An Investigation of the Regulation of RNA Polymerase II Transcription Termination by SMN, FUS, and TDP-43 in a pathway involving Symmetric Arginine Dimethylation (R1810) of the RNAPII CTD

by

Yanling Zhao

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy in the Graduate Department of Molecular Genetics University of Toronto

© Copyright by Yanling Zhao 2015
An Investigation of the Regulation of RNA Polymerase II Transcription Termination by SMN, FUS, and TDP-43 in a pathway involving Symmetric Arginine Dimethylation (R1810) of the RNAPII CTD

Yanling Zhao
Doctor of Philosophy in the Graduate Department of Molecular Genetics
University of Toronto 2015

Abstract

The C-terminal domain (CTD) of the largest RNA polymerase II (RNAPII) subunit, POLR2A, is a platform for modifications specifying the recruitment of factors that regulate transcription, messenger RNA (mRNA) processing, and chromatin remodeling. I found that a CTD arginine residue (R1810) is symmetrically dimethylated (me2s), allowing it to recruit the Tudor domain of the SMN protein, which is mutated in spinal muscular atrophy (SMA). SMN can oligomerize and interact with Senataxin (SETX), which is sometimes mutated in ALS4 (Amyotrophic Lateral Sclerosis). I showed that CTD R1810me2s and SMN, like SETX, are required for resolving R-loops created by RNAPII in transcription termination regions. A defect in this process is recapitulated in the SMA disease state. FUS and TDP-43 (TARDBP) are DNA/RNA binding proteins that are involved in numerous aspects of gene regulation, and they are known to interact with each other and with SMN. As well, they are sometimes mutated in neurodegenerative diseases such as ALS and FTD (Frontal Temporal Dementia). I was able to show that FUS and TDP-43 act downstream of the CTD R1810me2s-SMN pathway, and that defects in the recruitment of these proteins can influence RNAPII transcription termination through the accumulation of R-loops and also cause DNA damage. I propose that mis-regulation of RNAPII termination could lead to DNA damage, which may contribute to neurodegenerative disorders like SMA and ALS/FTD.
Acknowledgements

I feel tremendously grateful for having Dr. Jack Greenblatt as my supervisor. My graduate training with Jack, one of the best scientific thinkers that I know of, has been the most enriching and enlightening learning experience in my life. I also feel truly thankful for having Dr. Tony Pawson as my co-supervisor, for his extreme kindness in giving me a beginning in the graduate program at the University of Toronto. I feel extremely lucky for having Dr. Ben Blencowe and Dr. Dan Durocher in my thesis committee, for their support and guidance through the years, and for their amazing insights and tastes in science. I am also very grateful for my external thesis member, for taking parts in my defense, in spite of the short notice.

As well, I am indebted to many people who helped me during my studies in the Greenblatt and Pawson labs. I am especially grateful for Dr. Jerry Gish for technical support and scientific discussions, Guoqing Zhong for countless help with cell culture, Shuye Pu for help with sequencing analysis, Dr. Zuyao Ni for discussions, Dr. Karen Colwill, Monika Tucholska, Lorne Taylor, and Hongbo Guo for help with Mass Spectrometry, Dr. Sarang Kulkarni for training in microscopy, the lab managers Joyce Li, Edyta Marcon, Geraldine Mbamalu and the rest of the lab for being wonderfully supportive. I am also thankful for the collaborations from the Blencowe lab especially with help from Dr. Ulrich Braunschweig and Kevin Ha, and the assistance from Dr. Dax Torti and Danica Leung of the CCBR sequencing center. I am also thankful for the helps from several labs at the Structural and Genomic Consortium (Toronto), Dr. Frank Schmitges from the Hughes lab, and Dr. Yue Li from the Zhang lab at U of T. And finally, I’d like to acknowledge Dr. Danny Reinberg, Dr. James Manley, Dr. Dirk Eick, Dr.
Anne-Claude Gingras, Dr. Jason Moffat, Dr. Andres Aguilera, and Dr. Stephen Leppla for sharing the reagents.

Finally, I thank my parents for their unconditional love and support, my gratitude is beyond words! My gratitude for the amazing and lovely folks at U of T who supported and cared about me, directly or indirectly, during my time of grief and sorrow, was also beyond words! Thank you all for giving me the strength to carry on, I really appreciate it.
## Table of Contents

Summary ................................................................................................................................. i

Acknowledgements ............................................................................................................. ii

Table of Contents ................................................................................................................ iii

List of Figures ....................................................................................................................... vi

List of Tables ......................................................................................................................... viii

Abbreviations ....................................................................................................................... ix

### Chapter 1: Introduction .................................................................................................. 1

1.1 The RNAPII transcription machinery and the C-terminal domain (CTD) ................... 1

  1.1.1 The RNAPII CTD Code in transcription regulation ................................................ 2
  1.1.2 Reversible phosphorylation of the RNAPII CTD during the transcription cycle ...... 4
  1.1.3 Asymmetric dimethylation of the RNAPII CTD residue R1810 ............................. 6
  1.1.4 RNAPII promoter-proximal pausing and the switch into elongation ................... 8
  1.1.5 RNAPII CTD regulation of co-transcriptional alternative splicing ....................... 11
  1.1.6 Readers of the RNAPII CTD code ..................................................................... 12
  1.1.7 The roles of chromatin modifications, chromatin structure, and DNA looping in RNAPII transcription ................................................................. 14

1.2 RNAPII termination pathways and the formation of DNA:RNA hybrids (R-loops) .......... 16

  1.2.1 RNAPII termination pathways in metazoans ....................................................... 16
  1.2.2 RNAPII termination downstream of polyadenylation ......................................... 17
  1.2.3 R-loops as a cause of genome instability ............................................................. 19
  1.2.4 The roles of R-loops in transcriptional regulation ............................................... 21
  1.2.5 Surveillance pathways for R-loop suppression ................................................... 22
  1.2.6 Transcription coupled nucleotide excision repair (TC-NER) ............................. 24

1.3 Protein arginine dimethylation ....................................................................................... 27

  1.3.1 Types of arginine methylation and cross-talk involving this mark ....................... 27
  1.3.2 PRMT4/CARM1 in gene regulation .................................................................... 29
  1.3.3 PRMT5 in gene regulation .................................................................................. 30
  1.3.4 Arginine methylation readers: the Tudor domain containing proteins ............... 33
  1.3.5 Methylarginine binding TDRDs and the Arginine demethylase in gene regulation . 34
1.3.6 The Rme2s reader protein SMN and its role in protein complex assembly........36
1.3.7 Proposed roles of arginine methylation and SMN in RNAPII termination........37

1.4 Transcription and RNA processing related mis-regulations in neurodegeneration.....38

1.4.1 The pathology of Spinal muscular atrophy (SMA)........................................39
1.4.2 Low Complexity (LC) Domains and the pathogenesis of ALS-FTD..............41
1.4.3 The cellular function of TDP-43.................................................................45
1.4.4 The cellular function of FUS......................................................................46
1.4.5 Proposed roles of transcription induced R-loops and the consequent DNA damage as the cause of neurodegeneration.........................................................48

Chapter 2: Symmetric arginine dimethylation of the RNAPII CTD residue R1810 mediates regulation of transcription termination by SMN and SETX.............................................52

1. Introduction........................................................................................................53

2. Results..................................................................................................................55

2.1 The identification of R1810 symmetric dimethylation on the POLR2A CTD......55
2.2 Characterization and use of an R1810me2s specific antibody........................58
2.3 The identification of PRMT5 as the likely methyltransferase for R1810 in vivo...61
2.4 Methylation of R1810 on the RNAPII CTD by PRMT5 in vitro......................64
2.5 Recognition of RNAPII CTD R1810me2s by SMN in vitro.............................67
2.6 Recognition of RNAPII CTD R1810me2s by SMN in vivo.............................72
2.7 SMN enhances the interaction of SETX and RNAPII......................................75
2.8 RNAPII R1810me2s enhances the recruitment of SMN and SETX to RNAPII....78
2.9 SMN and R1810me2s regulate transcription termination by RNAPII..........83
2.10 SMN and R1810me2s regulate RNAPII termination through the resolution of R-loops..................................................93
2.11 R-loops induce DNA damage at the termination sites...............................98
2.12 RNAPII termination defects and R-loop accumulation in SMA disease states...98

2.13 Genome-wide characterization by ChIP-seq of the roles of R1810 and SMN in RNAPII termination.................................................................101
2.14 The effect of the CTD R1810A mutation on alternative splicing..................109
2.15 The effect of the CTD R1810A mutation on splicing shows some correlation with GC content.................................................................110
2.16 CTD R1810A effect on splicing does not correlate with RNAPII pause sites...113

3. Discussion.........................................................................................................116
Chapter 3: The regulation of RNAPII transcription termination by FUS and TDP-43 and its relevance to neurodegenerative disorders

1. Introduction

2. Results

   2.1 Identification of SMN-interacting proteins by AP-MS and BioID
   2.2 Interaction of FUS and TDP-43 with RNAPII and termination factors
   2.3 FUS and TDP-43 function downstream of the R1810me2s-SMN pathway
   2.4 SMN and CTD R1810me2s enhance the association of FUS and TDP-43 with transcribing RNAPII
   2.5 FUS and TDP-43 regulate transcription termination by RNAPII
   2.6 FUS and TDP-43 prevent the accumulation of R-loops and DNA damage at termination sites
   2.7 The use of SMN inhibitors mimics the effects of an RNAPII CTD R1810A mutation and an SMN mutant
   2.8 Genome-wide characterization of FUS and TDP-43 in RNAPII termination by ChIP-seq

3. Discussion

Chapter 4: Discussion, conclusions, and future directions

4.1 Discussion

4.1.1 The identification of the R1810me2s modification on the RNAPII CTD
4.1.2 SMN as an adaptor for 3’ end processing and RNAPII termination factors
4.1.3 The roles of FUS and TDP-43 in R-loop resolution and RNAPII termination
4.1.4 Possible mechanisms linking R-loop accumulation to neurodegeneration
4.2 Conclusions
4.3 Future directions

4.3.1 RNA-seq analysis for splicing and APA changes
4.3.2 Gene looping and promoter-proximal RNAPII termination
4.3.3 Is TDRD3 a component of TC-NER that regulates RNAPII degradation?

Chapter 5: Materials and methods

References
List of Figures

Figure 1 The Human RNAPII CTD and the modifiers.................................................................3
Figure 2 Arginine methylations and the potential reader SMN..................................................7
Figure 3 The detection of symmetric dimethylation of R1810 on the RNAPII CTD.......................56
Figure 4 Further characterization of the RNAPII CTD R1810me2s antibody...............................59
Figure 5 Symmetric dimethylation of R1810 on the RNAPII CTD by PRMT5 in vivo.....................62
Figure 6 Symmetric dimethylation of R1810 on the RNAPII CTD by PRMT5 in vitro......................65
Figure 7 R1810me2s in the RNAPII CTD is recognized by SMN in vitro.......................................68
Figure 8 CTD R1810me2s-SMN interaction not strongly affected by nearby phosphorylations...........70
Figure 9 The R1810me2s modification in the RNAPII CTD is recognized by SMN in vivo.................73
Figure 10 R1810me2s and SMN enhance Senataxin interaction with RNAPII.................................76
Figure 11 R1810me2s and SMN enhance Senataxin association with RNAPII on the beta-actin gene.................................................................................................................................79
Figure 12 R1810me2s and SMN enhances Senataxin association with RNAPII on the GAPDH gene....................................................................................................................................................81
Figure 13 SMN and R1810me2s regulate transcription termination by RNAPII on the beta-actin gene....................................................................................................................................................85
Figure 14 SMN and R1810me2s regulate transcription termination by RNAPII..............................87
Figure 15 R1810me2s regulates transcription termination by RNAPII (NRO).................................89
Figure 16 SMN and R1810me2s regulate transcription termination by RNAPII (Tech. replicates) ........91
Figure 17 SMN and R1810me2s are important for resolving R-loops created by elongating RNAPII....................................................................................................................................................94
Figure 18 The knock-out of SMN leads to increased RNAPII and R-loop accumulations..............96
Figure 19 SMN and R1810me2s are important for preventing R-loop induced DNA damage...........99
Figure 20 RNAPII and R-loop accumulations are observed in the SMA disease state.................102
Figure 21 SMN and POLR2A R1810 effects on transcription termination by RNAPII occur on a genome wide level........................................................................................................................................106-107
Figure 22 RNA-seq analysis for WT v. R1810A mutant RNAPII.................................................111
Figure 23 GC content analysis for cassette exons and intron retentions.....................................114
Figure 24 Correlation between cassette exon changes and RNAPII pausing patterns.................117
Figure 25 Model of the pathway that regulates R-Loop accumulation to prevent DNA damage at the RNAPII termination regions..................................................................................................121
Figure 26 FUS and TDP-43 localize to the nucleus......................................................................130
Figure 27 FUS and TDP-43 interact with RNAPII, SMN, and the termination factors.................133
Figure 28 FUS and TDP-43 interaction with SMN and the termination factors is independent of RNAPII, and their interaction with RNAPII is mediated by the CTD R1810me2s modification..........................................................135
Figure 29 SMN and PRMT5 mediate the interactions of FUS and TDP-43 with RNAPII............137
Figure 30 CTD R1810me2s and SMN enhance FUS and TDP-43 association with RNAPII.......140
Figure 31 FUS and TDP-43 regulate transcription termination by RNAPII on the beta-actin gene ..............................................................143
Figure 32 FUS and TDP-43 regulate transcription termination by RNAPII on the GAPDH gene ...........................................................................145
Figure 33 TDP-43 RRM is needed for its regulation of transcription termination by RNAPII ........................................................................................ 148
Figure 34 FUS and TDP-43 are important for resolving R-loops created by the elongating RNAPII ........................................................................................................................................................................150
Figure 35 TDP-43 RRM RNA binding domain is needed for its repression of R-loops...........152
Figure 36 FUS and TDP-43 are important for repressing R-loop induced DNA damage........155
Figure 37 SMN inhibitors block the Rme2s biding pocket of SMN........................................158
Figure 38 SMN inhibitors mimic the effects of an RNAPII CTD R1810A mutation and an SMN mutant (ChIP)........................................................................................................................................................................160
Figure 39 SMN inhibitors mimic the effects of an RNAPII CTD R1810A mutation and an SMN mutant (Co-IP)........................................................................................................................................................................162
Figure 40 FUS and TDP-43 regulate RNAPII Promoter and Terminator Stalling (ChIP-seq) ........................................................................................................................................................................165-166
Figure 41 Model: SMN as a nucleator of Rme2s proteins involved in transcription termination and neurodegeneration through the recruitment of FUS or TDP-43 to prevent R-loop formation and DNA damage at the RNAPII termination sites..............................................................169
Figure 42 SMN, FUS, and TDP-43 knock-outs lead to overlapping gene expression changes..185
Figure 43 TDRD3 regulates RNAPII degradation upon UV stress........................................192-193

List of Tables

Table 1 Genes implicated in the pathology of ALS.................................................................44
Table 2 SMA patient and control cell lines obtained from the Coriell Institute....................104
Table 3 RNAPII 3' processing and termination factors that contain Rme2 modifications as potential SMN binders.........................................................................................124
Table 4 SMN interactors identified by AP-MS and BioID.......................................................128
Table 5 GO term analysis of gene expression changes upon the knock-out of FUS, TDP-43, or SMN..............................................................................................................................187
Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ab</td>
<td>Antibody</td>
</tr>
<tr>
<td>ALS</td>
<td>Amyotrophic Lateral Sclerosis</td>
</tr>
<tr>
<td>ANG</td>
<td>Angiogenin Ribonuclease</td>
</tr>
<tr>
<td>AOA</td>
<td>Ataxia oculomotor apraxia</td>
</tr>
<tr>
<td>AP</td>
<td>Affinity purification</td>
</tr>
<tr>
<td>APA</td>
<td>Alternative polyadenylation</td>
</tr>
<tr>
<td>AS</td>
<td>Alternative splicing</td>
</tr>
<tr>
<td>BioID</td>
<td>Proximity-dependent biotin identification</td>
</tr>
<tr>
<td>BrUTP</td>
<td>Bromouridine-triphosphate</td>
</tr>
<tr>
<td>C9ORF72</td>
<td>Chromosome 9 Open Reading Frame 72</td>
</tr>
<tr>
<td>CDK</td>
<td>Cyclin dependent kinase</td>
</tr>
<tr>
<td>ChIP</td>
<td>Chromatin immunoprecipitation</td>
</tr>
<tr>
<td>CID</td>
<td>CTD Interaction Domain</td>
</tr>
<tr>
<td>CPSF</td>
<td>Cleavage and polyA specificity factor</td>
</tr>
<tr>
<td>CstF</td>
<td>polyA cleavage stimulatory factor</td>
</tr>
<tr>
<td>CTD</td>
<td>C-terminal domain</td>
</tr>
<tr>
<td>DDR</td>
<td>DNA damage response</td>
</tr>
<tr>
<td>DIP</td>
<td>DNA immunoprecipitation</td>
</tr>
<tr>
<td>DSBs</td>
<td>Double-stranded breaks</td>
</tr>
<tr>
<td>DSIF</td>
<td>DRB-sensitivity-inducing factor</td>
</tr>
<tr>
<td>EWSR1</td>
<td>EWS RNA-Binding Protein 1</td>
</tr>
<tr>
<td>FCP1</td>
<td>CTDP1 (CTD phosphatase subunit 1)</td>
</tr>
<tr>
<td>FET</td>
<td>FUS, EWSR1, TAF15 proteins</td>
</tr>
<tr>
<td>FITC</td>
<td>Fluorescein isothiocyanate</td>
</tr>
<tr>
<td>FMR</td>
<td>Fragile X mental retardation</td>
</tr>
<tr>
<td>FP</td>
<td>Fluorescence Polarization</td>
</tr>
<tr>
<td>FTD</td>
<td>Frontal Temporal Dementia</td>
</tr>
<tr>
<td>FUS</td>
<td>Fused in Sarcoma</td>
</tr>
<tr>
<td>GFP</td>
<td>Green fluorescent protein</td>
</tr>
<tr>
<td>GG-NER</td>
<td>Global genome nucleotide excision repair</td>
</tr>
<tr>
<td>GRN</td>
<td>Granulin</td>
</tr>
<tr>
<td>GRO</td>
<td>Global run-on</td>
</tr>
<tr>
<td>3H-SAM</td>
<td>S-[Methyl-$^3$H]-Adenosyl-L-methionine</td>
</tr>
<tr>
<td>hnRNPs</td>
<td>heterogeneous nuclear ribonucleoproteins</td>
</tr>
<tr>
<td>HR</td>
<td>Homologous recombination</td>
</tr>
<tr>
<td>IP</td>
<td>Immunoprecipitation</td>
</tr>
<tr>
<td>ITC</td>
<td>Isothermal Titration Calorimetry</td>
</tr>
<tr>
<td>Kd</td>
<td>Dissociation constant</td>
</tr>
<tr>
<td>LC</td>
<td>Low Complexity Domain</td>
</tr>
<tr>
<td>MMA</td>
<td>Monomethylation</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>MRN</td>
<td>MRE11, RAD50, and NBS1</td>
</tr>
<tr>
<td>MS</td>
<td>Mass spectrometry</td>
</tr>
</tbody>
</table>

x
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ncRNA</td>
<td>Non-coding RNA</td>
</tr>
<tr>
<td>NDRs</td>
<td>Nucleosome-depleted regions</td>
</tr>
<tr>
<td>NELF</td>
<td>Negative elongation factor</td>
</tr>
<tr>
<td>NHEJ</td>
<td>Nonhomologous end joining</td>
</tr>
<tr>
<td>NLS</td>
<td>Nuclear localization signal</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>NRO</td>
<td>Nuclear run-on</td>
</tr>
<tr>
<td>OPTN</td>
<td>Optineurin</td>
</tr>
<tr>
<td>P-TEFb</td>
<td>Transcription elongation factor-b</td>
</tr>
<tr>
<td>PAS</td>
<td>PolyAdenylation signal (AAUAAA)</td>
</tr>
<tr>
<td>PABP</td>
<td>polyA binding protein</td>
</tr>
<tr>
<td>PFN1</td>
<td>Profilin 1</td>
</tr>
<tr>
<td>POLR2A</td>
<td>RNA polymerase II subunit A</td>
</tr>
<tr>
<td>polyA</td>
<td>Polyadenylation</td>
</tr>
<tr>
<td>PRMT</td>
<td>Protein arginine methyltransferase</td>
</tr>
<tr>
<td>PSI</td>
<td>Percent spliced index</td>
</tr>
<tr>
<td>R-loops</td>
<td>DNA:RNA hybrid structure</td>
</tr>
<tr>
<td>RBP</td>
<td>RNA-binding protein</td>
</tr>
<tr>
<td>Rme2a</td>
<td>Asymmetric arginine dimethylation</td>
</tr>
<tr>
<td>Rme2s</td>
<td>Symmetric arginine dimethylation</td>
</tr>
<tr>
<td>RNAPII</td>
<td>RNA polymerase II</td>
</tr>
<tr>
<td>RNP</td>
<td>Ribonucleoprotein</td>
</tr>
<tr>
<td>RRM</td>
<td>RNA recognition motif</td>
</tr>
<tr>
<td>rRNA</td>
<td>Ribosomal RNA</td>
</tr>
<tr>
<td>SCX-HILIC</td>
<td>Strong cation exchange-hydrophilic interaction chromatography</td>
</tr>
<tr>
<td>s.e.m</td>
<td>Standard error of the mean</td>
</tr>
<tr>
<td>SETX</td>
<td>Senataxin</td>
</tr>
<tr>
<td>SILAC</td>
<td>Stable isotope labeling in cell culture</td>
</tr>
<tr>
<td>SMA</td>
<td>Spinal Muscular Atrophy</td>
</tr>
<tr>
<td>SMN</td>
<td>Survival of motor neurons</td>
</tr>
<tr>
<td>snoRNA</td>
<td>Small nucleolar RNA</td>
</tr>
<tr>
<td>snRNA</td>
<td>Small nuclear RNA</td>
</tr>
<tr>
<td>SPA</td>
<td>Scintillation proximity assay</td>
</tr>
<tr>
<td>TAF15</td>
<td>TATA Box Binding Protein (TBP)-Associated Factor 15</td>
</tr>
<tr>
<td>TARDBP</td>
<td>TAR DNA binding protein</td>
</tr>
<tr>
<td>TDP-43</td>
<td>TARDBP</td>
</tr>
<tr>
<td>TC-NER</td>
<td>Transcription coupled nucleotide excision repair</td>
</tr>
<tr>
<td>TDRD</td>
<td>Tudor domain containing proteins</td>
</tr>
<tr>
<td>TOP</td>
<td>Topoisomerases</td>
</tr>
<tr>
<td>tRNA</td>
<td>Transfer RNA</td>
</tr>
<tr>
<td>TSS</td>
<td>Transcription start site</td>
</tr>
<tr>
<td>TTS</td>
<td>Transcription termination site</td>
</tr>
<tr>
<td>UBQLN2</td>
<td>Ubiquilin 2</td>
</tr>
<tr>
<td>VCP</td>
<td>Valosin Containing Protein</td>
</tr>
<tr>
<td>WCL</td>
<td>Whole cell lysates</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

1.1 The RNAPII transcription machinery and the C-terminal domain (CTD)

In eukaryotes, amongst the three RNAP complexes, RNAPI and RNAPIII have more restricted transcriptional roles such that RNAPI synthesizes the 18S and 28S ribosomal RNAs (rRNAs), and RNAPIII generates short transcripts that code for structural or catalytic RNAs such as the transfer RNAs (tRNAs), 5S rRNA, and the U6 small nuclear RNAs (snRNA) [1, 2]. RNAPII is responsible for the synthesis of an extremely diverse set of RNAs, including both coding RNAs (mRNAs that then are used as templates for protein synthesis) or non-coding RNAs (ncRNAs) [5]. Even though three RNAP subunits are homologous to ancestral forms also found in bacteria, the regulation of eukaryotic RNAPII is much more complex, as exemplified by the roles of its CTD [2, 6, 7]. On the RNAPII core enzyme, the CTD is positioned next to the RNA exit channel, at a convenient location for it to coordinate transcription and RNA processing [8]. Wherever it has been tested, deletion of the CTD in eukaryotes is lethal, indicating the essentiality of its function in the regulation of RNAPII activity and other co-transcriptional events [9-11].

α-amanitin is a fungal toxin that binds to the POLR2A subunit of RNAPII and inhibits RNAPII at low concentration, inhibits RNAPIII at a higher concentration, and exerts no effect on RNAPI [12, 13]; therefore, it has been widely used for the study of RNAPII. This inhibitor led to the degradation of the POLR2A subunit of RNAPII in vivo, whereas an exogenously introduced POLR2A subunit [13] with a mutation in the Trigger Loop that causes resistance to α-amanitin is not degraded [14]. The Trigger Loop is a highly conserved mobile element that plays critical roles in NTP substrate selection, RNAPII elongation, and transcription fidelity [14]. The
binding of α-amanitin to POLR2A permits NTP entry to the active site but prevents the translocation of DNA and RNA needed to empty the site for the next round of catalysis [15].

1.1.1 The RNAPII CTD Code in transcription regulation

This introduction will focus on discussion of the mammalian RNAPII CTD. The CTD of the largest subunit of mammalian RNAPII (POLR2A) contains 52 heptapeptide repeats, with the N-terminal half containing mainly consensus repeats (Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7) and the C-terminal half containing more heterogeneous repeats (with Ser7 being the most variant residue) (Fig. 1A) [16, 17]. The CTD appears to be an unstructured flexible sequence that provides a platform for a plethora of modifications (e.g. phosphorylation, methylation, glycosylation, ubiquitination, proline isomerization) that then act as scaffolds for the recruitment of various reader proteins [1, 5]. Although it is possible to replace the yeast CTD with its mammalian counterpart and maintain viability [18], replacement of the mammalian CTD with a consensus CTD of the same length reduces cell viability [19]. This probably reflects the evolutionary replacement of consensus phosphorylatable residues with other residues that can be modified in different ways (e.g. acetylation or methylation) [2]. The non-consensus repeats that vary extensively from yeast to human may enable the recruitment of novel factors that are yet to be identified [2].

Transcription by RNAPII is usually divided into three phases, initiation, elongation, and termination, each associated with different sets of regulatory proteins [5, 17] (Fig.1B). Most CTD modifications (e.g. phosphorylation) are dynamic, reversible, and closely associated with a particular stage of the transcription cycle (Fig. 1C). As well, combinatorial modifications could recruit certain CTD binders with higher efficacy to potentiate transitions between phases [5, 20].
Figure 1 The Human RNAPII CTD and the CTD modifiers
To enable the accurate and robust cycling of the three phases of transcription that is particularly important for the regulation of gene expression, RNAPII also relies on cross-talk between modifications on the CTD and those on the transcribed chromatin [21]. In vertebrates, the C-terminal end of the CTD is also capped by a consensus 10 amino acid extension containing a site for casein kinase II phosphorylation, which is needed for RNAPII stability and activity [22, 23].

1.1.2 Reversible phosphorylation of the RNAPII CTD during the transcription cycle

*In vivo*, the CTD repeats can be reversibly phosphorylated on Tyr1, Thr4, and all three Ser residues (Ser2, Ser5, and Ser7), and glycosylation of Ser and Thr and isomerization of the two Pro residues are also known to occur [1, 5] (Fig. 1C). These activities depend on many enzymes that specifically write or erase a type of modification on the CTD. The hypophosphorylated form of the RNAPII CTD preferentially associates with the pre-initiation complex at the promoter [24, 25]. Ser5 phosphorylation by the cyclin-dependent kinase CDK7 of the general initiation factor TFIIH correlates with transcription initiation and recruitment of the methylguanosine capping enzyme complex for mRNA capping that occurs at the 5’ ends of genes [26-30]. It is also important for the histone H3K4 trimethylation that predominates near the 5’ ends of genes [31]. Ser2 phosphorylation by CDK9 of the positive transcription elongation factor-b (P-TEFb) complex and CDK12/13 is important for RNAPII elongation and occurs more prominently towards the 3’ ends of genes [16, 28, 32-35]. pSer2 allows the recruitment of enzymes needed for mRNA splicing and polyadenylation, RNAPII termination, and the histone H3K36 methylation that occurs along actively transcribed genes [16, 28, 32, 33, 36]. Ser7 phosphorylation (by CDK7) on the RNAPII CTD occurs from the 5’ to the 3’ ends of
genes, and has been shown to control snRNA transcription and the 3’ end processing of mRNAs [17, 37, 38].

Dynamic dephosphorylation of Ser2 and Ser5 by CTD-specific phosphatases enables the recycling of the RNAPII machinery [3, 5]. The levels of pSer2 are regulated at 3’ ends of genes by the Ser2 phosphatase FCP1 [3, 39-44]. There exist multiple Ser5 phosphatases in yeast, such as Ssu72, Scp1, and Rtr1 [3, 42, 45-47]; in higher eukaryotes, Ser5 is dephosphorylated by the Rtr1 homolog RNAPII-associated protein 2 (RPAP2) and CDC14 [17, 48-51]. Interaction of Ssu72 with the general initiation factor TFIIB in yeast causes looping of the 5’ and 3’ ends of genes [52, 53]. Inhibition of Ssu72 in yeast causes the direction of RNAPII transcription initiation to become mis-regulated, suggesting that the three phases of the transcription cycle are coordinated through the physical looping of the 5’ and 3’ ends of genes [53]. Ssu72 has recently been shown to also dephosphorylate pSer7, and this function may, as well, be needed to control proper looping and the directionality of transcription [17, 54]. Other non-canonical CTD phosphorylations on Ser2 and Ser5 can be carried out by CDK8, CDC2, ERK1, ERK2, and perhaps other proteins to create a hyperphosphorylated form of the CTD that may cause RNAPII to be transcriptionally inactive, or may initiate a signaling cascade as a DNA damage response (DDR), leading to the degradation of RNAPII [55-57].

Tyr1 phosphorylation of the CTD was also found to occur in vivo, and is conserved from yeast to human [58, 59]. It may be carried out by the c-ABL kinase [60]. This modification occurs on all active genes and correlates with pSer2 levels such that it is low at promoters but rises towards the transcription termination site (TTS) [59]. In contrast to the pSer2 mark, the pTyr1 mark is known to stimulate binding of the elongation factor Spt6 and impair the recruitment of termination factors such as Nrd1, Pcf11, and Rtt103 in yeast. It decreases (due to
dephosphorylation by the Glc7 phosphatase subunit of the yeast cleavage and polyadenylation factor) before the polyadenylation (polyA) site [59, 61]. In higher eukaryotes, pTyr1 is found at the promoters in an antisense orientation and is also specifically enriched at active enhancers, and this modification is also known to protect the RNAPII CTD from degradation by the proteasome [58, 62]. There also exists Thr4 phosphorylation on the CTD generated by the PLK3 and CDK9 kinases, and its dephosphorylation by the FCP1 phosphatase [63-65]. This modification was first identified to regulate histone mRNA 3’ end processing [63-65]. Later, it was also shown that Thr4 phosphorylation is required for activation-associated eviction of the histone variant Htz1 from the promoters of yeast Thr4-dependent genes [66].

### 1.1.3 Asymmetric dimethylation of the RNAPII CTD residue R1810

There exist two arginine residues, R1603 and R1810, on the human RNAPII CTD that are well conserved across higher eukaryotes (i.e. vertebrates) (Fig. 2A). In the 31st repeat of the human CTD, the Ser7 position is replaced by an arginine residue that was shown to be asymmetrically dimethylated (Rme2a) by the coactivator-associated arginine methyltransferase 1 (CARM1/PRMT4) [67]. As pSer2 or pSer5 modification inhibits CARM1 activity at this site in vitro, the authors suggested that this methylation occurs before CTD phosphorylation and is preserved along the active and elongating RNAPII through the transcription cycle [67]. The substitution of R1810 by alanine results by an unknown mechanism in the up-regulation of many classes of snRNAs and snoRNAs with comparatively little effect on the expression of protein-coding genes. The Tudor (Tud) domain-containing protein TDRD3 binds specifically to this modification in vitro, yet does not appear to mediate regulation of snRNA or snoRNA gene expression in vivo [67]. In a later NMR study, the TDRD3 Tudor domain was shown to bind
2A The amino acid sequence alignment of R1603 and R1810 containing CTD sequences across the vertebrates.

2B The types of arginine methylation. Type III PRMT generates monomethylation, Type II generates symmetric dimethylation, and Type I generates asymmetric dimethylation [265].

2C The Tudor domain of SMN with the structure shown in two orientations. The Tudor domain is composed of 4 beta-sheets (blue) that form a pocket with 3 aromatic amino acids (orange) that mediates the binding with dimethylated Arg. The domain compositions of SMN is shown below [327, 386].

Figure 2 Arginine methylations and the potential reader SMN
weakly to the R1810me2a modification on the CTD with a Kd of ~800uM [68].

