An investigation of the cellular functions of HNRNPUL1

By:

Llywelyn Griffith

A thesis submitted in partial fulfilment of the requirements for the degree
of Doctor of Philosophy

Department of Molecular Biology and Biotechnology

University of Sheffield

Sheffield, United Kingdom

September 2019

ACKNOWLEDGEMENTS

ABSTRACT

HNRNPUL1 displays many properties in common with proteins associated with the neurodegenerative disorder amyotrophic lateral sclerosis (ALS), such as RNA-binding capacity and a prion-like domain at its C-terminus. However, there is a limited understanding of its cellular functions, especially regarding RNA metabolism. Currently, ALS is an incurable condition and therefore there is a pressing need to further elucidate the molecular pathology underpinning this disease. During this study, we have identified HNRNPUL1 as a key component of the small RNA biogenesis pathway. Conditional depletion of HNRNPUL1 via the auxin-inducible degron system results in the clear downregulation of both snRNA and snoRNA expression levels. In addition, loss of HNRNPUL1 triggers the disintegration of Cajal bodies and loss of SMN-containing nuclear Gems – a hallmark of both ALS and the closely related neurodegenerative condition spinal muscular atrophy (SMA). We have also demonstrated that the prion-like domain of HNRNPUL1 is essential for forming the majority of its protein interactions, including with other ALS-causing factors such as FUS and TAF15, as well as RNA polymerase II. In addition, ALS patients with mutations in HNRNPUL1 have been identified. One of these patients, possessing a heterozygous S249N point mutation, displays reduced expression levels of HNRNPUL1 and an snoRNA biogenesis defect.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

AID – auxin-inducible degron

ALS - amyotrophic lateral sclerosis

ATP – adenosine triphosphate

CB – Cajal body

CBC – cap-binding complex

cDNA – complementary DNA

ChIP – chromatin immunoprecipitation

CoIP – co-immunoprecipitate

CTD – C-terminal domain

DNA – deoxyribonucleic acid

DSB – double-stranded break

DSE – distal sequence element

eCLIP – enhanced cross-linking and immunoprecipitation

GTP – guanosine triphosphate

HNRNP – heterogeneous nuclear ribonucleoprotein

HNRNPUL1-WA – HNRNPUL1 Walker A mutant

IEG – immediate-early gene

IP – immunoprecipitation

LB – Luria broth

LEC – little elongation complex

lncRNA – long non-coding RNA

miRNA – microRNA

mRNA – messenger RNA

NEXT complex – nuclear exosome targeting complex

NLS – nuclear localisation signal

NTP - nucleoside triphosphate

PCR – polymerase chain reaction

PrLD – prion-like domain

PSE – proximal sequence element

PTF - PSE-binding transcription factor

qRT-PCR - quantitative reverse transcription PCR

RBP – RNA-binding protein

RIP – RNA-immunoprecipitation

RNA – ribonucleic acid

RNAi – RNA-interference

RNAPII – RNA polymerase II

RRM – RNA-recognition motif

rRNA – ribosomal RNA

SB – Sleeping Beauty

scaRNA – small Cajal body-specific RNA

SD – standard deviation

SEC – super elongation complex

Ser2P – phosphorylated Serine-2

Ser5P – phosphorylated Serine-5

Ser7P – phosphorylated Serine-7

siRNA – small interfering RNA

SMA – spinal muscular atrophy

snoRNA – small nucleolar RNA

snRNA – small nuclear RNA

snRNP – small nuclear ribonucleoprotein

TB – terrific broth

TMG - trimethylguanosine

TSS – transcription start site

TTS – transcription termination site

CHAPTER 1 - INTRODUCTION

1.1 THE RNAPII CTD CODE

RNA polymerase II (RNAPII) is responsible for the transcription of all protein-encoding genes, as well as many long and short non-coding RNAs. The largest subunit of this complex - RBP1 - possesses a unique C-terminal Domain (CTD) that plays an essential role in the activity of the polymerase. In mammals, this CTD is composed of 21 copies of a consensus motif Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7, as well 31 non-consensus repeats (Eick and Geyer, 2013). Among the consensus repeats, residues Tyr1, Ser2, Thr4, Ser5 and Ser7 can be phosphorylated or glycosylated, while Pro3 and Pro6 can be isomerized (Zaborowska et al., 2016).

The process of transcribing a messenger RNA (mRNA) can be generally divided into three phases – initiation, elongation and termination. Phosphorylation and glycosylation of the CTD is critical in facilitating transition through these phases, as well as enabling sequential recruitment of RNA-binding factors. This allows coupling of transcription with processing of nascent mRNA transcripts. Accordingly, proteins required for the early steps of RNA-processing (e.g. mRNA capping enzymes) recognize CTD modifications that are most prevalent at the 5' end of genes (Fabrega et al., 2003), while factors involved in later stages interact with the CTD modifications found most abundantly at the 3' end. This is demonstrated by the case of polyadenylation factor Pcf11, that has specific affinity for phosphorylated Serine-2 (Ser2P) (Gu et al., 2013).

RNAPII is positioned on to the promoters of genes in a largely unphosphorylated state as part of a pre-initiation complex, along with six general transcription factors – TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH – and Mediator (Soutourina, 2018). For RNAPII to escape this complex and initiate transcription, it must be phosphorylated at Serine-5 by the CDK7 subunit of TFIIH (Søgaard and Svejstrup, 2007). CDK7 also has the ability to phosphorylate Serine-7 (Glover-Cutter et al., 2009). RNAPII then pauses approximately 30-50 nucleotides downstream of the transcription start site

(Zlotorynski, 2017), halted by DRB-sensitivity inducing factor (DSIF) and negative elongation factor (NELF) (Kwak and Lis, 2013). Once again, transition into the next phase of transcription – productive elongation – is facilitated by CTD phosphorylation. In this case, this is performed by the CDK9 subunit of P-TEFb, which targets Serine-2 of the CTD, as well as DSIF and NELF themselves (Peterlin and Price, 2006). NELF dissociates from the complex as a result, while DSIF is converted into a positive elongation factor (Figure 1.1).

Many genes that require rapid activation in response to external stimuli such as cellular infection, or internal stimuli such as DNA damage, are primarily regulated at the pause-release stage. One such class of genes is known as the immediate-early genes (IEGs), including EGR1, FOS and JUN. At many of these loci, RNAPII sits primed at the promoter-proximal site and therefore can be quickly activated into productive elongation, enabling synthesis of the mRNA in a matter of minutes (Bahrami and Drabløs, 2016). These IEGs encode transcription factors that can then trigger signalling cascades to allow the cell to rapidly respond to these external/internal stimuli. A key regulator of the pause-release transition is the 7SK snRNP complex (McNamara et al., 2016). This complex binds to and inhibits the activity of P-TEFb in a dynamic association dependent on the transcriptional needs of the cell. For example, in response to DNA-damaging agents, RNA-binding factor RBM7 binds 7SK stimulating release of P-TEFb from the inhibitory complex to trigger pause-release at genes involved in the DNA-damage response such as EGR1 and FOS (Bugai et al., 2019) (Figure 1.2).

Both Chromatin-Immunoprecipitation (ChIP) and mammalian Native Elongating Transcript sequencing (mNET-seq) have been utilized to analyse the profile of the various CTD modifications across the body of genes. Ser2P and Ser5P account for approximately 75% of all phospho-counts on the CTD in human cells (Schüller et al., 2016), and it is these two modifications that have been studied in most detail. ChIP studies have indicated that Tyr1P, Ser5P and Ser7P are most prominent at the 5'
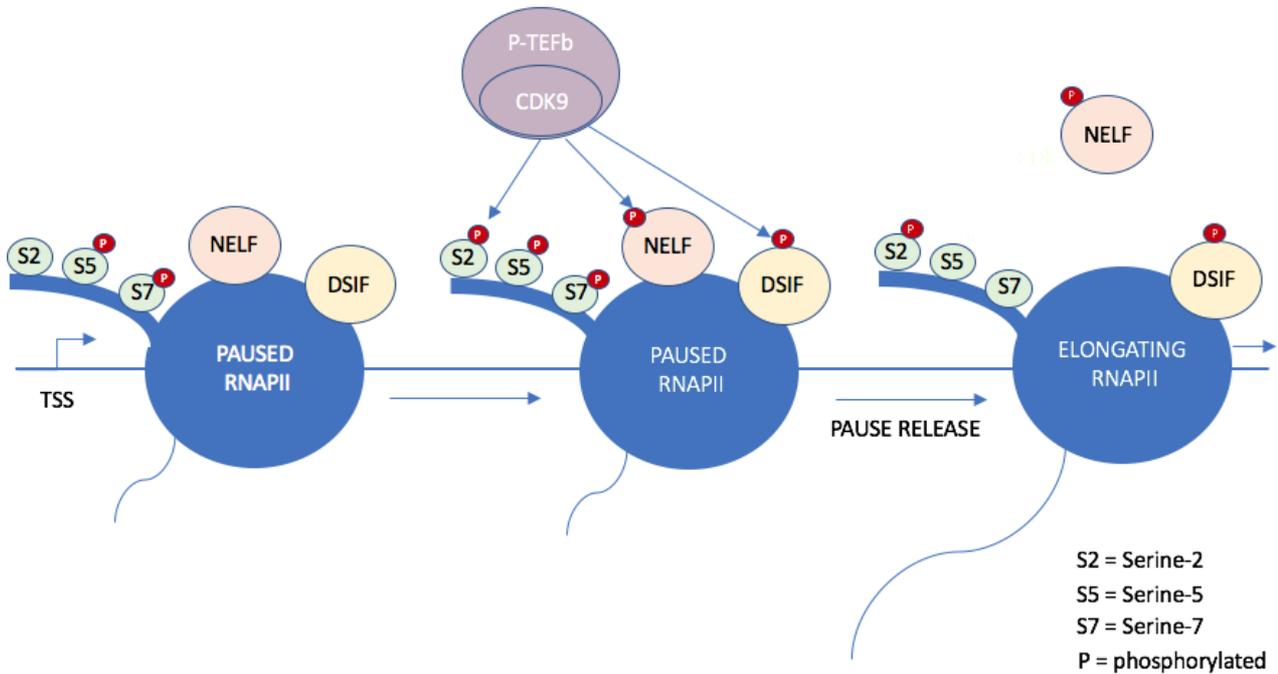
Figure 1.1 RNAPII pause-release is stimulated via CDK9-mediated phosphorylation events

RNAPII is held in a paused configuration by DSIF and NELF following the initial transcription of 20-60 nucleotides. The CDK9 subunit of P-TEFb phosphorylates Ser2 on the RNAPII CTD, as well as DSIF and NELF to facilitate pause-release and productive elongation.
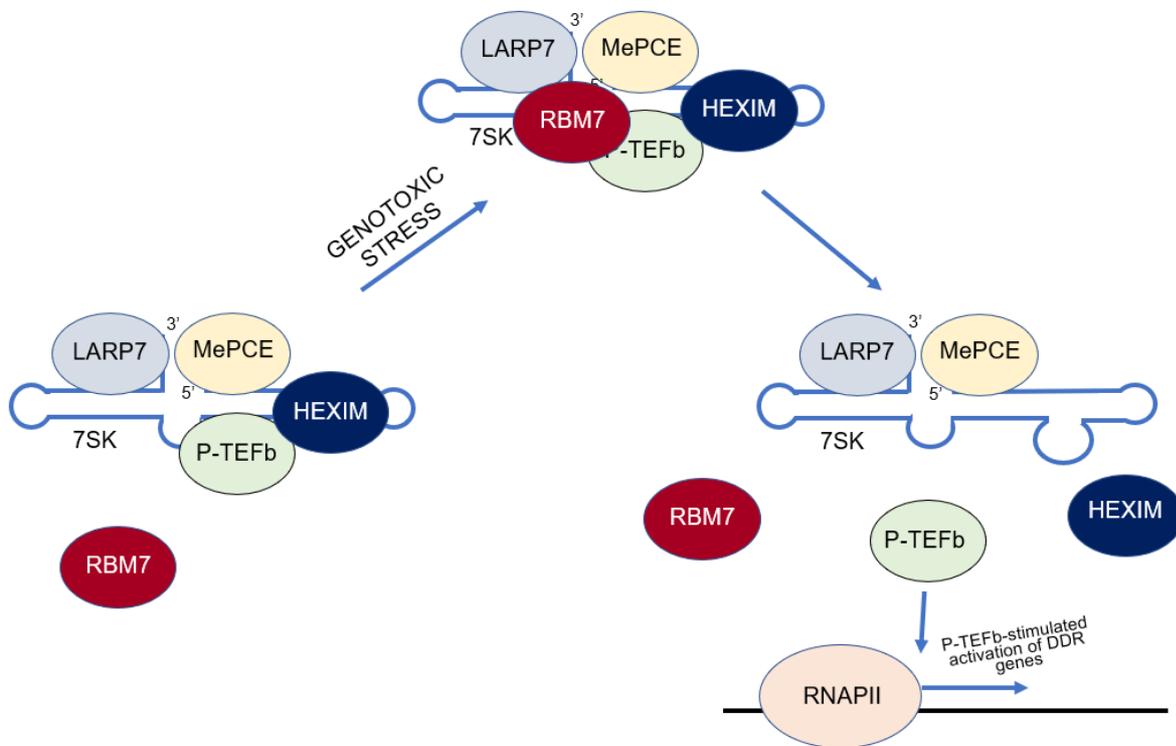
Figure 1.2 RBM7 releases P-TEFb from the inhibitory 7SK complex in response to genotoxic stress, in order to facilitate P-TEFb-dependent transcription of DNA-damage response genes

P-TEFb is prevented from stimulating transcription elongation via HEXIM, the inhibitory component of the 7SK snRNP complex. In response to genotoxic stress, RBM7 binds to the 7SK complex and releases P-TEFb from this inhibition. P-TEFb can then phosphorylate RNAPII to activate transcription of genes required for the DNA-damage response.

end of genes, while Ser2P and Thr4P appear more abundantly at the 3' end (Dias et al., 2015; Voss et al., 2015). mNET-seq analyses have found a similar Ser2P pattern, but diverge with ChIP analyses with respect to Ser5P. In their original mNET-seq paper, Nojima et al. (2015) report Ser5P as present across exons along the entire length of genes, with a particularly prominent peak at the end of exons thought to represent a splicing intermediate (Nojima et al., 2015). There are a couple of possible explanations for this discrepancy. Firstly, as mNET-seq is based on the extraction of RNA from the active site of RNAPII (RNA is then sequenced via linker ligation on the 3' end of RNAs), it could be the case that the hyper-phosphorylated Ser5 found at transcription start sites (TSS) has not yet synthesized the 35 nucleotides of RNA required to be recognized by mNET-seq. An alternative possibility is that as ChIP involves cross-linking via formaldehyde, this could potentially disrupt native chromatin conformation leading to misleading CTD isoform profiles.

It has been postulated that the RNAPII pausing at splice sites observed in these mNET-seq assays could ensure a window of opportunity for the proper assembly of the splicing machinery and subsequent splicing steps to occur, in light of previous studies demonstrating RNAPII elongation rates and splicing decisions to be intrinsically linked (Saldi et al., 2016). In 2018, Nojima et al. (2018) followed up their initial mNET-seq study by analysing via mass-spectrometry the protein components that co-immunoprecipitate with different RNAPII CTD isoforms (Nojima et al., 2018). Consistent with their Ser5P mNET-seq profile, they revealed that components of the spliceosome specifically co-immunoprecipitated with the Ser5P form of RNAPII (Nojima et al., 2018). Intriguingly however, functional knockdown of spliceosomal RNAs via antisense oligo causes a decrease in Ser2P at a global level, but leaves Ser5P unaffected (Koga et al., 2015). The exception to this finding was U1 snRNA inhibition, which triggered elevated levels of Ser5P (Koga et al., 2015).

In addition to CDK7 and CDK9, there are a growing number of kinases implicated in CTD phosphorylation. For example, CDK12 and CDK13 display the ability to

phosphorylate both Ser5 and Ser2 (Bartkowiak et al., 2010; Greifenberg et al., 2016). However, the biological significance of the relationships between these non-canonical kinases and their targets requires further in vivo investigation, with studies often hampered by the functional redundancy displayed among the many kinases involved.

It is becoming increasingly clear that RNAPII elongation rate itself can also affect CTD phosphorylation. By generating RNAPII mutants with reduced elongating speeds, Fong et al. (2017) demonstrated that an increased RNAPII dwell time at TSSs is correlated with elevated Ser2P levels at the 5' end of genes (Fong et al., 2017). Given the wide range of factors and complex structures that govern RNAPII pausing and/or elongation rates, such as nucleosomes (Jimeno-González et al., 2015) or R loops (Skourti-Stathaki and Proudfoot, 2014), this finding greatly expands the potential network of proteins that can directly or indirectly influence CTD phosphorylation.

The removal of CTD modifications is also a key feature of transcription regulation. For example the depletion of Ssu72, which dephosphorylates Ser5 and Ser7 (Krishnamurthy et al., 2004), not only results in elevated Ser5P and Ser7P levels at the 3' end of genes, but also produces a transcription termination defect (Zhang et al., 2012). Another Ser5 phosphatase RPAP2 has also been shown to direct pre-mRNA 3'-end formation (Wani et al., 2014), while also playing a particularly prominent role in snRNA transcription (Egloff et al., 2012).

1.2 snRNP BIOGENESIS

1.2.1 snRNA Transcription

The majority of protein-encoding genes contain introns, which are removed co-transcriptionally from nascent pre-mRNA transcripts via splicing. There are two multi-subunit RNP complexes that perform these reactions – the major spliceosome, comprising U1, U2, U4, U5 and U6 snRNPs, and the minor spliceosome consisting of

U11, U12, U4atac, U5 and U6atac snRNPs. The major spliceosome recognises GT/AG splice sites that flank a high proportion of introns and is therefore responsible for most of the splicing that occurs in the cell. In contrast, the minor spliceosome recognises AT/AC splice sites found in a small fraction of transcripts. With the exception U6, U6atac and 7SK snRNAs, which are all transcribed by RNA polymerase III (White, 2011), snRNAs are transcribed by RNAPII.

RNAPII-transcribed snRNAs have a simpler gene structure compared to their protein-encoding counterparts, containing no introns (Figure 1.3). A distal sequence element is located around 250 nucleotides upstream of the TSS. This acts as an enhancer, recognised by transcription factors Oct1, Staf, NF1 and Sp1 (Jawdekar and Henry, 2008). This is followed by a conserved snRNA-gene specific motif known as the proximal sequence element (PSE), located approximately 50 nucleotides upstream of the TSS (Egloff et al., 2008). Oct1 promotes snRNA gene transcription by stabilising the interaction between a complex known as PSE-binding protein/PSE-binding transcription factor/snRNA activating protein complex, PTF (sometimes referred to PBP and SNAPc) and the PSE (Murphy et al., 2015). The snRNA TAF complex (snTAFc), comprising the TATA-binding protein (TBP) plus the TBP-associated factors (TAFs), is then recruited by PTF along with the general transcription factors.

CDK7 phosphorylates Ser5 and Ser7 on the RNAPII CTD shortly after the initiation of transcription, and unlike protein-encoding genes, it is the phosphorylation of Ser7 that has been identified as essential for the transcription of snRNAs (Egloff et al., 2007). This is due to its role in recruiting RPAP2, the Serine 5 phosphatase mentioned previously. In addition to dephosphorylating Ser5 at snRNA loci, RPAP2 also recruits several subunits of the 3' end snRNA processing complex known as Integrator (Egloff et al., 2012). snRNAs are non-polyadenylated, instead featuring a region located approximately 200 nucleotides downstream of the TSS, known as the 3' box. When this sequence is recognised by Integrator, it triggers this 14 subunit complex to
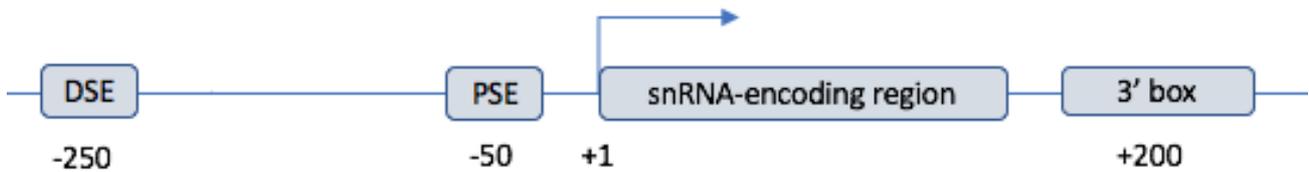
Figure 1.3 Structure of an snRNA gene

snRNA genes are composed of a distal sequence element (DSE), a proximal sequence element (PSE), the snRNA-encoding region, and a 3' box approximately 200 nucleotides downstream of the start of the coding region. The DSE acts as an enhancer, stabilising interactions between the transcription initiation machinery and the PSE promoter region. The 3' box governs the formation of pre-snRNA 3' ends. No introns are present.

endonucleolytically cleave the nascent pre-snRNA transcript (Baillat et al., 2005). Rather than promote pause-release as in the case of protein-coding genes, Ser2 phosphorylation via P-TEFb facilitates recruitment of the remaining Integrator subunits, Ints9 and Inst11, required for this reaction to occur (Egloff et al., 2010).

P-TEFb facilitates productive elongation during the transcription of mRNAs as part of a complex known as the super elongation complex (SEC), made up of ELL, AFF1, AFF4, AF9, and ENL (Smith et al., 2011a). In the case of snRNA gene transcription, elongation is promoted by another ELL-containing complex, referred to as the little elongation complex (LEC). This complex consists of proteins ELL, ICE1, ECE2, EAF and ZC3H8 (Smith et al., 2011b). Surprisingly, in 2017 Egloff et al. (2017) revealed that the 7SK snRNP complex is recruited to snRNA loci and is required for the integrity of the LEC, and as a result knockdown of 7SK components triggers reduced snRNA transcription (Egloff et al., 2017). It was also recently shown that Mediator, essential for transcription initiation of protein-encoding genes, is required for the recruitment of the LEC to snRNA loci (Takahashi et al., 2015).

DSIF and NELF are both also required for the proper transcription of snRNA genes, and interact with the Integrator complex at the 3' end of snRNA genes (Yamamoto et al., 2014). NELF inhibition was specifically shown to induce termination defects and 3'-end misprocessing (Yamamoto et al., 2014). Termination of snRNA transcription and 3'-end processing appear to be intrinsically linked, with knockdown of the catalytic subunits of Integrator – Ints9 and Inst11 – also causing disruptions to termination (O'Reilly et al., 2014).

1.2.2 snRNA Export and Cytoplasmic RNP Assembly

An m_7G cap is added to all RNAPII-transcribed RNAs soon after transcription is initiated (Cho et al., 1997). In the case of transcripts shorter than 250 nucleotides, such as snRNAs, the cap is bound by CBP80, CBP20, and ARS2, in a complex referred to as CBCA (Hallais et al., 2013). The CBCA is a critical component of the

snRNA biogenesis pathway, not only promoting efficient termination and 3'-end processing of snRNAs (Hallais et al., 2013), but also facilitating their nuclear export. It achieves this via an interaction with PHAX, forming the CBCAP complex that in turn recruits the export receptor chromosome region maintenance 1 (CRM1; sometimes referred to as exportin 1/XPO1). CRM1 binds nuclear pore proteins to promote export of pre-snRNAs into the cytoplasm in a Ran-GTP dependent manner (Fornerod et al., 1997). It is thought that pre-snRNAs traffic through nuclear structures known as Cajal bodies (CBs) prior to export (Matera and Wang, 2014). These structures will be discussed in greater detail in the 'Cajal bodies' section that follows.

Alternatively, the CBCA can interact with the nuclear-exosome-targeting complex (NEXT), comprised of MTR4, RBM7 and ZCCHC8, along with an additional cofactor known as ZC3H18 (Andersen et al., 2013). As the NEXT complex facilitates degradation of specific substrate transcripts by the nuclear exosome, the CBCA-NEXT interaction, referred to as the CBCN complex, is thought to promote RNA decay of these snRNA precursors. The competition between PHAX and ZC3H18 for CBCA binding (Figure 1.4) appears to dictate levels of snRNA nuclear transport versus degradation, although how this balance is regulated is still not completely understood (Giacometti et al., 2017).

snRNPs undertake an unusual assembly pathway, with several steps occurring in the cytoplasm prior to re-import into the nucleus, where the final maturation stages transpire. Following its initial export into the cytoplasm, dephosphorylation of PHAX triggers the disassembly of the pre-snRNA export complex (Kitao et al., 2008). This is followed by binding to the snRNAs by the survival motor neuron (SMN) protein complex, composed of SMN in tight association with a collection of associated proteins known as Gemins (Pellizzoni et al., 2002). In 2006, Gemin5 was identified as the SMN complex component that specifically recognises and binds snRNA transcripts in the cytoplasm (Battle et al., 2006). The SMN complex then orchestrates the assembly of a heptameric ring around the pre-snRNAs composed of 7 Sm-proteins, bound via a
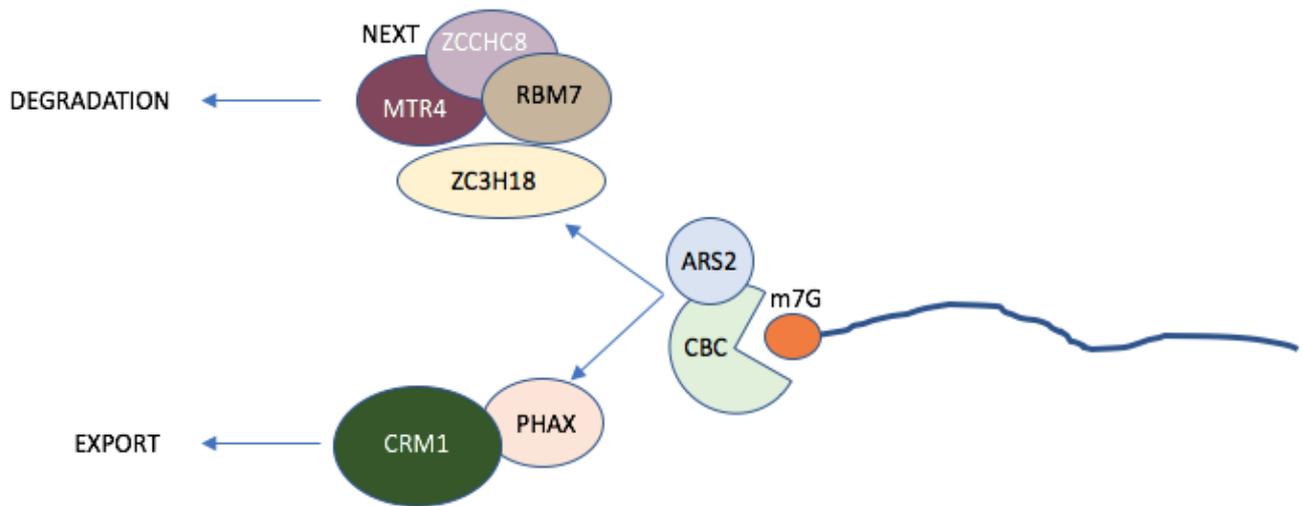
Figure 1.4 sn/snoRNA fate is specified by mutually exclusive CBC-ARS2-containing complexes

sn/snoRNA nuclear transport protein PHAX competes with ZC3H18, responsible for stimulating degradation of these RNAs, for binding to the CBC in association with ARS2.

consensus Sm-binding sequence AUUUUUG (Staněk, 2017). The Sm proteins initially form heterodimeric (SmD1–SmD2 and SmB–SmD3) or heterotrimeric (SmE–SmF–SmG) sub-complexes, and only rearrange into a heptamer upon association with the pre-snRNA (Raker et al., 1996). The protein Arg N-methyltransferase 5 (PRMT5) mediates the transfer of these sub-complexes to the SMN complex, methylating arginine residues present on SmB, SmD1 and SmD3 prior to their delivery (Meister et al., 2001). Gemin2 has also been identified as critical in promoting these reassembly reactions, directly binding five of the Sm proteins and stabilising sub-complex intermediates (Zhang et al., 2011). In addition, Sm core formation is also dependent on the ability of SMN to self-oligomerize, facilitated by the YG-box domain present at its C-terminus (Lorson et al., 1998). Once assembled, this Sm ring protects the pre-snRNA from degradation.

The SMN complex then recruits trimethylguanosine synthase 1 (TGS1) (Mouaikel et al., 2003), which catalyses the conversion of the 5'-end m_7G cap into a 2,2,7-trimethylguanosine (TMG) cap structure. Next, the final maturation step prior to re-import is 3'-end exonucleolytic trimming (Huang et al., 2015). Import is mediated by the snRNP-specific adaptor protein Snurportin-1. Snurportin-1 binds both the TMG cap and the import receptor importin-β, which then directs the import of the snRNP complexes in association with the SMN complex (Palacios, 1997). Once inside the nucleus, SMN dissociates from the snRNP, localising either to within Cajal bodies or to distinct subnuclear bodies known as Gems (Matera and Shpargel, 2006), depending on the cell type.

1.2.3 SMN and Spinal Muscular Atrophy

In 1995, it was discovered that mutations in the *SMN1* gene that result in reduced SMN protein expression cause a common neurodegenerative condition known as spinal muscular atrophy (SMA) (Lefebvre et al., 1995). SMA is characterized by the degeneration of the lower motor neurons as well as muscular atrophy, and is one of

the most common genetic causes of infant mortality (Arnold and Fischbeck, 2018). Although there is a second copy of the SMN gene – *SMN2* - a single nucleotide mutation in exon 7 causes this exon to regularly be skipped, resulting in low level SMN protein expression from this gene (Lorson et al., 1999). The reduction in SMN levels displayed in patient fibroblasts coincides with a loss of nuclear Gems, a hallmark of the condition. In addition, widespread pre-mRNA splicing defects are observed as well as a dysregulation of snRNA levels, consistent with the role of SMN in snRNP biogenesis (Zhang et al., 2008). Interestingly, it appears that minor spliceosome snRNAs are particularly affected by SMN loss (Gabanella et al., 2007), and accordingly a high proportion of U12-type introns are mis-spliced in SMA patient-derived cells as well as SMA mouse models (Doktor et al., 2017).

1.2.4 snRNP Nuclear Remodelling and Spliceosome Assembly

snRNP complexes undergo final maturation steps following re-import into the nucleus, with an additional 1 to 12 snRNP-specific proteins incorporated into each snRNP (Staněk, 2017). For example, the U2 snRNP sequentially incorporates proteins SNRNPA1 (U2A) and SNRNPB2 (U2B'), the SF3b complex, followed by the SF3a trimer complex (Krämer et al., 1999). This pathway involves the formation of three distinct complexes, known as 12S, 15S and 17S respectively. SPF30, DXH15, DDX46 and the U2AF dimer then associate with 17S U2 snRNP to form the mature particle (Will et al., 2002). These U2 snRNP-specific proteins are imported independently of the snRNA, as has been demonstrated for U1 snRNP-specific proteins (Romac et al., 2015).

Prior to incorporation into the spliceosome, U4 snRNP forms a di-snRNP, then a tri-snRNP particle with the U6 and U5 snRNPs respectively. The U6 snRNP undertakes a different biogenesis pathway to the RNAPII-transcribed snRNAs. It remains in the nucleus following its transcription, where it acquires a mono-methylated cap is at its 5'-end and a 2′,3′-cyclic phosphate moiety at its 3′ terminus (Licht et al., 2008). It is

then packaged together with a pre-assembled ring of 7 LSm 2–8 proteins (Zaric et al., 2005). These LSm proteins, along with a key recycling factor known as SART3, facilitate the formation of a U4:U6 snRNA duplex (Achsel et al., 1999; Bell et al., 2002). U4/U6 snRNP-specific proteins then assemble onto the duplex in a hierarchical manner, firstly with a factor known as 15.5K recognizing a specific K-turn on the U4 snRNA, followed by the recruitment of PRP31, then PRP3, PRP4 and CYPH (Liu et al., 2007). A U4/U6/U5 tri-snRNP structure is then formed via the interaction between PRP31 and U5-specific snRNP protein PRP6 (Liu et al., 2006), which can then engage in splicing as part of the spliceosome.

The first step in splicing is the recognition of the 5' splice site (SS) by U1 snRNP, which base-pairs with the pre-mRNA (Wahl et al., 2009). This is followed by recognition of the branch point adenosine by U2 snRNP and the formation of complex E, which is converted to complex A through interaction between U1 snRNP and U2 snRNP (Wahl et al., 2009). The pre-formed U4/U6/U5 tri-snRNP complex is then recruited to the splicing reaction, resulting in the formation of complex B. U4 and U6 snRNAs are extensively base-paired in the initial tri-snRNP conformation. This is a critical regulatory feature, as in effect it ensures that the U6 snRNA is delivered to the spliceosome in a catalytically inactive state (Mroczek and Dziembowski, 2013). U5 snRNP proteins are then crucial in facilitating structural re-arrangements that trigger catalytic activation of the complex. U5 snRNP RNA helicase Brr2 unwinds the U4/U6 helices, and U1 and U4 snRNPs are released from the spliceosome (Agafonov et al., 2016). Further structural rearrangements result in the base-pairing of U6 snRNA with U2 snRNA and the pre-mRNA substrate and the formation of a catalytically active complex - complex B'. This complex removes introns from pre-mRNAs via two transesterification reactions (Figure 1.5). Firstly, a nucleophilic attack on the 5' SS is performed by the 2'-OH group of the branch point adenosine, producing a free 5' exon and intron–exon 2 lariat intermediate. The first exon 3'-OH group then carries out a nucleophilic attack on the 3' SS, resulting in ligation of the two exons and intron removal. The remaining snRNPs – U2, U5 and U6 are then released and recycled
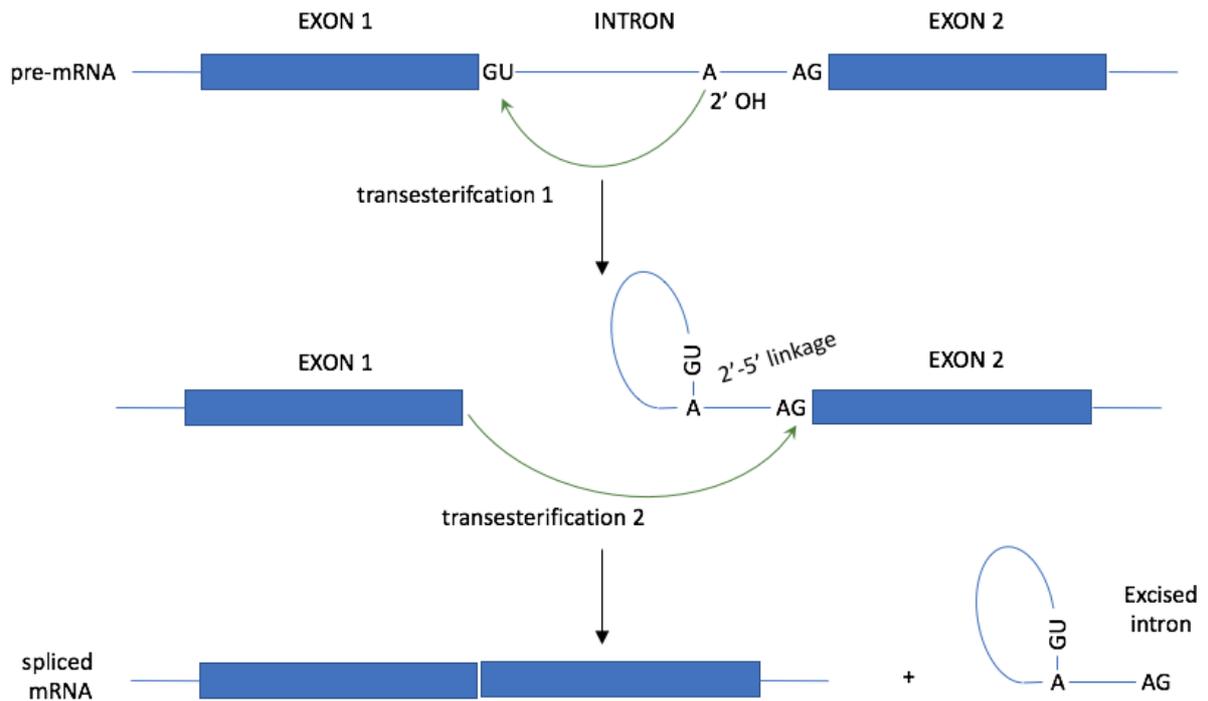
Figure 1.5 Splicing occurs via two transesterification reactions

The first step of the splicing reaction occurs via nucleophilic attack on the first nucleotide of the intron at the 5' splice site by the 2' OH of the branch point adenosine. A lariat intermediate is formed, and then released by a second transesterification involving the first intron nucleotide and the last nucleotide of the intron at the 3' splice site, thus joining the two exons together.

along with U1 and U4 to enable further rounds of splicing.

1.2.5 Cajal bodies

It is becomingly increasingly clear that subnuclear compartments known as Cajal bodies (CB) are intimately linked with every nuclear stage of the snRNP biogenesis pathway. First discovered in neurons by Ramón y Cajal at the beginning of the 20th century, interest in CBs increased when snRNAs were detected within these structures in 1991 (Carmo-Fonseca et al., 1991). Around the same time, a protein named coilin was identified as a marker and essential scaffold protein of CBs (Raška et al., 1990). In addition to neurons, Cajal bodies are also present in numerous dividing cell types, where their morphology changes throughout different stages of the cell cycle (Misteli, 2001). CBs are most frequent and at their largest during G1/S, and then disintegrate during mitosis (Misteli, 2001). This assembly/disassembly cycle appears to be regulated in part by the phosphorylation of coilin (Cantarero et al., 2015). Cantarero et al. (2015) demonstrated that VRK1 phosphorylates coilin in a cell-cycle dependent manner, and that this activity is required for CB formation as well as for the prevention of coilin degradation via the proteasome during mitosis (Cantarero et al., 2015).

In 2014, Machyna et al. (2014) performed coilin ChIP-seq assays, revealing that coilin is present at multiple RNAPII-transcribed snRNA loci (it is notably absent at the U6 locus) (Machyna et al., 2014). Consistent with this finding, multiple components of the snRNA transcription machinery, including proteins belonging to PTF and the LEC, have been identified in Cajal bodies (Hu et al., 2013; Schul et al., 1998). Moreover, the formation of Cajal bodies has been demonstrated by multiple groups to be dependent on ongoing snRNA transcription and the formation of nascent snRNA transcripts (Frey and Matera, 2001; Frey et al., 1999). Accordingly, environmental stresses that induce a transcriptional inhibition cellular response, such as exposure to UV-C radiation, also result in CB disintegration (Gridasova and Henry, 2005). Similarly,

snRNA genes are almost completely transcriptionally silent during mitosis, accounting for the disintegration of CBs witnessed at this cell cycle stage (Machyna et al., 2014).

snRNA transcription and Cajal body formation in fact appear to be a mutually dependent relationship. Disrupting Cajal bodies via the knockdown of essential CB protein components USPL1 or WRAP53 (Mahmoudi et al., 2010; Schulz et al., 2012) reduces the levels of nascent precursor U1 and U2 pre-snRNAs, indicating a direct role for CBs in snRNA transcription regulation (Wang et al., 2016). In 2016, Wang et al. (2016) used Genome-wide Chromosome Conformation Capture analysis (4C-seq) to assess the chromosomal regions that interact with CBs. They demonstrated that CBs facilitate the formation of CB-proximal inter-chromosomal gene clusters that are enriched in snRNA and intron-encoded snoRNA loci (Wang et al., 2016). They postulate that CBs act as a multi-chromosomal interface, promoting the spatial clustering of multiple sn/snoRNA loci in order to facilitate optimal transcription and processing (Sawyer et al., 2016). Interestingly, histone genes were also shown to be enriched among the CB-interacting loci, and as a result histone gene transcription was also downregulated upon RNAi-mediated disruption of CBs (Wang et al., 2016).

1.2.6 Cajal bodies and RNP Assembly

In addition to their role in snRNA transcription, it is thought that snRNAs traffic through CBs prior to their export to the cytoplasm. Firstly, Ints4 has been shown to localise to CBs, and knockdown of this Integrator subunit or Ints11 causes disintegration of CBs (Takata et al., 2012), suggesting a potential role for CBs in 3'-end processing. In addition, export factors PHAX and CRM1 also localise to CBs (Boulon et al., 2004), with PHAX knockdown similarly triggering CB disintegration (Lemm et al., 2006). Furthermore, disrupting snRNA export through PHAX inhibition causes pre-snRNAs to accumulate in frog oocyte CBs (Suzuki et al., 2010). Interestingly, knockdown of components involved in cytoplasmic snRNP maturation steps such SMN and TGS1 also results in CB disintegration (Lemm et al., 2006), although both have been

implicated in nuclear snoRNP biogenesis, which will be discussed in more detail in the 'snoRNA trafficking' section.

Once snRNAs have been re-imported into the nucleus, they initially accumulate at Cajal bodies (Sleeman and Lamond, 1999) where the addition of snRNP-specific protein components is thought to occur. This has been conclusively demonstrated in the case of U2 snRNP 17S complex formation (Nesic, 2004). Numerous studies have now established that the de novo formation and post-splicing reassembly of the U4/U6 di-snRNP and the U4/U6/U5 tri-snRNP also occurs in the Cajal body (Bell et al., 2002; Staněk and Neugebauer, 2004; Staněk et al., 2003) (Figure 1.6). SART3 is the protein responsible for targeting U6 snRNA to CBs and therefore promoting U4/U6 snRNA annealing (Staněk et al., 2003). It interacts with coilin and tethers immature snRNPs to this protein (Novotný et al., 2015). Novotny et al. (2015) also demonstrated that knocking down snRNP-specific proteins induces CB formation in cells normally lacking them, and that this process is dependent on expression of SART3 (Novotný et al., 2015). The accumulation of incomplete snRNPs in CBs, also observed when snRNA export or 3'-end processing is disrupted, strongly suggests that Cajal bodies act as centres for quality control during the snRNP biogenesis pathway. Following tri-snRNP assembly, SART3 is released prior to splicing, enabling it to direct recycling of individual U6 snRNPs released from the spliceosome (Bell et al., 2002).

1.2.7 RNA Modifications in Cajal bodies

In addition to the incorporation of snRNP-specific proteins following nuclear import, the nucleotides of snRNAs themselves are modified in Cajal bodies by a class of RNAs known as small CB-specific RNAs (scaRNAs). scaRNAs are a subclass of the small nucleolar RNA (snoRNA) family that localise permanently to CBs (Darzacq et al., 2002). Here, scaRNAs guide pseudouridylation of snRNAs in the case of H/ACA scaRNAs, while C/D scaRNAs catalyse 2′-O-methylation (Meier, 2017). U2 snRNA is the most post-transcriptionally modified out of all the snRNAs (Karijolich and Yu, 2010),
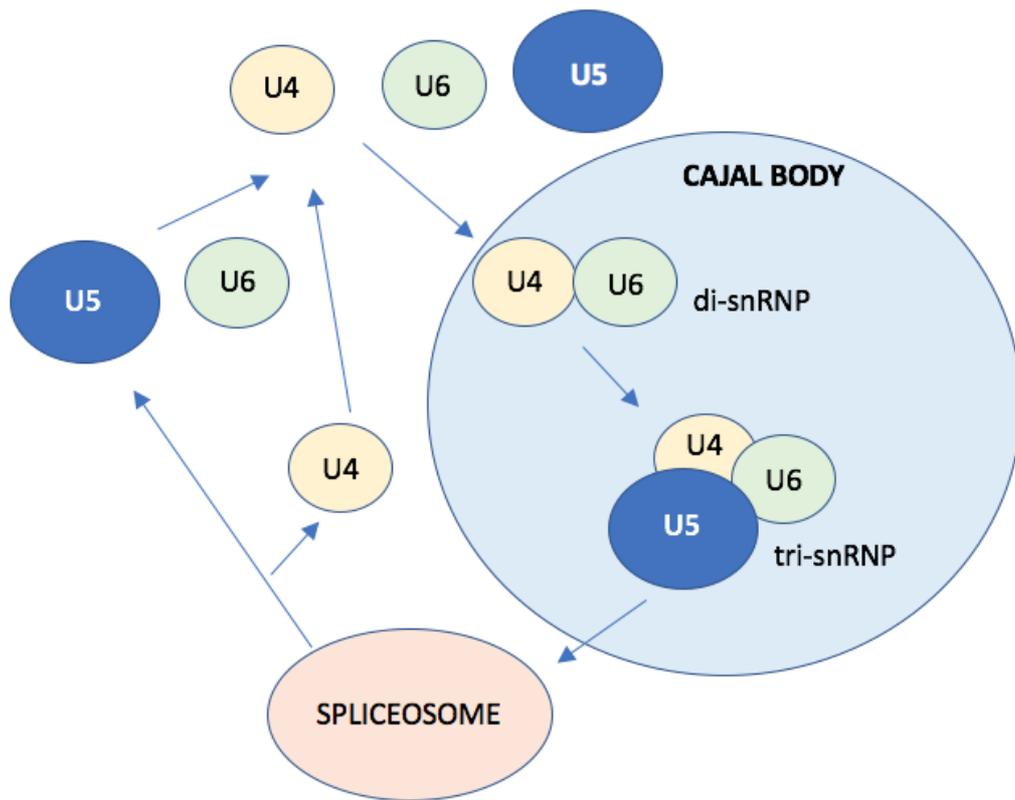
Figure 1.6 Assembly/disassembly cycle of the U4/U6 di-snRNP and U4/U6/U5 tri-snRNP complexes

U4 and U6 snRNAs anneal through base-pairing followed by the assembly of the U4/U6 di-snRNP complex within Cajal bodies. The U5 snRNP joins this complex to form the tri-snRNP, which can then enter the spliceosome. Structural rearrangements during splicing trigger the release of first U4 then U5 and U6. These individual snRNP components are then recycled into another round of re-assembly of the di-snRNP and tri-snRNP complexes.

and as a result the functional significance of these modifications has been studied in most detail. In 1998 it was demonstrated that U2 snRNAs lacking pseudouridine are unable to form functional 17S particles (Yu et al., 1998). Furthermore, Donmez et al. (2004) revealed that specific 2'-O-methylations at positions 1, 2, 12 and 19 are individually required for pre-mRNA splicing to operate successfully, facilitating the formation of complex E (Dönmez et al., 2004).

scaRNAs themselves are targeted to CBs by the essential CB component WRAP53 (Tycowski et al., 2009), which recognises a short sequence motif known as the CAB box (ugAG) present in H/ACA scaRNAs, and a G•U/U•G wobble stem in the case of C/D scaRNAs. Interestingly, WRAP53 is also required to mediate the interaction between coilin and SMN (Mahmoudi et al., 2010).

1.2.8 Cajal bodies and Neurodegeneration

Historically, loss of neuronal nuclear Gems – nuclear structures containing SMN but lacking coilin and snRNPs - has been considered a hallmark of SMA. However, it has been demonstrated that in postnatal and mature mammalian neurons, SMN in fact co-localises with coilin in the nucleus and coilin-negative Gems are absent (Carvalho et al., 1999; Pena et al., 2001). Many groups identify Gems via single immunostaining using an anti-SMN antibody, and this does not enable the distinction between coilin-lacking Gems and Cajal bodies containing coilin, SMN and snRNPs. Therefore, some argue that it is in fact Cajal body loss that is a defining feature of this condition (Lafarga et al., 2017), especially given that CB integrity is known to be dependent on ongoing snRNP biogenesis, which is disrupted in SMA (and in some forms of ALS). Consistent with this hypothesis, in 2012 the severe depletion of CBs in the motor neurons of a 3-month old SMA patient compared to age-matched control neurons was reported, while coilin-negative nuclear Gems were not present in either (Tapia et al., 2012). Interestingly, the remaining Cajal bodies in these patient neurons do not recruit SMN and snRNPs, suggesting a minimal role in snRNP biogenesis (Tapia et al., 2012). In

light of these findings, and given the literature demonstrating CB disintegration is induced by almost any defect in snRNP biogenesis, it appears highly likely neurodegenerative disorders such as SMA and ALS that are often underpinned by snRNP biogenesis defects will also be characterised by CB disruption.

1.3 snoRNA BIOGENESIS

1.3.1 snoRNAs

The scaRNAs described earlier are a subset of a larger family of approximately 750 small RNAs (Jorjani et al., 2016) known as small nucleolar RNAs (snoRNAs). scaRNAs are unusual among snoRNAs in that they catalyse base-modifications of snRNAs, whereas the remainder of this RNA family largely guide modifications of ribosomal RNAs (rRNAs). These RNAs, comprising between 70-200 nucleotides, can be classified into two categories based on conserved sequence and structural motifs – the C/D box, and the H/ACA box (Figure 1.7). The C/D box snoRNAs contain conserved C box and D box sequence elements at the 5' end and 3' end of the transcript respectively, which are brought within close proximity of each other as the molecule folds into its mature secondary structure (Darzacq and Kiss, 2002). These snoRNAs catalyse 2'-O-methylation of their rRNA targets, guided by specific base-pairing between their own nucleotides and that of the substrate (Dupuis-Sandoval et al., 2015). The H/ACA box snoRNAs are longer on average than their C/D box counterparts, formed of two stem loop structures separated by a conserved H box motif, with the ACA box located at the 3' end of the molecule (Dupuis-Sandoval et al., 2015). This type of snoRNA catalyses rRNA pseudouridylation, again guided by base-pairing with the substrate. This base-pairing mechanism has enabled the identification of rRNA targets for the majority of known snoRNAs, however approximately 70 snoRNAs are still considered orphan, without a known substrate (Jorjani et al., 2016). In addition to their distinct class of target, scaRNAs are also unusual among snoRNAs
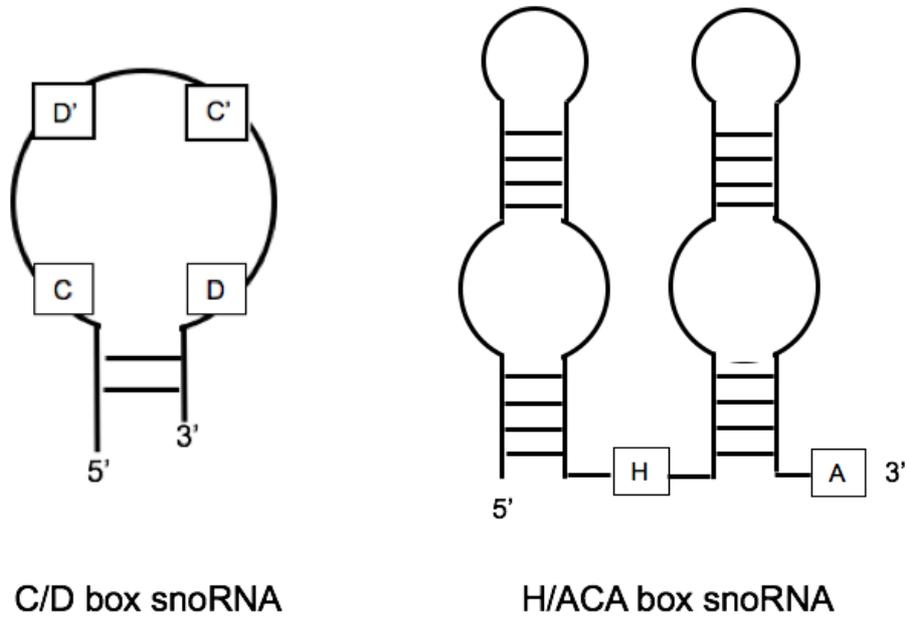
Figure 1.7 Schematic overview of C/D box and H/ACA box snoRNAs

C/D box snoRNAs possess C and D box motifs present near to the terminal stem, as well as internal C' and D' boxes. H/ACA box snoRNAs contain an internal H box motif sandwiched between two stem loop structures, along with an ACA box motif located at the 3' end.

in that some have been identified that are C/D box and H/ACA box hybrids, possessing both these structural motifs (Darzacq et al., 2002).

C/D box and H/ACA box snoRNAs each bind a set of four specific proteins - 15.5K, NOP56, NOP58 and fibrillarin in the case of C/D box snoRNAs (McKeegan et al., 2007) and Nhp2, Gar1, Nop10 and dyskerin in the case of H/ACA box snoRNAs (Lafontaine et al., 1998), forming a catalytically active snoRNP complex. The 15.5K protein is also found in the U4 snRNP, while NOP56 and NOP58 share structural similarities with U4 snRNP protein PRP31 (Bizarro et al., 2015). Interestingly, in 2015 it was reported that chaperones involved in box C/D snoRNP assembly, namely NUFIP, HSP9 and the R2TP complex, also promote U4 snRNP formation, revealing a common assembly machinery shared by these two types of small RNP particle (Bizarro et al., 2015).

The majority of snoRNAs are located within the introns of host pre-mRNAs or lncRNAs (Lykke-Andersen et al., 2018). The few exceptions to this rule that are transcribed from independent units – such as the highly conserved box C/D snoRNAs U3 and U8 – do not guide modifications of rRNAs, but instead promote rRNA endonucleolytic cleavage and folding (Kass et al., 1990; Tyc and Steitz, 1989). The abundance of the majority of snoRNAs is therefore dependent upon the transcription and splicing of the host gene in which the snoRNA resides. The RNA exosome complex degrades the excised intron of the host gene up to the snoRNA, which itself is protected from degradation by the snoRNPs that assemble upon it (Lykke-Andersen et al., 2018) (Figure 1.8). The exosome is targeted to the 3' end of these introns by a component of the nuclear exosome targeting (NEXT) complex known as RBM7 (Lubas et al., 2015). Thus, knockdown of exosome components results in elevated levels of 3'-end extended snoRNAs (Lubas et al., 2015). In the case of H/ACA box snoRNAs, it has been demonstrated that trimming of the last few nucleotides residing at the 3' end also requires poly(A) specific ribonuclease (PARN), which removes a small oligo(A) tail that is added by poly(A) polymerase PAPD5 (Berndt et al., 2012).

The expression levels of the host mRNA and the corresponding snoRNA are
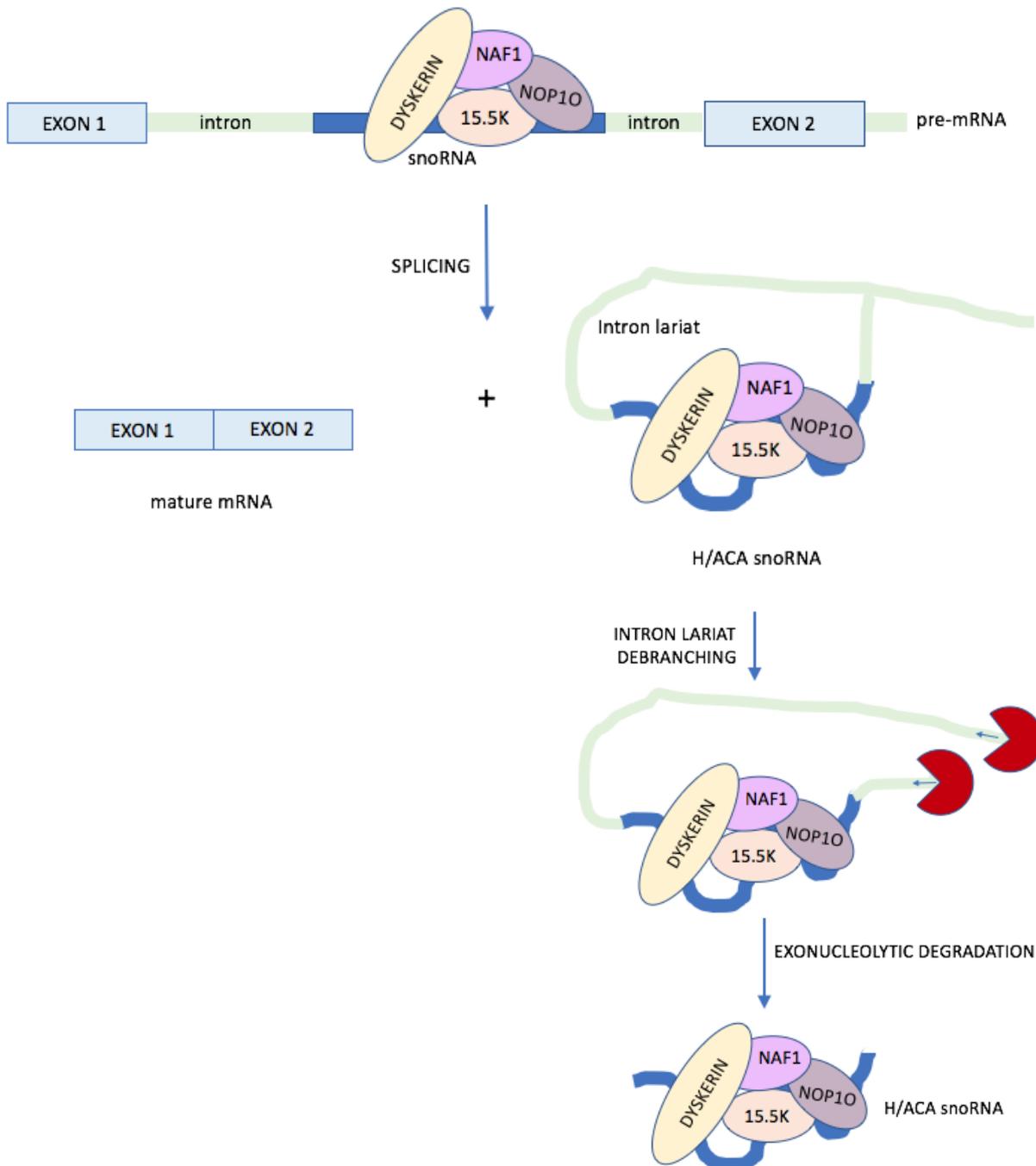
Figure 1.8 The majority of snoRNAs are processed from within introns of host pre-mRNAs

snoRNP proteins assemble upon snoRNAs located within introns co-transcriptionally. The intron surrounding the snoRNA is degraded exonucleolytically following splicing, with the snoRNA itself thought to be protected from degradation by the snoRNP proteins.

uncoupled through targeting of the host mRNA to the nonsense-mediated RNA decay pathway, with a high proportion of the host mRNA transcripts synthesized in these instances simply acting as a byproduct of the snoRNA biogenesis pathway (Lykke-Andersen et al., 2014). The stability of mature snoRNAs themselves are also regulated by the exosome. Interestingly, in this case it is DGCR8 that acts as a cofactor to target the exosome to these transcripts, independent from its role in micro-RNA processing as part of the microprocessor (Macias et al., 2015).

