</tr>
<tr>
<td>MARS</td>
<td>NCAPD2</td>
<td>PIPT1</td>
<td>PTMA</td>
<td>RPL26</td>
</tr>
<tr>
<td>MAT2A</td>
<td>NCL</td>
<td>PIH1D1</td>
<td>PTOV1</td>
<td>RPL27</td>
</tr>
<tr>
<td>MATR3</td>
<td>NDUFA4L2</td>
<td>PILRB</td>
<td>PTRH2</td>
<td>RPL27A</td>
</tr>
<tr>
<td>MAZ</td>
<td>NDUFA7</td>
<td>PIM3</td>
<td>PUS1</td>
<td>RPL28</td>
</tr>
<tr>
<td>MBD3</td>
<td>NDUFAF3</td>
<td>PMK</td>
<td>PUSL1</td>
<td>RPL29</td>
</tr>
<tr>
<td>MBLAC1</td>
<td>NDUFB10</td>
<td>PKMYT1</td>
<td>PXN-AS1</td>
<td>RPL3</td>
</tr>
<tr>
<td>MCCD1</td>
<td>NEAT1</td>
<td>PKN3</td>
<td>PYCR2</td>
<td>RPL30</td>
</tr>
<tr>
<td>MCL1</td>
<td>NEURL4</td>
<td>PLD3</td>
<td>QRICH1</td>
<td>RPL31</td>
</tr>
<tr>
<td>MCM3</td>
<td>NFE2L3</td>
<td>PLEKHG2</td>
<td>RAB11A</td>
<td>RPL32</td>
</tr>
<tr>
<td>MCM7</td>
<td>NFKBIE</td>
<td>PLEKMH1</td>
<td>RAB11FIP3</td>
<td>RPL35</td>
</tr>
<tr>
<td>MEA1</td>
<td>NFS1</td>
<td>PMAIP1</td>
<td>RAB24</td>
<td>RPL35A</td>
</tr>
<tr>
<td>MED22</td>
<td>NHP2</td>
<td>PNK</td>
<td>RAB26</td>
<td>RPL36</td>
</tr>
<tr>
<td>MED29</td>
<td>NKIRAS1</td>
<td>PNN</td>
<td>RABAC1</td>
<td>RPL36A</td>
</tr>
<tr>
<td>MEPE</td>
<td>NME3</td>
<td>PNRC1</td>
<td>RABEpk</td>
<td>RPL36AL</td>
</tr>
<tr>
<td>METTL23</td>
<td>NME4</td>
<td>PNRC2</td>
<td>RABGGTB</td>
<td>RPL37</td>
</tr>
<tr>
<td>MFSD3</td>
<td>NOC2L</td>
<td>POGZ</td>
<td>RAC3</td>
<td>RPL37A</td>
</tr>
<tr>
<td>MGC16275</td>
<td>NOL8</td>
<td>POLDIP3</td>
<td>RAD23A</td>
<td>RPL38</td>
</tr>
<tr>
<td>MGEA5</td>
<td>NOLC1</td>
<td>POLG</td>
<td>RAD54L</td>
<td>RPL39</td>
</tr>
<tr>
<td>MIDD</td>
<td>NOP56</td>
<td>POLR1C</td>
<td>RAN</td>
<td>RPL4</td>
</tr>
<tr>
<td>MLL2</td>
<td>NOP58</td>
<td>POLR2A</td>
<td>RANBP1</td>
<td>RPL41</td>
</tr>
<tr>
<td>MLL4</td>
<td>NPM1</td>
<td>POLR2I</td>
<td>RANGRF</td>
<td>RPL5</td>
</tr>
<tr>
<td>MLL5</td>
<td>NRARP</td>
<td>POLR2J</td>
<td>RAVER1</td>
<td>RPL6</td>
</tr>
<tr>
<td>MLLT11</td>
<td>NT5C</td>
<td>POLR2J2</td>
<td>RBM14</td>
<td>RPL7</td>
</tr>
<tr>
<td>MLST8</td>
<td>NTMT1</td>
<td>POLR2J3</td>
<td>RBM15</td>
<td>RPL7A</td>
</tr>
<tr>
<td>MMACHC</td>
<td>NUCKS1</td>
<td>POLR2L</td>
<td>RBM34</td>
<td>RPL7L1</td>
</tr>
<tr>
<td>MNF1</td>
<td>NUDCD2</td>
<td>POLR3H</td>
<td>RBM39</td>
<td>RPL8</td>
</tr>
<tr>
<td>MPDU1</td>
<td>NUDT19</td>
<td>POP4</td>
<td>RBM42</td>
<td>RPL9</td>
</tr>
<tr>
<td>MPI</td>
<td>NUDT4</td>
<td>PPAN-P2RY1</td>
<td>RBM48A</td>
<td>RPL9</td>
</tr>
<tr>
<td>MR1I</td>
<td>NUFlP2</td>
<td>PP1A</td>
<td>RBMX</td>
<td>RPL10</td>
</tr>
<tr>
<td>MRPL12</td>
<td>NUP133</td>
<td>PP1B</td>
<td>RCC2</td>
<td>RPL12</td>
</tr>
<tr>
<td>MRPL18</td>
<td>NUP210L</td>
<td>PPM1G</td>
<td>RCCD1</td>
<td>RPRD2</td>
</tr>
<tr>
<td>MRPL20</td>
<td>NUP85</td>