It was known that a mammalian RNAPII with a truncated CTD that contains 31 repeats and lacks the 31\textsuperscript{st} repeat and R1810 retains mRNA at the site of transcription, such that final maturation of spliced and 3' end polyadenylated mRNAs into exported ribonucleoproteins do not occur [69]. In agreement with this observation, I will show that the CTD residue R1810 is needed for RNAPII termination, a defect in which should cause RNAPII to be retained at the transcription site [70, 71]. I found that R1810 is also symmetrically dimethylated (Rme2s) by the arginine methyltransferase PRMT5 for the recruitment of the downstream adaptor protein SMN through its Tudor domain (Fig. 2B, 2C). I do not know if there exists an arginine demethylase that allows for the conversion between the Rme2a and the Rme2s modifications at the R1810 residue of the CTD during the various phases of the transcription cycle.

1.1.4 RNAPII promoter-proximal pausing and the switch into elongation

In higher eukaryotes, initiation and the processive elongation phase of transcription are separated by a checkpoint that is termed promoter-proximal pausing at 30-80bp downstream of the transcription start site (TSS) [72]. Advances in high-resolution sequencing technologies coupled with chromatin immunoprecipitation (ChIP) or global run-on (GRO) techniques allowed the precise determination of RNAPII pausing patterns on a genome-wide scale. Through GRO-seq, about 40–46% of active genes were shown to contain promoter-proximally paused RNAPII in mammalian cells [73-75].

Promoter-proximal pausing involves at least two transcription elongation factors, the negative elongation factor (NELF) and the DRB-sensitivity-inducing factor (DSIF) [76-78], which stabilize the paused RNAPII. The release of RNAPII is mediated by the P-TEFb complex.
[34, 79]. In cells, most P-TEFb appears to be inactive, because it is sequestered in complexes that also contain the 7SK ncRNA and the HEXIM1 protein [80, 81]. Recruitment of P-TEFb to target genes requires various changes in chromatin modification, such as deubiquitination of the histone H2Bub mark [82] and phosphorylation of the histone H3S10 residue [83], as well as the recruitment of the proto-oncoprotein c-MYC [84] and the bromo-domain-containing protein BRD4 [85, 86], which binds to the acetylated chromatin that marks active genes. Upon Ser2 phosphorylation by the p-TEFb complex, NELF is ejected from RNAPII, and DSIF, which consists of SPT4 and SPT5, becomes a positive elongation factor [34, 78, 87, 88].

The kinetics of promoter-proximal pausing is thought to be balanced between the RNAPII initiation force and the arrest by pausing factors that recognize certain DNA sequences at the promoters [72, 89]. The interaction model suggests that sequence-specific DNA binding factors interact with RNAPII to elicit pausing, whereas the barrier model suggests that nucleosomal factors on the chromatin block RNAPII from moving forward [76, 89]. The position of RNAPII pausing is determined by the strength of the promoter DNA, as stronger promoters tend to be associated with more promoter-proximal pausing (at +40) than weaker promoters (at +80) [72, 75, 89]. Thus, for promoters with stronger DNA elements, the interaction model predominates, whereas for promoters with weaker DNA elements, the nucleosomal barriers may stop RNAPII more distally [89]. Analogously to promoter-proximal pausing, paused RNAPII in the distal termination sites of a gene often correlates with high GC content in the DNA distal to the polyadenylation site (PAS) motif [90, 91]. It is possible that similar phenomena also regulate RNAPII termination, and our current understanding of termination is detailed in Section 2 of this introduction.
Promoter-proximally paused RNAPII may actually terminate transcription, followed by recycling or degradation of the enzyme [92-94]. It is known that RNAPII stalled at DNA damage sites can be ubiquitinated and degraded during transcription coupled nucleotide excision repair (TC-NER) [95, 96]. Otherwise, removal of RNAPII upon promoter-proximal termination may occur similarly to the events at distal termination sites after cleavage and polyadenylation of the mRNA, where RNAPII is released after the degradation of the nascent RNA by the 5′→3′ exonuclease XRN2 and the 3′→5′ exonuclease activity of the exosome [91, 97-99].

Once switched into a processive elongation mode, RNAPII elongates at an average rate of about 4.3 kb per minute, or 70 bp per second, and may extend the RNA to >1 Mb in length in vertebrates [100]. Elongation rates are highly dynamic, perhaps because certain DNA sequences tend to form secondary structures (DNA:RNA hybrids or G-quadruplexes) and thus are harder to transcribe, or perhaps because of barriers created by site-specific DNA binding proteins [72]. As well, variations in elongation rate may also be controlled by differential nucleosomal packing and modification, or the presence or absence of elongation factors [72, 89]. One of the elongation factors, the PAF complex, serves as a platform for recruiting other complexes, such as the SEC (super elongation complex, which contains P-TEFb), FACT, and various histone modifying enzymes, to the elongating RNAPII [101-104]. FACT facilitates elongation by disassembling and reassembling H2A-H2B dimers to overcome nucleosomal barriers [105, 106]. Also recruited by P-TEFb and the phosphorylated CTD is the positive elongation factor SPT6 [107, 108]. SPT6 is a histone chaperone that interacts with histones H3/H4 to displace nucleosomes in front of the RNAPII and reassemble them afterwards [109]; this function allows it to suppress cryptic transcription initiation within intragenic regions [110].
1.1.5 RNAPII CTD regulation of co-transcriptional alternative splicing

RNAPII elongation rates are known to affect co-transcriptional processes such as splicing and 3’ end processing of the mRNAs [72, 89]. Adding to the complexity of this issue is the cross-talk between the CTD and chromatin modifications (e.g. H3K36me3, H2Bub, H3K79me3) that also modulate the RNAPII elongation rate [72, 89, 111, 112]. RNAPII elongation rate may also be influenced by splicing, as RNAPII is known to pause near splice sites [113, 114]. Two models have been proposed to describe the role of the RNAPII CTD in the regulation of co-transcriptional splicing, although aspects of both models may be correct [5, 115]. According to the recruitment model, modification of the CTD would affect the recruitment of splicing factors that, in turn, would alter the splicing patterns of the transcripts [115, 116]. According to the kinetic model, changes in promoter strength and chromatin modifications occurring as a consequence of modifications on the RNAPII CTD would affect the elongation rate of the RNAPII, thereby altering the window of time available for co-transcriptional splicing reactions to occur [55, 115]. In the kinetic model, a slow moving RNAPII may favor alternative exon inclusion or exclusion, depending on the splicing factors that are recruited [115, 117, 118]. Both models received support from a study in which the application of a DNA damaging reagent led to the appearance of a hyperphosphorylated form of the RNAPII CTD [55]. Hyperphosphorylation of the CTD was accompanied by a slow RNAPII elongation rate, as well as changes in the co-transcriptional splicing of many genes that function in cell cycle control and apoptosis [55]. However, it is not clear how this form of CTD is related to the lack of recruitment of the splicing factor EWSR1, which led to similar changes in alternative splicing (AS) patterns in the examined genes [119, 120]. A second study involving the application of DRB (which blocks the activity of pTEF-b) to inhibit RNAPII elongation also supported the
kinetic model by showing that co-transcriptional splicing of many genes (such as splicing factors and RNA-binding protein encoding genes) is affected [121]. In this case, many splicing changes introduced premature termination codons that then elicited nonsense-mediated decay (NMD) of the mRNAs [121].

The coupling between splicing and RNAPII elongation could be mediated through the RNAPII CTD, as many splicing factors are known to interact with the CTD directly [89]. Also, the finding that phosphorylation of the CTD is important for the binding of splicing factors raises the possibility that differential CTD phosphorylation may contribute to the regulation of AS [122]. Among the phospho-CTD-binding splicing factors are yeast Prp40 and mammalian TCRG1/CA150 and U2AF65 [115, 123]. CA150 is known to bind to the elongating form of RNAPII via the hyperphosphorylated form of the CTD in vitro through its FF domains [124]; Prp40 prefers to bind a CTD phosphorylated on both Ser2 and Ser5 through its WW domains (which are related to FF domains) [125]; and U2AF65 prefers to bind the phosphorylated CTD and stimulates splicing in vitro through its association with the spliceosomal factor Prp19c [126]. On the other hand, the splicing and polypyrimidine tract binding protein-associated splicing factor p54nrb/PSF binds by recognizing the unphosphorylated and phosphorylated forms of the CTD equally well [127]. As well, the DBIRD (Dbc1-Znf326) complex involved in splicing was discovered to bind to the elongating RNAPII directly and to mediate splicing of exons embedded in AT rich DNA, as its depletion was shown to lead to exon skipping [128].

1.1.6 Readers of the RNAPII CTD code

Several studies have identified proteins that bind specifically to either the unphosphorylated CTD or the CTD with distinct patterns of modifications, and some readers
may be preferentially recruited by a combination of CTD modifications [5, 17]. The pSer5 modification of the CTD recruits the guanylyltransferase for the 5′ end capping of nascent transcripts [129], as well as the Set1/COMPASS histone methyltransferase complex that trimethylates histone H3K4 in the promoter regions of transcribed genes [104]. pSer5 on the CTD also recruits the Rtr1 and Scp1 pSer5 phosphatases, the Rpd3S histone deacetylase complex, the elongation factor PAF, and the Nrd1 component of the NRD termination complex for snRNA/snoRNA processing in yeast [46, 130-134]. Many CTD interactors, such as Nrd1, Pcf11, RPRD1A/1B, and a number of SR-containing splicing factors, bind to the modified CTD through a conserved CTD Interaction Domain (CID) [49, 135-137].

The CTD pSer2 mark that increasingly accumulates through the length of a gene recruits readers that facilitate histone modification, mRNA splicing, 3′ end formation, and RNAPII elongation and termination. Two yeast positive elongation factors, Npl3 and Spt6, and a termination factor for yeast, Rtt103, recognize pSer2 on the CTD specifically [17]. Differing somewhat from the pSer2 CTD mark, pTyr1 also accumulates through the length of the gene body but decreases toward the 3′ end in yeast; its presence on the CTD is known to stimulate the association of the elongation factor Spt6 and impair the recruitment of termination factors such as Nrd1, Pcf11, and Rtt103 [59]. For RNAPII termination, the CID of the cleavage factor Pcf11 is a reader for pSer2 on the CTD [136, 137], and so are the CPSF1 subunit of the cleavage and polyadenylation specificity factor (CPSF) and the cleavage stimulatory factor (CstF) [138].

There are many factors that recognize a combinatorial CTD code, such as pSer5 + pSer7, pSer2 + pSer5, or pSer2 + pSer7, to elicit a range of regulatory functions [5, 17]. The pSer5 and pSer7 marks in combination allow the recruitment of the integrator complex for the 3′ end processing of snRNAs [139, 140] and the RPAP2 phosphatase to snRNA genes [139]. As well,
Int11 of the integrator complex can recognize the pSer2 + pSer7 double mark [141]. The yeast
H3K36 methyltransferase Set2 is recruited by the pSer2 + pSer5 double mark to catalyze
methylation of histone H3K36 along the length of actively transcribed genes [142]. The splicing
factors Prp40 and U2AF65, as well as the proline isomerases Pin1 and Ess1, also recognize the
pSer2 + pSer5 double mark [126, 143, 144]. Two human proteins, RPRD1A and RPRD1B form
dimers through their coiled-coil domains and interact with CTD pSer2 + pSer7 via their CID
[49]. RPRD1A and RPRD1B also enable the recruitment of the RPAP2 phosphatase to pSer2 +
Ser7 double marks that bracket pSer5 residues for the efficient dephosphorylation of pSer5 by
RPAP2 [49].

Adding to the layers of complexity of the CTD code are the proline isomerases Pin1 and
Ess1 that isomerize the Pro3 and Pro6 of the CTD repeats [5, 17, 145], though it is not clear how
the isomerization cycles through the various phases of transcription. The polyadenylation factor
Pcf11 binds exclusively to pSer2 and prolines in trans [135]; the yeast pSer5 phosphatase Ssu72
reads the CTD with pSer5 and the downstream proline in cis [146]; and Nrd1 of the NRD
termination complex reads CTD pSer5 and requires the upstream pSer5-Pro6 bond of the CTD
to be in the cis conformation [147]. As the identification and characterization of CTD modifying
enzymes and binders continues, the complexity of the CTD code will no doubt expand [2].

1.1.7 The roles of chromatin modifications, chromatin structure, and DNA looping in
RNAPII transcription

A diverse set of chromatin modifications, chromatin remodelers, histone chaperones, and
histone variants are known to regulate nucleosomal dynamics for proper RNAPII transcription
[21]. During the initiation phase, chromatin remodelers remove or remodel nucleosomes on
active promoters [21, 148], and the nucleosome-depleted regions (NDRs) facilitate the
incorporation of the histone H2A.Z variant on the nearby nucleosomes [149]. H2A.Z facilitates RNAPII transcription by recruiting additional chromatin remodelers and transcription factors to active genes, and also antagonizes repressive DNA methylation [150, 151]. Acetylation of K14 on H2A.Z by the SAGA and the NuA4 complexes allows it to be retained and prevents its removal by the INO80 complex [152-155].

During the RNAPII elongation phase, the associated histone marks include the ubiquitination of histone H2BK20 by RNF20, acetylation of H3K56 by Rtt109, trimethylation of H3K36 by Set2, and dimethylation of H3K79 by Dot1 [102, 156-158]. These marks regulate RNAPII elongation by altering chromatin structure or by forming a platform to recruit factors such as the PAF elongation complex or histone chaperones (FACT, SPT6) for nucleosome remodeling [21, 72, 159]. The co-transcriptional methylation of H3K36 over the gene body prevents histone H3 exchange that would otherwise occur at the 5’ and 3’ ends of a gene; uninhibited H3 exchange is known to lead to internal initiation of transcription from cryptic promoters in the gene body (a process that is also kept under surveillance by SPT6) [160, 161].

Genome-wide studies have shown that many promoters from yeast to humans are bidirectional, and directional specificity is known to be influenced by the chromatin structure and modifications at the promoters [90]. There exist several pathways that suppress RNAPII bidirectional initiation, which would produce non-functional RNAs [90]. The upstream antisense ncRNAs often contain a higher density of polyadenylation signals (PAS -- AAUAAA) with a lack of U1 snRNP spliceosome recognition sites as compared to the sense mRNAs, suggesting that the spliceosome may facilitate the recognition and the synthesis of functional sense mRNAs [162]. The PAS may trigger proximal RNAPII termination by the CPSF and CstF complexes that also participate in distal RNAPII termination upon polyA site cleavage [162].
third pathway involves the formation of gene loops [163] mediated by the Ssu72 CTD phosphatase that is associated with both ends of a gene; the inactivation of Ssu72 can thus lead to increased divergent RNAPII initiation in yeast [53, 164]. Supporting the gene looping model in metazoans, co-transcriptional recruitment of processing factors, such as the 5’ capping complex, the elongation factor SPT5, and the 3’ end processing factors (CPSF, CstF), has been observed in ChIP assays to occur at both the 5’ and 3’ ends of many genes [165].

### 1.2 RNAPII termination pathways and the formation of DNA:RNA hybrids (R-loops)

#### 1.2.1 RNAPII termination pathways in metazoans

There exist three RNAPII transcription termination pathways in metazoans, which operate on the synthesis of snRNAs, histone encoding RNAs, and polyadenylated mRNAs, respectively [90]. snRNAs undergo 3’ end processing and termination by a specific pathway in higher eukaryotes that differs from the yeast Nrd1-Sen1-Nab3 (NRD) pathway, as the knock-down of SETX, the human homolog of yeast Sen1, does not lead to snRNA termination defects in metazoans [166]. In yeast, the Sen1 RNA/DNA helicase is a termination factor for coding and non-coding genes that is recruited by the pSer2 modification on the CTD at the 3’ ends of genes [167-169]. Instead, mammalian snRNA genes contain a conserved 13–16-nt sequence (3’ box) that is recognized by the integrator complex, which is recruited to the termination site by the pSer7 modification on the CTD [170]. Histone transcripts are mostly intronless and are not polyadenylated; instead, they form a stem-loop structure that then undergoes cleavage by the U7 snRNP and the CPSF3/73 endonuclease [171, 172].
I will focus this discussion on the cleavage-dependent RNAPII termination pathway for mRNAs, which involves the cleavage and polyadenylation specificity factors (CPSFs, PCF11, and WDR33) and the cleavage stimulatory factors (CstF) that are conserved across species from yeast to human [90, 173]. Through interacting with the nascent RNA and the RNAPII CTD [174], this set of factors also participates in promoter-proximal RNAPII termination for the regulation of gene expression and repression of bidirectional RNAPII initiation [97, 162].

1.2.2 RNAPII termination downstream of polyadenylation

Distal transcription termination typically occurs far downstream from cleavage and polyadenylation sites and requires either the cleavage signal and cleavage factors or else self-cleavage of the pre-mRNA with the CoTC sequence [99, 113, 175, 176]. The RNAPII CTD is needed for cleavage and polyadenylation *in vitro* and *in vivo* [174], and multiple interactions are known to occur between the 3’ end processing complex and the RNAPII. While CPSF4/30 can bind to the body of RNAPII [177], CstF1/50 and 3/77 [174, 178], CPSF1/160 and CPSF2/100[179, 180], and PCF11 [136, 137, 178, 181] interact specifically with the RNAPII CTD, often preferring the pSer2 modified form of the CTD [136, 137, 174]. Thus, the RNAPII CTD provides a docking site to stabilize the assembly of the CPSF and CstF complexes while they also interact with the 3’ UTR of the nascent transcript for 3’ end processing [174, 182-185]. In particular, RNAPII pauses and accumulates on the GU rich elements ~40 nt downstream of the polyadenylation site [186] prior to termination by interacting with CPSF4/30, which along with FIP1L1, also interacts with the polyU sequence upstream of the PAS motif on the nascent RNA [187-190]. CPSF4/30 and WDR33 then simultaneously recognize the PAS [191, 192].
Subsequently, CstF2/64 and PCF11 free the transcript from RNAPII [177, 193], and lastly, CPSF3/73 cleaves the RNA 18–30nt downstream of the PAS motif [194, 195].

In contrast to promoter-proximal pausing, RNAPII pausing at termination sites is not followed by immediate re-initiation, as the RNAPII has to be recycled. Pausing by RNAPII in the termination region is thought to enhance its release by increasing the ability of the 5’ → 3’ RNA exonuclease XRN2 to catch up to RNAPII (according to the Torpedo Model) [99, 175, 196, 197]. Pausing by RNAPII at termination sites is also associated with the formation of DNA:RNA hybrid structures (R-loops) when the nascent RNA anneals to the template DNA strand [91]. However, it is not clear whether R-loops enhance pausing or are a byproduct of paused transcription [90, 91]. The mammalian homolog of yeast Sen1, SETX, has been shown to enhance RNAPII termination after cleavage by resolving R-loops to facilitate the entry of the XRN2 nuclease [91]. A study on the circadian clock PER complex showed that PER proteins auto-regulate their own levels by influencing RNAPII termination [198]. During negative feedback, RNAPII accumulates at the termination sites of the Per and Cry genes, and the recruitment of the PER complex inhibits SETX action, thereby impeding RNAPII release and transcriptional re-initiation [198].

A recent study proposed that R-loops involved in RNAPII termination recruit the RNA interference pathway: the accumulation of R-loops induces antisense transcription, followed by the recruitment of the Dicer-Ago complex for the cleavage and removal of the resulting dsRNAs [199]. Dicer-Ago then promotes the deposition of histone H3K9me2 over termination sites which, in turn, recruits CBX3 to induce heterochromatization of the region and enhance RNAPII release [199]. A role for SETX in transcription termination may also be to facilitate cooperative action between the 5’ → 3’ exonuclease XRN2 and the 3’ → 5’ exosome for RNAPII release.
While XRN2 degrades the 5’ end of the nascent RNA behind RNAPII [99, 196], RNAPII may backtrack to free up the 3’ end of the RNA, thereby promoting its degradation by the 3’ → 5’ exosome [98]. There exists a SUMO modification site on SETX that allows it to interact with the exosome, suggesting it may recruit both XRN2 and the exosome complex to the nascent RNAs during the unwinding of R-loops [200]. In mammalian cells, studies of proximal termination using the transactivator response element (TAR) of HIV [97] showed that the transcription of TAR led to the recruitment of Drosha to cleave the nascent RNA by RNA interference, triggering the entry of XRN2, SETX, and the exosome (Rrp6). Thus, the RNA interference pathway, 5’ → 3’ XRN2, and 3’ → 5’ exosome also cooperate to regulate the removal of RNAPII following promoter-proximal pausing [97].

1.2.3 R-loops as a cause of genome instability

An R-loop is a triple-stranded nucleic acid structure that forms when the nascent RNA invades the DNA and pairs with the template DNA strand as soon as the RNA exits RNAPII [201]. It is known that GC content and high transcription activity at a locus lead to R-loop formation more readily [91, 201-204]. Defects in various pathways of RNA processing have been shown to contribute to R-loop accumulation, which precipitates genome instability and the consequent development of various diseases, including cancer [205, 206]. In fact, proper performance of many aspects of RNA biogenesis and processing, including transcription elongation, transcription termination, mRNA splicing, polyadenylation, RNA packaging/export, and degradation are important for the prevention of genome instability [205, 206].

The formation of R-loops could induce DNA damage through the generation of mutations on the coding DNA strand or by altering the expression of proteins needed for maintaining
genome stability. R-loops increase the likelihood of DNA breaks by creating unpaired ssDNA, which is thought to be less stable and prone to damage. Proteins like Cytidine Deaminase that can recognize R-loop structures may also be recruited to initiate DNA mutagenesis, although this protein has limited expression in mammalian cells. A recent study showed that the R-loops that accumulate in the absence of RNA processing factors such as the spliceosomal helicase Aquarius, SETX, or Topoisomerase I, lead to DNA double-strand breaks (DSBs) as a consequence of the function of transcription-coupled nucleotide excision repair (TC-NER) endonucleases (XPF/ERCC4, XPG/ERCC5). Also involved in the repair process is the TC-NER factor CSB/ERCC6, but not the global genome repair pathway protein, XPC. How involvement of the TC-NER pathway in the process of R-loop removal generates DSBs that then cause formation of γH2AX foci is not clearly understood.

Often, head-on collisions between R-loops created by the RNAPII transcription machinery and DNA replication forks cause DNA damage and genome instability. According to this model, as the stalled replication fork collapses, the consequent repair of the DSBs by homologous recombination (HR) tends to be mutagenic and contributes to transcription associated mutagenesis and recombination. Studies have shown that mutants with increased R-loops also contain elevated DNA hyper-recombination and chromosome loss. In yeast screens, it was observed that mutants affecting different steps of transcription and RNA processing show such phenotypes; these screens identified mutations in genes that code for the cleavage and polyadenylation factors (Rna14, Rna15, Pcf11, Fip11), exosome (Rrp6), the DNA:RNA helicase Sen1, and the THO/TREX complex needed for RNA packaging and export. An siRNA screen in...
mammalian cells based on monitoring the γH2AX mark characteristic of DSBs also identified related RNA-processing factors, including heterogeneous nuclear ribonucleoproteins (hnRNPs), splicing factors, polyadenylation factors (i.e. FIP1L1), and nuclear pore complex components [216, 217].

1.2.4 The roles of R-loops in transcriptional regulation

Genome-wide studies using antibodies (S9.6) to precipitate DNA:RNA hybrids have mapped many R-loop prone loci, showing widespread R-loop formation over CpG island promoters and 3’ end termination regions [91, 201-203]. Genes with high transcription frequency (e.g. rDNA for ribosomal RNA synthesis) [204] and regions associated with overlapping antisense transcripts that are suggestive of head-on collisions between transcriptional machineries [218] are also predisposed to form R-loops.

R-loops at CpG island promoters were shown to be correlated with protection against DNA methylation (by the DNMT3B1 methyltransferase) and transcriptional silencing [203]. These include many promoters with high GC skew (which tend to form thermodynamically stable G-rich RNA and C-rich DNA duplexes) [202]. R-loops in G-rich termination regions facilitate RNAPII pausing prior to termination in processes that are especially important for preventing read-through for genes that are closely positioned in tandem or in gene-dense regions [91, 202]. It is not clear how promoter associated R-loops are related to promoter-proximally paused RNAPII; it seems more likely that these R-loops may be byproducts of paused initiation, rather than being the signal for RNAPII to pause proximally. Recently, it has also been shown that UV induced DNA damage can stall RNAPII and lead to R-loop accumulation that then directly activates a non-cannonical ATM signaling pathway for the DNA damage response.
Thus, far from being harmful transcriptional byproducts, R-loop structures may have acquired positive evolutionary roles for the regulation of RNAPII transcription initiation, elongation, and termination [205, 206].

R-loops can also lead to chromatin repression via the recruitment of chromatin remodeling proteins, and they are known to be associated with repressive marks such as histone H3S10 phosphorylation and H3K9 methylation [199, 220]. As R-loops can lead to genome heterochromatization, increased R-loops near a gene may be detrimental to gene expression. Importantly, examples related to several neurological and neurodegenerative diseases have been identified: the trinucleotide repeat expansion of the FMR1 promoter that generates R-loops can silence the gene and cause fragile X syndrome [221, 222]; the trinucleotide repeat expansion in the gene body of the Frataxin gene generates R-loops to cause premature termination on the gene, leading to Friedreich’s ataxia [221, 223]; and the formation of truncated transcripts due to R-loop formation upon hexanucleotide repeat expansion in the C9ORF72 gene may result in ALS [224]. It may, therefore, be critical that the levels of R-loops be properly controlled so that RNAPII transcription functions optimally without creating heavy mutational burdens that are detrimental to the cell.

1.2.5 Surveillance pathways for R-loop suppression

There exist multiple surveillance pathways for R-loop elimination in cells [206, 225-227]. First, there are two RNase H enzymes in eukaryotes. RNase H1, in particular, is known to degrade the RNA in DNA:RNA hybrids during DNA replication and repair, and its over-expression has been employed in many R-loop related studies as a phenotype rescue [206]. Secondly, the core spliceosome and splicing factors (e.g. ASF/SF2, SC35, hnRNPs, the
spliceosomal helicase Aquarius) have been shown to suppress R-loops and prevent genome instability, possibly because genes in higher metazoans often contain long introns that tend to form R-loops readily if they are not properly spliced out [211, 216, 217, 219, 226, 228, 229].

Thirdly, mechanisms involving mRNA export are also important for suppressing R-loops. The THO/TREX complex that functions in mRNA packaging and export has been shown to affect transcriptional elongation and cause R-loop formation and hyper-recombination when mutated [225, 230, 231]. These factors may prevent R-loop accumulation by facilitating proper packaging of mRNA into ribonucleoprotein (RNP) complexes, thereby partitioning them from the DNA[201]. Fourth, the well conserved DNA Topoisomerases (TOP) are known to relax DNA supercoiling that otherwise causes R-loop formation; in the absence of TOP1, DNA supercoils and R-loops tend to build up behind the elongating RNAPII [204, 232, 233]. Another example is TOP3B, which is recruited to active promoters by histone arginine methylations recognized by TDRD3; the resulting complex then reduces R-loops at promoters to enhance transcription activation [227, 232].

In eukaryotes, R-loops can be unwound by several helicases, including the Pif1 DNA helicase, the spliceosomal RNA:DNA helicase Aquarius, the 3’→5’ DNA:RNA helicase DHX9, and the 5’→3’ DNA:RNA helicase Sen1/SETX [91, 201, 211, 215, 234, 235]. Notably, both SETX and DHX9 are known to interact with SMN, a protein whose mutation causes SMA [166, 234]. As well, the mammalian SETX is known to interact with RNAPII and certain RNA 3’ end processing factors such as the polyA-binding proteins 1 and 2 (PABP1/2), several hnRNPs, and Sap155/SF3B1 [166]. As was mentioned in the RNAPII termination model, SETX was proposed to resolve R-loops in termination regions to free the 5’ end of the nascent RNA for XRN2 degradation, leading to the release of RNAPII [91, 202].
SETX is similarly needed to resolve R-loops created by the collision between RNAPII and DNA replication forks [206], and it may also coordinate DNA repair in such an event, as it is known to be recruited to DNA repair foci [236-238]. Recently, it was found that SETX recruits the BRCA1 complex for DNA repair at transcription termination sites where R-loops accumulate [238]. Mutations in SETX are known to cause neuronal disorders such as ataxia oculomotor apraxia 2 (AOA2) and early onset ALS (ALS4) [91, 166, 239, 240], and these mutations either cause premature translational termination or interfere with the function of the helicase or N-terminal protein interaction domains of SETX [241, 242]. Because the DNA:RNA helicases unwind R-loops for the RNAs to be degraded by the exonucleases such as XRN2 and the exosome, the RNA exonucleases are therefore also known components in R-loop repression [243].

The last pathway that may eliminate R-loops involves the aforementioned RNA interference pathway, in which R-loops associated with paused RNAPII induce antisense transcription to recruit the Dicer-Ago complex, leading to heterochromatization of the locus[199]. R-loop-induced heterochromatization is also known to occur at centromeric repeats, as ncRNAs transcribed from these regions remain bound to DNA as R-loops, triggering similar formation of heterochromatin [244].

1.2.6 Transcription coupled nucleotide excision repair (TC-NER)

As mentioned in the previous section, recent studies showed that the nucleotide excision repair (NER) machinery can act to remove R-loops and, therefore, release the stalled RNAPII [211]. NER is a pathway conserved from bacteria to human that corrects various DNA lesions such as UV-induced damage and bulky structural changes [245]. NER can be divided into
transcription-coupled repair (TC-NER) and global genome repair (GG-NER) that differ in their lesion recognition steps. In GG-NER, specific proteins recognize the helix distortion, whereas in TC-NER, the stalled RNAP promotes the recruitment of the repair factors [245, 246]. Defective TC-NER is associated with autosomal-recessive disorders such as Cockayne syndrome, and patients often display increased UV-sensitivity, increased tumor frequency, premature aging, and progressive neurodegeneration [245, 247].

In TC-NER, CSA/ERCC8 and CSB/ERCC6 assemble the TC-NER complex through the recruitment of factors such as TFIIH (including the helicases XPB/ERCC2 and XPD/ERCC3), XPA, the endonucleases XPF/ERCC4-ERCC1 and XPG/ERCC5, and the scaffold protein XAB2 to the stalled RNAPII [246, 248, 249]. CSB is a SWI2/SNF2 helicase that can remodel chromatin to increase the accessibility of the DNA lesion to the repair complex [250]. CSB also recruits the histone acetyltransferase p300 and the nucleosome binding protein HMGN1 to loosen the nucleosome behind the RNAPII [249, 251]. CSA, like BRCA1, is a ubiquitin ligase that can ubiquitinate CSB to deactivate it [252]. Upon DNA damage, UVSSA interacts with the TC-NER complex, RNAPII, and the deubiquitinase USP7 to stabilize the complex by the deubiquitination of CSB [253-255].

DNA damage that stalls RNAPII either triggers the backtracking and recycling of the RNAPII for the entry of the TC-NER complex or leads to the degradation of the RNAPII POLR2A subunit [246]. Commonly, RNAPII backtracks from the lesion while degrading the nascent RNA through its 3’–5’ exonuclease activity that is activated by TFIIS [256]. The repair complex remains at the site to maintain the denatured DNA, as the XPF–ERCC1 endonuclease incises at the 5’ side of the bubble and XPG incises on the 3’ side to remove the lesion [211, 257, 258]. PCNA and the DNA polymerases fill in the gap, using the opposite DNA strand as the
template, and DNA ligation is carried out by the ligase I or III [257, 258]. CSA then facilitates resumption of transcription by removing and deactivating the TC-NER repair factors through ubiquitination [247]. If TC-NER fails to repair the lesion, transcriptional arrest can result in the degradation of RNAPII, as an E3 ubiquitin ligase adds K48-linked polyubiquitin chains to POLR2A [95, 96], followed by its extraction from the chromatin by the VCP ATPase and degradation by the proteasome [259]. There exist several potential ubiquitin ligases for POLR2A, including CSA [95], BRCA1/BARD1 [260, 261], WWP2 [262], NEDD4 [96], and the elongin E3 ligase complex [259, 263, 264], although not all ubiquitination reactions result in POLR2A degradation. It is also conceivable that RNAPII may be removed by the cooperative action between the 5’ → 3’ XRN2 nuclease and the 3’ → 5’ exosome. If not removed, the stalled RNAPII will block transcription and even trap additional polymerases, which will inactivate gene expression from the damaged allele [246].