The rRNA modifications catalyzed by snoRNAs promote proper ribosome biogenesis and function (Herter et al., 2015). Therefore, the expression levels of snoRNAs directly contribute to the translation capacity of the cell, and must be tightly regulated. Indeed, many types of cancer result from uncontrolled, upregulated ribosome activity, with oncogene MYC known to regulate the transcription of many genes involved in the ribosome biogenesis pathway (Van Riggelen et al., 2010).

Interestingly, snoRNA-containing genes themselves were reported as a novel class of MYC targets in 2015 (Herter et al., 2015). MYC binds the promoters of these snoRNA loci to promote their expression, while also regulating the transcription of the snoRNA-associated proteins (Herter et al., 2015). In this way, MYC acts a master regulator of snoRNP biogenesis. Another key protein regulating the transcription of snoRNA-containing genes is the RNA helicase DDX21, which promotes the release of CDK9 from the inhibitory 7SK complex at these loci in order to facilitate proper transcription elongation (Calo et al., 2015).

1.3.2 snoRNA Trafficking

While scaRNAs localize to Cajal bodies to perform snRNA base modification, targeted there via WRAP53, the rRNA-modifying snoRNAs traffic to the nucleolus to modify their RNA substrates. In the case of the independently transcribed snoRNAs U3, U8 and U13, their 5'-end m_7G cap is recognized by the same CBC-PHAX complex that binds to pre-snRNAs, which directs them to Cajal bodies for further maturation (Hallais

et al., 2013). Here these snoRNAs undergo trimming of their 3'- end, addition of the snoRNP-specific proteins and hypermethylation of their cap by trimethylguanosine synthase 1 (TGS1), which resides in CBs when in the nucleus (Machyna et al., 2013). Intriguingly, SMN directly interacts with snoRNP proteins fibrilllarin and Gar1, while U3 snoRNA accumulation in the nucleolus can be inhibited by a dominant negative form of SMN (Jones et al., 2001; Pellizzoni et al., 2001). This suggests that SMN also performs a role in the snoRNP maturation pathway in addition to its snRNP biogenesis function. Mature snoRNPs are then released to the nucleolus by CRM1, which displaces TGS1 from the complex (Boulon et al., 2004).

As in the case of snRNA loci, snoRNA genes are also located in the vicinity of CBs, indicating a direct role for CBs in their transcription (Wang et al., 2016). Indeed, it is becoming increasingly clear that it is not only the independently transcribed snoRNAs that associate with CBs at some stage during their maturation. Telomerase is an enzyme comprising telomerase RNA (TERC), telomerase reverse transcriptase (TERT) plus an associated set of factors, which is responsible for regulating telomere length. TERC possesses a CAB-box containing H/ACA box motif, which enables its localisation to CBs (Jády et al., 2004). In this sense, TERC could be considered a scaRNA. Mature TERC also possesses a TMG, which is likely added by TGS1 within CBs (Machyna et al., 2013).

Given that intron-encoded snoRNAs do not possess a 5'-end m_7G cap, it came as a surprise when these transcripts were also reported to interact with CB-localized TGS1 despite not being its substrate (Pradet-Balade et al., 2011). However, in accordance with this finding, coilin iCLIP performed by Machyna et al. (2014) revealed that hundreds of snoRNAs strongly interacted with this core CB component, the vast majority of which being intron-encoded transcripts (Machyna et al., 2014). Interestingly, coilin iCLIP tags were absent on the surrounding pre-mRNAs, suggesting that the protein interacts with snoRNAs after they are processed from their host transcript (Machyna et al., 2014). Furthermore, Machyna and colleagues injected fluorescently

labelled snoRNAs and tracked their localisation within the cell, conclusively demonstrating that all snoRNAs concentrate first in CBs prior to trafficking to the nucleolus. Recent iCLIP studies performed on ARS2 and PHAX demonstrated that PHAX binds mature, uncapped snoRNAs, whereas ARS2 did not (Giacometti et al., 2017). This indicates that PHAX can be recruited to these transcripts independently of CBC/ARS2 to enable their nuclear transport to CBs and then nucleolus, with this operation competing with the NEXT/nuclear exosome degradation pathway to define snoRNA fate.

1.4 HETEROGENOUS NUCLEAR RIBONUCLEOPROTEINS (hnRNPs)

1.4.1 hnRNPs

Messenger RNAs (mRNAs) undergo numerous processing steps both during and following their transcription, including 5' capping, splicing and polyadenylation. Hundreds of RNA-binding proteins (RBPs) associate with nascent mRNAs to direct and regulate these steps. One major class of RBP is the heterogeneous nuclear ribonucleoproteins (hnRNPs). The members of this family, termed A through to U, perform roles in multiple stages of nucleic acid metabolism ranging from transcription regulation to RNA transport throughout the cell. The first to be identified were hnRNP-A/B and hnRNP-C, purified via sucrose density gradients along with associated RNAs as part of the 40S particle (Beyer et al., 1977). Subsequently, immunoprecipitation of hnRNP-C co-purified 20 interacting partners, enabling the compilation of a definitive list of hnRNP family members (Piñol-Roma et al., 1988).

While predominantly located in the nucleus, hnRNPs have the ability to shuttle between the nucleus and the cytoplasm (Piñol-Roma and Dreyfuss, 1992), enabling them to perform functions in both these cellular compartments and therefore throughout the entire gene expression pathway. hnRNPs interact with RNA via four

types of RNA-binding domain: the RNA-recognition motif (RRM), the quasi-RRM, the KH domain, and the RGG box. Each hnRNP consists of different combinations of these domains, often joined by linker regions as well as further auxiliary domains that enable protein-protein interactions. The modular composition of the various family members facilitates RNA-substrate specificity, as well as enabling the broad functional diversity observed among hnRNPs. For example, hnRNP family member HNRNPD has a high affinity for AU-rich sequences located within the 3' UTR of mRNAs, and is therefore associated with mediating mRNA decay (Fialcowitz et al., 2005). Conversely, HNRNPL has been shown to specifically bind exons within *CD45* mRNA, triggering changes in its splicing patterns during the immune response (Melton et al., 2007). In addition, many members of the hnRNP family can exist in multiple different isoforms generated via alternative pre-mRNA splicing, which also contributes to this functional variation.

In recent years, it has become apparent that this functional diversity is even larger than previously thought, with several members of the hnRNP family now identified as performing functions in pathways distinct from mRNA processing, such as DNA repair (Wang et al., 2013) and chromatin remodelling (Mahajan et al., 2005). For example, in 2007 it was discovered that HNRNPA1 binds to and promotes the efficient processing of the microRNA (miRNA) precursor molecule pri-miR-18a by Drosha (Guil and Cáceres, 2007). Secondly, in addition to its role in regulating the splicing of *SMN* pre-mRNA, hnRNP family member HNRNPU also regulates the maturation of U2 snRNP complex formation (Xiao et al., 2012). Xiao et al. (2012) demonstrated that HNRNPU binds all types of snRNA, and that knockdown of the protein triggers significantly increased numbers of Cajal bodies in the nucleus. Indeed, HNRNPU displays a quite remarkable functional repertoire. Not only is it essential for the chromosomal localisation of Xist lncRNA (Hasegawa et al., 2010), but it can also act as a repressor of transcription elongation through inhibiting RNAPII CTD kinase CDK7 (Kim and Nikodem, 1999). Given the wide-ranging activities of the hnRNP family, it is unsurprising that many of these proteins have been implicated in diseases such as

cancer (Carpenter et al., 2006) as well as several neurodegenerative conditions (discussed in further detail in the upcoming 'RNA-binding proteins and ALS' section).

1.4.2 HNRNPUL1

HNRNPUL1 is a member of the hnRNP family first identified as an interactor of adenovirus type 5 (Ad5) early 1B 55-kDa protein (E1B-55kDa) (Gabler et al., 1998), and hence initially referred to as E1B-AP5. However, given its significant structural and sequence homology to HNRNPU, it was renamed HNRNPU-Like-1 (HNRNPUL1). The HNRNPUL1 gene is located on chromosome 19 at the region 19:41.26-41.31Mb (19q13.2), consisting of 15 exons. Once translated, HNRNPUL1 comprises 856 amino acids, and like other hnRNPs it possesses several distinct domains joined by linker regions (Figure 1.9). At its N-terminus is a SAP domain that possesses DNA-binding capacity, while its C-terminus is made up of an RGG box and a poly-proline rich region. Described first in HNRNPU, RGG boxes consist of several arginine-glycine-glycine repeats clustered closely together, and facilitate RNA-binding as well as protein-protein interactions (Thandapani et al., 2013). Several of the arginine residues present in the RGG box of HNRNPUL1 are targets of arginine methylation via protein arginine methyltransferase 1 (PRMT1), and these modifications affect the interactions of HNRNPUL1 with its protein binding partners (Gurunathan et al., 2015).

Located in the central region of HNRNPUL1 is a SPRY domain that is also thought to mediate protein-protein interactions (Woo et al., 2006), next to a domain identified by Gabler et al. (1998) as a putative nucleoside triphosphate (NTP)-binding motif. This region contains a Gxxxx-GKS/T sequence (amino acids 428-435) of residues known as the Walker A motif. Walker A motifs are present in a large class of nucleotide-binding proteins known as P-loop NTPases. The motif forms a characteristic loop structure known as a P-loop (phosphate-binding loop), containing a pair of conserved glycine residues that facilitate hydrogen-bonding interactions with nucleotide phosphoryl groups. A second motif known as the Walker B motif is also found in P-
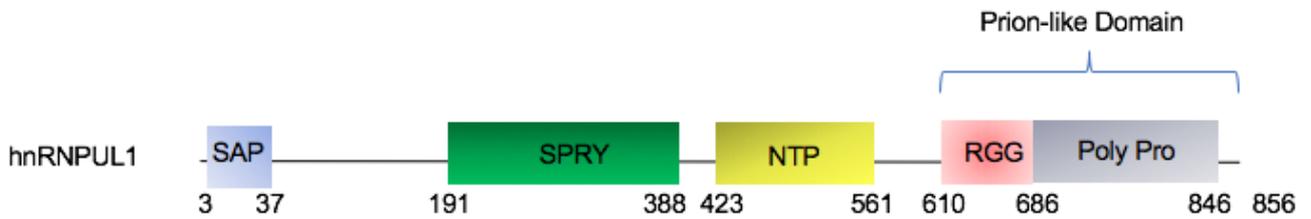
Figure 1.9 Domain organisation of HNRNPUL1

Schematic diagram of the domains of HNRNPUL1. Known functions of each domain:
SAP = DNA-binding, SPRY = protein-interactions, NTP = nucleotide-binding. RGG =
ssDNA/RNA/protein-binding, Poly-pro = protein interactions.

loop NTPase proteins, comprising four hydrophobic residues followed by a conserved aspartic acid or glutamic acid, i.e. hhhhD/E. This motif binds the co-substrate Mg_{2+} ion and is essential for catalysis of the nucleotide (Matte and Delbaere, 2006). HNRNPUL1 also possesses a putative Walker B motif (amino acids 501-505, sequence: NYILD) approximately 70 residues downstream of its Walker A motif. Gaber et al. (1998) postulate that therefore HNRNPUL1 is a GTP-binding protein, noting its similarities with the small GTP-binding protein Ran. In agreement with this notion, HNRNPUL1 appears to bind GTP preferentially over ATP in vitro, although no GTP hydrolysis activity has been identified (Wilson lab, unpublished). However, Expasy online software tools reveal that the NTP region of HNRNPUL1 shares most structural homology with polynucleotide kinase phosphatase (PNKP). Following DNA damage, this enzyme uses ATP to phosphorylate the 5' ends of DNA molecules as part of the DNA repair pathway (Bernstein et al., 2005).

Interestingly, the central NTP-binding region of HNRNPU was recently shown to possess ATPase, but not GTPase, activity (Nozawa et al., 2017). Nozawa et al. (2017) also demonstrated that this ATPase activity was increased in the presence of RNA, with RNA-binding facilitated by the RGG box of HNRNPU downstream of the central NTP-binding region.

Functionally, HNRNPUL1 was first linked to mRNA-processing when Gaber et al. (1998) revealed that stable expression of the protein overcame E1B-dependent inhibition of cytoplasmic host mRNA accumulation in Ad-infected cells. In addition, HNRNPUL1 has been reported to interact with mRNA export receptor NXF1 (Bachi et al., 2000), again suggesting a role in RNA metabolism. Intriguingly, HNRNPUL1 was also identified via mass spectrometry as a strong interactor of ARS2 – a component of the CBCA complex described in the 'snRNA Export and Cytoplasmic RNP Assembly' section (Hallais et al., 2013). This complex regulates 3' end processing of several RNA families such as pre-snRNAs and pre-snoRNAs, while also regulating the balance between their nuclear export or degradation via mutually exclusive interactions with

PHAX or NEXT (Giacometti et al., 2017). Therefore, this reported interaction suggests HNRNPUL1 may also participate in small RNA biogenesis in addition to any role relating to mRNA metabolism.

However, in recent years HNRNPUL1 has been implicated more directly in transcription regulation, as well as the DNA damage response. Overexpression of the protein has been demonstrated to repress basal transcription from a range of promoters, with a transcription factor known as BRD7 forming a complex with HNRNPUL1 to modulate this activity (Kzhyshkowska et al., 2003). Immunoprecipitation/Mass Spectrometry assays performed on RNA polymerase II revealed HNRNPUL1 as one of the top interacting hits, suggesting a direct role in the regulation of this enzyme's activity (Chi et al., 2018a). In addition, HNRNPUL1 directly binds transcription factor p53 and inhibits its transcriptional activity following UV radiation (Barral et al., 2005). In 2012, HNRNPUL1 was also linked to the regulation of histone gene transcription (Ideue et al., 2012). Histone gene transcription is activated during S phase to ensure newly synthesized DNA is packaged into chromatin, and then highly repressed in other phases of the cell cycle. Ideue et al. (2012) revealed that the U7 snRNP mediates this transcriptional repression by recruiting HNRNPUL1 to histone gene loci during non-S phase periods.

In the same year, Polo et al. (2012) demonstrated that HNRNPUL1 was recruited to sites of DNA damage by the DNA-double stranded break sensor complex known as MRN, along with a highly homologous member of the hnRNP family called HNRNPUL2 (Polo et al., 2012). At these sites of DNA damage, HNRNPUL1 and HNRNPUL2 mediate recruitment of Bloom syndrome helicase to promote repair of the double-stranded breaks. HNRNPUL1 can also be recruited to sites of DNA damage in a PARP1-mediated poly-(ADP-ribosyl)ation dependent manner (Hong et al., 2013). In addition, HNRNPUL1 forms a complex with a long non-coding RNA (lncRNA) known as *DDSR1* at sites of DNA-damage to modulate the activity of DNA repair protein BRCA1 (Sharma et al., 2015).

HNRNPUL1 interacts with another lncRNA known as *NEAT1* in the context of paraspeckles (Naganuma et al., 2012). Paraspeckles are subnuclear bodies comprising *NEAT1* and approximately 40 RNA-binding proteins. These structures are not present under basal conditions in every cell type, but can be induced by cellular stress such as proteasome inhibition (Hirose et al., 2014) or viral infection (Imamura et al., 2014). The formation of paraspeckles in effect sequesters paraspeckle proteins such as HNRNPUL1, and given that many normally perform functions related to transcription regulation or alternative splicing, paraspeckle formation is associated with gene expression changes (Imamura et al., 2014). *NEAT1* is essential for paraspeckle formation along with seven paraspeckle proteins, including NONO, SFPQ and RBM14 (Naganuma et al., 2012). Naganuma et al. (2012) showed that knockdown of HNRNPUL1 caused a substantial, but not total, decrease in paraspeckle numbers. Intriguingly, several paraspeckle proteins such as FUS, TAF15 and HNRNPA1 are associated with the neurodegenerative disorder amyotrophic lateral sclerosis (ALS), and paraspeckles have been identified in the motor neurons of ALS patients (Nishimoto et al., 2013).

1.5 AMYOTROPHIC LATERAL SCLEROSIS (ALS)

1.5.1 ALS Pathogenesis

ALS is a neurodegenerative disorder characterised by injury and death to both the upper motor neurons in the motor cortex and the lower motor neurons in the brain stem and spinal cord. As the disease progresses, the loss of these neurons results in widespread failure of the neuromuscular system and as a result most patients die within 3-5 years of diagnosis due to respiratory failure (Taylor et al., 2016). In populations of European descent, ALS occurs at an incidence of around 3 in 100000, but is less frequent in South and East Asian populations (0.7-0.8 cases per 100000) (Chiò et al., 2013). Around 5-10% of cases are hereditary, usually transmitted in an

autosomal dominant manner, with the rest of cases occurring in a sporadic fashion (Taylor et al., 2016).

Mutations in over 30 genes have been identified as causing ALS (Walsh et al., 2015). The first gene to be discovered encodes an enzyme known as superoxide dismutase (SOD1), with mutations in this gene responsible for approximately 20% of inherited cases. This protein converts toxic superoxide into hydrogen peroxide and oxygen. ALS-causing mutations have been identified throughout the protein, and it appears that the resulting disease phenotype arises independently of any change to dismutase activity (Taylor et al., 2016). Instead, a toxic gain of function mechanism has been proposed. SOD1 mutants often fail to fold properly, and as a result form large ubiquitinylated protein aggregates in the cytoplasm. Multiple deleterious cellular events occur due to these toxic aggregates - such as disruption of cytoskeleton, impaired protein autophagy pathways, and mitochondrial dysfunction, all resulting in progressive neurodegeneration (Ferraiuolo et al., 2011). Interestingly, mutations in genes involved in these processes, such as *OPTN* in the case of autophagy (Maruyama et al., 2010) and *DCTN1* in the case of axonal cytoskeleton organisation (Puls et al., 2003), have also been identified as ALS-causing. This indicates a common molecular pathology underpinning different genetic subtypes of familial ALS.

The most common genetic cause of ALS – a hexanucleotide (GGGGCC) repeat expansion within the first upstream intron of *C9orf72* - was discovered in 2011 (DeJesus-Hernandez et al., 2011). *C9orf72* encodes a guanine nucleotide exchange factor (GEF) that regulates specific Rab GTPases involved in autophagy (Webster et al., 2016). While some argue that reduced C9orf72 protein function underpins the molecular pathology of *C9orf72*-related ALS, a more widely held view is that toxic gain of function at the RNA and/or protein level is a more likely cause (Garcia-Santibanez et al., 2018). *C9orf72* intronic repeats form nuclear RNA foci that have high affinity for numerous RNA-binding factors (Lee et al., 2013). These foci sequester these proteins, leading to widespread dysregulation of RNA metabolism in patient neurons.

Additionally, repeat-associated non-AUG (RAN) translated dipeptide repeats (DPR) are synthesized based on the hexanucleotide expanded region, which form cytoplasmic aggregates that may also be neurotoxic.

1.5.2 RNA Metabolism and ALS

A major advance in ALS research occurred in 2006, when Neumann et al. reported mislocalization of RNA-binding protein TDP-43 in ALS patients from the nucleus to ubiquitin-positive cytoplasmic aggregates (Neumann et al., 2006). Since this breakthrough, the presence of TDP-43 cytoplasmic inclusions in patient neurons has become a hallmark of ALS pathology in the majority of sporadic and inherited forms of ALS. TDP-43 performs multiple roles in RNA metabolism, ranging from alternative splicing (Lagier-Tourenne et al., 2012) to mRNA axonal transport (Yoshimura et al., 2006). It binds to thousands of RNA targets (Tollervey et al., 2011), and therefore ALS-causing mutations that perturb the localisation of TDP-43 consequently affect RNA processing on a massive scale. Interestingly, Lagier-Tourenne et al. (2012) revealed that the RNAs most affected by TDP-43 loss were enriched with genes that encode proteins required for neuronal integrity (Lagier-Tourenne et al., 2012), suggesting a potential explanation for the specific susceptibility of neuronal cells to RNA-binding protein mutations.

Following the initial discovery of TDP-43, mutations in numerous other RNA-binding proteins such as FUS (Vance et al., 2009), HNRNPA1 (Kim et al., 2013) and MATRIN-3 (Johnson et al., 2014) have also been identified as ALS-causing, confirming the link between dysregulated RNA metabolism and ALS pathogenesis (Figure 1.10). Interestingly, many of these proteins have also been identified as strong interactors of the *C9orf72* GGGGCC RNA hexanucleotide repeats (Haeusler et al., 2014), again suggesting a common pathological mechanism underpinning different forms of familial ALS. Along with TDP-43, FUS is the RNA-binding factor that most commonly causes ALS (both account for approximately 4-5% of inherited cases each). FUS belongs to

RNA-binding protein	Gene Locus	Discovery date of linkage to ALS	Proportion of ALS familial cases
TDP43	1p36	2008	5%
FUS	16p11	2009	5%
hnRNPA1	12q13	2013	<1%
hnRNPA2/B1	7p15	2013	<1%
MATRIN-3	5q31	2014	<1%
SETX	9q34	2004	<1%
TAF15	17q12	2011	<1%
ELP3	8p21	2009	<1%
ATXN2	12q24	2010	<1%
ANG	14q11	2004	<1%
EWSR1	22q12	2012	<1%
TIA1	2p13	2017	<1%

Figure 1.10 Numerous RNA-binding proteins are associated with ALS

A table of the RNA-binding proteins so far discovered that have been unequivocally implicated in the pathogenesis of ALS. Information sourced from Renton et al., 2014; Taylor et al., 2016; and Zhao et al., 2018 (Renton et al., 2014; Taylor et al., 2016; Zhao et al., 2018).

the FET family of proteins along with TAF15 and EWSR1. It is a 526-amino acid protein, comprising a glutamine, glycine, serine and tyrosine-rich region (QGSY) at its N-terminus, followed by a glycine-rich region, an RRM, and two RGG boxes separated by a zinc-finger motif. A nuclear-localisation signal (NLS) resides at its C-terminus, and the majority of ALS-causing mutations are located in this region of the protein, along with the glycine-rich region (Lagier-Tourenne et al., 2010) (Figure 1.11). These NLS mutations cause the mislocalization of FUS into cytoplasmic aggregates that also sequester other RNA-binding factors, triggering broad changes to RNA processing in ALS patient neurons (Fujioka et al., 2013).

As with other RNA-binding proteins, the reported functional repertoire of FUS has grown in recent years, ranging from miRNA biogenesis (Zhang et al., 2018) to the DNA damage response (Rulten et al., 2014). In 2015, it was demonstrated that FUS is essential to maintain the interaction with RNA polymerase II and U1 snRNP, in effect coupling transcription with splicing (Yu and Reed, 2015). FUS further regulates RNAPII by preventing Ser2 hyper-phosphorylation of the CTD by CDK9, ensuring successful elongation from transcription start sites (Schwartz et al., 2012). In addition, FUS is also one of the few paraspeckle proteins essential for their formation (Naganuma et al., 2012), and interestingly ALS-causing FUS mutants prevent the formation of paraspeckles in patient neurons (Shelkovnikova et al., 2014)

FUS also directly interacts with SMN (Yamazaki et al., 2012). As discussed earlier, mutations in SMN cause the neurodegenerative condition spinal muscular atrophy (SMA), with loss of nuclear SMN-containing structures known as Gems a hallmark of the condition. Yamazaki et al. (2012) revealed that FUS expression in HeLa cells was essential for Gem formation, while an ALS-causing mutant FUS R495X also reduced Gem numbers in patient fibroblasts (Yamazaki et al., 2012). Given the fundamental role of SMN in snRNP biogenesis, it is therefore possible that disrupted snRNP formation is one cause of the splicing alterations observed in FUS-related ALS.

These findings also indicate a common molecular pathway underpinning the
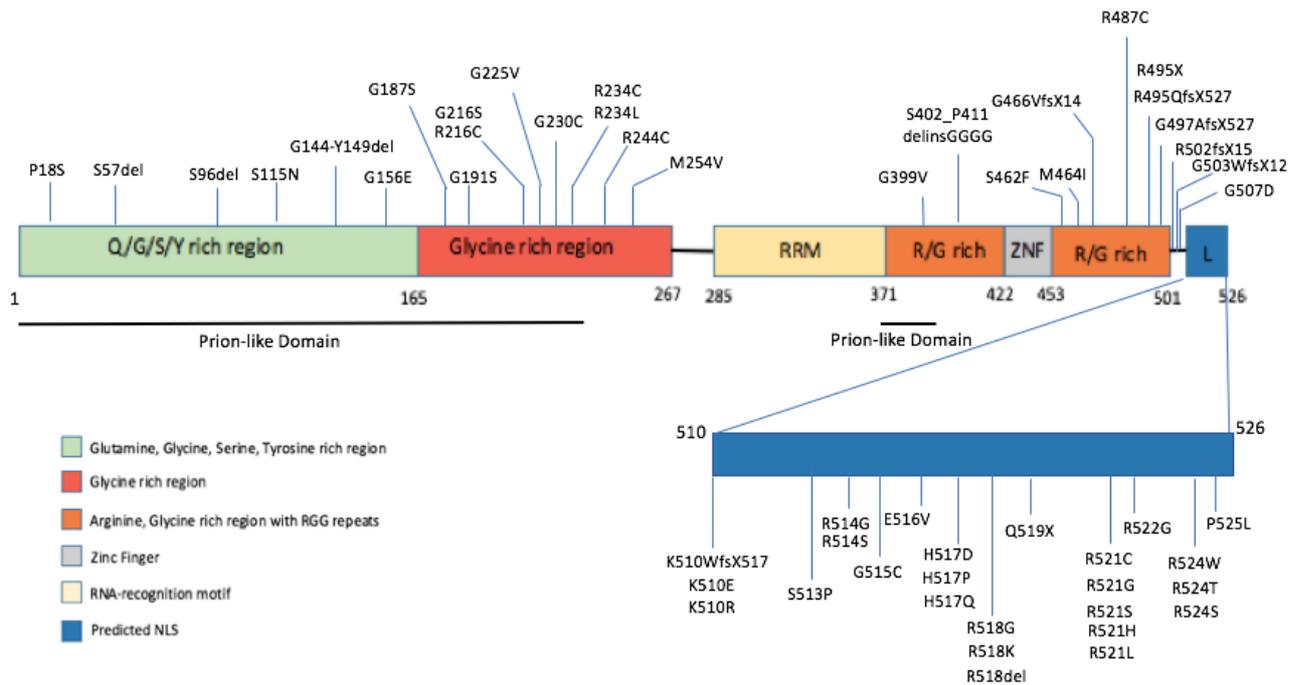
Figure 1.11 ALS-causing FUS mutations are primarily located within its glycine-rich region and NLS

Schematic diagram of the domains of FUS and the locations of ALS-causing mutations so far identified. Adapted from Shang and Huang, 2016 (Shang and Huang, 2016).

pathology of ALS and SMA. Consistent with this hypothesis, ALS mouse models display reduced SMN protein expression (Turner et al., 2008), while decreases in minor spliceosome snRNA levels have also been reported in both ALS and SMA (Onodera et al., 2014). Interestingly, ALS-causing forms of SOD1 also alter the nuclear localisation of SMN through disrupting the interaction between SMN and the core Cajal body component coilin (Kariya et al., 2012). This results in the loss of nuclear Gems.

1.5.3 Prion-like domains and ALS

RNA-binding proteins in association with their RNA substrates often assemble into higher-order complex particles, driven by liquid-phase transition (Lin et al., 2015). The formation of these RNP granules is facilitated in part by the presence of prion-like domains (PrLDs) within these RNA-binding proteins. PrLDs, named due their similarities to the yeast prion protein, are low complexity regions enriched with uncharged polar amino acids and glycines. These regions can be identified using a Hidden Markov Model known as PLAAC (prion-like amino acid composition), first described by Lancaster and colleagues in 2014 (Lancaster et al., 2014). In recent years, it has become apparent that PrLDs are essential in facilitating the functional aggregation of RNA-binding proteins. One important example is the formation of stress granules in the cytoplasm in response to environmental stress such as infection or heat shock. During times of cellular stress, the cell responds by limiting translation to only the most essential proteins to help conserve energy and facilitate recovery. Therefore, many non-essential RNA transcripts are sequestered in temporary, membrane-less organelles known as stress granules via RNA-binding proteins such as TDP-43 (McDonald et al., 2011). In 2011, it was reported that the PrLD of TDP-43 is essential for this ability to drive RNP stress granule accumulation (Dewey et al., 2011). A second example is that of paraspeckles, with Hennig et al. (2015) demonstrating in 2015 that the PrLDs of both RBM14 and of FUS are required for the formation of these subnuclear bodies (Hennig et al., 2015).

In addition to TDP-43 and FUS, several other RNA-binding proteins associated with ALS such as EWSR1, HNRNPA2B1 and TAF15 all also possess prion-like domains (Harrison and Shorter, 2017). While the ability to assemble into higher order structures is normally functionally advantageous, mutations in these domains can result in irreversible, excessive protein aggregation with cytotoxic consequences. Indeed, a high proportion of ALS-causing mutations found in these RBPs including FUS (Figure 1.11) are located within their PrLDs (Taylor et al., 2016), consistent with the hypothesis that these domains are the causative factor underpinning the pathology of RBP-mediated ALS.

PLAAC PrLD prediction software analyses reveal that HNRNPUL1 also possesses a PrLD, located at its C-terminal region from amino acids 615-856. In 2013, a study that analysed the prion propensity of all RBPs (utilizing the algorithm outlined in Alberti et al. (2009)) ranked HNRNPUL1 10th in this criteria, notably above known ALS-causing proteins such as HNRNPA1 and TDP-43 (FUS ranked 2nd on this list) (Li et al., 2013).This feature, along with its links to RNA-processing and its interaction with *C9orf72* GGGGCC RNA repeats (Cooper-Knock et al., 2018), marks HNRNPUL1 as a strong candidate for an ALS-associated gene. Moreover, HNRNPUL1 has also been reported to directly bind to FUS (Raczynska et al., 2015), strongly supporting the notion that they function in a common pathway. The HNRNPUL1-FUS interaction was confirmed in 2018, when Chi et al. (2018) performed immunoprecipitation/mass spectrometry assays on four ALS-associated RBPs - FUS, EWSR1, TAF15 and MATRIN-3 (Chi et al., 2018b). HNRNPUL1 was identified as a co-immunoprecipitating interactor of all four proteins, and was particularly abundant in the case of EWSR1 (Chi et al., 2018b). In the same paper, U1 snRNP immunoprecipitation/mass spectrometry was performed, and again HNRNPUL1 was a particularly prominent interactor, notably more so than FUS (Chi et al., 2018b). This points to a previously unreported role for HNRNPUL1 in splicing or snRNP regulation, which is further investigated in this study.

1.6 AIMS OF THE STUDY

HNRNPUL1 shares multiple characteristics with numerous ALS-causing RNA-binding proteins, and has been reported to interact with several of them, including FUS, TAF15 and MATRIN-3. However, there is little known about its functions relating to RNA metabolism. Given the lack of effective treatments for ALS, there is a pressing need to improve our understanding of the cellular roles of ALS-associated proteins. Therefore, the aims of this study are to determine the functions of HNRNPUL1 and investigate its relationship with ALS. To that end, cell lines will be generated that facilitate the rapid and controlled depletion of HNRNPUL1 to study the consequences of its absence from the cell. The prion-like domain located at the C-terminus of HNRNPUL1, which is present in many other ALS factors, is of particular interest and functional assays involving HNRNPUL1 mutants lacking this region will be performed. These studies will also involve HNRNPUL1 NTP-binding domain mutants, building on previous work in the Wilson lab that demonstrated that this region enabled HNRNPUL1 to bind nucleotides *in vitro*. The *in vivo* consequences of disrupting this nucleotide-binding capacity will be explored, including how it affects the interactome of the protein, its RNA-binding ability and its cellular localisation.

In addition, the relationship between HNRNPUL1 and FUS will be studied in detail, as part of the investigation into the links between HNRNPUL1 and ALS. Concomitantly, the cellular consequences of any HNRNPUL1 ALS mutations discovered via our collaboration with the Sheffield Institute for Translational Neuroscience (SITraN) will be examined, both through complementation studies and assays performed in ALS patient lymphoblastoid cell lines.

CHAPTER 2 - MATERIALS AND METHODS

2.1 MATERIALS:

2.1.1 BACTERIAL STRAINS AND MEDIA

The *E.coli* strain DH5α (Invitrogen) was utilised for cloning, plasmid purification and amplification, while expression of recombinant proteins was performed in the *E.coli* strain BL21 (Invitrogen).

Growth media was prepared using Millipore water, followed by autoclaving at 126°C.

Luria broth (LB): 10 g/l Tryptone, 10 g/l NaCl, 5 g/l Yeast extract.

LB Agar plates: LB as above with 2% agar.

Terrific Broth (TB): 12 g/L Tryptone, 24 g/l Yeast extract, 15.54 g/l K_2HPO_4, 2.31 g/l KH_2PO_4, 4 ml/l Glycerol.

Antibiotic Selection Conditions

ANTIBIOTIC	CONCENTRATION
Ampicillin	100 µg/ml

2.1.2 PLASMIDS

PLASMID	Resistance	Source
pX330	Amp	Steven West
pUC18	Amp	Steven West
pcDNA5-FRT	Amp	Invitrogen

pPGKFLPobpA	Amp	Addgene
pMK243 (Tet-OsTIR1-PURO)	Amp	Addgene
AAVS1 T2 CRIPR in pX330	Amp	Addgene
pGEX-6P1	Amp	Amersham

2.1.3 TISSUE CULTURE

Cell Lines

HeLa - Human cervical epithelial carcinoma cell line.

293T - Human embryonic kidney cell line that expresses the SV40 large T antigen.

Flp-In T-Rex 293 - Human embryonic kidney cell line, containing a single integrated FRT site.

HCT116 - Human colorectal carcinoma cell line.

HCT116-TIR1 - Human colorectal carcinoma cell line expressing the TRANSPORT INHIBITOR RESPONSE 1 protein (TIR1) from Oryza Sativa (os) (Eaton et al., 2018).

Lymphoblastoid cell lines - peripheral B lymphocytes derived from individual ALS patients/controls, which have been transformed by Epstein-Barr virus (EBV).

Growth Media

HeLa, 293T, HCT116 - Dulbecco's Modified Eagle Medium (DMEM, life technologies) supplemented with 10% FCS (v/v) (life technologies) and 1% Penicillin-Streptomycin (v/v) (Pen/Strep, Invitrogen).

Flp-In T-Rex 293 - DMEM supplemented with 10% Tet-free FCS (v/v) (life technologies), 1% Pen/Strep (v/v), 100 μg/ml Zeocin (Invitrogen) and 15 μg/ml Blasticidin (Invitrogen). 100 μg/ml Hygromycin (Invitrogen) was added following the generation of an inducible Flp-In cell line to maintain selection.

HCT116-TIR1 – DMEM supplemented with 10% FCS (v/v), 1% Pen/Strep (v/v) and 10 µg/ml Blasticidin. Following the integration of the HNRNPUL1-AID tag, selection was maintained via the addition of 150 µg/ml Hygromycin, 800 µg/ml G418/Neomycin (ThermoFisher), and 1 µg/ml Puromycin (ThermoFisher) to the media.

Lymphoblastoid cell lines – RPMI 1640 Medium supplemented with 20% FCS (v/v) (life technologies) and 200 µM L-glutamine.

2.1.4 BUFFERS AND SOLUTIONS

Miscellaneous Buffers

1X PBS (Phosphate-Buffered Saline): 137 mM NaCl, 2.7 mM KCL, 1.47 mM KH_2PO_4, 4.3 mM NaH_2PO_4, HCl used to bring to pH7.4.

mRNP Capture Buffers

2X Denaturing Buffer: 20 mM HEPES-NaOH pH 7.5, 1 M NaCl, 0.2 mM EDTA pH8, 1% SDS.

mRNP Lysis Buffer: 50 mM HEPES-NaOH pH7.5, 100 mM NaCl, 1 mM EDTA pH8, 1 mM DTT, 0.5% Na-deoxycholate, 0.5% Igepal Ca-630/NP-40, 10% glycerol, protease inhibitors [SigmaFAST, Sigma].

1X Denaturing Buffer: 2X binding buffer and mRNP lysis buffer mixed at a 1:1 volume ratio.

mRNP Elution Buffer: 10 mM Tris pH7.5, 1 mM EDTA pH8, 50 µg/ml RNase A.

Protein IP Buffers

IP Lysis Buffer: 50 mM HEPES-NaOH pH7.5, 100 mM NaCl, 0.5% Triton X-100, 1

mM EDTA pH 8.0, 10% Glycerol, 1 mM DTT, protease inhibitors [SigmaFAST, Sigma].

High Salt IP Lysis Buffer: 50mM HEPES-NaOH pH 7.5, 1 M NaCl, 0.5% Triton X-100, 1 mM EDTA pH 8.0, 10% Glycerol, 1 mM DTT.

Arginine Elution Buffer: 1 M Arginine-HCl pH 3.5.

DNA buffers

6X DNA loading buffer: 0.25% Bromophenol blue, 0.25% Xylene cyanol, 30% glycerol.

5X TBE: 4.4 M Tris, 4.4 M Boric Acid, 0.1 M EDTA pH8.

Chromatin-IP Buffers

ChIP Lysis Buffer 1: 50 mM HEPES-NaOH pH7.5, 140 mM NaCl, 1 mM EDTA, 0.5% NP-40, 0.25% Triton X-100, 10% glycerol, protease inhibitors [SigmaFAST, Sigma].

ChIP Buffer 2: 10 mM Tris-HCl pH7.3, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, protease inhibitors [SigmaFAST, Sigma].

ChIP Lysis Buffer 3: 10 mM Tris-HCl pH7.3, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.5% N-lauroylsarcosine, 0.1% Na-deoxycholate. protease inhibitors [SigmaFAST, Sigma].

ChIP RIPA Wash Buffer: 50 mM HEPES-NaOH pH7.5, 500 mM LiCl, 1% NP40, 1 mM EDTA, 0.1% N-lauroylsarcosine, 0.7% Na-deoxycholate.

ChIP Final Wash Buffer: 10 mM Tris-HCl pH7.3, 50 mM NaCl, 1 mM EDTA.

ChIP Elution Buffer: 50 mM Tris-HCl pH8.0, 10 mM EDTA, 1% SDS.

RNA-IP Buffers

RIP Lysis Buffer: 50 mM HEPES-HCl pH7.5, 150 mM NaCl, 0.1% SDS, 1% NP-40, 0.5% sodium deoxycholate, 10% glycerol, protease inhibitors [SigmaFAST, Sigma].

RIP High Salt Wash Buffer: 50 mM HEPES-HCl pH7.5, 1 M NaCl, 0.1% SDS, 1% NP-40, 0.5% sodium deoxycholate, 10% glycerol.

3X Reverse Crosslinking Buffer: 3X PBS, 6% N-lauroyl sarcosine, 30 mM EDTA pH8, 15 mM DTT.

Chromatin/Nucleoplasm/Cytoplasm Fractionation Buffers

Sucrose lysis buffer: 10 mM Tris-HCl pH8, 0.5 M sucrose, 2 mM $MgCl_2$, 3 mM $CaCl_2$, 10% glycerol, 0.5% Triton X-100, 1mM DTT, protease inhibitors [SigmaFAST, Sigma].

NRB buffer: 20 mM HEPES pH7.5, 75 mM NaCl, 1 mM DTT, 50% Glycerol, protease inhibitors [SigmaFAST, Sigma].

NUN buffer: 20 mM HEPES pH7.5, 300 mM NaCl, 1 M Urea, 10 mM $MgCl_2$, 1% NP-40, 1 mM DTT.

Buffer A: 10 mM HEPES pH7.5, 10 mM KCl, 4 mM $MgCl_2$, 10% glycerol, 1 mM DTT, protease inhibitors [SigmaFAST, Sigma].

Fractionation RIPA buffer: 50 mM HEPES pH7.5, 15 mM NaCl, 1% NP-40, 0.5% Na-deoxycholate, 0.1% SDS, 10% glycerol, 1 mM DTT, protease inhibitors [SigmaFAST, Sigma].

SDS PAGE/Western Blot Buffers

4X SDS-PAGE Loading Buffer: 200 mM Tris-HCl pH 6.8, 1% Bromophenol blue, 10% Sodium dodecyl sulphate (SDS), 50% Glycerol.

4X SDS-PAGE Stacking Gel Buffer: 0.5 M Tris HCl pH 6.8, 0.15% SDS.

4X SDS-PAGE Resolving Gel Buffer: 1.5 M Tris-HCl pH 8.8, 0.15% SDS.

SDS-PAGE Running Buffer: 25 mM Tris, 250 mM Glycine, 0.1% SDS.

Coomassie Brilliant Blue Stain: 0.1% Coomassie Brilliant Blue R-250, 40% Methanol, 10% Acetic Acid.

Destain Solution: 40% Methanol, 10% Acetic Acid.

BioRad TurboBlot Transfer Buffer: 200 ml 5X Commercial Stock combined with 600 ml H_2O and 200 ml ethanol.

10x TBS: 0.2 M Tris-HCl pH7.6, 1.37 M NaCl.

TBST: 1X TBS, 0.2% Tween-20.

5% Blocking Solution: 2.5 g Powdered Milk, 5 ml 10X TBS, 0.2% Tween-20.

ECL1: 100 mM Tris-HCl pH8.5, 400 µM p-coumaric acid, 2.5 mM Luminol.

ECL2: 100 mM Tris-HCl pH8.5, 5.3 mM Hydrogen Peroxide.

Immunostaining Buffers

FIX solution: 1X PBS, 4% Formaldehyde, 0.5% Triton X-100.

Immunostaining blocking solution: 1X PBS, 1% BSA.

GST Pull Down Buffers

GST Lysis buffer: 1X PBS, 0.1% Tween-20, 2 mM EDTA pH8, 10% glycerol, protease inhibitors [SigmaFAST, Sigma].

GST Wash Buffer: 1X PBS, 0.1% Tween-20

GST Elution Buffer: 50 mM Tris pH8, 40 mM GSH, 200 mM NaCl, 10% glycerol.

In Vitro Kinase Assay Buffers

Kinase Assay Buffer – 50 mM Tris-HCl (pH7.5), 10 mM $MgCl_2$, 0.1 mM ATP, 1 µCi [$\gamma_{32}P$]-ATP.

2.1.5 MOLECULAR BIOLOGY KITS

Small scale plasmid DNA extraction and purification: Qiagen Mini Spin Preparation kit.

Medium scale plasmid DNA extraction and purification: Qiagen Midi Spin Preparation kit.

DNA extraction from agarose gels: Qiagen Gel Extraction kit.

Molecular Cloning via the Gibson Assembly method: The Gibson Assembly Master Mix Kit (New England Biolabs).

2.1.6 ANTIBODIES

ANTIBODY	SOURCE	SPECIES	CLONALITY
HNRNPUL1	In house	Rabbit	polyclonal
TUBULIN	Sigma-Aldrich T5168 - Clone B-5-1-2	Mouse	monoclonal
FLAG	Sigma-Aldrich F3165 – Clone M2	Mouse	monoclonal
UAP56	In house	Rabbit	polyclonal
COILIN	Bethyl Laboratories A303-760A	Rabbit	polyclonal
SMN	Abcam [2B1] (ab5831)	Mouse	monoclonal
SART3	Bethyl Laboratories A301-521A	Rabbit	polyclonal
PRP31	Abcam [EPR14587] (ab188577)	Rabbit	monoclonal
U1C	Sigma-Aldrich Clone 4H12	Rat	monoclonal
U1A	Abcam (ab155054)	Rabbit	polyclonal
SF3B3	Bethyl Laboratories A302-508A	Rabbit	polyclonal
HNRNPU	Abcam (ab10297) – Clone 3G6	Mouse	monoclonal
FUS	Novus Biologicals (NB100-565)	Rabbit	polyclonal

TAF15	Abcam (ab134916)	Rabbit	monoclonal
EWSR1	Bethyl Laboratories A300-417A	Rabbit	polyclonal
Total Pol II	MBL Life science MABI0601	Mouse	monoclonal
Unphosphorylated Pol II	Abcam [8WG16] (ab817)	Mouse	monoclonal
Pol II Ser2P	Abcam (ab5095)	Rabbit	polyclonal
Pol II Ser5P	Abcam (ab5131)	Rabbit	polyclonal
Pol II Ser7P	Active Motif – Clone 4E12	Rat	monoclonal
SSRP1	Biolegend - Clone 10D1	Mouse	monoclonal
HISTONE H3	Abcam (ab1791)	Rabbit	polyclonal
ARS2	Abcam (ab192999)	Rabbit	polyclonal
MTR4	Abcam (ab187884)	Rabbit	polyclonal
ZCCHC8	Abcam (ab68739)	Rabbit	polyclonal
LARP7	Bethyl Laboratories A303-723A	Rabbit	polyclonal
MEPCE	Proteintech (14917-1-AP)	Rabbit	polyclonal
HEXIM	Bethyl Laboratories A303-113A	Rabbit	polyclonal
CDK9	Abcam (ab76320)	Rabbit	polyclonal
RBM7	Atlas (HPA013993)	Rabbit	polyclonal
TBP	Abcam (ab818)	Mouse	monoclonal
α-RABBIT	Promega W4011, HRP Conjugate	Goat	polyclonal
α-MOUSE	Promega W4021, HRP Conjugate	Goat	polyclonal
α-RAT	ThermoFisher 62-9520, HRP Conjugate	Goat	polyclonal

2.1.7 PRIMERS

PRIMER	SEQUENCE	DESCRIPTION
5' rev pcDNA5-FRT-3XFLAG	tgcggccgcaagcttgtcatc	Used to generate linearised pcDNA5-FRT-3XFLAG vector
3' fwd pcDNA5-FRT-3XFLAG	ctcgagtctagagggcccgtttaaacc	Used to generate linearised pcDNA5-FRT-3XFLAG vector
HNRNPUL1 FL Fwd	gatgacaagcttgcggccgca ATGGATGTGCGCCGTCTG	Used to clone full length HNRNPUL1 into pcDNA5-FRT-3XFLAG vector
HNRNPUL1 FL Rev	acgggccctctagactcgag CTACTGTGTACTTGTGCCACCC	Used to clone full length HNRNPUL1 into pcDNA5-FRT-3XFLAG vector
HNRNPUL1 ΔCTD Rev	acgggccctctagactcgag *TTA*GTCAAAGCGCTTTTCAGGG	Used to clone HNRNPUL1 ΔCTD into pcDNA5-FRT-3XFLAG vector
pUC18 FWD	TCGCGCGTTTCGGTGATG	Used to generate linearised pUC18 vector
pUC18 REV	GACGAAAGGGCCTCGTGATAC	Used to generate linearised pUC18 vector
HNRNPUL1 Homology Arms Fwd	CGAGGCCCTTTCGTCG ACAGCCTGCTGGCCTCG	Used to clone HNRNPUL1 homology arms in pUC18
HNRNPUL1 Homology Arms Rev	CACCGAAACGCGCGAAG GGTCTGGGAAGCTGAGG	Used to clone HNRNPUL1 homology arms in pUC18
pUC18/Hom Arms Fwd	CCAGTGTGACCCAGAGGC	Used to generate linearised pUC18/Hom Arms vector
pUC18/Hom Arms Rev	CTGTGTACTTGTGCCACCCTG	Used to generate linearised pUC18/Hom Arms vector

AID Fwd	AGGGTGGCACAAGTACACAGATGA TGGGTAGTGTGGAGCTGAAC	Used to clone AID-HYG and AID-NEO into pUC18/Hom Arms vector
AID Rev	GAGCCTCTGGGTCACACTGGTAAGA TACATTGATGAGTTTGGACAAACCA CAACTAGAATGCAGTGAAAAAA ATGCTTTATTTG	Used to clone AID-HYG and AID-NEO into pUC18/Hom Arms vector

2.1.8 qRT-PCR PRIMERS

Transcript/ Locus	FORWARD	REVERSE
18S	GTGGAGCGATTTGTCTGGTT	CGGACATCTAAGGGCATCAC
U1 snRNA	ACCTGGCAGGGGAGATACCA	GGGGAAAGCGCGAACGCAGT
EGR1	GAAGAACTTGGACATGGCTGT TTC	CCTCCCTCTCTACTGGAGTGG AA
DCTPP1	AGACCCATTCTCGTGTTCC	CATCTGGACTCTTACAGCCTT C
MT-ND1	TCGCCCTATTCTTCATAGCC	GAGTTGGTCGTAGCGGAATC
MT-ND2	ATCATCCCCACCATCATAGC	GTAGGAGTAGCGTGGCAAGG
MIR17HG	AACTCAAACCCCTTTCTACACA	ATCCCCACCAAACTCAACAG
EGR1 TSS (ChIP)	CAGAACTTGGGGAGCCGC	GGGGAACACTGAGAAGCGT
EGR1 +400b (ChIP)	CTGGAGGAGATGATGCTGCT	CCTGAGGGTTGAAGGTGCT
EGR1 +800b (ChIP)	TTGGATGGAGAGCTCTGGAG	AGGATCACGGTCCTTCCTTG
EGR1 +1.4kb (ChIP)	TTACCCCAGCCAAACCACTC	ACTGACCAAGCTGAAGAGGG
FOS TSS (ChIP)	TCGTACTCCAACCGCATCTG	AGAACATCATCGTGGCGGTTA

FOS +400b (ChIP)	CAGGTAAGGCTGGCTTCCC	CTTACTATGGCAAGCGTGCG
FOS +750b (ChIP)	CGTTCTGAGCAACCTCTGGT	GAAACTGCCTTACACACCCG
FOS +1.5kb (ChIP)	GCCCATTCCATCCCAACTCA	CTGTGAAACCATTTCTGACCTGC
FOSB	CGATGGTCTCTCTCCCTCTG	AAAATCTCATGTCCCCAACG
FOS	CCAACCTGCTGAAGGAGAAG	AGATCAAGGGAAGCCACAGA
MT-ATP8	TGCCCCAACTAAATACTACCG	GCAATGAATGAAGCGAACAG
ATF5	GAGTGGCGACAGGATAGAGC	TTTAGCCTCCCTCCCTTAGC
U2 snRNA	GATTTTTGGAGCAGGGAGATGG	TACTGCAATACCAGGTCGATGC
U4 snRNA	TATCCGAGGCGCGATTATTG	CCAGTGCCGACTATATTGCAAG
U5 snRNA	TCTGGTTTCTCTTCAGATCGCA	GAGTTGTTCCTCTCCACGGA
U6 snRNA	TGCTCGCTTCGGCAGCACAT	AATATGGAACGCTTCACGAA
7SK snRNA	ACGACCATCCCCGATAGA	CCTCATTTGGATGTGTCTGG
SNORD16	TGCAATGATGTCGTAATTTGCG	GCTCAGTAAGAATTTTCGTCAACC
pre-SNORD16	TGCCTGCTGTCAGTAAGCTG	GGCCTCCACGACACATCTAT
RPL4	TGTGCTCGCCCACTGATATC	TTTGCGCAAGTTGGTGTGAA
SNORD10	ATGCGTGTCATCTGAGCCTC	ACTGATCCTTTGCCCAGGAC
pre-SNORD10	CGCTTTCCAGTCTTTCAGCG	CAACAGCCCTGGGAAGTAGG
EIF4A1	TTGGCTCAGCAGATACAGAA	GGGACAGGTATCTCCGGTTA
U3 snoRNA	CGTGTAGAGCACCGAAAACC	ACTCAGACCGCGTTCTCTC
scaRNA2	CCGCCTCGTCTATCTGATCA	CGGCCTCGTCTATCTGATCA
SNORA9B	TGCTTGGGTCTGCAGTGA	TCTACATGGAGCACGGCAA
ACT B (ChIP)	TATTCTCGCAGCTCACCAT	TCCTGTGCAGAGAAAGCG
U1 snRNA uncleaved	TACCTGGCAGGGGAGATACC	GCGTACGGTCTGTTTTTGAAACTC
U2 snRNA uncleaved	GCAGGTGCTACCGTCTCTCAC	ACGTCCTCTATCCGAGGACAATA

U4 snRNA uncleaved	CGTAGCCAATGAGGTCTATCCG	CCTCTGTTGTTCAACTGCAAGAAA
HNRNPA2B1	GGTGGCTTAAGCTTTGAAACCA	CATGGCAGCATCAACCTCAG
TAF15	GAGGGGCTACAGAGGTCGT	CCCACTTCTATCTCCGCTGT
NEAT1_2	GGCCAGAGCTTTGTTGCTTC	GGTGCGGGCACTTACTTACT
pre-SP4	CCTGTTGTCGTGTGTGTGTG	AGCCTGAGGTTTTGGGTTTT
ATF4	TCCTGTCCTCCACTCCAGAT	AGGGATCATGGCAACGTAAG
GAPDH	GAAGGTGAAGGTCGGAGTC	GAAGATGGTGATGGGATTTC
NR4A1	CTGCCTGTCACGTCTGTTG	CTTGTCAATGATGGGTGGAG

2.2 METHODS

2.2.1 MOLECULAR BIOLOGY

Polymerase Chain Reaction (PCR) – PCR reaction mixtures consisted of 50 ng template DNA, 1X reaction buffer, 0.5 mM dNTPs, 0.2-1 µM forward and reverse primers, and 2.5-5 U DNA polymerase depending on the manufacturer's requirements. Reactions were cycled 35 times. Annealing temperatures ranged from 58-65°C depending on the primers. An extension temperature of 72°C was used, with the extension time dependent on the polymerase and the length of the amplicon.

Colony PCR – To screen *E.coli* transformants following cloning/ligation attempts, a P20 pipette tip was used to pick colonies and shake them into a PCR tube. PCR would then be performed as described above with an extended denaturation step of 10 minutes at 98°C.

Site-Directed Mutagenesis – Site-directed mutagenesis PCRs were performed using

Pfu Turbo Polymerase (Stratagene). 50 µl reaction mixtures consisted of 50 ng template plasmid, 125 ng sense primer, 125 ng antisense primer, 0.2 mM dNTPs, 1X Pfu Turbo buffer and 2.5 U Pfu Turbo. Primer length was dictated by the number of codons that were being mutated – 1 codon = 18 bp on either side, 2 codons = 21 bp on either side, and 3 codons = 24 bp on either side. Reactions were started with a 95 °C step for 30 seconds, followed by 26 cycles of: 95 °C – 30 seconds, 55 °C – 1 minute, 68 °C – 15 minutes. This was followed by a final 10 minute incubation at 68 °C. 10 U of Dpn1 (Roche) was then added to reaction mixtures to digest template DNA followed by a 1 hour incubation at 37 °C. 10 µl of the reaction was used to transform *E.coli* following the protocol outlined in the *'E.coli* transformations' section below.

E.coli transformations – Competent *E.coli* cells were thawed on ice, followed by the addition of approximately 20-100 ng DNA. This mixture was then kept on ice for 20 minutes, before heat shock at 42 °C for 30 seconds, and then returned to ice for a further 2 minutes. Next, 900 µl LB medium was added to cells followed by incubation at 37 °C for 1 hour whilst rotating at 200 rpm. 1/10 of these cells were then plated onto selective medium and incubated at 37 °C overnight.

Plasmid DNA isolation from *E.coli* – For small scale plasmid isolation, *E.coli* was grown in 5 ml LB cultures overnight at 37 °C whilst rotating at 200 rpm, whereas for larger scale preps, 50 ml LB cultures were used. The following morning, QIAGEN mini or midi-prep kits were used in accordance with the instructions specified by the manufacturers.

Agarose Gel Electrophoresis of DNA – Agarose gels were produced by dissolving 0.5-2% agarose (depending on the size of the DNA fragment intended to be visualised) in 0.5X TBE, followed by the addition of ethidium bromide (BIO-RAD) to a final

concentration of 10 µg/ml. Gels were run in BIO-RAD electrophoresis tanks at approximately 100 V, with 0.5X TBE used as running buffer. DNA bands were visualized via UV light exposure using a BIO-RAD Chemidoc transilluminator.

Gel Extraction – DNA bands were extracted and purified from DNA agarose gels using a Qiagen Gel Extraction Kit according to the manufacturer's instructions.

Phenol:chloroform DNA purification – DNA samples were made up to 100 µl with H_2O, before the addition of an equal volume of phenol:chloroform pH 6.7. Next, samples were vortexed for 1 minute followed by centrifugation at 12000 x g for 5 minutes at 4 °C. The upper phase would then be transferred to a tube containing 10 µl 3M Sodium acetate (pH 5.3), 300 µl 100% ethanol and 5 µg glycogen, and incubated at -20 °C for 30 minutes. Samples were then centrifuged for 20 minutes at 12000 x g, before pellets were washed once with 70 % ethanol. This was followed by air drying the pellets, and then resuspension in the required volume of H_2O.

Molecular Cloning – The Gibson Assembly Master Mix Kit (New England Biolabs) was used when cloning DNA fragments into all vectors used in this study, according to the manufacturer's instructions. Primers used to amplify vectors and inserts are listed in the primers table in the Materials section. Transformations were carried out as described above in the '*E.coli transformations*' section.

2.2.2 PROTEIN EXPRESSION AND BIOCHEMISTRY

Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis of proteins (SDS-PAGE) - SDS-PAGE gels were run in BIO-RAD Electrophoresis Chambers at 30 mA for approximately 1 hour. Recipes for stacking gels and resolving gels are listed below. A range of 8-12% acrylamide was used depending on the size of the protein interest.