<td>PPP1R10</td>
<td>RFC4</td>
<td>RPS11</td>
</tr>
<tr>
<td>MRPL24</td>
<td>NXF1</td>
<td>PPP1R15A</td>
<td>RGL2</td>
<td>RPS12</td>
</tr>
<tr>
<td>MRPL36</td>
<td>OAZ1</td>
<td>PPP1R35</td>
<td>RGS16</td>
<td>RPS13</td>
</tr>
<tr>
<td>MRPL51</td>
<td>ORC3</td>
<td>PPP2R5D</td>
<td>RHBDL1</td>
<td>RPS14</td>
</tr>
<tr>
<td>MRPL55</td>
<td>PA2G4</td>
<td>PPRC1</td>
<td>RHOA</td>
<td>RPS15</td>
</tr>
<tr>
<td>MRPS12</td>
<td>PABPC1</td>
<td>PRC1</td>
<td>RHOB</td>
<td>RPS15A</td>
</tr>
<tr>
<td>MRPS14</td>
<td>PABPC4</td>
<td>PRDX1</td>
<td>RNF167</td>
<td>RPS16</td>
</tr>
<tr>
<td>MRPS16</td>
<td>PAF1</td>
<td>PRDX2</td>
<td>RNF40</td>
<td>RPS17</td>
</tr>
<tr>
<td>MRPS18B</td>
<td>PAIP2</td>
<td>PRDX3</td>
<td>RNPS1</td>
<td>RPS17L</td>
</tr>
</tbody>
</table>
RPS18  SHMT2  SRSF7  TRA2B  XPO5
RPS19  SIN3A  SRSF9  TRAF7  XPO5
RPS2  SKIV2L  SSBP4  TRIM11  XRC6
RPS21  SLC16A1  ST7-AS1  TRIM28  YARS
RPS23  SLC1A5  STC2  TRIM41  YBX1
RPS24  SLC25A1  STRAP  TRIM52  YDJC
RPS25  SLC25A10  STUB1  TRIP6  YRDC
RPS26  SLC25A3  SUP75H  TRMT2A  YWHAE
RPS27  SLC25A44  SURF1  TRNAU1AP  ZBTB37
RPS27A  SLC29A2  SWSAP1  TROVE2  ZBTB4
RPS3  SLC30A1  SYNCRIP  TRPV2  ZC3H10
RPS3A  SLC35B2  TAF1D  TSIX  ZCWPW1
RPS4X  SLC35G6  TAOK2  TSPAN31  ZDHHC12
RPS5  SLC38A2  TARDBP  TSPY1  ZEB2-AS1
RPS6  SLC39A7  TARS  TTC18  ZFP30
RPS7  SLC3A2  TBCB  TTI2  ZFP36
RPS8  SLC7A5P2  TBB  TUBA1B  ZFP36L2
RPS9  SLFLN1  TCEB2  TUBA1C  ZNF146
RPSA  SMARCC2  TCP1  TUBB  ZNF224
RRM2  SMCR7L  TDG  TUBB4B  ZNF225
RRNAD1  SMG1  TEAD3  TUFM  ZNF226
RSRC2  SMG9  THAP7  TUT1  ZNF227
RUFY1  SNAPC5  THAP8  TXLNA  ZNF228
RUVBL2  SNAPIN  THAP9-AS1  U2AF1L4  ZNF229
RWDD2B  SNHG1  THOC6  UBA2  ZNF230
S100A1  SNHG12  THUMPD3  UBA2P2  ZNF231
SAMD11  SNHG15  TIAL1  UBB  ZNF232
SAMD4B  SNHG3  TICRR  UBC  ZNF233
SAR1A  SNHG5  TIMM13  UBE2M  ZNF234
SARS  SNHG6  TIMM44  UBE2S  ZNF235
SARS2  SNHG7  TIMM8B  UBXN6  ZNF236
SAT1  SNRNP70  TK1  UNCA45A  ZNF237
SCAMP3  SNRP4A  UQCRH  UCN529  ZNF238
SCAND1  SNRPB  UQCRQ  ZNF239
SCD  SNRPD2  TLE3  ZNF546  ZNF240
SDE2  SNX17  TM7SF2  ZNF547  ZNF241
SDHAF1  SNX22  TMBIM6  ZNF548  ZNF242
SEC24C  SON  TMEM107  USMG5  ZNF243
SENP3  SOX4  TMEM147  USP5  ZNF244
EIF4A1  SPG7  TMEM167B  VDAC1  ZNF245
SEPHS2  SPHK2  TMEM55B  VDAC2  ZNF246
SEPT1  SPRTN  TMEM69  VEGFA  ZNF247
SERBP1  SP5B2  TMEM88  VIM  ZNF248
SERHL2  SP5B3  TMPO  VPS29  ZNF249
SERTAD1  SRCAP  TMPO-AS1  VPS52  ZNF250
SERTAD3  SRCP  TMPOA1  WDR24  ZNF251
SESN2  SRPK1  TNFRSF12A  WDR26  ZNF252
SET  SRRM2  TNPO2  WDR38  ZNF253
SETD5  SRRT  TOB2  WDR46  ZNF254
SETD6  SRSF1  TOE1  WDR6  ZNF255
SF3A2  SRSF10  TOMM20  WDR74  ZNF256
SF3B3  SRSF2  TOMM6  WTPA  ZNF257
SFPPQ  SRSF3  TP53  XBP1  ZNF258
SFXN2  SRSF5  TPI1  XIST  ZNF259
SHC1  SRSF6  TPT1  XPO1  ZWILCH