G-rich DNA tracts that lead to R-loops are frequently found at promoters, termination regions, telomeres, centromeres, immunoglobulin hypermutation sites, and in the trinucleotide expansions that are responsible for various hereditary diseases [221-224]. Such tracts contribute to RNAPII pausing and induce mutations, and their R-loops are known to be stabilized by knock-down of components of the TC-NER pathway such as CSB, TFIIS, XPA, XPB, and XPD [211, 265, 266]. It is likely that aberrant structures such as R-loops that impede RNAPII movement directly trigger TC-NER in the absence of DNA damage [245], though it is still not clear how the repair pathway causes DSBs upon clearing of the R-loops.
1.3 Proteinarginine dimethylation

1.3.1 Types of arginine methylation and cross-talk involving this mark

There exist nine protein arginine methyltransferases (PRMTs) conserved in mammalian cells that catalyze monomethylation (MMA) or asymmetric (Rme2a) or symmetric (Rme2s) dimethylation [267] (Fig. 2B). PRMTs have been shown to play regulatory roles in signal transduction, the DDR, gene expression and splicing, and their mis-regulation has been linked to cancer and various other diseases [267]. PRMT1 and PRMT5 are the major PRMTs catalyzing Rme2a and Rme2s, respectively, and the complete loss of either protein in an animal is lethal [268, 269]. The type 1 PRMTs (1-4, 6, and 8) are known to catalyze the Rme2a reaction, and the type II PRMTs (5 and 9) catalyze the Rme2s [267, 270] modification. MMA is the reaction intermediate for both types of PRMTs, and it is the end point for PRMT7 [271, 272]. It is not uncommon for PRMTs to compete for the deposition of me2a and me2s modifications on the same substrate[273]. For example, PRMT5 antagonizes PRMT1’s activating histone H4R3me2a mark with the repressive me2s marks on histones H4R3 and H3R8 [274-276], and an unknown PRMT antagonizes PRMT6’s repressive histone H3R2me2a mark with an activating me2s mark [277, 278]. As well, alternative modifications of SPT5 R698 by PRMT1 and PRMT5 are known to regulate its role in RNAPII elongation [279], and modifications of the splicing factors CA150, SmB, U1c, and SF3B49/Saf49 are known to occur by both the CARM1 and PRMT5 methyltransferases [280]. PRMT1 and PRMT5 also compete for methylation of the transcription factor E2F1, leading to two functional consequences: formation of Rme2a by PRMT1 augments E2F-1-dependent apoptosis, whereas formation of Rme2s by PRMT5 favors proliferation; upon cyclin A binding to E2F-1, it becomes preferentially methylated by PRMT5 [281].
Arginine methylations exhibit multiple levels of cross-talk with other modifications, such as phosphorylation and acetylation, thereby altering the kinetics of protein-protein interactions. One example involves methylation of the RNA binding protein (RBP) SAM68 by PRMT1, which prevents its interactions with the SH3 domain binding proteins (FYN, LCK and ITK), while permitting its interaction with WW domain containing proteins [282]. Another example involves the EGFR receptor activation that is blocked by PRMT5 methylation at R1175, which promotes EGFR Y1173 phosphorylation to recruit the SHP1 phosphatase to inhibit downstream ERK signaling [283]. A third example involves transcription factor C/EBPβ, which can be methylated by CARM1, and yet the interaction is abolished upon the phosphorylation of C/EBPβ in RAS signaling [284]. Arginine methylation also antagonizes the AKT phosphorylation of substrates such as BCL-2 and FOXO [285, 286]. FEN1 exo/endonuclease methylation by PRMT5 (at R192) is antagonized by its phosphorylation (at S187), and only methylated FEN1 interacts with PCNA for its localization to DNA replication and repair foci for cell cycle regulation [287]. As for RNAPII R1810 methylation by CARM1, it was found to be blocked in vitro by either pSer2 or pSer5 modification on the CTD [67].

With respect to histone arginine methylation, histone H3K18 acetylation can prime H3R17 for dimethylation by CARM1 [288] and blocks H3R8 dimethylation by PRMT5, which would lead to gene repression [275]. Histone H4K5 acetylation can also make H4R3 be a better substrate for PRMT5 than for PRMT1, thereby turning the activating Rme2a mark into a repressive Rme2s-mediated response [289]. MLL4, a transcriptional co-activator that trimethylates histone H3K4, needs first to bind to histone H4 through its PHD domain [290]. The H4R3me2s modification blocks MLL4 from binding, thus reducing the level of the H3K4me3 mark for gene activation [290]. Another example comes from PRMT6, which
represses transcription through its methylation of H3R2, which blocks the recruitment of MLL for H3K4 trimethylation that signals for gene activation [277, 291, 292]. Many effectors that bind H3K4me3 are blocked from docking if H3R2me2a is present, and binding of certain effectors to HeK4me3 is enhanced when H3R2me2s (presumably modified by PRMT5 or 9) is present in conjunction with H3K4me3 [270, 278, 291, 293, 294]. Antibodies developed to recognize the dual H3R2me2sK4me3 mark show that it is present at high levels and is conserved from yeast to human [278, 294]. H3R2me2s can interact with the WDR5 subunit of MLL through hydrophobic interactions [278], and thus the dual mark may prevent the demethylation of H3K4me3 and preserve the activation mark for long periods [294].

1.3.2 PRMT4/CARM1 in gene regulation

Of the 9 PRMTs, I will focus my discussion on PRMT4 and PRMT5 with respect to gene regulation, due to their relevance to my study. CARM1/PRMT4 catalyzes Rme2a modification with a preference for a PGM motif in a diverse set of substrates, including histones, transcriptional factors and co-regulators, splicing factors, the RNAPII CTD, and many other proteins [267, 280]. Therefore, CARM1 participates in the regulation of a wide range of cellular processes, including transcription, splicing, cell signaling, cell cycle, and the DDR [67, 267, 280].

CARM1 was first described as a transcriptional activator that associates with the p160 steroid receptor co-activators (SRC-1, GRIP1/TIF2, and p/CIP) in transcription [295], and later it was found to be a promoter-specific regulator for NF-κB -dependent gene expression, because of its interaction with the p65 subunit of NF-κB [296]. From these studies, it became clear that transcriptional activation by CARM1 was not limited to its methylation of transcription factors
but also involved its methylation of the histones H3R17 and R26 [291, 295, 296]. More recently, CARM1 was also found to methylate the transcription factor PAX7 to enable it to form a complex with H3K4 methyltransferase MLL for gene activation during myogenesis [297]. CARM1 is also known to methylate the transcription factor RUNX1, leading to the assembly of a repressor complex that blocks the myeloid differentiation of human stem/progenitor cells, leading to acute myelogenous leukemia [298].

Though CARM1 over-expression can activate a number of oncogenic pathways (FOS, E2F1, WNT/β-catenin, AIB1, and others), CARM1 is also known to function as a checkpoint for cell cycle regulation [267, 299]. CARM1 methylates the histone acetyltransferase p300/CBP for its activation by auto-acetylation [300], and amongst the multiple sites for p300 methylation, methylation of R754 is required for the induction of the expression of cell cycle regulators (p21 and GADD45) for cell cycle arrest and the DDR [299, 301, 302]. The R754me2a mark is known to uniquely interact with and recruit the BRCT domain of the tumor suppressor BRCA1, though this domain family is well known for phosphate binding [301, 303].

Another class of substrates that is favored by CARM1 is involved in RNA processing and splicing. These proteins include PABP1, HuD, HuR, CA150/Tcerg1, SAP49/SF3B4, SmB, and U1c [267, 272, 280, 304]. The HuR protein is known to bind to AU-rich mRNAs to increase their stability, and the presence of methylation on HuR correlates with stability of these mRNAs [305].

1.3.3 PRMT5 in gene regulation

Although PRMT5 is ubiquitously expressed, its level can be regulated at the transcript level by miRNAs [276, 306], through its binding to the Heat Shock 90 (Hsp90) protein for
stabilization [307], or phosphorylation by JAK2, which reduces its activity [308]. There exist many PRMT5 substrates, including various transcriptional factors and chromatin-associated proteins, and PRMT5 is known to be a transcriptional co-activator and co-repressor [267]. The PRMT5 substrates include E2F1, p53, GATA4, FEN1, HBX-A9, NF-κB, and RAD9, as well as the SPT5 subunit of the RNAPII elongation factor DSIF [279, 309-315]. As a PRMT5 substrate, the Rme2s modification on E2F-1 can alter its ability to suppress cell growth and apoptosis [314]. PRMT5 methylation of the transcription factor GATA4 inhibits its activity and blocks the p300-mediated acetylation of GATA4 in cardiomyocytes [309]. PRMT5 also methylates NF-κB p65 on multiple Arg residues to promote its activation of gene expression after TNF-α-mediated induction [310, 311]. PRMT5 methylation of the transcription factor HOXA9 is important for the cytokine-induced expression of pro-inflammatory genes such as E-selectin and VCAM-1 [312]. RAD9 is also known to be methylated by PRMT5, a process that is critical for RAD9 regulation of the S/M and G2/M cell cycle checkpoints [315]. And in the DDR, p53 is methylated by PRMT5 at R333/R335/R337, and a triple R-to-K mutant lacks the ability to arrest the cell cycle at G1 [313].

Similarly, methylation of histones by PRMT5 can lead to both activating and repressive forms of histone marks [267]. PRMT5 is thought to be a transcriptional repressor through generating the H3R8me2s and H4R3me2s histone marks [274-276] via its interaction with the transcriptional repressor BLIMP1 [316]. In embryonic development, the PRMT5-dependent repressive H4R3me2s modification on transposons is needed for their silencing during epigenetic reprogramming [317]. Under the regulation of transcription repressor PAX2/GRG4, the H4R3me2s modification is required for the subsequent recruitment of the PRC2 complex for gene silencing [318]. H4R3me2s may also serve as a binding site for DNMT3A for DNA
methylation and gene repression [319]. PRMT5 is also known to maintain an active euchromatic mark through the symmetrical dimethylation of H3R2, which excludes repressive complexes such as Sin3a, NURD, and PRC2, and enhances the interaction with WDR5, a component found in co-activator complexes such as MLL, SET1A, SET1B, and so on [278].

WDR77 (MEP50) is the best known adaptor for PRMT5, and it is also a regulatory subunit for PRMT5 substrate specificity [267]. Additional adaptors for PRMT5 include pICln, which stimulates WDR77/PRMT5’s activity on the Sm proteins in spliceosome assembly [320, 321]; RIOK1, which exclusively recruits WDR77/PRMT5 for the methylation of nucleolin [322]; and the SWI/SNF chromatin remodeler, which recruits WDR77/PRMT5 to histone substrates [275]. Another adaptor that alters substrate selectivity for PRMT5 is COPR5, which is known to bind to histone H4 to guide WDR77/PRMT5 to methylate histone H4R3 more preferentially than H3R8 [323]. Recently, the Mediator complex subunits CDK8 and CDK19 that function in activating transcription by RNAPII were also found to recruit the PRMT5/WDR77 methylosome complex [324].

Besides methylating Arginine in GAR (glycine, alanine rich) or PGM (proline, glycine, and methionine rich) motifs [267], PRMT5 is known to methylate arginines in other non-canonical motifs, such as histone H3R8 (ARK), EGFR R1199 (LRV), and the Golgi protein Gm130 R6/18/23 (PRL, PRP, TRQ) [267, 273, 275, 325-328]. Because this involves a change of its substrate selectivity, it is likely that different adaptors are needed for different situations. Since the RNAPII CTD residue R1810 is embedded in the sequence PRY, making it a plausible substrate of PRMT5, we suspect that a known or unknown adaptor for PRMT5 has yet to be characterized for PRMT5 recruitment to R1810 on the CTD.
1.3.4 Arginine methylation readers: the Tudor domain containing proteins

Methylated arginine-containing motifs are binding sites for protein readers that contain Tudor (Tud) domains, and the mammalian genome contains more than 30 Tud domain-containing proteins (TDRDs) [329]. Based on sequence and structure, Tudor domains are related to other chromatin-binding proteins with chromo, MBT, PWWP or Agenet domains that recognize methyllysine. Together they are termed the Royal Domain superfamily [330]. Members of this family possess a barrel fold comprised of 3 to 5 antiparallel β-sheets [329, 330] that form 2 to 4 aromatic residue-containing pocket. This pocket provides electrostatic and hydrophobic contacts to accommodate the methylated ligand [331]. Methylation increases the size and hydrophobicity of the arginine, thus enhancing the non-electrostatic (cation-π) contacts between the aromatic cage of the domain and the cationic carbon of the methylarginine [331].

TDRD proteins are specific for methylarginine or methyllysine binding, with the aromatic cages for methylarginine binding being narrower [331]. Methyllysine binding Tud domains are commonly linked to other chromatin binding domains and are involved in histone modification and chromatin remodeling [329]. These include three lysine demethylases (KDM4A/B/C), a histone lysine methyltransferase SETDB1, several chromatin binders that also contain PHD domains for methyllysine binding [329], and TP53BP1, which is involved in the DDR through its binding to the histone H4K20me2 modification [332]. The TDRDs that recognize methylarginine include SMN, SPF30, SND1, and the extended Tudor domain (eTud) containing proteins TDRD1-12. Of these, the Tud domain in SMN is the ancestral fold upon which the eTud domain that preferentially binds to Rme2s in RG repeats evolved an additional α-helix and two β-strands at the N-terminal and several helices and strands at the C-terminal [329, 333].
TDRDs have important roles in many cellular processes, including RNA metabolism, germ cell development, transposon silencing, the DDR, histone modification, and chromatin remodeling [267, 329]. The germline specific TDRDs (TDRD1-12 other than 3) contain the eTud domain and are involved along with PRMT5 and the PIWI endonucleases in the regulation of transposon silencing through a germline specific RNA interference (piRNA) pathway [329, 334-338]. PRMT5 participates in the assembly of the Nuage (an electron-dense perinuclear structure) through the methylation of the PIWI nucleases and helicases that function in the pathway for transposon cleavage [334, 336, 337]. The Rme2s modifications on these proteins are read by the eTud TDRDs that tend to contain multiple eTud domains in tandem [333, 335-337]. Through multiple interactions, a condensed protein-RNA aggregate forms. Many eTud TDRDs also contain RNA-binding domains (e.g. KH, Zinc finger) and helicase domains, presumably to enhance the trapping and processing of transposon RNAs in the Nuage [329].

1.3.5 Methylarginine binding TDRDs and the Arginine demethylase in gene regulation

Many TDRDs participate in transcription regulation. One example is TDRD3, which is recruited to transcription start sites (TSSs) by the histone H3R17me2a and H4R3me2a marks that are generated by CARM1 and PRMT1, respectively [227]. Thereafter, it was shown that TDRD3 in complex with the TOP3B topoisomerase enhances transcription initiation through unwinding R-loop structures at the TSSs [232]. The authors also suggested that the RNAPII CTD R1810me2a mark might recruit TDRD3 for transcription activation and for the recruitment of additional regulators [67, 232]. TDRD3 also harbors a UBA domain, and thus it could also be directed to specific genomic loci where ubiquitin signals are enriched, presumably upon genotoxic stress [291]. Recently, the TDRD3-TOP3B heteromer was shown to interact with the
fragile X mental retardation protein (FMRP) complex to facilitate active translation, especially in neurons, and, in this case, the topoisomerase activity of TOP3B is important for the translational activation [339, 340]. So far, TDRD3 is the only Tud domain containing protein that preferentially binds to the Rme2a modification, whereas the other TDRDs prefer the Rme2s modification (though the Tud domain in SMN also has a weak affinity for Rme2a) [227, 280, 341]. A structural study of TDRD3 bound to the RNAPII CTD R1810me2a mark showed that a tyrosine (Y566) residue in its Tud domain is critical for the Rme2a specificity, and when this residue is altered to tryptophan, the specificity for Rme2a is lost [68].

SPF30/SMNDC1 is the Tud containing protein sharing the most homology with SMN (~50% identity), and it also binds preferentially to Rme2s containing motifs [342, 343]. SPF30 is a component of the spliceosome [344], and it is essential for spliceosome assembly from the U4/U5/U6 tri-snRNPs [345]. Another protein with a similar Tud fold is the SND1/TDRD11 protein, which harbors a single Tud domain followed by four repeats of staphylococcal nuclease-like domains [291]. SND1 functions as a transcriptional co-activator with transcription factors such as EBNA2, STAT6 and E2F1 [281, 346, 347], and often these interactions are mediated by arginine methylation [347]. SND1 can facilitate the acetylation of histones for gene activation, leading to a spreading of euchromatin [348]. SND1 is also known to regulate mRNA splicing and processing, as it interacts with various RBPs, including the RISC complex for miRNA processing, the Sm proteins for spliceosome assembly, and SAM68 for mRNA splicing [349-351].

An earlier report suggests JMJD6 may be the demethylase for H4R3me2a and H4R3me2s [352], but JMJD6 was later shown to be a lysine-hydroxylase for the splicing factor U2AF65 [353]. Additional structural studies of JMJD6 also suggested that it is not an arginine
demethylase [354, 355]. More recently, JMJD6 was shown to be an arginine demethylase that is recruited by BRD4 to distal enhancers to erase the H4R3me2s repressive mark [356]. Via long-range interactions, JMJD6 allows RNAPII to be released from the promoter-proximal pause as H4R3 demethylation releases the 7SK inhibition of p-TEFb [356]. Therefore, JMJD6 may be a candidate demethylase for the RNAPII CTD R1810me2a and me2s modifications.

1.3.6 The Rme2s reader protein SMN and its role in protein complex assembly

SMN is a 38 kD protein that is both nuclear and cytoplasmic, and its Tud domain is composed of ~60 amino acids that form a barrel structure via 4 antiparallel β-strands (Fig. 2C) [331, 342]. SMN associates with GEMINs 2–8 and UNRIP to form a multimeric complex (the SMN complex) that is predominantly localized to the nuclear GEM bodies [357-359]. SMN can self-aggregate through its N-terminal K-rich domain and C-terminal YG box that allow it to cluster into oligomers, ranging from dimers to octamers [360]. Consequently, SMN can serve as an adaptor for the assembly of multiple Rme2s-containing proteins.

In eukaryotes, RNA splicing involves 9 snRNPs that form the spliceosome to target the 5’ and 3’ splice sites, as well as and the branch points on introns [361-368]. Each snRNP contains one snRNA and various accessory proteins, and all snRNPs except U6 contain 7 common Sm core proteins (SmB/B’, D1, D2, D3, E, F and G) that form a heptameric ring around the U-rich Sm sites on snRNAs [369-373]. Sm (SmB/B’, SmD1 and SmD3) proteins first interact with the PRMT5/WDR77 methylosome and the adaptor pICln for the methylation of their C-terminal RG-rich region prior to binding to SMN [320, 363, 374, 375]. Thereafter, the SMN complex binds both snRNAs and the Sm proteins for their assembly in an ATP-dependent manner [320, 368, 371].
In addition to SMN’s role in the assembly of snRNPs involved in splicing, SMN assembles other ribonucleoprotein (RNP) complexes, including the LSM complex that regulates mRNA decay, stabilization, and transport [376]. SMN also assembles the small nucleolar ribonucleoproteins (snoRNPs) that are important for methylation and pseudouridylation of rRNAs, tRNAs, and snRNAs [377]. And SMN is also known to interact with RNAPII in a complex with an RNA helicase DHX9, though the exact function of this complex has not been elucidated [234].

SMN also interacts with many RBPs that are known to regulate axonal mRNA transport, stability, and local translation in neurons, such as for the β-actin mRNA [378-380]. These complexes also include RBPs such as hnRNP U, hnRNP Q, FMR, TDP-43, ATXN2, FUS, and EWSR1 that are known to be mutated in neuronal diseases such as fragile X syndrome, ALS, and FTD [378, 381-387]. As SMN serves as an assembly factor for the spliceosome, and is also present in cytosolic RNPs that transport mRNAs, these two functions have been proposed to contribute to the SMA disease phenotype when SMN is mutated [388].

1.3.7 Proposed roles of arginine methylation and SMN in RNAPII termination

Complete loss of SMN is lethal in yeast and embryonic lethal in mammals, and in particular, mutations in SMN1’s oligomerization and Tud domains cause the SMA phenotype [383, 389]. Analogously to its role in snRNP assembly through binding to Rme2s modified Sm proteins, I propose that SMN acts similarly to assemble a transcription termination complex on the RNAPII CTD through binding to CTD R1810me2s, as well as Rme2s modified termination factors and 3’ end mRNA processing proteins.
SMN is known to interact with Senataxin [166], a DNA:RNA helicase that is mutated in AOA2 and ALS4 and is important for transcription termination by RNAPII [91]. Factors involved in 3’ end transcript processing and RNAPII termination often contain methylated arginine (me2s and/or me2a), and such modifications may allow them to interact with Tud domain-containing proteins such as SMN. Rapid advances in affinity purification and mass spectrometry (AP-MS) have allowed the demonstration that many RNAPII transcription and termination factors contain dimethylated Arg in their GAR or PGM motifs [390-394]. These include XRN2, three subunits of CPSF (1, 5, and 6), CstF2/64, three PolyA-binding proteins (PABP1, 2, and 4), RBBP6, WDR33, PCF11, SPT5, CTDP1, DHX9, TDP-43, the FET proteins (FUS, TAF15, and EWSR1), and many others. Among these candidate SMN interactors, I examined the methylarginine containing proteins FUS and TDP-43 for their roles in RNAPII termination. I hypothesize that SMN may dock multiple arginine-methylated factors to assemble 3’ end transcription processing and RNAPII termination machineries on the RNAPII CTD, the disruption of which may contribute to neurodegenerative disorders such as SMA and ALS/FTD.

1.4 Transcription and RNA processing related mis-regulations in neurodegeneration

SMA together with ALS and FTD are some examples of neurological disorders associated with genes that play important roles in RNA metabolism [388, 395-401]. SMA is caused by homozygous mutations or deletions in the SMN1 gene, and is one of the leading causes of infant mortality due to progressive spinal motor neuron loss, muscle denervation, axonal degeneration, and death from respiratory failure [382, 388, 402]. Although SMA is caused by reduced levels of a ubiquitously expressed protein, its primary symptom first occurs in
the lower motor neurons [388]. ALS is similar to SMA symptomatically but is adult-onset and characterized by rapidly progressive paralysis and death from respiratory failure with ~90% of all cases being sporadic [403]. There exist a diverse set of mutations contributing to ALS pathology that include mutations in SOD1, TDP-43, FUS, OPTN, VCP, UBQLN2, C9ORF72, GRN, ANG, and PFN1 [403-414]. FTD is a common form of dementia caused by a progressive loss of neurons from the frontal and temporal cortex, thus leading to a decline of mental ability [398]. Increasing understanding suggests that ALS and FTD are related diseases at the molecular level due to the mutation of the same genes [398, 403].

1.4.1 The pathology of Spinal muscular atrophy (SMA)

While the human genome contains two copies of the SMN gene (SMN1, SMN2), the mutation of one copy of SMN1 can lead to a sufficient decrease of the SMN level to cause the SMA disease phenotype, with the severity of the disease reflecting the insufficiency of the remaining SMN for its full range of functions [388]. SMN2 differs from SMN1 by a C-T nucleotide change in exon 7, generating mostly an incorrect splice isoform with only 10% of the SMN2 transcripts producing functional SMN protein [382, 383, 415]. It has been suggested that SMN plays a dual function in neurons by favoring the assembly not only of the spliceosome, but also of RBPs and their target transcripts in regulated axonal mRNA transport and translation [388, 416].

Structure studies have shown that SMN Tud domain missense mutations that contribute to the SMA disease prevent interaction of SMN with Rme2s containing peptides [331, 417]. The SMN Tud domain residue Glu134 has been shown to stabilize the aromatic cage through hydrogen bonding [331, 417], and the SMA E134K mutation abolishes the binding of SMN to
Rme2s peptides [331]. The SMA W92S and Q136E mutations cause the Tud domain to misfold, also disrupting its ability to bind to dimethylated arginine [331, 418]. As SMN can oligomerize in vitro and in vivo, mild SMA mutations can produce proteins that still associate with the wild-type SMN to form oligomers, whereas severe mutations such as the loss of the last exon disrupt SMN’s ability to oligomerize [383, 419, 420]. When SMN fails to oligomerize, it is degraded, as was found in the case of the SMA Y272C mutation [383, 419, 421, 422]. Collectively, disrupted SMN Tud domains and the loss of the SMN C-terminal oligomerization domain are often found to occur in SMA patients [378, 383-386].

In the SMA disease model, defects in spliceosome assembly and stability and genome-wide alternations of gene expression and pre-mRNA splicing (i.e. changes in exon inclusion for CHODI, USPL1, HIF3A, COL5A1, and UACA) have been observed in expression and exon array analyses [423, 424]. One study observed splicing changes in 259, 73, and 633 genes from spinal cord, brain, and kidney, respectively, in SMA animals [423]. Though defective spliceosome assembly and splicing are seen in the SMA models [423-427], it is still difficult to conclude whether these changes that occur late in the disease progression of the animals represent a secondary effect of cell injury or are a direct result of the lack of SMN [424]. Recently, it was shown that SMN deficiency perturbs U12-dependent minor spliceosome (U11, U12, U4atac/U6atac, and U5) assembly [423] and decreases the expression of U12 intron-containing genes in mammalian cells and in flies, leading to the mis-splicing of a target gene (Stasimon) that is needed for motor circuit function [426, 427]. These findings indicate that defective splicing of critical neuronal genes due to SMN deficiency may be important for the selective pathology of SMA in motor neurons [427].
Because the spliceosome assembly in spinal cord from mice with severe or mild mutations of SMN was similar [425], it was suggested that SMN’s functions in the assembly of other RBPs also contribute to SMA [382]. SMN is known to associate with cytoskeletal proteins and RBPs in mRNA transport, stability, and local translation [382]. In SMA mice, a significant reduction in the levels of β-actin and neuritin mRNA and protein translation at the axon in motor neurons was observed [379, 428]. The SMN-associated actin bundling protein Plastin 3 acts as a protective modifier of SMA [382, 429], and its overexpression rescues axon outgrowth defects in cultured neurons from SMA animals. Therefore, SMN’s contribution to RNP assembly for regulated transport and translation is also compromised in SMA.

1.4.2 Low Complexity (LC) Domains and the pathogenesis of ALS-FTD

Aggregations of RBP-containing granules have been observed in a wide range of cell types and organisms, forming distinct structures such as the Cajal bodies, PML bodies, GEM bodies, paraspeckles, splicing speckles, and nuclear factories that organize events such as rRNA biogenesis, snRNP assembly, sno/snRNA modifications, RNA splicing, and RNAPII transcription [430, 431]. In the cytoplasm, analogous RNP structures similarly form, including the worm germ cell P granules, fly germline Nuage, fly embryo polar granules, stress granules from yeast to human that form upon nutrient deprivation, and neuronal granules that shuttle mRNAs along dendrites for translation at the synapses [430, 431]. Many proteins within these bodies contain RNA-binding KH or RRM domains or RNA helicase activities, as well as low complexity (LC) domains [430, 431]. The LC domain is both necessary and sufficient for a reversible phase transition between a soluble state and an amyloid-like granule state, and LC domains are known to be enriched in RNA- and DNA-binding proteins [431-433]. Similarly,
LC domains also precipitate the formation of pathological aggregates or inclusion bodies that are characteristic of neurodegenerative disorders such as ALS/FTD and Alzheimer’s, Parkinson’s, and Huntington’s diseases [395, 396, 434, 435].

The addition of biotinylated isoxazole, a drug that triggers differentiation of progenitor cells into mature neurons, was found to precipitate hundreds of RBPs with LC domains, and many of these RBPs are implicated in neurodegeneration, including FMR1, FXR1/2, FUS, TDP-43, EWSR1, TAF15, ATXN2 and hnRNPA1 [430, 431]. One evolutionarily conserved form of LC domain that occurs predominantly in the aforementioned RBPs is the RGG/RG repeats that are known to interact with RNA secondary structures (G-quartets), leading to their increased stability [436-438]. RGG/RG repeats are also suitable substrates for methylation by PRMTs (as PRMT1 and 5 are known to methylate GAR-rich residues) [267, 280], and this process is often required for the nuclear localization of many of these proteins, including nucleolin, FUS, SERBP1, and SAM68 [436, 439, 440]. One possibility is that the addition of methyl groups creates increased fluidity in the LC domains that would otherwise pack into a gel-like solid phase, and that nuclear transport may consequently occur by a passive movement.

Condensed gel-like structures can form from the LC domain of FUS in vitro, and they are capable of retaining proteins with the FUS LC domain (homotypic trapping) or an LC domain derived from other RBPs (heterotypic trapping) [433]. The FUS LC domain may function as a transcriptional activation domain through similar interaction with the CTD of RNAPII in its unphosphorylated state, forming a polymeric fiber [433, 441]. This provides a mechanistic basis for how proteins such as FUS could participate in the regulation of RNAPII initiation, as the CTD switches from a hypophosphorylated form to a hyperphosphorylated form upon activation, which frees the CTD from a gel-like state [433].
The abnormal expansion of C9ORF72 is a signature found in familial ALS (~35-40% of cases) and FTD (~25% of cases) [414], providing a genetic basis for two clinical outcomes. C9ORF72 expansion is a dominant mutation in which the hexanucleotide GGGGCC expands from 2-23 copies in normal individuals to 700-1600 copies in the disease state [414]. Abnormal foci (containing sense or antisense transcripts) are observed in C9ORF72 patient cells, potentially implicating in disease disrupted RNA metabolism due to the sequestration of RBPs into inclusion bodies [403]. C9ORF72 expanded transcripts also form stable G-quadruplex RNAs, which may re-hybridize with DNA to form R-loops, thus affect genome stability, RNA transcription, splicing, translation and transport [442]. Recently, it has been shown that the sense and antisense transcripts of an expanded C9ORF72 locus are also translated into dipeptide repeats (poly-GA, -GP, -GR) [443], which then act like LC domains to trap RBPs such as TDP-43 [404, 414, 444-446], subsequently leading to cell death as a sign of neurodegeneration [443, 447].

Other proteins implicated genetically in ALS seem heterogeneous in character [448] (Table 1). One example is mitochondrial SOD1, which contributes to ALS pathology through the formation of inclusion bodies, rather than the loss of its dismutase activity [449-451]. Therefore, mutations in LC domains that induce protein aggregation seem central to the pathological state of ALS/FTD, as well as other similar neurodegeneration conditions. For my study related to ALS/FTD, I will focus on RBPs such as FUS [410, 412] and TDP-43[411, 452] that are known to regulate RNAPII transcription and co-transcriptional splicing [453, 454].
<table>
<thead>
<tr>
<th>ALS Mutated Gene</th>
<th>Frequency</th>
<th>Proteins trapped in cytoplasmic inclusion bodies</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>TDP-43</td>
<td>FUS</td>
</tr>
<tr>
<td>C9ORF72</td>
<td>35%</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>SOD1</td>
<td>20%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRN</td>
<td>low</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>ANG</td>
<td>low</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>UBQLN2</td>
<td>low</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>PFN1</td>
<td>low</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>VCP</td>
<td>low</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>hnRNPA1/B1</td>
<td>low</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>SQSTM1</td>
<td>low</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>ATXN2</td>
<td>low</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>OPTN</td>
<td>low</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>TDP-43</td>
<td>low</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>FUS</td>
<td>low</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>TAF15</td>
<td>low</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>EWSR1</td>
<td>low</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>SETX</td>
<td>low</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1 Genes implicated in the pathology of ALS
1.4.3 The cellular function of TDP-43

A major improvement in the understanding of ALS and FTD came from the discovery that the TDP-43 protein is trapped in ubiquitin-positive neuronal inclusions in many familial forms and almost all spontaneous forms of disease, also supporting the hypothesis that ALS and FTD are related diseases [403, 404]. The mutations in TDP-43 itself that cause disease are autosomal dominant in ALS (~4% of familial cases) and FTD [403, 404] and occur mostly in the glycine-rich prion-like LC domain in the C-terminus, a region also involved in RNA-binding and splicing [396, 403, 455]. Mutant TDP-43 is more prone to form aggregates when compared to the wild type (possibly due to an increased tendency to solidify), and may trap the wild type TDP-43 and other LC domain-containing proteins in the aggregates, leading to a depletion and loss of their function in the nucleus [398, 456]. TDP-43 is found trapped in cytoplasmic inclusions in many forms of ALS/FTD, including those caused by mutation in GRN [404], ANG[405], UBQLN2 [406], PFN1 [407], and VCP [408]. It has become clear that TDP-43-mediated neuronal loss is not caused by toxic gain of function by cytoplasmic aggregates that may interfere with cellular transport, but rather by a loss of its normal function in the nucleus [398], as similar TDP-43 loss-of-function studies in flies, fish, and mice also showed axonal loss, impairment of locomotion and synaptic transmission, and embryonic lethality [398, 457-460].