5% Stacking gel:

COMPONENT	VOLUME
H_2O	6.3 ml
30% Acrylamide/0.8% Bisacrylamide	1.2 ml
Tris pH 6.8	2.5 ml
10% APS	110 µl
TEMED	20 µl

10% Resolving gel:

COMPONENT	VOLUME
H_2O	4.1 ml
30% Acrylamide/0.8% Bisacrylamide	3.33 ml
Tris pH 8.8	2.5 ml
10% APS	110 µl
TEMED	20 µl

Western Blots – SDS-PAGE gels were transferred to nitrocellulose membranes via a BIO-RAD fast transfer machine, programmed at 25 volts for 15 minutes. Membranes were then incubated in 5% blocking solution for 1 hour with shaking. This was followed by the addition of the primary antibody in 5% blocking solution. Incubation periods with the primary antibody ranged from 1-3 hours depending on the antibody. Membranes were then washed 3X in TBST for 30 seconds, followed by three more TBST washes for 10 minutes each. Membranes were incubated with the appropriate secondary

antibody in 5% blocking solution for 30 minutes. Next, the membranes were washed 3 x 30 seconds with TBST and 3 x 10 minutes with TBST. Protein bands would then be visualised using Enhanced ChemiLuminescence (ECL). This involved mixing ECL1 and ECL2 solutions at a 1:1 v/v ratio and adding them to the blot. Following a one minute incubation, membranes would then be exposed using a BIO-RAD Chemidoc system.

Endogenous protein immunoprecipitation (IP) – Four 15cm dishes were seeded per IP condition with 5000000 cells/dish. 100 μl protein-G Dynabeads per condition were washed twice in IP lysis buffer, before being resuspended in 400 μl IP lysis buffer + 1% BSA (from 20% BSA solution stock, 0.2 μm filtered) along with 4-10 μg of the relevant antibody. This was followed by rotation at room temperature for 1 hour. After blocking, beads were then washed 3 times in IP lysis buffer prior to the addition of sample lysate. Cells were washed once with 1X PBS and lysed in IP lysis buffer. This was followed by syringing samples up and down 5 times through a 0.6 mm x 25 mm needle to sheer cell nuclei in the extracts. Samples were cleared by centrifugation (13.2k rpm, 5 minutes, 4 °C), and lysate concentrations were measured by Bradford assay. A small aliquot would be then isolated and stored as an input sample. Equal amounts of protein extract per condition were then added to the beads, and volumes matched via the addition of IP lysis buffer to the most concentrated samples. IPs were carried out overnight at 4 °C with rotation. The beads were then washed 3 times with IP lysis buffer. This was followed by the addition of Arginine Elution buffer, with samples incubated on ice for two minutes to elute the immunoprecipitated complexes. Eluates were neutralized via the addition 1 M Tris-HCl pH8.8 at a ratio of Arginine elution buffer to Tris-HCl pH 8.8 of 25:1 v/v. Inputs and eluates were then analysed via SDS-PAGE/Western Blot.

FLAG-tagged Protein Immunoprecipitation (IP) – Tetracycline was added to the stable

cell lines to induce expression of FLAG-tagged proteins for 48 hours prior to the IP. 50 µl FLAG-agarose beads (Sigma-Aldrich) per condition were washed twice in IP lysis buffer, followed by blocking in IP lysis buffer + 1% BSA for 2 hours at 4 °C. Beads were then washed twice more in IP lysis buffer prior to the addition of sample lysates. Processing of cells and IP conditions were performed as described in the 'Endogenous protein immunoprecipitation' section above. After the final wash, a gel loading tip was used to remove all liquid from beads, and 60 µl Flag peptide solution (10 µl 2mM Flag peptide (sigma) + 190 µl IP lysis buffer) was added to each sample. Immunoprecipitated complexes were eluted for 1 hour at 4 °C whilst rotating. Inputs and eluates were then analysed via SDS-PAGE/Western Blot.

High Salt Protein Immunoprecipitation (IP) – High salt protein immunprecipitations were performed as described in the 'FLAG-tagged Protein Immunoprecipitation' section up until the washing of beads following their incubation with the lysates. In this protocol beads were washed twice with IP lysis buffer, then twice with High Salt IP Lysis Buffer, and finally twice more with IP lysis buffer. Elution steps were then carried out as described in the 'FLAG-tagged Protein Immunoprecipitation' section above.

FLAG-tagged Protein Immunoprecipitation (IP) for Mass Spectrometry analysis – For IPs performed for mass spectrometry analysis, 10 x 15 cm dishes were seeded per condition. IPs were then carried out as described in the 'FLAG-tagged Protein Immunoprecipitation' section up until the washing of beads following their incubation with the lysates. In this case, beads were washed twice with IP lysis buffer, followed by two washes with IP lysis buffer without glycerol and Triton X-100. Immunoprecipitated complexes were then eluted via the low pH Arginine Elution buffer method outlined in the 'Endogenous protein immunoprecipitation' section above. Mass spectrometry analysis was performed by Mark Dickman at the Department of Biological and Chemical Engineering, University of Sheffield, using nano-flow liquid

chromatography (U3000 RSLCnano, Thermo Scientific) coupled to a hybrid quadrupole-orbitrap mass spectrometer (Q Exactive HF, Thermo Scientific). Peptides were separated on an Easy-Spray C18 column (75 μm x 50 cm) using a 2-step gradient from 97% solvent A (0.1% formic acid in water) to 10% solvent B (0.1% formic acid in 80% acetronitrile) over 5 min then 10% to 50% B over 75 min at 300 nL min-1. The mass spectrometer was programmed for data dependent acquisition with 10 product ion scans (resolution 30,000, automatic gain control 1e5, maximum injection time 60 ms, isolation window 1.2 Th, normalised collision energy 27, intensity threshold 3.3e4) per full MS scan (resolution 120,000, automatic gain control 1e6, maximum injection time 60ms) with a 20 second exclusion time. Database searching MaxQuant (version 1.5.2.8) software was used for database searching with the *.raw MS data file using standard settings. The data was searched against the Homo sapiens Uniprot proteome database (taxa id: 9606, downloaded 25 November 2018, 73101 entries), using the following settings: Digestion type: trypsin; Variable modifications: Acetyl (Protein N-term); Oxidation (M); MS scan type: MS2; PSM FDR 0.01; Protein FDR 0.01; Site FDR 0.01; MS tolerance 0.2 Da; MS/MS tolerance 0.2 Da; min peptide length 7; max peptide length 4600; max mis-cleavages 2; min number of peptides 1.

Purification of GST-tagged proteins – BL21 *E.coli* cells were transformed with GST-tagged construct plasmids and plated on selective media overnight. Colonies were picked and grown in TB media until they reached an OD_{600} of 1. Protein expression was then induced via the addition of IPTG to a final concentration of 0.5mM. Cultures were grown for 3 hours at 37°C, before centrifugation at 4000 x g for 15 minutes. Bacterial pellets were resuspended in 1 ml GST lysis buffer per 0.4g of pellet, and lysed via sonication (5 x [30s-ON/30s-OFF]). Lysates were then cleared by centrifugation at 16100 x g for 15 minutes at 4 °C. 50 μl GSH beads (GE Healthcare) were washed 3X with GST wash buffer. Cleared lysates were then added to the washed beads and incubated for 1 hour at 4°C with rotation. This was followed by

pelleting the beads and three more washes with GST wash buffer. After the final wash, a gel loading tip was used to remove residual wash buffer from the beads, and 60 µl GST elution buffer was added. This was followed by incubation at 4 °C for 30 minutes with rotation. Eluted samples were then removed and analysed via SDS-PAGE and Western blotting.

In vitro **Kinase Assay** – Purified GST-CTD and FLAG-UL1/GFP were mixed in a 1:1 molar ratio in kinase assay buffer along with 100 ng CDK7, and left at room temperature for 5 minutes. Reactions were terminated via the addition of SDS loading buffer, and samples were then run on SDS-PAGE gels. Next, dried gels were exposed to a phosphor film. A Typhoon FLA 7000 laser scanner was used to expose films and analysis was performed using Quantity One software.

Chromatin Immunoprecipitation (ChIP) – Between one and four 15cm dishes (depending on the target protein) were seeded/ChIP condition with 5000000 cells/dish. Protein-DNA complexes were crosslinked through incubating cells with 20ml PBS-formaldehyde (1%). Cell pellets were lysed in ChIP Lysis Buffer 1 and rotated for 5 minutes at 4 °C. Nuclei were then pelleted via centrifugation (3000 x g, 5 minutes at 4 °C). This was followed by resuspension in ChIP Buffer 2 and rotation for 10 minutes at room temperature. Nuclei were pelleted via centrifugation (1500 x g, 5 minutes at 4 °C) and resuspended in ChIP Lysis Buffer 3. Sonication was performed using a Bioruptor (High, 20 x [30s-ON/30s-OFF]), generating chromatin fragments of 250-300 nucleotides. Samples were cleared by centrifugation (16100 x g, 15 minutes, 4 °C) and lysate concentrations were measured by Bradford assay. Equal concentrations of chromatin were incorporated into the IPs. IPs were carried out overnight at 4 °C using 5 µg of antibody. 100 µl blocked protein-G Dynabeads were then added to samples and incubated for 2 hrs at 4 °C. Following incubation, beads were washed 4 times with ChIP RIPA Wash Buffer and once with ChIP Final Wash Buffer. Complexes were

eluted by adding ChIP Elution buffer and incubated for 30 minutes at 65 °C. NaCl was added to a final concentration of 200mM and cross-links were reversed overnight at 65 °C. Samples were then treated with RNase A (0.2 mg/ml final) for 2 hrs at 37 °C, followed by proteinase K (0.2 mg/ml final) for 2 hrs at 55 °C. DNA was purified via phenol-chloroform extraction and ethanol precipitation, and then resuspended in H_2O.

mRNP capture – 100 µl oligo-d(T)25 Dynabeads (New England Bioscience) per condition were washed two times in 1X binding buffer, and then stored in 1X binding buffer while samples were being processed. One 15 cm dish was seeded per condition with 5000000 cells/dish. Cells were washed once with 1X PBS, before being crosslinked on ice with 300 mJ/cm_2 whilst covered in cold 1X PBS. A non-crosslinked control condition would also be processed during each experiment. Following cross-linking, cells were lysed in mRNP capture lysis buffer supplemented with Ribosafe RNase inhibitors (Bioline), or mRNP capture lysis buffer lacking RNase inhibitors and supplemented with RNase A in the case of the plus RNase A control condition. Samples were then cleared by centrifugation at 16100 x g for 15 minutes at 4 °C, and lysate concentrations measured by Bradford assay. 1-10% of the total sample would be isolated and kept as an input fraction. This was followed by the addition of pre-warmed 2X denaturing buffer to the remaining samples at a 1:1 v/v ratio, to denature the extracts. Equal amounts of total protein per condition were then added to the beads, and volumes were equalized through the topping up of the most concentrated samples with 1X denaturing buffer. Samples were then rotated at 25 °C for 1 hour, followed by three washes with 1X denaturing buffer. Once washed, RNP complexes were eluted via the addition of mRNP elution buffer with incubation at 25 °C for 30 minutes. Inputs and eluates were then analysed via SDS-PAGE/Western Blot.

2.2.3 MAMMALIAN CELL BIOLOGY

Mammalian Tissue Culture – Cell lines were grown at 37 °C with 5% CO_2, and

passaged twice a week. During each passage, cells were washed once in 1X PBS, and then detached from flasks using pre-warmed 0.25% Trypsin/EDTA solution. The addition of DMEM then deactivated the trypsin, before cells were distributed in appropriate amounts into flasks and dishes.

Generation of FLAG-tagged FlpIn-293 Stable Cell Lines – 1×10^5 FlpIn-293 cells were seeded into one 6cm dish in DMEM containing Tet free-FCS and no antibiotics, and incubated at 37 °C for 24 hours. Next, the cells were transfected with 3.6 μg FlpIn recombinase construct (pPGKFLPobpA) and 2.4 μg FRT vector, using the Turbofect protocol outlined in the 'DNA transfections' section below. Two days after transfecting, the cells were split into two 10 cm dishes. Selection medium comprising DMEM containing Tet free-FCS and two antibiotics - Blasticidin (15 μg/ml final) and Hygromycin (0.1 mg/ml final) – was applied to the cells 6 hours later. The selection medium was then renewed after 48 hours. This step was repeated a further 5 days later, facilitating the removal of any dead cells. Once colonies had formed, cloning disks soaked in trypsin were used to transfer individual colonies to one well each of a 24 well plate. These colonies were then expanded into multiple wells to allow screening via Western blotting.

DNA transfections – Cells were seeded 24 hours prior to transfection, in amounts necessary to produce 60-70% confluency upon transfection. When generating stable cell lines, Turbofect (Thermo Scientific) was used as the transfection reagent, in accordance with the manufacturer's instructions. For all other transfections, polyethylenine (PEI) was used. In these instances, the transfection protocol followed that of Turbofect transfections, but with Turbofect replaced by PEI used at a final concentration of 3.5 μg/μl.

Colony Formation Assay – HCT116 and HNRNPUL1-AID line cells were treated with Doxycycline (1 μg/ml) and Auxin (500 μM) for 48 hours and 24 hours respectively, ensuring HNRNPUL1 was depleted in the degron line upon seeding. A 6-well plate per condition was seeded at 200 cells/well. Each well was adjusted to 2 ml DMEM and cells were left to grow for 14 days at 37 °C. After 14 days, cells were washed with 1X PBS and stained with 1ml of crystal violet (0.5 % giemsa powder in methanol) / well for 5 minutes. This was followed by two washes with 1 ml deionized water. Wells were then left to dry for 30 minutes.

Immunostaining – Cells seeded onto coverslips were washed once with 1X PBS, and then incubated with FIX solution for 20 minutes at room temperature. This was followed by two more washes with 1X PBS. Cells were then incubated with immunostaining blocking solution for 1 hour. Next, the primary antibody/antibodies were added to the cells in immunostaining blocking solution, followed by an incubation period of 1 hour. Cells were then washed three times with 1X PBS, and the secondary fluorescent antibody/antibodies diluted in immunostaining blocking buffer were added to the cells. An incubation period of 30 minutes followed. Then, cells were washed 3X with PBS, and coverslips were transferred to slides with mounting medium containing DAPI (Vectashield). The coverslips were sealed on the slides by painting around their edges with nail varnish.

2.2.4 RNA BIOLOGY

Total RNA extraction – Total RNA was extracted using TRIzol reagent (Invitrogen) as per the manufacturer's instructions. RNA samples were resuspended in H_2O and subjected to DNase treatment via the addition of 4 units of TurboDNase (Thermofisher) plus TurboDNase buffer and RNase inhibitors, with incubation at 37 °C for 1 hour. RNA acidic phenol (pH 4.5) was added to samples at a 1:1 v/v ratio and left to incubate at room temperature for 5 minutes. RNA in the aqueous phase would then be

precipitated via the addition of 86 mM sodium acetate pH 5.8, 100% ethanol and glycogen along with storage at -20 °C for 2 hours. This was followed by washing the RNA pellets once in 75% ethanol and resuspension in H_2O.

RNA Immunoprecipitation (RIP) - One 6 cm dish was used per RIP, with cells at 80% confluency on the day of experiment. 100 µl protein-G Dynabeads per condition were washed twice in RIP lysis buffer, before being resuspended in 400 µl RIP lysis buffer + 1% BSA (from 20% BSA solution stock, 0.2 µm filtered) along with 4-10 µg of the relevant antibody. This was followed by rotation at room temperature for 1 hour. After blocking, beads were then washed 3 times in RIP lysis buffer prior to the addition of sample lysate. Protein-RNA complexes were crosslinked through incubation with 3 ml PBS-formaldehyde (0.1%) for 10 mins, followed by quenching via the addition of glycine (0.125 M final concentration). Cells were then washed 3X with PBS, and lysed in RIP lysis buffer supplemented with Ribosafe RNase Inhibitors (Bioline) and Turbo DNase (Ambion). Sonication was performed using a Bioruptor (High, 5 x [30s-ON/30s-OFF]), generating RNA fragments of 300-400 nucleotides. Samples were cleared by centrifugation (16100 x g, 15 minutes, 4 °C), and 10% of the total lysate was isolated to keep as an input fraction. The remaining lysates were then added to the beads and IPs were carried out whilst rotating for 2 hours at 4 °C. This was followed by washing the beads twice with RIP lysis buffer, then twice with RIP high salt wash buffer, and finally two more washes with RIP lysis buffer. Input and IP samples were then made up to 56 µl with $H_2O_{RNaseFree}$, followed by the addition of 33 µl 3X reverse crosslinking buffer, 10 µl proteinase K (Roche, 19mg/ml) and 1 µl Ribosafe RNase Inhibitors (Bioline). Samples were incubated for 1 hour at 42 °C and then 1 hour at 55 °C, both whilst shaking at 1100 rpm, to reverse the crosslinks and digest proteins in the samples. RNA was then isolated from the input and IP samples via TRIzol extraction (detailed above), and converted into cDNA prior to qRT-PCR analysis.

Chromatin/Nucleoplasm/Cytoplasm Fractionation – One 10 cm dish was used per condition. Cells were washed once in 1X PBS, then trypsinised and transferred to a 15 ml falcon. This was followed by 2 more washes with 1X PBS. Cell pellets were then lysed in 5-10X cell pellet volume of Sucrose Lysis Buffer, and centrifuged at 500 x g for 5 minutes at 4 °C. The resulting supernatant was removed and centrifuged at 16100 x g at 1 minute, with the supernatant generated following this spin stored as the cytoplasmic fraction. The pellet produced from the first 500 x g spin was washed twice with Sucrose Lysis Buffer, then once with PBS + 0.5 M sucrose, followed by a final wash with Sucrose Lysis Buffer. The pellet was then resuspended in 350 µl NRB buffer, followed by the addition of 350 µl of NUN buffer. This mixture was incubated on ice for 5 minutes with inversion every minute. This was followed by centrifugation at 1200 x g for 5 minutes at 4 °C. The resulting supernatant was removed and stored as the nucleoplasmic fraction. The pellet was resuspended in Buffer A, followed by centrifugation at 1200 x g for 5 minutes at 4 °C. The pellet would then be resuspended in Fractionation RIPA buffer plus 500 U Benzonase, and left at room temperature for 45 minutes. This was followed by centrifugation at 16100 x g for 10 minutes at 4 °C, with the resulting supernatant then removed and stored as the chromatin fraction.

Reverse transcription and quantitative Polymerase Chain Reaction (qRT- PCR) – Reverse transcription reactions were performed using the Bioscript kit (Bioline). 0.5 mM dNTPs and 200 ng poly(dN)$_6$ random primer would be added to 1 µg RNA and reaction mixtures were incubated for 5 minutes at 70°C, followed by 1 minute on ice. Next, 100 units of Bioscript Reverse Transcriptase enzyme along with 1X reaction buffer would be added. The following program conditions were used for the reverse transcription reaction: 25 °C for 10 minutes, 42 °C for 1 hour and 85 °C for 5 minutes. cDNA produced from these reactions was diluted 3-6X with $H_2O_{RNaseFree}$ prior to being used in qRT-PCR. qRT-PCR reaction mixtures consisted of 2 µl diluted cDNA, 1X SensiMix (Bioline), 500 nM primers (see table), topped up to 10 µl with $H_2O_{RNaseFree.}$

qRT-PCR was performed using a Rotor-gene 6000 (Qiagen) with the following program conditions: 95 °C for 10 minutes and then 45 cycles of 95 °C for 10 seconds, 59 °C for 15 seconds and 72 °C for 25 seconds.

Statistical Testing – All statistical analyses were performed using a two-tailed unpaired T-test, except during the Sleeping Beauty complementation assays where ANOVA was used. The P value generated from these analyses is displayed on graphs using the following asterisk marks:

P-value	Number of asterisks
$p > 0.05$	Not indicated on graphs
$0.01 < p < 0.05$	*
$0.001 < p < 0.01$	**
$p < 0.001$	***

CHAPTER 3 – THE GENERATION OF HNRNPUL1-AID AND FLAG-HNRNPUL1 STABLE CELL LINES

Despite being discovered over 20 years ago, our understanding of the role of HNRNPUL1 in RNA metabolism is still incomplete. It has been reported to interact with mRNA export adaptor NXF1 (Bachi et al., 2000), along with a range of other RNA-binding proteins such as FUS and TAF15 (Chi et al., 2018b). Many of these interacting partners perform functions at multiple stages of the gene expression pathway, including alternative splicing regulation and mRNA nucleo-cytoplasmic transport (Lagier-Tourenne et al., 2010). Despite these interactions, no direct role in mRNA splicing or export has yet been established, although previous work performed in the Wilson lab has demonstrated that overexpression of HNRNPUL1 results in the export of a poorly spliced transcript (Wilson, unpublished). The highly homologous protein HNRNPU has been implicated in a diverse range of RNA-related processes, including splicing regulation (Xiao et al., 2012), snRNP biogenesis (Xiao et al., 2012), and lncRNA nuclear localisation (Hasegawa et al., 2010). However, the extent of functional redundancy between these two proteins remains unknown.

Both HNRNPU and HNRNPUL1 possess a central nucleoside triphosphate (NTP)-binding motif that enables these proteins to bind nucleotides (Nozawa et al., 2017; Wilson, unpublished). The role that nucleotide binding plays with regards to the function of HNRNPUL1 in the cell has also yet to be determined. The C-terminal prion-like domain of HNRNPUL1 composed of an RGG box and a poly-proline region has been reported to facilitate its recruitment to sites of DNA damage (Hong et al., 2013). PrLDs found in other proteins (including those associated with ALS) form associations with a wide range of protein interactors, and enable these factors to assemble into higher order structures composed of RNA/protein complexes. Whether the PrLD of HNRNPUL1 also mediates the majority of its interactions with protein and RNA partners is another outstanding question.

Therefore, the aims of the work described in this chapter are to generate the tools that will enable me to further investigate the functions of HNRNPUL1, and specifically determine what roles the NTP and PrLD regions play with regards to those functions. With these intended goals, I firstly set out to generate a HNRNPUL1 conditional knockout cell line using the auxin-inducible degron system. In addition, I aimed to create stable cell lines expressing tetracycline-inducible FLAG-HNRNPUL1 wild-type, as well as two stable FLAG-HNRNPUL1 mutant lines – one that was unable to bind nucleotides: FLAG-HNRNPUL1 Walker A mutant (FLAG-HNRNPUL1 WA), and one that lacked the PrLD at its C-terminus: FLAG-HNRNPUL1 ΔCTD.

3.1 HNRNPUL1 RNAi does not induce complete knockdown of the protein

An HNRNPUL1 RNAi line was generated previously in the Wilson lab to assess the consequences of HNRNPUL1 depletion within the cell. However, the knockdown of HNRNPUL1 observed in this cell line is incomplete, and requires 48 hours to occur following the addition of tetracycline (Figure 3.1). The advent of CRISPR/Cas9 technology as a tool for genome editing has radically expanded the landscape of genetic alterations that molecular biologists can introduce into mammalian genomes. In recent years, this technology has been utilised to tag target proteins with a sequence known as an auxin-inducible degron (AID). Once tagged with an AID, the addition of indole-3-acetic acid (referred to as auxin from here on in) triggers the degradation of the target protein, mediated by plant protein TIR1 (Natsume et al., 2016; Nishimura et al., 2009). In plants, the presence of auxin triggers TIR1 to interact with the SCF complex and mediates ubiquitinylation of degron-containing proteins, followed by their rapid degradation by the proteasome (Ramos et al., 2007). This SCF complex is highly conserved between plants and humans, whereas the TIR1 protein needs to be introduced via genome editing in order to adapt this system for human cells. In contrast to RNAi technology, the AID system has been reported to facilitate depletion of target proteins to virtually undetectable levels as quickly as 20 minutes
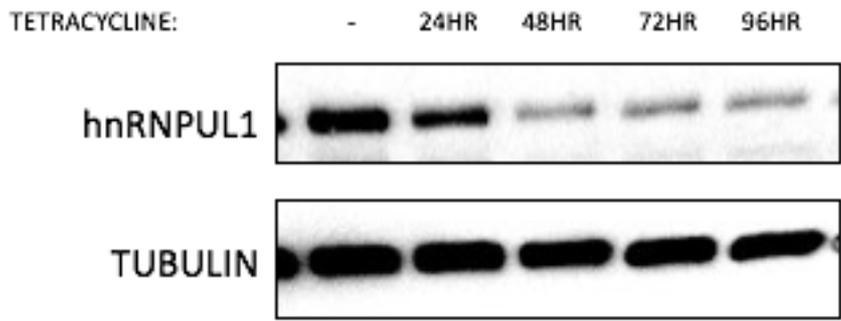
Figure 3.1 Time-course in HNRNPUL1 RNAi line demonstrating the extent of HNRNPUL1 knockdown

Western analyses of HNRNPUL1 levels in HNRNPUL1 RNAi line following treatment with tetracycline. Tubulin is shown as a loading control. Tetracycline was added to cells at a final concentration of 1 µg/ml and cells were harvested every 24 hours.

post auxin-addition (Natsume et al., 2016). Therefore, I set out to generate a HNRNPUL1-AID cell line, as in theory this method would enable the study of the immediate cellular consequences of the complete loss of HNRNPUL1 as directly as possible.

3.2 Tagging HNRNPUL1 with an AID

HCT116 cells were chosen as the parental line in which to undertake this experimental work as they are diploid in nature, unlike other commonly used tissue culture cell lines that often display polyploidy of varying severities, and therefore the number of HNRNPUL1 alleles needing to be tagged is limited to two. An HCT116 cell line already possessing a constitutively expressed TIR1 gene was donated as a gift from Steven West. TIR1 from *Oryza sativa* (rice) was incorporated as opposed to *Arabidopsis thaliana* TIR1 to optimise the functionality of the protein at the higher temperatures at which human cells are grown (Nishimura et al. 2009).

To tag the AID onto the end of HNRNPUL1, a plasmid containing the AID sequence flanked by regions homologous to the end of the HNRNPUL1 coding region (homology arms) must be transfected into the cell along with the Cas9 expression cassette. This ensures that when the Cas9 enzyme catalyses a double-stranded break (DSB) at the end of the HNRNPUL1 coding region, the cell favours homologous recombination as a means to repair this DSB and uses these homology arms as a repair template. This can result in the incorporation of the homology arms along with the AID into the genome at the site of Cas9-induced cleavage. To direct the Cas9 enzyme to cleave at the appropriate region - in this case the end of the HNRNPUL1 gene - a guide RNA that is complementary to a sequence of DNA within this region must be cloned into the Cas9-containing vector. The efficiency at which the guide RNA directs the enzyme to the appropriate locus can be tested via a Surveyor assay. Firstly, three guide RNAs were designed using the online Benchling tool (Benchling 2018) and cloned into the Cas9 expression vector (Ann Ran et al., 2013). These plasmids were then transfected

into HCT116 cells, as well as a positive control guide RNA targeting CPSF73 that was donated as a gift by Steven West. If the guide RNAs operated effectively, the Cas9 enzyme would be expected to cleave at the HNRNPUL1 gene and in the absence of a repair template the cell will repair this DSB via non-homologous end joining. This is an error-prone process, often resulting in an insertion or deletion of bases (indel). As a consequence, PCR amplification of the HNRNPUL1 coding region from genomic DNA isolated from these cells will be made up of a heterozygous population of molecules, with some containing indels and some maintaining the wild type sequence. Denaturing and reannealing these DNA molecules results in the annealing of some indel-containing and non-indel-containing DNA strands. These mismatches are recognised and cleaved by the Surveyor nuclease S, the products of which can be visualised on an agarose gel. Therefore, PCR primers were designed to amplify the region of the HNRNPUL1 gene that surrounded the three guide RNA target sites, along with primers to amplify the corresponding region of the CPSF73 gene (Figure 3.2A). These PCR amplified regions were then denatured, reannealed and incubated with the Surveyor nuclease. Although Surveyor nuclease digestion was not apparent in the case of the CSPF73 positive control (Figure 3.2B), smaller DNA fragments were clearly visible in the HNRNPUL1 guide RNA B lane of the agarose gel, indicating significant nuclease digestion (Figure 3.2B). As these bands appeared less abundant in the case of guide RNAs A and C, guide RNA B was chosen to proceed with when attempting to generate the HNRNPUL1-AID cell line, as this suggests that it is this guide RNA that is most effective at targeting Cas9 to the HNRNPUL1 locus.

Two vectors containing the HNRNPUL1 homology arms surrounding the AID cassette were generated, each containing a different antibiotic resistance gene – either hygromycin or neomycin – so that cells in which both alleles were tagged with the degron could be selected for. The degron sequence was composed of three copies of a small region of the IAA17 degron (3X mini-AID) (Natsume et al., 2016; Nishimura et al., 2009). This 3X mini-AID was separated by the antibiotic resistance gene sequence by a self-cleaving P2A cleavage site (Liu et al., 2017; Wang et al., 2015).

A.

HNRNPUL1:

HNRNPUL1 gene

5'
3'
guide RNA target site: A B C
Surveyor PCR primers:

CPSF73:

CPSF73 gene

5'
3'
guide RNA target site: +
Surveyor PCR primers:

B.

CPSF73 gRNA   HNRNPUL1 gRNAs

-    +   A   B   C

Products of cleaved mismatches
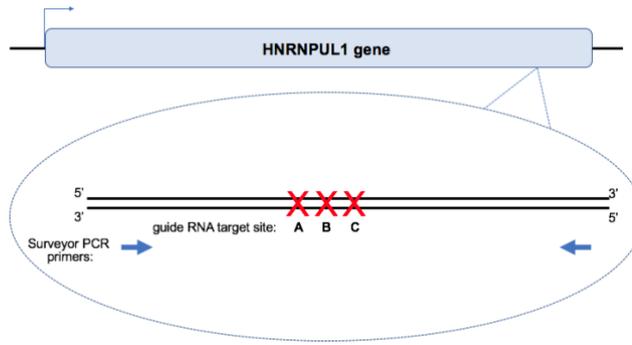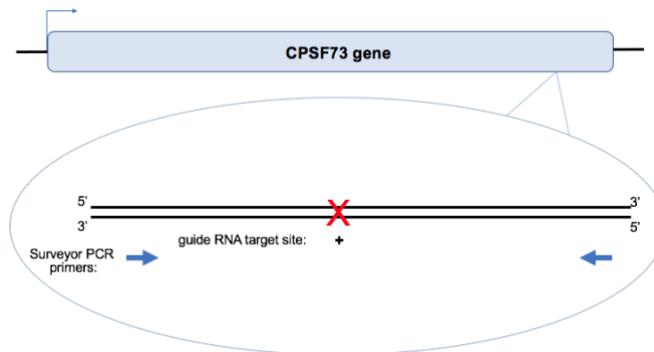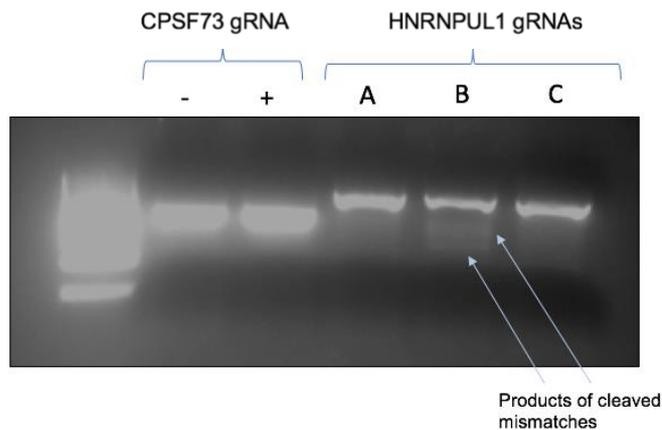
Figure 3.2 Guide RNA B is most efficient at directing Cas9 to HNRNPUL1 locus to generate DSBs, and was therefore chosen to guide Cas9 expression cassette

A - Schematic illustration of the location of the SURVEYOR PCR primers used in relation to the guide RNA target sites within the HNRNPUL1 and CPSF73 genes. B - DNA gel electrophoresis of SURVEYOR assay products using the IDT Surveyor mutation detection kit. An untransfected (-) sample and functioning CPSF73 gRNA transfected (+) sample were used as negative and positive controls respectively. A, B and C represent the three HNRNPUL1 guide RNAs tested.

The P2A cleavage site enables the antibiotic resistance protein to function separately from the HNRNPUL1-AID protein following their expression as a single polypeptide. A flexible glycine-rich linker sequence was located upstream of the degron to separate it from the C-terminal domain of HNRNPUL1, along with a 3X-FLAG tag to aid identification of successful clones during the screening process. It is also important to note that silent mutations were introduced into the homology arm sequence at the guide RNA region, so that the Cas9 enzyme did not target and cleave this sequence within the vector itself (Figure 3.3A).

These vectors, along with the Cas9/gRNA B-expressing vector, were then transfected into the HCT116/TIR1 cell line, and cells that incorporated the AID into both HNRNPUL1 alleles of their genome were selected via the addition of hygromycin and neomycin to the growth media. Colonies were then screened plus and minus auxin via Western analyses. Although it was expected that auxin may facilitate depletion of the tagged protein within 20 minutes, cells were incubated with auxin for 24 hours during the screening process to maximise the contrast between the plus and minus auxin conditions. Two of the three clones screened – clones 1 and 2 - appeared to have successfully incorporated the degron, as evidenced by the higher HNRNPUL1 band present on the Western blot in the minus auxin condition lanes, which disappears following the addition of auxin (Figure 3.3B). The same clones were also probed using an anti-FLAG antibody, which revealed that those higher bands were indeed hnRNPUL1-AID (Figure 3.3C). However, the expression levels of HNRNPUL1-AID appear significantly weaker than endogenous HNRNPUL1, even in the absence of auxin. In addition, there is a band the size of endogenous HNRNPUL1 on the anti-FLAG blot, suggesting that the mini-AID tag may be cleaved from the HNRNPUL1 protein by endogenous proteases. Combined, the result resembles a constitutive HNRNPUL1 knock-out cell line, as opposed to the inducible knock-out cell line that was intended. Therefore, alterations in degron sequence and parental cell line background were introduced with the goal of increasing the expression levels of steady state tagged HNRNPUL1. This process is described in the next section.

A.
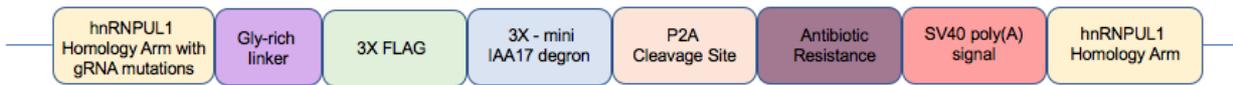
B.
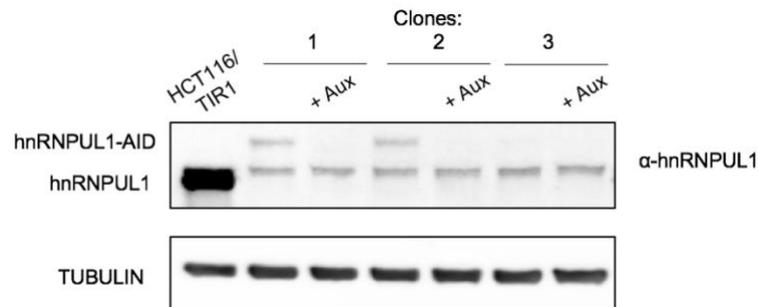
C.
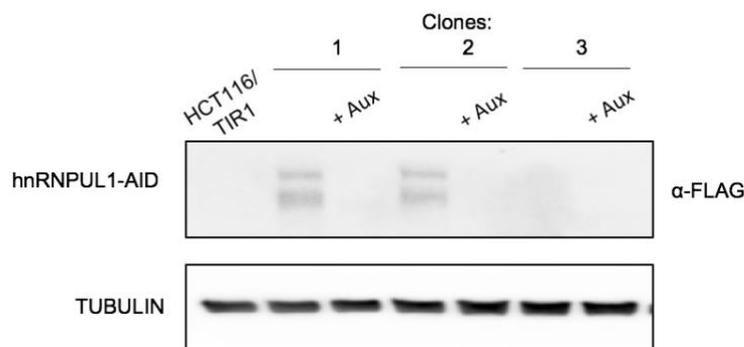
Figure 3.3 Tagging HNRNPUL1 with a mini-AID in the presence of constitutively expressed TIR1 causes severe degradation prior to the addition of auxin

A – Schematic diagram of the AID plus homology arms sequence that will act as a repair template following Cas9 cleavage. B – α-hnRNPUL1 Western Blot analyses of selected clones plus and minus auxin treatment. Tubulin is shown as a loading control. B - α-FLAG Western Blot analyses of selected clones plus and minus auxin treatment. Tubulin is shown as a loading control.

3.3 Creating a HNRNPUL1-AID line in a Dox-osTIR1 background

Two strategies were employed to improve HNRNPUL1-AID expression levels in the absence of auxin. Firstly, changes to the AID itself were introduced. The glycine-rich linker region was removed, and the 3X mini-AID was replaced by one copy of the full length IAA17 degron followed by a C-terminal FLAG-tag (Figure 3.4A). Secondly, rather than use an HCT116 cell line expressing constitutive osTIR1, HNRNPUL1 would firstly be tagged with the AID in standard HCT116 cells, and then a doxycycline-inducible osTIR1 (Natsume et al., 2016) would be incorporated into this cell line subsequently.

Thus, the HNRNPUL1 repair template vectors containing the new AID along with the Cas9/gRNA B-expressing vector were transfected into HCT116 cells and once again cells with the AID integrated into both alleles were selected for via the addition of hygromycin and neomycin. Clones were then screened via Western analyses. In this instance, three out of four clones successfully integrated the degron (Figure 3.4B). In addition, the expression levels of this form of HNRNPUL1-AID were significantly closer to endogenous HNRNPUL1. Clone 3 appeared to have the highest expression levels of HNRNPUL1-AID, so this clone was chosen to be expanded and subsequently used as a parental line for TIR1 integration.

Expression of osTIR1 in the HNRNPUL1-AID line could be achieved through introducing an osTIR1 expression vector at the safe-harbour locus AAVS1 (Smith et al., 2008) via CRISPR/Cas9. The osTIR1 expression vector used contained the TIR1 gene under the control of a conditional Tet promoter, and a puromycin resistance gene to enable selection of clones that had integrated the cassette (Natsume et al., 2016). This sequence was flanked by AAVS1 locus homology arms that would act as a repair template following Cas9 cleavage. This vector was transfected into the HNRNPUL1-AID line along with a plasmid that expressed Cas9 plus a guide RNA to direct the enzyme to the AAVS1 locus. Puromycin was then added to growth media to facilitate selection of positive clones. Colonies were then subject to three conditions –

A
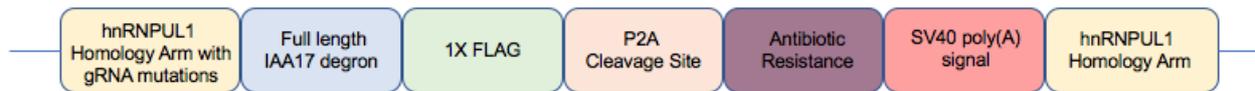
B
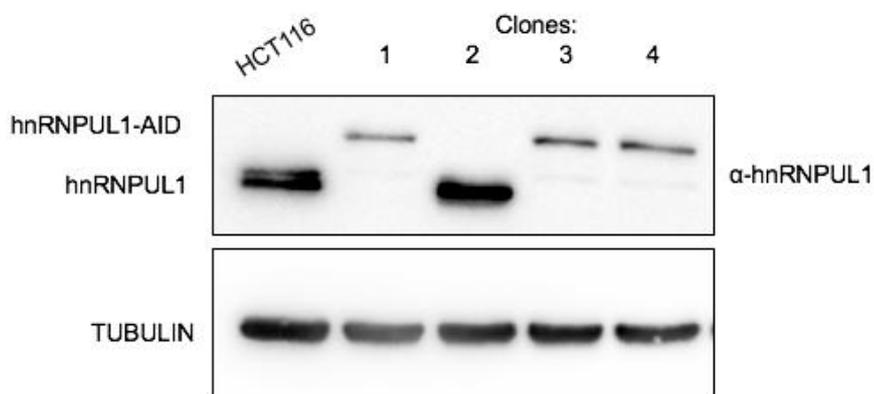
Figure 3.4 Tagging HNRNPUL1 with an AID containing the full length degron in the absence of TIR1 facilitates higher expression levels of HNRNPUL1-AID

A – Schematic diagram of new AID sequence, containing the full length IAA17 degron. B – Western analyses of clones grown in selective media following transfection of the Cas9/gRNA B plasmid along with the new repair templates. Tubulin is displayed as a loading control.

untreated, doxycycline-treated, and doxycycline- and auxin-treated, before being screened via Western analyses. In the case of the doxycycline and auxin condition, doxycycline was added to the media 48 hours prior to the addition of auxin so that maximal TIR1 expression would be induced prior to the introduction of the hormone. Three out of the four clones screened in this instance appeared to have successfully integrated the dox-OsTIR1 gene, as evidenced by the clear decrease in HNRNPUL1-AID expression levels in the doxycyclin and auxin-treated conditions (Figure 3.5A). Surprisingly, significant degradation of HNRNPUL1-AID was also observed in the doxycycline-only treated condition of the successful clones. Clone 2 was chosen as the HNRNPUL1-AID/Dox-TIR1 line to expand and used for future experiments. Comparison of HNRNPUL1 levels in knockdown conditions between this line and the previously generated HNRNPUL1 RNAi line revealed that a much greater depletion could be achieved via the AID system (Figure 3.5B).

Next, a time-course assay was performed to assess how quickly HNRNPUL1-AID could be degraded in this cell line upon the addition of doxycycline and auxin. During this assay, cells were treated with doxycycline and auxin and then harvested at various timepoints. This revealed that a 48 hour doxycycline incubation combined with a 2 hour auxin incubation is required to deplete HNRNPUL1-AID to virtually undetectable levels (Figure 3.6).

The original appeal of the AID system was the potential for rapid degradation of the target protein. Given that the 48-hour depletion period in the HNRNPUL1-AID line is no faster than can be achieved in the previously generated HNRNPUL1 RNAi line, the major difference between these two lines going forward was that the extent of HNRNPUL1 knockdown following doxycycline and auxin addition in the HNRNPUL1-AID line is considerably greater than via RNAi. At this stage it remained unnknown whether this difference in knockdown levels had functionally significant consequences for the cell, and this question was examined in the next section.
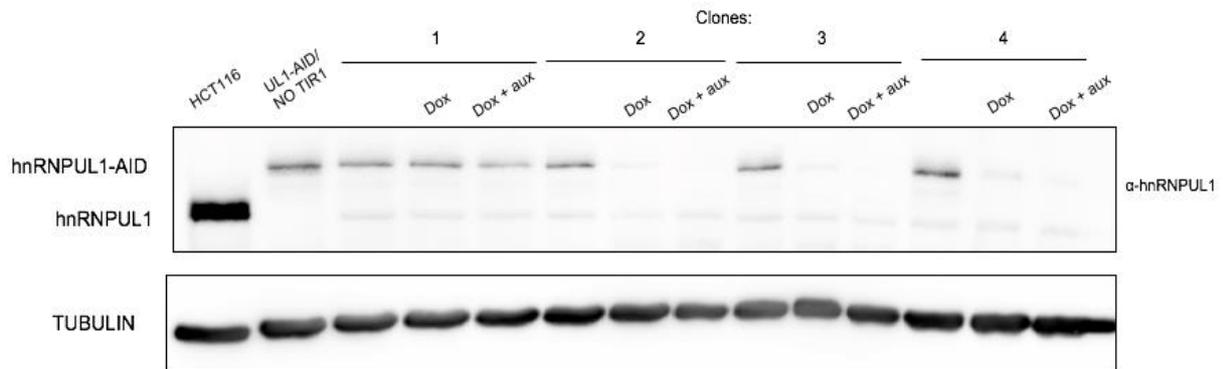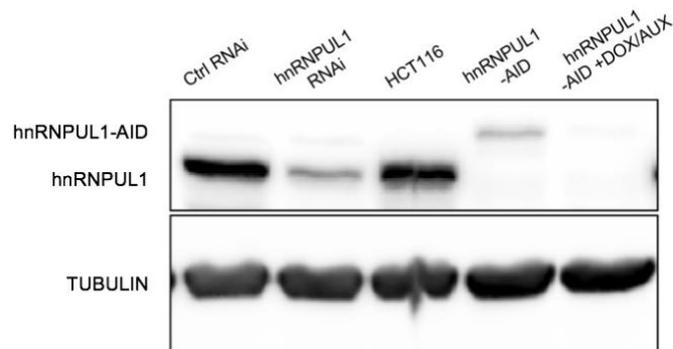
Figure 3.5 Integration of dox-OsTIR1 at the AAVS1 safe harbour locus facilitates significant degradation of HNRNPUL1-AID in the presence of doxycycline and auxin

A -Western analyses of clones grown in selective media following transfection of the dox-osTIR1 and AAVS1 Cas9 expression vectors. Cells were either untreated, treated with doxycycline for 72 hours, or treated with doxycycline for 72 hours and auxin for 24 hours. Tubulin is displayed as a loading control. B – Western analyses comparing level of HNRNPUL1 knockdown achieved in HNRNPUL1 RNAi and HNRNPUL1-AID lines.
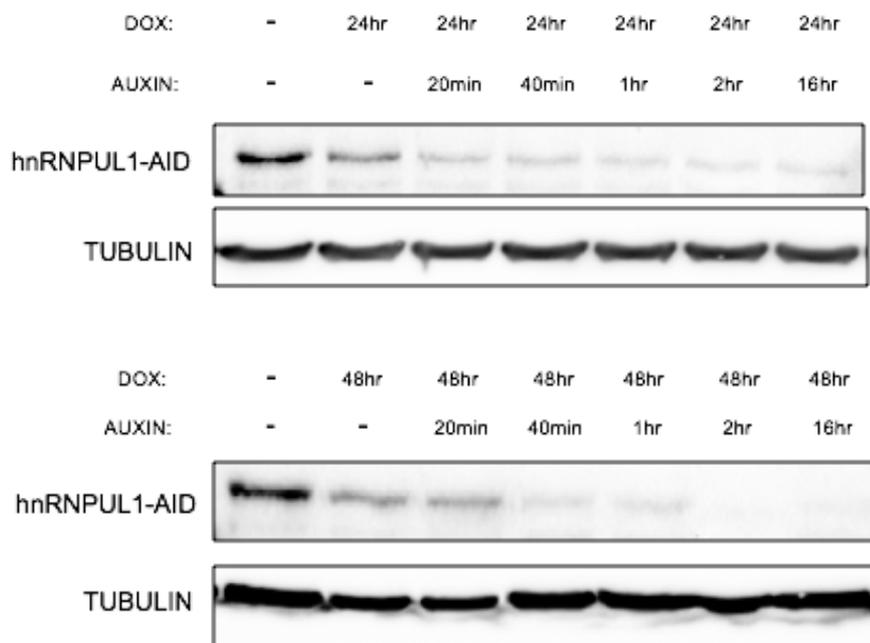
Figure 3.6 Time-course assay revealing that a 48-hour doxycycline incubation and 2-hour auxin incubation is necessary to cause complete loss of HNRNPUL1-AID levels

Western analyses of time-course assay comparing HNRNPUL1-AID levels at various time-points following the addition of doxycycline and auxin. Tubulin is shown as a loading control.

<u>3.4 Validating the HNRNPUL1-AID/Dox-TIR1 line</u>

Following the generation of the HNRNPUL1-AID/Dox-TIR1 cell line, it was important to test whether knockdown of HNRNPUL1 in this line produced expected phenotypes based on what has been previously published regarding the functions of HNRNPUL1. In addition, the severity of these phenotypes compared to the HNRNPUL1 RNAi line previously generated in the Wilson lab had yet to be assessed.

Therefore, HNRNPUL1 RNAi Encode RNA-seq data was analysed by Dr Ian Sudbery to generate a list of transcripts whose levels significantly increased or decreased upon HNRNPUL1 knockdown. RNA was extracted from the Wilson lab-generated HNRNPUL1 RNAi line and the HNRNPUL1-AID line and converted to cDNA. The levels of a selection of the transcripts identified by Dr. Sudbery were then analysed via qRT-PCR to examine whether the same phenotypes observed in the Encode HNRNPUL1 RNAi dataset could be replicated in both of our cell lines. A control RNAi cell line was used as the baseline condition to compare against the HNRNPUL1 RNAi line, while both parental HCT116 and untreated HNRNPUL1-AID line samples were analysed to compare against HNRNPUL1-AID treated with doxycycline and auxin (Figure 3.7A). Minimal changes were observed for all transcripts tested in the case of the HNRNPUL1 RNAi line compared to the control condition (Figure 3.7B). However, in the case of the HNRNPUL1-AID line treated with doxycycline and auxin, the abundance of every transcript tested changed in the expected direction compared to the HCT116 condition (Figure 3.7C). Specifically, EGR1, MT-ND1 and MT-ND2 all increased, while DCTPP1 and MIR17HG decreased. These changes all occur in the Encode RNA-seq dataset. U1 snRNA was originally chosen as an internal normaliser transcript, however to our surprise its levels were consistently lower in the HNRNPUL1-AID treated with doxycycline and auxin condition (Figure 3.7C), and therefore 18S rRNA was chosen as a normaliser for each sample tested. This U1 snRNA finding is pursued in more detail in Chapter 4. In the case of some transcripts, there was no observable difference between the untreated
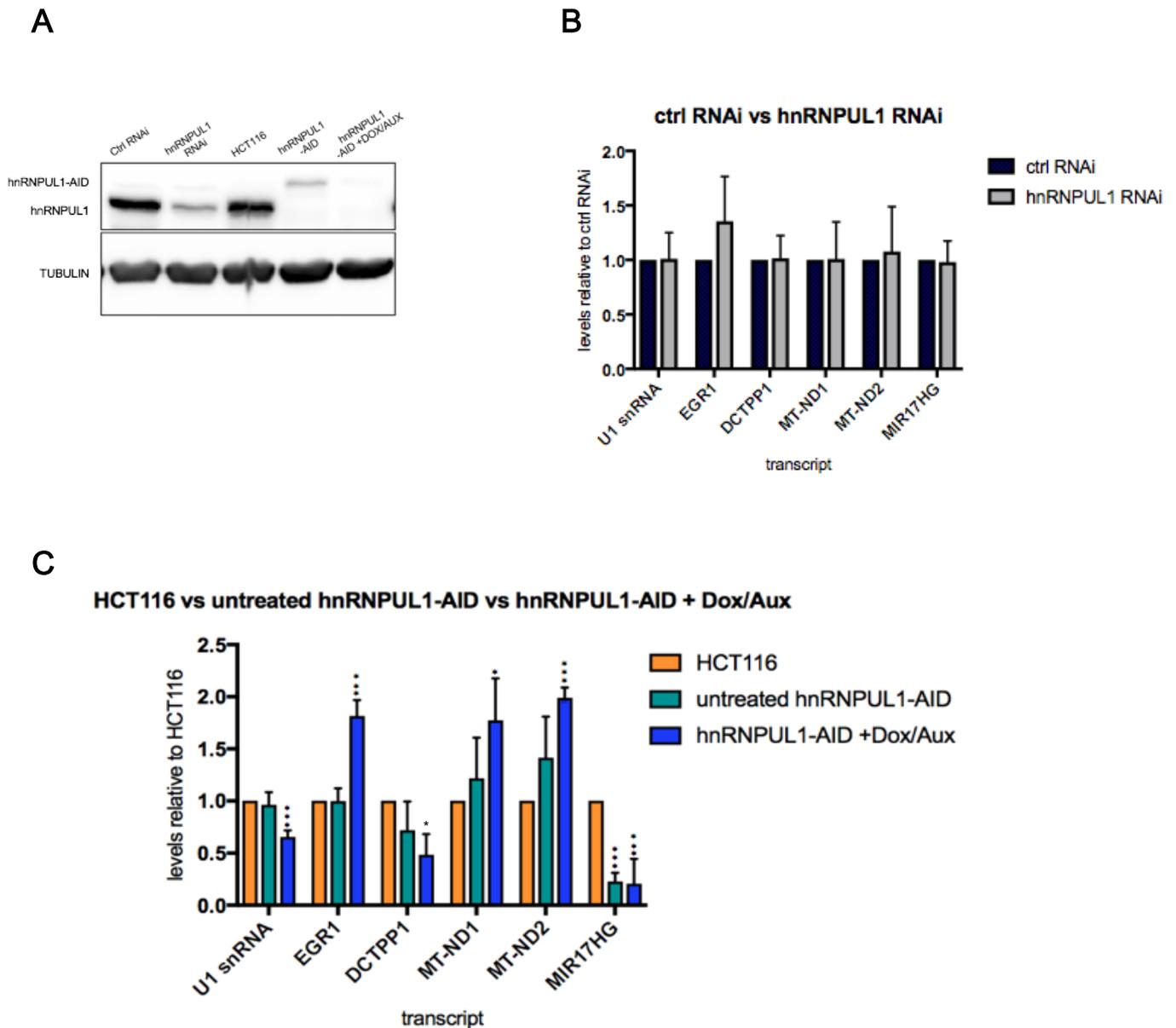
Figure 3.7 HNRNPUL1-AID treated with doxycycline and auxin displays phenotypes consistent with HNRNPUL1 RNAi Encode RNA-seq data

A – Western analyses demonstrating HNRNPUL1 levels in the various conditions incorporated into qRT-PCR analyses. B – qRT-PCR analyses on selected Encode transcripts in control RNAi line vs HNRNPUL1 RNAi line. Errors bars represent the SD. Transcripts were internally normalised to 18S rRNA levels. C – qRT-PCR analyses on selected transcripts in HCT116 vs untreated HNRNPUL1-AID vs HNRNPUL1-AID treated with doxycycline and auxin (+Dox/Aux) lines. Errors bars represent the SD. Transcripts were internally normalised to 18S rRNA levels.

HNRNPUL1-AID line and HCT116 cells, while in other cases there were mild phenotypes that are less pronounced than those displayed in the plus doxycycline and auxin condition. Combined with the minimal effects observed in the HNRNPUL1 RNAi line in the qRT-PCR analyses, this suggests that a large depletion of HNRNPUL1 levels is required to effectively disrupt its function.

3.5 HNRNPUL1 is required for cell proliferation

In addition to qRT-PCR analyses, colony formation assays in the HNRNPUL1-AID/Dox-TIR1 line were performed to assess the impact of the loss of HNRNPUL1 levels upon cell viability. Once again, three conditions – HCT116, untreated HNRNPUL1-AID, and doxycycline and auxin-treated HNRNPUL1-AID – were incorporated into the experiment. Interestingly, a drastic decrease in the number of colonies after 14 days – from a mean of 148 to a mean of 26.33 - was observed in the doxycycline and auxin-treated condition (Figure 3.8A, Figure 3.8B). Firstly, this indicates that HNRNPUL1 is required for efficient cell proliferation. Secondly, it appears that the moderate depletion of HNRNPUL1 observed in the untreated HNRNPUL1-AID line does not significantly impact the ability of HNRNPUL1 to perform its cellular functions, a hypothesis supported by the qRT-PCR analysis on untreated HNRNPUL1-AID and the HNRNPUL1 RNAi line described earlier.

Therefore, the HNRNPUL1-AID/Dox-TIR1 line was used to perform future functional assays on HNRNPUL1, given that the largest HNRNPUL1 knockdown can be achieved in this line. The parental HCT116 line was chosen as the control condition as opposed to untreated HNRNPUL1-AID, due the untreated line already displaying moderate depletion of HNRNPUL1 as well as partial phenotypes in the qRT-PCR assays. The control HCT116 line was treated with doxycycline and auxin at the same concentrations and time-points as the HNRNPUL1-AID line in all future functional assays to ensure that any observed phenotypes are not as a result of the effects of these compounds themselves.

A

HCT116 +Dox/Aux

untreated hnRNP-UL1-AID

hnRNP-UL1-AID +Dox/Aux

B

Colony formation assay

Colonies per well

HCT116 + Dox/Aux
hnRNPUL1-AID untreated
hnRNPUL1-AID + Dox/Aux

HCT116 + Dox/Aux
hnRNPUL1-AID untreated
hnRNPUL1-AID + Dox/Aux
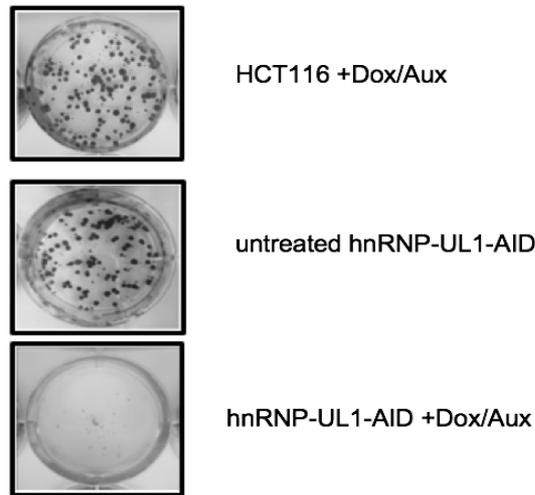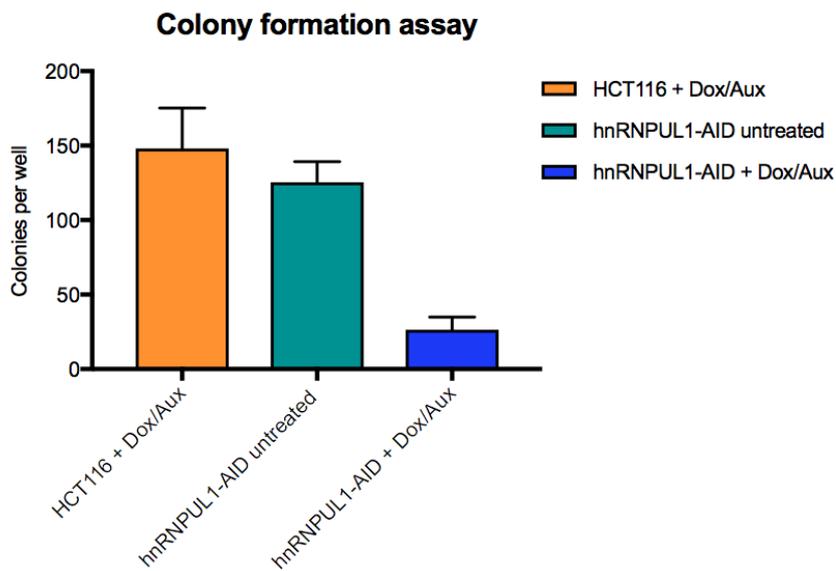
Figure 3.8 Colony formation assay demonstrating a significant decrease in cell viability following total depletion of HNRNPUL1

A – Photograph of one of the six wells seeded per condition. B – Quantification of the mean number of colonies grown after 14 days per condition. Errors bars represent the SD.