TDP-43 gain-of-function studies have also been carried out in diverse species, showing a tight control of TDP-43 expression is much needed, which is partially achieved by its auto-regulation through RNA degradation [461, 462]. Overexpression of TDP-43 also phenocopies the ALS/FTD disease state that is characterized by neuronal degeneration and the accumulation of ubiquitin-positive TDP-43 aggregates [398, 404, 463, 464]. TDP-43 neurotoxicity upon its over-expression occurs only when its RNA binding domain is included, suggesting that RNA
metabolism rather than aggregate formation alone is compromised in the disease state [465, 466]. As TDP-43 can oligomerize, the mutant may sequester the wild type proteins, the bound RNAs, and other interactors such as SMN in a dominant-negative manner, also depleting them from their site of function [398, 467, 468]. Through the identification of RNAs immunoprecipitated with TDP-43, it was found that >6,000 RNAs expressed in brain could be targeted by TDP-43 [396, 401], which is known to preferentially bind to long clusters of UG rich sequences in introns [400]. Splicing arrays revealed that the levels of >60 mRNAs (including FUS, Progranulin and other neuronal disease associated transcripts) were changed, and >900 splicing events (including Sortilin, the receptor for Progranulin) were also altered, upon the depletion of TDP-43 from mouse adult brains [400, 401]. Like many splicing factors including PTBP1, hnRNP F, HuR, MBNL, p54nrb/PSF, and PABPN1 that are implicated in altering mRNA3’ end formation [173], TDP-43 is also known to directly compete with CstF for UG rich sequences downstream of the PAS motif, thus regulating the site of polyadenylation [461].

1.4.4 The cellular function of FUS

FUS contains a transcriptional activation domain, a Gln-Gly-Ser-Tyr-rich region, a Gly-rich prion-like LC domain, three RGG repeats, RNA/DNA binding domains, and a C-terminal nuclear localization signal (NLS) [395]. FUS mutations account for a small percentage of familial cases (~4%) of ALS, and mutations that cluster in its LC and NLS regions cause FUS to be trapped in cytoplasmic inclusions[469, 470]. FUS and TDP-43 are known to interact with each other and are localized to the nucleus [469-472]. Besides FUS, two related RBPs of the FET protein family (TAF15 and EWSR1) also tend to accumulate in cytoplasmic aggregates in ALS/FTD [473], and they have the same domain structure as FUS and are known to interact with
FUS [395, 474, 475]. In FTD, EWSR1 and TAF15 co-accumulate in FUS-positive cytoplasmic inclusions, showing a nuclear reduction in all three proteins [473, 476]. The formation of the pathological inclusion bodies is thought to be affected by arginine methylation, which also regulates the nuclear import of proteins such as FUS [471, 477]. FUS aggregation may also contribute to the SMA phenotype, as SMN can interact with FUS and may be trapped in FUS containing inclusion bodies through Rme2s binding [399, 478, 479].

FUS is known to be an RNA- and DNA-binding protein that can regulate gene expression [480-482] through its association with various transcription factors and RNAPII [441, 453, 483]. Mouse fibroblasts with FUS knock-outs show elevated chromosomal instability and radiation sensitivity [484, 485], which is suggestive of possible transcription induced R-loop accumulation (see below). FUS also plays important roles in splicing regulation and RNA maturation [486-488]: it is known to bind to the Tau pre-mRNA in brain, and the loss of FUS leads to a Tau splicing defect that is reminiscent of certain pathological states [487]. Another study using mouse neurons showed that FUS tends to bind to stable RNA secondary structures around alternatively spliced exons that are associated with neurodegeneration (e.g. the MAPT, CAMK2A, and FMR1 transcripts) [481]. Genome scale studies found that FUS binds to RNAs from >5,500 genes in brain, with a preference for a GUGGU motif and with a sawtooth-like binding pattern [454, 482]. Depletion of FUS from the adult brain altered the levels or splicing in >950 mRNAs, most of which are distinct from TDP-43-affected splicing patterns, but as well contain long introns that code for proteins essential for neuronal integrity [454, 482]. The loss of FUS has been linked to the appearance of intron retention in 3-5% of transcripts, including ones that are important for developmental processes such as gastrulation and mesodermal differentiation (e.g. the FGF8, FGFR2, and CDH2 RNAs) [489]. A recent study reveals that
FUS can regulate RNAPII transcription termination, and can also bind to the 3’ end of mRNAs to alter the sites of polyadenylation through the recruitment of CPSF1/160 [490].

Using animal models such as fly, worm, fish, and rodents, FUS loss-of-function showed pathological phenotypes such as reduced viability, shortened lifespan, impaired locomotion, disrupted neuromuscular junctions, and degeneration of motor neurons similar to the ALS disease state[491-494]. As well, many studies have described neurodegeneration phenotypes that are associated with over-expression of wild-type or mutant FUS [491, 494-498]. Like the ALS-causing TDP-43 mutations, ALS-causing FUS mutations are autosomal dominant and may function to sequester the wild type FUS, other RBPs, and RNAs into the inclusions, leading to disrupted RNA metabolism overall [476, 479, 499, 500].

1.4.5 Proposed roles of transcription induced R-loops and the consequent DNA damage as the cause of neurodegeneration

Once formed during development, post-mitotic neurons need to maintain genome integrity for proper function for the rest of the organisms’ lifetime. DNA damage and subsequent cell death are known to be elevated in disorders such as SMA, ALS/FTD, and Alzheimer’s and Parkinson’s diseases, though the underlying mechanism is unclear [501-505]. It has been shown that DNA damage escalates around the promoters of various genes, leading to their reduced expression in aged cortex, and that these genes are often essential for synaptic plasticity, vesicular transport, and mitochondrial function [506]. As well, natural increases in neuronal activities can lead to DSBs in neurons even in young mice, and this situation is exacerbated in the Alzheimer’s model, leading to elevated synaptic dysfunction [507].

Multiple DNA repair pathways function in the nervous system for the repair of various types of DNA lesions including DNA single- or double-stranded breaks, DNA intra-strand
crosslinks, and helix distortions [508, 509]. The disruption of each of the repair pathways is known to be associated with neuropathology, including defects in neurodevelopment, microcephaly, brain tumors, and neurodegeneration [508, 509]. Unlike neural progenitor cells, the differentiated post-mitotic neurons only rely on the error-prone non-homologous end joining (NHEJ) pathway for the repair of DSBs with the involvement of DNA ligase LIG4 for the direct ligation of the broken ends [508, 509]. The more accurate homologous recombination (HR) repair pathway for DSBs only occurs in the S and G2 phases of the mitotic cell cycle in dividing cells, as it requires the sister chromatid as the repair template [509, 510].

DSB repair signaling first occurs through the detection of the breaks by the MRN complex (MRE11, RAD50, and NBS1), which also activates and recruits PI3 kinases for the phosphorylation of the histone H2A variant X at Ser139 (resulting in γH2AX) at the break sites [229, 511, 512]. In neurons, three PI3 kinases, ATM, ATR, and DNA-PK, are known to be activated for the phosphorylation of H2AX in the vicinity of the lesions upon the detection of DSBs (ATM and DNA-PK) and ssDNAs (ATR) [509, 513]. It is estimated that these kinases phosphorylate more than 700 proteins as result of DDR, including p53, NBS1, BRCA1, CHEK1, and CHEK2 [514]. γH2AX is recognized by MDC1, which coordinates the DDR through the recruitment of various proteins such as ATM, TP53BP1, and CHEK2 for DNA repair, cell cycle arrest, or apoptosis [509, 513]. Mutations of DNA repair and signaling components including ATM, MRE11, NBS1, ATR, LIG4, and BRCA2, are known to cause neurodegeneration [509].

Of the genes implicated in ALS, it known that a dominant mutation of SETX causes ALS4 [515]. In a recent study, mice with disrupted ATM, TDP1, SETX, or APTX genes, which cause Ataxia Telangiectasia, Spinocerebellar Ataxia with Axonal Neuropathy 1, AOA2, and AOA1, respectively, were examined for R-loop formation. Indeed, R-loops were found to be
enriched in proliferating cells such as the testes, but were not found in brain tissues that correlated with high levels of DNA damage and apoptosis [516]. This result is surprising [517], as R-loops are known to occur in post-mitotic neuronal cells, and are known to be associated with neuronal disorders [518]. As well, transcription induced R-loops are implicated in DNA damage in neurons, and have been shown to contribute to neurodegeneration in patient cells [519].

Another link of DNA damage to ALS comes from studies of the RBP FUS. FUS, which stands for Fused in Sarcoma, was shown to be important for genomic stability, as FUS<sup>−/−</sup> mice suffer from high levels of genomic instability, defective B-lymphocyte development, and sterility [484, 485]. FUS is also a component of the DDR, as it localizes to sites of DNA damage upon UV stress [520], and motor cortexes of ALS patients with FUS mutations were also shown to contain elevated DNA damage compared to control samples [520, 521]. These studies suggested that dysfunction of FUS may contribute to the progression of both ALS and cancer, due to transcription induced R-loop formation and the consequent DNA damage. A third link between R-loops, DNA damage, and ALS comes from the observation that upon hexanucleotide expansion of C9ORF72 in ALS/FTD that led to over-production of non-functional transcripts, elevated R-loops and DNA damage were also observed [522]. As well, mutations of several subunits of RNase H2 and Topoisomerase1 (TOP1) also support the link between R-loop accumulation, increased genome instability, and neurodegeneration [523, 524].

In summary, after the loss of RBPs such as FUS and TDP-43 or the RNA:DNA helicase SETX, transcription induced R-loops and DNA damage are elevated and therefore may be implicated in neurodegenerative disease, but additional experiments are needed to solidify such a model. I therefore propose that elevated R-loop accumulation that occurs following the loss of
interactions of SMN, SETX, FUS, and TDP-43 with RNAPII CTD R1810me2s during RNAPII transcription termination may be of critical importance for neurodegeneration.
Chapter 2: Symmetric arginine dimethylation of the RNAPII CTD residue R1810 mediates regulation of transcription termination by SMN and SETX

“SMN and Symmetric Arginine Dimethylation of RNAP II CTD Control Termination”
Dorothy Yanling Zhao, Gerald Gish, Ulrich Braunschweig, Yue Li, Zuyao Ni, Frank W. Schmitges, Guoqing Zhong, Ke Liu, Weiguo Li, Jason Moffat, Masoud Vedadi, Jinrong Min, Tony J. Pawson, Benjamin J. Blencowe, Jack F. Greenblatt. *Nature* (accepted)

Author contributions
Dr. Gerald Gish prepared the FITC peptides. Dr. Frank Schmitges performed WT v. R1810A RNAPII ChIPseq experiments. Dr. Ulrich Braunschweig performed computational data analysis for ChIP-seq and RNA-seq. Vectors for shRNAs were provided by Dr. Jasson Moffat, and Guoqing Zhong generated stable MAPLE tagged and shRNA-mediated knock-down cell lines. Dr. Zuyao Ni generated CRISPR knock-out for SMN in HEK293. Dr. Ke Liu provided the purified Tudor domains and performed Fluorescence Polarization assays. Dr. Weiguo Li performed Isothermal Titration Calorimetry assays. Dr. Masoud Vedadi provided the purified PRMT5/WDR77 complex.
Chapter 2: Symmetric arginine dimethylation of the RNAPII CTD residue R1810 mediates regulation of transcription termination by SMN and SETX

1. Introduction

The CTD of the mammalian RNAPII subunit POLR2A contains 52 heptapeptide repeats, the N-terminal half containing mostly consensus heptads (Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7) and the C-terminal half many more heterogeneous repeats [16]. The CTD is a platform for modifications that specify the recruitment of factors for transcription regulation, mRNA processing, and chromatin remodeling. In vivo, these repeats can be phosphorylated on Tyr1, Thr4, and all three Ser residues (Ser2, Ser5, and Ser7) [1, 5]. Ser5 phosphorylation by TFIIH correlates with transcription initiation and recruits capping enzymes for RNA 5’-ends and termination factors for snRNAs and snoRNA genes [2, 16, 26, 27]. Ser2 phosphorylation by P-TEFb, CDK12/13, and BRD4 downstream of the promoter is important for RNAPII escape from promoter-proximal pause sites and interactions with enzymes needed for mRNA 3’ end formation, transcription termination, and histone H3K36 methylation in transcribed regions [2, 16, 28, 32, 33]. Ser7 and Ser5 phosphorylations recruit the Integrator complex that processes snRNAs [525, 526], whereas Tyr1 phosphorylation stimulates association of the elongation factor Spt6 and impairs recruitment of yeast termination factors Nrd1, Pcf11, and Rtt103 [59]. Thr4 phosphorylation is required specifically for histone mRNA 3’ end processing [63].

Two non-consensus human CTD Arg residues, R1603 and R1810, are conserved in vertebrate species. In the 31st repeat of the human CTD, the Ser7 position is replaced by an arginine residue (R1810) that was shown to be asymmetrically dimethylated (Rme2a) by the CARM1/PRMT4 methyltransferase [67]. It was found that the R1810me2a modification of the CTD in human inhibits the expression of snRNA and snoRNA genes [67], and it was also shown
that this R1810me2a mark can be bound by the Tudor domain of TDRD3 in vitro [67]. I now show that this CTD R1810 residue is also symmetrically dimethylated (me2s) by the arginine methyltransferase PRMT5 in vitro and in vivo. This particular modification on the RNAPII CTD enables the CTD to interact with the Tudor domain of the SMN protein in vitro, and enhances the recruitment of SMN to the transcribing RNAPII in vitro. Also observed was the enhanced interaction between POLR2A and the DNA:RNA helicase SETX, which is likely mediated by the CTD R1810me2s modification and the recruitment of the adaptor protein SMN.

Pausing by RNAPII at termination sites is known to be associated with the formation of DNA:RNA hybrid structures (R-loops) when the nascent RNA anneals to the template DNA strand [91]. SETX has been shown to enhance RNAPII termination by resolving R-loops to facilitate the entry of the XRN2 exonuclease [91]. I further showed that, like SETX, the RNAPII CTD R1810me2s modification, PRMT5, and SMN are needed for R-loop resolution and the proper release of RNAPII at the 3’ ends of genes. The problem with R-loop accumulation that stalls RNAPII at the 3’ ends of genes is also observed in the SMA disease state. Lastly, persistence of the R-loops is known to lead to DNA damage and genome instability, as has been shown to be the case when SETX is depleted [238]. When the CTD R1810me2s-SMN-SETX pathway becomes compromised, I also observed an increase of DNA damage (as marked by the level of γH2AX) in regions where unresolved R-loops and stalled RNAPII accumulate. We therefore propose that transcription induced R-loops and the consequent DNA damage that occurs following the loss of the recruitment of SMN and SETX to RNAPII during termination may contribute to the pathology of neurodegeneration.
2. Results

2.1 The identification of R1810 symmetric dimethylation on the POLR2A CTD

I performed co-IP using anti-FLAG antibody and FLAG-tagged TDRD3 or the POLR2D subunit of RNAPII and observed that both tagged proteins could pull down POLR2A with the Rme2a modification (ASYMM24 antibody) as detected by western blotting, whereas only the precipitation of POLR2D pulled down POLR2A with an Rme2s modification (SYMM10 antibody) (Fig. 3A). This strongly suggested that the Rme2s modification on POLR2A was likely to be on the same CTD residue, R1810 [67], as the Rme2a modification. To determine whether the Rme2s modification actually involved R1810, Raji cells stably expressing α-amanitin-resistant, HA-tagged, wild type or R1810A mutant POLR2A were used [67]. After treating these cells with α-amanitin (2ug/ml for 3 days) to deplete endogenous RNAPII, HA-tagged RNAPII was pulled down using anti-HA antibody, and blotted with R1810me2a antibodies [67] and Y12 antibodies (for Rme2s) to reveal that the R1810A mutation causes loss of both modifications (Fig. 3B).

That IP of TDRD3 from HEK293 whole cell lysates (WCL) retrieved RNAPII with the R1810me2a mark, but not RNAPII with R1810me2s (Y12), indicated that there exist distinct dimethylation marks on R1810 that presumably recruit different readers (Fig. 3A). This result also indicated that R1810me2a is recognized by TDRD3 in vivo, as was previously shown in vitro [67], although the TDRD3 ChIP pattern and the effect of TDRD3 on the expression of snRNA and snoRNA genes indicate that it does not mediate the effects of the R1810me2a modification on snRNA/snoRNA genes [67, 227].
Figure 3 The detection of symmetric dimethylation of R1810 on the RNAPII CTD
Figure 3 The detection of symmetric dimethylation of R1810 on the RNAPII CTD

A. POLR2A carries Rme2a and Rme2s modifications. Whole cell lysates (WCL) from HEK293 cells stably expressing FLAG-tagged TDRD3 or the POLR2D subunit of RNAPII were used for immunoprecipitation (IP) using beads conjugated with M2 anti-FLAG antibody, and the precipitates were western blotted with the indicated antibodies. Cells expressing FLAG-GFP were used as negative control. Precipitated TDRD3 and POLR2D contained POLR2A with the Rme2a modification (ASYMM24 antibody), whereas precipitated POLR2D, and not TDRD3, contained POLR2A with the Rme2s modification (SYMM10 and Y12 antibodies).

B. IP with the indicated antibodies from Raji whole cell lysates (WCL) stably expressing the HA-tagged wild-type or R1810A mutant POLR2A upon α-amanitin treatment that abolishes the endogenous POLR2A. Anti-HA was used to precipitate HA-tagged wild-type (WT) or R1810A mutant POLR2A, followed by western blotting with the indicated antibodies. Relative quantification of the western blots is shown under the blots.

C. Slot blots illustrating that the Y12 and R1810me2s antibodies specifically recognize peptides containing RNAPII R1810me2s. The amounts labeled on the right indicate the biotin-labeled 7mer CTD peptides bracketing R1810 with no modification, Rme2a, and Rme2s that were blotted before incubating the membrane with the R1810me2s or Y12 antibodies for detection.

D. IP with the N20 antibody for POLR2A from the HEK293WCL followed by treatment with phosphatase or phosphatase inhibitor to show that the detection of POLR2A by Y12 and the R1810me2s antibodies improved upon alkaline phosphatase treatment. The blots indicate that the antibody recognition of the epitope can be blocked by nearby phosphorylated residues.
2.2 Characterization and use of an R1810me2s specific antibody

We made a rabbit polyclonal antibody against a 7mer peptide that contained CTD residue R1810 with an Rme2s modification, and the serum was purified by passing it through columns that were conjugated with CTD R1810 peptides carrying either no modification or an R1810me2a modification to enrich for the fraction with R1810me2s binding specificity [67]. The specificity of the R1810me2s and the Y12 antibodies was tested with 7mer peptides containing R1810 with no modification, Rme2a, or Rme2s, as illustrated in the dot blots of Fig. 3C: they recognize a peptide bracketing CTD R1810me2s but bind much less well to peptides without methylation or with Rme2a. The specificity of the R1810me2s antibody in the detection of POLR2A in whole cell lysate or in a POLR2A IP was further shown in western blots, as the detected bands were abolished upon α-amanitin treatment of the cells, which causes the degradation of POLR2A (Fig. 4A). Moreover, the R1810A mutation in POLR2A reduced its detectability by the R1810me2 antibody in western blots (Fig. 3B). However, the R1810me2s antibody detects additional non-specific bands that are not abolished by growth of the cells in α-amanitin, showing that it does have cross-reactivity with additional proteins in the cell.

Notably, the detection of POLR2A by both Y12 and the R1810me2s antibodies improved upon phosphatase treatment of immunoprecipitated RNAPII, presumably because phosphate groups present on residues near R1810 interfere with the binding between the antibodies and the epitope (Fig. 3D). Therefore, all experiments involving the use of R1810me2s or the Y12 antibodies with immunoprecipitated RNAPII samples were conducted with phosphatase treatment of the samples. The R1810me2s antibody was also shown to be capable of precipitating POLR2A in IP experiments (Fig. 4B), and the specificity of the antibody for the
Figure 4 Further characterization of the RNAPII CTD R1810me2s antibody
Figure 4 Further characterization of the RNAPII CTD R1810me2s antibody

**A.** The R1810me2s antibody can detect POLR2A in WCL and IP samples
IP with the N20 antibody for POLR2A from the HEK293 WCL with and without the α-amanitin treatment that abolishes the endogenous POLR2A, followed by western blotting with the indicated antibodies (labeled below the blots). The R1810me2s antibody is shown to recognize POLR2A in WCL and in N20 IP, as well as non-specific bands of smaller sizes that are not abolished by the α-amanitin treatment.

**B.** The R1810me2s antibody can immunoprecipitate POLR2A
The R1810me2s antibody was used for IP for POLR2A from the HEK293 WCL, followed by western blotting with R1810me2s or N20 antibodies for POLR2A. The R1810me2s antibody is shown to be able to pull down POLR2A, and the interaction is weakened by the pre-incubation of the antibody with R1810 containing peptide with the Rme2s modification, but not with Rme2a or no modification, prior to the IP reaction.
R1810me2s modification was also demonstrated by pre-incubating the antibody with R1810 containing peptides that contained no modification or the me2a or me2s modification. Only with incubation of the antibodies with R1810me2s peptides did the binding of the antibody to POLR2A become weakened (Fig. 4B).

2.3 The identification of PRMT5 as the likely methyltransferase for R1810 in vivo

PRMT5 is the major enzyme that symmetrically dimethylates arginine in vivo and has many substrates that play roles in transcription and gene regulation [279]. I performed IP with antibodies against POLR2A and SMN, and observed that PRMT5 co-precipitated with both proteins (Fig. 5A). Therefore, to test whether the methylosome, which is comprised of PRMT5/WDR77, may be the enzyme that symmetrically dimethylates R1810 in vivo, stable knock-down cell lines for CARM1, PRMT5, and GFP were generated, and endogenous RNAPII was precipitated (with N20 or 4H8 antibodies). Western blotting revealed that the stable knock-down of CARM1 caused a reduction of the R1810me2a mark, whereas the stable knock-down of PRMT5 only caused a reduction of the R1810me2s mark (Fig. 5B). With the R1810me2s and Y12 antibodies, upon transient knock-down of PRMT5 by siRNAs, similar reduction of the R1810me2s modification was seen after precipitation of equal amounts of POLR2A across samples (as detected by the N20 and 8WG16 antibodies) (Fig. 5C, 9F), whereas over-expression of FLAG-tagged PRMT5 increased the R1810me2s signal on POLR2A (Fig. 9E). These experiments indicated that PRMT5 is required in vivo for the R1810me2s modification on the RNAPII CTD.
Figure 5 Symmetric dimethylation of R1810 on the RNAPII CTD by PRMT5 in vivo
**Figure 5 Symmetric dimethylation of R1810 on the RNAPII CTD by PRMT5 *in vivo***

A. IP with the indicated antibodies from HEK293 WCL, followed by western blotting with the indicated antibodies. The blots showed that POLR2A and SMN interact with each other, and with PRMT5, and the termination factors SETX and XRN2. Also shown is that POLR2A contains the R1810me2s modification that is recognized by SMN.

B. IP with the 4H8 antibody for POLR2A from HEK293 WCL after stably knocking down of the indicated proteins followed by western blotting with the indicated antibodies. Relative quantification of the western blots is shown under the blots. The knock-down of CARM1 led to a loss of R1810me2a signal, and the knock-down of PRMT5 led to a loss of R1810me2s signal. The efficiency of CARM1 knock-down is confirmed in qPCR shown below.

C. IP with the N20 antibody for POLR2A from HEK293 WCL after transient knock-downs of PRMT5 followed by western blotting with the indicated antibodies. Relative quantification of the western blots is shown under the blots. Knock-down of PRMT5 led to a loss of R1810me2s signal as detected by both Y12 and R1810me2s antibodies, whereas N20 and 8WG16 antibodies detected equal amount of POLR2A precipitated in each sample.
2.4 Methylation of R1810 on the RNAPII CTD by PRMT5 in vitro

Because the requirement in vivo of PRMT5 for the symmetric dimethylation of R1810 could be indirect, it was important to test whether PRMT5 could directly methylate R1810 in vitro. To test whether PRMT5 can directly dimethylate the RNAPII CTD on arginine, I incubated the recombinant methylosome complex, PRMT5/WDR77, in vitro with \(^3\)H-SAM and either recombinant GST-N-CTD containing CTD repeats 1-29 that includes R1603, or GST-C-CTD containing CTD repeats 24-52 that includes R1810. Scintillation proximity assay (SPA) was employed to assay for \(^3\)H-SAM incorporation following glutathione pull down of the GST proteins after the methylation reactions. SPA revealed that GST-C-CTD and GST-N-CTD are both methylated well above the background generated by GST alone (which contains 14 arginine residues), PRMT5 alone, or \(^3\)H-SAM alone (Fig. 6A). When the assays were performed on biotinylated 13mer peptides that contained CTD residues R1810 or R1603, methylation was observed on both R1603 and R1810 in SPA (Fig. 6B) and by mass spectrometry (not shown) [527]. Endogenous RNAPII immunoprecipitation followed by the methylation assay with PRMT5/WDR77 and \(^3\)H-SAM in vitro also showed that POLR2A is a PRMT5 substrate as detected using fluorography (Fig. 6C).

PRMT5 is best known for methylating proteins with Arg embedded in GAR (glycine and arginine rich) or PGM (proline, glycine, and methionine rich) motifs [267, 528]. It is possible that both R1063 and R1810 of CTD may be PRMT5 substrates in vivo, as the RNAPII CTD residues R1603 and R1810 are embedded in the sequences PRS and PRY, respectively. Besides the well-known canonical PRMT5 substrates that include the SmB, D1 and D3 proteins present in various snRNPs, non-canonical sites of PRMT5 methylation that include histone H3R8
Figure 6 Symmetric dimethylation of R1810 on the RNAPII CTD by PRMT5 \textit{in vitro}
Figure 6 Symmetric dimethylation of R1810 on the RNAPII CTD by PRMT5 \textit{in vitro}

A. Quantification of methylation with $^3$H-SAM of the indicated GST-CTD fusion proteins, as shown in the Coomassie-stained SDS gel of the insert, with the indicated concentrations of recombinant PRMT5/WDR77 labeled on the X-axis. Error bars denote s.e.m. (n=3).

B. Quantification of methylation with $^3$H-SAM of 13-mer CTD peptides with the indicated concentrations of recombinant PRMT5/WDR77 labeled on the X-axis. Error bars denote s.e.m. (n=3).

C. Quantification of methylation with $^3$H-SAM following incubation of the endogenous POLR2A (immunoprecipitated with the N20 antibody from HEK293 WCL) with the indicated concentrations of PRMT5/WDR77 \textit{in vitro}. The Coomassie-stained SDS gel below shows the amount of POLR2A present in each lane.
(ARK), EGFR R1199 (LRV), GM130 R6/18/23 (PRL, PRP, TRQ) are known to exist [267, 273, 275, 325-328].

2.5. Recognition of RNAPII CTD R1810me2s by SMN in vitro

As dimethylated Arg residues are often recognized by the Tudor (Tud) domain containing proteins [274, 529], I reasoned that an important clue about the function of CTD R1810me2s might be obtained by identifying its protein binder in vitro and in vivo. Because the Tudor domains of SMN and SPF30, as well as TDRD1, 2, and 4-12 (extended Tudor domains), have specific affinity for Rme2s [529], we used them for fluorescence polarization (FP) binding assays to identify which Tudor domains could bind an FITC-13mer CTD peptide that contains dimethylated R1810. Only the SMN Tud domain exhibited binding (Kd~ 130uM for R1810me2s), with a preference in the order R1810me2s > R1603me2s ~ R1810me2a > R1603me2a (Figs. 7A, 7B). In contrast, the TDRD3 Tud domain showed very weak but observable affinity only for R1810me2a > R1603me2a above background (no modification or Rme2s) (not shown) [67]. Fig. 7C presents the same experiment done with the same FITC-CTD peptides (me2s, me2a, and me0) and the Tud domains of SPF30/SMNDC1, TDRD1, TDRD2, TDRD9, and TDRD11, showing a lack of binding. Therefore, the Tud domain of SMN is specific for binding the R1810me2s CTD peptide in vitro.

The CTD is phosphorylated on various types of genes and at various points in the transcription cycle on Tyr1, Ser2, Thr4, Ser5, and Ser7 [1, 5]. Such phosphorylations have the potential of strengthening or weakening the interaction of SMN with R1810me2s. Therefore, we then carried out a FP binding screen using the SMN Tud domain and nine FITC-15mer R1810me2s containing peptides, each carrying a phosphate residue in the indicated position
Figure 7 R1810me2s in the RNAP II CTD is recognized by the SMN Tudor domain *in vitro*.
Figure 7 R1810me2s in the RNAPII CTD is recognized by SMN Tudor domain *in vitro*

**A-B.** Fluorescence polarization (FP) peptide binding assays. Recombinant SMN Tudor domain was incubated with FITC-labeled 13mer CTD peptides bracketing either R1810 or R1603. SMN preferentially bound CTD peptides in the order R1810me2s > R1603me2s ~ R1810me2a > R1603me2a, and exhibited no detectable affinity for the unmodified peptides. Fluorescence polarization assays showing binding of SMN to the indicated FITC-labeled CTD peptides containing R1810 or R1603.

**C.** FP assay: FITC-CTD R1810me2s or FITC-CTD R1810me2a is not recognized by other recombinant Tudor domains from proteins such as SMNDC1/SPF30, TDRD1, TDRD2, TDRD9, or TDRD11/SND1.
Figure 8 CTD R1810me2s-SMN interaction not strongly affected by nearby phosphorylations
Figure 8 CTD R1810me2s-SMN interaction not strongly affected by nearby phosphorylations

A-B. FP assay: Recombinant SMN Tudor domain was incubated with FITC-labeled 15-mer CTD peptides bracketing R1810me2s also containing additional phosphorylated residue in the indicated position. Peptides with Y1P, S2P, or both upstream of the R1810me2s modification showed slightly preferential binding to the SMN Tudor domain when the phospho-modification(s) were present.

C. Isothermal titration calorimetry (ITC) assays showed that the recombinant SMN Tudor domain has no enhanced binding to R1810me2s containing peptides also carrying S2P or both Y1P and S2P.
We found that the presence of pTyr1 (Y1p) or pSer2 (S2p) upstream of R1810me2s slightly enhanced SMN-R1810me2s binding (Fig. 8A). Other phosphorylations near R1810 seemed to have no effect on SMN binding in vitro. We thus generated a FITC-peptide carrying both Y1p and S2p and found that the SMN-R1810me2s interaction was somewhat further enhanced (Fig. 8B). Yet upon performing Isothermal Titration Calorimetry (ITC) to measure the binding affinity between the SMN Tud domain and R1810me2s containing peptides with and without Y1p or S2p, or both, no significant effect was detected, as shown in the Kd values (Fig. 8C), thus indicating that SMN association with R1810me2s is not greatly influenced by CTD phosphorylation(s).

2.6 Recognition of RNAPII CTD R1810me2s by SMN in vivo

The specificity of SMN-R1810me2s interaction in vitro suggested that SMN might recognize R1810me2s in vivo. To test whether SMN also recognizes RNAPII with the R1810me2s modification in vivo, I first showed that SMN and RNAPII localize to similar regions in the nucleus upon immuno-staining (Fig. 9A), and then tested whether upon IP of POLR2A or SMN, each protein co-precipitated with the other. This was indeed the case (Fig. 5A) and, consistently, IP of SMN from HEK293 WCL co-precipitated endogenous RNAPII with the R1810me2s modification (Fig. 5A) and phosphorylation on Ser2 or Ser5, but not with the R1810me2a modification (Fig. 9B).

To test whether R1810 is important for the association of RNAPII with SMN in vivo, Raji cell lines expressing HA-tagged wild type (WT) or mutant (R1810A) RNAPII were used for RNAPII IP. Western blotting showed that the R1810A mutation causes loss of association of
Figure 9 The R1810me2s modification in the RNAPII CTD is recognized by the SMN in vivo
Figure 9 R1810me2s modification in the RNAPII CTD is recognized by the SMN in vivo

A. SMN localizes with active RNAPII in similar regions of the nucleus. HEK293 cells were used for co-staining of SMN (green) and POLR2A with its CTD phosphorylated on Ser2 (red). Hoechst was used to stain the nucleus (blue).

B. SMN associates with many isoforms of RNAPII. Immunoprecipitation with SMN antibody, but not control IgG, from HEK293 WCL, co-precipitated RNAPII with unmodified CTD repeats (8WG16 antibody) and CTD repeats phosphorylated on Ser5 and Ser2 as detected by western blotting, but not RNAPII with the R1810me2a modification.

C. IP with the indicated antibodies from RAJ1 WCL stably expressing the HA-tagged wild-type or R1810A mutant POLR2A upon α-amanitin treatment that abolishes the endogenous POLR2A. Anti-HA was used to precipitate HA-tagged wild-type (WT) or R1810A mutant POLR2A, followed by western blotting with the indicated antibodies. Relative quantification of the western blots is shown under the blots. The interactions of SMN, TDRD3, and SETX with RNAPII are weakened upon mutating R1810 to alanine.

D. IP with the 4H8 antibody for POLR2A from HEK293 WCL after stable knock-downs of the indicated proteins followed by western blotting with the indicated antibodies. Relative quantification of the western blots is shown under the blots. Only the knock-down of PRMT5, but not CARM1 leads to a loss of SMN interaction with POLR2A.

E. IP with the N20 antibody for POLR2A from HEK293 WCL. PRMT5 over-expression increases the R1810me2s modification and the SMN and Senataxin associations with RNAPII. Relative quantification of the western blots is shown under the blots.

F. IP with the N20 antibody for POLR2A from HEK293 WCL upon transient knock-downs of PRMT5 or SMN. Western blots were performed with the indicated antibodies. PRMT5 knock-down causes loss of R1810me2s on RNAPII (as detected by Y12 and R1810me2s antibodies), as well as reduced interaction of SMN and Senataxin with RNAPII. SMN knock-down causes reduced interaction of Senataxin with RNAPII. Relative quantification of the western blots is shown under the blots.
RNAPII with both SMN and TDRD3 (Fig. 9C). IP of RNAPII from HEK293 cells after stable knock-down of GFP, CARM1, or PRMT5 revealed that only the knock-down of PRMT5, but not CARM1, reduced co-precipitation of SMN (Fig. 9D). Consistent with this, in cells with transient siRNA-mediated knock-down of PRMT5, the interaction of SMN with RNAPII was also reduced (Fig. 9F), whereas the overexpression of PRMT5 led to increased interaction between SMN and RNAPII (Fig. 9E). Taken together, these results indicated that SMN interacts with RNAPII in a manner that depends on R1810 and its methylation by PRMT5.

2.7 SMN enhances the interaction of SETX and RNAPII

SMN can interact with SETX [166], a DNA:RNA helicase that is important for transcription termination by RNAPII [91, 166]. This suggested that R1810, PRMT5, and SMN might also be involved in RNAPII termination. I confirmed the interaction of SMN with SETX by co-IP (Fig. 10A), and the interaction persisted even when RNAPII was abolished by growing the cells for 3 days in the presence of α-amanitin (Fig. 10B). I also observed that the SMN-SETX interaction is reduced when PRMT5 is stably knocked down, indicating that the interaction is likely mediated through the Tud domain of SMN binding to an as yet unidentified Rme2s modification(s) on SETX (Fig. 10C). SETX co-precipitated with α-amanin-resistant, wild type RNAPII from α-amanitin-treated cells, and the interaction was reduced by the R1810A mutation on RNAPII (Fig. 9C), indicating that SMN stabilizes the interaction between SETX and RNAPII. Consistent with this, when SMN or PRMT5 is stably knocked down, the association of SETX with RNAPII is reduced in the POLR2A IP (Fig. 10D, 10E). Similarly, siRNA knock-down of PRMT5 or SMN also led to a loss of SETX association with POLR2A (Fig. 9F), whereas over-expression of PRMT5 appears to increase the POLR2A R1810me2s
Figure 10 R1810me2s and SMN enhance Senataxin association with RNAPII
Figure 10 R1810me2s and SMN enhance Senataxin association with RNAPII

A. IP with the SMN antibody from HEK293 WCL, followed by western blotting detection for SETX.

B. IP with the SETX antibody from HEK293 WCL followed by western blotting for SMN. The interaction persists upon 3 days of α-amanitin treatment (2ug/ml) that abolished the endogenous POLR2A.

C. IP with the SETX antibody from HEK293 WCL, upon shRNA stable knock-down of the indicated protein, followed by western blotting for SMN, indicating that SMN-SETX interaction is bridged by PRMT5 mediated methylation.

D-E. IP with the N20 antibody for POLR2A from HEK293 WCL, upon stable knock-down of CARM1, PRMT5, or SMN. Western blots were performed with the indicated antibodies. PRMT5 or SMN knock-down causes the loss of SETX interaction with RNAPII. Relative quantification of the western blots is shown under the blots.
modification, as well as stronger association with RNAPII of SMN and SETX (Fig. 9E). Together, these observations suggested that R1810, PRMT5, SMN, and SETX might function coordinately in a pathway important for termination by RNAPII.

2.8 RNAPII R1810me2s enhances the recruitment of SMN and SETX to RNAPII

To determine whether SMN and SETX associate with RNAPII during transcription in vivo, I carried out ChIP experiments. The ChIP graphs are displayed with error bars representing standard errors of the mean (s.e.m.) from at least 3 biological replicates unless stated otherwise. ChIP involving cell lines expressing α-amanitin resistant WT or R1810A mutant RNAPII was done after a 3-day treatment of the cells with α-amanitin to deplete the endogenous RNAPII. SMN and SETX ChIP were performed with antibodies against the endogenous proteins (Fig 11, 12). For cells expressing WT RNAPII, the SMN or SETX ChIP data is displayed both as percent input and as a ratio to POLR2A for the β-Actin and GAPDH genes, showing signals all the way from the promoters to the termination regions (left panels of Fig. 11A, 12A). Stable knock-down of PRMT5, but not CARM1 or GFP, caused strong reductions of the SMN ChIP signals, which also decreased in the RNAPII R1810A mutant (right panels of Fig 11A, 12A) (not shown). In these cases, the ChIP data for SMN was expressed as ratios to POLR2A, with the WT or control knock-down set to 1 at each position along the β-actin or GAPDH gene (right panels of Fig. 11A, 12A). The specificity of the SMN antibody for ChIP was validated in these ChIP experiments by using a knock-down cell line for SMN (Fig. 11A top right). These experiments indicated that R1810me2s recruits SMN to RNAPII all along the β-actin and GAPDH genes. Consistent with the observation that SMN is recruited from the promoters to the 3’ ends of
Figure 11  R1810me2s and SMN enhance Senataxin association with RNAPII on the beta-actin gene.
Figure 11 R1810me2s and SMN enhance Senataxin association with RNAPII on the β-actin gene

A. Quantification of chromatin immunoprecipitation (ChIP) data in HEK293 or Raji cells, expressed as percent input or SMN/POLR2A ratio, using the indicated primer positions for qPCR along the β-actin gene. Also shown are the relative effects of knocking down PRMT5 or SMN (top right) or mutating R1810 to alanine (bottom right). Error bars denote biological replicates s.e.m. (n=3).

B. Quantification of ChIP data in HEK293 or Raji cells, expressed as percent input or SETX/POLR2A ratio, using the indicated primer positions for qPCR along the β-actin gene. Also shown are the relative effects of knocking down PRMT5, SMN, or SETX (bottom right) or mutating R1810 to alanine (top right). Error bars denote biological replicates s.e.m. (n=3).
Figure 12  R1810me2s and SMN enhance Senataxin association with RNAPII on the GAPDH gene
Figure 12 R1810me2s enhances SMN and Senataxin association with RNAPII on the GAPDH gene

A. Quantification of chromatin immunoprecipitation (ChIP) data in HEK293 or Raji cells, expressed as percent input or SMN/POLR2A ratio, using the indicated primer positions for qPCR along the GAPDH gene. Also shown are the relative effects of knocking down PRMT5 or SMN (top right) or mutating R1810 to alanine (bottom right). Error bars denote biological replicates s.e.m. (n=3).

B. Quantification of ChIP data in HEK293 or Raji cells, expressed as percent input or SETX/POLR2A ratio, using the indicated primer positions for qPCR along the GAPDH gene. Also shown are the relative effects of knocking down PRMT5, SMN, or SETX (bottom right) or mutating R1810 to alanine (top right). Error bars denote biological replicates s.e.m. (n=3).
genes, IP of SMN co-precipitated RNAPII with pSer5 and pSer2 forms of the CTD as detected by western blotting, the pSer5 and pSer2 modifications marking the initiation and elongation phases of RNAPII, respectively (Fig. 9B).

Because SETX interaction with RNAPII in co-IP experiments decreased upon SMN knock-down or introduction of the R1810A mutation, I also used ChIP to test whether SMN and R1810 are needed for the recruitment of SETX to the β-actin and GAPDH genes. SETX ChIP in cells expressing wild type or R1810A mutant RNAPII revealed that R1810 enhances the recruitment of SETX (top right panels of Fig. 11B, 12B). Following stable knock-down of SMN or PRMT5 in HEK293 cells, ChIP on SETX revealed that SMN and PRMT5 also enhance the recruitment of SETX (bottom right panels of Fig. 11B, 12B). In these cases, the ChIP data for SETX were expressed as ratios to POLR2A, with the WT or control knock-down set to 1 at each position along the β-actin and GAPDH genes (right panels of Fig. 11B, 12B). The specificity of the SETX antibody for ChIP was validated by ChIP with a knock-down cell line for SETX (bottom right panel of Fig. 11B). I concluded from these experiments that SMN is recruited to RNAPII in a manner that depends on R1810 and PRMT5, and also that the recruitment of SETX to RNAPII depends on R1810, PRMT5, and SMN, consistent with the hypothesis that they all function in a common pathway.

2.9 SMN and R1810me2s regulate transcription termination by RNAPII

Because of the involvement of SETX in transcription termination [91, 166], and because SMN interacts with SETX [166] and is important for its recruitment to RNAPII (Fig. 11B, 12B), I investigated the importance of SMN and the R1810me2s modification for transcription
termination. RNAPII ChIP was carried out with Raji cells expressing α-amanitin-resistant wild type or R1810A mutant RNAPII 3 days after initiating treatment of the cells with α-amanitin to cause the degradation of the endogenous RNAPII. There was strong enrichment of the R1810A mutant RNAPII compared to the wild type RNAPII downstream of the cleavage sites where RNAPII pauses on the β-actin and GAPDH genes as detected by anti-POLR2A antibodies (4H8, H224, or 8WG16) (Fig. 13, 14). RNAPII similarly accumulated in termination regions when SETX was depleted (Fig. 13A), as was previously demonstrated [91]. In all of these RNAPII ChIP experiments, as well as for the R-loop DIP experiments in the following section, the intron 3 (1671) and intron 5 (2436) positions of the β-actin and GAPDH genes, respectively, were used for normalization across the samples. In Fig. 13A and 14, the primary RNAPII ChIP data are shown in the graphs on the left, whereas the ratios of the mutant or knock-downs to the controls are shown on the right, with signals from the control normalized to 1 at each position along the length of the genes.

The effect of the R1810A mutation on transcription termination in RNAPII ChIP experiments was confirmed by carrying out nuclear run-on experiments in which nuclei were incubated with BrUTP and short run-on RNAs were isolated by binding to anti-BrU antibodies [530]. Here again, the R1810A mutation led to hyper-accumulation of active RNAPII downstream of the cleavage sites on the β-actin and the GAPDH genes (Fig. 15). Consistent with these observations, stable shRNA knock-down of SMN or PRMT5 also led to the accumulation of RNAPII in the termination regions of the β-actin (Fig. 13A, 16A) and GAPDH (Fig. 14, 16C) genes in RNAPII ChIP experiments. Similar results were also seen following transient knock-down of SMN and PRMT5 with siRNAs (Fig. 16B, 16D). These results indicated that release of RNAPII in the termination regions of β-actin and GAPDH depends on
Figure 13 SMN and R1810me2s regulate transcription termination by RNAPII on the β-actin gene
Figure 13 SMN and R1810me2s regulate transcription termination by RNAPII on the β-actin gene

A. Quantification of RNAPII ChIP using POLR2A antibody (4H8, 8WG16, N20) in HEK293 (top) or Raji (bottom) cells, using the indicated primer positions for qPCR along the β-actin gene, after knocking down PRMT5, SMN, or Senataxin (top left), with GFP knock-down as a negative control, or mutating R1810 to alanine (bottom left). Normalized ratios to control POLR2A (set to 1) are displayed to the right. Error bars denote biological replicates as s.e.m. (n= 3-5).

B. Quantification of RNAPII ChIP using POLR2A antibody (4H8, 8WG16, H224) in Raji cells, using the indicated primer positions for qPCR along the β-actin gene. Error bars denote technical replicates as s.e.m (n=3).
Figure 14 SMN and R1810me2s regulate transcription termination by RNAPII on the GAPDH gene
Figure 14 SMN and R1810me2s regulate transcription termination by RNAPII on the GAPDH gene

Quantification of RNAPII ChIP using POLR2A antibody (4H8, 8WG16, N20) in HEK293 (top) or Raji (bottom) cells, using the indicated primer positions for qPCR along the GAPDH gene, after knocking down PRMT5, or SMN (top left), with GFP knock-down as a negative control, or mutating R1810 to alanine (bottom left). Normalized ratios to control POLR2A (set to 1) are displayed to the right. Error bars denote biological replicates as s.e.m. (n= 3-5).
Figure 15 R1810me2s regulates transcription termination by RNAPII as determined by the Nuclear run on assays
Figure 15 R1810me2s regulates transcription termination by RNAPII as determined by the Nuclear run on assays

A-B. Nuclear run-on experiments in which nuclei from Raji cells expressing wild type or mutant (R1810A) POLR2A 3-days after α-amanitin treatment to eliminate endogenous POLR2A were incubated with BrUTP for 30 minutes, and short run-on RNAs were isolated by binding to anti-BrU antibodies. The R1810A mutation led to over-accumulation of active RNAPII in the region downstream of the poly(A) site on the β-actin and the GAPDH genes. Error bars represent technical repeats as s.e.m. (n=3).
Figure 16 SMN and R1810me2s regulate transcription termination by RNAPII (with the error bars representing technical replicates)
Figure 16 SMN and R1810me2s regulate transcription termination by RNAPII (with the error bars representing technical replicates)

A-B. Quantification of RNAPII ChIP using POLR2A antibody (4H8, 8WG16) in HEK293 cells, using the indicated primer positions for qPCR along the β-actin gene, after stably knocking down PRMT5, CARM1, or SMN (top), with GFP knock-down as a negative control, or transiently knocking down PRMT5 or SMN (bottom). Error bars denote technical replicates as s.e.m. (n=3).

C-D. Quantification of RNAPII ChIP using POLR2A antibody (4H8, 8WG16) in HEK293 cells, using the indicated primer positions for qPCR along the GAPDH gene, after stably knocking down PRMT5 or SMN (top), with GFP knock-down as a negative control, or transiently knocking down PRMT5 or SMN (bottom). Error bars denote technical replicates as s.e.m. (n=3).
R1810, PRMT5, and SMN, as well as SETX, again consistent with the idea that R1810me2s, SMN, and SETX function together in a process important for RNAPII termination.

### 2.10 SMN and R1810me2s regulate RNAPII termination through the resolution of R-loops

It was recently shown that both the formation of R-loops behind the elongating RNAPII over G-rich pause sites downstream of cleavage signals and their resolution by the SETX helicase are important for recruitment of the 5'→3' exonuclease XRN2 and termination by RNAPII [91]. We used the monoclonal antibody S9.6 to detect R-loops in the DNA immunoprecipitation (DIP) experiment shown in Fig. 17A. Fig. 17A shows that the R-loop signal is specific for the R-loop structure, in that it is abolished upon treatment with RNase H, which is specific for the RNA in RNA/DNA hybrids. The R-loop signals tend to build up and accumulate in the 3’ ends of genes, an indication that as the nascent RNA extends its tendency to hybridize with the DNA template strand increases (Fig. 17A). Like the depletion of SETX (Fig. 17B) [91], mutation of R1810 or knock-down of SMN or PRMT5 also led to the hyper-accumulation of R-loops in the termination regions of the β-actin gene (Fig. 17B). The regions where RNAPII accumulated correlated quite well with where R-loops accumulated both for the R1810A mutant and the knock-downs of PRMT5, SMN, and SETX (Fig. 17B).

To further address the validity of R-loop detection with the S9.6 antibody, a second method with a GFP fusion construct that includes the RNase H1 R-loop binding domain (GFP-HB) was used [230]. When control and CRISPR system-mediated SMN knock-out cells (confirmed in the western blot of Fig. 18A) were used that stably express GFP-HB under G418 selection, increased accumulation of RNAPII (as detected by the N20 antibody) and R-loops (as detected by the GFP antibody) were seen in the termination regions in SMN knock-out cells
Figure 17  SMN and R1810me2s are important for resolving R-loops created by the elongating RNAPII
Figure 17 SMN and R1810me2s are important for resolving R-loops created by the elongating RNAPII

A. Quantification of DNA immunoprecipitation (DIP) with the S9.6 antibody in HEK293 cells with or without RNase H treatment prior to DIP.

B. Quantification of DIP with the S9.6 antibody in HEK293 cells with or without knocking down of SMN, PRMT5, or Senataxin, or in Raji cells after mutating R1810 to alanine. Error bars denote biological replicates as s.e.m. (n=3 to 5). Normalized ratios to control POLR2A are displayed to the right.
Figure 18 The knock-out of SMN leads to increased RNAPII and R-loop accumulations
Figure 18 The knock-out of SMN leads to increased RNAPII and R-loop accumulations

A-B. ChIP quantification of RNAPII with the N20 antibody (A), and R-loops with the GFP antibody (B), with the indicated primer positions for qPCR along the β-actin gene, after the knocking out of SMN by CRISPR, or with scramble guide RNA as a negative control. Error bars denote biological replicates as s.e.m. (n=3). Western blotting that verified the knocking out of SMN is displayed to the right.

C. Microscopy images showing that the SMN KO cells appear to be physiologically normal in comparison to the control KO on the left.
(Fig. 18). These experiments suggest that R1810me2s, PRMT5, SMN, and SETX are in a common pathway important for R-loop resolution and transcription termination. DHX9/RHA, a known 3' → 5' helicase for DNA:RNA hybrids that also associates with RNAPII and SMN and participates in the DDR, may play a similar role in the pathway [234].

2.11 R-loops induce DNA damage at the termination sites

Mutations in the SMN1 and SETX genes can lead to neurodegenerative disorders (SMA, AOA2, and ALS4). Therefore, an important issue is why R-loop accumulation in termination regions, due to the loss of SMN or SETX, may contribute to neurodegeneration. One possibility is that genome instability caused by less efficient removal of R-loops [201] may contribute to the neurodegenerations characteristic of SMA and ALS. To test whether R-loop accumulation in the β-actin gene termination region would lead to DNA damage, I used ChIP to assay the effects of the R1810A mutation, or the depletion of SMN or PRMT5, on the accumulation of the phosphorylated form of H2AX (γH2AX), using the H2AX signal as a control [238] (Fig. 19). As was previously observed following SETX knock-down [238], I observed similar accumulations of γH2AX and increased γH2AX/H2AX ratios in the termination region where RNAPII and R-loops accumulate when R1810 is mutated or SMN or PRMT5 is depleted. Therefore, the R-loops that accumulate as a consequence of defects in the RNAPII termination pathway involving R1810me2s, SMN, and SETX do lead to DNA damage that might contribute to neurodegeneration.

2.12 RNAPII termination defects and R-loop accumulation in SMA disease states

As I observed that the R1810me2s modification in the RNAPII CTD recruits SMN,
Figure 19  SMN and R1810me2s are important for preventing R-loop induced DNA damage
Figure 19 SMN and R1810me2s are important for preventing R-loop induced DNA damage

A-B. ChIP quantification of γH2AX as percent input (A) or as γH2AX/H2AX ratio (B) in HEK293 or Raji cells, along the length of the β-actin gene, after the knocking down of PRMT5 or SMN, with GFP knock-down as a negative control, or mutating R1810 to alanine. Error bars denote biological replicates as s.e.m. (n=4).
which, in turn, recruits SETX to resolve R-loops for RNAPII release, I also employed SMA patient cells to check if similar phenotypes are seen in the SMA disease state. Six human cell lines (3 fibroblast, 3 B lymphocyte) were obtained from the Coriell Institute (Table 2) that included cells from two SMA children and their unaffected parents. RNAPII ChIP and R-loop DIP were performed as before on the β-actin gene (Fig. 20). When compared to the controls (average of the two parents), RNAPII in the SMA disease cells from both children over-accumulated in the termination regions (Fig. 20A), which also correlated somewhat with sites of increased R-loop formation (Fig. 20B). Thus, in the actual SMA disease state, RNAPII termination release may be compromised due to unresolved R-loops.

2.13 Genome-wide characterization by ChIP-seq of the roles of R1810 and SMN in RNAPII termination

Because my initial experiments were carried out only on the β-actin and the GAPDH genes, I also carried out RNAPII ChIP-seq to determine whether SMN and the R1810me2s modification had more general effects on transcription termination. We performed ChIP-seq to establish that the loss of SMN or the mutation of R1810 to alanine on CTD lead to RNAPII stalling in termination regions, in a genome wide pattern (Fig. 21). RNAPII ChIP-seq (with 4H8 and 8WG16 antibodies) was carried out for shGFP and shSMN knock-down (Raji) cells that produced ~10 million unique reads. RNAPII ChIP-seq was also performed for WT and R1810A RNAPII (with the N20 antibody) after 3 day α-amanitin treatment to deplete the endogenous RNAPII that then also generated ~10 million unique reads per sample. In agreement with the recent literature, we did not observe that termination region pause sites were particularly enriched with GC sequences [199]. For the analysis, we used genes with reliable ChIP-seq
Figure 20  RNAPII and R-loop accumulations are observed in the SMA disease state
Figure 20  RNAPII and R-loop accumulations are observed in the SMA disease state

A. ChIP quantification of RNAPII with the N20 and 8WG16 antibodies in B lymphocytes (bottom) or Fibroblasts (top) obtained from the Corriell Institute, with the indicated primer positions for qPCR along the β-actin gene, with the average of the parents as the negative control. Error bars denote biological replicates as s.e.m. (n=5).

B. Quantification of DIP with the S9.6 antibody for R-loops, with the indicated primer positions for qPCR along the β-actin gene, with the average of the parents as the negative control. Error bars denote biological replicates as s.e.m. (n=5).
<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Cell Type</th>
<th>Origin</th>
<th>Affected</th>
<th>Family</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMA1</td>
<td>Fibroblast</td>
<td>Male (Child)</td>
<td>Yes</td>
<td>553</td>
</tr>
<tr>
<td>SMA1</td>
<td>Fibroblast</td>
<td>Female (Parent)</td>
<td>No</td>
<td>553</td>
</tr>
<tr>
<td>SMA1</td>
<td>Fibroblast</td>
<td>Male (Parent)</td>
<td>No</td>
<td>553</td>
</tr>
<tr>
<td>SMA1</td>
<td>B-Lymphocyte</td>
<td>Female (Child)</td>
<td>Yes</td>
<td>3042</td>
</tr>
<tr>
<td>SMA1</td>
<td>B-Lymphocyte</td>
<td>Female (Parent)</td>
<td>No</td>
<td>3042</td>
</tr>
<tr>
<td>SMA1</td>
<td>B-Lymphocyte</td>
<td>Male (Parent)</td>
<td>No</td>
<td>3042</td>
</tr>
</tbody>
</table>

Table 2 SMA patient and control cell lines obtained from the Coriell Institute
signals (top 5 or 10% by expression; ~800-2000 genes), as we ran into issues with sequencing depth and background interference when more genes were included in the analysis.

Upon normalizing reads to the gene body and comparing to the control (grey), shSMN (purple) causes RNAPII hyper-accumulation in termination regions in a genome-wide pattern (Fig. 21A). Similar patterns were observed when antibodies against either phospho (4H8) or non-phospho (8WG16) RNAPII were used (not shown). The number of genes analyzed is labeled on the bottom right. We also observed a loss of RNAPII accumulation in promoters when shSMN was compared to the control knock-down, an outcome that may be explained by compromised spliceosome assembly due to the lack of SMN. It is known that defective spliceosome formation can lead to increased RNAPII accumulation in introns of the gene body and therefore, a comparative loss of promoter signal [531, 532]. Alternatively, the loss of promoter signal may be an outcome of compromised RNAPII termination, as seen in the recent literature [199]. RNAPII ChIP-seq peaks with or without knock-down of SMN are displayed for several housekeeping genes in Fig. 21D.

Between the WT and R1810A RNAPII ChIP-seq samples, when compared to the WT RNAPII (blue), R1810A RNAPII (red) also showed hyper-accumulation in the termination regions in a genome wide pattern (Fig. 21B). Notably, the RNAPII R1810A mutant also seemed to display a decrease in transcription directionality at the promoter peak, a phenotype observed as an outcome of compromised gene looping that may co-occur when RNAPII fails to be released from the termination sites after polyA cleavage [53, 164]. This phenotype may also be explained by a defect in upstream proximal RNAPII termination (that is needed for suppressing the intrinsic bidirectional nature of many promoters) through the same pathway as the polyA mediated termination [162]. It may or may not depend on SMN’s regulation of RNAPII
Figure 21  SMN and POLR2A R1810 effects on transcription termination by RNAPII occur on a genome wide level
Figure 21 SMN and POLR2A R1810 effects on transcription termination by RNAPII occur on a genome wide level
Figure 21 SMN and POLR2A R1810 effects on transcription termination by RNAPII occur on a genome-wide level

A. shGFP control v. shSMN: ~10 million unique RNAPII ChIP-seq reads (4H8, 8WG16) were obtained from GFP (grey) or SMN (purple) stable knock-down cells (Raji). SMN knock-down (purple) led to increased RNAPII accumulation in the termination regions across 5-10% most highly expressed genes (with reliable ChIP-seq signals) analyzed. The number of genes is displayed on the bottom right. On the right, western blotting verified the efficient knocking down of SMN for the ChIP-seq experiment; shGFP was used as a control.

B. WT control v. R1810A: ~10 million unique RNAPII ChIP-seq reads (N20) were obtained from WT (blue) or R1810A (red) RNAPII samples (Raji cells) upon 3 days of α-amanitin treatment. The R1810A mutant RNAPII (red) showed an increased accumulation in the termination regions for 5-10% most highly expressed genes (with reliable ChIP-seq signals) analyzed. The number of genes is displayed on the bottom right. Also observed was a compromised gene looping phenotype in which the directionality of RNAPII at the promoter was reduced. Western blotting using N20 and HA antibodies verified the expression of the transgene and the effect of the α-amanitin treatment for cells used for the ChIP-seq experiment.

C. ~10 million unique SMN ChIP-seq reads showed that SMN is enriched near promoter and termination regions with (bottom) or without (top) background subtraction, similar to previous observation made with ChIP on individual gene such as β-actin.

D. RNAPII ChIP-seq results for several housekeeping genes were displayed in detail using the Integrative Genomics viewer. The promoter peaks are displayed to the left, and the RNAPIII termination regions are displayed with red underlines. IgG ChIP-seq was used as negative control.
termination, as it is known that R1810 of the CTD can also be asymmetrically dimethylated by CARM1 [67], and that the Rme2a modification may recruit the Tud domain containing protein TDRD3 in a complex with TOP3B for R-loop resolution [232] at promoters for transcription activation. RNAPII ChIP-seq peaks for WT and R1810A RNAPII are displayed for several housekeeping genes in Fig. 21D.

We also performed SMN ChIP-seq to generate ~10 million reads with GFP antibody against inducible GFP-SMN or with SMN specific antibody in HEK293 cells. We observed enriched but weak SMN ChIP-seq signals at the promoter and the termination regions (Fig. 21C); the pattern is reminiscent of the SMN ChIP pattern on individual genes such as β-actin or GAPDH.

2.14 The effect of the CTD R1810A mutation on alternative splicing

SMN is known to affect splicing [423, 424, 427]. This might only reflect its role in snRNP assembly [321, 357, 368, 417, 533], but it might also reflect a role for SMN in splicing via its participation in a pathway for R-loop resolution involving R1810 on the RNAPII CTD. To examine this possibility, I carried out RNA-seq on RNAs extracted from Raji cells that were stably expressing α-amanitin resistant wild type or R810A mutant RNAPII after α-amanitin treatment (2ug/ml) to deplete the endogenous RNAPII. In contrast to a previous RNA-seq experiment with the same types of cells that used a 10 day α-amanitin treatment [67], our samples were treated with α-amanitin for only 3 days. ~100 million reads were generated from each sample for analyses of gene expression, alternative splicing (AS), and alternative polyadenylation (APA), done in collaboration with the Blencowe lab. In comparison to the dataset from Sims et al. [67], we observed similar increases in many snoRNAs (Fig. 22A)
in the mutant when compared to the wild type; and likewise, we observed few gene expression changes for messenger RNA-encoding genes (Fig. 22B). However, we observed 412 AS events with at least 15% changes in percent spliced index (dPSI>15%), which included 208 cassette exons, 104 intron retentions, 54 alternative 5’ splice site selections, and 46 alternative 3’ splice site selections (Fig. 22C). Some of these events were validated by PCR (Fig. 22D). Although these results might mean that the pathway involving the R1810me2s modification and SMN regulates AS, these effects on AS might also be a consequence of the effects of the R1810A mutation on the accumulation of snoRNAs (Fig. 22A) [67].

2.15 The effect of the CTD R1810A mutation on splicing shows some correlation with GC content

It has been shown that stalling by RNAPII leads to changes in co-transcriptional alternative splicing [55], presumably because a slowly moving RNAPII may favor alternative exon inclusion or exclusion, depending on the splicing factors recruited [117, 118]. Since we hypothesized that the RNAPII R1810A mutant might accumulate in G-rich regions because of increased R-loop formation, we expected that we might observe regions with cassette exon inclusion changes or intron retention changes [534] that correlate with higher GC content. When stratifying the AS events from loss to gain of inclusion (blue to yellow) for cassette exons, higher GC content (in the cassette exon and downstream intron) was associated with cassette exon exclusion in the R1810A mutant (Fig. 23A). When stratifying the AS events from loss to gain of intron retention (blue to yellow), higher GC content (in the retained intron) was associated with intron retention in the R1810A mutant (Fig. 23B). The introns that gained or lost retention between the wild type and mutant tended to be more GC-rich compared to non-change.
Figure 22 RNA-seq analysis for WT v. R1810A mutant RNAPII
Figure 22 RNA-seq analysis for WT v. R1810A mutant RNAPII

A. Increased snoRNA levels are seen when RNAPII R1810 is mutated to alanine (p<0.00018).

B. No significant global mRNA expression changes are observed between WT and R1810A samples.

C. Observed changes in different classes of alternative splicing events (schematics below).

D. RT-PCR validation of some predicted changes in alternative splicing.
events (grey), which might reflect the observation that retained introns typically tend to be more GC rich than non-retained ones [534].

Many genes contain high G content after the polyadenylation site that plays a likely role in enhancing RNAPII termination, especially in gene dense regions [91, 202]. It is known that RNAPII stalls in termination regions upon the inhibition of polyadenylation [70, 71], and so we expected we might see changes in alternative polyadenylation site (APA) usage between wild type and the R1810A mutant. Proliferating and cancerous cells express mRNAs that tend to terminate at upstream polyadenylation sites, thereby hypothetically forming transcripts capable of prolonged protein synthesis due to shortened 3’ UTRs that are targeted by fewer microRNAs [535, 536]. Preliminary analysis with dAPA >15% between WT and the R1810A mutant showed that approximately 550 genes have altered APA usage. In these genes, although we tried to correlate GC or G content in the 3’ UTRs and regions downstream (50-200bp) of the cleavage sites with the effect of the mutation on APA, no bias was observed (not shown). We also tried to detect if the R1810A mutant prefers a shorter or longer 3’ UTR length as compared to the wild type, again seeing no bias (not shown). Along with our observations from the aforementioned RNAPII ChIP-seq experiments, it seems that RNAPII termination defects in the R1810A mutant do not only happen on G-rich tracts or contribute to APA changes in G-rich regions; instead, the defect is likely to happen regardless of GC content and is coupled with subtle APA changes in the mutant.

2.16 CTD R1810A effect on splicing does not correlate with RNAPII pause sites

Since we hypothesized that sites where the R1810A mutant RNAPII accumulate in the
Figure 23 GC content analysis for cassette exons and intron retentions
Figure 23 GC content analysis for cassette exons and intron retentions

The GC contents of the entire exon, upstream and downstream introns were considered.

A. When stratifying the events from loss to gain of inclusion (blue to yellow) for cassette exons, the higher GC content (in the cassette exon and downstream intron) is associated with cassette exon exclusion in the R1810A mutant. When comparing CG content in each region between events with dPSI < -5 vs. dPSI > 5, the differences in the cassette exon (p=0.011) and downstream introns (p=0.014) are significant. dPSI: difference in percent splicing index

B. When stratifying the events from loss to gain of intron retention (blue to yellow), higher GC content (in the retained intron) is associated with retention in the R1810A mutant. When comparing CG content in the each region between events with dPSI < -5 vs. dPSI > 5, the differences in the retained intron (p=0.024) is significant.
gene body might affect splice site selection, we expected that we might observe a correlation between sites of cassette exon changes and RNAPII peak changes. Peaks were called in individual ChIP-seq experiments using MACS, with all peak locations combined (~60,000). These peaks overlapped 1,697 of ~ 75,000 cassette exons, whereas 10,353 overlapped with their upstream introns and 16,269 with their downstream introns. Differential RNAPII accumulation between wild type and the R1810A mutant RNAPII was assayed by comparing ChIP reads in peak regions using the binomial test. The cumulative distribution of PSI differences was plotted for cassette exons with overlapping peaks but showed no significant correlation with the change in RNAPII accumulation (Fig. 24). Similarly, differential RNAPII accumulation between shGFP and shSMN samples also showed no correlation with cassette exon changes (Fig. 24). We, however, have not searched for a correlation between sites of RNAPII pausing and intron retention due to technical difficulties in computation. In any case, our study suggests that the AS changes for cassette exons in the R1810A mutant may be better described by the recruitment model rather than by the RNAPII kinetic model. Therefore, as described in Chapter 3, we looked for additional RBPs or splicing factors that might be recruited by SMN to the RNAPII CTD so as to then generate the observed cassette exon changes [115].