3.6 Immediate-early genes are upregulated upon HNRNPUL1 depletion

One of the transcripts included in the initial qRT-PCR analysis validation of the HNRNPUL1-AID line was EGR1. EGR1 is one of the most significantly increased transcripts in the HNRNPUL1 RNAi Encode RNA-seq dataset, and a clear upregulation was also observed in the HNRNPUL1-AID line upon HNRNPUL1 depletion (Figure 3.7C). EGR1 is an example of an immediate-early gene (IEG). Expression of IEGs are rapidly activated within minutes of stimulation by internal or external factors such as cytokines or serum (Bahrami and Drabløs, 2016). This is facilitated in many cases by RNAPII already bound at the promoters of these genes awaiting pause-release activation (Li and Gilmour, 2011). Many IEGs are transcription factors such as FOS and FOSB that trigger the activation of a diverse range of signalling pathways, and are therefore critical for several cellular processes such as differentiation and the immune response (Bahrami and Drabløs, 2016). IEGs are also induced in response to DNA damage, a pathway in which HNRNPUL1 has been previously reported to play a role (Polo et al., 2012). IEG expression is typically fairly transient, for example FOS expression levels peak around 30-60 minutes post-activation before returning to basal levels after 90 minutes (Greenberg and Ziff, 1984).

To assess whether other IEGs are upregulated in response to HNRNPUL1 knockdown, further qRT-PCR analyses were performed on total RNA that had been extracted from HNRNPUL1-AID cells and converted into cDNA. These revealed that, in addition to EGR1, both FOS and FOSB are significantly upregulated upon HNRNPUL1 depletion (Figure 3.9A). Time-course assays were performed subsequently whereby expression of IEGs was activated by adding 20% FCS to cells previously starved of serum, with RNA extracted at various time-points during this process. These assays demonstrated that elevated IEG mRNA transcript levels occur upon HNRNPUL1 knockdown throughout the cycle of IEG expression activation and repression (Figure 3.9B, Figure 3.9C).

We next set out to identify whether the increased IEG mRNA levels were brought

A – Immediate Early Genes in hnRNPUL1-AID line

B – EGR1 serum time course in HNRNPUL1-AID line

C – FOS serum time course in HNRNPUL1-AID line
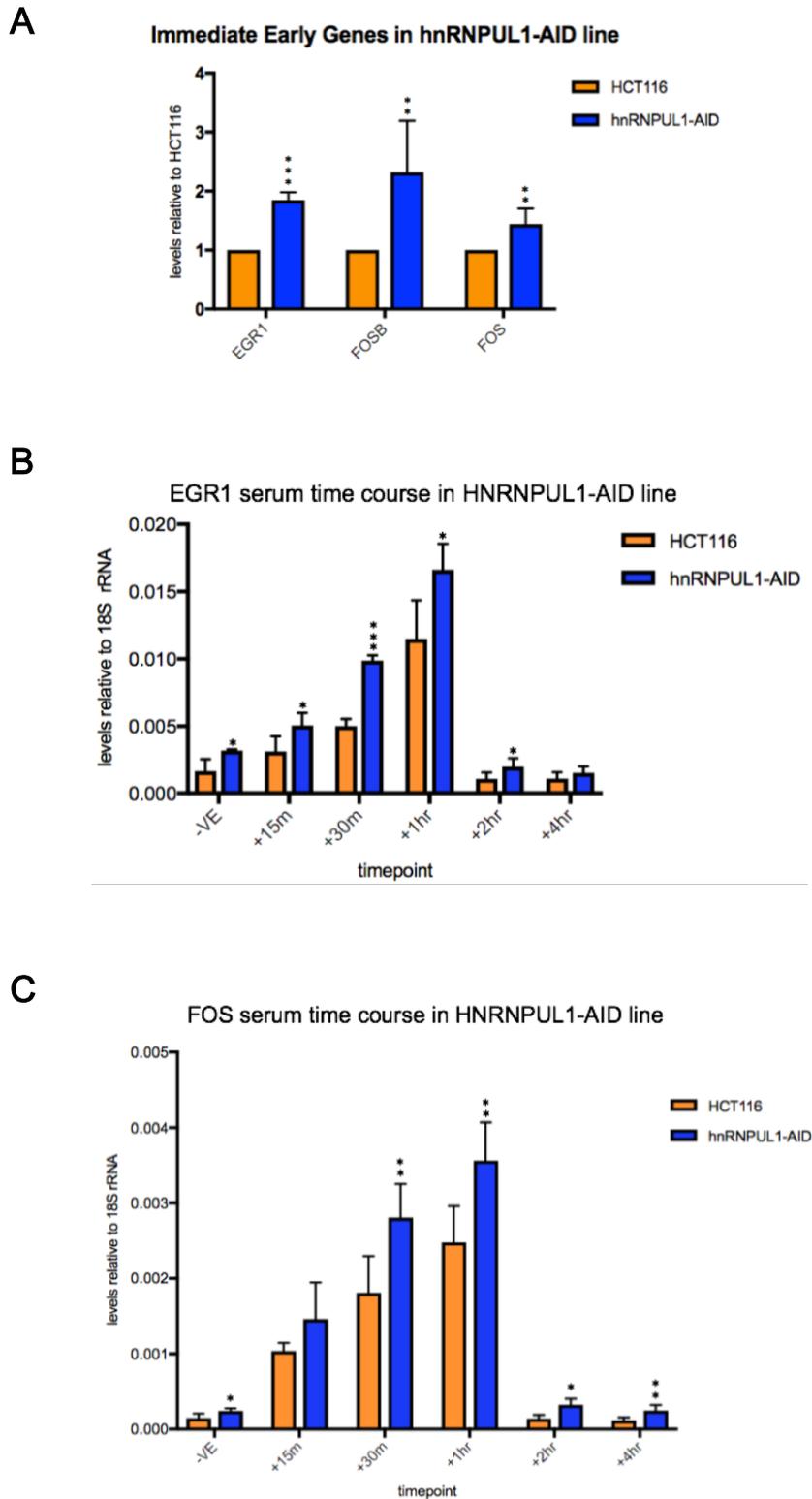
Figure 3.9 IEGs are upregulated at the mRNA level upon HNRNPUL1 knockdown

A – qRT-PCR analyses of total RNA levels of IEGs in HNRNPUL1-AID line. Error bars represent the SD. B – qRT-PCR analyses of EGR1 mRNA levels at various time-points over the course of a serum-starvation time-course assay. Cells were starved of serum

for 24 hours prior to the addition of 20% FCS to the media. Error bars represent the SD. C – qRT-PCR analyses of FOS mRNA levels at various time-points over the course of a serum-starvation time-course assay. Cells were starved of serum for 24 hours prior to the addition of 20% FCS to the media. Error bars represent the SD.

about via an upregulation in transcription. Therefore, RNAPII ChIP assays were performed in the HNRNPUL1-AID line and the RNAPII signal at various loci throughout the EGR1 and FOS genes was analysed via qRT-PCR. These revealed a clear increase in RNAPII signal at downstream regions of these genes (Figure 3.10A, Figure 3.10B), consistent with an upregulation of transcription via RNAPII pause-release from the promoter-proximal regions of these genes. This phenotype observed via ChIP was confirmed by mNET-seq analysis performed by Ivaylo Yonchev in the Wilson lab, which also identified elevated RNAPII signal in the body of both the EGR1 and FOS genes (Figure 3.11A, Figure 3.11B).
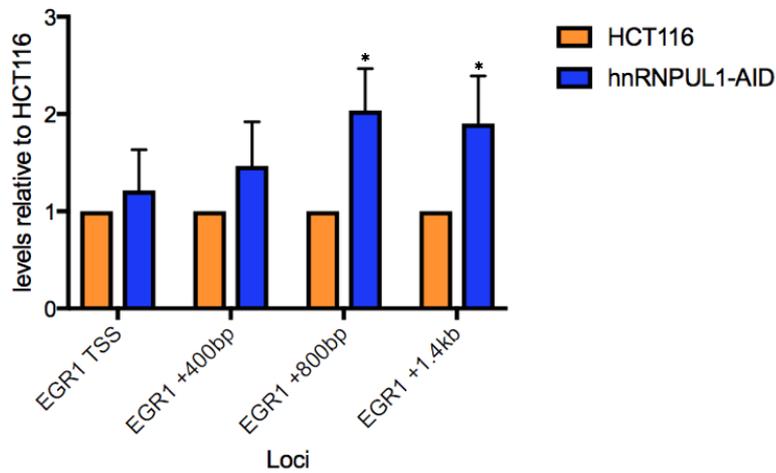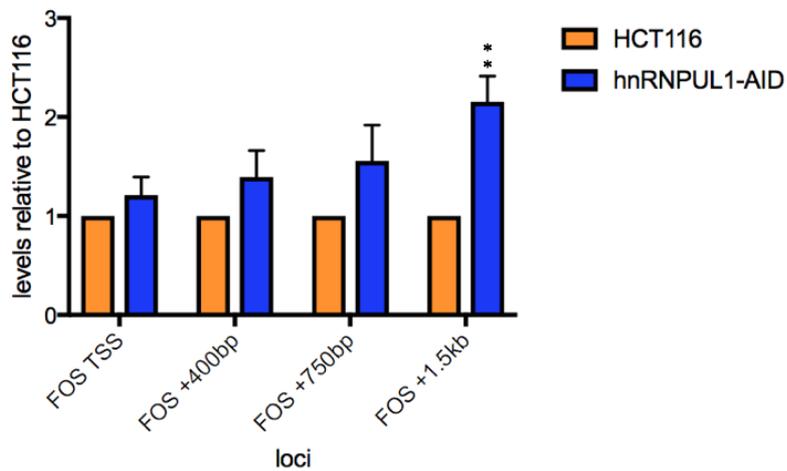
Figure 3.10 RNAPII ChIP analyses reveal upregulation of IEGs at the transcription level upon HNRNPUL1 knockdown

A – qRT-PCR analysis of RNAPII signal at EGR1 gene loci following ChIP performed plus and minus HNRNPUL1. Errors bars represent the SD. B - qRT-PCR analysis of RNAPII signal at FOS gene loci following ChIP performed plus and minus HNRNPUL1. Errors bars represent the SD.
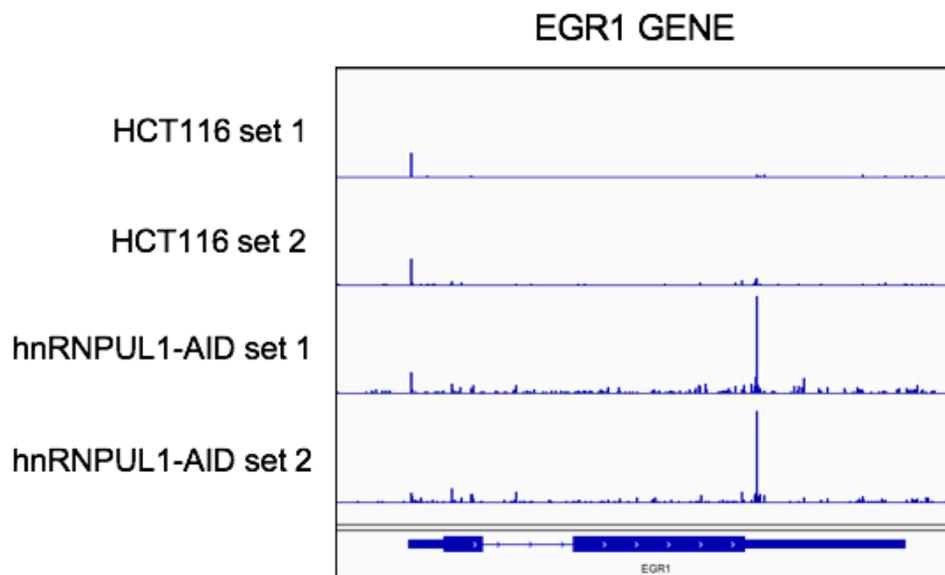
A – EGR1 GENE
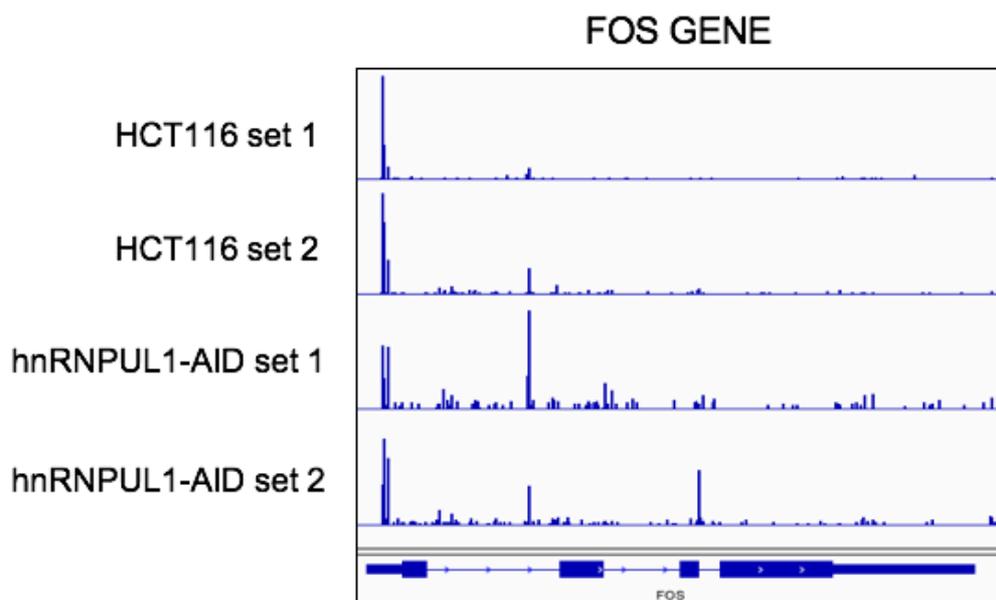
B – FOS GENE

Figure 3.11 mNET-seq analysis confirms activation of IEG transcription upon HNRNPUL1 knockdown

A – IGV tracks of the RNAPII signal at the EGR1 gene following mNET-seq performed by Ivaylo Yonchev. B - IGV tracks of the RNAPII signal at the FOS gene following mNET-seq performed by Ivaylo Yonchev.

3.7 The generation of stable FLAG-HNRNPUL1 cell lines

As well as investigating the cellular consequences of HNRNPUL1 knockdown, dissecting the roles of the individual domains within this protein was also a key priority. Therefore, in addition to the HNRNPUL1-AID line, we also attempted to generate stable cell lines expressing wild-type and mutant forms of FLAG-tagged HNRNPUL1 using the Flp-In T-Rex System (Thermofisher). Under this method, a vector expressing the gene of interest, along with a vector expressing an enzyme known as Flp recombinase, are transfected into a host cell line that possesses an FRT site. This facilitates Flp-recombinase-mediated integration of the host gene into the FRT target site. The host cell line also expresses the Tet repressor, enabling expression of the gene of interest under the control of tetracycline.

Introducing a triple amino-acid substitution into the Walker A motif of HNRNPUL1 disrupts its ability to bind nucleotides (Wilson lab, unpublished data). Three FRT/3X-FLAG vectors – one expressing this nucleotide-binding mutant form of HNRNPUL1, one expressing wild-type HNRNPUL1 and another expressing a C-terminal truncated form of HNRNPUL1 that lacked its PrLD (Figure 3.12A) – were transfected into host Flp-In T-REx 293 cells. Following antibiotic selection, colonies were treated plus and minus tetracycline and screened via Western analyses. Every colony screened had integrated the FLAG-tagged forms of HNRNPUL1, with a significant band appearing at the size of HNRNPUL1 in the 'plus tetracycline' condition (Figure 3.12B).

It was clear from this initial screening Western blot that the different forms of FLAG-HNRNPUL1 displayed different expression levels when induced at the same 2µg/ml concentration of tetracycline. Therefore, a tetracycline titration assay was carried out to identify the tetracycline concentration that induced equal expression levels among all three forms of HNRNPUL1. This would ensure that any effects identified in upcoming functional assays would be as a result of the mutations themselves, not due to expression level differences. Western analyses of samples harvested following induction with various concentrations of tetracycline revealed that the most similar
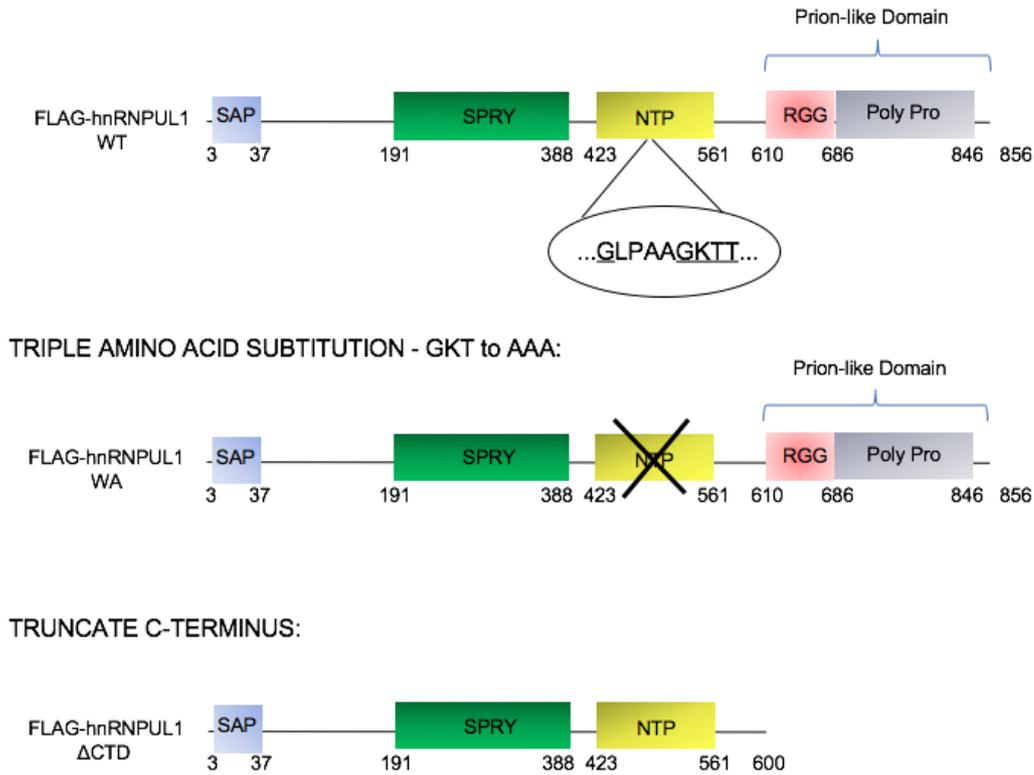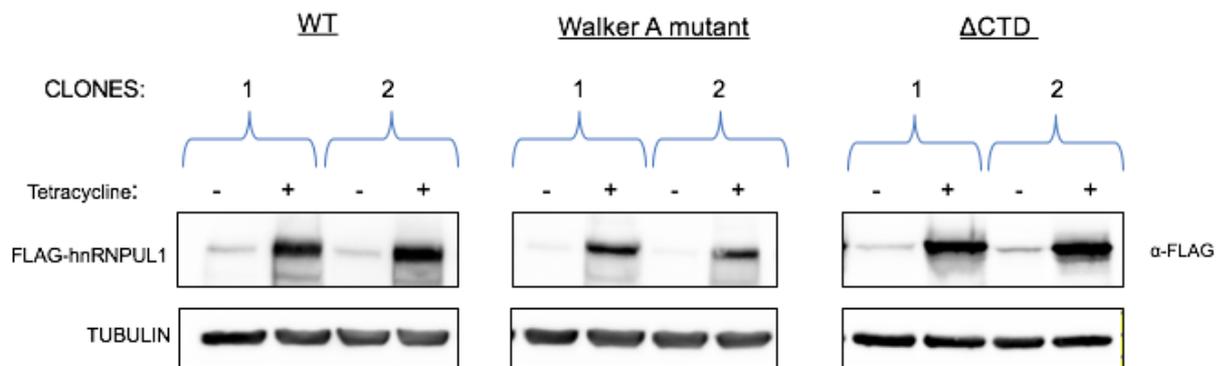
Figure 3.12 Generation of tetracycline-inducible FLAG-HNRNPUL1 WT, FLAG-HNRNPUL1 WA and FLAG-HNRNPUL1 ΔCTD cell lines

A – Schematic diagram of FLAG-HNRNPUL1 WT, FLAG-HNRNPUL1 WA and FLAG-HNRNPUL1 ΔCTD. B – Western analyses of clones grown in selective media following transfection of the pcDNA5/FRT/FLAG-HNRNPUL1 and pOG44 Flp recombinase expression vectors. Expression of FLAG-tagged proteins was induced by addition of tetracycline (2µg/ml) for 48 hours. Tubulin is displayed as a loading control.

expression levels were brought about following tetracycline induction at 2 µg/ml in the case of HNRNPUL1 WA, and 0.0167 µg/ml in the case of HNRNPUL1 WT and HNRNPUL1 ΔCTD (Figure 3.13A). These tetracycline concentrations also resulted in expression of FLAG-HNRNPUL1 that is comparable to the endogenous form of the protein, as revealed by Western analyses using an endogenous HNRNPUL1 antibody on a FLAG-HNRNPUL1 WT sample induced at this 0.0167 µg/ml tetracycline concentration (Figure 3.13B). Hence, these concentrations were used to induce FLAG-HNRNPUL1 expression in all subsequent experiments involving these cell lines.

3.8 Walker A and C-terminus mutations do not disrupt the mRNA-binding capacity of HNRNPUL1

A defining characteristic of HNRNPUL1 is its ability to bind RNA. It has been widely established that RGG boxes interact with RNA (Thandapani et al., 2013), however whether the other domains of HNRNPUL1 contribute to its RNA-binding capacity remains unknown. It was recently demonstrated that the central SPRY domain plus NTP region of HNRNPU possessed RNA-binding ability (Panhale et al., 2019), which may also be true in the case of HNRNPUL1 given the high degree of homology between the two proteins.

To address the impact of the Walker A and ΔCTD mutations on HNRNPUL1 RNA-binding at a global level *in vivo*, an mRNP capture assay was performed involving the FLAG-HNRNPUL1 WT, FLAG-HNRNPUL1 WA and FLAG-HNRNPUL1 ΔCTD stable cell lines. This assay involves the UV-crosslinking of protein to RNA followed by the purification of mRNP complexes via poly d(T) coated beads. A non-UV crosslinked condition along with a crosslinked condition treated with RNase A were used as controls for the capture. Interestingly, this capture revealed no significant differences in the amount of FLAG-HNRNPUL1 co-purifying with mRNA transcripts between the wild-type form of the protein and the two mutants (Figure 3.14). This suggests that other HNRNPUL1 domains in addition to the RGG box at the C-terminus possess the
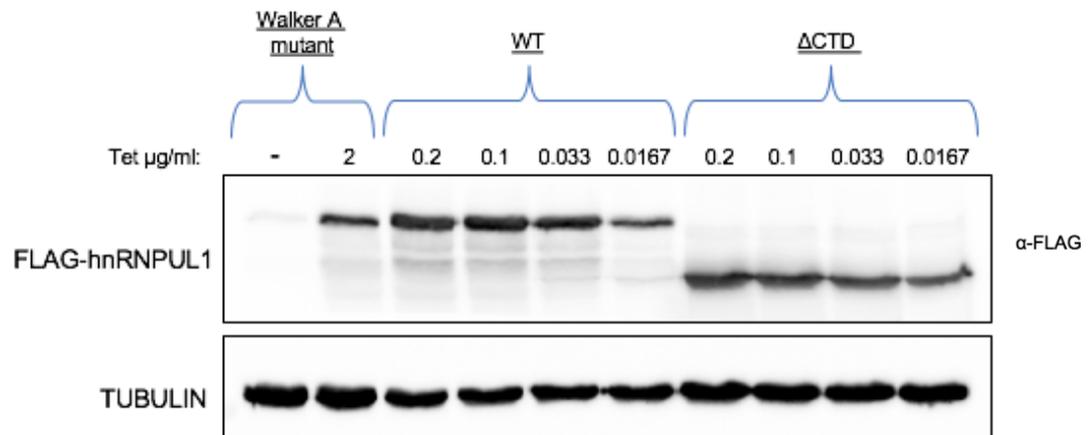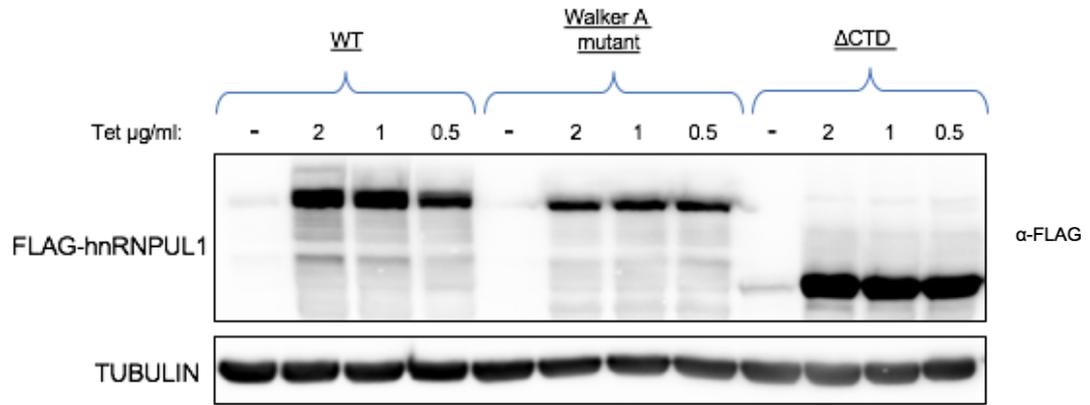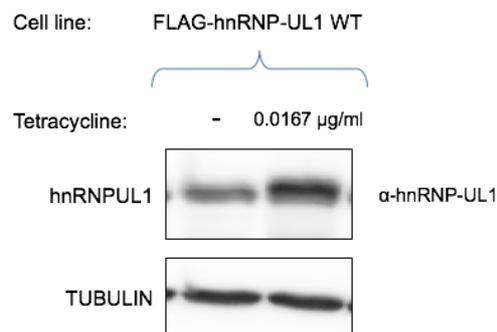
Figure 3.13 The effects of varying tetracycline concentration on expression levels of FLAG-HNRNPUL1 WT and mutants in the stable FLAG lines

A - Western analyses of tetracycline titration assay examining the expression levels of FLAG-HNRNPUL1 WT and mutants following the addition of tetracycline at various concentrations. Tubulin is shown as a loading control. B – Western analyses comparing the expression levels of FLAG-UL1 WT induced at 0.0167 µg/ml to endogenous HNRNPUL1. Tubulin is shown as a loading control.
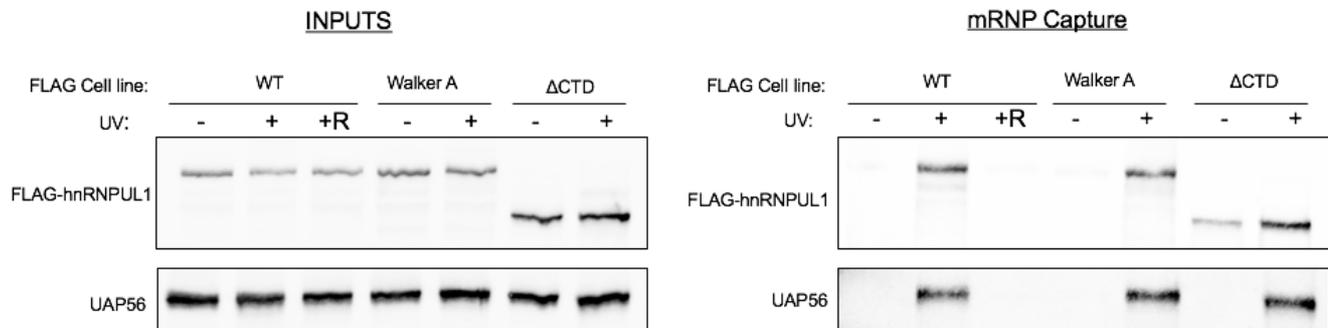
Figure 3.14 The mRNA-binding ability of HNRNPUL1 is unaffected by Walker A and ΔCTD mutations

Western analyses measuring the levels of FLAG-HNRNPUL1 WT, FLAG-HNRNPU-UL1 WA, and FLAG-HNRNPUL1 ΔCTD RNA-binding following mRNP capture assays. UAP56 is shown as a control for mRNP capture efficiency in each condition. +R = UV-crosslinked and RNase A-treated condition.

ability to bind RNA, and therefore mutating or deleting only one or two of these domains is not sufficient to hinder its RNA-binding capacity. This finding also indicates that nucleotide-binding is not a prerequisite for HNRNPUL1 association with RNA. Whether nucleotide-binding affects the association of HNRNPUL1 with its protein interacting partners will be explored in subsequent chapters.

3.9 Summary

In this chapter, a HNRNPUL1 auxin-inducible degron cell line has been generated, facilitating the depletion of this protein to extremely low levels following the addition of doxycycline and auxin. During this process, it became apparent that in the case of HNRNPUL1, a 3X mini-AID tag added in the presence of constitutively expressed TIR1 triggers degradation of HNRNPUL1 regardless of auxin addition. Therefore, alterations to the degron sequence were introduced, along with placing TIR1 expression under the control of doxycycline, in order to improve HNRNPUL1-AID expression levels under basal conditions. Although the doxycycline-inducible component of this cell line means that knockdown of HNRNPUL1 is not as rapid as previously expected, the greater extent to which the protein can be depleted relative to alternative methods such as RNAi justifies continuing with this AID system to study the cellular effects of loss of HNRNPUL1 expression. This was conclusively demonstrated by qRT-PCR analyses of RNA expression changes in HNRNPUL1 knockdown conditions in both the AID line and the RNAi line. These revealed much stronger phenotypes in the case of the HNRNPUL1-AID line. Immediate-early genes are particularly affected by HNRNPUL1 depletion in the HNRNPUL1-AID line, and are upregulated at the level of transcription throughout their rapid gene activation and repression cycle.

Colony formation assays performed in the newly generated HNRNPUL1-AID line revealed that HNRNPUL1 is required for efficient cell proliferation. These assays also demonstrated that low levels of HNRNPUL1 appear to be sufficient for the protein to function, given that there was minimal change in colony formation numbers in the untreated HNRNPUL1-AID cell line condition.

In addition to a conditional HNRNPUL1 knockout cell line, stable cell lines expressing FLAG-HNRNPUL1 wild type, Walker A mutant and ΔCTD were also generated using the Flp-In system. Perhaps surprisingly, the HNRNPUL1 ΔCTD mutant lacking its RGG box bound RNA as well as the wild type form of the protein in

mRNP captures performed in these cell lines. This indicates that either the SAP, SPRY or NTP regions also possess RNA-binding capacity. The HNRNPUL1 Walker A mutant also displayed no mRNA-binding defect, suggesting that nucleotide binding is not essential for HNRNPUL1:RNA interactions, at least when an RGG box domain is also present.

CHAPTER 4 – HNRNPUL1 AND SMALL RNA BIOGENESIS

Following the successful generation of a HNRNPUL1-AID line and stable FLAG-HNRNPUL1 lines, the function of HNRNPUL1 with regards to RNA metabolism can be investigated. A surprising result from the qRT-PCR validation assays of the HNRNPUL1-AID line performed in Chapter 3 was the downregulation of U1 snRNA levels upon HNRNPUL1 knockdown. While HNRNPUL1 has been reported to co-immunoprecipitate with U1 snRNP (Chi et al., 2018b), and HNRNPU has been demonstrated to negatively regulate U2 snRNP maturation (Xiao et al., 2012), a role for HNRNPUL1 in snRNP biogenesis has not been established previously. Given the widely reported association between snRNP biogenesis defects and neurodegeneration (Coady and Lorson, 2011; Tsuiji et al., 2013), investigating a potential role for HNRNPUL1 in this cellular pathway was of great interest, and therefore pursued throughout the work described in this chapter.

4.1 snRNA levels are downregulated upon HNRNPUL1 knockdown

In addition to the reduction in U1 snRNA levels observed via qRT-PCR upon HNRNPUL1 knockdown, mNET-seq assays performed in the HNRNPUL1-AID line by Ivaylo Yonchev in the Wilson lab revealed a global reduction in the amount of snRNAs co-purifying with RNAPII in the HNRNPUL1 knockdown condition (Figure 4.1A). Therefore, to assess whether other snRNAs were also affected by HNRNPUL1 knockdown, a panel of snRNAs were analysed via qRT-PCR following RNA extraction from the HNRNPUL1-AID line. Although efficient knockdown of HNRNPUL1 can be achieved in the HNRNPUL1-AID cell line following 48hrs doxycycline-addition and 2hrs auxin-addition, cells were treated with doxycycline and auxin for an additional 48hrs in this assay due to the long half-lives of snRNAs (Sauterer et al., 1988), which could potentially mask any defects in their biogenesis if analysed at a short time point. This revealed that U2, U4 and U5 in addition to U1 snRNA were all downregulated as

A

mNET-seq signal over all snRNA genes

B

snRNA levels in hnRNPUL1-AID line
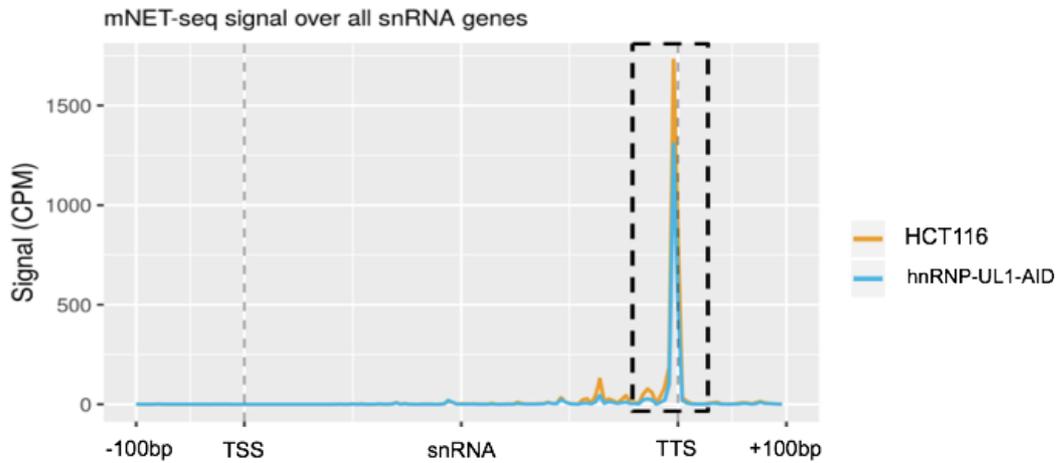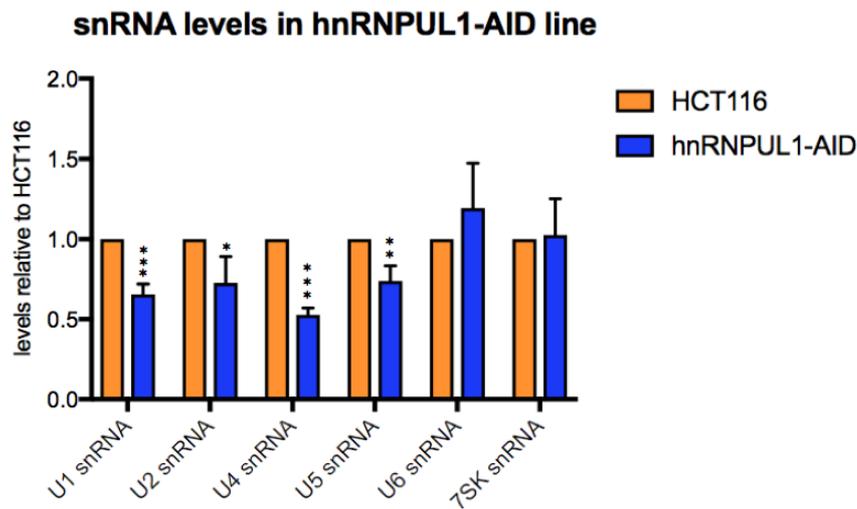
Figure 4.1 RNAPII-transcribed snRNAs are downregulated upon HNRNPUL1 knockdown

A – mNET-seq signal across snRNA genes following mNET-seq analyses in HNRNPUL1-AID line performed by Ivaylo Yonchev. CPM = Counts per million. TSS = Transcription start site, TTS = Transcription termination site. B – qRT-PCR analyses of snRNA levels following HNRNPUL1 depletion in HNRNPUL1-AID line. Errors bars represent the SD.

a result of HNRNPUL1 depletion, while U6 and 7SK snRNAs were not affected (Figure 4.1B). As U6 and 7SK are transcribed by RNA polymerase III, this result suggests that the snRNA biogenesis defect observed is exclusive to RNAPII-transcribed snRNAs. As in the case of other transcripts screened following HNRNPUL1 knockdown in Chapter 3, the extent of the reduction of snRNA abundance appears to be sensitive to the extent of HNRNPUL1 depletion, with untreated and just doxycycline-treated cells displaying a reduced snRNA defect phenotype compared to doxycycline and auxin-treated cells (Figure 4.2A, Figure 4.2B).

4.2 HNRNPUL1 interacts with snRNAs, binding most strongly to U4

Following the observation that snRNAs are downregulated in the HNRNPUL1-AID line, we next wanted to establish whether HNRNPUL1 plays a direct role in snRNP biogenesis, and to identify which stage or stages of snRNP biogenesis were specifically disrupted upon HNRNPUL1 depletion.

Therefore, in order to firstly establish whether HNRNPUL1 bound snRNAs, HNRNPUL1 was immunoprecipitated from cells using an endogenous HNRNPUL1 antibody and the presence or absence of snRNAs in the IPed fraction was assessed via qRT-PCR. This revealed that HNRNPUL1 interacts with a variety of snRNAs including all the components of the major spliceosome, as well the 7SK snRNA component of the 7SK snRNP complex (Figure 4.3), as measured by fold enrichment over a negative IP using an anti-FLAG antibody. U4 snRNA bound particularly strongly, an interesting result in light of the qRT-PCR assays described previously, which revealed that this transcript was the most downregulated upon HNRNPUL1 knockdown amongst all the snRNAs (Figure 4.1B). These interactions are consistent with the hypothesis that HNRNPUL1 plays a direct role in at least one stage of the life cycle of an snRNA.
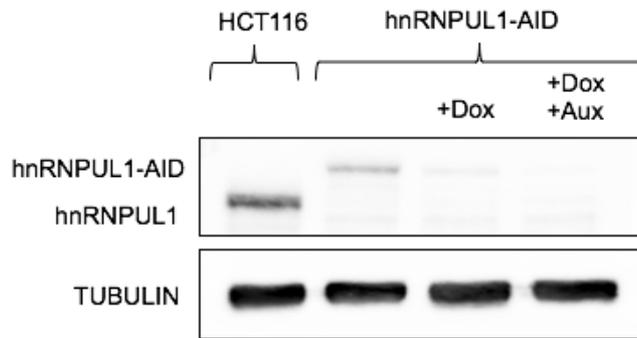
A – Western analyses comparing HNRNPUL1 levels in HCT116 cells vs untreated, doxycycline-treated (+Dox) and doxycycline and auxin-treated (+ Dox+Aux) HNRNPUL1-AID cells. Tubulin is displayed as a loading control. B - qRT-PCR analyses of snRNA levels following varying degrees of HNRNPUL1 depletion. Errors bars represent the SD.
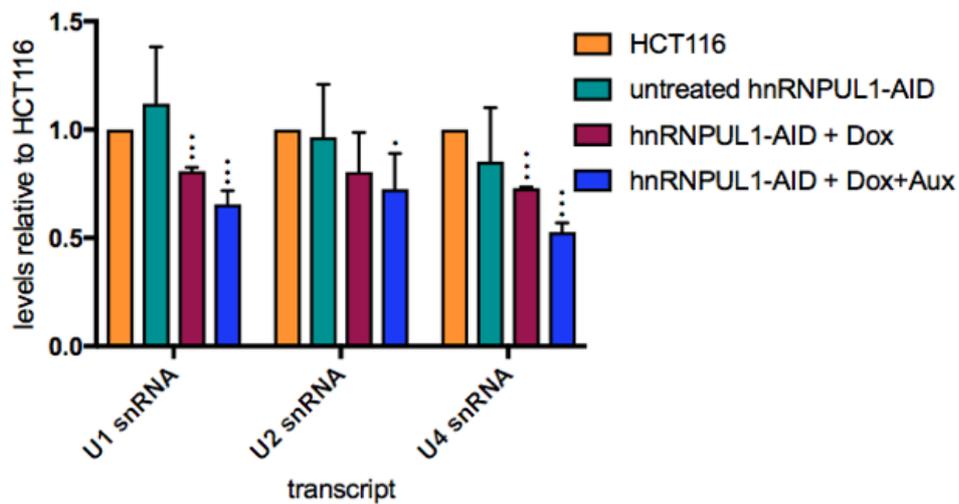
Figure 4.2 A greater knockdown of HNRNPUL1 results in a stronger snRNA defect phenotype
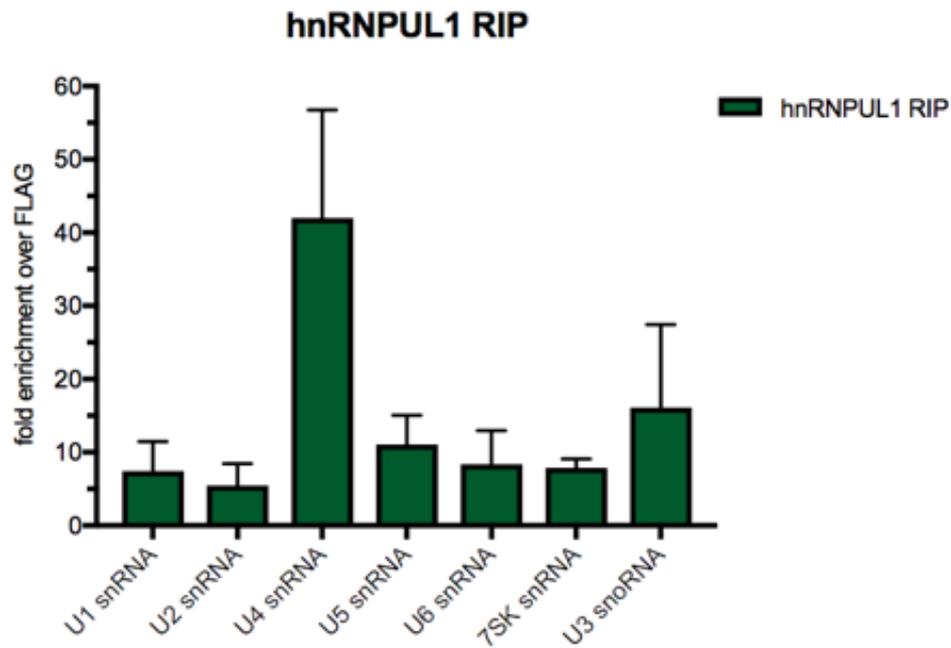
Figure 4.3 HNRNPUL1 interacts with snRNAs

qRT-PCR analyses following a HNRNPUL1 RIP assay using an endogenous HNRNPUL1 antibody. A FLAG-antibody was used as a negative control. Transcripts in the immunoprecipitated fraction of both IPs were measured as a percentage of input, and then a fold enrichment over the negative IP was calculated. Error bars represent the SD.

4.3 HNRNPUL1 knockdown downregulates transcription of RNAPII-transcribed snRNAs

Although the HNRNPUL1 RIP assay demonstrated HNRNPUL1 binding to RNAPIII-transcribed as well as RNAPII-transcribed snRNAs (Figure 4.3), the fact that the levels of U6 and 7SK snRNAs were unaffected in the HNRNPUL1-AID line suggests that the defect is specific to the RNAPII-transcribed snRNA pathway. A previous study has reported HNRNPUL1 to be a strong interactor of RNAPII (Chi et al., 2018a), indicating a direct role for HNRNPUL1 in transcription regulation. To confirm this interaction, RNAPII was immunoprecipitated from HCT116 cells and Western analyses were carried out to identify whether HNRNPUL1 co-immunoprecipitated (coIPed) with this complex. IPs were performed plus and minus DNase and RNase A to assess the importance of DNA and RNA with regards to the interaction. These revealed that HNRNPUL1 coIPs with RNAPII both in the presence and absence of DNA and RNA, at comparable levels with known strong interactors of RNAPII such as FUS and U1 snRNP (Figure 4.4A).

To assess whether HNRNPUL1 is present specifically at sites of snRNA transcription, publicly available HNRNPUL1 ChIP-seq data was analysed by Ian Sudbery who generated a meta-gene averaging the HNRNPUL1 signal across all snRNA loci. This demonstrated a clear HNRNPUL1 enrichment over snRNA genes, peaking over the body of the gene as well as the transcription termination site (Figure 4.4B). Combined with the RNAPII co-immunoprecipitation, this suggests HNRNPUL1 could be playing a direct role in the regulation of snRNA transcription.

To test whether there was indeed a defect in snRNA transcription upon HNRNPUL1 knockdown, RNAPII ChIP assays were performed in the HNRNPUL1-AID line and the RNAPII signal occupying U1, U4 and U3 sn/snoRNA loci was assessed via qPCR. This revealed a significant decrease in snRNA transcription at all three loci tested occurring upon HNRNPUL1 depletion (Figure 4.5). The RNAPII signal at the beta-actin gene was also analysed to identify whether the observed transcription phenotype
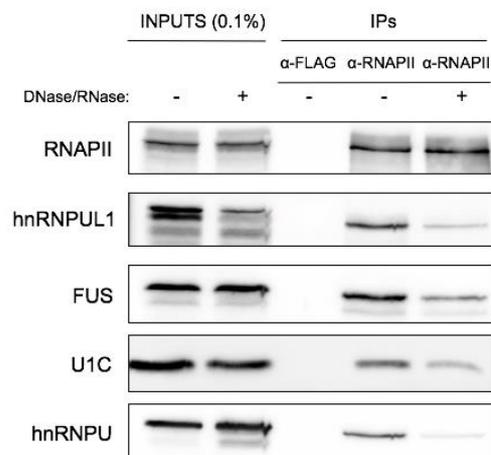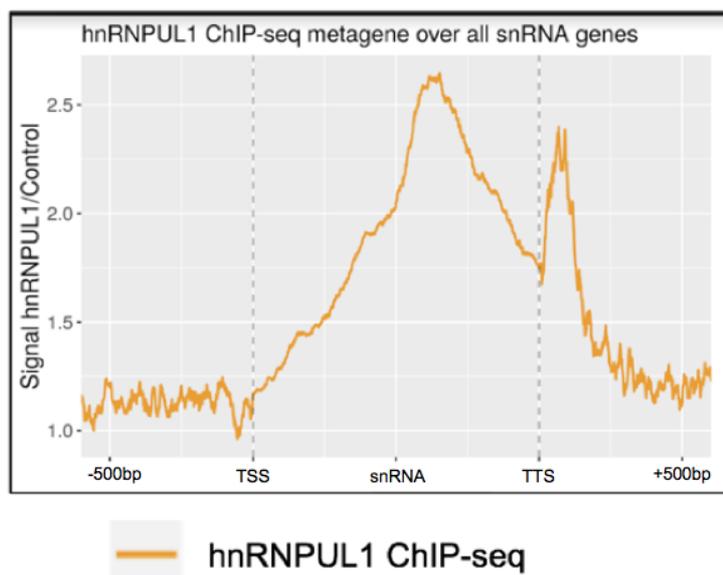
Figure 4.4 HNRNPUL1 interacts with RNAPII and ChIPs at snRNA loci

A – CoIP of HNRNPUL1 and RNAPII plus and minus DNase/RNase A treatment. B - Meta-gene analysis of HNRNPUL1 ChIP-seq signal at snRNA loci. TSS = transcription start site, TTS = transcription termination site.
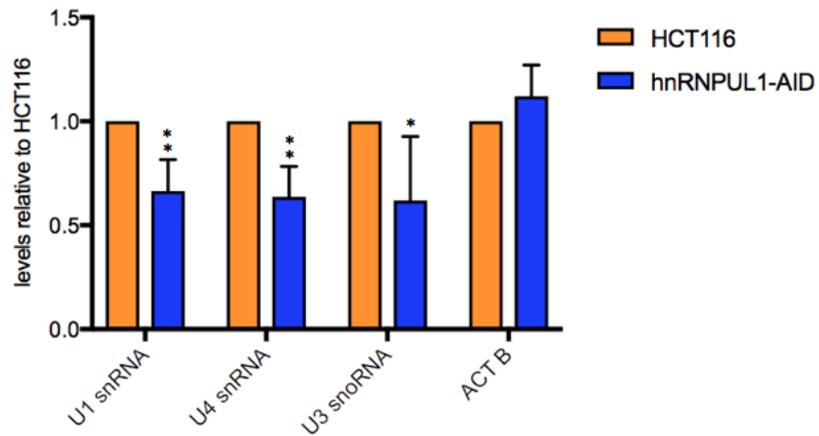
Figure 4.5 HNRNPUL1 is required for efficient snRNA transcription

ChIP-PCR assays performed using a RNAPII antibody in HCT116 and HNRNPUL1-AID lines. Errors bars represent the SD.

was exclusive to snRNA genes. This revealed no decrease in transcription at this locus, suggesting that this was indeed an snRNA gene-specific phenomenon. This indicates that HNRNPUL1 plays a direct role in the transcription of RNAPII-transcribed snRNA genes.

4.4 Analysis of snRNA transcription termination and 3'-end processing in the HNRNPUL1-AID line

As the HNRNPUL1 ChIP-seq meta-gene analysis suggests that HNRNPUL1 is present at the 3' end of snRNA genes, we hypothesised that HNRNPUL1 may specifically function in the stages of snRNA transcription and processing that occur at this region of genes – transcription termination and 3'-end processing. If this was the case, it is likely that depletion of HNRNPUL1 in the HNRNPUL1-AID line would induce observable defects in these processes. Further evidence supporting this notion came from the strong interaction between HNRNPUL1 and ARS2 reported by Hallais et al. in 2013 (Hallais et al., 2013). ARS2 binds pre-snRNAs and pre-snoRNAs and can recruit PHAX to promote nuclear transport of these transcripts, or alternatively associate with the NEXT complex to trigger their degradation. Hallais and colleagues also demonstrated that ARS2 knockdown results in elevated levels of 3'-end extended snRNAs, as well as increased RNAPII signal downstream of snRNA genes consistent with improper transcription termination (Hallais et al., 2013).

Therefore, ChIP primers were designed to amplify a region approximately 400 base pairs downstream of the end of the U1 snRNA coding region (Figure 4.6A), and the level of RNAPII occupancy at this locus was analysed relative to its abundance over the U1 snRNA promoter (to ensure that a reduction in overall transcription did not mask the phenotype), plus and minus HNRNPUL1. This revealed a mild increase in RNAPII signal in the HNRNPUL1 knockdown condition, however this increase was not statistically significant (Figure 4.6B).

Alongside the ChIP assays, the levels of 3'-end extended snRNA transcripts in the
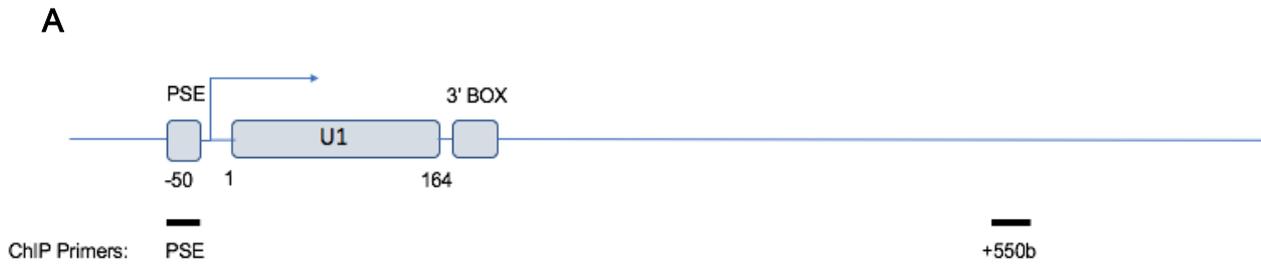
A

B
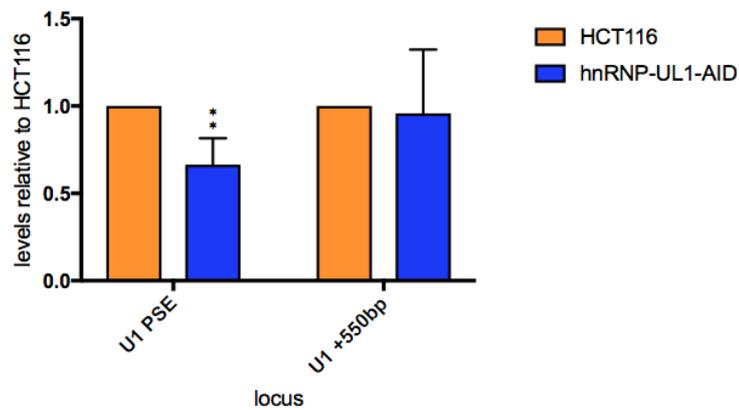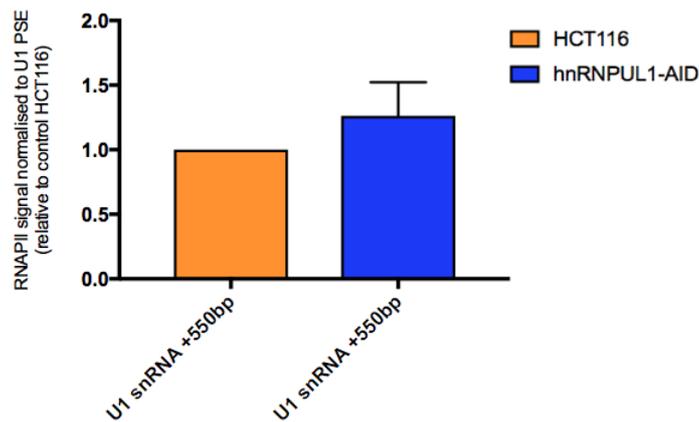
Figure 4.6 HNRNPUL1 knockdown does not result in significant snRNA transcription termination defect

A – Schematic illustration of U1 snRNA gene and the location of ChIP primers used.
B - ChIP-PCR analysis of RNAPII signal at U1 snRNA +550bp locus in HCT116 and HNRNPUL1-AID line. Errors bars represent the SD.

HNRNPUL1-AID line were measured via total RNA extraction and qRT-PCR, again using primers that amplified regions downstream of the U1, U2, U4 and U5 snRNA coding sequences (Figure 4.7A). Previous studies have demonstrated that knockdown of 3'-end snRNA processing factors such as DSIF and NELF result in a 3-6 fold increase in the levels of these uncleaved transcripts (Yamamoto et al., 2014) and therefore if HNRNPUL1 also functioned in this process one would expect to see similarly elevated levels in the HNRNPUL1-AID line upon HNRNPUL1 depletion. qRT-PCR analysis of these transcripts revealed a mixed picture, with a small increase observed in the amount of 3'-end extended snRNAs in the HNRNPUL1- AID line in the case of U1 and U2 snRNA, whereas there did not appear to be a phenotype in the case of U4 and U5 (Figure 4.7B). Once the proportion of uncleaved transcripts was normalised against the total level of each snRNA (Figure 4.7C), a small but statistically significant increase in the relative abundance of U1, U2 and U5 uncleaved transcripts was revealed, but still not the 3-6 fold increases that have been reported upon knockdown of other 3'-end processing factors.

Combined, the ChIP and total RNA extraction assays hint at a role for HNRNPUL1 in snRNA transcription termination and/or 3'-end processing. However, given the mild phenotypes demonstrated, we are unable to confidently conclude that HNRNPUL1 specifically functions in these processes at this stage.

4.5 Cajal bodies and SMN-containing nuclear Gems are disrupted upon HNRNPUL1 knockdown

It has been previously reported that Cajal bodies require ongoing snRNA transcription for their integrity (Lemm et al., 2006), with coilin ChIP-seq analyses also demonstrating that this core Cajal body component, like HNRNPUL1, is present across the body of all RNAPII–transcribed snRNA genes (Machyna et al., 2014). In addition, knockdown of several individual components of the snRNP biogenesis pathway such as PHAX, TGS1 or INTS11 has also been shown to be sufficient to disrupt Cajal bodies

A
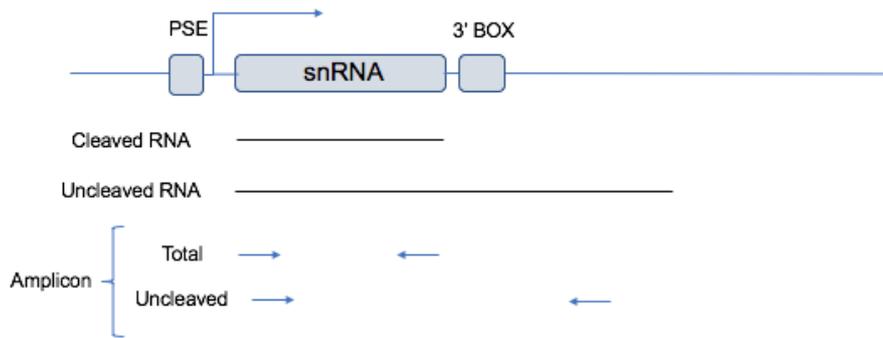
B

C

Figure 4.7 HNRNPUL1 knockdown results in small snRNA 3'-end processing defect

A - Schematic illustration of an snRNA gene and the location of qRT-PCR primers used. B – qRT-PCR analyses of total and uncleaved snRNA levels in the HNRNPUL1-AID line. Errors bars represent the SD. C - qRT-PCR analyses of uncleaved snRNA levels in the HNRNPUL1-AID line normalised to total snRNA level. Errors bars represent the SD.

(Lemm et al., 2006; Takata et al., 2012). Therefore, given that an snRNA transcription defect has previously been demonstrated in this chapter, the morphology of Cajal bodies in the HNRNPULI-AID line upon HNRNPUL1 depletion was analysed via immunostaining. At the same time, the effect of HNRNPUL1 knockdown on nuclear SMN-containing Gems was also assessed. These frequently co-localise with Cajal bodies within the nucleus, and their absence is a hallmark of the neurodegenerative disorder spinal muscular atrophy (SMA) as well as some SOD1-related forms ALS (Kariya et al., 2012). Furthermore, loss of SMN is associated with a downregulation of snRNA levels comparable to the decrease observed in the HNRNPUL1-AID line (Zhang et al., 2008). HCT116 and HNRNPUL1-AID cells were co-stained with coilin and SMN antibodies following auxin depletion of HNRNPUL1 and then examined under a microscope. This revealed a clear decrease in the number of Cajal bodies per cell in the HNRNPUL1-AID line, and remarkably almost a total loss of SMN-containing nuclear Gems (Figure 4.8A, Figure 4.8B).

To confirm this dysregulation of Cajal bodies and Gems, a coilin IP was carried out following depletion of HNRNPUL1 in the HNRNPUL1-AID line. The levels of a panel of snRNP biogenesis components co-immunoprecipitating with coilin in each condition was compared via Western blot. Firstly, the input fractions of these IPs demonstrated that coilin and SMN levels were not downregulated in the HNRNPUL1-AID line, indicating that this was not the reason for disintegration of Cajal bodies and Gems (Figure 4.9). Secondly, this revealed a clear disruption of the interaction between coilin and SMN upon HNRNPUL1 knockdown. Interestingly, coilin binding to SART3 – a key U4/U6 di-snRNP recycling factor – was also reduced, as well as its interaction with U1 snRNP (Figure 4.9). HNRNPUL1 also bound coilin in the HCT116 cell line, although the amount co-immunoprecipitating was significantly less than observed in the case of other snRNP biogenesis components (Figure 4.9). These IP data along with the immunostaining assays conclusively demonstrate that HNRNPUL1 is required for the integrity of Cajal bodies and SMN-containing nuclear Gems.
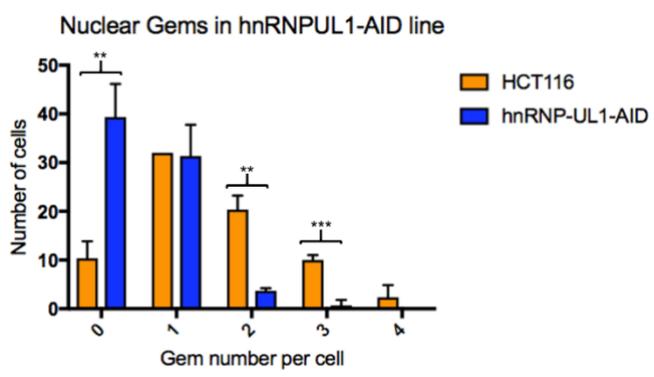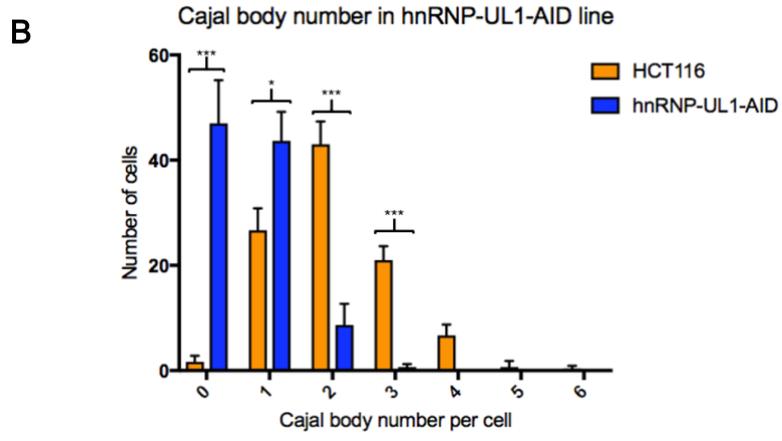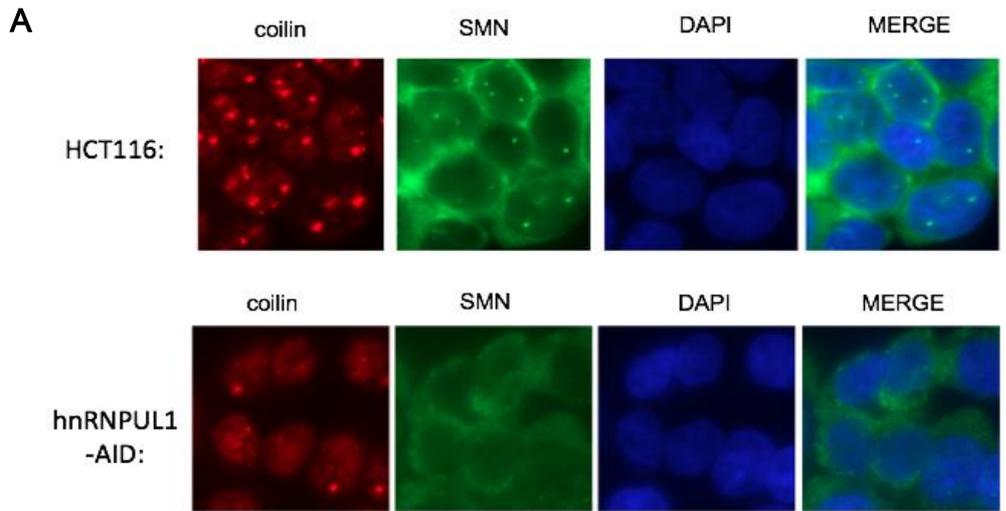
Figure 4.8 – HNRNPUL1 knockdown results in disintegration of Cajal bodies and loss of nuclear SMN-containing Gems

A – Immunostaining analyses in HNRNPUL1-AID line using anti-coilin and anti-SMN antibodies. Pictures were taken and Cajal bodies/Gems were counted by Dr. Helen

Knight. B – Tally chart quantification of Cajal body and Gem number decrease in HNRNPUL1-AID line. The experiment was repeated three times, with 100 cells counted per replicate.
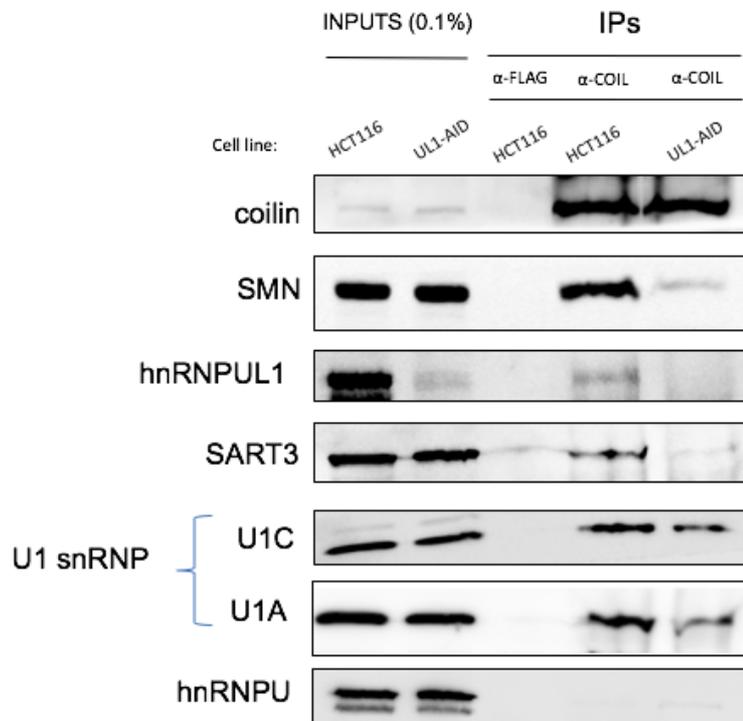
Figure 4.9 – Coilin interactome is disrupted upon HNRNPUL1 knockdown, while its abundance in the cell remains unaffected

Western analyses of coilin IP plus and minus HNRNPUL1, showing a clear decrease in the amount of snRNP biogenesis components co-immunoprecipitating with coilin as a result of HNRNPUL1 depletion. Lysates were not treated with RNase A during these IPs to ensure that RNA-dependent interactions could also be assessed.

4.6 HNRNPUL1 interacts with components of the snRNP maturation machinery but does not exclusively reside in Cajal bodies

While ongoing snRNA transcription is a prerequisite for Cajal body integrity, it has also been reported that Cajal bodies themselves are required for proper snRNA transcription (Wang et al., 2016). Therefore, it was plausible at this stage that HNRNPUL1 was in fact an essential Cajal body component, and that the Cajal body disintegration resulting from its depletion has the knock-on effect of reducing the levels of snRNA transcription. This would also be consistent with the overlap observed between the ChIP-seq profile of HNRNPUL1 and coilin. In addition, reduced levels of RNAPII-transcribed snRNAs observed in a HNRNPU auxin- inducible degron line generated in the Wilson lab were not sufficient to disrupt Cajal bodies. In fact, an increase in number per nuclei was observed (Wilson lab, unpublished data). Alternatively, if HNRNPUL1 regulated an snRNP biogenesis component known also to be essential for Cajal body integrity such as SMN or SART3 (Lemm et al., 2006, Novotný et al., 2015), then this could also account for the loss of Cajal bodies and subsequent snRNA transcription defect upon HNRNPUL1 depletion. Interestingly, analyses of publicly available RNA-binding protein eCLIP datasets by Ivaylo Yonchev revealed that, in contrast to other RBPs, HNRNPUL1 was most enriched on small Cajal body-specific RNAs (scaRNAs) (Figure 4.10), As their name suggests, this class of snoRNA permanently reside in the Cajal body, where they catalyse base modifications of snRNAs. The fact that it is this class of RNA that is most strongly bound by HNRNPUL1 suggests that this protein may spend at least some of its life within these nuclear bodies.

To identify whether HNRNPUL1 was indeed a Cajal body component, the localisation of HNRNPUL1 with respect to coilin was analysed via co-immunostaining HeLa cells with endogenous HNRNPUL1 and coilin antibodies. This revealed distinctive localisation patterns for the two proteins (Figure 4.11). HNRNPUL1 appears diffusely stained throughout the nucleus, with no clear concentration of the protein in
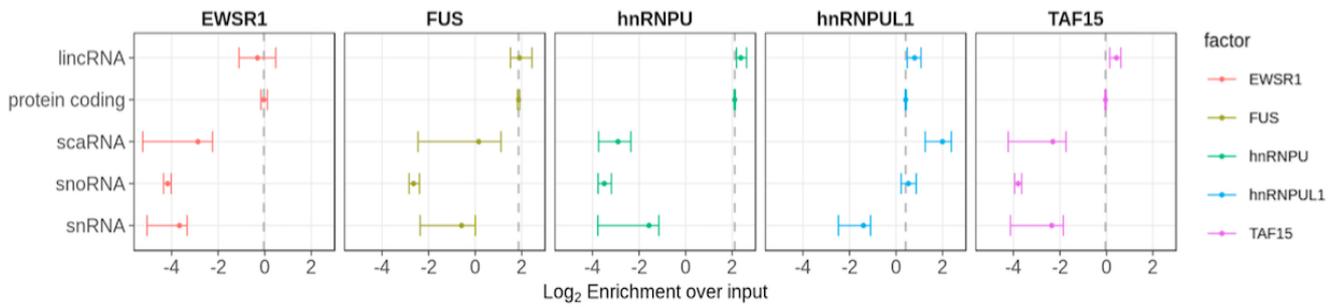
Figure 4.10 HNRNPUL1 interacts most strongly with Cajal body-associated RNAs

Analyses of the eCLIP profiles of HNRNPUL1 and a selection of RBPs produced by Ivaylo Yonchev, revealing a specific enrichment on scaRNAs in the case of HNRNPUL1.
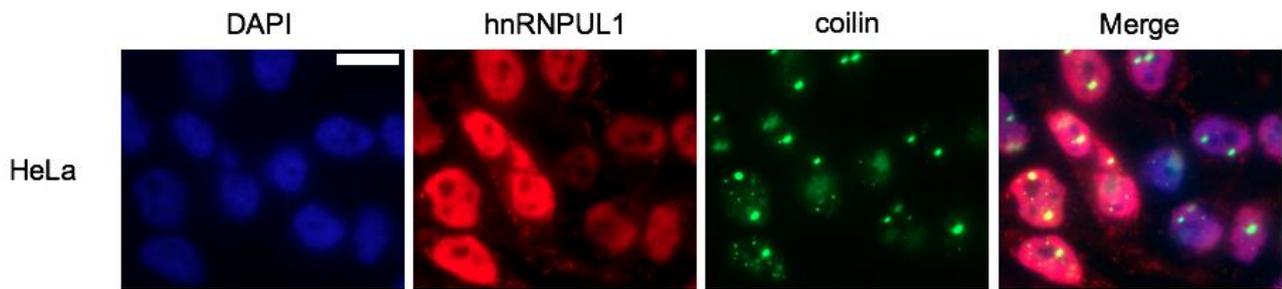
Figure 4.11 HNRNPUL1 does not concentrate in Cajal bodies

Immunostaining analyses using endogenous anti-HNRNPUL1 and anti-coilin antibodies, revealing distinctive localisation patterns for the two proteins.

Cajal bodies which were clearly identified as discrete nuclear bodies with coilin staining. However, HNRNPUL1 is clearly present in Cajal bodies, but is not exclusively localised there. Combined with the weak HNRNPUL1/coilin interaction observed in the coilin IP, this suggests that although HNRNPUL1 is found in Cajal bodies it is not a core component of these structures and probably plays many other nuclear roles not associated with Cajal bodies.

Although the interaction between HNRNPUL1 and coilin had been examined via IP and immunostaining, the relationship between HNRNPUL1 and other components of the snRNP biogenesis machinery so far affected by HNRNPUL1 knockdown such as SMN and SART3 was still unknown. Therefore, FLAG IPs using the stable FLAG cell lines generated in Chapter 3 were performed, incorporating FLAG-HNRNPUL1 WT, FLAG-HNRNPUL1 WA, and FLAG-HNRNPUL1 ΔCTD. Including the mutant cell lines would reveal the importance of nucleotide-binding and the C-terminal prion-like domain to any interactions identified. In addition, FLAG-HNRNPU was also immunoprecipitated during this experiment from a stable FLAG-HNRNPU line generated by a previous student in the Wilson lab (Figure 4.12A). This would enable the identification of snRNP maturation component interactions that were exclusive to HNRNPUL1, and therefore more likely to be relevant to the specific phenotypes observed in the HNRNPUL1-AID line. RNase A was not added to the lysate during these IPs, given that many of the protein complexes formed in these pathways are RNA-dependent.

Firstly, this revealed that HNRNPUL1 interacts with numerous snRNP proteins, with a clear band identifiable in the FLAG-HNRNPUL1 WT IP lane in the case of U1A (U1 snRNP), SF3B3 (U2 snRNP) and PRP31 (U4 snRNP) (Figure 4.12B). Although it was demonstrated previously in this chapter via RNA-IP that HNRNPUL1 interacts with snRNAs, that assay did not differentiate between nascent snRNA transcripts and snRNAs present within fully mature snRNP particles. The fact that HNRNPUL1 also binds snRNP proteins suggests that this protein does in fact bind to mature snRNPs.
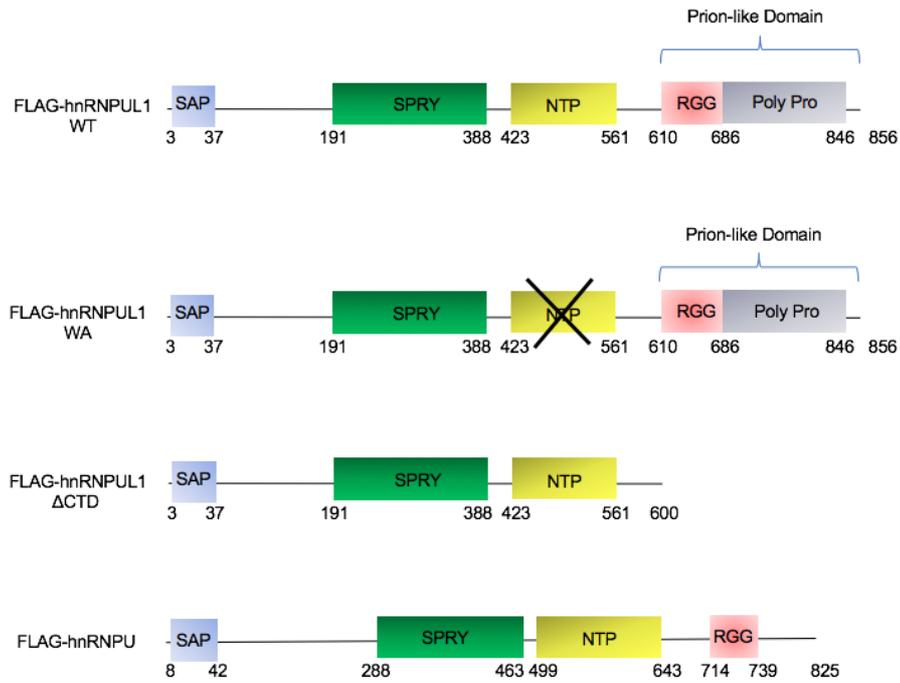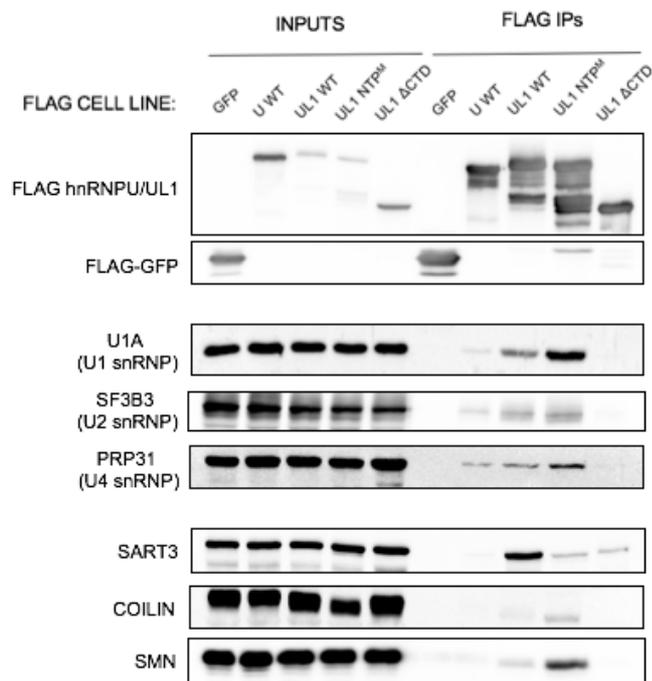
Figure 4.12 HNRNPUL1 interacts with snRNPs and snRNP biogenesis components
A – Schematic diagram of the HNRNPU/UL1 proteins included in the FLAG IP assay.
B – Western analyses following FLAG-HNRNPU and FLAG-HNRNPUL1 IP, demonstrating that HNRNPUL1 specifically interacts with snRNP components and snRNP biogenesis factors.

Surprisingly, the HNRNPUL1 Walker A mutant interacted with snRNP proteins more strongly than HNRNPUL1 WT for all three snRNP proteins tested (Figure 4.12B). In contrast, the ΔCTD mutation almost completely abolished the interaction between HNRNPUL1 and the snRNP proteins (Figure 4.12B). This pattern of increased interaction with the Walker A mutant and disrupted interaction with the ΔCTD mutant was also observed in the case of coilin and SMN (Figure 4.12B). These findings suggest that the prion-like domain located at the C-terminus of HNRNPUL1 is essential for forming many of the protein's interactions, including several components of the snRNP biogenesis pathway.

The interaction between HNRNPUL1 and coilin was weak, consistent with the endogenous coilin IP and coilin/HNRNPUL1 co-immunostaining performed as previously described in this chapter. So too was the interaction between HNRNPUL1 and SMN, suggesting that it is doubtful that HNRNPUL1 co-operates with or regulates the localisation of SMN during cytoplasmic snRNP assembly and re-import into the nucleus. Therefore, the loss of SMN-containing nuclear Gems observed upon HNRNPUL1 knockdown is more likely an indirect consequence of a defect in a different stage of the snRNP biogenesis pathway.

Interestingly, the strongest interaction observed was between HNRNPUL1 and SART3 (Figure 4.12B), with SART3 being highly enriched in the HNRNPUL1 WT IPed fraction relative to its input levels. Furthermore, this interaction appeared specific to HNRNPUL1, as there was considerably less SART3 present in the FLAG-HNRNPU IPed fraction (Figure 4.12B). The SART3/HNRNPUL1 interaction also appeared unique in the respect that, in contrast to the association with the snRNP components, the HNRNPUL1 Walker A mutation severely disrupted the interaction (Figure 4.12B). This suggests that HNRNPUL1 nucleotide-binding is required for this interaction to occur.

4.7 SART3 mis-localisation and U4/U6/U5 tri-snRNP assembly defect upon HNRNPUL1 knockdown

SART3 targets the U6 snRNA to Cajal bodies, promoting U4:U6 snRNA annealing, and therefore subsequent U4/U6 di-snRNP and U4//U6/U5 tri-snRNP formation within these structures (Bell et al., 2002; Staněk et al., 2003). In addition, SART3 is also required for the induction of Cajal body formation (Novotný et al., 2015). Given that HNRNPUL1 bound most strongly to U4, U5 and U6 snRNAs relative to U1 and U2 in the RNA-IP assay performed earlier in the chapter (Figure 4.1), the strong HNRNPUL1/SART3 interaction further pointed to HNRNPUL1 playing a direct role in this stage of snRNP biogenesis.

 To confirm that SART3 normally localises to Cajal bodies in HCT116 cells, and that this localisation would be disrupted as a result of the Cajal body disintegration observed upon HNRNPUL1 knockdown, HCT116 and HNRNPUL1-AID cells were co-immunostained with SART3 and coilin following doxycycline and auxin treatment. This demonstrated that SART3 does indeed concentrate in Cajal bodies in HCT116 cells, and that this co-localisation between SART3 and coilin is lost upon HNRNPUL1 depletion (Figure 4.13). This corroborates the coilin IP described earlier in the chapter, which revealed a clear disruption of the interaction between the two proteins in the HNRNPUL1-AID line (Figure 4.9).

 Although the U4/U6 di-snRNP and the U4/U6/U5 tri-snRNP can assemble in the nucleoplasm in the absence of Cajal bodies, it has been previously reported that this process occurs approximately 10-fold faster within Cajal bodies (Novotný et al., 2011). In light of the observed dysregulation of SART3 localisation and disintegration of Cajal bodies upon HNRNPUL1 knockdown, coupled with the finding that HNRNPUL1 strongly interacts with U4, U6 and SART3, we hypothesised that HNRNPUL1 depletion may also trigger defects in U4/U6 di-snRNP and U4/U6/U5 tri-snRNP assembly. To test this idea, RNA-IP was performed on U4 snRNP protein PRP31 (Figure 4.14A) in the HNRNPUL1-AID line following auxin depletion of HNRNPUL1.
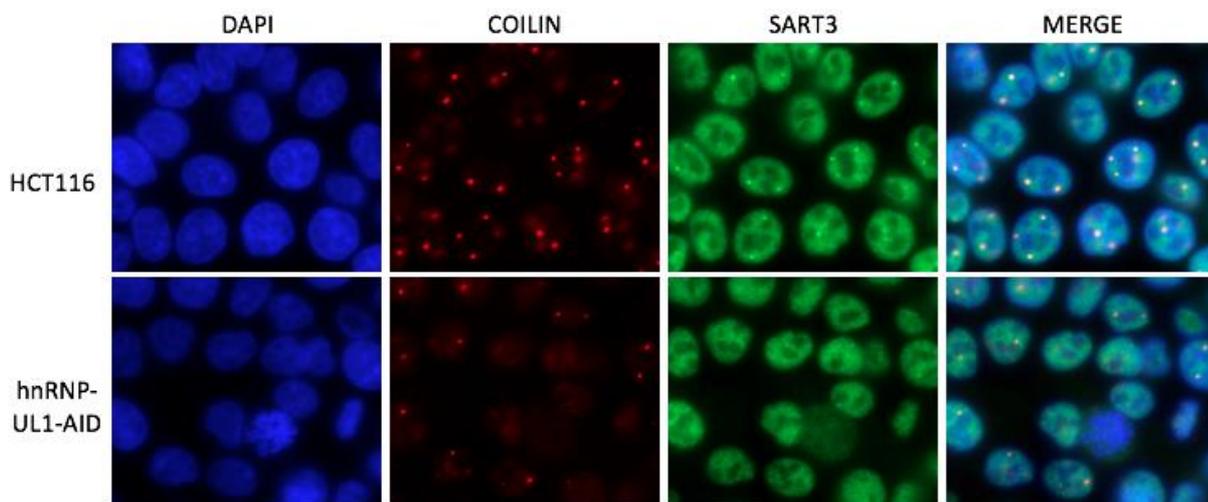
Figure 4.13 – Cajal body disintegration caused by HNRNPUL1 knockdown disrupts SART3 localisation

Immunostaining analyses using endogenous anti-SART3 and anti-coilin antibodies in the presence and absence of HNRNPUL1. SART3 normally concentrates in Cajal bodies in HCT116 cells, which are disrupted upon HNRNPUL1 knockdown.

A

B

PRP31 RIP in hnRNPUL1-AID line
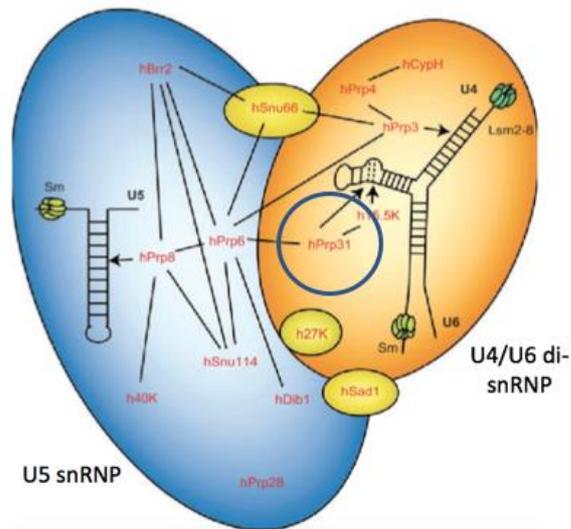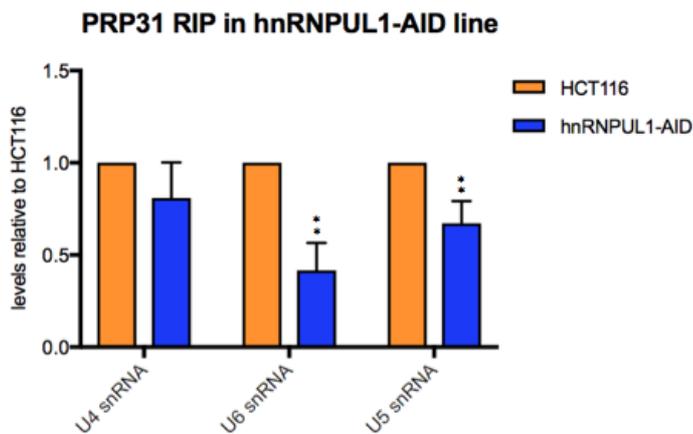
Figure 4.14 HNRNPUL1 knockdown causes defects in U4/U6 di-snRNP and U4/U6/U5 tri-snRNP assembly

A – Structure of the U4/U6/U5 tri-snRNP. PRP31 protein immunoprecipitated in RIP assay is circled. This figure is adapted from Bleichert and Baserga, 2010. B – RIP assay comparing U4, U6 and U5 levels co-immunoprecipitating with PRP31 in HNRNPUL1-AID line relative to HCT116 cells. snRNAs in each RIP sample were internally normalised to their input levels prior to the comparison of HCT116 and HNRNPUL1-AID conditions. Assays were performed three times, errors bars represent the SD.

This revealed that HNRNPUL1 knockdown results in a significant decrease in the amount of U6 snRNA and U5 snRNA co-immunoprecipitating with PRP31 (Figure 4.14B), indicating that U4/U6 di-snRNP and U4/U6/U5 tri-snRNP assembly is indeed disrupted in the absence of HNRNPUL1.

4.8 HNRNPUL1 depletion induces limited global splicing changes

To examine the impact of the defects in snRNP biogenesis so far discovered in the HNRNPUL1-AID line upon splicing at a transcriptome-wide level, we performed RNA-seq on nuclear poly (A)+ RNA extracted from cells following auxin depletion of HNRNPUL1. Triplicate sets of libraries were prepared and sequenced by Novogene, Hong Kong, in paired-end 150 mode on Illumina technology with ~25 million paired reads per sample. Reads were mapped to hg38 using STAR using default parameters. In order to assess splicing changes, gene models were obtained from Ensembl v85 and divided into minimally spanning transcript chunks using the gtf2gtf program from the CGAT suite and annotated as either constitutive exon, constitutive intron, annotated retained intron or alternate. Read counts were calculated using featureCounts from the subread package and differential chunk usage calculated using DEXSeq. Intron chunks with an adjusted p-value less than 0.1 and a log2 fold change between control and hnRNPUL1-AID > 1 were designated as significantly more retained. The pipeline used is available at https://github.com/sudlab/pipeline_retained_introns. To our surprise, limited splicing changes were detected, and in the case of all three classes of intron, decreased intron retention was observed upon HNRNPUL1 knockdown (Figure 4.15A). This indicates that the decrease in snRNA levels identified in the HNRNPUL1-AID line is not severe enough to trigger an immediate reduction in splicing efficiency, given that in fact there is more intron removal occurring in the HNRNPUL1 depleted cells. We have noticed some striking examples of HNRNPUL1 enrichment upon introns in the publicly available HNRNPUL1 eCLIP dataset, such as the introns surrounding FUS exon 7 (Figure

A
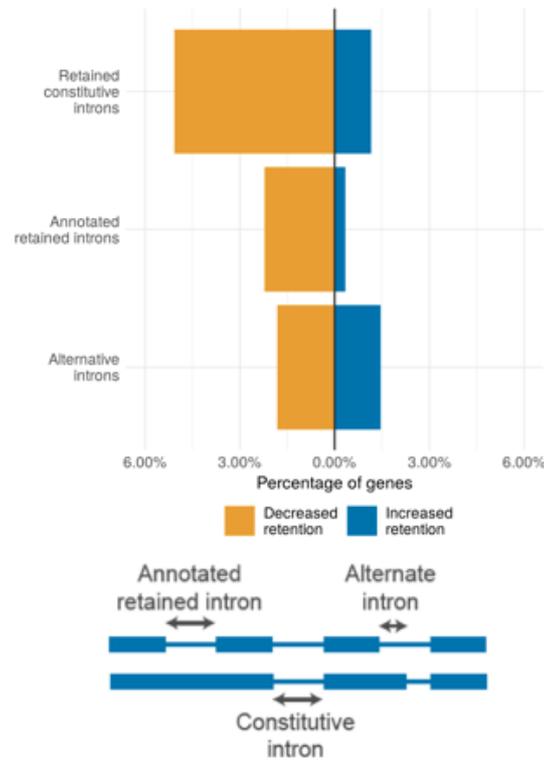
B
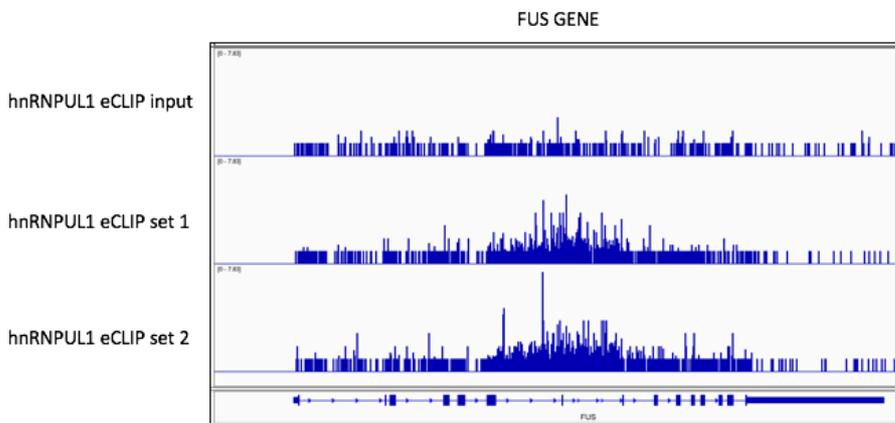
Figure 4.15 HNRNPUL1 knockdown triggers mild decrease in intron retention events

A – Effect of HNRNPUL1 knockdown upon intron retention. The number of gene products with evidence of an increase or decrease in intron retention (adjp<0.1;logFC>1) was measured as a fraction of gene products that contained introns of that type. Analysis performed by Ivaylo Yonchev. B – Example of HNRNPUL1 intron binding from publicly available HNRNPUL1 eCLIP dataset.

4.15B). Therefore, it is possible that HNRNPUL1 directly promotes the retention of introns to which it binds, and that its depletion would thus trigger decreased retention of these introns, counteracting any effects of the decrease in snRNP levels. However, the relationship between HNRNPUL1 intron binding and intron retention requires further global analysis to confirm this hypothesis. Alternatively, these splicing changes may be the product of increased RNAPII CTD phosphorylation that also occurs upon HNRNPUL1 depletion – a phenomenon that will be described in detail in Chapter 5.

4.9 snoRNA levels are downregulated upon HNRNPUL1 knockdown

As in the case of snRNAs, all snoRNAs traffic through Cajal bodies during their biogenesis pathway (Machyna et al., 2014), while scaRNAs reside within these structures permanently. Given the disintegration of Cajal bodies upon HNRNPUL1 depletion previously identified in this chapter, we speculated that snoRNA biogenesis would also be affected in the HNRNPUL1-AID line. This would also be consistent with the HNRNPUL1 eCLIP profile displayed in Figure 4.10, which revealed that HNRNPUL1 binds most strongly to these classes of RNAs. As there are approximately 750 known snoRNAs (Jorjani et al., 2016), we adopted a global approach to analyse the levels of these transcripts by performing small-RNA sequencing on RNA harvested from the HNRNPUL1-AID line following auxin depletion of HNRNPUL1. Triplicate nuclear 50-250bp size selected libraries were prepared and sequenced by Novogene, Hong Kong in single-end 150 mode on Illumina technology with ~12.5 million reads per sample. Transcript abundances were estimated by mapping to the RNAcentral non-coding RNA sequence database in order to avoid duplicated small RNA gene sequences, using bwa aln with parameters -l10 -k2 -n5. An average of 92% of reads were mapped across all samples. Reads were extracted using samtools idxstats and merged into a transcript abundance table which was then fed into DESeq2. A meta-gene averaging the signal density of all snoRNAs in this small RNA-seq sample in the HNRNPUL1-AID line compared to the HCT116 line revealed a significant, global

decrease in snoRNA levels upon HNRNPUL1 knockdown (Figure 4.16A). We also analysed the signal over all snRNA genes from the small RNA-seq samples, and as in the case of the snoRNAs, a clear reduction in abundance of these transcripts was observed (Figure 4.16B). This finding corroborates the decrease in snRNA levels identified by qRT-PCR that was detailed earlier in the chapter.

To confirm that this decline in small RNA levels was specific to these classes of transcript, the levels of total mRNA in the HNRNPUL1-AID line was examined from the poly (A)+ RNA sequencing data obtained previously. Transcript abundances were estimated using salmon against the Ensembl v85 annotation (hg38) and aggregated to gene level using tximport. Differential expression analysis was performed with limma using RLE normalization to account for gene length. Genes were considered differential upon showing a Bonferonni-Hochberg adjusted p value < 0.05 and an absolute fold change of > 1.5x. A meta-gene plot averaging the signal over all protein-coding genes demonstrated that HNRNPUL1 knockdown does not induce a global decrease in mRNA levels (Figure 4.17A), with in fact more mRNA transcripts upregulated (1715) than downregulated (1613) in the HNRNPUL1-AID line relative to HCT116 cells (Figure 4.17B). Interestingly however, the average abundance of mRNAs produced from pre-mRNAs containing snoRNAs within their introns significantly decreased in the HNRNPUL1-AID line compared to HCT116 cells (Figure 4.18), consistent with the decline in snoRNAs previously noted.

To validate the downregulation of snoRNAs observed in the small RNA-seq data, the levels of a panel of snoRNAs that displayed strong phenotypes in that dataset were analysed via qRT-PCR, from total RNA extracted from the HNRNPUL1-AID line following auxin depletion of HNRNPUL1. Primers were designed to measure the levels of mature snoRNAs, mature mRNAs produced from snoRNA-containing pre- mRNAs, and the levels of the precursor pre-mRNAs themselves (Figure 4.19A). Examples of both H/ACA and C/D snoRNAs were incorporated, as well as two independently transcribed snoRNAs – U3 and scaRNA2. These analyses confirmed the

A
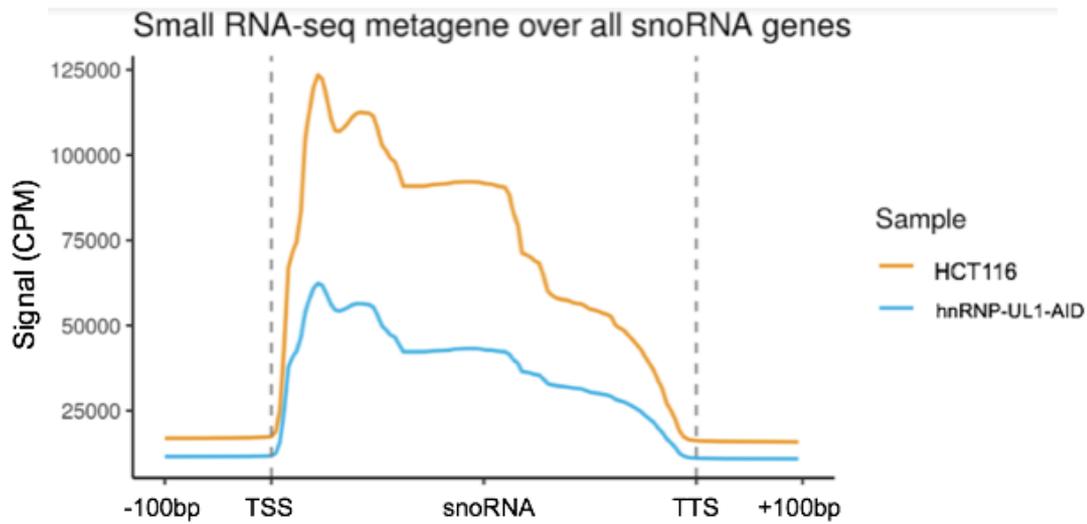
B
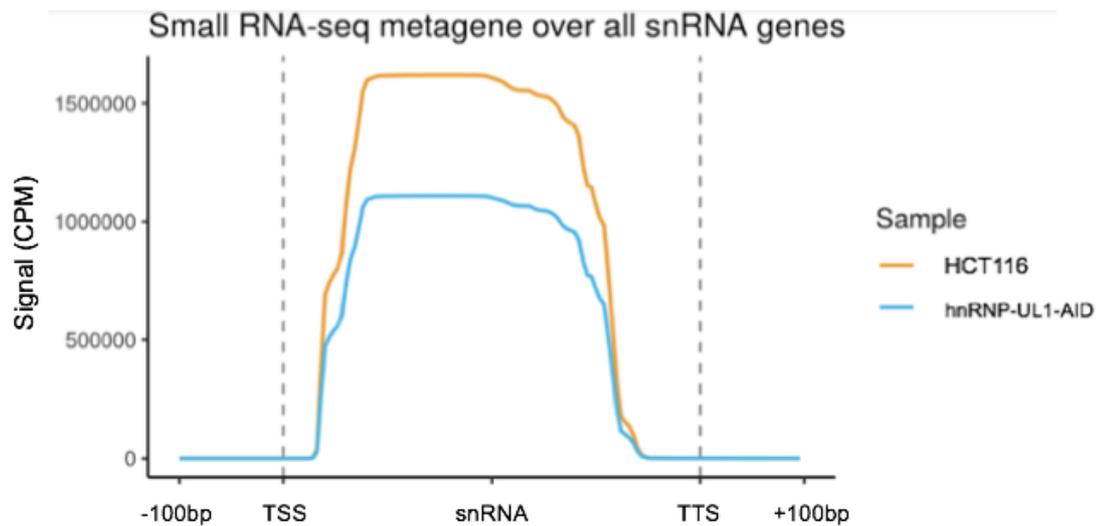
Figure 4.16 Global decrease in snoRNA and snRNA levels upon HNRNPUL1 knockdown

A - Meta-gene analysis of small RNA-seq signal averaged across all snoRNA genes. CPM = counts per million. B- Meta-gene analysis of small RNA-seq signal averaged across all snoRNA genes. CPM = counts per million. Analyses performed by Ivaylo Yonchev.

A – RNA-seq gene metagene over exons within all expressed, nonoverlapping proteincoding genes.
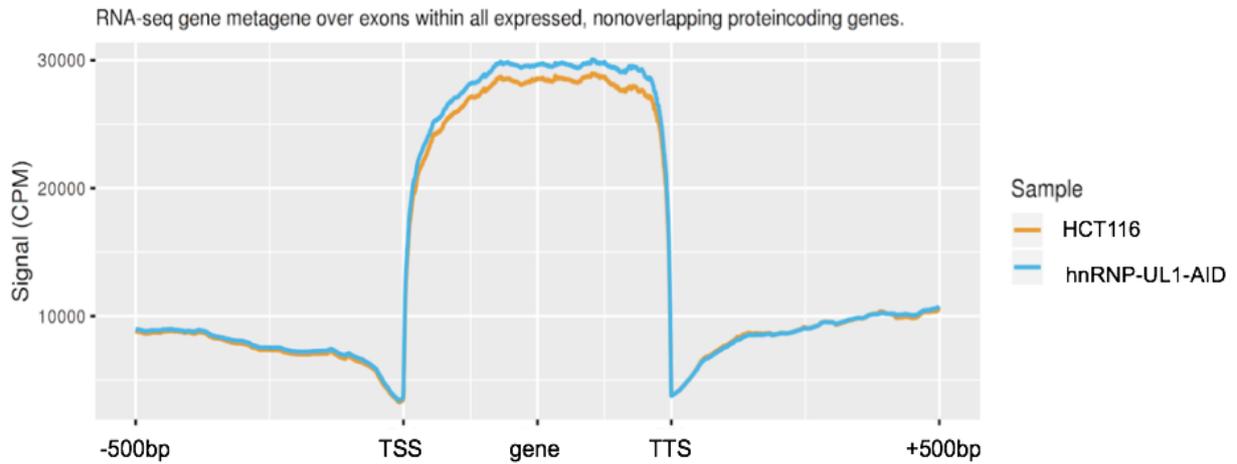
B – mRNA transcript expression changes in HNRNPUL1-AID line relative to HCT116

1613 >1.5x Down

1715 >1.5x Up
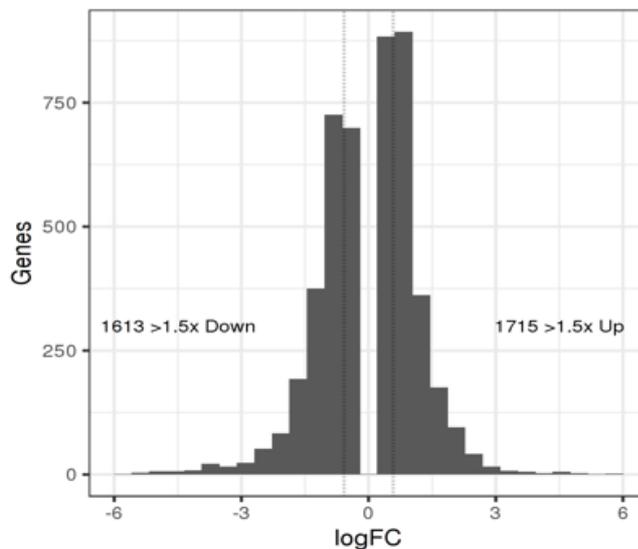
Figure 4.17 HNRNPUL1 depletion does not trigger global downregulation of mRNAs

A – Meta-gene analysis of poly (A)+ RNA-seq signal averaged across all protein-encoding genes. CPM = Counts per million. B – Histogram analysis of poly(A)+ RNA transcript expression changes in the HNRNPUL1-AID line following depletion of HNRNPUL1 relative to HCT116 control. The dotted line indicates a log2 fold change significance threshold of 0.5849625, indicating a 1.5 fold increase or decrease in gene abundance.

RNA-seq gene body metagene over all snoRNA containing genes.
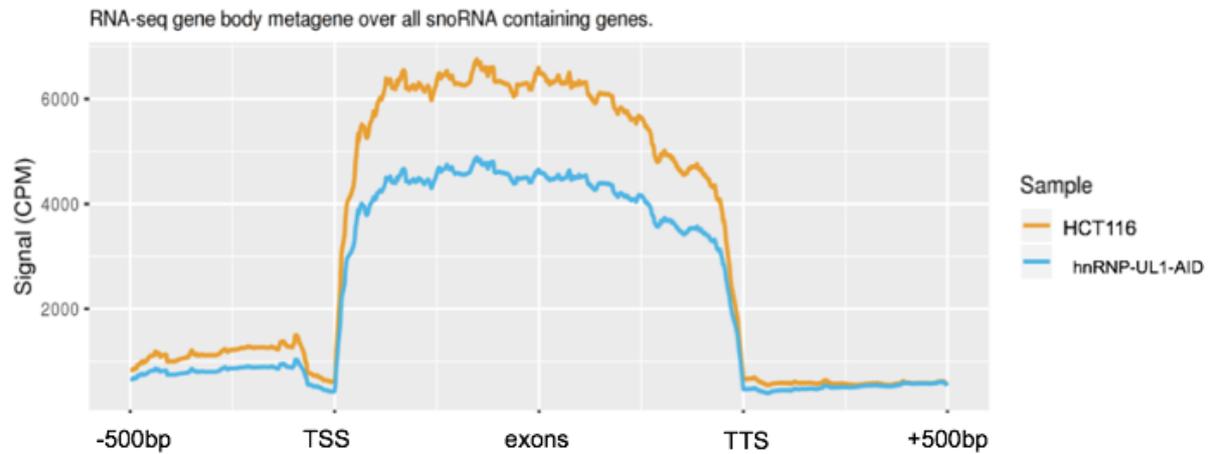
Figure 4.18 HNRNPUL1 depletion triggers specific downregulation of mRNAs produced from snoRNA-containing pre-mRNAs

Meta-gene analysis of poly (A)+ RNA-seq signal averaged across all snoRNA-containing genes. CPM = Counts per million.
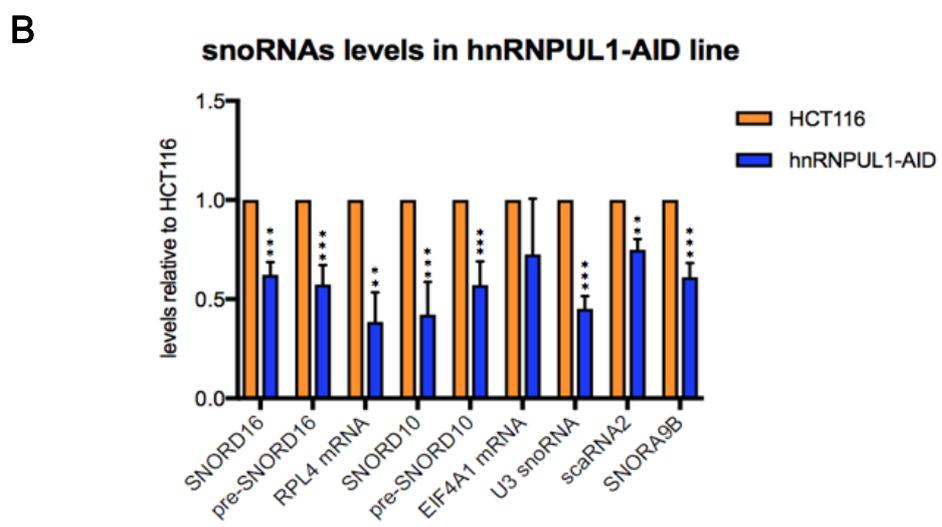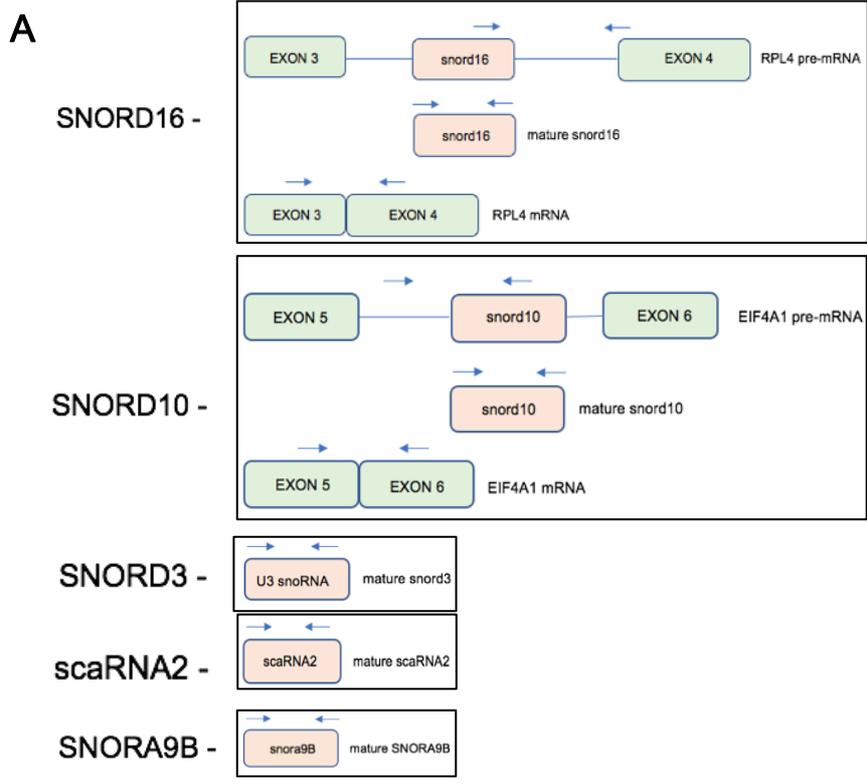
Figure 4.19 qRT-PCR validation of small RNA-seq data demonstrating decrease in snoRNA levels upon HNRNPUL1 knockdown

A - Schematic illustration of the snoRNAs analysed and the location of qRT-PCR primers used. B - qRT-PCR analyses of snoRNA levels in the HNRNPUL1-AID line. Errors bars represent the SD.

downregulation of snoRNA levels following HNRNPUL1 knockdown (Figure 4.19B). It appeared from these analyses that all types of snoRNA are affected by the depletion of HNRNPUL1, given that examples of H/ACA, C/D, scaRNA, and independently transcribed snoRNAs are all downregulated (Figure 4.19B). Although precursor transcripts levels were also reduced, mNET-seq analyses performed by Ivaylo Yonchev in the Wilson lab revealed that there is no global transcriptional downregulation of snoRNA-containing genes upon HNRNPUL1 depletion (Figure 4.20). This indicates that, in contrast to snRNAs, the snoRNA biogenesis defect resulting from HNRNPUL1 knockdown occurs post-transcriptionally.

4.10 HNRNPUL1 interacts with the NEXT complex responsible for degrading snRNA and snoRNA precursors

As HNRNPUL1 specifically binds snoRNA and scaRNAs, it is plausible that it plays a direct role in the processing of these species from their host pre-mRNA transcripts. However, if this were the case one might expect the levels of the precursor transcripts to increase upon HNRNPUL1 knockdown, whereas the levels of two snoRNA-containing pre-mRNAs analysed via qRT-PCR in the HNRNPUL1-AID line in fact declined (Figure 4.19). This finding, coupled with the knowledge that transcription is occurring at the same rate, suggests that these transcripts are subjected to increased degradation following HNRNPUL1 depletion. Under normal conditions, the fate of nascently-transcribed snoRNAs and snRNAs appears to be determined by competition between nuclear transport and degradation pathways. In the case of snRNAs and independently transcribed, capped snoRNAs such as U3 snoRNA, the CBC-ARS2 complex recruits PHAX to mediate their nuclear transport (Boulon et al., 2004). However this complex can also interact with the NEXT complex, composed of RBM7, ZCCHC8 and MTR4, which targets these RNAs to the nuclear exosome for degradation (Giacometti et al., 2017). In the case of uncapped snoRNAs located within introns, PHAX appears to be recruited in an ARS2-independent manner. Similarly,

mNET-seq exon metagene over all snoRNA containing genes.
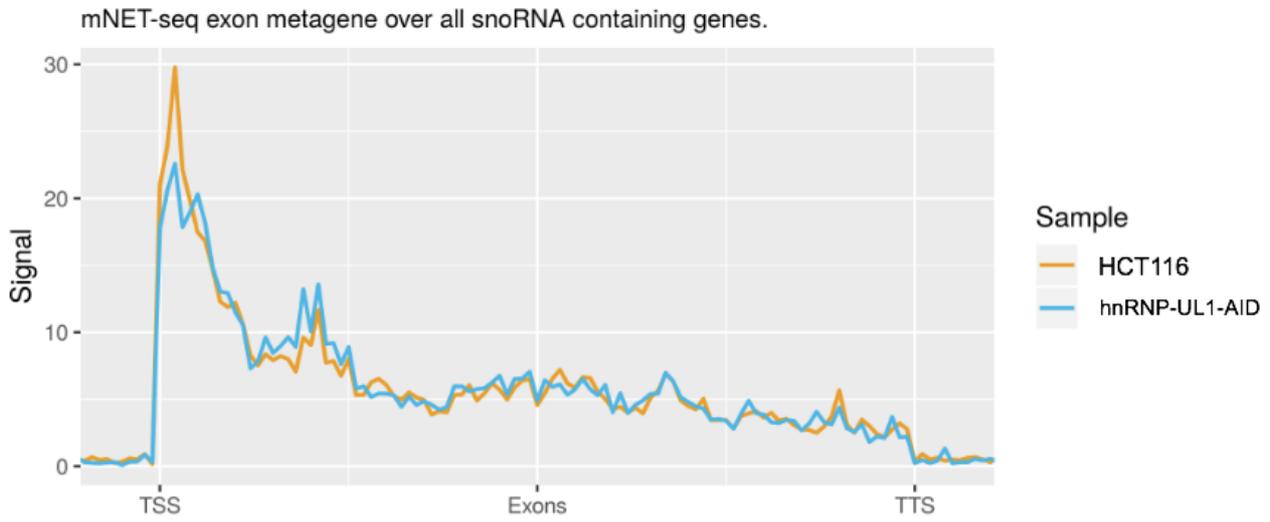
Figure 4.20 snoRNA-containing genes are not downregulated at the level of transcription upon HNRNPUL1 depletion

mNET-seq RNAPII signal across snoRNA-containing genes following mNET-seq analyses in HNRNPUL1-AID line performed by Ivaylo Yonchev.

RBM7 - the RNA-binding component of NEXT – also directly binds to pre-snRNAs (Hrossova et al., 2015) and to the 3'-end of snoRNA-containing introns, which are subsequently degraded up to the mature snoRNA by the exosome. The snoRNA itself is thought to be protected from further degradation via the assembly of snoRNP-specific proteins upon it, as well as PHAX recruitment. However, how this competition between assembly/transport and degradation is regulated is still not completely understood (Kufel and Grzechnik, 2019).

It has been previously demonstrated that all snoRNAs traffic through Cajal bodies en route to the nucleolus during their maturation (Machyna et al., 2014). The loss of Cajal bodies caused by HNRNPUL1 knockdown will therefore disrupt the snoRNA transport pathway, and this could potentially trigger increased targeting of these precursor transcripts to the nuclear exosome. Alternatively, it could be the case that aberrant, upregulated degradation of these snoRNAs via a different mechanism, combined with the snRNA downregulation observed upon HNRNPUL1 depletion, is what causes Cajal body disintegration in the first place.