3. Discussion

I have found that R1810 in the RNAPII CTD is symmetrically dimethylated by the arginine methyltransferase PRMT5. The presence of this modification on the CTD is substantiated by a number of observations: first, immunoprecipitated RNAPII is detected in western blots by two different antibodies, SYMM10 and Y12, that are specific for Rme2s; second, RNAPII is precipitated by a specific antibody raised against the R1810me2s containing
Figure 24 Correlation between cassette exon changes and RNAPII pausing patterns
Figure 24 Correlation between cassette exon changes and RNAPII pausing patterns.

Peaks were called in individual ChIP-seq experiments using MACS (Model-based Analysis of ChIP-Seq), and all peak locations were combined. Differential RNAPII accumulation between WT and R1810A mutant RNAPII (top) or between shGFP and shSMN samples (bottom) was assayed by comparing ChIP reads in peak regions using the binomial test. The cumulative distribution of PSI differences was plotted for cassette exons with overlapping peaks but showed no significant correlation with the change in RNAPII accumulation.
peptide, and this antibody also recognizes RNAPII in western blots; third, the presence of the modification recognized by the antibodies depends on R1810; fourth, the presence of the modification depends on PRMT5; and fifth, R1810 and PRMT5 are needed for the association and the recruitment of SMN to the transcribing RNAPII, whose Tudor domain is specific for Rme2s. These observations also suggest that symmetric dimethylation of R1810 in the RNAPII CTD causes the direct recruitment of SMN and indirect recruitment of the RNA:DNA helicase SETX, followed by the resolution of R-loops for the efficient termination by RNAPII (see model in Fig. 25). From this study, it was also observed that the interaction between SMN and SETX is mediated by PRMT5, likely through the SMN Tudor domain binding with an Rme2s modification on SETX (Fig. 10C).

PRMT5 appears to be the only type II PRMT that generates symmetric dimethylation marks on arginine residues embedded in PGM containing sequences [267]. FCP1, a phosphatase for the dephosphorylation of RNAPII CTD, as well as two Mediator CDKs (CDK8 and CDK19) are known to recruit PRMT5/WDR77 to RNAPII [324, 537]. It is likely that PRMT5 is similarly recruited to methylate the CTD at R1810 through the elongation phase of the transcription cycle. Using the SMN ChIP pattern as a readout for the CTD methylation status, I observed that SMN is enriched at the 3’ ends of genes, though it is also present at promoters and throughout the gene body (Fig. 11A, 12A, 21C).

I have shown that SMN recognizes RNAPII CTD R1810me2s in vitro and in vivo. SMN can self-aggregate through its N-terminal K-rich domain and its C-terminal YG box to form a multimeric adaptor to assemble Rme2s containing proteins into a complex [388, 538, 539]. Because the SMN complex serves as an assembly factor for the snRNPs involved in spliceosome assembly and is also present in cytosolic RNP complexes for axonal mRNA transport, these two
functions have been proposed to contribute to the SMA disease phenotype [388]. Analogous to its role in the assembly of Rme2s containing proteins in the spliceosome, I believe that SMN acts similarly to assemble a termination complex on the RNAPII CTD, in this case through interacting with Rme2s modified termination factors.
Fig 25 Model of the CTD1810me2s-SMN-SETX pathway that regulates R-Loop accumulation to prevent DNA damage at the RNAPII termination regions.
Chapter 3: The regulation of RNAPII transcription termination by FUS and TDP-43 and its relevance to neurodegenerative disorders

“The Regulation of RNA polymerase II Transcription Termination by SMN, FUS, and TDP-43”

Dorothy Yanling Zhao, Shuye Pu, Ulrich Braunschweig, Frank Schmitges, Hongbo Guo, Guoqing Zhong, Jinrong Min, Tony Pawson, Ben Blencowe, Jack Greenblatt

Manuscript in preparation

Author contributions

Dr. Frank Schmitges performed ChIPseq experiments. Shuye Pu and Dr. Ulrich Braunschweig performed computational data analysis for ChIP-seq. Hongbo Guo performed Mass Spectrometry. Guoqing Zhong generated stable shRNA knock-down cell lines. Dr. Jinrong Min provided SMN inhibitors.
Chapter 3: The regulation of RNAPII transcription termination by FUS and TDP-43 and its relevance to neurodegenerative disorders

1. Introduction

In the previous chapter, I showed that the arginine residue R1810 of the RNAPII CTD is symmetrically dimethylated (me2s) by the PRMT5 methyltransferase, allowing it to recruit the Tudor domain of the SMN protein, which is mutated in SMA. SMN can self-associate through its N- and C-terminal domains to form a multimeric adaptor for the nucleation of Rme2s containing proteins [360, 388]. Because SMN participates in spliceosome assembly and is also present in cytosolic ribonucleoprotein complexes for axonal mRNA transport in neurons, the mis-regulations of these two roles have been proposed to contribute to SMA [388]. One SMN interactor is known to be Senataxin (SETX) [166], a DNA:RNA helicase that is sometimes mutated in ALS4, and both proteins participate in the regulation of transcription termination by RNAPII [91]. In the previous Chapter, I confirmed that SMN interacts with SETX, and this interaction is mediated by PRMT5 and independent of RNAPII.

Among the ~150 human proteins that have been shown by mass spectrometry to contain dimethylarginine in their GAR or PGF motifs, a remarkable number of them (>50) are involved in RNAPII transcription and termination [390-394]. A fraction of these is shown in Table 3. It is likely that the CTD R1810me2s-SMN pathway may enhance the assembly of these methylated factors onto the RNAPII CTD for the proper regulation of RNAPII transcription and termination, as well as pre-mRNA splicing and 3’ end processing. Among the factors that are candidates for SMN binding, I focused on showing that, like SETX, the recruitment of FUS and TDP-43 to RNAPII is enhanced by CTD R1810me2s and the adaptor protein SMN. FUS and TDP-43 are methylated DNA/RNA binding proteins [394, 471] that are involved in numerous aspects of
Table 3  RNAPII 3’ processing and termination factors that contain Rme2 modifications as potential SMN binders.
Candidates detected in the experiments (AP-MS or Bio-ID shown later) as SMN interactors are shown in red.
gene regulation, and are known to interact with each other [470] and with SMN [399, 479]. As well, they are sometimes mutated in neurodegenerative diseases such as ALS and FTD [473, 476, 540].

It is known that the ALS disease phenotypes caused by FUS and TDP-43 mutations may be related to their RNA binding activities [474, 541] and could be involved in RNAPII transcription and splicing [401, 454]. TDP-43 preferentially binds to GU-rich sequences and predominantly affects the expression of transcripts with long introns, such that its loss may lead to intron retention that would introduce premature stop codons and promote RNA degradation. TDP-43 likely also plays a role in RNAPII termination, as it has been shown to form a complex with the 5' → 3' exonuclease XRN2 that is known to induce RNAPII termination and dissociation [99, 542]. FUS was shown to associate with the RNAPII CTD and to affect the CTD phospho-Ser2 level. It was also shown to regulate 3' end cleavage and polyadenylation and to promote RNAPII termination [453, 490]. The loss of FUS is known to lead to intron retention in 3-5% of transcripts [543]. FUS also has a preference for GGUG-rich RNAs that resemble the consensus 5′ splice site (GGUG/A), and thus it may interfere with proper splicing at 5′ splice sites [543]. It is possible that, in the absence of RBPs such as FUS or TDP-43, the accumulation of R-loops would interfere with splicing and RNA processing, thereby leading to an increased likelihood of intron retention. Recently, it was shown that the low complexity (LC) domain of FUS allows it to oligomerize and to interact with the RNAPII CTD for transcriptional regulation [433, 441]. Because such studies were done in vitro, it is not known whether FUS and other LC domain containing proteins that are linked to ALS/FTD (TDP-43, TAF15, EWSR1) would form heteromers through their LC domains in vivo or whether such structures would have a more stable association with the transcribing RNAPII [431].
In the previous Chapter, I showed that CTD R1810me2s and SMN, like SETX, are required for resolving R-loops that otherwise would stall RNAPII and lead to DNA damage in the termination regions of many genes, including the β-actin gene. This chapter summarizes my evidence that FUS and TDP-43 act downstream of the CTD R1810me2s-SMN pathway, and that defects in the recruitment of these proteins can also influence genome-wide RNAPII transcription termination through the accumulation of R-loops. As a consequence of the loss of FUS or TDP-43, DNA damage was also observed to accumulate in the termination regions of genes such as β-actin. Similar phenotypes are recapitulated when a truncated form of TDP-43 that lacks the RNA binding RRM domain is overexpressed. These observations suggest that the binding of RBPs such as FUS or TDP-43 to the nascent RNAs plays a role in preventing R-loop formation. In summary, I propose that the mis-regulation of RNAPII termination due to R-loop accumulation can lead to DNA damage, which may contribute to neurodegenerative disorders such as SMA and ALS/FTD.

2. Results

2.1 Identification of SMN-interacting proteins by AP-MS and BioID

Transcription termination typically occurs far downstream from cleavage and polyadenylation sites and requires either the cleavage factors or self-cleavage of the pre-mRNA [99, 175]. After cleavage, the 5’→3’ exonuclease XRN2 degrades the downstream RNA behind RNAPII in a process important for termination [99, 196]. Another mechanism by which RNAPII can terminate is through backtracking and freeing up the 3’ end of the nascent RNA, thereby promoting its degradation by the 3’ → 5’ exosome complex for the release of the RNAPII [98]. It is thought that SETX may coordinate the actions of XRN2 and the exosome for RNAPII
termination [97, 544]. In order to identify other proteins that might function in the R1810me2s-SMN termination pathway, I used tagged version of SMN in AP-MS and BioID experiments to identify putative SMN-interacting proteins.

HEK293 Flp-In T-REx cell lines expressing inducible GFP- or BirA (promiscuous biotin ligase)-tagged SMN were used to identify SMN interactors through AP-MS and BioID, respectively [545]. BioID relies on the tagged BirA to covalently conjugate biotin to lysine residues on nearby proteins in live cells, after which the biotinylated proteins are enriched and identified by MS [545, 546]. Combining BioID with AP-MS often produces complementary datasets that include weak and transient interactions for a more complete interactome [545]. Both approaches identified FUS and TDP-43 as SMN interactors, as well as other known interactors, including certain Sm and Gemin proteins (highlighted in orange in Table 4). BioID identified additional interactors in the 3’ processing and RNAPII termination pathways, including the 5’→3’ riboexonuclease (XRN2), three subunits of the 3’→5’ exosome complex (Exosc), 3 subunits of CPSF (1, 2, and 6), FIP1L1, WDR33 [191, 192], and RBBP6 [547], as well as many other proteins that regulate transcription and splicing (Table 4). Notably, also identified were several interactors whose mutated forms are known to cause neuronal disorders such as Fragile X mental retardation syndrome (FMR/FXR complex), Spinocerebellar Ataxia 2 (ATXN2L), and ALS/FTD (FUS, TDP-43, EWSR1, and hnRNPA3). Therefore, the SMN interactome shown in Table 4 also supports the idea that mis-regulations of transcription and RNA metabolism are the underlying basis of neuronal diseases [399]. Another known factor involved in RNAPII termination is CBX3 (HP1γ), which was also identified as an SMN interactor from both AP-MS and Bio-ID approaches. CBX3 enables proper RNAPII termination
Table 4: SMN interactors identified by both AP-MS and Bio-ID. SMN interactors identified by both AP-MS and Bio-ID are listed in the middle. Listed hits are seen multiple times in BioID or AP-MS. Red = RNA processing or 3'end processing factors; yellow = known SMN interactors. Negative control for BioID is cells expressing BirA alone, and negative control for AP-MS is cells expressing GFP alone.

|---------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|

Legend:
- Red: RNA processing or 3'end processing factors
- Yellow: Known SMN interactors
through the heterochromatization of termination regions by an RNA interference mechanism after the elimination of R-loops [199]. Therefore, the recruitment of CBX3 to the termination sites of RNAPII may also be enhanced through its interaction with SMN.

2.2 Interaction of FUS and TDP-43 with RNAPII and termination factors

I focused on elucidating the roles of the RBPs FUS and TDP-43 as possible downstream factors in the CTD R1810me2s-SMN pathway because it has been proposed that ALS and SMA are similar diseases [399]. FUS and TDP-43 are known to play multiple roles in RNAPII transcription and co-transcriptional splicing, yet it is not known whether they have roles in preventing R-loop formation, thus enhancing RNAPII termination like SMN and SETX.

It has been shown that more than 50 proteins that contain dimethylated arginine (in GAR or PGM motifs) function either in the regulation of RNAPII elongation and termination, or in co-transcriptional splicing or the 3’ end processing of mRNAs [267, 280, 390-394]. These include well known RNAPII 3’ end processing and termination factors such as XRN2, three subunits of CPSF (1, 5, and 6), CstF2, RBBP6, WDR33, and PCF11, as well as other RNAPII regulatory proteins such has SPT5, CTDP1, DHX9, three of the PolyA-binding proteins (PABP1, 2, and 4), TDP-43, and the FET proteins (FUS, TAF15, and EWSR1) (Table 3). Therefore, defects in the CTD R1810me2s-SMN pathway may introduce splicing or termination defects due to the formation of a less stable RNAPII termination complex that contains many of the aforementioned Rme2s-containing proteins.

To test whether FUS and TDP-43 might also function in RNAPII termination pathways, I first used immuno-staining to show that these proteins co-localize in the nucleus with each other,
Figure 26 FUS and TDP-43 localize to the nucleus
Figure 26 FUS and TDP-43 localize to the nucleus

Immuno-staining of FUS and TDP-43 in HEK293 cells.

The staining shows that FUS, TDP-43, the termination factors such as SETX and XRN2, and POLR2A (pSer2, pSer5) localize to similar regions of the nucleus.
with the phosphorylated forms of POLR2A, and with RNAPII termination factors such as XRN2 or SETX (Fig. 26). Next, from multiple co-IP experiments, I observed that SMN and POLR2A interact with the Rme2-containing proteins FUS, TDP-43, and XRN2 in a network (Fig. 27). I was also able to show that many of these interactions occur independently of RNAPII, as they persisted upon 3 days of α-amanitin treatment (2ug/ml) that eliminated the bulk of the RNAPII (Fig. 28A). The persistent interactions included SMN-XRN2, SMN-FUS, SETX-XRN2, SETX-FUS, FUS-TDP-43, and SMN-SETX (Chapter 2), indicating that the RBPs and the termination factors pre-form a complex through direct interactions prior to being recruited to RNAPII for transcription termination. It is also likely that FUS, TDP-43, and XRN2 contain Rme2s modifications and are PRMT5 substrates, because I found them to interact with PRMT5 in various co-IP experiments, even when RNAPII is abolished by treating the cells with α-amanitin (Fig. 27, 28A).

2.3 FUS and TDP-43 function downstream of the R1810me2s-SMN pathway

In the previous chapter, I showed that the CTD R1810me2s modification of RNAPII recruits SMN and SETX to resolve R-loops and to facilitate RNAPII termination. Here, I will present evidence that the association of FUS and TDP-43 with RNAPII is also enhanced by SMN and the presence of the RNAPII CTD R1810 residue in vivo. First, I found by co-IP that the interactions of FUS and TDP-43 with RNAPII were reduced when R1810 was mutated to alanine on the RNAPII CTD (Fig. 28B, 28C). The experiments were performed in Raji cells after 3 days of α-amanitin treatment that eliminated the bulk of the endogenous RNAPII. Interactions were detected by immunoprecipitating the α-amanitin resistant HA-tagged WT or
Figure 27 FUS and TDP-43 interact with RNAPII, SMN, and the termination factors
Figure 27 FUS and TDP-43 interaction with RNAPII, SMN, and the termination factors

A. IP with the indicated antibodies from HEK293 WCL, followed by western blotting with the indicated antibodies. The blots showed that POLR2A and SMN interact with each other, and with PRMT5, and the termination factor XRN2, as well as the RBPs such as FUS and TDP-43.

B. IP with the indicated antibodies from HEK293 WCL, followed by western blotting with the indicated antibodies. The blots showed that POLR2A, SMN, SETX, FUS, and TDP-43 interact as a network.
Figure 28 FUS and TDP-43 interaction with SMN and the termination factors is independent of RNAPII, and their interaction with RNAPII is mediated by the CTD R1810me2s modification.
Figure 28  FUS and TDP-43 interaction with SMN and the termination factors is independent of RNAPII, and their interaction with RNAPII is mediated by the CTD R1810me2s modification

A. IP with the indicated antibodies from HEK293 WCL, followed by western blotting with the indicated antibodies with or without α-amanitin treatment. Many of the interactions occur independently of RNAPII as they persist with α-amanitin treatment that eliminated the bulk of RNAPII. These include SMN-XRN2, SMN-FUS, SETX-XRN2, SETX-FUS, and FUS-TDP-43.

B-C. IP with the indicated antibodies from Raji WCL stably expressing the HA-tagged wild-type or R1810A mutant POLR2A upon α-amanitin treatment that abolishes the endogenous POLR2A. Anti-HA was used to precipitate HA-tagged wild-type (WT) or R1810A mutant POLR2A, followed by western blotting with the indicated antibodies.
Figure 29 SMN and PRMT5 mediate the interactions of FUS and TDP-43 with RNAPII
Figure 29 SMN and PRMT5 mediate the interactions of FUS and TDP-43 with RNAPII

A-B. IP with the N20 and 8WG16 antibodies for POLR2A from HEK293 WCL upon stably knocking down of PRMT5 or SMN (A), or knocking out of SMN by CRISPR (B). Western blots were performed with the indicated antibodies. The knocking down of PRMT5 or SMN caused a loss of FUS and TDP-43 interaction with RNAPII. SMN knock-out caused reduced interaction of TDP-43 and SETX with RNAPII.
the R1810A mutant POLR2A using anti-HA antibodies. Similarly, by immunoprecipitating the RNAPII, I found that the interaction between POLR2A and TDP-43 was consistently reduced when SMN or PRMT5 was knocked down, or SMN was knocked out in HEK293 cells (Fig. 29). The interaction of FUS with RNAPII was often but not always weakened upon the loss of SMN in co-IP experiments (Fig. 29B), perhaps because FUS is able to interact directly with the non-phosphorylated isoform of RNAPII (as immunoprecipitated by the 8WG16 antibody) [453, 490], as well as with SMN [399].

2.4 SMN and CTD R1810me2s enhance the association of FUS and TDP-43 with transcribing RNAPII

To provide further evidence for the roles of SMN and RNAPII CTD R1810me2s in recruiting FUS and TDP-43 to RNAPII, ChIP experiments were performed. ChIP involving α-amanitin resistant WT and R1810A RNAPII mutant cell lines was done after 3 days of treatment with α-amanitin at 2μg/ml to deplete the endogenous RNAPII. In WT cells, FUS and TDP-43 ChIP are represented as percent input or as ratio with respect to RNAPII. I observed that FUS and TDP-43 are associated with the β-actin gene throughout its length from the promoter to the termination regions (Fig. 30A). ChIP also showed that both FUS/POLR2A and TDP-43/POLR2A signals along the length of the β-actin gene are reduced when CTD R1810 is mutated to alanine (Fig. 30B), or when PRMT5 or SMN is knocked down, with the ratio in the control sample set to 1 (Fig. 30C). Collectively, these observations suggested that FUS and TDP-43 association with RNAPII is enhanced through SMN and the R1810me2s modification on the CTD.

With the control ratio normalized to 1, the ChIP signals of FUS/POLR2A and TDP-43/POLR2A along the length of the β-actin gene were reduced when either FUS or TDP-43 was
Figure 30  CTD R1810me2s and SMN enhance FUS and TDP-43 association with RNAPII

140
Figure 30 CTDR1810me2s and SMN enhance FUS and TDP-43 association with RNAPII

A. Quantification of ChIP data in HEK293 cells, expressed as percent input, or as FUS/POLR2A or TDP-43/POLR2A ratio, using the indicated primer positions for qPCR along the β-actin gene. Error bars denote biological replicates s.e.m. (n=3).

B. Quantification of ChIP data in Raji cells as FUS/POLR2A or TDP-43/POLR2A ratio to show the relative effects of mutating R1810 to alanine. Error bars denote biological replicates s.e.m. (n=3).

C. Quantification of ChIP data in HEK293 cells as FUS/POLR2A or TDP-43/POLR2A ratio to show the relative effects of knocking down of PRMT5 or SMN. Error bars denote biological replicates s.e.m. (n=3).

D. Quantification of ChIP data in HEK293 cells as FUS/POLR2A or TDP-43/POLR2A ratio to show the relative effects of knocking down FUS or TDP-43. Error bars denote biological replicates s.e.m. (n=3).
knocked down (Fig. 30D) or knocked out (not shown). These findings suggested that FUS and TDP-43 may form heteromers, possibly through their low complexity (LC) domains, and that such structures may be more stably associated with the RNAPII CTD than when either protein is present alone. Because of these interdependencies among R1810, SMN, FUS, and TDP-43, it therefore seemed likely that SMN, FUS, and TDP-43 would function similarly in the regulation of RNAPII termination.

2.5 FUS and TDP-43 regulate transcription termination by RNAPII

Because SMN and RNAPII CTD R1810me2s are important for RNAPII termination (Fig. 13, 14), I investigated whether the loss of FUS or TDP-43 also influences RNAPII termination using similar assays. RNAPII ChIP and R-loop DIP (see below) were performed and the results displayed following the style mentioned previously in Chapter 2. It was observed previously that upon mutating the RNAPII CTD residue R1810 to alanine or knocking down PRMT5, SMN, or SETX, RNAPII accumulated downstream of the polyadenylation sites on genes such as β-actin and GAPDH in ChIP experiments (Fig 13, 14). Using HEK293 cells, similar observations were made for stable shRNA-mediated knock-downs (Fig. 31A, 32A), transient-mediated siRNA knock-downs (not shown), or CRISPR-mediated knock-outs (Fig. 31B, 32B) of FUS or TDP-43. The control samples in these experiments were HEK293 cells with stable shRNA-mediated knock-down of GFP, transient knock-down using scrambled siRNAs, or transfection of scrambled guide RNAs with the CRISPR system, respectively.

It is known that the RNA-binding ability of the FUS RNA recognition motif (RRM) is essential for its function and is linked to the neurodegenerative phenotype in vivo [548]. Since we hypothesized that the RNA binding ability of FUS or TDP-43 could prevent R-loop
Figure 31 FUS and TDP-43 regulate transcription termination by RNAPII on the beta-actin gene
Figure 31 FUS and TDP-43 regulate transcription termination by RNAPII on the β-actin gene

A-B. Quantification of RNAPII ChIP using POLR2A antibody (4H8, N20) in HEK293 cells, using the indicated primer positions for qPCR along the β-actin gene, after stably knocking down FUS or TDP-43 (A), with GFP knock-down as a negative control, or knocking out of FUS or TDP-43 by the CRISPR system, using scrambled guide RNA as negative control (B). Normalized ratios to control POLR2A (set to 1) are displayed to the right. Error bars denote biological replicates as s.e.m. (n= 3-5).

C. Western blotting with the indicate antibodies showing gene knock out of TDP-43 or FUS by CRISPR.
Figure 32  FUS and TDP-43 regulate transcription termination by RNAPII on the GAPDH gene
Figure 32 FUS and TDP-43 regulate transcription termination by RNAPII on the GAPDH gene

A-B. Quantification of RNAPII ChIP using POLR2A antibody (4H8, N20) in HEK293 cells, using the indicated primer positions for qPCR along the GAPDH gene, after stably knocking down FUS or TDP-43 (A), with GFP knock-down as a negative control, or knocking out of FUS or TDP-43 by CRISPR, using scrambled guide RNA as negative control (B). Normalized ratios to control POLR2A (set to 1) are displayed to the right. Error bars denote biological replicates as s.e.m. (n= 3-5).

C. Microscopy images showing that the knock-outs of FUS or TDP-43 appear to be physiologically normal in comparison to the control KO shown on the left.
formation, thus reducing the tendency of RNAPII to stall, HEK293 Flp-In T-REx cell lines expressing GFP-tagged wild type TDP-43 or a truncation mutant that lacks the RRM domain were generated to test this hypothesis (Fig. 33A). Compared to the control and with the ChIP signals normalized in the gene body, the overexpression of the TDP-43 truncation mutant led to increased RNAPII accumulation in the termination regions of the β-actin and the GAPDH genes (Fig.33B). This result suggested that the ability of TDP-43 to bind RNA is important for its ability support termination by RNAPII.

2.6 FUS and TDP-43 prevent the accumulation of R-loops and DNA damage at termination sites

In the previous chapter, I showed that, like the depletion of SETX (Fig. 17B) [91], the mutation of the RNAPII CTD R1810 residue to alanine or the knock-down of SMN or PRMT5 also led to the hyper-accumulation of R-loops in the termination regions of the β-actin gene (Fig. 17B). The monoclonal antibody S9.6 that I used for these DIP experiments was shown to specifically bind to the R-loop structure (Fig. 17A) in the DIP experiment, with the signals abolished upon RNase H treatment. The regions where RNAPII stalled correlated well with sites of R-loop accumulation both for the R1810A mutant and the knock-downs of PRMT5, SMN, and SETX (Fig. 17B). Because FUS and TDP-43, like SMN and SETX, also affected termination by RNAPII, I tested whether they also influence R-loop accumulation. I found that the knock-down of FUS or TDP-43 led to similar accumulation of R-loops in the termination regions on the β-actin gene (Fig. 34A). The regions where RNAPII accumulated correlated well with where R-loops accumulated upon the knock-down of FUS or TDP-43 (Fig. 34A). As well, compared to the control cells that over-express full length GFP-TDP-43, the expression of the
Figure 33 TDP-43 RRM is needed for its regulation of transcription termination by RNAPII
Figure 33 TDP-43 RRM is needed for its regulation of transcription termination by RNAPII

A. Top: Immuno-staining verifies the expression pattern of full length or the truncated GFP-TDP-43 that were induced upon 1 day of doxycycline treatment. Bottom: western blotting with GFP antibody to show the levels and sizes of the GFP fusion proteins extracted from HEK293 WCL. The truncated TDP-43 lacks the RRM (RNA recognition motifs) at the C-terminus of the protein (schematics shown at the bottom).

B. Quantification of RNAPII ChIP using the POLR2A antibody (N20) in HEK293 cells, with cells overexpressing wild type GFP-TDP-43 or the TDP-43 RRM- truncation mutant, using the indicated primer positions for qPCR along the β-actin (top) or GAPDH (bottom) genes. Error bars denote biological replicates as s.e.m. (n= 3).
Figure 34  FUS and TDP-43 are important for resolving R-loops created by the elongating RNAPII
Figure 34 FUS and TDP-43 are important for resolving R-loops created by elongating RNAPII

A. Quantification of DIP with the S9.6 antibody in HEK293 cells with or without knocking down of FUS or TDP-43, using the indicated primer positions for qPCR along the β-actin gene. Error bars denote biological replicates as s.e.m. (n=3 to 5).

B. ChIP quantification of R-loops with the GFP antibody, with the indicated primer positions for qPCR along the β-actin gene, after knocking out of FUS or TDP-43 by CRISPR or with scramble guide RNA as a negative control. Error bars denote biological replicates as s.e.m. (n=3).
Figure 35: TDP-43 RRM RNA binding domain is needed for its repression of R-loops.
Figure 35 TDP-43 RRM RNA binding domain is needed for its repression of R-loops

A-B. Quantification of DIP with the S9.6 antibody in HEK293 cells, with cells overexpressing wild type GFP-TDP-43 or the TDP-43 RRM- truncation mutant, with the indicated primer positions for qPCR along the β-actin (A) or the GAPDH (B) genes. Error bars denote biological replicates as s.e.m. (n=3).
truncation mutant missing the RRM domain of TDP-43 led to further accumulation of R-loops in the termination regions of the β-actin (Fig. 35A) and the GAPDH (Fig. 35B) genes.

To further address the validity of R-loop detection, a second method with a GFP Fusion construct that includes the RNase H1 R-loop binding domain (GFP-HB) was used [230]. The cells stably expressed GFP-HB under G418 selection. Compared to the mock knock-out cells, cells with increased accumulation of R-loops as detected by GFP ChIP were seen in the termination regions of the β-actin gene in the FUS and TDP-43 gene knock-outs using the CRISPR system (Fig. 34B). These experiments indicated that R1810me2s, PRMT5, SMN, SETX, FUS, and TDP-43 act in a common pathway important for R-loop resolution and transcription termination by RNAPII.

As was the case for the loss of SMN or SETX, an important issue is why R-loop accumulation in termination regions, due to the loss of FUS or TDP-43, may contribute to neurodegeneration. Because increased genome instability is known to be associated with neurodegeneration, I again checked if there exists an elevation of DNA damage in the termination regions when FUS or TDP-43 is depleted from the transcribing RNAPII [201]. As before, I used γH2AX ChIP to assay for DNA damage [238] (Fig. 36). Similar to my previous observations, I observed an accumulation of γH2AX and increased γH2AX/H2AX ratio in the termination region where RNAPII and R-loops accumulated upon the loss of FUS or TDP-43.

2.7 The use of SMN inhibitors mimics the effects of an RNAPII CTD R1810A mutation and an SMN mutant

To further validate the observation that the Tudor domain of SMN is the nucleator that enables the proper termination of RNAPII through the recruitment of FUS and TDP-43, we obtained two SMN inhibitors (1190, 237) and a negative control (1194) for use in Co-IP and
Figure 36 FUS and TDP-43 are important for repressing R-loop induced DNA damage
Figure 36 FUS and TDP-43 are important for repressing R-loop induced DNA damage

ChIP quantification of γH2AX as percent input (Top) or as γH2AX /H2AX ratio (Bottom) in HEK293 cells, along the length of the β-actin gene, after knocking down FUS or TDP-43, with the knock-down of GFP as a negative control. Error bars denote biological replicates as s.e.m. (n=4).
ChIP assays. The SMN inhibitors were generated at the Structural Genomic Consortium (SGC) in Toronto, and have been shown to occupy the binding pocket of the SMN Tudor domain with a low micromolar Kd (as measured by ITC or FP in vitro binding assays), thereby potently inhibiting its binding to Rme2s containing peptides in vitro (Fig. 37A). A crystal structure was also generated at the SGC, showing a complex of the inhibitor (1190) with the Tudor domain of SMN (Fig. 37B). The inhibitor resides in the aromatic cage of the Tudor domain, stacked between the aromatic rings of W102 and Y130, and forms a hydrogen bond between its imine group and the side chain of N132.

The inhibitors and the negative control were used on HEK293 cells at 3uM for 2 days prior to co-IP and ChIP experiments. The HEK293 cells also stably expressed the HB-GFP construct for the detection of R-loops by GFP ChIP. Compared to the control, the addition of the inhibitors led to an accumulation of RNAPII as detected by the POLR2A (N20 antibody) ChIP (Fig. 38A) in the termination regions of the β-actin gene. Similarly, R-loop ChIP also showed an increased signal in these regions upon the addition of the SMN inhibitors (Fig. 38B). With the SMN/POLR2A, SETX/POLR2A, FUS/POLR2A, and TDP-43/POLR2A ratios of the control treatment normalized to 1, the addition of the inhibitors also led to a loss in the association of these proteins with RNAPII along the length of the β-actin gene (Fig. 38C). These findings from the ChIP assays were supported by co-IP experiments, in which the addition of the SMN inhibitors was shown to weaken interactions among SMN, TDP-43, FUS, XRN2, and POLR2A upon immunoprecipitation of SMN or POLR2A (with the 4H8 antibody) (Fig. 39). In almost every case, the effect of inhibitor 237 was stronger than that of inhibitor 1190, even though these inhibitors bound equally well to SMN in vitro (Fig. 37A), perhaps because inhibitor
Figure 37 SMN inhibitors block the Rme2s binding pocket of SMN
Figure 37 SMN inhibitors block the Rme2s binding pocket of SMN

A. Inhibitors show specific blocking of the SMN Tudor domain as measured by Fluorescence Polarization and Isothermal Titration Calorimetry assays. The structurally related 1194 that showed negligible binding was used as a negative control for the ChIP and co-IP experiments.

B. Co-crystal structure of the inhibitor 1190 binding to the SMN Tudor domain (left), blocking its supposed interaction with an Rme2s modification (right).
Figure 38 SMN inhibitors mimic the effects of an RNAPII CTD R1810A mutation and an SMN mutant (ChIP)
Figure 38 SMN inhibitors mimic the effects of an RNAPII CTD R1810A mutation and an SMN mutant (ChIP)

A. ChIP quantification of POLR2A (N20 antibody), with the indicated primer positions for qPCR along the β-actin gene, after the addition of the SMN inhibitors or the negative control. Error bars denote biological replicates as s.e.m. (n=3).

B. ChIP quantification of R-loops with the GFP antibody, with the indicated primer positions for qPCR along the β-actin gene, after the addition of the SMN inhibitors or the negative control. Error bars denote biological replicates as s.e.m. (n=3).