To examine whether HNRNPUL1 associates with the NEXT complex and may therefore be directly involved in the regulation of snoRNA degradation, further FLAG IPs were performed utilising the stable FLAG-HNRNPUL1 WT, FLAG-HNRNPUL1 WA, FLAG-HNRNPUL1 ΔCTD and FLAG-HNRNPU cell lines (Figure 4.21A). Western blot analyses detected the presence of all three NEXT components in the FLAG-HNRNPUL1 WT immunoprecipitated fraction, with particular enrichment relative to its own input observed in the case of RBM7 (Figure 4.21B). Unlike the interaction with snRNP biogenesis components, HNRNPUL1 binding to the NEXT complex appears to be unaffected by the Walker A domain mutation (Figure 4.21B). The prion-like domain at the C-terminus of HNRNPUL1 is clearly required to maintain these interactions, although low level binding to RBM7 was observed with this mutant. ARS2 was also probed for, and this too was detected as a co-immunoprecipitant, however only weak binding to FLAG-HNRNPUL1 WT was observed in this case (Figure 4.21B).
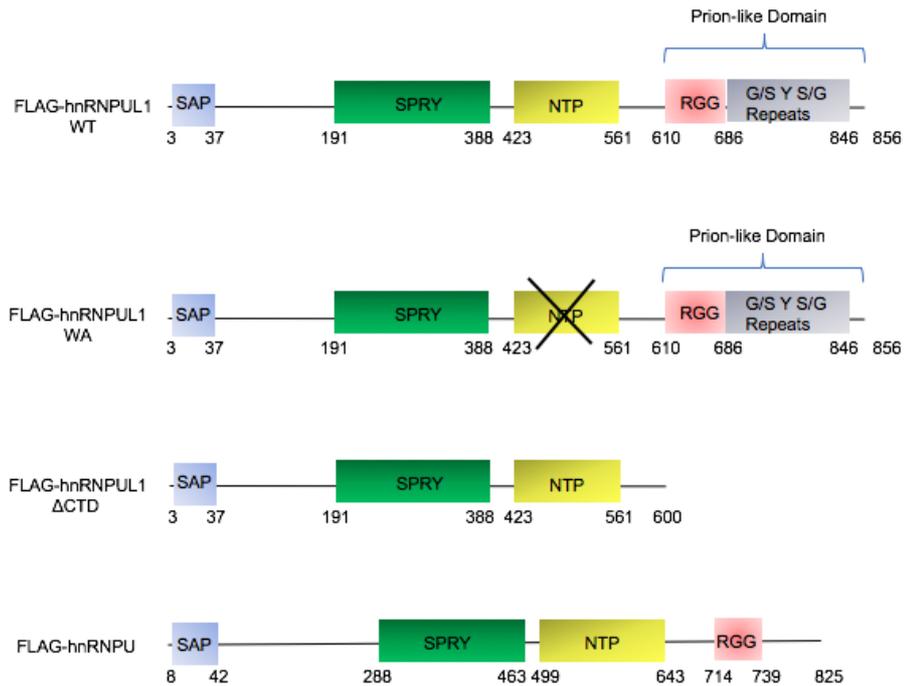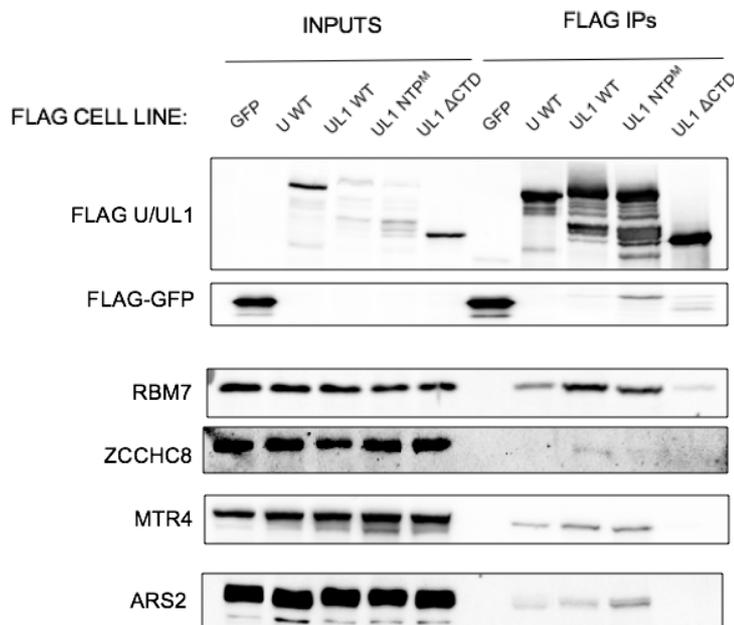
Figure 4.21 HNRNPUL1 interacts with the NEXT complex

A – Schematic diagram of the HNRNPU/UL1 proteins included in the FLAG IP assay. B – Western analyses following FLAG-HNRNPU and FLAG-HNRNPUL1 IP, demonstrating that HNRNPUL1 specifically interacts with the NEXT complex.

This was somewhat surprising given the strong interaction reported between the two proteins previously described (Hallais et al., 2013), but consistent with the lack of severe snRNA transcription termination phenotype upon HNRNPUL1 knockdown described previously in this chapter. This weak interaction implies that it is unlikely that HNRNPUL1 directly mediates the interaction between CBC-ARS2 and NEXT.

As we have shown that HNRNPUL1 is specifically enriched on snoRNAs (Figure 4.10), the strong interaction with RBM7 – the RNA-binding component of NEXT that specifically associates with the precursors of these RNAs - is of particular note. Whether HNRNPUL1 may prevent RBM7/NEXT from accessing these substrates, or disrupt its ability to form an active complex with the exosome, are intriguing possibilities to be investigated in the future.

4.11 Summary

Throughout this chapter, the HNRNPUL1-AID line and stable FLAG-HNRNPUL1 lines generated as described in Chapter 3 have been utilised to investigate the relationship between HNRNPUL1 and small RNA biogenesis. It has been demonstrated that the presence of HNRNPUL1 at RNAPII-transcribed snRNA loci is necessary to maintain the efficient transcription of these genes, with the downregulation of transcription upon HNRNPUL1 depletion resulting in a reduction in overall abundance of RNAPII-transcribed snRNAs. Consistently, we have shown that Cajal bodies disintegrate upon HNRNPUL1 knockdown, resulting in the mislocalization of snRNP biogenesis components SMN and SART3 within the nucleus. HNRNPUL1 strongly interacts with SART3 in a nucleotide binding-dependent manner, promoting the efficient assembly of the U4/U6 di-snRNP and U4/U6/U5 tri-snRNP. We have also identified the prion-like domain located at the C-terminus of HNRNPUL1 as fundamental in enabling the protein to interact with several components of this cellular pathway. The disruption to the snRNP biogenesis pathway appears to have limited immediate consequences for splicing within the cell, although a small number of intron retention changes were identified.

It has also been revealed in this chapter that HNRNPUL1 specifically interacts with all classes of snoRNA, and that loss of the protein results in lowered expression of these RNAs as well as their precursor transcripts. As we demonstrate no transcriptional downregulation of these genes upon HNRNPUL1 depletion, this indicates a role for HNRNPUL1 in specifying the post-transcriptional fate of these RNAs. In agreement with this notion, a strong interaction between HNRNPUL1 and NEXT component RBM7 – responsible for targeting snRNA and snoRNA precursors for degradation via the nuclear exosome complex – was uncovered, again mediated by the prion-like domain located at the C-terminus of HNRNPUL1.

CHAPTER 5 – HNRNPUL1 AND ALS

In addition to investigating the cellular functions of HNRNPUL1, we also aimed to further understand its relationship with the neurodegenerative condition ALS over the course of this study. Therefore, in this chapter the association between HNRNPUL1 and other ALS-causing RBPs is studied in detail. In addition, a further investigation into the function of the prion-like domain located at the C-terminus of HNRNPUL1 will be performed, alongside assays assessing the importance of the nucleotide-binding ability of HNRNPUL1. Finally, the impact of HNRNPUL1 mutations present in ALS patients, uncovered by our collaborators at SITraN, will be explored. Of particular interest is how these mutations affect the cellular functions of HNRNPUL1 discovered in Chapters 3 and 4, namely the regulation of immediate-early gene expression and small RNA biogenesis.

5.1 HNRNPUL1 IP/mass spectrometry identifies numerous ALS-causing RBPs as interactors

The FLAG-HNRNPUL1 IPs performed in the previous chapter revealed that the Walker A and C-terminal truncation mutations strongly affect many of the interactions between HNRNPUL1 and its protein binding partners. To assess the impact of these mutations upon the entire interactome of HNRNPUL1, FLAG-HNRNPUL1 WT, FLAG-HNRNPUL1 WA and FLAG-HNRNPUL1 ΔCTD were all immunoprecipitated from the stable FLAG cell lines and the IPed fractions were then subjected to mass spectrometry analysis. In addition, FLAG-GFP was immunoprecipitated from a stable FLAG-GFP cell line, and the IPed fraction was also analysed via mass spectrometry as a negative control. One replicate was performed per condition. RNase A was added to the lysates during the IP in order to identify direct protein interactors, rather than interactions bridged by RNA. 5% of each IPed sample was first assessed via Coomassie staining prior to mass spectrometry analysis. This revealed high yields of the expected IPed protein in each condition (Figure 5.1), and therefore the rest of
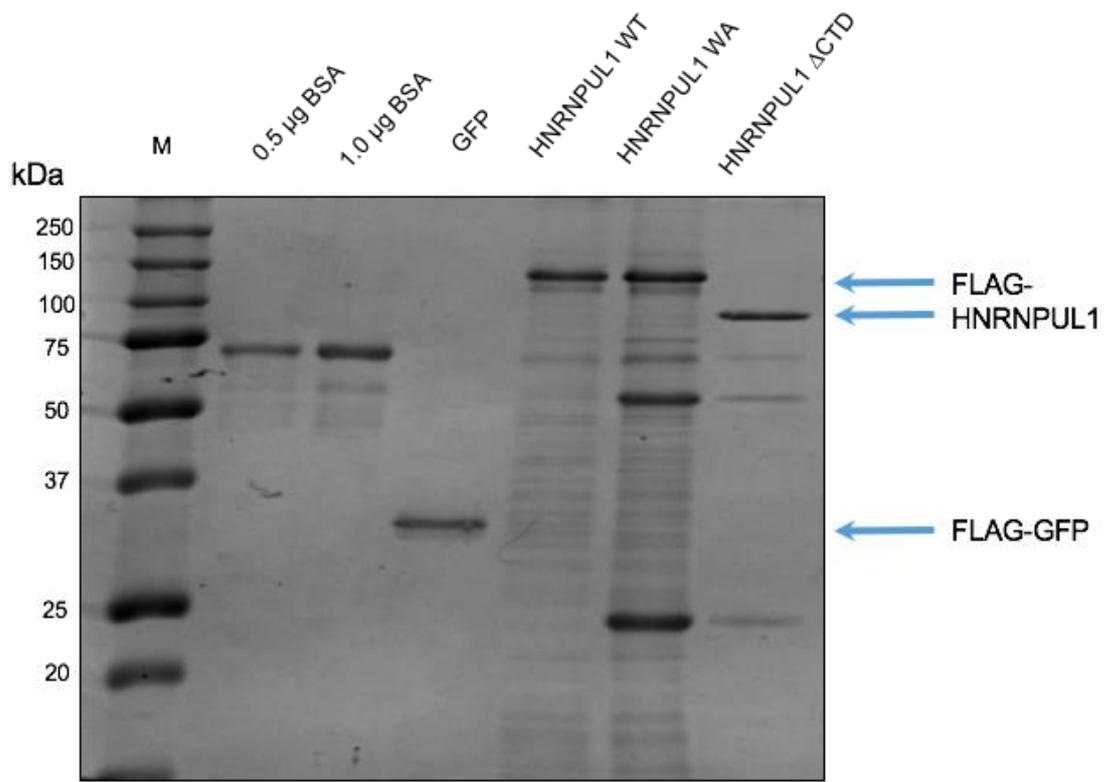
Figure 5.1 Immunoprecipitation of FLAG-GFP, FLAG-HNRNPUL1 WT, FLAG-HNRNPUL1 WA and FLAG-HNRNPUL1 ΔCTD for mass spectrometry analysis

Coomassie staining of FLAG-GFP, FLAG-HNRNPUL1 WT, FLAG-HNRNPUL1 WA and FLAG-HNRNPUL1 ΔCTD IPed fractions prior to mass spectrometry analysis. FLAG-tagged proteins were purified from Flp-In T-REx 293 cells via anti-FLAG M2 agarose beads and eluted using arginine-HCl (pH3.5) solution.

each sample was incorporated into the mass spectrometry analyses. Detailed methodology of the mass spectrometry procedure can be found in the 'FLAG-tagged Protein Immunoprecipitation (IP) for Mass Spectrometry analysis' section of the Materials and Methods chapter. Peptides were identified using MaxQuant software and interactors were ranked according to iBAQ score (the sum of intensities of all tryptic peptides for each protein divided by the number of theoretically observable peptides). 59 proteins were identified in the FLAG-HNRNPUL1 WT IPed fraction (Figure 5.2). The majority of these interacting proteins were other RNA-binding factors, with functions relating to various stages of gene expression such as transcription and splicing, or RNA stability and transport (Figure 5.3). In addition, a large number of ribosomal proteins were present, suggesting that HNRNPUL1 performs a previously unreported role in ribosomal assembly or protein translation. snRNP proteins were also identified, consistent with the findings of the previous chapter linking HNRNPUL1 to these complexes. Interestingly, many of the HNRNPUL1 interactors with the highest iBAQ scores - HNRNPA2B1 (1st), FUS (2nd), HNRNPA1 (4th), EWSR1 (10TH), TAF15 (28th) - are DNA/RNA-binding factors that have been previously associated with ALS, providing further evidence that HNRNPUL1 may operate in a common pathway with these proteins.

5.2 HNRNPUL1 Walker A mutant displays increased interactome and is enriched on chromatin

A much larger group of interacting factors was identified via mass spectrometry in the case of the FLAG-HNRNPUL1 Walker A mutant compared to HNRNPUL1 WT (116 hits versus 59 hits in the case of the wild-type) (Figure 5.4). Almost every functional group of interacting proteins increased in number as a result of this mutation (Figure 5.5), suggesting that preventing HNRNPUL1 from binding nucleotides improves its ability to form protein-protein interactions in a general fashion. In addition, FLAG-HNRNPUL1 WA bound more strongly to its interactors compared to the wild-type, as

Gene names	Number of peptides	iBAQ
HNRNPUL1 WT	62	764635000
HNRNPA2B1	25	48046000
FUS	11	32775000
RPL29	5	25476500
HNRNPA1;HNRNPA1L2	17	24029000
RPLP2	8	13016000
ERH	3	10837450
HNRNPA3	13	10416500
RPL6	12	10041150
RBMX	18	9799000
EWSR1	5	9736600
HNRNPC	14	9392900
RABEP1	2	8246650
YBX1	5	7341600
HNRNPH1	8	6368300
RPS28	4	6263000
SNRPA	7	6066150
HIST1H1E;HIST1H1D	7	4525850
RPL12	5	4509550
TXNDC12	7	4407900
HIST1H2BN;HIST1H2BL	4	4339300
RBM14	14	4251900
HIST1H2AJ;HIST1H2AH	5	4048100
HNRNPAB	8	4016150
HNRNPH3	5	3461550
RPLP0;RPLP0P6	8	3246200
SNRPE	3	3176500
HNRNPL	7	2995350
TAF15	2	2674500
RBM3	3	2478300
HNRNPA0	4	2184800
YBX3	6	2140450
HNRNPD	4	1954550
HSPA1B;HSPA1A	14	1831750
CLNS1A	2	1811400
ARL6IP4	4	1740150
CIRBP	5	1576450
HNRNPK	10	1397450
RPL27	3	1322900
HNRNPDL	6	1269400
RPL9	3	1244400
RPL14	2	1230500
RPS10;RPS10-NUDT3	3	1148750
HNRNPF	4	1127805
SNRPD2	4	1096700
PRPS1;PRPS1L1	5	1055565
RPL21	2	1040250
IGKV A18;IGKV2D-26	2	934815
ELAVL1	4	829825
HSPA8	10	803685
RPS3	5	778070
RPL5	3	448045
DDX17	6	394250
DDX5	5	383275
ACTG1;ACTB	2	222875
HNRNPR	3	173230
PRPSAP2	2	165585
HNRNPM	3	150128.5
SRSF1	2	142145
PLOD1	3	65577.5

Figure 5.2 RNA-independent interactors of HNRNPUL1 WT

Table of HNRNPUL1-associated proteins identified by mass spectrometry, ranked by iBAQ score.
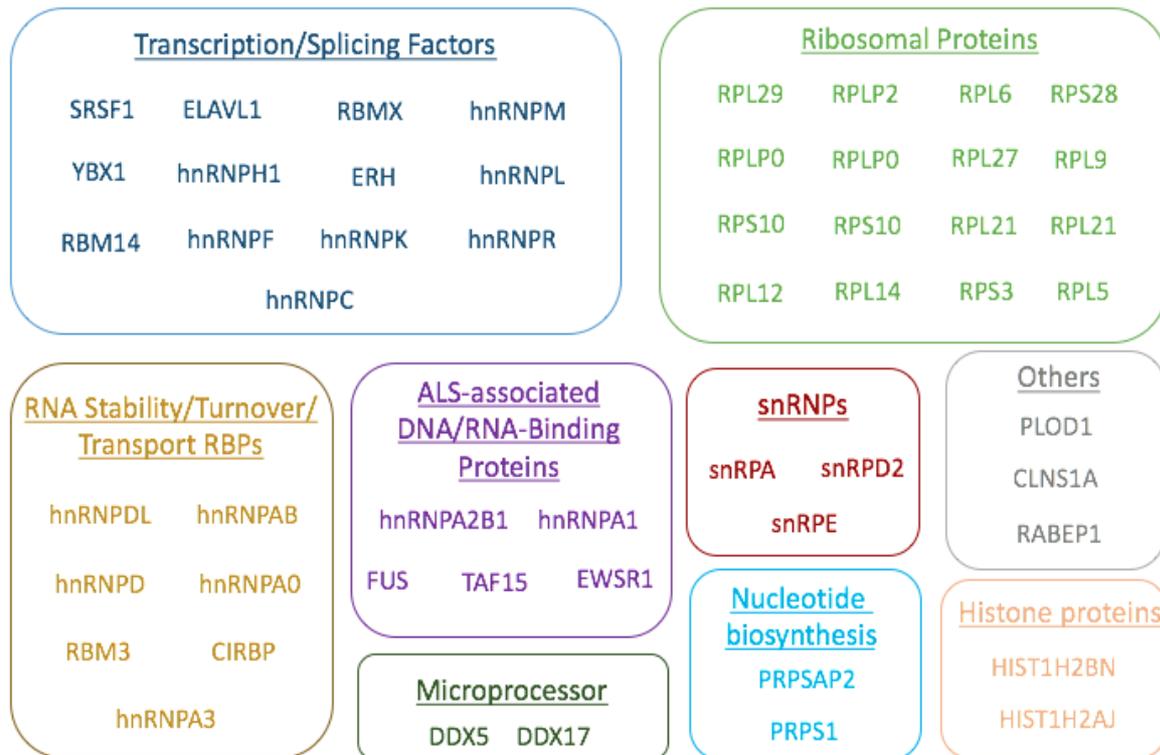
Figure 5.3 RNA-independent interactors of HNRNPUL1 categorised by function

Summary of HNRNPUL1-associated proteins, identified via mass spectrometry following FLAG IP in the presence of RNase A. Mass spectrometry analysis was performed by Mark Dickman and Caroline Evans. Proteins identified in the FLAG-HNRNPUL1 WT sample that were also present in the FLAG-GFP sample have been removed from this visualisation, along with the contaminant protein Actin.

Gene names	Number of Peptides	iBAQ
HNRNPUL1 WA	51	242885000
HNRNPA2B1	32	158335000
RPLP2	14	103710500
FUS	15	70994500
RPL29	5	57098000
HNRNPC	20	56220500
HNRNPA1;HNRNPA1L2	25	54429000
HNRNPA3	24	44878500
EWSR1	8	35438500
RPLP1	2	34175500
HNRNPH1	17	30804500
RPL6	15	21974500
RBMX	25	19555000
HIST1H1E;HIST1H1D	12	17666000
RPS28	6	16767000
HNRNPH3	9	15366500
ERH	4	14931000
HNRNPL	15	13994000
RPLP0;RPLP0P6	14	13545500
HNRNPAB	11	13344000
RPL21	6	13327000
HIST1H2AJ;HIST1H2AH	6	12393000
SNRPA	8	12365500
RBM14	20	11490500
SRSF3	3	11409500
SRSF7	3.5	10825000
RPL12	6	10783550
SNRPE	5.5	10657500
TAF15	4	10609500
YBX1	5	9449800
HNRNPD	8	8715800
SNRPD2	10	8600400
RPL14	3	8289350
CIRBP	9	7784850
RBM3	3	7554150
HNRNPA0	12.5	7413300
HNRNPDL	13	7377100
HNRNPK	7	6630650
HIST1H2BN;HIST1H2B	4	6094400
RPL5	10	5591300
YBX3	10	5509800
RPL9	5	5053700
CLNS1A	7.5	4684750
RPL34	6.5	4659800
KHDRBS1	6	4359550
RPL13	8	4344300
RPS14	4	4301300
HNRNPF	8	4159800
RPS5	8	3920250
RPS10;RPS10-NUDT3	7	3871100
ELAVL1	4	3782450
RPL11	3	3449150
SNRPG;SNRPGP15	3	3315900
RPS11	2	2705600
RPS8	6	2637200
RPS3	10	2542250
SNRPD3	2	2495550
HNRNPR	13	2448900
SRSF2	3	2351550

(Figure 5.4, continued on next page)

RPL27	4	2205700
PCNP	4	2122600
ARL6IP4	8	2120050
RPS20	17	2073300
RPL7A	3	2057850
RPS3A	8	2055500
HNRNPU	7.5	2042350
HSPA1B;HSPA1A	3.5	2040950
RALY	10	1982500
RPS19	2	1899700
IGKV A18;IGKV2D-26;	2	1888500
HNRNPC	4	1859650
SNRPN;SNRPB	2	1851400
SNRPC	9.5	1793200
TXNDC12	5	1513650
RPS12	2	1422750
HSPA8	14	1281050
DDX17	11	1084285
SF3B4	3	1081200
IGF2BP1	11	991440
PABPN1	4	929070
PABPC1	2.5	892725
ILF2	3.5	864685
SRSF1	4	754050
ZRANB2	3	749385
ALYREF	9	740300
SF3A1	12	717810
MRPS23	2.5	711865
RPL24	3	682225
PRMT1	4	664975
RPS4X	4	644385
RPSA	3	590285
DDX5	5	559850
RPL7	3	537830
ACTG1;ACTB	3	494120
MRPS28	3	478025
PRPS2	2	477050
C14orf166	3	434975
RPS13	2	410090
CKMT1A;CKMT1B	2.5	386790
PRPS1;PRPS1L1	3	376715
KHSRP	3	329285
SRSF9	2	306930
SPIN1	6.5	291620
DDX3X;DDX3Y	7	258540
ILF3	2	251205
PLOD1	7	233070
SYNCRIP	2	215925
HNRNPDL	2	214300
PRPSAP2	3	210975
CLASP2	7	179380
BCLAF1	2	177890
SF3B2	3	160670
XRN2	3	87390
HSPA5	2	83030.5
RBM28	2	83030
MAP1B	5	59770

Figure 5.4 RNA-independent interactors of HNRNPUL1 Walker A mutant

Table of HNRNPUL1 Walker A mutant-associated proteins identified by mass spectrometry, ranked by iBAQ score.

Figure 5.5 RNA-independent interactors of HNRNPUL1 Walker A mutant categorised by function

Summary of HNRNPUL1 Walker A mutant-associated proteins, identified via mass spectrometry following FLAG IP in the presence of RNase A. Mass spectrometry analysis was performed by Mark Dickman and Caroline Evans. Proteins identified in the FLAG-HNRNPUL1 WA sample that were also present in the FLAG-GFP sample have been removed from this visualisation, along with the contaminant protein Actin.

evidenced by the reduced ratio of bait to prey iBAQ scores in this IPed fraction (Figure 5.6A, Figure 5.6B). Only one protein was identified in the wild-type mass spec that was absent in the Walker A mass spec – RABEP1, a factor involved in ER to Golgi trafficking.

Given the clear increase in binding between the FLAG-HNRNPUL1 Walker A mutant and a range of chromatin-associated transcription/pre-mRNA processing factors relative to the wild-type form of the protein, we hypothesised that the cellular localisation of this mutant would display an increased chromatin bias. To test this possibility, fractionation of cells expressing FLAG-HNRNPUL1 WT and FLAG-HNRNPUL1 WA was performed in order to isolate chromatin, nucleoplasm and cytoplasmic fractions. The abundance of HNRNPUL1 WT and HNRNPUL1 WA in each fraction was then assessed via Western blot analysis. SSRP1, TUBULIN and HISTONE-H3 were used as markers to assess the efficiency of fractionation. Consistent with the previous IP results, FLAG-HNRNPUL1 WA was significantly more enriched in the chromatin fraction compared to FLAG-HNRNPUL1 WT (Figure 5.7).

5.3 The prion-like domain of HNRNPUL1 is required for the majority of its interactions

In contrast to the Walker A mutation, the C-terminal truncated form of HNRNPUL1 displayed a severely reduced interactome, with only 16 proteins identified via mass spectrometry in this IPed fraction (Figure 5.8, Figure 5.9). This suggests that the prion-like domain of HNRNPUL1 is critical for forming the majority of its protein-protein interactions.

To confirm that HNRNPUL1 binds several ALS-causing RBPs via its prion-like domain, and that these interactions are enhanced when the nucleotide-binding ability of HNRNPUL1 is disrupted, FLAG-HNRNPUL1 IPs were repeated using the stable FLAG-HNRNPUL1 lines, and the presence of FUS, TAF15 and EWSR1 was assessed in each IPed fraction via Western blot. Consistent with the mass spectrometry findings, each ALS-causing RBP interacted with FLAG-HNRNPUL1 WT, and this interaction

A

RANK	PROTEIN	BAIT:INTERACTOR	
1	HNRNPUL1 (WT)	1.0	BAIT
2	HNRNPA2B1	15.9	
3	FUS	23.3	
4	RPL29	30.0	
5	HNRNPA1;HNRNPA1L2	31.8	
6	RPLP2	58.7	
7	ERH	70.6	
8	HNRNPA3	73.4	
9	RPL6	76.2	
10	RBMX	78.0	
11	EWSR1	78.5	

B

RANK	PROTEIN	BAIT:INTERACTOR	
1	HNRNPUL1 (WA mutant)	1.0	BAIT
2	HNRNPA2B1	1.5	
3	RPLP2	2.3	
4	FUS	3.4	
5	RPL29	4.3	
6	HNRNPC	4.3	
7	HNRNPA1;HNRNPA1L2	4.5	
8	HNRNPA3	5.4	
9	EWSR1	6.9	
10	RPLP1	7.1	
11	HNRNPH1	7.9	

Figure 5.6 Comparison of iBAQ bait to prey ratios in the FLAG-HNRNPUL1 WT and FLAG-HNRNPUL1 Walker A mutant IP fractions

A – Table displaying the iBAQ bait to interactor ratios of the top 10 proteins identified via mass spectrometry following IP of FLAG-HNRNPUL1 WT. B - Table displaying the iBAQ bait to interactor ratios of the top 10 proteins identified via mass spectrometry following IP of FLAG-HNRNPUL1 WA.
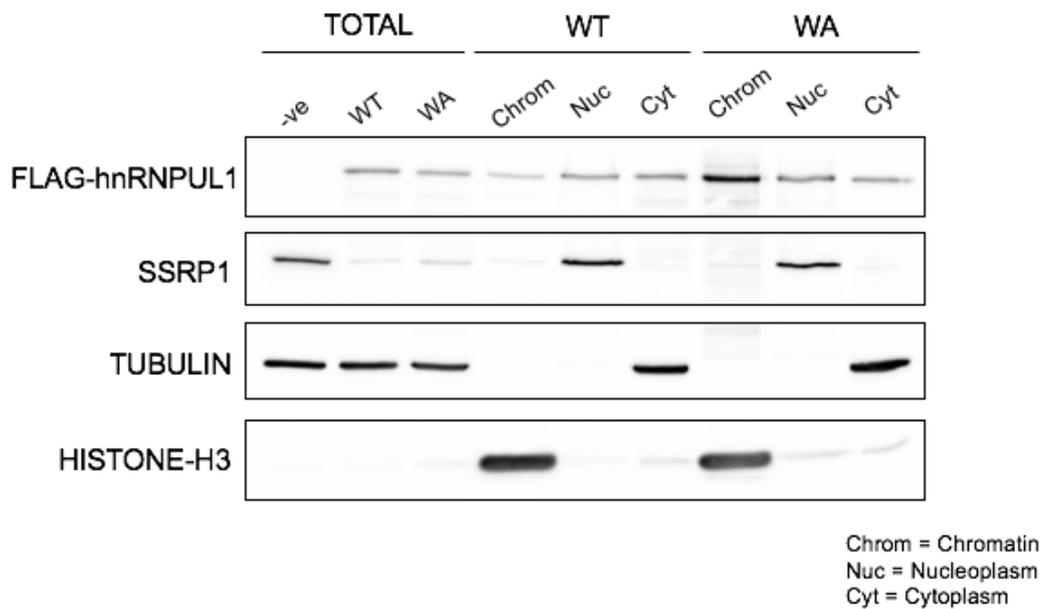
Figure 5.7 FLAG-HNRNPUL1 Walker A mutant is enriched in chromatin fraction

Western blot analyses of the levels of FLAG-HNRNPUL1-WT and FLAG-HNRNPUL1-WA in each sub-cellular fraction. SSRP1, TUBULIN and HISTONE-H3 are shown as nucleoplasmic, cytoplasmic, and chromatin markers respectively.

Gene names	Number of Peptides	iBAQ
HNRNPUL1 ΔCTD	58	1789250000
RABEP1	2	33778500
HNRNPA2B1	15	7599100
RPL6	6	5866450
HNRNPA1;HNRNPA1L2	8	5305950
SNRPA	5	3966100
RPLP2	4	3096650
ARL6IP4	5	3011900
HSPA1B;HSPA1A	15	2467950
TXNDC12	4	1778100
HSPA8	14	1537450
SNRPE	2	1533750
HIST1H2BN;HIST1H2BL;H	3	1382000
SNRPD2	4	1311805
HIST1H1E;HIST1H1D	3	1083405
HNRNPA3	5	1022015
HNRNPU	5	907165
ACTG1;ACTB	4	878800
RBMX	6	790380
HNRNPC	4	717730
SNRPN;SNRPB	2	490840
SPIN1	3	449315
SSB	4	395830
QPCT	3	249540

Figure 5.8 RNA-independent interactors of HNRNPUL1 ΔCTD mutant

Table of HNRNPUL1 ΔCTD mutant-associated proteins identified by mass spectrometry, ranked by iBAQ score.
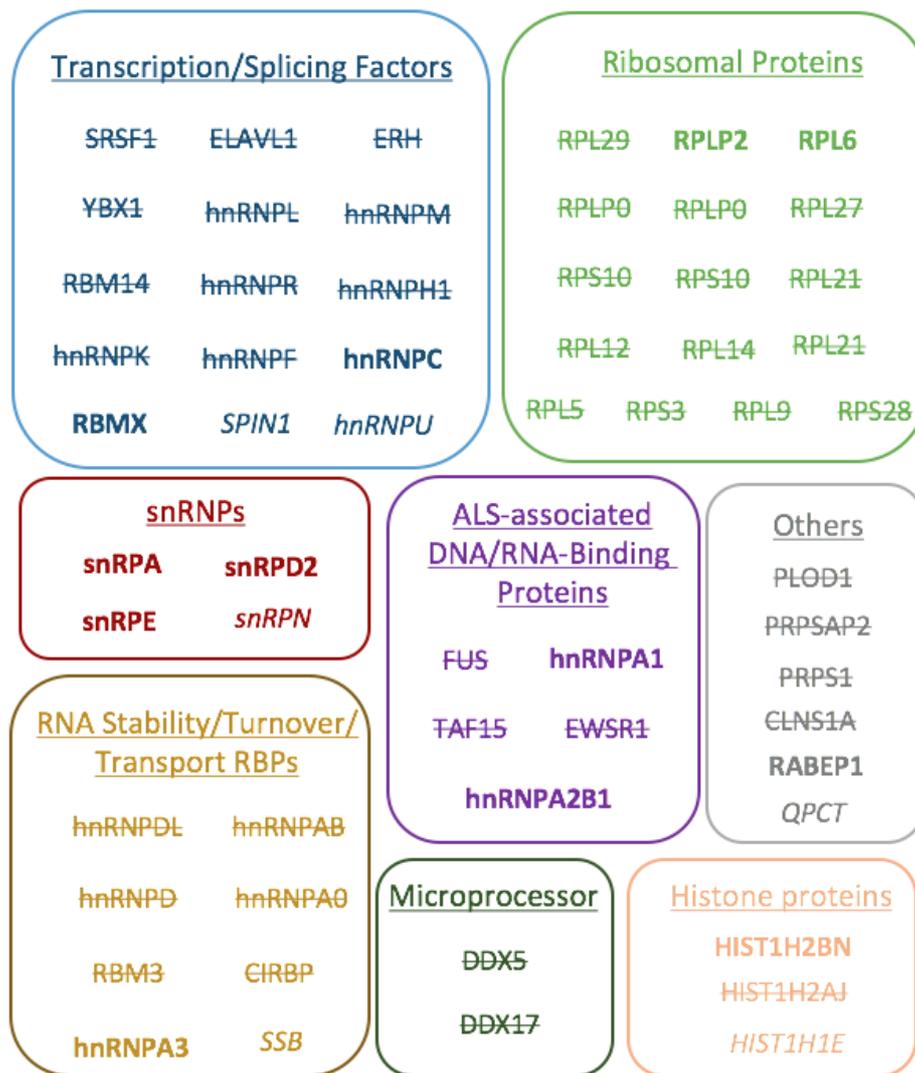
Figure 5.9 RNA-independent interactors of HNRNPUL1 ΔCTD mutant categorised by function.

Summary of HNRNPUL1 ΔCTD mutant-associated proteins, identified via mass spectrometry following FLAG IP in the presence of RNase A. Mass spectrometry analysis was performed by Mark Dickman and Caroline Evans. Proteins identified in the FLAG-HNRNPUL1 ΔCTD sample that were also present in the FLAG-GFP sample have been removed from this visualisation, along with the contaminant protein Actin.

was stronger in the case of FLAG-HNRNPUL1 WA and significantly disrupted in the case of FLAG-HNRNPUL1 ΔCTD (Figure 5.10). RNAPII was also probed for, as although it was not identified via mass spectrometry we have demonstrated a clear interaction between this complex and the endogenous form of HNRNPUL1 in the previous chapter. This interaction displayed a similar profile to the ALS-causing RBPs, with the Walker A mutant binding RNAPII more strongly than the wild-type and the ΔCTD mutant unable to bind the complex (Figure 5.10).

5.4 HNRNPUL1 protein-protein interactions relating to small RNA biogenesis are RNA-dependent

Although a few snRNP proteins were identified in the HNRNPUL1 WT IPed fraction via mass spectrometry, many factors involved in small RNA biogenesis were noticeably absent given the findings of the previous chapter linking HNRNPUL1 to this pathway. We therefore repeated the IP of FLAG-HNRNPUL1 followed by mass spectrometry analyses, this time without the addition of RNase A to the lysates in order to assess whether HNRNPUL1 interacted with these proteins in an RNA-dependent manner. As previously, FLAG-GFP was also IPed as a negative control. As expected, a much larger group of proteins co-immunoprecipitated with FLAG-HNRNPUL1 WT under these conditions, with 236 interactors identified compared to 59 in the plus RNase A condition (Figure 5.11). Among this group of RNA-dependent interactors, several were related to snRNP and snoRNP biogenesis (Figure 5.12). Components of U1, U2, U4/U6 snRNPs and Box C/D snoRNPs were present, as well as members of the NEXT complex and nuclear exosome previously demonstrated to interact with HNRNPUL1 in Chapter 4. Interestingly the exonuclease XRN2, which in addition to its role in transcription termination processes the 5'-end of pre-snoRNAs, was ranked 10th in iBAQ score out of all HNRNPUL1 interactors. The presence of both 5'-end and 3'-end snoRNA processing and degradation factors in the HNRNPUL1 IPed fraction again strongly supports the notion that HNRNPUL1 plays a role in the snoRNA
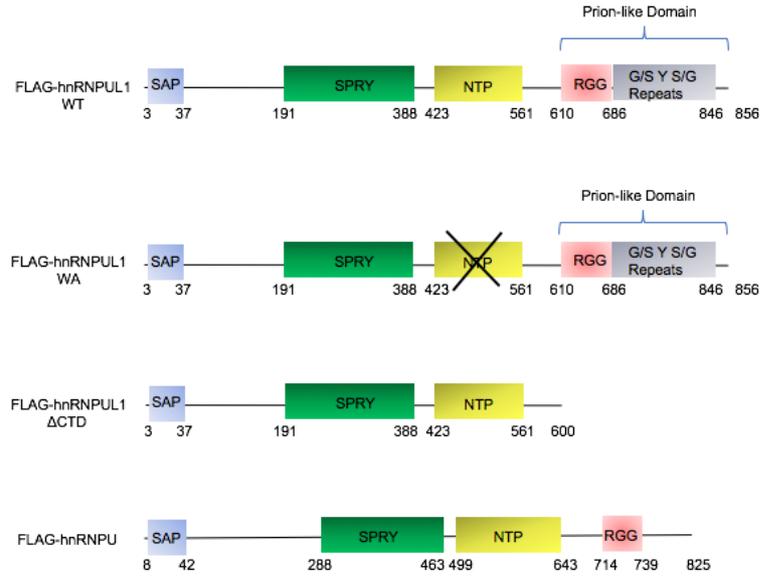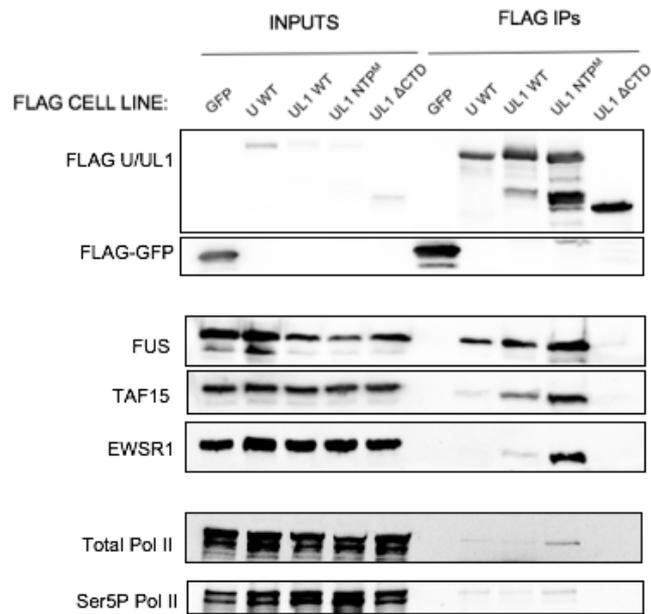
Figure 5.10 HNRNPUL1 interacts with ALS-causing RBPs and RNAPII via its prion-like domain

A – Schematic diagram of the HNRNPU/UL1 proteins included in the FLAG IP assay. B – Western analyses following FLAG-HNRNPU and FLAG-HNRNPUL1 IP, demonstrating that HNRNPUL1 specifically interacts with FUS, TAF15, EWSR1, and RNAPII.

Gene names	Number of Peptides	iBAQ
HNRNPUL1 WT	37	470530000
RPL29	4	13724000
RPS30	3	12768000
YBX1	6	12527000
NUCLEOLIN	31	10480000
RPLP2	7	9097300
TRYP	4	7861200
HISTH1E	8	7133800
RPL13	9	6244000
XRN2	24	5901800
SART3	21	5744600
HNRNPA2B1	11	5674300
NFKB	19	5587200
YBX3	11	4937000
SRSF1	8	4612200
HNRNPK	13	4545400
HNRNPA1	2	4492200
BSAP	6	4388500
RBM14	17	4271200
RPS28	3	4188000
RPL26	5	4048700
RPL39S	1	3874300
RPL7	4	3119900
RPL7A	6	3028700
RPL9	3	2741500
SRSF7	4	2549600
NPM1	6	2385300
HNRNPH1	5	2341400
GRASP	1	2236400
SSB	13	2148100
HNRNPC	3	2145700
HNRNPAB	6	1881500
RPL35	3	1727100
STK38	5	1673600
RPL36	3	1671300
RPS14	2	1662500
RPL6	7	1659600
HIST1H2B	1	1598300
SRSF9	5	1596400
HIST1H2AJ	1	1510000
RPLP1	1	1489800
RPL8	4	1460100
ALYREF	5	1448900
HSPA1B	13	1391700
RPS10	2	1387800
HNRNPU	8	1312100
RPS11	2	1295000
SNRPD2	5	1207300
RPL5	6	1177500
RPL4	4	1166400
PABPC1	14	1165700
RPL19	3	1153400
RPS19	5	1109000
HNRNPD	1	1100800
H1FX	5	1087100
MRPL55	1	1033000
RPL13AP3	2	1004000
RPL23A	2	952810
EEF1A1	3	897850

(Figure 5.11, continued on next page)

MRPL12	1	885060
HNRNPL	7	878880
RPS8	3	850220
HNRNPR	5	846180
HNRNPH3	4	827050
SNRPE	3	782490
RPL24	1	765940
SNRPD1	2	720060
SRP14	3	696060
RPL27A	1	667760
HNRNPF	3	662120
RPL12	3	656200
RPS29	1	631070
RPS3	5	606860
RBMX	8	605690
IGF2BP1	4	581960
RPS25	2	578980
UBB	1	565060
SRSF2	1	561950
DHX15	7	544910
MRPL13	1	541360
KHSRP	9	538200
PRMT5	3	507940
DDX17	8	507110
RPS6	3	500490
IGKV2-29	1	476560
ERH	1	468830
HIST1H4A	2	447900
MYCBP	1	440010
RPL11	2	429190
TAOK1	1	411670
ILF3	9	408160
MRPL49	2	407640
MRPS7	2	404510
RPS3A	3	399570
PABPC4	4	398950
ACTB	4	394360
RPS17	1	393490
MRPS23	3	385750
TRIM21	5	385450
SNRPA1	4	377010
LARP1	7	367850
DHX9	6	364470
HNRNPA0	1	357830
ARL6IP4	2	347400
RBM10	8	342830
DDX5	3	342570
ELAVL1	3	336000
STAU1	5	333890
SYNCRIP	4	327230
RPS4X	3	325980
PABPN1	2	306260
FUS	1	295690
MRPS28	1	294070
MRPL11	4	293720
MTR4	2	290900
RPL31	1	286700
ZNF169	1	280600

(Figure 5.11, continued on next page)

RPSA	3	278350
RPL32	1	274340
KPNA2	1	268480
HNRNPDL	3	262640
EBNA1BP2	4	258000
QKI	1	257710
MRPS31	3	256920
MEPCE	2	250880
RPL27	1	247650
DDX21	6	240600
TUBA1B	1	235810
RPL21	1	223180
RPS21	1	218680
RPL3	2	209870
SNRPA	1	197780
SF3A1	5	194320
MRPS6	1	193790
RPL15	1	189490
MRPL41	2	188610
MRPL40	2	187440
SRSF10	2	186770
CAS	1	182520
HSPA8	3	177600
RBM17	3	173780
MRPL17	2	173400
TRIM28	3	172710
RBMS1	1	161100
YTHDF2	1	160760
SF3B2	4	158000
NONO	4	156330
HNRNPM	5	146650
RPL18	2	143360
PRPS2	1	136550
GADD45GIP1	2	135520
DHX36	4	134010
MRPL34	1	131250
MRPS22	1	130920
ZCCHC8	3	130680
RRP40	1	128600
CLNS1A	1	128290
YTHDC2	6	125390
RBM4	1	122550
SFPQ	2	116820
MRPL48	1	110330
OFD1	2	108520
KHDRBS1	1	108310
SNRPN	1	107740
HNRNPA3	1	105380
PRPF4	3	104680
DAP3	1	104030
RBM3	1	102820
RPS12	2	101840
SF1	1	100400
LSM8	1	98607
HSPA5	3	98402
SERBP1	1	97008
RBM7	1	96056
KIF11	5	95467

(Figure 5.11, continued on next page)

CCAR2	3	93842
C8orf33	2	86996
SRSF5	1	86268
TUBB	2	84052
PTBP1	2	83846
HSPA9	4	80350
MRPL28	1	74866
EWSR1	1	73495
SF3A3	1	72202
EXOSC4	1	71919
MRPS26	2	71442
RPS18	1	69837
FBL	1	66794
MRPS9	1	66235
TUBB4A	1	61762
SF3B1	5	61658
PRPF3	2	60757
NOP2	2	60579
MRPL23	1	59946
MRPL15	1	57730
MATR3	2	54796
DHX30	4	52797
RSL1D1	1	52467
LARP7	1	52257
DDX39B	1	52186
RPL10A	1	50908
HNRNPH2	1	43553
PCNP	1	41073
MRPL38	1	41044
THRAP3	2	40969
SRPK1	2	40709
PHAX	1	39181
FAM120A	1	38634
MRPS34	2	38223
RPS7	1	37127
TFAM	1	36770
EIF2S1	1	36537
EXOSC6	1	34851
TXNDC12	1	33233
HSPA6	1	31876
IGF2BP2	2	30976
SNRPB2	1	30080
GLA	1	29625
IGF2BP3	1	27371
MAP1B	3	26104
SNRNP70	2	25728
PARP1	1	25313
ZFR	2	24656
TUT1	1	23784
DDX50	2	21931
FUBP1	1	20584
CLASP2	3	19958
TOE1	1	17937
TSEN2	1	17913
ZC3HAV1	2	17482
ZNF326	1	16978
PLOD1	2	14115
MYBBP1A	1	13383
KRT1	1	9767.9
APOB	1	7328.5
MOV10	1	6879.4
EXOSC10	1	6652.9
EPRS	1	5333.8
ADAR	1	3187.2

Figure 5.11 RNA-dependent interactors of HNRNPUL1 WT

Table of HNRNPUL1-associated proteins identified by mass spectrometry, ranked by iBAQ score. IPs were performed without RNase A treatment.
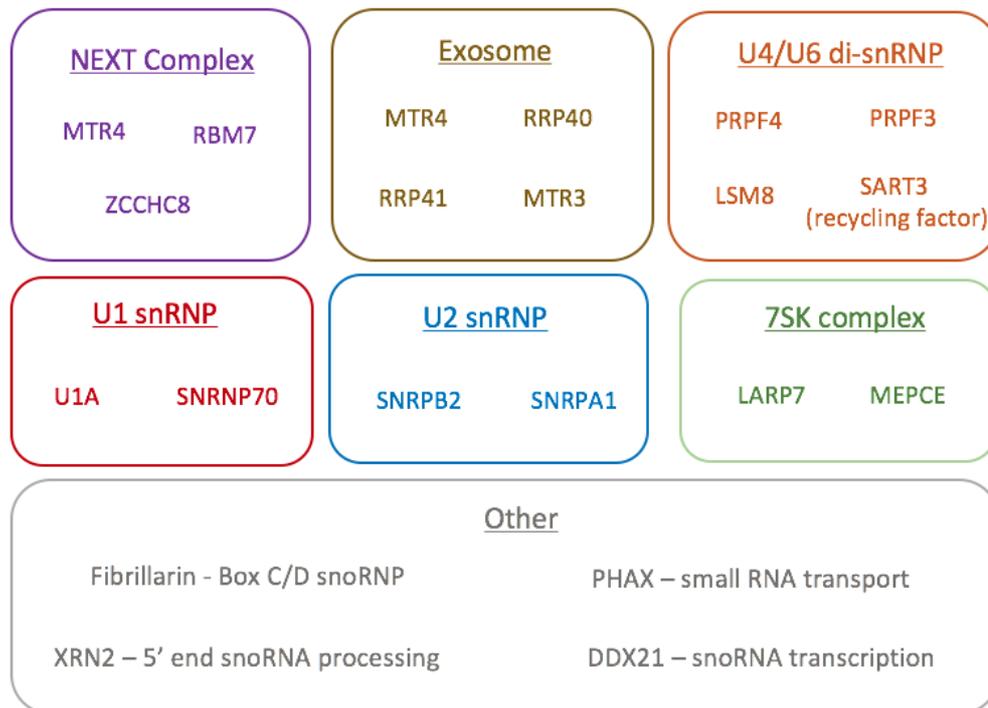
Figure 5.12 The interactions between HNRNPUL1 and the snRNP/snoRNP biogenesis machinery are mediated by RNA

Summary of snRNP/snoRNP biogenesis components that interact with HNRNPUL1, identified via mass spectrometry following FLAG IP in the absence of RNase A. Mass spectrometry analysis was performed by Mark Dickman and Caroline Evans.

maturation process, especially given the reduction in snoRNA levels observed upon HNRNPUL1 depletion. That these interactions are mostly RNA-dependent suggests that it is more likely that HNRNPUL1 is competing with these proteins for binding to their snoRNA substrates as opposed to directly inhibiting these factors through direct protein-protein interaction.

SART3 was identified under these conditions as an interactor of HNRNPUL1 with the 11th highest iBAQ score, consistent with the strong co-IP between this protein and HNRNPUL1 demonstrated in the previous chapter. In addition to its role as a U4/U6 di-snRNP recycling factor, SART3 has also been reported as a component of the 7SK complex (Jeronimo et al., 2007; McNamara et al., 2013). This complex is composed of a core of protein subunits MePCE and LARP7 associated with 7SK RNA, and regulates the transcription of numerous classes of gene such as snRNAs and immediate-early genes (Figure 1.2). Interestingly, MePCE and LARP7 were also present in the FLAG-HNRNPUL1 IPed fraction (Figure 5.11). This is consistent with the HNRNPUL1 RIP assays performed in Chapter 4 that demonstrated that HNRNPUL1 also interacts with the 7SK RNA transcript itself (Figure 4.3).

To validate this IP/mass spectrometry finding, FLAG-HNRNPUL1 IPs were repeated using the stable FLAG-HNRNPUL1 lines, and the presence of 7SK components was assessed in each IPed fraction via Western blot. In agreement with the mass spec, each 7SK component analysed was clearly identifiable in the wild-type IPed lane, and strongly depleted in the FLAG-HNRNPUL1 WA IP (Figure 5.13). The increased interaction observed between FLAG-HNRNPUL1 WA and FUS with respect to wild-type HNRNPUL1 demonstrates that this phenomenon is specific to the 7SK complex.

5.5 ALS-causing FUS mutant does not disrupt the cellular localisation of HNRNPUL1

The DNA/RNA-binding protein FUS is a major cause of ALS, accounting for approximately 5% of familial cases (Mackenzie et al., 2010). In light of the strong
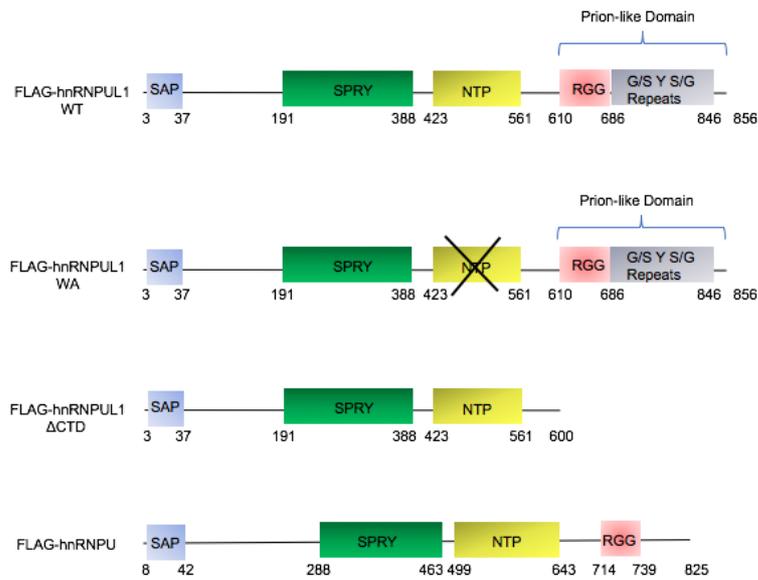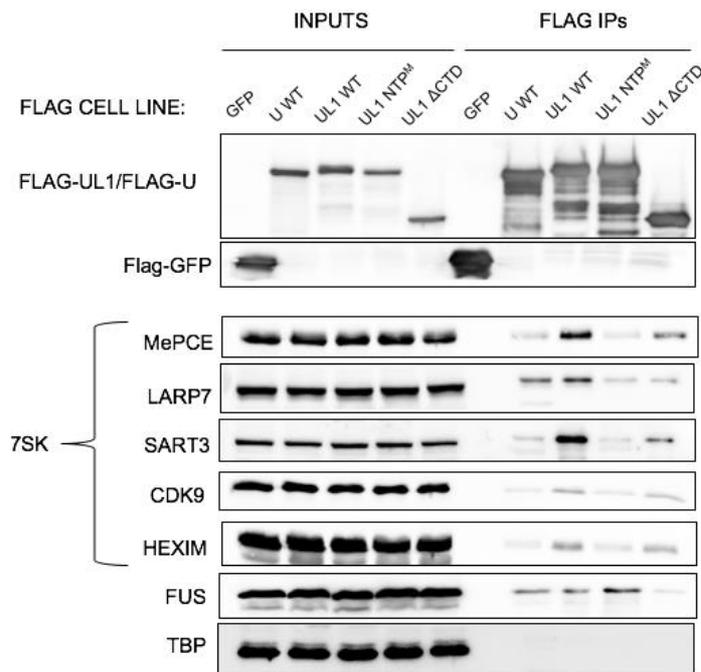
Figure 5.13 HNRNPUL1 interaction with the 7SK complex requires an intact Walker A motif

A – Schematic diagram of the HNRNPU/UL1 proteins included in the FLAG IP assay. B – Western analyses following FLAG-HNRNPU and FLAG-HNRNPUL1 IP, demonstrating that HNRNPUL1 specifically interacts with the 7SK complex.

interaction identified between HNRNPUL1 and FUS via mass spectrometry, further investigating the relationship between these two proteins was therefore of great interest. It has previously been reported that ALS-causing FUS mutants that accumulate in the cytoplasm can also sequester other proteins and RNA into these aggregates (Jun et al., 2017; Yamazaki et al., 2012; Yu et al., 2015). The resulting loss of nuclear function of these interacting partners is thought to contribute to the pathology of FUS-related proteinopathy. Therefore, we next examined the effect of ALS-causing FUS mutant R495X on the cellular localisation of HNRNPUL1. FUS R495X lacks its C-terminal nuclear localisation signal (Figure 5.14A) and accumulates in the cytoplasm as a result. Therefore, vectors expressing FLAG-tagged wild type FUS and FLAG-tagged FUS R495X, kindly sent to us by Robin Reed, were transfected into HeLa cells and immunostaining was performed using anti-FLAG and endogenous HNRNPUL1 antibodies. This firstly confirmed that the R495X mutation does result in cytoplasmic mis-localisation of FUS (Figure 5.14B). However, no accompanying increase in HNRNPUL1 cytoplasmic signal was observed (Figure 5.14B), indicating that the nuclear localisation of HNRNPUL1 is unaffected by this ALS-causing FUS mutant.

5.6 Depletion of FUS results in elevated HNRNPUL1 RNA binding

As demonstrated earlier in the chapter, FUS interacts with the C-terminal region of HNRNPUL1. This region contains an RGG box, a motif known to possess RNA-binding activity. Therefore, to further investigate the relationship between HNRNPUL1 and FUS, the impact of FUS on the RNA-binding ability of HNRNPUL1 was next examined. RNA-IP was performed on HNRNPUL1 using an antibody against endogenous HNRNPUL1 in the presence and absence of FUS, using a FUS knockout cell line donated to us by Robin Reed (Figure 5.15A). The levels of transcripts co-immunoprecipitating with HNRNPUL1 were assessed via qRT-PCR. snRNAs, mRNAs and lncRNAs were all incorporated in order to ascertain whether any phenotype

A

Wild-type:

R495X mutant:
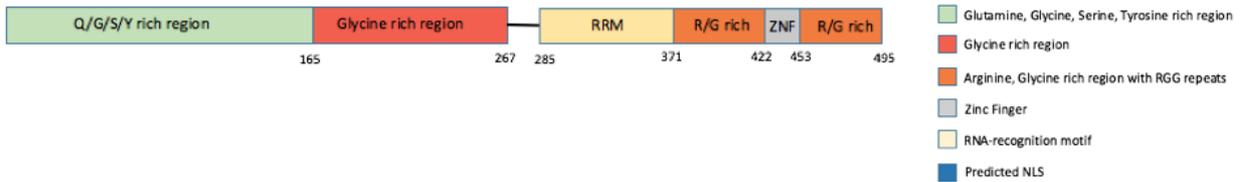
B
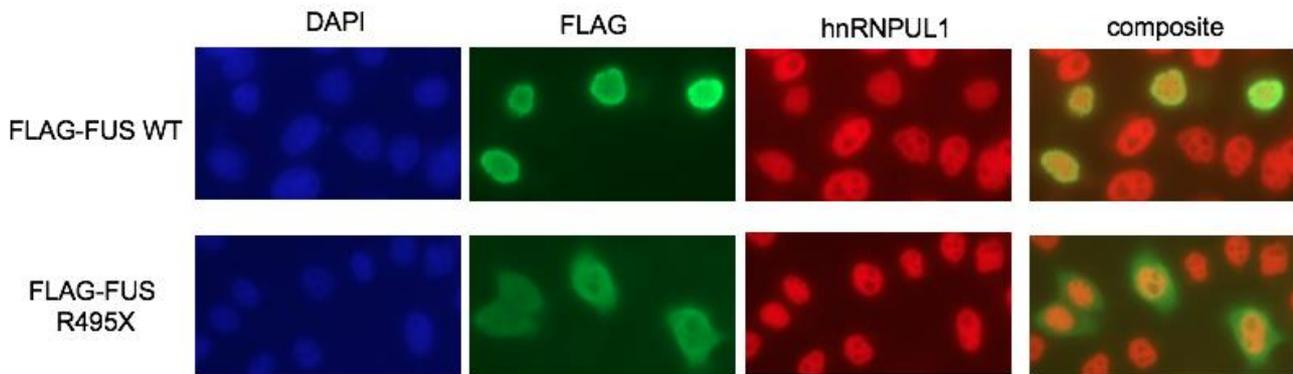
Figure 5.14 The cellular localisation of HNRNPUL1 is unaffected by expression of an ALS-causing FUS R495X mutant

A – Schematic diagram comparing the wild-type form of FUS with the R495X mutant. B – Immunostaining using endogenous anti-HNRNPUL1 and anti-FLAG antibodies in cells transfected with either FLAG-FUS WT or FLAG-FUS R495X.