C. ChIP quantification shows that the ratios of SMN/POLR2A, SETX/POLR2A, FUS/POLR2A and TDP-43/POLR2A were reduced upon the addition of SMN inhibitors, along the length of the β-actin gene, with the control inhibitor set to 1. Error bars denote s.e.m. (n=2 for biological replicates).
Figure 39 SMN inhibitors mimic the effects of an RNAP II CTD R1810A mutation and an SMN mutant (Co-IP)
Figure 39 SMN inhibitors mimic the effects of an RNAPII CTD R1810A mutation and an SMN mutant (Co-IP)

A-B. IP with the SMN and POLR2A (4H8) antibodies from the HEK293WCL, after 2-day treatment with the control or the SMN inhibitors. Western blots were performed with the indicated antibodies. The addition of the inhibitors weakened the interactions among RNAPII, SMN, FUS, TDP-43, and XRN2.
237 entered HEK293 cells more efficiently than inhibitor 1190. These observations indicated that the Tudor domain of SMN is important for recruiting SETX, FUS, and TDP-43 to RNAPII, likely through the Rme2s mediated interactions.

2.8 Genome-wide characterization of FUS and TDP-43 in RNAPII termination by ChIP-seq

Because my initial experiments were carried out only on the β-actin and GAPDH genes, I also performed RNAPII ChIP-seq to determine whether, like mutating SMN or the R1810 residue of CTD to alanine, mutating FUS or TDP-43 had more general effects on transcription termination. We performed RNAPII ChIP-seq (with the N20 antibody), generating >10 million unique reads from samples with FUS or TDP-43 knocked out using the CRISPR system. The control sample was HEK293 cells transfected with a scrambled guide RNA. Our ChIP-seq results indicated that the loss of FUS or TDP-43 association with RNAPII causes RNAPII to stall in termination regions in a genome-wide pattern (Fig. 40). For the analysis, we only used genes with reliable ChIP-seq signals (top 5 or 10% by expression; ~800-2000 genes) as we ran into issues with sequencing depth and background interference when more genes were included in the analysis.

For the calculations, the promoter stalling ratio was calculated as promoter reads normalized to gene body reads, whereas the termination stop ratio was calculated as termination reads normalized to gene body reads (Fig. 40A). When the ratio for the FUS or TDP-43 knock-out sample is compared to the control knock-out, a shift of the curve to the right indicates increased RNAPII accumulation at promoter or termination regions with respect to the gene body [549]. We observed that the knock-out of FUS indeed caused an accumulation of RNAPII
Figure 40 FUS and TDP-43 regulate RNAPII Promoter and Terminator Stalling (ChIP-seq)
Figure 40 FUS and TDP-43 regulate RNAPII Promoter and Terminator Stalling (ChIP-seq)
Figure 40 FUS and TDP-43 regulate RNAPII Promoter and Terminator Stalling (ChIP-seq)

A. Schematics for the calculation of the stalling ratio at promoter and termination regions.

B. RNAPII ChIP-seq (with the N20 antibody in HEK293 cells). Compared to the scrambled KO, FUS or TDP-43 KO led to increased RNAPII accumulation in the promoter regions of genes. Top 1000 genes by expression were analyzed.

C. RNAPII ChIP-seq (with the N20 antibody in HEK293 cells). Compared to the scrambled KO, FUS or TDP-43 KO led to increased RNAPII accumulation in the termination regions of genes. Top 1000 genes by expression were analyzed.

D. Another way of displaying the ChIP-seq analysis. When normalizing the reads to the gene body, the knock-out of FUS or TDP-43 led to increased RNAPII accumulation in the promoter and termination regions genome-wide. 5-10% most highly expressed genes with reliable ChIP-seq signals were analyzed, with the number of genes displayed on the bottom right of the graph.

E. RNAPII ChIP-seq results for several housekeeping genes were displayed in detail using the Integrative Genomics viewer. The promoter peaks are displayed to the left, and the RNAPII termination regions are displayed with red underlines.
at promoters, in agreement with the literature (Fig 40B, 40D) [453]. Supporting the roles of FUS and TDP-43 in regulating RNAPII termination, we observed genome-wide accumulation of RNAPII in the termination regions when either FUS or TDP-43 was knocked out (Fig. 40C, 40D). RNAPII ChIP-seq results for several housekeeping genes are displayed in Fig. 40E. These observations further supported our hypothesis that FUS and TDP-43 are involved in preventing R-loop accumulation, acting downstream of the CTD R1810me2s-SMN pathway for the regulation of RNAPII termination (model in Fig.41).

3. Discussion

Amongst the SMN interactors that we and others have identified, the RBPs FUS and TDP-43 are known to be mutated in familial ALS, a disease akin to SMA. Therefore, I focused on elucidating if like SMN and SETX, FUS and TDP-43 also regulate RNAPII termination. I hypothesized and showed that the R1810me2s modification of the RNAPII CTD does play a role in stabilizing the interactions of FUS and TDP-43 with RNAPII, likely through SMN. Based on Co-IP and ChIP experiments, CTD R1810, PRMT5, and SMN contribute substantially to the recruitment of FUS and TDP-43 to RNAPII (Fig. 29, 30).

FUS and TDP-43 are known to play multiple roles in RNAPII transcription; it has however, not been shown that they regulate RNAPII termination by preventing R-loop formation, and if this function is associated with the ALS disease state. It would not be surprising if these proteins can prevent R-loop formation during splicing or upon RNAPII termination, as several splicing factors are already known to resolve R-loops and prevent DNA damage [206, 211, 226]. The ALS disease phenotype caused by FUS and TDP-43 mutations is
Figure 41 Model: SMN as a nucleator of Rme2s proteins involved in transcription termination and neurodegeneration (top), through the recruitment of FUS or TDP-43 to prevent R-loop formation and DNA damage at the RNAPII termination sites (bottom).
at least partially related to their RNA binding activities [444, 550]. Therefore, I also tested whether the expression of a truncated TDP-43 lacking its RNA binding domains would prevent R-loop formation like full-length wild type TDP-43, and found that it could not. In summary, my results point to the likelihood that the CTD R1810me2s modification allows SMN to recruit Rme2s-containing proteins such as FUS and TDP-43 to the transcribing RNAPII. Defects in FUS and TDP-43 recruitment to RNAPII may lead to R-loop accumulation that stalls RNAPII and causes DNA damage, which may contribute to neurodegeneration.

Besides FUS and TDP-43, other RBPs (e.g. TAF15, EWSR1, hnRNPA3, hnRNPA2/B1, and hnRNP1) were also found to be associated with ALS/FTD by pathological mutations or as components of inclusion bodies [540]. Thus they are likely performing similar functions in the repression of R-loop formation and DNA damage, and these proteins will be good candidates for follow-up studies.
Chapter 4: Discussion, conclusions, and future directions

4.1 Discussion

4.1.1 The identification of the R1810me2s modification on the RNAPII CTD

I have found that R1810 in the RNAPII CTD is symmetrically dimethylated by the arginine methyltransferase PRMT5. The presence of this modification was substantiated by a number of observations with generic antibodies against Rme2s (SYMM10 and Y12) and a specific antibody raised against a CTD R1810me2s-containing peptide. Additionally, the presence of the CTD R1810 residue and PRMT5 are needed for the association of SMN with RNAPII and the recruitment to transcribing RNAPII of SMN, whose Tudor domain is specific for Rme2s binding. My study also indicated that symmetric dimethylation of R1810 in the RNAPII CTD causes the direct recruitment of SMN and indirect recruitment of the RNA:DNA helicase SETX, which is need for resolving R-loop structures in RNAPII termination regions, thereby facilitating the efficient release of the RNAPII (see model in Fig. 25).

Recently, PRMT9 was characterized as a type II PRMT, and it has been shown to methylate SAP145 (at R508 that embedded in the WCFKRKYLYQ sequence) of the U2 snRNP in vivo [270]. PRMT9 appears to be a much more specific Rme2s methyltransferase than PRMT5, requiring the arginine to be bracketed in the KRK sequence for the methylation in vivo [551]. Additionally, as estimated from RNAseq data (not shown), PRMT9 appears to be expressed at a much lower level (<50 fold) in comparison to PRMT5 in most cell lines [551], including the ones that I worked with. Therefore, PRMT5 may be the only type II PRMT that is important for generating the Rme2s mark on arginine residues embedded in PGM-containing sequences (which includes the CTD R1810 residue that is embedded in a PRY sequence) [267].
FCP1 (CTDP1), a phosphatase that is known to stimulate transcription elongation through the dephosphorylation of pSer2 and pSer5 on the RNAPII CTD, as well as two Mediator components (CDK8 and 19), are known to recruit PRMT5/WDR77 to RNAPII [324, 537]. It is likely that PRMT5 is similarly recruited to methylate the CTD at R1810 through the elongation phase of the transcription cycle. To test if FCP1 or CDK8/19 may recruit PRMT5 to the RNAPII CTD \textit{in vivo}, knock-down or knock-out cell lines for these proteins could be generated to test if there are changes in the association of PRMT5 with RNAPII, in the level of the CTD R1810me2s modification, and in the recruitment of SMN to RNAPII. The cell lysates of these knock-down lines could be incubated with recombinant RNAPII CTD and PRMT5 in an \textit{in vitro} methylation assay to identify which of these adaptor component could enhance PRMT5’s activity on the CTD R1810 residue \textit{in vitro}.

CARM1 and PRMT5 are known to methylate arginine embedded in GAR and PGM motifs [267, 273, 275, 325-328] and, notably, the RNAPII CTD residue R1603 is embedded in the sequence PRS, which makes it a possible substrate for PRMT5 and CARM1. Though it is known from the literature that CARM1 was not found to methylate R1603 \textit{in vitro} or \textit{in vivo} in the tested conditions [67], we observed that both R1603 and R1810 of the RNAPII CTD are PRMT5 substrates \textit{in vitro}. It is possible that, \textit{in vivo}, the presence of adaptors such as FCP1 or a mediator subunit (CDK8 or 19) can enhance PRMT5’s methylation activity preferentially on the RNAPII CTD R1810 site. To test the function of these adaptors \textit{in vitro}, GST recombinant CTD-N terminal (that contains R1603) and CTD-C terminal (that contains R1810) could be incubated with purified PRMT5 together with FCP1, CDK8 or 19 to determine if R1810 methylation can be enhanced above that of R1603 by the addition of these purified adaptors \textit{in vitro}.
PRMT5 has many known substrates in gene regulation, including the SPT5 subunit of the RNAPII regulator DSIF that can function as a positive or a negative regulator in RNAPII elongation [279]. An SPT5 mutant that is defective in PRMT1 and PRMT5 methylation led to a loss of RNAPII promoter-proximal pausing, which was shown to be caused by the increased binding of the mutant SPT5 to RNAPII [279]. It was not conclusive, however, from this study whether the Rme2a modification that is generated by PRMT1 and the Rme2s modification that is generated by PRMT5 play antagonizing or synergistic roles in controlling SPT5’s role in RNAPII elongation. It is not uncommon for Type I and Type II PRMTs to compete for the deposition of me2a and me2s modifications on the same arginine substrates [273], thereby altering downstream protein-protein interactions. There exist many examples of alternative modifications of the same arginine residue, as mentioned in the introduction [274-276]. Similarly, I cannot conclude from my study whether the R1810me2a mark generated by CARM1 and the R1810me2s mark generated by PRMT5 on the RNAPII CTD antagonize each other in RNAPII regulation, though I found in ChIP experiments that CARM1 is not needed for RNAPII termination or the recruitment of SMN to RNAPII (Fig. 16A). It seems likely that these two marks play synergistic roles in the resolution of R-loops, with TDRD3-TOP3B recruited by the R1810me2a modification on the RNAPII CTD playing a more dominant role at promoters [227, 232].

So far, JMJD6 is the only known possible arginine demethylase, and several reports have suggested the possibility that JMJD6 may demethylate H4R3me2a and H4R3me2s, leading to a release of RNAPII promoter-proximal pausing [352, 356]. It is conceivable that an arginine demethylase such as JMJD6 may regulate the relative levels of R1810me2s and R1810me2a on the RNAPII CTD. A study to investigate this could first be done in vitro to determine if a CTD
with methylated R1810 is a JMJD6 substrate or not, and if R1810me2a or R1810me2s is a better substrate for JMJD6. In vivo, RNAPII could be immunoprecipitated from wild type or JMJD6 knocked-out cell lines to determine if there exists a change in the R1810me2s or R1810me2a levels, using Rme2a- and Rme2s-specific antibodies. As well, the recruitment of SMN to RNAPII, the release of RNAPII from termination sites, and the resolution of R-loops may all be affected upon the loss of JMJD6.

4.1.2 SMN as an adaptor for the 3’ end processing and RNAPII termination factors

I have shown that SMN recognizes the RNAPII CTD with the R1810me2s modification in vitro and in vivo. SMN can self-aggregate through its N-terminal K-rich domain and its C-terminal YG box to potentially form a multimeric adaptor to assemble Rme2s containing proteins into a complex [388, 538, 539]. While the human genome contains two copies of the SMN gene (SMN1 and SMN2), the mutation of one allele of SMN1 can lead to a decrease of total SMN protein sufficient to cause the SMA disease phenotype [388]. In particular, mutations in SMN1’s oligomerization and Tudor domains sometimes occur in cases of SMA [383, 388]. Because the SMN complex assembles snRNPs for the spliceosome and cytoplasmic RNPs, these two functions have been proposed to contribute to SMA pathology when SMN is defective [388]. Analogous to its role in spliceosome and RNP assembly through binding to Rme2s-containing proteins, we believe that SMN acts similarly to assemble a termination complex on the RNAPII CTD through binding to Rme2s modified termination factors. Though the binding of the SMN Tudor domain to R1810me2s containing peptides is not particularly strong in in vitro measurements, we believe that cooperative interactions of various Rme2s modified proteins (e.g. FUS) with both SMN and the RNAPII CTD could lead to the formation of a stable
complex. In this scenario, the R1810me2s-SMN interaction would play an enhancement role in recruitment of the termination complex to the RNAPII CTD.

We and others have found that the SMN and SETX interact and participate in a common pathway that is important for R-loop disassembly and transcription termination by RNAPII (Fig. 25). From my study, I observed that the interaction between SMN and SETX requires PRMT5, presumably because the SMN Tudor domain binds to an Rme2s modification on SETX (Fig. 10C). Because SETX is the locus for ALS4 and AOA2 mutations [239, 240], this suggests that defects in the RNAPII termination pathway due to R-loop accumulation may be related to various types of neurodegenerative diseases. In fact, based on the interaction between SMN and FUS, it has already been argued that ALS and SMA are linked in a common molecular pathway and are, therefore, related diseases [399].

In my study, I first observed that the loss of SMN led to increased RNAPII and R-loop accumulation in the termination regions of the β-actin and the GAPDH genes, as was previously shown upon the loss of SETX [91]. A similar phenotype was detected upon the addition of SMN inhibitors that block its binding to Rme2s-containing motifs, indicating that the function of the SMN Tudor domain is indispensable for the regulation of RNAPII termination and R-loop resolution. With the addition of SMN inhibitors, I also observed a loss of interactions among RNAPII, SMN, FUS, TDP-43, and XRN2 by co-IP (Fig. 39), as well as a loss of recruitment of SMN, FUS, TDP-43, and SETX to RNAPII by ChIP (Fig. 38C). These findings indicated that the Tudor domain of SMN indeed acts as a scaffold for the assembly of the termination machinery on the RNAPII CTD. With the SMA disease cell lines, with mutations in SMN, I consistently observed an increase of RNAPII and R-loop accumulation in termination regions, suggesting that the importance of SMN for RNAPII termination may be associated with disease
pathology in neurodegeneration. I also showed that stalling by RNAPII in termination regions due to the loss of SMN or the RNAPII R1810A mutation occurs on a genome-wide scale, which further substantiated the importance of the CTD R1810me2s modification and SMN for the regulation of RNAPII termination.

Among ~150 human proteins that have been shown by mass spectrometry to contain dimethylarginine [392], a remarkably large number of them are involved in RNA metabolism, transcription elongation, and termination by RNAPII. These include RBPs such as FUS, EWSR1, TAF15, and TDP-43 that tend to be mutated in the ALS/FTD disease state, the 5’ → 3’ exonuclease XRN2, 4 subunits of the cleavage and polyadenylation specificity factors (CPSF1, 5, and 6; PCF11), one subunit of the cleavage stimulatory factor (CstF2), 3’ end processing factors such as RBBP6 and WDR33 (which is known to bind to the PAS motif) [191, 192, 547], the RNAPII elongation factor SPT5, the CTD phosphatase FCP1 (CTDP1), the DNA:RNA helicase DHX9, which is known to interact with SMN and SETX [198, 234], and a number of polyA-binding proteins (PABP1, 2, and 4). Among these proteins that contain dimethylarginine, it is known that POLR2A, FUS, EWSR1, CPSF (1, 5, and 6), FCP1, and PABP (1 and 2) contain symmetrical dimethylation modifications on arginine [391, 552], whereas it is unknown if the others contain Rme2s or Rme2a (or both). It is likely that Rme2a modifications, if present on these proteins, play little role in the regulation of RNAPII termination, because, unlike the knock-down of SETX, SMN, or PRMT5, upon CARM1 knock-down, RNAPII did not accumulate in the 3’ ends of genes (Fig. 16A). Another possibility could be that there exists an additional type I PRMT (e.g. PRMT1) that can methylate these proteins in a redundant manner with CARM1, as it is known that PRMT1 and CARM1 indeed play synergistic roles in the regulation of gene expression, such as for STAT5-controlled genes [274, 553].
4.1.3 The roles of FUS and TDP-43 in R-loop resolution and RNAPII termination

Like SETX, the RBPs FUS and TDP-43 are sometimes mutated in familial ALS (ALS6 and ALS10, respectively), they co-localize with RNAPII, and they interact with SMN [399, 554, 555]. It is therefore reasonable to hypothesize that the R1810me2s modification of the RNAPII CTD may enhance the recruitment of FUS and TDP-43 through SMN, thereby allowing these RBPs to prevent the nascent RNAs from re-hybridizing with ssDNA, and thereby facilitating transcription termination. It appears that the CTD R1810 residue, PRMT5, and SMN contribute substantially to the recruitment of FUS and TDP-43 to RNAPII, as seen in co-IP (Fig. 29) and ChIP (Fig. 30A-C) experiments. In co-IP, the TDP-43-RNAPII interaction is consistently seen to be disrupted upon the loss of SMN, whereas the FUS-RNAPII interaction appears to be stronger and sometimes remains unaffected after the loss of SMN (Fig. 29). It is likely that FUS may associate with RNAPII in multiple ways, as it had already been shown that the disordered LC domain in FUS can form oligomers with the RNAPII CTD in vitro [433, 441].

One may suspect that FUS and other LC domain-containing proteins can form heteromers and be recruited to the RNAPII CTD through LC domain-mediated interactions in vivo [431]. If so, it is possible that, through binding to the CTD and SMN, FUS may also stabilize the CTD R1810me2s-SMN interaction. To further stress the importance of RNAPII regulation by a set of LC domain-containing proteins, other LC domain-containing RBPs (e.g. TAF15, EWSR1, hnRNPA3, hnRNPA2/B1, and hnRNP1) are also known to interact with RNAPII and SMN, and losses of these proteins are also implicated in ALS/FTD pathology [431, 540] (Table 1). In my ChIP experiments, after shRNA-mediated stable knock-downs, siRNA-mediated transient knock-downs, or CRISPR-mediated knock-out of FUS or TDP-43 (Fig. 30D), I consistently observed that the recruitment of FUS and TDP-43 to RNAPII was interdependent. These observations
were further supported by ChIP experiments with ALS patient cell lines, in which the associations of FUS and TDP-43 with RNAPII also tended to be reduced simultaneously (not shown). These in vivo observations substantiated the formation of a FUS-TDP-43 heteromer in the regulation of RNAPII transcription.

FUS was known to be associated with the RNAPII CTD and to regulate the pSer2 level of the CTD, and was also known to promote RNAPII termination and to cause 3’end-APA changes [453, 490]. The loss of FUS can lead to intron retention [543] and DNA damage [513]. It seems conceivable that the role of FUS in repressing R-loops may be the mechanistic basis for all the aforementioned observations. Like FUS, TDP-43 also binds to and sequesters nascent transcripts, and its loss also leads to intron retention that may be explained by un-resolved R-loop formation during splicing [401]. It will not be surprising if FUS and TDP-43 can prevent R-loop formation during splicing or upon RNAPII termination, as several splicing factors are already known to repress R-loops and, thereby, prevent DNA damage [206, 211, 226]. Though FUS and TDP-43 are known to play multiple roles in RNAPII transcription, mine is the first study to show that they regulate RNAPII termination by preventing the buildup of R-loops, thereby preventing DNA damage. Such functions may commonly be compromised in the ALS disease state.

4.1.4 Possible mechanisms linking R-loop accumulation to neurodegeneration

R-loops form predominantly on a subset of GC rich promoters, in termination regions, on genes with high transcription rates, at collision sites between RNAPII and the replication machinery, and on genes with anti-sense transcription where head-on collisions between RNAPII molecules tend to occur [91, 199, 202, 230, 238, 517]. RNAPII stalling due to R-loop
accumulation is known to alter gene expression and splicing and is linked to DNA damage [201, 206, 211, 215, 226]. Upon the mutation of RNAPII CTD residue R1810 to alanine, or upon the loss of SMN, SETX, FUS, or TDP-43, I observed an accumulation of RNAPII in the termination region of genes such as β-actin. In neurons, the loss of β-actin expression is associated with SMA pathology in particular, leading to impaired F-actin bundle formation and the loss of RNP transport to growth cones [388]. The overexpression of Plastin3, which is known to increase F-actin levels, can rescue the SMN-dependent SMA disease phenotype [429]. Although SMN, SETX, FUS, and TDP-43 are ubiquitously expressed proteins, and we observed genome-wide effects on transcription termination upon their losses (Fig. 21, 40), it is possible that termination defects on a subset of genes (including the highly transcribed β-actin gene) may be more directly linked to the pathology of neurodegeneration in post-mitotic neurons.

Indeed, it is known that defects in transcriptional termination can lead to reduced protein expression in human cells, with the argument being that efficient termination is needed for proper splicing and polyadenylation of the nascent RNA, and if these processes are compromised, the exosome may degrade the aberrantly retained RNA at the termination regions [556]. However, RNA-seq revealed that there do not appear to be drastic gene expression changes, aside from snoRNAs, between the WT and R1810A mutant RNAPII samples after 3-10 days of α-amanitin treatment to get rid of endogenous RNAPII [67] (Fig.22B). This is not to say that, when compared to the WT, the R1810A mutant RNAPII in post-mitotic neurons, after decades of genotoxic stress, would function equally well in the maintenance of gene expression and cell survival in vivo. Because it had been shown that RNAPII stalling leads to changes in co-transcriptional alternative splicing [55], and a slow moving RNAPII may favor alternative exon inclusion or exclusion, depending on the splicing factors recruited [117, 118], we therefore
also hypothesize that difficulty in R-loop resolution may alter cassette exon splicing and lead to intron retention. It is also known that selective deletion of PRMT5 or SMN in neural cells can lead to aberrant splicing changes and postnatal death in mice, though it is not known if stalled RNAPII due to the build up of R-loops may be linked to the aberrant splicing changes that are linked to SMA [423, 424, 427, 557].

Through RNA-seq, we observed a higher GC content to be correlated with exon exclusion and intron retention in the R1810A mutant when compared to the WT, though we do not know what splicing factors caused these changes (Fig. 23). Between the WT and the R1810A samples, our preliminary analysis also revealed many APA changes, but we did not observe a bias in GC or G content nor detect a change in the median 3’UTR length. We also looked for a correlation between sites of cassette exon changes and RNAPII peak changes between the WT and R1810A samples, but observed no significant correlation. We, however, have not done similar correlational analysis for paused RNAPII and intron retentions. In summary, our RNA-seq analysis suggested that the differential AS observed for cassette exons between the WT and the R1810A mutant may be better described by changes in the recruitment of splicing factors rather than by changes in RNAPII kinetics between these samples [115]. Therefore, we suspect that RBPs such as FUS and TDP-43 may be the likely splicing factors for the AS changes observed between WT and the R1810A mutant [115].

Another outcome of difficulty in R-loop resolution may be genome instability, as the stalled RNAPII may collide with replication forks to generate a cellular response akin to the DDR [201, 212-214]. Given that ALS typically develops after several decades of life, it is plausible that the accumulation of R-loops and DNA damage in post-mitotic neurons may lead to mutations and genome rearrangements that ultimately result in neurodegeneration.
others have examined and found that the loss of SETX, SMN, FUS, or TDP-43, or the RNAPII R1810A mutation is associated with an increase of DNA damage in termination regions wherever R-loops and RNAPII accumulate [91, 238]. From two recent studies, it is likely that R-loops induce DSBs through the aforementioned TC-NER pathway [211, 238], and importantly, mutations in certain TC-NER components are known to cause progressive neurodegeneration [245, 246]. However, understanding the exact mechanism by which TC-NER leads to DSBs requires further clarification. As well, another recent study showed that stalled RNAPII and the accumulation of R-loops due to DNA damage can directly trigger a non-canonical ATM signaling pathway, which leads to profound changes in the maturation of the late-stage spliceosome and consequently alters the splicing patterns (including intron retentions) of many genes [219]. So far we have only used γH2AX as a marker for DNA damage, but different assays, including the comet tail assay, staining and ChIP experiments with other DDR markers, should be used to better characterize the exact nature of the DNA damage that is triggered by the accumulation of R-loops in the termination regions of genes.

In summary, since R-loops may appear at promoters, within gene bodies, or in termination regions, and can lead to the silencing of a particular locus, we think that it is likely that R-loop-induced DNA damage affects gene expression and splicing changes in the long run, especially in post-mitotic neurons. Indeed, R-loops at promoters, gene bodies, and termination regions are known to contribute to neuronal disorders such as FMR, ALS/FTD, and Friedreich’s Ataxia [91, 221-224].
4.2 Conclusions

I have found that the RNAPII CTD arginine residue R1810 is symmetrically dimethylated (me2s), allowing it to recruit the Tudor domain of the SMN protein, which is sometimes mutated in SMA. SMN can oligomerize and interact with SETX, which is mutated in ALS4 and AOA2. From the literature, it is known that many RNAPII termination factors contain dimethylated arginine residues that could potentially interact with SMN. Among these candidates, I studied FUS and TDP-43, which are sometimes mutated in ALS/FTD and are involved in numerous aspects of gene regulation. I showed that FUS and TDP-43 act downstream of the CTD R1810me2s-SMN pathway. I showed that like SETX, RNAPII CTD R1810me2s, SMN, FUS, and TDP-43 are all important for repressing the buildup of R-loops that would otherwise stall RNAPII in termination regions and lead to DNA damage. Additional candidate proteins that may prevent R-loop formation include EWSR1 and TAF15, which are similar to FUS in domain structure, interact with FUS, and are also known to sometimes be mutated in ALS/FTD. Another likely candidate in this pathway is DHX9, a 3’ → 5’ helicase that resolves RNA:DNA hybrids, which also interacts with SMN and the RNAPII machinery [234]. DHX9 could also be indispensable for resolving transcription induced R-loops, and may be needed to coordinate the degradation of the nascent RNA by the 3’ → 5’ exosome for the proper release of RNAPII. Because DHX9, SETX, and BRCA1 interact together and are implicated in DNA repair, it will be interesting to test if DHX9 also facilitates the entry into TC-NER upon the detection of unresolved R-loops [238, 558].
4.3 Future directions

4.3.1 RNA-seq analysis for gene expression, splicing, and APA changes

Although RNA-seq revealed that there do not appear to be drastic gene expression changes between the WT and the R1810A mutant RNAPII samples after 3-10 days of α-amanitin treatment [67] (Fig. 22B), this may be caused by the limited duration of the expression of the α-amanitin resistant RNAPII prior to sample collection. We, therefore, propose to use RNA-seq to examine the long term effects of the stable CRISPR-mediated knock-outs of SMN, FUS, and TDP-43, and the CRISPR-mediated knock-in of R1810 to alanine on the RNAPII CTD in HEK293 cells. We previously only had the RNAPII WT and R1810A mutant expressed in Raji cells; therefore, we are currently using the CRISPR system to make the R1810A mutant in HEK293 cells. We hope to observe a significant overlap of changes in gene expression, AS, and APA across the FUS, TDP-43, and SMN knock-outs and the RNAPII R1810A mutant to consolidate the idea that these components function in the same pathway in RNAPII transcriptional regulation.

From the literature, it is known that the loss of FUS or TDP-43 will alter gene expression and AS somewhat differently, as was previously shown in brain [454, 482]. We predict that the loss of SMN or the CTD R1810A mutation will cause more gene expression and AS changes in the long run that will bracket those changes due to the loss of either FUS or TDP-43. The reason behind this expectation is that SMN is the adaptor that enhances the association of both FUS and TDP-43 with RNAPII during elongation and termination.

So far, we have obtained gene expression data for the SMN, FUS, and TDP-43 knock-out samples, and compared to the control, we observed both drastic up-regulation and down-regulation of the expression of various genes in each sample (Fig. 42). A large fraction of these
gene expression changes do overlap among the knock-out samples. Not surprisingly, GO analysis suggested that many of the genes involved in the regulation of neural development and differentiation are significantly down-regulated and that genes involved in DNA repair and cell cycle control are significantly up-regulated in the knock-out samples (Table 5). It is already known that DNA damage escalation is linked to reduced gene expression in aged cortex, causing the reduction of the expression of genes that are essential for synaptic plasticity, vesicular transport, and mitochondrial function [506], and this situation is exacerbated in neurodegenerative diseases [507]. The preliminary observations from our knock-out RNA-seq samples do support the idea that defects in RNAPII transcription termination and consequent DNA damage can be a mechanistic basis for gene down-regulation in neurodegeneration [501-505]. Interestingly, upon the knock-out of SMN, I also observed increased expression for many of the spliceosomal components such as the Sm proteins, therefore suggesting a compensatory mechanism to overcome the loss of the spliceosome assembly adaptor SMN. The loss of SMN also led to the upregulation of many genes that are known to be involved in the pathology of neurodegeneration, including Parkinson’s, Alzheimer’s, and Huntington’s diseases (Table 5).

This is somewhat intriguing, and may be explained by the possibility that similar compensatory mechanisms exist to up-regulate other RNP’s, which normally require SMN for the assembly into complexes. The RNP’s are often over-represented in the pathology of neurodegeneration, and their over-expression and the consequent formation of aggregates upon the loss of SMN may in fact exacerbate the degenerative phenotype of the SMA disease.

In addition to its role in regulating pre-mRNA splicing through the recognition and cleavage of the 5’ splice site, the U1 snRNP has been characterized to regulate pre-mRNA premature cleavage and polyadenylation in a separate process called U1 telescripting [559, 560].
Figure 42 SMN, FUS, and TDP-43 knock-outs lead to overlapping gene expression changes.
Figure 42 SMN, FUS, and TDP-43 knock-outs lead to overlapping gene expression changes

The numbers indicate down-regulated (A) and up-regulated (B) genes in the KO samples in comparison with the control KO sample, which was generated with scrambled guide RNA.

The numbers indicating the numbers of genes with significant expression changes (without cutoff) in each sample are shown on the left. The numbers on the right represent the number of genes with significant expression changes when the cutoff is set to >5 FPKM, only considering the more robustly expressed genes.
Table 5 GO term analysis of gene expression changes upon the knock out of FUS, TDP-43, or SMN
This process requires the U1 snRNA-pre-mRNA base pairing to suppress cleavage from a nearby cryptic PAS located in an intron, and the process is independent from the catalytic splicing activity of the U1 snRNP [560]. And recently, it’s also been suggested that RNAPII, U1, SMN, and FUS form a complex that is then trapped in the cytoplasm when FUS is mutated (R495X) in ALS cases [561]. This then implies the possibility that SMN and the RNAPII CTD R1810me2s modification may, as well, play a role in the regulation of APA, promoting the lengthening of 3’UTR in a complex with FUS and the U1 snRNP. Therefore, it is a possibility that RNAPII termination defects seen in SMN, FUS, or TDP-43 knock-downs, or the mutation of R1810 to alanine on POLR2A may be linked to a general bias towards a shortened form of 3’ UTR on the corresponding mRNA transcripts [559, 560].

We have previously examined APA changes due to the mutation of R1810 to alanine on the RNAPII CTD, and observed no obvious changes in the median 3’UTR length as compared to the wild type control. It is likely that patterns of APA changes will be more apparent when the most highly transcribed transcripts are examined, because R-loops and stalled RNAPII are known to accumulate there more readily. Our ChIP-seq analysis did indeed support the idea that the highly transcribed regions tend to stall RNAPII more readily, after the loss of SMN or in the presence of the mutation of CTD R1810. Therefore, we hypothesize that it is more likely to observe APA changes with the predicted shortened 3’UTR usage on the most highly expressed genes due to the buildup of R-loops that caused the defective termination of RNAPII.
4.3.2 Gene looping and promoter-proximal RNAPII termination

When I looked at the RNAPII ChIP-seq signals at the promoters, when RNAPII CTD R1810 is mutated to alanine, or when SMN, FUS, or TDP-43 was knocked down, there appeared to be various changes that may be explained by changes in gene looping or promoter-proximal RNAPII termination. From the ChIP-seq analysis, it seems likely that SMN, FUS, and TDP-43 have different roles in the regulation of RNAPII promoter-proximal pausing, which may or may not be affected by their regulation of RNAPII termination at the 3’ ends. Specifically, when comparing the most highly expressed genes with robust ChIP-seq signals, the CTD R1810A mutation led to a loss of RNAPII promoter directionality similar to that observed after the loss of the CTD Ser5 phosphatase Ssu72 in yeast, which suggested compromised gene looping [53]. In contrast, the loss of SMN led to a reduction of the RNAPII promoter signal that is suggestive of a defect in RNAPII termination, which may compromise RNAPII release and re-initiation. This loss of RNAPII promoter signal was also observed in a previous study when RNAPII termination was compromised [199]. However, again in contrast to these phenotypes, the loss of FUS or TDP-43 led to an increase of the RNAPII promoter signal, which seems to suggest a possible increase in RNAPII initiation or promoter-proximal RNAPII pausing.

Transcription may initiate in both directions from a promoter region because promoter regions are nucleosome depleted [562], and bi-directional promoters for RNAPII initiation occur commonly from yeast to humans. Such processes can lead to the synthesis of unwanted transcripts that may negatively affect gene expression [562]. As it has been suggested that promoter-proximal RNAPII termination and gene looping can enhance promoter directionality, it is worth testing if there is a gene looping defect, and an increase in bi-directional transcription, when RNAPII doesn’t properly terminate distally. Functional and physical interactions between
RNAPII initiation and termination factors are known to occur in various species, supporting the generality of gene looping [163, 562]. Currently, chromosome conformation capture (3C) analyses of single genes or genome-wide studies by ChIA-PET are often performed to examine changes in gene looping. Such studies could be carried out after the knock-down of SETX, XRN2, SMN, TDP-43, or FUS, or the mutation of R1810 to alanine on the RNAPII CTD. If single gene analysis by 3C is informative, a genome-wide study could then be performed.

From published work, it is known that cleavage and polyadenylation factors, as well as termination factors like XRN2, SETX, and the exosome, participate in promoter-proximal RNAPII termination[91, 97-99]. Therefore, it is important to examine if the CTD R1810A mutation or the loss of SMN, FUS, or TDP-43 affect promoter-proximal RNAPII termination. To do this, one may check for expression changes of divergent RNAs derived from the bi-directional promoters, after the knock-down of SETX, SMN, TDP-43, or FUS, or when RNAPII CTD R1810 is mutated to alanine, on genes with and without gene looping. If these divergent RNAs are uncapped and are naturally unstable, such experiments may need to be performed in a background in which the 5’ → 3’ exonuclease XRN2 is knocked out.

4.3.3 Is TDRD3a component of TC-NER that regulates RNAPII degradation?

It is known that TDRD3 recognizes the RNAPII CTD R1810me2a modification in vitro [67], and I showed that the TDRD3-RNAPII interaction occurs in vivo and is further enhanced after UV (40 J/M2)-induced stress (Fig. 43A, 43B). TDRD3, a protein that contains a Tudor domain for Rme2a binding and a UBA domain for binding to polyubiquitinated lysine (Fig. 43B), increases its protein level in cells with increased γH2AX staining for elevated DNA damage upon UV stress (Fig. 43C, Fig. 43I). These observations suggest that TDRD3 may
participate in the aforementioned TC-NER pathway for recycling and/or degrading stalled RNAPII upon DNA damage. Interestingly, based on immuno-staining, TDRD3 increases in cells with DNA damage (γH2AX), and these cells also show a consistent loss of active RNAPII (with pSer2, pSer5, or pSer7 CTD modifications) but not the non-phosphorylated form of RNAPII (Fig. 43D). However, upon loss of NEDD4 or CARM1, or both, these TDRD3 positive cells no longer show a decrease of active forms of RNAPII (that are marked with pSer2 or pSer5 CTD modifications) (Fig. 43E-G). These observations suggest that TDRD3 regulates the level of active RNAPII in response to DNA damage, and that both the UBA and Tudor domains of TDRD3 that read the marks generated by NEDD4 and CARM1, respectively, are needed for this function.

Supporting the idea that R1810me2a and plausible lysine ubiquitination in its vicinity could regulate TDRD3 binding to the CTD, previous studies using mass spectrometry have identified the presence of multiple ubiquitination sites on the RNAPII CTD near the R1810 site in mammalian cells [262]. It is known that upon DNA damage, stalled RNAPII can become ubiquitinated by ubiquitin ligases (e.g. NEDD4, WWP2, and the elongin E3 ligase complex) [96, 262, 263] for recycling and degradation. Based on western blotting, unlike the knock-down of NEDD4 (Fig. 43H), which led to an increased level of RNAPII 3-6 hours post UV stress, the knock-down of TDRD3 did not block RNAPII degradation (Fig. 43I, 43J). It seems enigmatic that the TDRD3-RNAPII interaction led to the removal of active phosphorylated forms of POLR2A, but did not direct POLR2A for degradation.

Therefore, I hypothesize that, following DNA damage, TDRD3 in complex with TOP3B may function to resolve R-loops and to channel stalled RNAPII to backtrack and be recycled,
Figure 43. TDRD3 regulates RNAPII degradation upon UV stress
Figure 43  TDRD3 regulates RNAPII degradation upon UV stress
Figure 43. TDRD3 regulates RNAPII degradation upon UV stress

A. IP with the N20 antibody for POLR2A from HEK293 WCL, 6 hours post UV treatment (40J/M2). Western blots were performed with the indicated antibodies.

B. IP with the N20 antibody for POLR2A from HEK293 WCL, without UV, 1, 3, or 6 hours post UV treatment (40J/M2). Western blots were performed with the indicated antibodies. The schematic of the domain composition of TDRD3 is displayed below.

C. Immuno-staining of TDRD3 (red) and γH2AX (green) 3 and 6 hours post UV (40J/M2) treatment in control HEK293 cells.

D. Immuno-staining of TDRD3 (green) and pSer5, pSer2, pSer7, or non-phospho (8WG16) forms of RNAPII CTD (red) without UV, or 6 hours post UV (40J/M2) treatment in control HEK293 cells.

E. Immuno-staining of TDRD3 (green) and pSer5 form of RNAPII CTD (red) without UV, or 6 hours post UV (40J/M2) treatment in HEK293 cells with shCARM1 or with siNEDD4+shCARM1 knock-downs.

F. Immuno-staining of TDRD3 (green) and pSer2 form of RNAPII CTD (red) 6 hours post UV (40J/M2) treatment in HEK293 cells with shCARM1, or siNEDD4 or with siNEDD4+shCARM1 knock-downs.

G. Statistics and graph of the effect of UV (40J/M2) treatment on control cells, or cells lacking CARM1, NEDD4, or both.

H-J. IP with the N20 antibody for POLR2A from HEK293 WCL with control, NEDD4, or TDRD3 2-day siRNA knock down, followed by +/- UV treatment (40J/M2). Western blots were performed with the indicated antibodies.
thereby facilitating the entry of TC-NER factors for the repair of the DNA damage upon UV stress. If TC-NER fails, the E3 ubiquitin ligase NEDD4 adds K63-linked polyubiquitin chains to POLR2A, which are shortened to monoubiquitylated forms to generate a substrate for the elongin E3 ubiquitin ligase complex, which then adds K48-linked polyubiquitin chains to POLR2A, marking it for degradation by the proteasome [246]. The binding of TDRD3 to ubiquitylated POLR2A through its UBA domain may block the ubiquitin chain extension for POLR2A degradation (Fig. 43I, 43J). It will thus be interesting to determine whether, following UV stress, with the RNAPII CTD R1810A mutation or with the loss of the TDRD3-TOP3B interaction, the TC-NER pathway becomes mis-regulated and if the series of steps leading to POLR2A degradation become compromised.
Chapter 5 Materials and methods

In vitro PRMT5 methyltransferase assays

In vitro methylation assays (30μl) containing 100μM of biotinylated 13mer CTD peptides centered around R1810 or R1603, 0.1 – 2μM PRMT5/WDR77 complex, 100μM tritiated SAM (PerkinElmer, Cat # NET155V250UC), 20mM Tris-HCl pH8, 0.01% TritonX-100 (Sigma Cat # T8532), and 5mM DTT were incubated at room temperature for 2 hours. The biotinylated peptides were then precipitated with 20μl Streptavidin-agarose (Invitrogen Cat # SA100), washed with buffer (above), mixed with 5ml ScintiVerse BD Cocktail (Fisher Chemical, Cat # SX18-4), and counted using a Beckman Liquid Scintillation Counter (LS 6500). Other similar reactions contained recombinant GST proteins (GST alone, GST-CTD-N, or GST-CTD-C), 0.1-2μM PRMT5/MEP50 and 100μM tritiated SAM. These reactions were precipitated with glutathione beads (Invitrogen Cat #G2879), and the beads were washed and eluted with 500μl 20 mM L-glutathione (pH 8) for counting.

Peptides, GST recombinant proteins, recombinant Tudor domains, fluorescence polarization and isothermal titration calorimetry assays

FITC- and biotin-labeled CTD peptides containing R1603 (SPAYEPRSPGGYT) and R1810 (YSPSSPRYTPQSP) were prepared on a Prelude peptide synthesizer (Protein Technologies, Inc., Tucson AZ) using Fmoc (9-fluorenyl methoxycarbonyl) solid-phase chemistry. Dimethyl Arginine derivatives were prepared using Fmoc-SDMA(Boc)2-ONa or Fmoc-ADMA (Pbf)-OH reagents (Nova biochem, Germany). Peptides were purified using C18 reverse-phase HPLC and authenticated using mass spectrometry.
Constructs for GST recombinant protein expression (GST-CTD-N: contains repeats 1-29; or GST-CTD-C: contains repeats 24-52) were expressed in BL21 bacteria and purified following the standard glutathione bead purification protocol. Bacterial expression constructs and purified Tudor domains from TDRD3, SMN, SPF30/SMNDC1, TDRD1, TDRD2, TDRD9, and TDRD11/SND1 were described previously [341].

Fluorescence polarization assays were carried out as described before [341]. The buffer used in the fluorescence polarization assay was 20 mM Tris pH 7.5, 50 mM NaCl, 1 mM DTT and 0.01% Triton X-100. An excitation wavelength of 485 nm and an emission wavelength of 528 nm were used. The data were obtained at 25°C and corrected by subtracting the label-free peptide background. The data were collected by the Synergy 2 (BioTec, USA) fluorescence polarization program and were fitted to a one-site binding model using Origin 7 (MicroCal, Inc). The Kd values are from the average of three measurements.

For isothermal titration calorimetry, the concentrated protein was diluted in 20 mM Tris, pH 7.5, 150 mM NaCl. The lyophilized peptides were dissolved in the same buffer and pH was adjusted by adding NaOH. Peptide concentrations were estimated from the molecular weight. All the measurements were performed at 25 °C, using a VP-ITC microcalorimeter. Protein with a concentration of 50 uM was placed in the cell chamber, and the peptides with a concentration of 1 mM in syringe were injected in 25 successive injections with a spacing of 180s and a reference power of 13mcal/s. Data were fitted using the single-site binding model within the Origin software package (MicroCal, Inc.).
Cell Culture, shRNA knock-downs, siRNA knock-downs, and CRISPR knockouts

Raji cells were cultured in RPMI (SLRI media facility) plus 10% Tetracycline free FBS (Gibco) and 1% Glutamate, and stably transduced cells were maintained with 500μg/ml G418 (Gibco Cat #11811031). Stably transduced cell lines with shRNAs [563] were maintained with 2μg/ml puromycin. HEK293 cells were grown in DMEM (SLRI media facility) plus FBS (Sigma Cat # F1051), and stably MAPLE tagged or shRNA knocked down transduced cell lines were maintained with 2μg/ml puromycin (Sigma Cat # p8833) [563]. Flp-In™ 293 T-REx cell lines were obtained from Life Technologies (Cat # R780-07), and stable Flp-In™ 293 T-REx cell lines after the incorporation of the transgene were maintained with hygromycin (Life Technologies Cat # 10687010) at 2μl/ml. For the induction of a GFP- or BirA-FLAG-tagged transgene in Flp-In™ 293 T-REx cells, doxycycline (Sigma-Aldrich: cat # D9891) was used at 2 μg/ml for 1 day. For BioID experiments, the cells were also treated with 50 μM biotin for 1 day.

shRNAs in lentivirus vectors were used to stably transduce cell lines using an established protocol [563]. siRNA knock-downs for HEK293 cells were performed with 50nM siRNAs with PepMute siRNA transfection reagent (SigmaGen Laboratory Cat #SL100566) for 3 days.

SMARTpool On-Target plus siRNAs against human PRMT5 (Cat # L-015817) and SMN (Cat # L-011108) and scrambled control were purchased from Thermo Scientific. siRNAs against human FUS (Cat # NM_001010850) and TDP-43 (Cat # NM_007375) were purchased from Sigma.

For CRISPR-mediated gene knock-outs, CRISPR/Cas9 plasmids (pCMV-Cas9-GFP) were purchased from Sigma-Aldrich to express the scrambled guide RNA, or guide RNAs for SMN, TDP-43, or FUS. 2 μg of the plasmids were transfected into HEK293 cells, and 1 day after transfection, cells were sorted by BD FACS AriaTM flow cytometry (Donnelly Centre,
University of Toronto) and single GFP-positive cells were plated into a 48-well plate. The expression levels of SMN in each clone were detected by western blotting.

**Cell transfection and electroporation**

Raji cells with stable expression of HA-tagged wild type or R1810A POLR2A constructs were generated by electroporation (10μg of plasmid DNA/10^7 cells) followed by selection (for 10 days) and maintenance with G418 (0.5mg/ml). α-amanitin treatment was carried out with 2μg/ml α-amanitin for 3 days for Co-IP and ChIP experiments involving HA-tagged wild type or R1810A POLR2A.

The transfections of the CRISPR/Cas9 plasmids (pCMV-Cas9-GFP), GFP-tagged (pDEST pcDNA5/FRT/TO-eGFP), BirA-FLAG-tagged (BirA-FLAG pcDNA5/FRT/TO), and the GFP-HB R-loop detector transgene into the Flp-In™ 293 T-REx cell lines were performed with FuGENE Transfection Reagent (Roche, Cat # E269A).

**Disease cell lines**

SMA disease relevant and control fibroblast and B-lymphocyte cell lines were obtained from the Coriell Institute (Family 553: GM03813, GM03814, GM03815; Family 3042: GM23686, GM23687, GM23688), and were grown in conditions as instructed by the Coriell Institute. The cells were collected and fixed for RNAPII ChIP and R loop DIP.

**Immuno-staining**

For staining, cells were fixed with 4% paraformaldehyde (Sigma Cat # P6148) and washed with PBS + 0.1% Triton X 100 (Sigma Cat # T8532). Primary antibodies 1:50-1:100 in
PBST + 30mg/ml BSA (Roche Cat # 10735108001) were added to cells for staining overnight at 4 °C. Cells were washed with PBST and stained with Alexa Fluor 488 or 594 (Invitrogen) at 1:1000 and Hoechst 33342 (Thermo Scientific) in PBST with 5% goat serum (Sigma Cat # G9023) at room temperature for 1 hour, followed by washes with PBS + 0.2% Triton and mounting with ImmunoMount (Cat # GTX30928). Images were taken using an Olympus Upright Microscope BX61 with Optigrid function and processed using Volocity OpenLab Software (PerkinElmer).

Immunoprecipitation (IP) and western blots

IP was performed with RIPA buffer (140mM NaCl, 10mM Tris pH7.6-8, 1% Triton, 0.1% Na deoxycholate, 1mM EDTA) containing protease inhibitors (Roche Cat # 05892791001) and Benzonase (Sigma E1014). 1-2x10^7 cells were lysed on ice for 25 minutes by vortexing and forcing them through a 27 gauge needle. After centrifuging at 13,000rpm for 10 minutes at 4 °C, the supernatant was incubated with 25μl (1:10 dilution) protein G beads (Invitrogen Cat # 10-1243, Cat # 10003D) and 1-2μg of antibodies for 4 hours to overnight. The samples were washed 3 times with RIPA buffer and boiled in SDS gel sample buffer. To detect R1810me2s or R1810me2a modifications on POLR2A, alkaline phosphatase (Roche Cat # 10108138001) treatment (5μl) at 37 °C for 30 minutes was performed for POLR2A IP samples before boiling. Samples were run using 7.5-10% SDS-PAGE and transferred to PVDF membranes (Bio-Rad Cat # 162-0177) using a Trans-Blot® SD Semi-Dry Electrophoretic Transfer Cell (BioRad Cat # 170-3940). Primary antibodies were used at 1:250 to 1:1000 dilutions for incubation overnight, and horseradish peroxidase-conjugated goat anti-mouse, -rabbit, or –rat secondary antibodies were
used at 1:10,000 (Dako Cat # P0450). Blots were developed using SuperSignal West Pico or Femto (Thermo Scientific Cat # 34079, 34094). Blots were quantified using ImageJ software.

A Hoefer Slot blot system (Fisher Scientific Cat #11509543) was used to assay R1810me2s antibody specificities following the manufacturer’s protocol.

**Chromatin immunoprecipitation (ChIP) and DNA Immunoprecipitation (DIP)**

ChIP was performed using the EZ-ChIP™ A - Chromatin Immunoprecipitation Kit (Millipore Cat# 17-371) or similar homemade solutions according to the manufacturer’s instructions. Antibodies were used in the 1-2μg range, and IgG was used as a background control. DIP was performed according to El Hage et al with minor modifications [564]. DIP was performed following the ChIP protocol except that, after the nuclear lysis and sonication, genomic DNA was de-crosslinked in ChIP elution buffer containing 5M NaCl at 65°C overnight. DNA was purified with the Qiaex II kit (Qiagen Cat # 20021) for PCR product purification and eluted in water. DIP was carried out overnight with 25μl of Dynabeads® protein G beads (Invitrogen Cat #100-03D) and 1μg of antibody purified from the S9.6 hybridoma cell line [565] that recognizes RNA/DNA hybrids. Immunoprecipitated and input DNAs were used as templates for qPCR. DIP RNase-sensitivity analysis was carried out by adding 50 U of RNase H (Invitrogen Cat # 18021-014) in 10X RNase H buffer (75mM KCl, 50mM Tris pH8.3, 3mM MgCl2, 10mM DTT) with 4% glycerol and 20μg/ml BSA prior to immunoprecipitation. The RNase H treatment was performed for 2 hours at 37°C.

For comparing POLR2A ChIP and S9.6 DIP signals on the beta-actin gene, wild type or control knock-down signals were normalized to 1, and the R1810A mutant or knock-down samples were adjusted such that the ratio for the intron 3 (1671) position was set to 1. Similarly,
for the Gapdh gene, the ratio for the intron 5 (2436) position was set to 1. ChIP data for Senataxin and SMN were expressed as ratios to the ChIP data for POLR2A. Error bars represent biological replicates, except where indicated otherwise.

**Nuclear run-ons (NRO)**

NRO was performed according to Skoutri-Stathaki et al. with modifications[91, 530]. ~10^7 cells were incubated on ice in swelling buffer (10mM Tris-Cl pH7.5, 2mM MgCl2, 3mM CaCl2) for 5 minutes, and were pelleted. Pellets were resuspended in 1ml lysis buffer (swelling buffer containing 0.5% NP40, 10% glycerol, and 2U/ml RNaseOUT (Invitrogen Cat # 10777-019)) and pipetted for lysis, followed by centrifugation. The pellet was resuspended in 1ml freezing buffer (50mM Tris-Cl pH8.3, 40% glycerol, 5mM MgCl2, 0.1mM EDTA). Reactions contained 100μl resuspended nuclei, 100μl reaction buffer (40mM Tris pH7.9, 300mM KCl, 10mM MgCl2, 40% glycerol, 2mM DTT), 500μM rNTPs (ATP, CTP, GTP) (GE Cat # 27-2025-01), including 125μM UTP as a negative control or Br-UTP (Invitrogen Cat # B21551) for 30 minutes at 30°C. 3μl BrdU antibody (Sigma #B8434) was pre-conjugated to 30μl Dynabeads® protein G beads with 10μg tRNA (Invitrogen Cat # 115401) as block in 100-RSB buffer (10mM Tris pH7.4, 100mM NaCl, 2.5mM MgCl2, 0.4% Triton X-100) for 2 hours at 4°C. The RNA was extracted using Trizol (Invitrogen Cat #15596-026) and was heat fragmented at 95°C for 8 minutes. RNA was then mixed with beads and BrdU antibody for 2 hours at 4°C in 500μl 100-RSB buffer, 100U/ml RNase OUT, 400U/ml DNase I (Invitrogen Cat #18047-019). Immunoprecipitated RNA was washed three times with 100-RSB buffer.
Primer information

Primers used in ChIP, DIP, and NRO are listed here [91, 165].

For the β-actin gene:

-72.fw CCGAAAGTTGCCTTTTATGCC, -72.rev CAAAGGCGAGGCTCTGTGTC;
332.fw CGGGGTCTTTTGTCTGAGC, 332.rev CAGTTAGCGCCCAGGAGAC;
1671.fw TAACACTGCTGCCTGTGACAA, 1671.rev AAGTGCAAAGAAACAGGCTAA;
2661.fw GGAGCTGCCACATCCAGGGTC, 2661.rev TGCCACTGCGCTGTGACCAA;
2911.fw TGCAGAGAAAACAAGATGAG, 2911.rev GTACCTTTACCCGTTCCAGT;
3560.fw TTACCACAGTGCAGGTGTG, 3560.rev CCCCATAAGCAGGACAGAG;
3752.fw GGGACTATTGGGGGTTGTCT, 3752.rev TCCCATAGGTCAGGCAAG;
4657.fw TGGGCTCTTAATCACTTCAAC, 4657.rev CTCACTTTCCAGACTGACAGC;
5590.fw CAGTGGTGCTGGTGATCTTGTG, 5590.rev GGCAAAACCCTGTATCTGTG.

For the Gapdh gene:

55.fw CTCCTGTTCGACAGTCAGC, 55.rev TTCAGGCCGTCCCTAGC;
1407.fw CACCCCTGCTCTGAGGTAAATAG, 1407.rev GTGGGAGCACAGGTAAGT;
2436.fw ATAGGGGAGATCCCTCCAA, 2436.rev TGAAGACGCCAGTGGAC;
3882.fw CCCTGTGCTCAACCAGT, 3882.rev CTCACCTTGACACCAAGCC;
4511.fw AGATGTCAGGGTGACTTAT, 4511.rev TAGGTCCTCAGCTACAGCG;
5196.fw GTTCAGGTGTATGACAGACACG, 5196.rev TGTATGTCGCTCACAGG.

Chromatin immunoprecipitation and Sequencing analysis (ChIPseq)

Chromatin immunoprecipitation was performed as before [566]. In brief, $10^7$-$10^8$ cells were cross-linked for 10 min in 1% formaldehyde. Lysates were sonicated to a DNA fragment
length range of 200-300 bp using a Bioruptor (Diagenode). RNAPII was immunoprecipitated with 2μg of antibodies and Dynabeads Protein G (Invitrogen). Subsequently, crosslinks were reversed at 65°C overnight and bound DNA fragments were purified (EZ-10 Spin Column PCR Product Purification kit, Bio Basic). Sequencing libraries were constructed using the TruSeq ChIP Sample Prep Kit (Illumina) according the manufacturer's instructions. Libraries were sequenced (single end reads) on the Illumina HiSeq 2500 to a minimum depth of 30 million reads, obtaining at least 10 million unique reads per sample.

ChIPseq analysis for the display of pileups was performed chiefly as before [534]. For ChIPseq, reads in FASTQ format were mapped to the human genome (hg19) using Bowtie 2 [567] with local alignment, duplicate reads were removed, and reads were extended to 300 bp. Pileups – the number of fragments overlapping each genomic bp – were calculated, and were normalized by million mappable reads in the ChIP-seq library. Normalized pileups from different replicates were then averaged to create FPM (fragments per million reads). Reads from the input samples are not subtracted from the RNAPII ChIPseq pileup displays as it creates negative values in the FPM scale (although the difference between the wild type and mutant is preserved in the termination region).

ChIPseq analysis for the generation of quantile plots and the calculation of stalling and stopping ratios was as follows: single-end sequence reads (51bp) were aligned to human reference genome hg19 using bowtie2 [567] with default parameters. Aligned reads were further processed to remove low quality ones as follows: duplicated reads were marked with Picard tools, and reads that are unmappable, duplicates or mapped to multiple genomic loci were removed using Samtools [568]. Only uniquely mapped reads were kept for further analysis. Signal density was calculated using the SPP [569] R package (bandwidth 250bp, step 100bp).
Background from input samples was first scaled based on sequencing depth (total number of unique reads) and then subtracted from sample signal. Signal density in different experimental groups was quantile normalized (all values from each treatment were first ranked separately, then values with the same rank in all groups were replaced by the average of corresponding values) to allow direct comparison between mutants and wild-type controls.

The coordinates of protein coding genes were obtained from the human reference genome hg19. Promoter region was defined as -30bp to +300bp around the transcription start site (TSS). The rest (from +301 to the end of annotated gene (EAG)) was defined as the gene body. Termination region was defined as the 2000bp segment downstream from the EAG site. Signals in each of these regions were summed up, divided by the length of the region and multiplied by 1000 to get a normalized value. Stalling ratio was defined as the ratio of promoter region signal to gene body signal, and stopping ratio was defined as the ratio of termination region signal to gene body signal. Only genes with signal in gene body greater than 1 were included in the analysis. Quantiles of promoter signals, body signals and termination signals or stalling ratios and stopping ratios were computed for mutants and wild-type controls separately and plotted against increasing level of signals or ratios. Statistically significant differences between the quantiles were determined using the Kolmogorov Smirnov test. Level of significance was set at p < 0.05. Only the top 20% of genes (at least 1000 genes) by mRNA expression levels were considered. Signal density data were processed with Perl scripts developed in house. Statistical analysis and data plotting were performed using R.
**RNA extraction, qPCR, and Illumina sequencing**

RNA was used for cDNA synthesis with the SuperScript® VILO™ Kit (Invitrogen Cat # 11754). PCR was performed using the Phusion-high fidelity PCR kit (Thermo Scientific Cat # F-553S), and qPCR was performed with Fast SYBR Green Master qPCR mix, using the Applied Biosystems 7300 Real Time PCR System (Cat # 4406984). qPCR consisted of 40 cycles of 95°C for 15 s and 55 °C for 30 s, and a final cycle (95 °C for 15 s and then 60 °C) generated a dissociation curve. Input DNA or RNA reverse transcribed into cDNA were used to calculate the percent enrichment in the immunoprecipitated samples. RNA sequencing was performed according to Illumina standard RNA-seq protocol with RNA collected after 3 days of α-amanitin treatment.

**BioID and AP-MS**

BioID was performed as described before [545] and Mass Spectrometry analysis and protein/peptide identification were done as before. Briefly, cell pellets from four 150-mm plates were incubated and agitated at 4°C in RIPA buffer [50 mM tris-HCl (pH 7.5), 150 mM NaCl, 1% NP-40, 1 mM EDTA, 1 mM EGTA, 0.1% SDS, with protease inhibitors and benzonase] for 1 hour. The lysates were sonicated with 10 pulses (High) in a Bioruptor. Lysates were centrifuged and supernatants were mixed with a 30-μl bed volume of streptavidin-magnetic beads (Life Technologies, Cat# 11205D) for 4 hours. The beads were then washed twice in RIPA buffer, twice in TAP lysis buffer [50 mM Heps-KOH (pH 8.0), 100 mM KCl, 10% glycerol, 2 mM EDTA, 0.1% NP-40], and three times in 50 mM ammonium bicarbonate (ABC; pH 8.0). Beads were resuspended in 50 μl of ABC with 1 μg of trypsin (Promega, Cat # V5111) and incubated at 37°C overnight with agitation. Beads were pelleted, and the supernatant was
transferred and dried in a vacuum centrifuge. Tryptic peptides were resuspended in 10 μl of 5% formic acid for MS analysis.

Affinity purifications for GFP-taged proteins were done as before with slight modification [570]. Mass Spectrometry analysis and protein/peptide identification were done as before [571]. Briefly, all peptides were reconstituted in 1% Formic acid, then desalted in a C18 2.5 cm silica capillary trap column and separated for 90 minutes in a 10 cm silica capillary analytical column packed with 3 μm Luna C18 stationary phase (Phenomenex) using the EASY-nLC system (Proxeon). Eluted peptides were directly sprayed into an Orbitrap Velos mass spectrometer (Thermo Fisher Scientific) with collision induced dissociation (CID). 16 ms/ms CID data-dependent scans were acquired simultaneously with one profile mode full scan mass spectra. The full scan was performed in 60000 resolutions with 1E6 AGC. An exclusion list was enabled to exclude a maximum of 500 ions for 22.5s. For protein and peptide identification, RAW files were extracted from the mass spectrometry data with the ReAdW program and submitted to database search using SEQUEST v2.7 and a UniProt/SwissProt protein database FASTA file containing 22,491 human proteins. Matched peptides were further filtered at the precursor ion spectra level using a 20 ppm cutoff. 99% confidence level was set up for the protein hits, as filtered using the StatQuest (1) program.

**Antibodies, constructs and reagents**

Anti-CTD R1810me2s antibody was raised in rabbits using a KLH-conjugated CTD peptide from POLR2A (amino acids 1806-1813) that carried an R1810me2s modification. KLH conjugation was performed using an N-terminal cysteine residue (Cedarlane). R1810me2s-
specific antibodies were enriched by flowing the serum through a column containing an R1810me0 peptide conjugated to SulfoLink Coupling Resin (Thermo Scientific Cat # 20401). GST fusion constructs containing CTD N-terminal repeats 1-29 and C-terminal repeats 33-52 were kindly provided by Dr. James Manley [572]. A Flp-in TREx GFP-HB fusion construct that contains the R-loop binding domain of RNase H was provided by Dr. Andres Aguilera [230]. The ORFs for POLR2D, TDRD3, PRMT5, SMN and TDP-43 came from the Plasmid collection at Harvard. CMV promoter-driven Flag-tagged POLR2D, TDRD3, GFP, and PRMT5 constructs for HEK293 cell culture were generated using the MAPLE system as previously described [563]. Doxycycline inducible constructs with the BirA-FLAG tag were kindly provided by Dr. Anne-Claude Gingras. BirA-FLAG and GFP constructs were generated via Gateway cloning into pDEST5 BirA-FLAG pcDNA5/FRT/TO and pDEST pcDNA5/FRT/TO-eGFP, respectively. α-amanitin resistant wild type and R1810A mutant POLR2A constructs were kindly provided by Dr. Dirk Eick [67]. RNAPII R1810me2a antibody was kindly provided by Dr. Danny Reinberg [67]. We obtained the POLR2A pSer2 and pSer5 antibodies from the Eick laboratory (S2P: 3E10; S5P: 3E8). We obtained the S9.6 antibody for R-loop detection from Dr. Stephen Leppla. 8WG16 antibody against unphosphorylated CTD repeats of POLR2A was prepared in the lab.

Commercial antibodies were as follows: Y12 (Abcam mAb Cat #ab3138); Symm10 (Millipore pAb Cat # 07-412); Asymm24 (Millipore pAb Cat # 07-414); TDRD3 (Santa Cruz pAb Cat # C-20); HA (Sigma mAb Cat # H9658); PRMT5 (Upstate pAb Cat # C7-405, Santa Cruz mAb Cat # sc-22132); SMN (Santa Cruz pAb Cat # H-195); SETX (for ChIP and IP (Novus Biologicals pAb Cat #NB100-57543) and for western blots (Bethyl Lab pAb Cat # A301-104A)); XRN2 (Santa Cruz pAb Cat # sc-99237); FUS (Santa Cruz mAb Cat # sc-47711);
TDP-43 (Bethyl lab pAb Cat # A303-233A); POLR2A N20 (Santa Cruz pAb Cat # sc-899);
POLR2A 4H8 (Abcam mAb Cat # ab5408); POLR2A H224 (Santa Cruz pAb Cat # sc9001);
gammaH2Ax (Millipore, Cat # 05-636); H2Ax (Millipore, Cat # 07-627); Tubulin (Sigma mAb
Cat # T8328); FLAG (Sigma mAb Cat # F1804); GFP (Abcam pAb Cat# 290); IgG negative
controls for ChIP and IP (Millipore pAb Cat # 12-370). α-amanitin was purchased from Sigma
(Cat #23109-05-9).

References


237