A
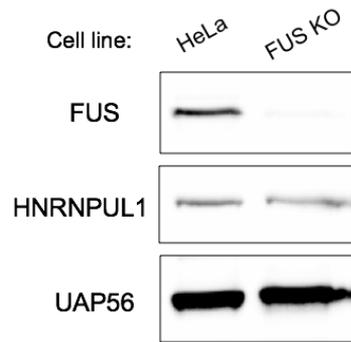
B
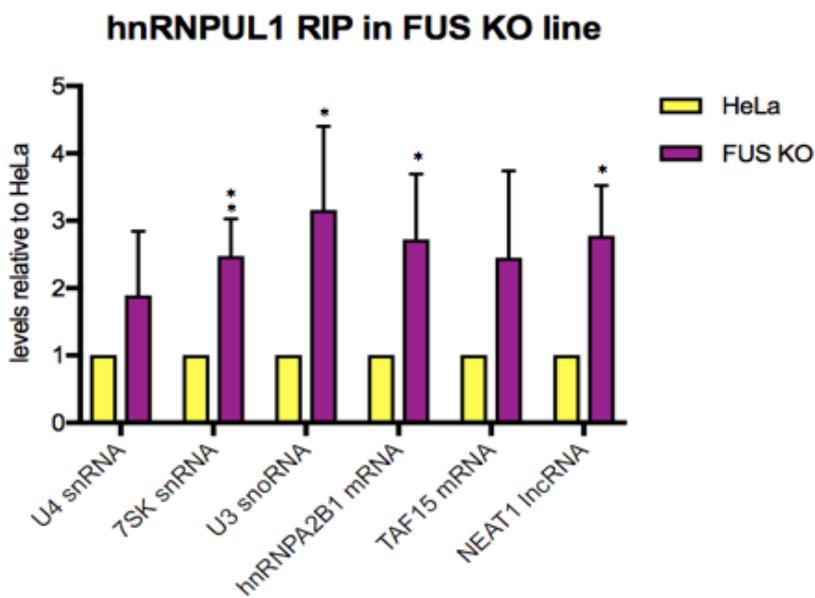
Figure 5.15 FUS knockdown results in enhanced association between HNRNPUL1 and its RNA targets

A – Western analyses demonstrating absence of FUS protein in the FUS KO line. UAP56 is shown as a loading control. B – RNA-IP analyses comparing the levels of RNA co-immunoprecipitating with HNRNPUL1 in FUS knockout cells compared to HeLa. Error bars represent the SD.

observed was specific to a certain class of RNA. The specific examples chosen were based on analysis of HNRNPUL1 eCLIP dataset as well as the HNRNPUL1 RIP assays performed as described in the previous chapter, enabling the identification of transcripts that HNRNPUL1 was likely to bind. Interestingly, this revealed an increase in HNRNPUL1 RNA-binding to all transcripts tested in FUS knockout conditions relative to the HeLa control (Figure 5.15B). This suggests that FUS binding to the C-terminal region of HNRNPUL1 has a negative impact on its ability to bind RNA. Alternatively, given that FUS also interacts with these RNAs, it may compete with HNRNPUL1 for binding to these targets and hence the absence of FUS would enable HNRNPUL1 to associate to a greater degree.

To assess whether this effect on RNA binding was reciprocal, RNA-IP was then performed on FUS in the HNRNPUL1-AID line following auxin depletion of HNRNPUL1. Again, transcripts displaying strong FUS enrichment in the publicly available FUS eCLIP dataset were incorporated into the assay. In this instance however, knockdown of HNRNPUL1 resulted in no significant increase in FUS RNA-binding (Figure 5.16). This indicates that HNRNPUL1 does not suppress the RNA-binding capacity of FUS.

5.7 RNAPII CTD is hyper-phosphorylated upon HNRNPUL1 depletion

In addition to their roles in RNA processing, various ALS-associated RNA-binding proteins have also been reported to directly affect transcription via their impact on the phosphorylation status of the CTD of RNAPII (Gorthi et al., 2018; Schwartz et al., 2012). For example, the depletion of FUS results in elevated levels of CTD Serine-2 phosphorylation and increased RNAPII pausing at transcription start sites (Schwartz et al., 2012). Interestingly, this accumulation of RNAPII that is hyper-phosphorylated at Serine-2 has also been demonstrated in the fibroblasts of patients possessing ALS-causing FUS mutations (Schwartz et al., 2014). In addition, knockdown of another member of the FET family of RNA-binding proteins linked to ALS – EWSR1 has been
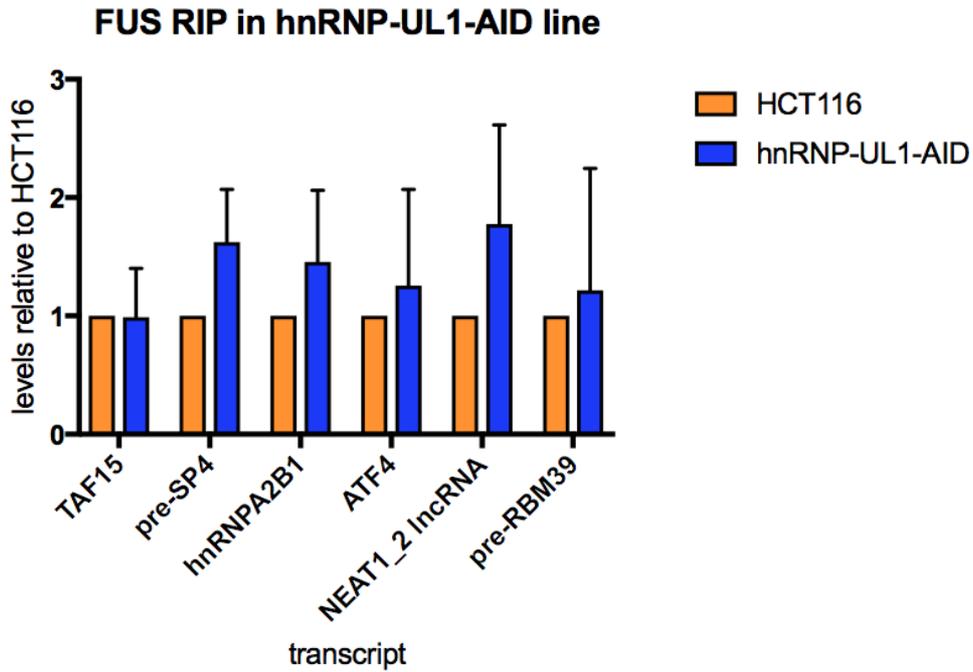
Figure 5.16 HNRNPUL1 knockdown does not result in significant increase in FUS RNA binding

RIP analyses comparing the levels of RNA co-immunoprecipitating with FUS in the HNRNPUL1-AID line relative to HCT116 cells. Error bars represent the SD.

,

shown to cause elevated levels of RNAPII CTD Serine-5 (Gorthi et al., 2018).

Given the HNRNPUL1/RNAPII interaction demonstrated earlier in the chapter and in Chapter 4, as well as previous studies reporting that this binding is comparable to the affinity of other ALS-causing proteins for RNAPII (Chi et al., 2018a), we hypothesised that knockdown of HNRNPUL1 may also affect the phosphorylation status of the CTD. To test this notion, Western analyses using various CTD antibodies were performed on whole cell lysates extracted from the HNRNPUL1-AID line following auxin depletion of HNRNPUL1. This revealed a clear upregulation of Serine-5 and Serine-7 phosphorylation (Ser5P, Ser7P) in the HNRNPUL1 knockdown condition (Figure 5.17). A marginal increase in Serine-2 phosphorylation (Ser2P) was also observed, but not to the same extent as the Ser5P and Ser7P increase. Consistently, the unphosphorylated form of the RNAPII CTD appeared downregulated (Figure 5.17). These increased Ser5P and Ser7P levels did not reflect an increase in overall expression of RNAPII, as Western blot analyses using an antibody that recognises all CTD peptides demonstrated no change in RNAPII abundance in the HNRNPUL1-AID line compared to HCT116 cells (Figure 5.17). This suggests that the increased Ser5P and Ser7P signal is as a result of genuine, increased phosphorylation.

Inactive RNAPII is largely unphosphorylated, and initiation of transcription involves the phosphorylation of Serine-5 and Serine-7 by CDK7, followed by Serine-2 phosphorylation by CDK9 to stimulate productive elongation (Zlotorynski, 2017). Therefore, we considered the possibility that the elevated phosphorylation status of RNAPII observed upon HNRNPUL1 knockdown reflected increased levels of transcription in the cell. However, meta-gene analyses of RNAPII signal across all protein-encoding genes in the HNRNPUL1-AID line following mNET-seq assays in fact demonstrated no change in RNAPII distribution upon HNRNPUL1 depletion (Figure 5.18). This indicated that there is not a global upregulation of transcription that would account for the increased RNAPII phosphorylation.
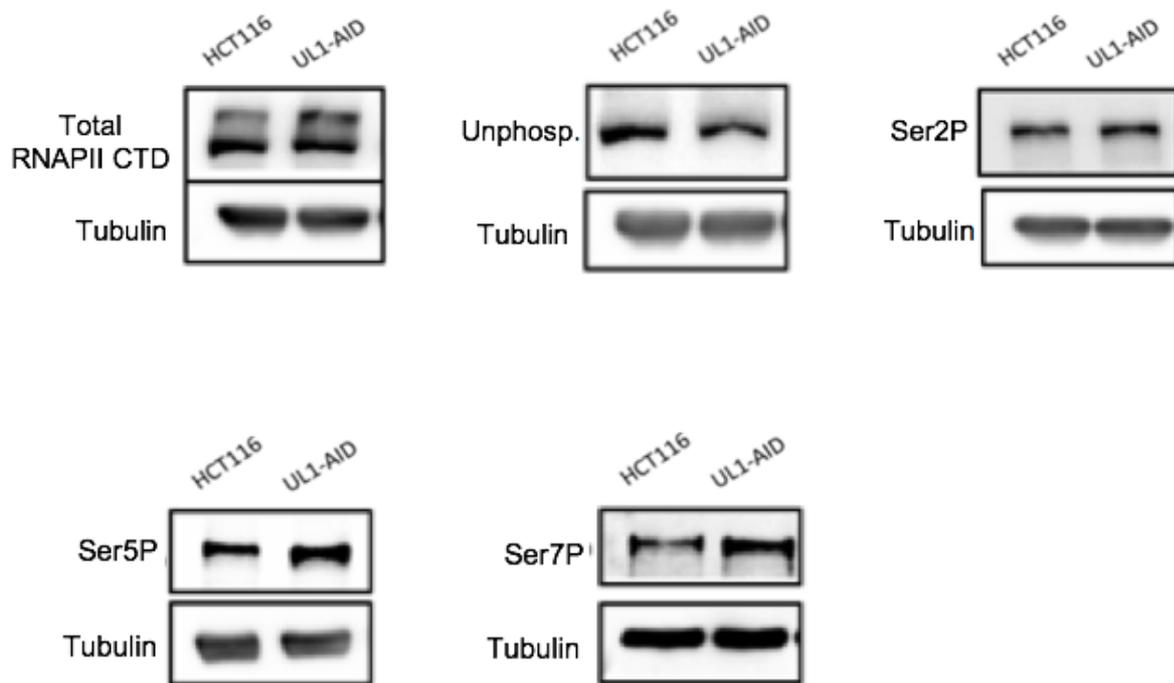
Figure 5.17 RNAPII CTD is hyper-phosphorylated upon HNRNPUL1 knockdown

Western analyses of the phosphorylation status of RNAPII CTD in the HNRNPUL1-AID line compared to HCT116 cells, using antibodies against total, unphosphorylated, Serine-2, Serine-5 and Serine-7-phosphorylated forms of the CTD. Tubulin is displayed as a loading control.

Figure 5.18 No global increase in transcription upon HNRNPUL1 knockdown

mNET-seq RNAPII signal across all protein-encoding genes following mNET-seq analyses in HNRNPUL1-AID line performed by Ivaylo Yonchev. The global distribution of RNAPII is unchanged upon HNRNPUL1 knockdown.

5.8 HNRNPUL1 is a CDK7 substrate, Walker A mutant reduces RNAPII phosphorylation by CDK7

HNRNPU – another member of the HNRNP family that displays significant structural and sequence homology to HNRNPUL1 – has been previously reported to inhibit the kinase activity of CDK7 (Kim and Nikodem, 1999). We therefore speculated that HNRNPUL1 may also be a CDK7-inhibitor, as this would be consistent with the observed upregulation of Ser5P and Ser7P. In this model, knockdown of HNRNPUL1 would result in less suppression of CDK7, and thus an increase in phosphorylation of its targets would occur, as is displayed in the HNRNPUL1-AID line. Therefore, the ability of HNRNPUL1 to inhibit CDK7-phosphorylation of RNAPII CTD was assessed via an *in vitro* kinase assay.

In order to obtain purified HNRNPUL1 to incorporate into these assays, FLAG-HNRNPUL1 WT was transfected into 293T cells and immunoprecipitated using anti-FLAG antibody-coated agarose beads. RNase A and DNase were added to the lysate to degrade the nucleic acids present, and beads were washed twice in a high salt wash buffer prior to elution to remove any protein binding partners that would otherwise co-IP with HNRNPUL1. Given that the FLAG-HNRNPUL1 WA mutant interacts more strongly with RNAPII than the wild-type, this mutant was also purified in the same manner to enable it to be incorporated into the kinase assays as well. FLAG-GFP was also purified to be used as a negative control. Coomassie staining analyses on the IPed fractions indicated that the three FLAG-tagged proteins had been purified cleanly with no obvious contaminants co-purifying (Figure 5.19).

As a substrate for the kinase assay, 25 copies of the consensus RNAPII CTD heptad repeat fused to a GST tag (GST-CTD25) was purified from *E.coli* cells (Figure 5.19). The expression vector containing this construct was kindly donated as a gift from Shona Murphy. Purified CDK7 was purchased from Sigma, as part of its active complex with Cyclin H1 and MNAT1.

Purified GST-CTD25, CDK7 and each FLAG-tagged protein were incubated with
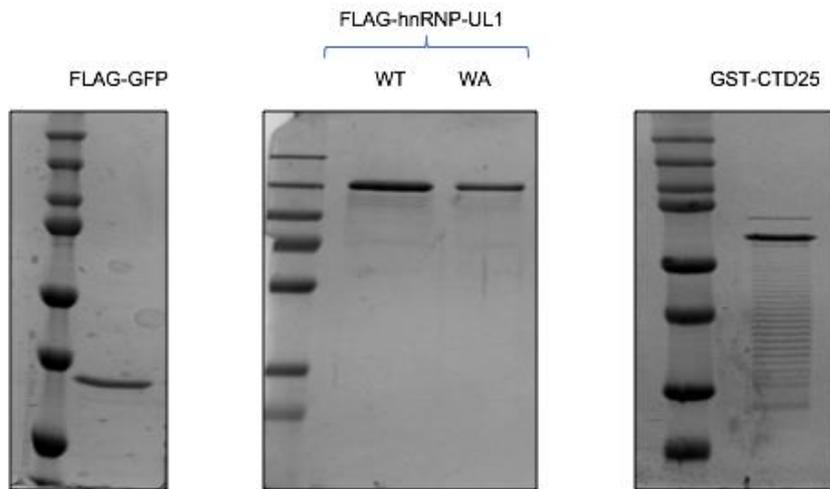
Figure 5.19 Purification of FLAG-HNRNPUL1 WT, WA, FLAG-GFP and GST-CTD25

Coomassie staining of FLAG-GFP, FLAG-HNRNPUL1 WT, FLAG-HNRNPUL1 WA and GST-CTD25. FLAG-tagged proteins were purified from mammalian cells via FLAG IP with high salt washing, while GST-CTD25 was purified from *E.coli* using glutathione sepharose beads.

[γ32P]-ATP for 5 minutes, and the level of GST-CTD25 phosphorylation was measured subsequently via phosphor-imager. This firstly revealed that CDK7 phosphorylates HNRNPUL1, as evidenced by the clear band present at the appropriate size of FLAG-HNRNPUL1 in the case of both the wild type and Walker A mutant (Figure 5.20A). This came as a surprise, as although HNRNPUL1 is known to be regulated by arginine methylation (Gurunathan et al., 2015), CDK7-mediated phosphorylation of HNRNPUL1 has not been previously reported. Whether this phosphorylation of HNRNPUL1 by CDK7 influences transcription initiation at the specific genes that HNRNPUL1 regulates is an outstanding question. Secondly, there appeared to be a minimal reduction in CTD phosphorylation when the wild-type form of HNRNPUL1 was present in the assay, whereas the Walker A mutant did appear to inhibit CTD phosphorylation (Figure 5.20A). Interestingly, the Walker A mutant itself appeared more phosphorylated than the wild-type form of the protein (Figure 5.20A).

To ensure that the phosphorylation occurring was indeed mediated by CDK7, and not as a result of HNRNPUL1 phosphorylating the CTD and/or itself, a repeat of the kinase assay was performed with a 'no CDK7' control included that did not contain the CDK7 enzyme. No bands were present in this lane when visualised by the phosphor-imager, confirming that CDK7 was the active kinase in these assays (Figure 5.20B). Once again, the Walker A mutant appeared phosphorylated to a greater extent than the wild type in the assays that did include CDK7. This result is interesting in light of the finding that the WA mutant also binds RNAPII with greater affinity than the wild-type. It could be the case that this form of the protein is masking the CTD, preventing CDK7 from accessing it as a substrate. Alternatively, it may also interact with CDK7 with greater affinity than the wild-type and therefore compete more successfully with the CTD for enzyme binding. Overall however, as there is limited inhibition of CDK7 in the case of FLAG-HNRNPUL1 WT, it appears unlikely that the upregulation in Ser5P and Ser7P observed in the HNRNPUL1-AID line is due directly as a result of reduced inhibition of CDK7 upon HNRNPUL1 depletion.
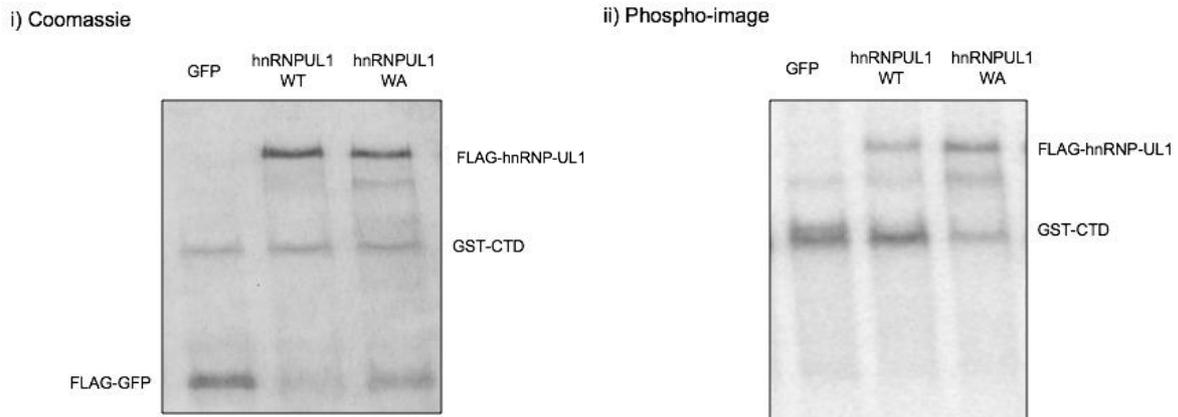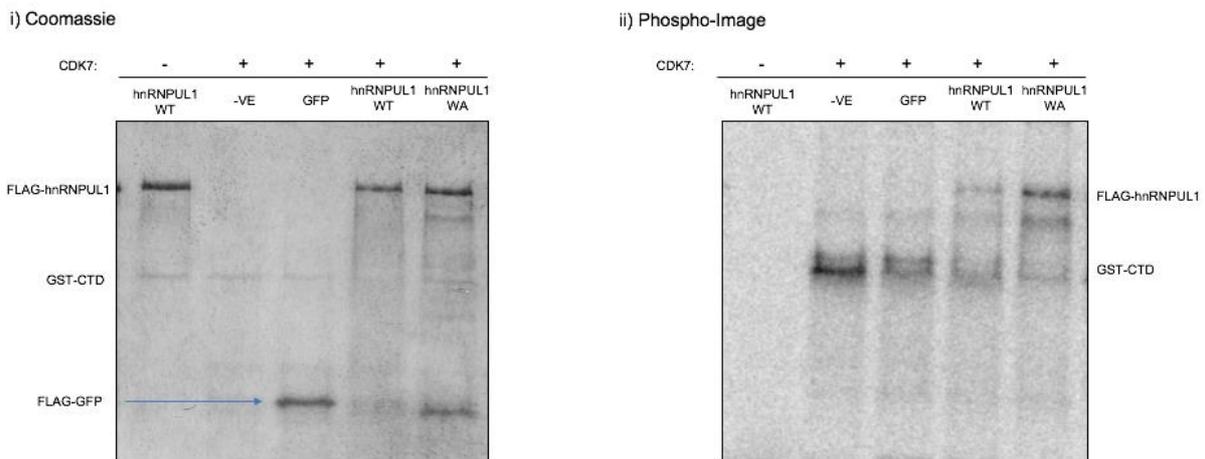
Figure 5.20 HNRNPUL1 is phosphorylated by CDK7, but only the Walker A mutant inhibits its ability to phosphorylate RNAPII CTD

A – Kinase assay including CDK7, GST-CTD25 and FLAG-HNRNPUL1 WT and WA. Purified proteins were incubated with [γ32P]-ATP for 5 minutes, before reactions were stopped following the addition of SDS-loading buffer and boiling. Samples were run on a gel, which was then dried and analysed using a phosphor-imager. B – Kinase assay repeat, including a 'no CDK7' control.

5.9 RNAPII CTD hyper-phosphorylation triggers increased RNAPII-U1 snRNP interaction

It has recently been demonstrated both in *S. cerevisiae* and human cells that Ser5P levels peak at actively spliced exons and aid recruitment of the spliceosome to nascent pre-mRNA (Harlen et al., 2016; Nojima et al., 2018). mNET-seq assays in combination with mass spectrometry performed on Ser2P, Ser5P and Thr4P isoforms of the CTD revealed that components of the spliceosome complex, along with a range of splicing factors, specifically co-immunoprecipitate with the Ser5P isoform (Nojima et al., 2018). In addition, inhibition of U1 snRNA via antisense oligo triggers upregulation of Ser5P levels (Koga et al., 2015), further cementing the link between splicing and this CTD isoform. It is therefore possible that the reduced levels of U1 snRNA observed in HNRNPUL1 knockdown cells could mimic U1 snRNA inhibition and thus be the cause of the elevated levels of Ser5P, given the lack of obvious CDK7 inhibition demonstrated by HNRNPUL1 in the *in vitro* kinase assays.

In light of this relationship, we reasoned that spliceosome recruitment to RNAPII may be affected by the upregulation in Ser5P levels displayed in the HNRNPUL1 AID cell line. Therefore, RNAPII IPs were performed in HCT116 and HNRNPUL1-AID lines following auxin depletion of HNRNPUL1. The levels of U1 snRNP co-immunoprecipitating with RNAPII in both conditions were assessed via Western blot. This revealed a large increase in RNAPII-U1 snRNP binding in HNRNPUL1 knockdown conditions (Figure 5.21A). The interaction between FUS and RNAPII was unaffected by the depletion of HNRNPUL1, suggesting that this increased binding was specific to the spliceosome. To confirm this finding, the reciprocal experiment was performed, this time immunoprecipitating U1 snRNP component U1C from cells following auxin depletion of HNRNPUL1. Again, there was a clear increase in the interaction between the U1 snRNP and RNAPII in the HNRNPUL1-AID line (Figure 5.21B). This could be recognised by probing with the pan-specific RNAPII antibody and was particularly noticeable in the case of Ser5P, with the enrichment observed
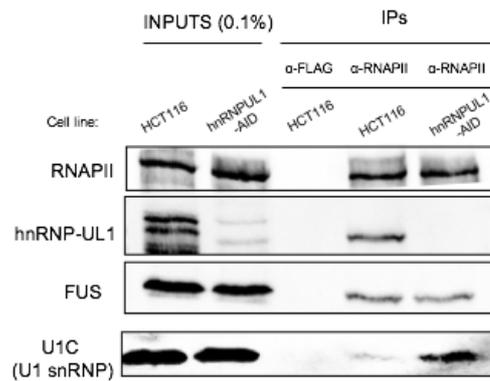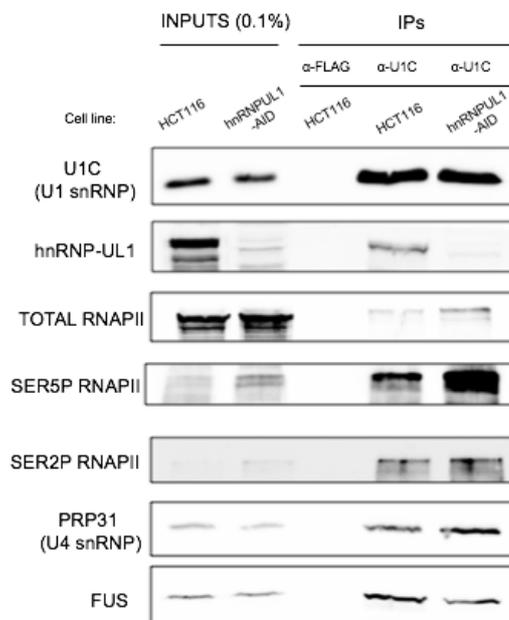
Figure 5.21 HNRNPUL1 knockdown triggers increased RNAPII-U1 snRNP interaction

A - Western analyses demonstrating increased U1C binding to RNAPII in an RNAPII IP plus and minus HNRNPUL1, using a pan-specific RNAPII antibody. B – Western analyses of U1C IP plus and minus HNRNPUL1, revealing a clear increase in RNAPII interaction and a specificity for the SER5P form of the CTD.

corroborating previous reports of the specific affinity displayed by the spliceosome for this CTD isoform. Once more, this effect appeared specific to RNAPII-U1 snRNP as FUS-U1 snRNP binding did not increase in HNRNPUL1 knockdown conditions.

Interestingly, there also appeared to be more U4 snRNP co-immunoprecipitating with U1 snRNP in the HNRNPUL1-AID line (Figure 5.21B). This suggests that the upregulation of Ser5P caused by HNRNPUL1 depletion triggers increased formation of active spliceosome complexes. This could therefore account for the global decrease in intron retention observed in the poly (A)+ RNA-seq assays performed in the HNRNPUL1-AID line in the previous chapter, as the increased affinity for RNAPII displayed by the spliceosome seemingly overrides the disruptions to snRNP biogenesis that are also occurring when HNRNPUL1 levels are depleted.

5.10 ALS patients with HNRNPUL1 mutations identified

As part of the investigation into the relationship between HNRNPUL1 and ALS, our collaborator at the Sheffield Institute for Translational Neuroscience – Dr. Johnathan Cooper-Knock - analysed the exomes of hundreds of ALS patients using information available from national and international ALS patient databases, in order to screen them for mutations in HNRNPUL1. Once identified, for a HNRNPUL1 mutation to be considering potentially ALS-causing, several criteria were imposed. Firstly, this mutation must be exceedingly rare among the healthy population. Furthermore, mutations were analysed and ranked by CADD score, an algorithm that predicts the potential deleteriousness of mutations. Only mutations that produced high CADD scores were included. In addition, the patients possessing these mutations must also have no other mutations present in known ALS-causing genes. Interestingly, 28 HNRNPUL1 mutations were identified that met all these criteria. There was no obvious clustering of these mutations within a particular domain of HNRNPUL1, instead they reside fairly uniformly throughout the protein (Figure 5.22).
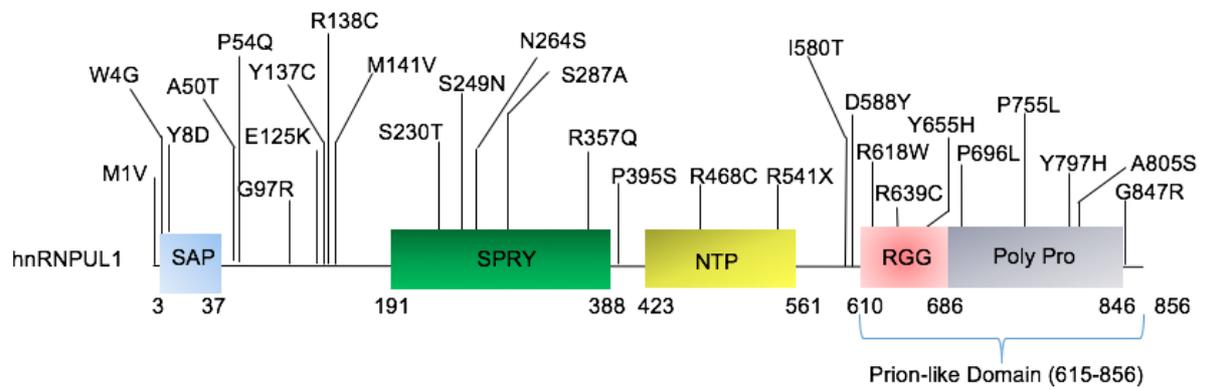
Figure 5.22 HNRNPUL1 mutations identified in ALS patients

Schematic diagram of the domains of HNRNPUL1 and the locations of the potentially ALS-causing mutations identified through screening of national and international ALS patient databases.

5.11 S249N mutation reduces HNRNPUL1 expression levels but does not impact RNAPII CTD Serine-5 phosphorylation

We obtained a lymphoblastoid cell line from Project MinE UK that was generated from one of these patients possessing a HNRNPUL1 mutation, in this case S249N (Figure 5.23A). This enabled us to study whether this patient also displayed any of the phenotypes that we have observed upon HNRNPUL1 knockdown in the HNRNPUL1-AID line. This would indicate whether the S249N mutation affects the ability of the protein to function, and therefore aid in assessing the likelihood that this was an ALS-causing mutant. We also ordered a lymphoblastoid line generated from a healthy individual of similar age to the ALS patient to serve as our control line.

We first wanted to determine whether this patient was homozygous or heterozygous for the S249N mutation. Therefore, genomic DNA was extracted from the LCL of the S249N patient as well as the wild-type control, followed by PCR amplification of the HNRNPUL1 gene and DNA sequencing. In the case of the S249N patient, two peaks were identified in the sequencing tracks at the 2nd nucleotide of the S249 codon (Figure 5.23B). One peak was black, representing the wild-type guanine nucleotide, while the second was green, representing an adenosine. This confirms that the S249N patient is in fact a heterozygote, possessing one of copy of the HNRNPUL1 gene encoding the wild type Serine-249 residue (AGC), and another copy encoding N-249 (AAC).

We next examined the expression of HNRNPUL1 at the protein level in the S249N patient line via Western blot on whole cell lysates. Interestingly, this revealed a clear reduction in the expression level of HNRNPUL1 in the S249N patient compared to control cells (Figure 5.24). This suggests that the S249N mutation destabilises HNRNPUL1, especially as the expression level observed via Western reflects a combination of wild-type and S249N protein due to the heterozygotic status of the patient. We also analysed the levels of RNAPII CTD Ser5 phosphorylation as this is upregulated upon HNRNPUL1 depletion in the HNRNPUL1-AID line. However, we observed no change in the S249 patient line compared to the control (Figure 5.24).
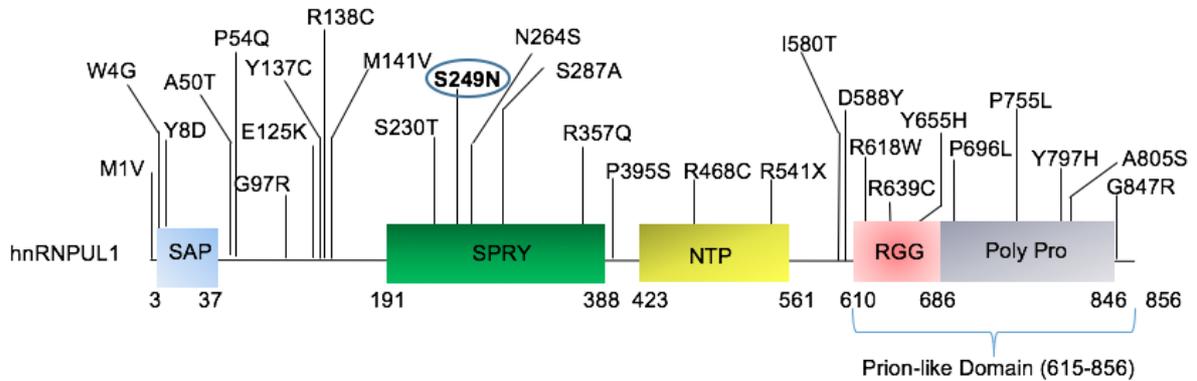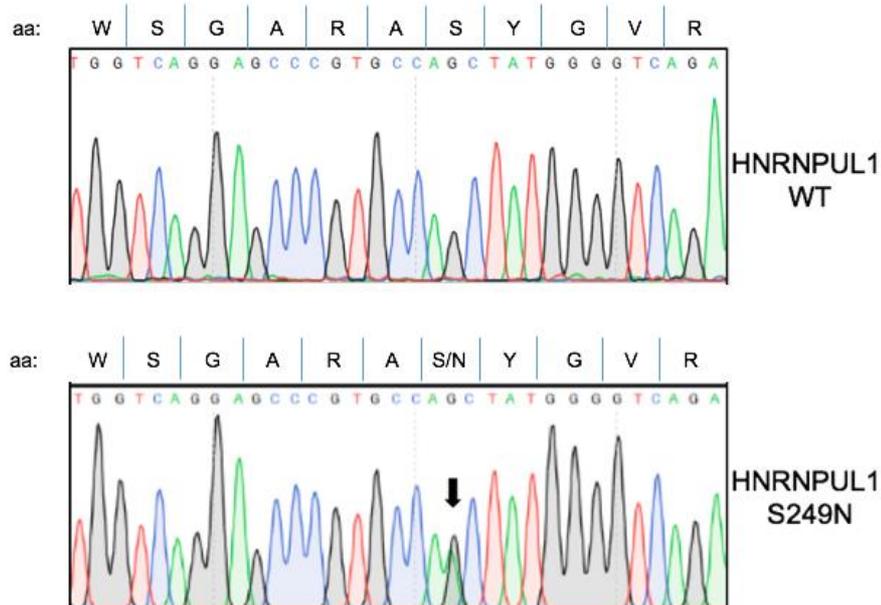
Figure 5.23 Patient S249N is a heterozygote

A – Schematic diagram of HNRNPUL1 highlighting the location of the S249N patient mutation. B – DNA Sanger sequencing traces of genomic DNA isolated from the S249N patient and healthy control LCLs, encompassing the HNRNPUL1 gene region encoding residues 243-253. The presence of both a mini green peak and a mini black peak at the middle nucleotide of residue S249 in the case of the S249N patient indicates that this patient is a heterozygote.
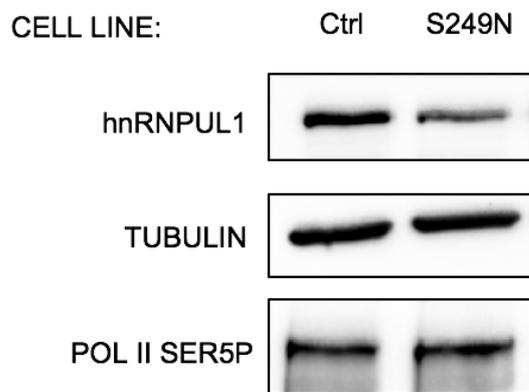
Figure 5.24 HNRNPUL1 protein expression levels are reduced in the S249N patient LCL, while Ser5P RNAPII levels are unchanged

Western analyses of HNRNPUL1 and Ser5P RNAPII levels in the S249N lymphoblastoid cell line. TUBULIN is displayed as a loading control.

5.12 snRNA levels are unchanged in the S249N patient LCL, while snoRNA levels are downregulated

The reduction in snRNA and snoRNA levels observed upon HNRNPUL1 knockdown in the HNRNPUL1-AID line led us to also assess their abundance in the S249N patient LCL. Therefore, total RNA was extracted from this cell line as well as from the healthy control, and a panel of snRNAs and snoRNAs were analysed via qRT-PCR. This revealed that there appears to be no snRNA biogenesis defect as a result of the S249N mutation (Figure 5.25A). In contrast, three out of five snoRNAs tested displayed significant downregulation in the S249N patient LCL compared to the control sample (Figure 5.25B). U3 snoRNA, responsible for processing of ribosomal RNA, was particularly affected. As GAPDH and the snRNAs are unaffected, this suggests this defect is specific to snoRNAs.

5.13 Complementation analyses in the HNRNPUL1-AID line

In addition to analysing the effects of one of the ALS HNRNPUL1 mutations in the cells of the ALS patient itself, the ability of various potentially ALS-causing HNRNPUL1 mutants to complement phenotypes observed in the HNRNPUL1-AID line was also assessed. The Sleeping Beauty (SB) transposon system (Kowarz et al., 2015) was utilised to integrate stable, constitutively expressed forms of HNRNPUL1 within the HNRNPUL1-AID line. Six lines were generated - one HNRNPUL1-AID line expressing the wild-type form of HNRNPUL1 (WT), as well as five mutants (Figure 5.26A, Figure 5.26B). Three of these mutants – A50T, R468C, R639C – are present in the ALS patients identified by Dr. Cooper-Knock, while a HNRNPUL1-AID/SB-HNRNPUL1-Walker A mutant (WA) line and a HNRNPUL1-AID/SB-HNRNPUL1-ΔRGG domain deletion mutant (ΔRGG) line were also created in order to assess the effects of these mutations on the ability of HNRNPUL1 to complement the HNRNPUL1-AID phenotypes as well. In addition, we aimed to generate a HNRNPUL1-AID/SB-HNRNPUL1-S249N line, but unfortunately no successful clones were isolated after

A
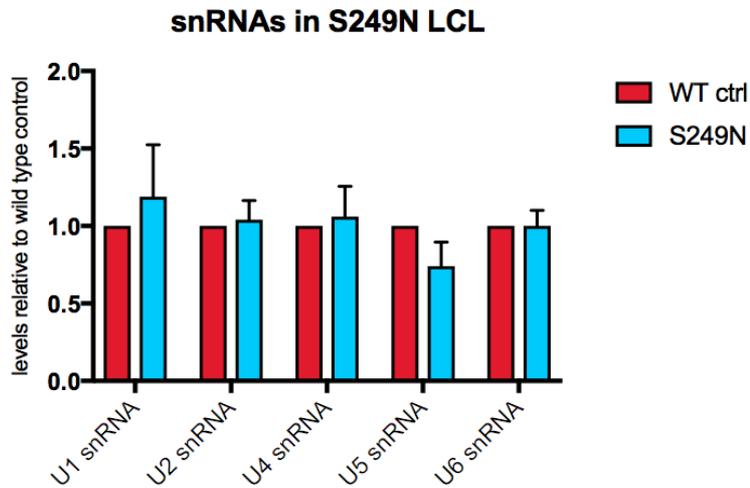
B
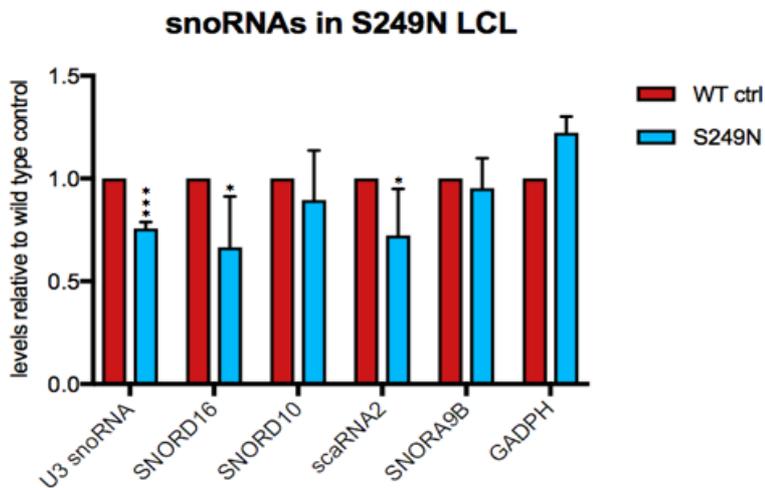
Figure 5.25 snoRNA levels are reduced in the S249N patient LCL, while snRNAs levels are unchanged

A – qRT-PCR analysis of snRNA levels in the S249N patient line compared to an age matched control. Error bars represent the SD. B – qRT-PCR analysis of snoRNA levels in the S249N line compared to an age matched control. Error bars represent the SD.
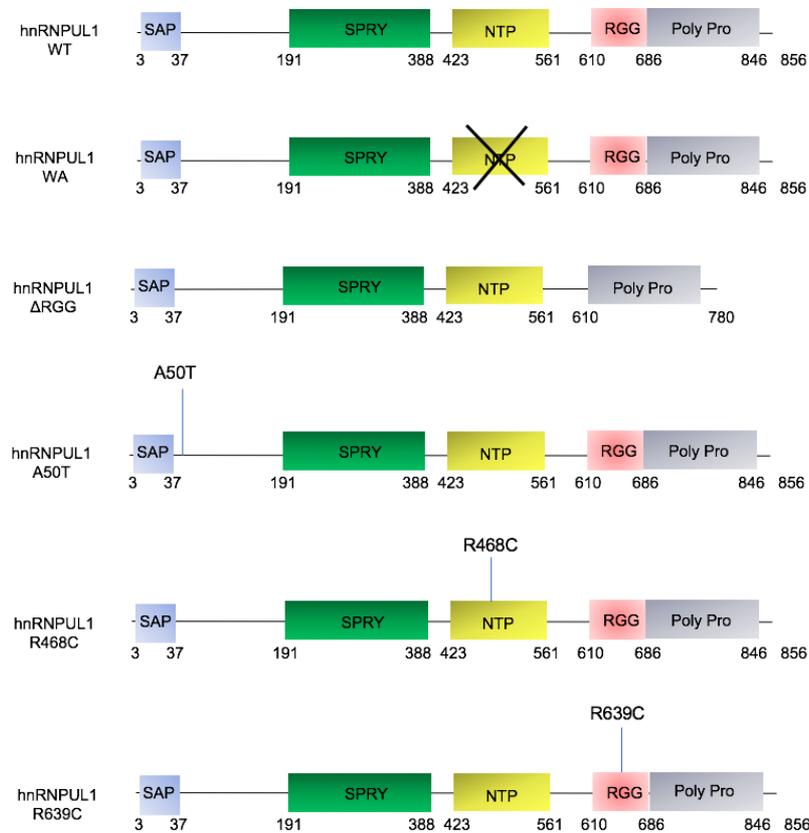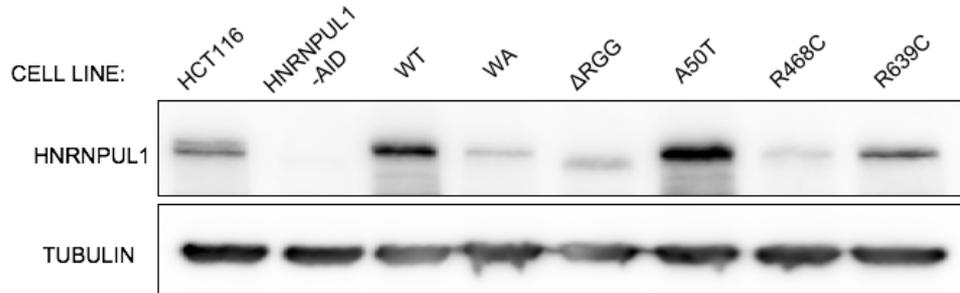
Figure 5.26 Stable expression of HNRNPUL1 WT and various mutants in HNRNPUL1-AID line via the Sleeping Beauty transposon system

A – Schematic diagram the HNRNPUL1 mutants integrated into the HNRNPUL1-AID line via the Sleeping Beauty transposon system. B – Western analyses of the expression levels of these HNRNPUL1 mutants integrated into the HNRNPUL1-AID line. TUBULIN is shown as a loading control.

several attempts and therefore this mutation could not be included in subsequent complementation studies.

To analyse the ability of these various forms of HNRNPUL1 to restore snRNA and snoRNA levels in the HNRNPUL1-AID line when HNRNPUL1-AID itself had been depleted, RNA was extracted from these HNRNPUL1-AID/SB-HNRNPUL1 lines following treatment with doxycycline and auxin. As the SB transposon-mediated integrated copies of HNRNPUL1 are untagged, their expression levels are unaffected by the addition of these hormones, and therefore these constructs were the only form of HNRNPUL1 expressed in the cell at the point of harvesting and RNA extraction. Isolated RNA was converted to cDNA via reverse transcription and the levels of two snRNAs – U1 and U4, and two snoRNAs – scaRNA2 and snoR9AB – that we have previously shown are downregulated upon HNRNPUL1 depletion, were assessed via qRT-PCR. The ability of each form of SB-HNRNPUL1 to complement the small RNA phenotype relative to the original HNRNPUL1-AID line to a statistically significant degree was assessed via ANOVA tests. In the case of two out of four small RNAs, the wild-type form of HNRNPUL1 did partially complement the HNRNPUL1-AID phenotype (Figure 5.27). This suggests that the defects in small RNA biogenesis observed upon HNRNPUL1 depletion in the HNRNPUL1-AID line are a direct result of the absence of the protein as opposed to an off-target effect. In contrast, with the exception of U1 snRNA, each mutant did not complement the HNRNPUL1-AID small RNA phenotypes to a statistically significant degree. Interestingly, the HNRNPUL1 ALS mutant possessing a mutation within its prion-like domain – R639C - did not induce restoration of the levels of any of the transcripts assessed (Figure 5.27). However, given the relatively large error bars, more replicates are required before strong conclusions can be drawn.

We next assessed the levels of three immediate-early genes identified as strongly upregulated in our poly(A+) RNA-seq analysis on the HNRNPUL1-AID line – EGR1, FOSB and NR4A1 - using the same cDNA samples generated for the small RNA
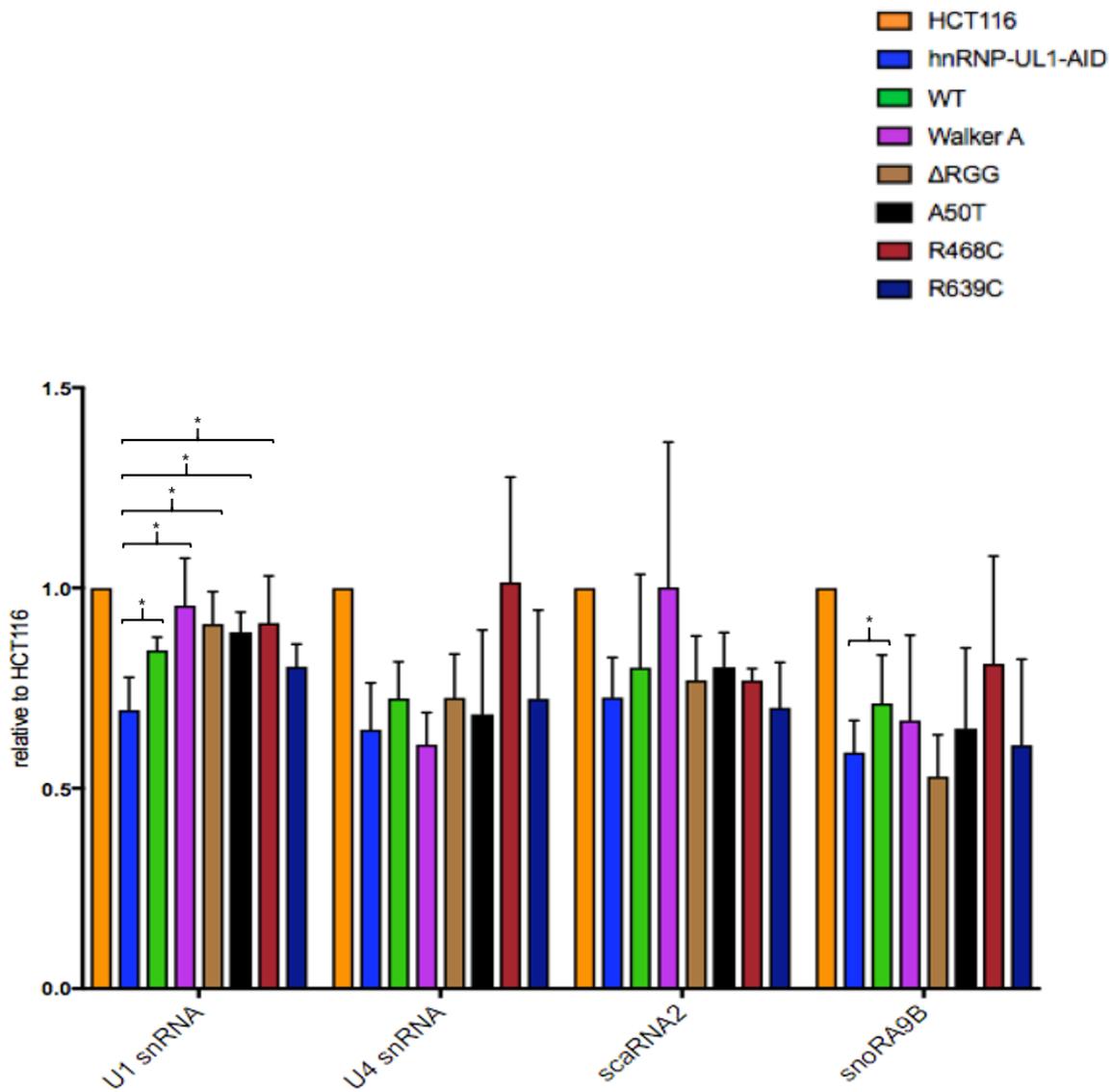
Figure 5.27 Small RNA complementation analyses in the HNRNPUL1-AID/SB-HNRNPUL1 lines

qRT-PCR analyses of snRNA and snoRNA levels following HNRNPUL1-AID depletion in the HNRNPUL1-AID/SB-HNRNPUL1 lines. Errors bars represent the SD.

analysis. In this instance, the wild-type SB-HNRNPUL1 protein rescued the overexpression defect in all three cases (Figure 5.28), again indicating that this phenotype in the HNRNPUL1-AID line is directly related to the absence of HNRNPUL1 upon the addition of auxin. Strikingly, the SB-HNRNPUL1 Walker A mutant exacerbated the HNRNPUL1-AID phenotype of every transcript tested (Figure 5.28). This strongly indicates that nucleotide-binding by HNRNPUL1 is required for it to mediate the repression of immediate-early gene expression. SB-HNRNPUL1-R468C rescued the defect in the case of two transcripts, SB-HNRNPUL1-ΔRGG and SB-HNRNPUL1-R639C each rescued the defect in the case of one transcript, while SB-HNRNPUL1-A50T did not rescue any immediate early gene transcript phenotype. This suggests that these ALS mutations do impact the ability of HNRNPUL1 to regulate immediate-early gene transcription, although once again the large error bars among each sample mean that more sets are required before this is conclusively proven.
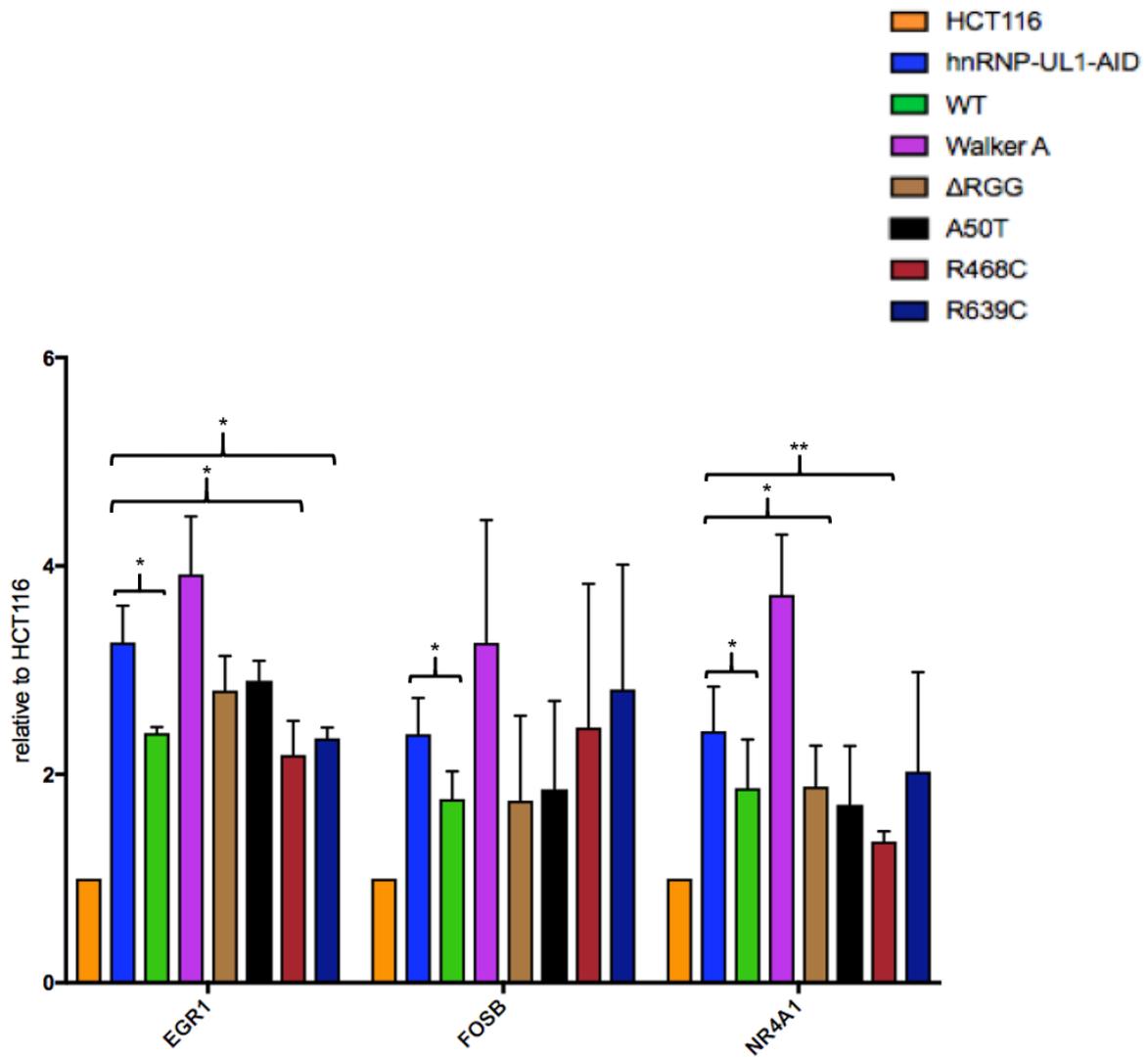
Figure 5.28 – Immediate-early gene complementation analyses in the HNRNPUL1-AID/SB-HNRNPUL1 lines

qRT-PCR analyses of immediate-early gene mRNA levels following HNRNPUL1-AID depletion in the HNRNPUL1-AID/SB-HNRNPUL1 lines. Errors bars represent the SD.

5.14 Summary

During the work described in this chapter, the relationship between HNRNPUL1 and ALS has been explored in detail. Firstly, IP/mass spec analysis performed on HNRNPUL1 identified multiple ALS-causing factors as strong, RNA-independent interactors of the protein. These assays revealed that the prion-like domain of HNRNPUL1 is critical for its ability to form these interactions, as is the case in the majority of its binding partners. In contrast, it has been demonstrated that disrupting the nucleotide-binding ability of HNRNPUL1 strongly enhances its capacity to form protein-protein interactions, with the notable exceptions of the 7SK complex and SART3. HNRNPUL1 nucleotide binding has also been shown to affect the localisation of this protein within the cell, with the Walker A mutant showing a clear enrichment upon chromatin compared to the wild-type form of the protein. It has also been uncovered that knockdown of FUS, mutations of which are a common cause of ALS, results in enhanced association between HNRNPUL1 and its RNA binding partners.

Like other ALS-causing RBPs, HNRNPUL1 was also found to affect the phosphorylation status of RNAPII. However, *in vitro* kinase assays did not detect a direct inhibition of CDK7 by HNRNPUL1, despite it being a substrate of the enzyme itself. It was subsequently demonstrated that the increased phosphorylation of RNAPII CTD Serine-5 induced by HNRNPUL1 depletion triggers an increased RNAPII-U1 snRNP interaction. An increased U1 snRNP-U4 snRNP interaction was also detected, suggesting that the enhanced phosphorylation of RNAPII CTD promotes increased formation of active spliceosome complexes. This is consistent with the mild increases in intron removal observed upon HNRNPUL1 depletion described in the previous chapter. Further experiments are required to determine whether this increased CTD phosphorylation is therefore a cellular response to the snRNP biogenesis defects also caused by HNRNPUL1 knockdown.

We have also identified numerous ALS patients possessing mutations in HNRNPUL1 who do not also have mutations in any known ALS-causing gene. These

HNRNPUL1 mutations are also exceedingly rare among the normal population. Assays performed in a lymphoblastoid cell line of one of these patients, who possessed a heterozygous S249N mutation, revealed that their mutation disrupted the expression levels of HNRNPUL1 and caused a mild snoRNA biogenesis defect.

It was also demonstrated that expressing a wild-type copy of HNRNPUL1 in the HNRNPUL1-AID line can partially rescue the small RNA biogenesis and immediate-early gene expression defects observed in this line, providing further evidence for a direct role for HNRNPUL1 in these pathways. ALS patient HNRNPUL1 mutations A50T, R468C and R639C appeared to mildly impact the ability of the protein to complement the HNRNPUL1-AID line in these assays. However more work is required to establish the significance of these effects. Interestingly, the Walker A mutant form of HNRNPUL1 exacerbated the increased immediate-early gene expression phenotype observed in the HNRNPUL1-AID line, highlighting the importance of the nucleotide-binding capacity of HNRNPUL1 with regards to its function in the cell.

CHAPTER 6 – DISCUSSION

6.1 The Auxin-Inducible Degron as a system to deplete HNRNPUL1

In Chapter 3, the auxin-inducible degron system was utilised to generate a cell line in which HNRNPUL1 could be depleted on a conditional basis. The level of HNRNPUL1 knockdown enabled by this cell line is significantly greater than that achieved by RNAi. Functional assays performed in both a previously generated HNRNPUL1 RNAi line and the HNRNPUL1-AID line demonstrate the importance of maximal depletion when investigating the function of the protein. The HNRNPUL1 RNAi cell line displayed barely any of the phenotypes reported in the ENCODE HNRNPUL1 RNAi RNA-seq data – despite Western analyses showing reasonable knockdown of the protein by RNAi standards (Figure 3.1). In contrast, the HNRNPUL1-AID line did display the expected phenotypes (Figure 3.7). The difference between partial and complete knockdown was also highly apparent in the colony formation assays performed in Chapter 3 (Figure 3.8), which revealed that cell viability is severely reduced upon total HNRNPUL1 knockdown, but was relatively unaffected in the untreated condition, when HNRNPUL1 levels were similar to those achieved following RNAi knockdown.

However, a drawback of the auxin degron system is that even in non-auxin treated conditions, expression of HNRNPUL1-AID was reduced compared to endogenous HNRNPUL1 levels. Although there was not a clear effect on cell viability in the non-auxin treated condition, subtle reductions in HNRNPUL1 functionality may create additional cellular stresses that could impact the outcome of functional assays, while also generating a selection pressure on cells to adapt. Accordingly, we used only low passage number HNRNPUL1-AID cells in order to minimise the likelihood that this issue would affect our results. Alterations to the tag and the use of a doxycycline-inducible form of TIR1 did improve expression levels under basal conditions. However, this meant that a longer induction period was required to deplete the protein. In addition, this also meant that we were unable to use tetracycline-inducible Sleeping

Beauty constructs for subsequent complementation assays, which reduced our ability to control the expression levels of these proteins.

6.2 Upregulation of immediate-early gene expression is triggered by HNRNPUL1 loss

The qRT-PCR analyses performed in the HNRNPUL1-AID line clearly demonstrated that several immediate-early genes (IEGs) are strongly upregulated upon HNRNPUL1 knockdown (Figure 3.9). This finding is consistent with the ENCODE HNRNPUL1 RNAi data, which also displays this phenomenon. Subsequent mNET-seq analyses carried out in the HNRNPUL1-AID line revealed that these genes are activated at the level of transcription when HNRNPUL1 is depleted, which was then confirmed by RNAPII ChIP assays (Figure 3.10).

The cause of this upregulation is still to be determined. It appears unlikely that HNRNPUL1 acts as a direct repressor of transcription at these loci, as ENCODE HNRNPUL1 ChIP-seq data demonstrates that EGR1 is the only one of these genes at which there is significant HNRNPUL1 ChIP signal. A time-course assay was also performed in Chapter 3 with the goal of isolating the stage of IEG expression that was specifically altered upon HNRNPUL1 knockdown, i.e. whether the genes are being induced inappropriately prior to stimulation or whether there is defective repression of these genes following stimulation. However, these assays revealed that the IEGs were upregulated at every phase of the activation and repression cycle (Figure 3.9).

The transcription of these genes is regulated primarily at the stage of pause-release, with RNAPII occupying the TSSs of these genes even in the absence of external or internal stimulus, awaiting phosphorylation via P-TEFb to facilitate productive elongation. This is apparent in the RNAPII mNET-seq tracks in the HCT116 line, which display a large spike in RNAPII signal over the beginning of the IEGs (Figure 3.11). Therefore, the expression of these genes is regulated in large part by the availability of active P-TEFb. Inactive P-TEFb is sequestered by the 7SK complex, preventing inappropriate levels of transcription elongation.

We have demonstrated a clear RNA-dependent interaction between HNRNPUL1 and the 7SK complex, as well as CDK9 – the enzyme component of P-TEFb (Figure 5.8). It is therefore possible that HNRNPUL1 plays a role in regulating the association between this complex and P-TEFb, and that elevated levels of active P-TEFb underpin the upregulation of IEGs observed upon HNRNPUL1 knockdown. These IP assays also revealed that the HNRNPUL1-7SK interaction is dependent on an intact HNRNPUL1 Walker A motif (Figure 5.8). This result is interesting in light of the subsequent complementation assays described in Chapter 5, which demonstrated that the SB-HNRNPUL1 Walker A mutant was unable to suppress the elevated IEG expression phenotype observed in the HNRNPUL1-AID line (Figure 5.23). While the SB-HNRNPUL1-Walker A mutant displayed a lower expression level than wild-type, this expression was sufficient for it to complement the U1 snRNA phenotype, and is comparable to the expression levels of SB-HNRNPUL1-ΔRGG and SB-HNRNPUL1-R468C which both restored IEG levels to a similar degree to SB-HNRNPUL1-WT. This indicates that decreased expression levels do not account for the exacerbated IEG phenotype displayed by the HNRNPUL1-AID/SB-HNRNPUL1-Walker A mutant line. This provides further evidence that the HNRNPUL1-7SK interaction is functionally relevant to the HNRNPUL1-mediated regulation of IEG expression. However, it is important to note that the Walker A mutation also distorts many other HNRNPUL1 interactions as well as its release from chromatin, and therefore further experiments are required to prove this hypothesis. For example, a comparison of the levels of CDK9 co-immunoprecipitating with the 7SK complex in the presence and absence of HNRNPUL1 could be assessed by performing a CDK9 IP in the HNRNPUL1-AID line.

The IEGs that are upregulated upon HNRNPUL1 depletion all play key roles in the cellular response to DNA damage (Christmann and Kaina, 2013; Malewicz et al., 2011; Quiñones et al., 2003), and are therefore switched on in response to elevated levels of DNA damage. HNRNPUL1 has also been implicated in this pathway, whereby it promotes ATR-dependent signalling to trigger repair of double-stranded DNA breaks (Polo et al., 2012). Therefore, an alternative, plausible explanation for the IEG

upregulation is that defective DNA repair caused by HNRNPUL1 knockdown results in increased DNA damage in the HNRNPUL1-AID line, and this in turn causes the increased expression of IEGs. One could test whether there was indeed greater DNA damage in the HNRNPUL1-AID line by analysing the levels of γH2AX – a marker for DNA damage – via Western blot on total lysates.

Interestingly, RBM7 mediates the activation of IEGs in response to DNA damage independently from its role in the NEXT complex, through binding to 7SK in order to promote the release of P-TEFb from this complex (Bugai et al., 2019). As we have demonstrated a strong RNA-dependent HNRNPUL1 interaction with RBM7 (Figure 4.21), one could next analyse the association between RBM7 and 7SK/P-TEFb via RBM7 IP in the HNRNPUL1-AID line to establish whether this axis is dysregulated upon HNRNPUL1 knockdown. However, the fact that the HNRNPUL1 Walker A mutant can bind RBM7 as efficiently as HNRNPUL1 WT (Figure 4.21) but cannot complement the HNRNPUL1-AID line (Figure 5.23) argues against the hypothesis that the IEG upregulation phenotype is related to RBM7.

6.3 The role of HNRNPUL1 in small RNA biogenesis

Over the course of this study, we have identified numerous small RNA biogenesis defects caused by the knockdown of HNRNPUL1 (Figure 6.1). We have demonstrated that HNRNPUL1 is required for efficient transcription of RNAPII-transcribed snRNAs (Figure 4.5), and that this downregulation observed upon HNRNPUL1 depletion causes a reduction in overall snRNA levels (Figure 4.1, Figure 4.17). This is consistent with HNRNPUL1 ENCODE ChIP-seq data that demonstrates a clear HNRNPUL1 ChIP signal over snRNA genes, as well as the strong interaction between HNRNPUL1 and RNAPII reported in Chapter 4 (Figure 4.4).

Given that the HNRNPUL1 ChIP signal peaks over the middle of snRNA loci as well as the TTSs, it appears unlikely that HNRNPUL1 plays a direct role in the initiation of transcription of these genes. There was also an absence of any snRNA transcription
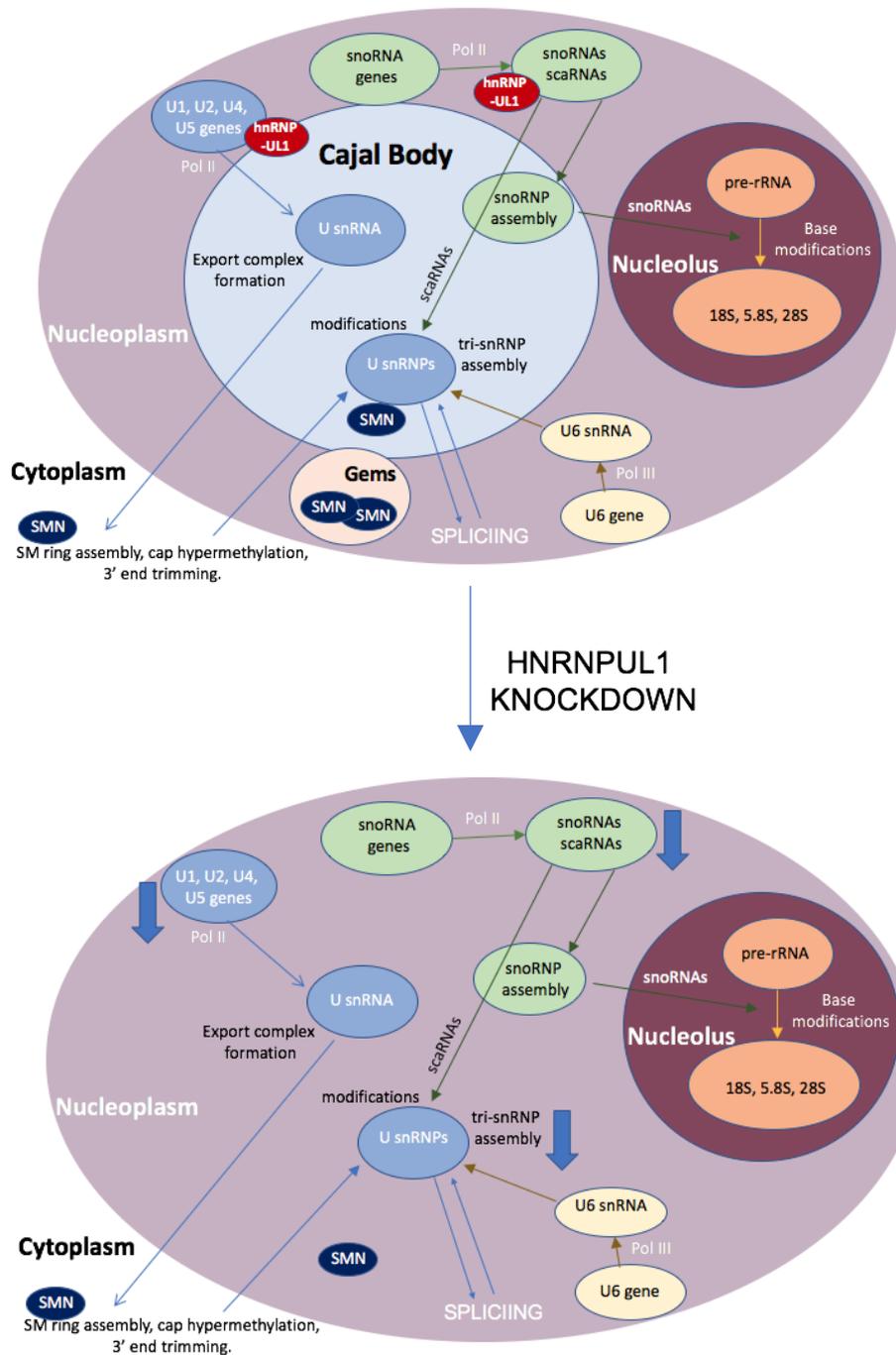
Figure 6.1 – Summary of small RNA biogenesis defects observed upon HNRNPUL1 depletion

HNRNPUL1 is required for efficient snRNA transcription and the integrity of Cajal bodies and nuclear Gems. The disintegration of these structures upon HNRNPUL1 knockdown is accompanied by defects in U4/U6 di-snRNP and U4/U6/U5 tri-snRNP assembly. HNRNPUL1 strongly interacts with snoRNAs, which are also downregulated in the absence of the protein.

204

initiation factors in the HNRNPUL1 IP/mass spec analyses performed in Chapter 5. As in the case of the IEG upregulation phenotype observed upon HNRNPUL1 knockdown, it is plausible that the HNRNPUL1-7SK interaction identified in Chapter 5 underpins the regulation of snRNA transcription by HNRNPUL1. The 7SK complex mediates the recruitment of the little elongation complex to snRNA genes to promote transcription elongation at these loci (Egloff et al., 2017). Knockdown of this complex triggers downregulation of snRNA transcription as a result (Egloff et al., 2017). The complex performs this function while not in association with HEXIM or P-TEFb. The FLAG IPs performed in Chapter 5 clearly demonstrate that HNRNPUL1 interacts more strongly with 7SK core components MePCE and LARP7 than HEXIM and P-TEFb (Figure 5.8), perhaps indicating that snRNA genes are the site of HNRNPUL1-7SK interaction. As the SB-HNRNPUL1-Walker A mutant could restore U1 snRNA levels in the HNRNPUL1-AID line but not U4 snRNA, it is still unclear how significant this domain (and therefore the 7SK interaction) is with regards to the snRNA biogenesis function of HNRNPUL1. Testing a larger panel of snRNAs in future complementation assays may resolve this uncertainty.

In the case of the IEGs (and other large protein-encoding genes) one could distinguish between a transcription initiation defect and an elongation defect via RNAPII ChIP or mNET-seq. However, the resolution of the RNAPII ChIP assays performed in Chapter 4 (Figure 4.5) was not great enough to distinguish between a reduction in RNAPII signal over an snRNA promoter compared to a reduction in the body of the gene given the small size of snRNA genes. Attempts to analyse RNAPII occupancy over snRNA loci via mNET-seq analyses were complicated by snRNA transcripts themselves co-purifying with RNAPII and distorting the signal over these genes. An alternative approach could be to perform ChIP on an initiation factor such as TBP as well as ChIP on an elongation factor such as a 7SK protein component in the HCT116 and HNRNPUL1-AID lines. A large reduction in TBP recruitment to snRNA loci upon HNRNPUL1 knockdown would suggest that the snRNA transcription defect preceded the elongation phase.

Despite the HNRNPUL1 ChIP-seq peak over snRNA transcription termination sites, no significant defects in transcription termination and 3'-end processing were detected upon HNRNPUL1 knockdown, suggesting that HNRNPUL1 does not directly operate in these processes. However, 3'-end extended snRNAs are rapidly targeted for degradation by the NEXT complex. Therefore, it may be advisable to repeat this assay while depleting this complex via RNAi, in order to ensure that any increase in the levels of these 3'-end extended transcripts caused by HNRNPUL1 knockdown is not masked by an accompanying increase in degradation.

Another potential cause of the snRNA transcription defect is the elevated levels of Ser5P on RNAPII CTD resulting from HNRNPUL1 knockdown identified in Chapter 5 (Figure 5.12). At snRNA genes, RPAP2 dephosphorylates Serine-5 and recruits Integrator subunits to trigger productive elongation, with knockdown of this phosphatase also resulting in reduced snRNA transcription (Egloff et al., 2012). Persistent Serine-5 phosphorylation is thought to decrease RNAPII transcription re-initiation rates as well as prevent the dissociation of the capping enzyme (Reyes-Reyes and Hampsey, 2007), both which cause reduced transcription. In addition, the recruitment of Integrator subunit Ints11 to snRNA genes is repressed by the presence of Ser5P, which therefore acts as a barrier to 3'-end processing and transcription termination events (Egloff et al., 2010). It is currently unclear whether the global upregulation in Ser5P levels observed upon HNRNPUL1 depletion is specific to particular genes/region of genes, or if this defect is distributed uniformly. This could be determined by performing Ser5P ChIP or Ser5P mNET-seq and comparing the signal with that of total RNAPII over specific genes, e.g. at snRNA loci.

Assessing whether Ser5P upregulation occurs within intronless genes in these assays would also help determine whether the Ser5P upregulation phenotype was related to or independent of the splicing process. Ser5P interacts with spliceosome components, and given the downregulation of snRNAs observed in the HNRNPUL1-AID line we considered the possibility that the increased Ser5P levels are a

compensatory cellular response to promote enhanced spliceosome recruitment. However, RNAPII IPs revealed an increased RNAPII-spliceosome interaction upon HNRNPUL1 knockdown (Figure 5.16), and this explanation would not account for the upregulated Ser7P levels also displayed in the HNRNPUL1-AID line (Figure 5.12). Given that snRNPs have long half-lives and are in vast excess (Lund and Dahlberg, 1984; Tardiff and Rosbash, 2006), it would be interesting in the future to assess whether this increased RNAPII-spliceosome interaction can be maintained after several days of HNRNPUL1 knockdown, as it may require this long to reduce snRNP levels to a significant degree. Similarly, although a longer post-auxin addition time-point was chosen to harvest cells during the poly(A) RNA-seq experiment, the lack of clear splicing defect suggests that there were sufficient levels of snRNP complexes within the cell at that time to maintain splicing fidelity.

As both Ser5 and Ser7 are CDK7 targets, *in vitro* kinase assays were performed to assess the ability of HNRNPUL1 to inhibit this enzyme. This revealed that HNRNPUL1 itself was a CDK7 substrate (Figure 5.15), but only the HNRNPUL1 Walker A mutant clearly reduced RNAPII CTD phosphorylation. It may be the case that additional proteins are required to enable HNRNPUL1 WT to efficiently inhibit the kinase and/or interact with RNAPII CTD, which were not present in the *in vitro* assay. One could also investigate the relationship between HNRNPUL1 and Ssu72 – a Ser5P and Ser7P phosphatase, as decreased activity of this protein upon HNRNPUL1 knockdown could alternatively account for the increased phosphorylation phenotype.

It has previously been reported that Cajal bodies mediate specific genome organisation events to promote snRNA transcription (Wang et al., 2016). Knockdown of a core Cajal body component that triggers the disintegration of these structures therefore also results in decreased snRNA transcription rates (Wang et al., 2016). The clear reduction in Cajal body number observed in the HNRNPUL1-AID line (Figure 4.8) led us to consider the possibility that HNRNPUL1 was a previously unreported core Cajal body component, and that this disintegration of CBs upon HNRNPUL1

knockdown explained the reduction in snRNA transcription. However, given the lack of strong coIP between HNRNPUL1 and coilin (Figure 4.9), as well as the lack of clear HNRNPUL1 enrichment at CBs when assessed by immunostaining (Figure 4.11), this appears unlikely. It has been widely reported that knockdown of factors involved in snRNA transcription results in the reduction of CB numbers (Lemm et al., 2006; Takata et al., 2012), and in light of these findings, it seems more probable that an snRNA transcription defect caused by the absence of HNRNPUL1 then triggers the disintegration of CBs.

We have also demonstrated that there is a clear reduction in the assembly rates of the U4/U6 di-snRNP as well as the U4/U6/U5 tri-snRNP upon HNRNPUL1 knockdown (Figure 4.14). This is to be expected given the disintegration of CBs, where the assembly of these complexes has been reported to occur at a 10-fold faster rate compared to the nucleoplasm (Novotný et al., 2011). It is presently unclear whether HNRNPUL1 also directly impacts this assembly process. We have demonstrated that HNRNPUL1 strongly interacts with U4 snRNA (Figure 4.5), along with key U4/U6 di-snRNP recycling factor SART3 (Figure 5.8), suggesting a direct role in this pathway. However, the unusual HNRNPUL1-SART3 binding profile displayed in the HNRNPUL1 FLAG IPs (Figure 5.8) – mutating the Walker A motif almost completely disrupts the interaction – closely resembles other 7SK components as opposed to the HNRNPUL1-U4 snRNP binding pattern (Figure 4.15). This suggests that the HNRNPUL1-SART3 interaction occurs within the context of the 7SK-related function of SART3. Assessing the levels of U4 and U6 snRNAs interacting with SART3 via RIP in the HNRNPUL1-AID line could determine whether HNRNPUL1 does indeed directly impact the U4/U6 di-snRNP assembly process. In addition, determining whether the HNRNPUL1-SART3 interaction is 7SK dependent could be performed through coIP following siRNA knockdown of 7SK snRNA.

Unlike snRNAs, HNRNPUL1 appears to regulate snoRNA biogenesis at the post-transcription stage. mNET-seq analyses demonstrated that there is no reduction in

transcription of snoRNA-containing pre-mRNAs upon HNRNPUL1 knockdown (Figure 4.20), and in contrast to snRNAs there is not significant HNRNPUL1 ChIP signal over these genes. However, HNRNPUL1 does strongly interact with snoRNA transcripts themselves, especially scaRNAs, as evidenced by a clear enrichment of HNRNPUL1 eCLIP signal on these classes of RNA (Figure 4.10). This interaction appears to promote the expression of these transcripts, as there is a clear reduction in their abundance upon HNRNPUL1 knockdown (Figure 4.16, Figure 4.19). qRT-PCR and small RNA-seq analyses did not detect elevated levels of 5'-end or 3'-end extended precursors (Figure 4.16, Figure 4.19), suggesting that HNRNPUL1 does not directly impact on the processing of pre-snoRNAs. Alternatively, mis-processed precursors may be immediately degraded via the NEXT/exosome pathway and therefore would be undetectable in those assays. To assess whether increased degradation of precursor and mature snoRNAs account for the decline in their levels upon HNRNPUL1 knockdown, one could inactivate the degradation pathway via siRNA-mediated knockdown of one or more of the NEXT/exosome complex components in conjunction with HNRNPUL1 depletion and assay whether this restores snoRNA levels.

As many of the interactions between HNRNPUL1 and the RNA degradation machinery appear RNA-dependent (Figure 5.7), this would suggest that these proteins interact with and potentially compete for the same RNA substrates, as opposed to HNRNPUL1 directly binding these complexes to inhibit their activity. It is also striking that the strongest interactor between HNRNPUL1 and the NEXT/exosome complex components is RBM7 (Figure 4.21), which is the component responsible for binding the pre-snoRNA substrates themselves and targeting them for degradation (Lubas et al., 2015), while the interaction between HNRNPUL1 and other components of the complex is much weaker (Figure 4.21). This again indicates that HNRNPUL1 encounters the NEXT complex via the RNA substrate. Cross-referencing HNRNPUL1 eCLIP profiles with an analysis of the transcripts downregulated in the HNRNPUL1-AID line poly(A) and small RNA-seq samples may further clarify whether there is a

general inhibition of NEXT mechanism occurring or if competition for substrate binding is key. RBM7 RIP assays in the HNRNPUL1-AID line could also determine whether increased NEXT recruitment to pre-snoRNAs occurs upon HNRNPUL1 knockdown, although these assays could be complicated by the fact that the RNA substrates will be degraded shortly after HNRNPUL1 depletion.

Increased degradation of snoRNAs could also potentially be triggered by the disintegration of CBs that results from HNRNPUL1 knockdown. As all snoRNAs traffic through these structures during their biogenesis (Machyna et al., 2014), disrupting this transport pathway could cause these transcripts to be targeted for degradation, given that it is becoming increasingly clear that the levels of these small RNAs are determined by a balance between competing degradation and transport pathways (Giacometti et al., 2017). Whether disintegration of CBs is sufficient to induce the reduction in snoRNA levels observed in the HNRNPUL1-AID line could be tested through siRNA-mediated knockdown of a core CB component such as WRAP53 or USPL1 followed by qRT-PCR analyses.

During the complementation assays performed in Chapter 5, the SB-HNRNPUL1-WT protein restored the levels of two out of four of the small RNA transcripts tested to a statistically significant degree, providing further evidence that HNRNPUL1 plays a direct role in the small RNA biogenesis pathway. However, the restoration was not complete. One potential explanation for this is that the SB-HNRNPUL1-WT protein was overexpressed relative to the levels of endogenous HNRNPUL1. If more time was available, more clones would have been screened during the generation of the cell line in order to select one that displayed SB-HNRNPUL1 WT expression that was closer to the endogenous levels of the protein.

6.4 HNRNPUL1 as an ALS protein

There is a growing body of literature linking defects in snRNP biogenesis to both ALS and SMA [reviewed in (Cauchi, 2014)]. Loss of nuclear SMN-containing Gems has

long been recognized as a hallmark of SMA, and in recent years several studies have also identified this phenomenon as a feature of ALS (Ishihara et al., 2013; Kariya et al., 2012; Tsuiji et al., 2013; Yamazaki et al., 2012). ALS-causing forms of TDP-43 and FUS have been shown to both disrupt SMN function and trigger a decrease in Gem number (Groen et al., 2013; Ishihara et al., 2013), while cytoplasmic mis-localised FUS also causes snRNPs to accumulate in the cytoplasm (Groen et al., 2013). The work in this study clearly identifies HNRNPUL1 as a key component of the snRNP biogenesis pathway. Like other ALS-causing RNA-binding factors, we have shown that the loss of HNRNPUL1 results in a dramatic reduction in the number of Gems per cell (Figure 4.8). Consistent with this observation, the interaction between core CB component coilin and SMN is significantly reduced in HNRNPUL1-depleted cells (Figure 4.9). Interestingly, it has also been reported that the disruption of this interaction is also triggered by ALS-causing mutant forms of SOD1 (Kariya et al., 2012). It is striking how closely the cellular consequences of HNRNPUL1 loss resemble the phenotypes produced by ALS-causing proteins, and these findings suggest that mutant forms of HNRNPUL1 are likely to be associated with neurodegeneration.

We have demonstrated in this study that the prion-like domain of HNRNPUL1 performs a similar function to the PrLDs found in other ALS-causing binding factors. IP/mass spectrometry assays detailed that this region is critical for the formation of many of HNRNPUL1's interactions, including with other ALS-causing RBPs that bind HNRNPUL1 in an RNA-independent manner (Figure 5.1, Figure 5.5). Interestingly, given the widespread literature on the role of RGG boxes in RNA-binding, mRNP captures assays described in Chapter 3 demonstrated that HNRNPUL1 can efficiently bind mRNA without its PrLD (Figure 3.14). This experiment revealed that disrupting the ability of HNRNPUL1 to bind nucleotides also has a limited impact in the RNA-binding capacity of the protein. These findings strongly imply that more than one domain of HNRNPUL1 has RNA-binding capacity. A recent study revealed that the NTP-binding region of HNRNPU, which is highly homologous in sequence to that of HNRNPUL1, is indeed capable of binding RNA (Panhale et al., 2019). It is also

important to note that the mRNP capture assay is global and non-quantitative, so subtle effects on RNA-binding ability may not be recognised. It could well be the case that the Walker A and PrLD mutations alter which mRNAs HNRNPUL1 targets, and these changes would not be detected by this assay. In addition, only messenger RNAs are analysed, and given the role of HNRNPUL1 in small RNA biogenesis identified in this study this is clearly an incomplete representation of the overall RNA interactome of HNRNPUL1. Incorporating these mutants into RIP/RIP-seq assays would enable both the identification of minor changes in RNA-binding affinities as well as the assessment of transcript interactions on an individual basis. Isolated HNRNPUL1 domains could also be purified and tested in *in vitro* RNA-binding assays for RNA-binding capacity.

We have also demonstrated that the interaction between HNRNPUL1 and ALS-causing RBP FUS inhibits the RNA-binding capacity of HNRNPUL1 (Figure 5.10). These assays were performed in a cell line where FUS expression is permanently knocked out. Thus, there is a very strong selective pressure on these cells to develop adaptations to counter the loss of this important protein. Therefore, confirming this finding by repeating the assay in a system where FUS levels have been downregulated via an alternative mechanism such as siRNA may be prudent. If confirmed, it would be interesting to test in future whether ALS-causing mutations in FUS disrupt this ability to regulate the RNA-binding capacity of HNRNPUL1. If this is indeed the case, then dysregulated HNRNPUL1 RNA-binding may be an as yet unrecognised component of the molecular pathology underpinning cases of FUS-related ALS.

Disrupting the nucleotide-binding capacity of HNRNPUL1 strongly increases its ability to associate with several ALS-causing proteins, as well as a variety of other transcription and splicing factors (Figure 5.2, Figure 5.3). Although this mutant also displays an increased chromatin association (Figure 5.4), the enhanced binding between HNRNPUL1 and these chromatin-associated factors is unlikely to be simply

due to increased proximity, as previous work in the Wilson lab has demonstrated that the HNRNPUL1 Walker A mutant also binds TAF15 more strongly *in vitro* (Wilson lab, unpublished). Therefore, the increased binding observed reflects a genuine increase in affinity between HNRNPUL1 and these proteins in the absence of an associated nucleotide.

In light of the many properties shared between HNRNPUL1 and other ALS-causing RNA-binding factors, the identification of ALS patients possessing HNRNPUL1 mutations by our collaborators at SITraN is an important discovery. Several criteria were imposed to decrease the likelihood that the mutations identified were not random polymorphisms. However, to ultimately prove that these mutations are ALS-causing, for each case a family pedigree that demonstrated clear linkage between the inheritance of the mutation and the manifestation of the condition would need to be generated. This would not be possible for the small percentage of these HNRNPUL1 cases identified that appear sporadic, however the A50T, S249N, R468C and R639C mutations investigated in this study are all present in patients with the familial form of the condition, and therefore the generation of a family pedigree tracking these mutations could be undertaken.

Interestingly, HNRNPUL1 expression levels were reduced in the LCL derived from the ALS patient possessing the S249N mutation (Figure 5.19), and a minor snoRNA defect was identified (Figure 5.20). It would be interesting to further investigate this snoRNA defect by performing small RNA-seq in this patient LCL in the future, to assess whether this downregulation occurs globally. It is perhaps unsurprising that this phenotype was modest compared to the HNRNPUL1-AID line, given that this patient was identified as a heterozygote, possessing one normal copy of the HNRNPUL1 gene (Figure 5.18). This may also account for the lack of Ser5P phosphorylation defect observed in Western analyses performed on total cell lysates harvested from these cells. In studies examining the effect of FUS knockdown on CTD phosphorylation, CTD phosphorylation changes were also too subtle to be identified

via this method. Instead, ChIP-seq assays were performed using antibodies specific to phosphorylated isoforms of RNAPII, and only then were defects recognised (Schwartz et al., 2012). Therefore, it may be advisable to perform Ser5P ChIP-seq assays in this LCL to definitively determine whether the S249N HNRNPUL1 mutation does affect the levels and distribution of this post-translational modification.

HNRNPUL1-AID/SB-HNRNPUL1 complementation lines have been generated for three ALS mutants – A50T, R468C and R639C. As the wild-type form of the protein did not fully complement the HNRNPUL1-AID line with regards to the RNA transcripts tested via qRT-PCR, likely as a result of overexpression, it is difficult to draw strong conclusions concerning the effects of these patient mutations. Increasing the number of replicates in the assay would likely reduce the error bars generated and therefore more subtle effects will be more easily identifiable.

Nevertheless, it is interesting to note that the SB-HNRNPUL1-R639C mutant did not complement any of the small RNAs analysed, and only EGR1 to a statistically significant degree (Figure 5.23). This mutation lies within the PrLD of HNRNPUL1 that we have shown is critical for forming many of HNRNPUL1's interactions. This residue has also been identified as a likely arginine methylation target (Gurunathan et al., 2015). Therefore, it would be of great interest to examine the impact of this mutation on the interactome of HNRNPUL1 in future studies.

However, it is possible that the mechanism by which these mutations inhibit the function of HNRNPUL1 is via destabilisation of the protein and resulting decreased expression, as displayed by the S249N mutation in the LCL derived from the ALS patient possessing that mutation. Hence, this reduction in expression would have to be mimicked in functional assays involving that mutant for any defects to be observed. In addition, ALS is in most cases an adult-onset disorder, suggesting that ALS-causing mutations initially trigger only minor cellular insults, which then build up over a long period of time before becoming fully neurotoxic. Therefore, experiments performed with relatively young cells, such as the complementation assays described in Chapter

5, may have been conducted over an inappropriately short timescale to identify the adverse effects of these mutations. Finally, as ALS symptoms are generated by injury and cell death of motor neurons, it would be advisable to conduct future functional assays in neuronal or neuronal-like cells.

REFERENCES

Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M., and Lührmann, R. (1999). A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. EMBO J. 18, pp. 5789-5802.

Agafonov, D.E., Kastner, B., Dybkov, O., Hofele, R. V., Liu, W.T., Urlaub, H., Lührmann, R., and Stark, H. (2016). Molecular architecture of the human U4/U6.U5 tri-snRNP. Science 351, pp. 1416-1420.

Andersen, P.R., Domanski, M., Kristiansen, M.S., Storvall, H., Ntini, E., Verheggen, C., Schein, A., Bunkenborg, J., Poser, I., Hallais, M., et al. (2013). The human cap-binding complex is functionally connected to the nuclear RNA exosome. Nat. Struct. Mol. Biol. 20, pp. 1367-1376

Ann Ran, F., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nature. 8, pp. 2281-2308

Arnold, E.S., and Fischbeck, K.H. (2018). Spinal muscular atrophy. In Handbook of Clinical Neurology, 148, pp. 591-601.

Bachi, A., Braun, I.C., Rodrigues, J.P., Panté, N., Ribbeck, K., Von Kobbe, C., Kutay, U., Wilm, M., Görlich, D., Carmo-Fonseca, M., et al. (2000). The C-terminal domain of TAP interacts with the nuclear pore complex and promotes export of specific CTE-bearing RNA substrates. RNA. 6, pp. 136-158.

Bahrami, S., and Drabløs, F. (2016). Gene regulation in the immediate-early response process. Adv. Biol. Regul. 62, pp. 37–49.

Baillat, D., Hakimi, M.A., Näär, A.M., Shilatifard, A., Cooch, N., and Shiekhattar, R. (2005). Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. Cell. 123, pp. 265–276.

Barral, P.M., Rusch, A., Turnell, A.S., Gallimore, P.H., Byrd, P.J., Dobner, T., and

Grand, R.J.A. (2005). The interaction of the hnRNP family member E1B-AP5 with p53. FEBS Lett. 579, pp. 2752-2758.

Bartkowiak, B., Liu, P., Phatnani, H.P., Fuda, N.J., Cooper, J.J., Price, D.H., Adelman, K., Lis, J.T., and Greenleaf, A.L. (2010). CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. Genes Dev. 24, pp. 2303–2316.

Battle, D.J., Lau, C.K., Wan, L., Deng, H., Lotti, F., and Dreyfuss, G. (2006). The Gemin5 Protein of the SMN Complex Identifies snRNAs. Mol. Cell. 23, pp. 273–279.

Bell, M., Schreiner, S., Damianov, A., Reddy, R., and Bindereif, A. (2002). p110, a novel human U6 snRNP protein and U4/U6 snRNP recycling factor. EMBO J. 21, pp. 2724-2735.

Berndt, H., Harnisch, C., Rammelt, C., Stöhr, N., Zirkel, A., Dohm, J.C., Himmelbauer, H., Tavanez, J.P., Hüttelmaier, S., and Wahle, E. (2012). Maturation of mammalian H/ACA box snoRNAs: PAPD5-dependent adenylation and PARN-dependent trimming. RNA.18, pp. 958–972.

Bernstein, N.K., Williams, R.S., Rakovszky, M.L., Cui, D., Green, R., Karimi-Busheri, F., Mani, R.S., Galicia, S., Koch, C.A., Cass, C.E., et al. (2005). The molecular architecture of the mammalian DNA repair enzyme, polynucleotide kinase. Mol. Cell. 17, pp. 657-670.

Beyer, A.L., Christensen, M.E., Walker, B.W., and LeStourgeon, W.M. (1977). Identification and characterization of the packaging proteins of core 40S hnRNP particles. Cell. 11, pp. 127–138.

Bizarro, J., Dodré, M., Huttin, A., Charpentier, B., Schlotter, F., Branlant, C., Verheggen, C., Massenet, S., and Bertrand, E. (2015). NUFIP and the HSP90/R2TP chaperone bind the SMN complex and facilitate assembly of U4-specific proteins. Nucleic Acids Res. 43, pp. 8973–8989.

Boulon, S., Verheggen, C., Jady, B.E., Girard, C., Pescia, C., Paul, C., Ospina, J.K.,

Kiss, T., Matera, A.G., Bordonné, R., et al. (2004). PHAX and CRM1 are required sequentially to transport U3 snoRNA to nucleoli. Mol. Cell. 16, pp. 777–787.

Bugai, A., Quaresma, A.J.C., Friedel, C.C., Lenasi, T., Düster, R., Sibley, C.R., Fujinaga, K., Kukanja, P., Hennig, T., Blasius, M., et al. (2019). P-TEFb Activation by RBM7 Shapes a Pro-survival Transcriptional Response to Genotoxic Stress. Mol. Cell. 74, pp. 254-267.

Calo, E., Flynn, R.A., Martin, L., Spitale, R.C., Chang, H.Y., and Wysocka, J. (2015). RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. Nature. 518, pp. 249–253.

Cantarero, L., Sanz-García, M., Vinograd-Byk, H., Renbaum, P., Levy-Lahad, E., and Lazo, P.A. (2015). VRK1 regulates Cajal body dynamics and protects coilin from proteasomal degradation in cell cycle. Sci. Rep. 5:10543.

Carmo-Fonseca, M., Pepperkok, R., Sproat, B.S., Ansorge, W., Swanson, M.S., and Lamond, A.I. (1991). In vivo detection of snRNP-rich organelles in the nuclei of mammalian cells. EMBO J. 10, pp. 1863-1873.

Carpenter, B., MacKay, C., Alnabulsi, A., MacKay, M., Telfer, C., Melvin, W.T., and Murray, G.I. (2006). The roles of heterogeneous nuclear ribonucleoproteins in tumour development and progression. Biochim. Biophys. Acta - Rev. Cancer. 42, pp. 295–309.

Carvalho, T., Almeida, F., Calapez, A., Lafarga, M., Berciano, M.T., and Carmo-Fonseca, M. (1999). The spinal muscular atrophy disease gene product, SMN: A link between snRNP biogenesis and the Cajal (coiled) body. J. Cell Biol. 147, pp. 715–728.

Cauchi, R.J. (2014). Gem depletion: Amyotrophic lateral sclerosis and spinal muscular atrophy crossover. CNS Neurosci. Ther. 20, pp. 574-581.

Chi, B., O'Connell, J.D., Iocolano, A.D., Coady, J.A., Yu, Y., Gangopadhyay, J., Gygi, S.P., and Reed, R. (2018a). The neurodegenerative diseases ALS and SMA are linked at the molecular level via the ASC-1 complex. Nucleic Acids Res. 46, pp. 11939–11951.

Chi, B., O'Connell, J.D., Yamazaki, T., Gangopadhyay, J., Gygi, S.P., and Reed, R.

(2018b). Interactome analyses revealed that the U1 snRNP machinery overlaps extensively with the RNAP II machinery and contains multiple ALS/SMA-causative proteins. Sci. Rep. 8:8755.

Chiò, A., Logroscino, G., Traynor, B.J., Collins, J., Simeone, J.C., Goldstein, L.A., and White, L.A. (2013). Global epidemiology of amyotrophic lateral sclerosis: A systematic review of the published literature. Neuroepidemiology. 41, pp. 118–130.

Cho, E.J., Takagi, T., Moore, C.R., and Buratowski, S. (1997). mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. Genes Dev. 11, pp. 3319-3326.

Christmann, M., and Kaina, B. (2013). Transcriptional regulation of human DNA repair genes following genotoxic stress: Trigger mechanisms, inducible responses and genotoxic adaptation. Nucleic Acids Res. 41, pp. 8403–8420.

Coady, T.H., and Lorson, C.L. (2011). SMN in spinal muscular atrophy and snRNP biogenesis. Wiley Interdiscip. Rev. RNA. 2, pp. 546-564.

Cooper-Knock, J., Robins, H., Niedermoser, I., Wyles, M., Heath, P.R., Higginbottom, A., Walsh, T., Kazoka, M., Ince, P.G., Hautbergue, G.M., et al. (2018). Targeted Genetic Screen in Amyotrophic Lateral Sclerosis Reveals Novel Genetic Variants with Synergistic Effect on Clinical Phenotype. Front. Mol. Neurosci. 10:370.

Darzacq, X., and Kiss, T. (2002). Processing of Intron-Encoded Box C/D Small Nucleolar RNAs Lacking a 5',3'-Terminal Stem Structure. Mol. Cell. Biol. 20, pp. 4522-4531.

Darzacq, X., Jády, B.E., Verheggen, C., Kiss, A.M., Bertrand, E., and Kiss, T. (2002). Cajal body-specific small nuclear RNAs: A novel class of 2′-O-methylation and pseudouridylation guide RNAs. EMBO J. 21, pp. 2746–2756.

DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., Boxer, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.C.A., Flynn, H., Adamson, J., et al. (2011). Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72

Causes Chromosome 9p-Linked FTD and ALS. Neuron. 72, pp. 245–256.

Dewey, C.M., Cenik, B., Sephton, C.F., Dries, D.R., Mayer, P., Good, S.K., Johnson, B.A., Herz, J., and Yu, G. (2011). TDP-43 Is Directed to Stress Granules by Sorbitol, a Novel Physiological Osmotic and Oxidative Stressor. Mol. Cell. Biol. 31, pp. 1098–1108.

Dias, J.D., Rito, T., Triglia, E.T., Kukalev, A., Ferrai, C., Chotalia, M., Brookes, E., Kimura, H., and Pombo, A. (2015). Methylation of RNA polymerase II non-consensus Lysine residues marks early transcription in mammalian cells. Elife. 4, e11215.

Doktor, T.K., Hua, Y., Andersen, H.S., Brøner, S., Liu, Y.H., Wieckowska, A., Dembic, M., Bruun, G.H., Krainer, A.R., and Andresen, B.S. (2017). RNA-sequencing of a mouse-model of spinal muscular atrophy reveals tissue-wide changes in splicing of U12-dependent introns. Nucleic Acids Res. 45, pp. 395-416.

Dönmez, G., Hartmuth, K., and Lührmann, R. (2004). Modified nucleotides at the 5′ end of human U2 snRNA are required for spliceosomal E-complex formation. RNA. 10, pp. 1925-1933.

Dupuis-Sandoval, F., Poirier, M., and Scott, M.S. (2015). The emerging landscape of small nucleolar RNAs in cell biology. Wiley Interdiscip. Rev. RNA. 6, pp. 381-397.

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

Egloff, S., O'Reilly, D., and Murphy, S. (2008). Expression of human snRNA genes from beginning to end. Biochem. Soc. Trans. 36, pp. 590–594.

Egloff, S., Szczepaniak, S.A., Dienstbier, M., Taylor, A., Knight, S., and Murphy, S. (2010). The integrator complex recognizes a new double mark on the RNA polymerase II carboxyl-terminal domain. J. Biol. Chem. 285, pp. 20564–20569.

Egloff, S., Zaborowska, J., Laitem, C., Kiss, T., and Murphy, S. (2012). Ser7 phosphorylation of the CTD recruits the RPAP2 ser5 phosphatase to snRNA genes.

Mol. Cell. 45, pp. 111–122.

Egloff, S., Vitali, P., Tellier, M., Raffel, R., Murphy, S., and Kiss, T. (2017). The 7SK snRNP associates with the little elongation complex to promote snRNA gene expression. EMBO J. 36, pp. 934–948.

Eick, D., and Geyer, M. (2013). The RNA Polymerase II Carboxy-Terminal Domain (CTD) Code. Chem. Rev. 113, pp. 8456–8490.

Fabrega, C., Shen, V., Shuman, S., and Lima, C.D. (2003). Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. Mol. Cell. 11, pp. 1549-1561.

Ferraiuolo, L., Kirby, J., Grierson, A.J., Sendtner, M., and Shaw, P.J. (2011). Molecular pathways of motor neuron injury in amyotrophic lateral sclerosis. Nat. Rev. Neurol. 7, pp. 616–630.

Fialcowitz, E.J., Brewer, B.Y., Keenan, B.P., and Wilson, G.M. (2005). A hairpin-like structure within an AU-rich mRNA-destabilizing element regulates trans-factor binding selectivity and mRNA decay kinetics. J. Biol. Chem. 280, pp. 22406–22417

Fong, N., Saldi, T., Sheridan, R.M., Cortazar, M.A., and Bentley, D.L. (2017). RNA Pol II Dynamics Modulate Co-transcriptional Chromatin Modification, CTD Phosphorylation, and Transcriptional Direction. Mol. Cell. 66, pp. 546-557.

Fornerod, M., Ohno, M., Yoshida, M., and Mattaj, I.W. (1997). CRM1 is an export receptor for leucine-rich nuclear export signals. Cell. 90, pp. 1051–1060.

Frey, M.R., and Matera, A.G. (2001). RNA-mediated interaction of Cajal bodies and U2 snRNA genes. J. Cell Biol. 154, pp. 499–509.

Frey, M.R., Bailey, A.D., Weiner, A.M., and Matera, A.G. (1999). Association of snRNA genes with coiled bodies is mediated by nascent snRNA transcripts. Curr. Biol. 9, pp. 126–135.

Fujioka, Y., Ishigaki, S., Masuda, A., Iguchi, Y., Udagawa, T., Watanabe, H., Katsuno,

M., Ohno, K., and Sobue, G. (2013). FUS-regulated region- and cell-type-specific transcriptome is associated with cell selectivity in ALS/FTLD. Sci. Rep. 3, 2388.

Gabanella, F., Butchbach, M.E.R., Saieva, L., Carissimi, C., Burghes, A.H.M., and Pellizzoni, L. (2007). Ribonucleoprotein assembly defects correlate with spinal muscular atrophy severity and preferentially affect a subset of spliceosomal snRNPs. PLoS One. 2, p. e921.

Gabler, S., Schütt, H., Groitl, P., Wolf, H., Shenk, T., and Dobner, T. (1998). E1B 55-kilodalton-associated protein: a cellular protein with RNA-binding activity implicated in nucleocytoplasmic transport of adenovirus and cellular mRNAs. J. Virol. 72, pp. 7960-7971.

Garcia-Santibanez, R., Burford, M., and Bucelli, R.C. (2018). Hereditary Motor Neuropathies and Amyotrophic Lateral Sclerosis: a Molecular and Clinical Update. Curr. Neurol. Neurosci. Rep. 18: 93.

Giacometti, S., Benbahouche, N.E.H., Domanski, M., Robert, M.C., Meola, N., Lubas, M., Bukenborg, J., Andersen, J.S., Schulze, W.M., Verheggen, C., et al. (2017). Mutually Exclusive CBC-Containing Complexes Contribute to RNA Fate. Cell Rep. 18, pp. 2635–2650.

Glover-Cutter, K., Larochelle, S., Erickson, B., Zhang, C., Shokat, K., Fisher, R.P., and Bentley, D.L. (2009). TFIIH-Associated Cdk7 Kinase Functions in Phosphorylation of C-Terminal Domain Ser7 Residues, Promoter-Proximal Pausing, and Termination by RNA Polymerase II. Mol. Cell. Biol. 29, pp. 5455–5464.

Gorthi, A., Romero, J.C., Loranc, E., Cao, L., Lawrence, L.A., Goodale, E., Iniguez, A.B., Bernard, X., Masamsetti, V.P., Roston, S., et al. (2018). EWS-FLI1 increases transcription to cause R-Loops and block BRCA1 repair in Ewing sarcoma. Nature. 555, pp. 387–391.

Greenberg, M.E., and Ziff, E.B. (1984). Stimulation of 3T3 cells induces transcription of the c-fos proto-oncogene. Nature. 311, pp. 433-438.

Greifenberg, A.K., Hönig, D., Pilarova, K., Düster, R., Bartholomeeusen, K., Bösken, C.A., Anand, K., Blazek, D., and Geyer, M. (2016). Structural and Functional Analysis of the Cdk13/Cyclin K Complex. Cell Rep. 14, pp. 320–331.

Gridasova, A.A., and Henry, R.W. (2005). The p53 Tumor Suppressor Protein Represses Human snRNA Gene Transcription by RNA Polymerases II and III Independently of Sequence-Specific DNA Binding. Mol. Cell. Biol. 25, pp. 3247–3260.

Groen, E.J.N., Fumoto, K., Blokhuis, A.M., Engelen-Lee, J.Y., Zhou, Y., van den Heuvel, D.M.A., Koppers, M., van Diggelen, F., van Heest, J., Demmers, J.A.A., et al. (2013). ALS-associated mutations in FUS disrupt the axonal distribution and function of SMN. Hum. Mol. Genet. 22, pp. 3690–3704.

Gu, B., Eick, D., and Bensaude, O. (2013). CTD serine-2 plays a critical role in splicing and termination factor recruitment to RNA polymerase II in vivo. Nucleic Acids Res. 41, pp. 1591-1603.

Guil, S., and Cáceres, J.F. (2007). The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. Nat. Struct. Mol. Biol. 14, p. 591.

Gurunathan, G., Yu, Z., Coulombe, Y., Masson, J.Y., and Richard, S. (2015). Arginine methylation of hnRNPUL1 regulates interaction with NBS1 and recruitment to sites of DNA damage. Sci. Rep. 5, 10475.

Haeusler, A.R., Donnelly, C.J., Periz, G., Simko, E.J., Shaw, P.G., Kim, M., Maragakis, N.J., Troncoso, J.C., Pandey, A., Sattler, R., et al. (2014). C9orf72 nucleotide repeat structures initiate molecular cascades of disease. Nature *507*, 195–200.

Hallais, M., Pontvianne, F., Andersen, P.R., Clerici, M., Lener, D., Benbahouche, N.E.H., Gostan, T., Vandermoere, F., Robert, M.C., Cusack, S., et al. (2013). CBC-ARS2 stimulates 3′-end maturation of multiple RNA families and favors cap-proximal processing. Nat. Struct. Mol. Biol. 20, pp. 1358-1366.

Harlen, K.M., Trotta, K.L., Smith, E.E., Mosaheb, M.M., Fuchs, S.M., and Churchman, L.S. (2016). Comprehensive RNA Polymerase II Interactomes Reveal Distinct and

Varied Roles for Each Phospho-CTD Residue. Cell Rep. 15, pp. 2147–2158.

Harrison, A.F., and Shorter, J. (2017). RNA-binding proteins with prion-like domains in health and disease. Biochem. J. 474, pp. 1417-1438.

Hasegawa, Y., Brockdorff, N., Kawano, S., Tsutui, K., Tsutui, K., and Nakagawa, S. (2010). The matrix protein hnRNP U is required for chromosomal localization of xist RNA. Dev. Cell. 19, pp. 469–476.

Hennig, S., Kong, G., Mannen, T., Sadowska, A., Kobelke, S., Blythe, A., Knott, G.J., Iyer, K.S., Ho, D., Newcombe, E.A., et al. (2015). Prion-like domains in RNA binding proteins are essential for building subnuclear paraspeckles. J. Cell Biol. *210*, 529–539.

Herter, E.K., Stauch, M., Gallant, M., Wolf, E., Raabe, T., and Gallant, P. (2015). snoRNAs are a novel class of biologically relevant Myc targets. BMC Biol. 13 1 17.

Hirose, T., Virnicchi, G., Tanigawa, A., Naganuma, T., Li, R., Kimura, H., Yokoi, T., Nakagawa, S., Bénard, M., Fox, A.H., et al. (2014). NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies. Mol. Biol. Cell. 25, pp. 169–183.

Hong, Z., Jiang, J., Ma, J., Dai, S., Xu, T., Li, H., and Yasui, A. (2013). The Role of hnRPUL1 Involved in DNA Damage Response Is Related to PARP1. PLoS One. 8, e60208.

Hrossova, D., Sikorsky, T., Potesil, D., Bartosovic, M., Pasulka, J., Zdrahal, Z., Stefl, R., and Vanacova, S. (2015). RBM7 subunit of the NEXT complex binds U-rich sequences and targets 3′-end extended forms of snRNAs. Nucleic Acids Res. 43, pp. 4236–4248.

Hu, D., Smith, E.R., Garruss, A.S., Mohaghegh, N., Varberg, J.M., Lin, C., Jackson, J., Gao, X., Saraf, A., Florens, L., et al. (2013). The little elongation complex functions at initiation and elongation phases of snRNA gene transcription. Mol. Cell. 51, pp. 493–505.

Huang, Q., Jacobson, M.R., and Pederson, T. (2015). 3' processing of human pre-U2

small nuclear RNA: a base-pairing interaction between the 3' extension of the precursor and an internal region. Mol. Cell. Biol. 17, pp. 7178-7185.

Ideue, T., Adachi, S., Naganuma, T., Tanigawa, A., Natsume, T., and Hirose, T. (2012). U7 small nuclear ribonucleoprotein represses histone gene transcription in cell cycle-arrested cells. Proc. Natl. Acad. Sci. 109, pp. 5693-5698.

Imamura, K., Imamachi, N., Akizuki, G., Kumakura, M., Kawaguchi, A., Nagata, K., Kato, A., Kawaguchi, Y., Sato, H., Yoneda, M., et al. (2014). Long Noncoding RNA NEAT1-Dependent SFPQ Relocation from Promoter Region to Paraspeckle Mediates IL8 Expression upon Immune Stimuli. Mol. Cell *53*, 393–406.

Ishihara, T., Ariizumi, Y., Shiga, A., Kato, T., Tan, C.F., Sato, T., Miki, Y., Yokoo, M., Fujino, T., Koyama, A., et al. (2013). Decreased number of gemini of coiled bodies and U12 snRNA level in amyotrophic lateral sclerosis. Hum. Mol. Genet. 22, pp. 4136–4147.

Jády, B.E., Bertrand, E., and Kiss, T. (2004). Human telomerase RNA and box H/ACA scaRNAs share a common Cajal body-specific localization signal. J. Cell Biol. 164, pp. 647–652.

Jawdekar, G.W., and Henry, R.W. (2008). Transcriptional regulation of human small nuclear RNA genes. Biochim. Biophys. Acta - Gene Regul. Mech. 1779, pp. 295–305

Jeronimo, C., Forget, D., Bouchard, A., Li, Q., Chua, G., Poitras, C., Thérien, C., Bergeron, D., Bourassa, S., Greenblatt, J., et al. (2007). Systematic Analysis of the Protein Interaction Network for the Human Transcription Machinery Reveals the Identity of the 7SK Capping Enzyme. Mol. Cell *27*, 262–274.

Jimeno-González, S., Ceballos-Chávez, M., and Reyes, J.C. (2015). A positioned +1 nucleosome enhances promoter-proximal pausing. Nucleic Acids Res. 43, pp. 3068-3078.

Johnson, J.O., Pioro, E.P., Boehringer, A., Chia, R., Feit, H., Renton, A.E., Pliner, H.A., Abramzon, Y., Marangi, G., Winborn, B.J., et al. (2014). Mutations in the Matrin 3 gene

cause familial amyotrophic lateral sclerosis. Nat. Neurosci. 17, pp. 664–666.

Jones, K.W., Gorzynski, K., Hales, C.M., Fischer, U., Badbanchi, F., Terns, R.M., and Terns, M.P. (2001). Direct Interaction of the Spinal Muscular Atrophy Disease Protein SMN with the Small Nucleolar RNA-associated Protein Fibrillarin. J. Biol. Chem. 276, pp. 38645–38651.

Jorjani, H., Kehr, S., Jedlinski, D.J., Gumienny, R., Hertel, J., Stadler, P.F., Zavolan, M., and Gruber, A.R. (2016). An updated human snoRNAome. Nucleic Acids Res. 44, pp. 5068-5082.

Jun, M.H., Ryu, H.H., Jun, Y.W., Liu, T., Li, Y., Lim, C.S., Lee, Y.S., Kaang, B.K., Jang, D.J., and Lee, J.A. (2017). Sequestration of PRMT1 & Nd1-L mRNA into ALS-linked FUS mutant R521C-positive aggregates contributes to neurite degeneration upon oxidative stress. Sci. Rep. 7, p. 40474.

Karijolich, J., and Yu, Y.T. (2010). Spliceosomal snRNA modifications and their function. RNA Biol. 7, pp. 192-204.

Kariya, S., Re, D.B., Jacquier, A., Nelson, K., Przedborski, S., and Monani, U.R. (2012). Mutant superoxide dismutase 1 (SOD1), a cause of amyotrophic lateral sclerosis, disrupts the recruitment of SMN, the spinal muscular atrophy protein to nuclear cajal bodies. Hum. Mol. Genet. *21*, 3421–3434.

Kass, S., Tyc, K., Steitz, J.A., and Sollner-Webb, B. (1990). The U3 small nucleolar ribonucleoprotein functions in the first step of preribosomal RNA processing. Cell. 60, pp. 897-908.

Kim, M.K., and Nikodem, V.M. (1999). hnRNP U Inhibits Carboxy-Terminal Domain Phosphorylation by TFIIH and Represses RNA Polymerase II Elongation. Mol. Cell. Biol. 19, pp. 6833–6844.

Kim, H.J., Kim, N.C., Wang, Y.-D., Scarborough, E.A., Moore, J., Diaz, Z., MacLea, K.S., Freibaum, B., Li, S., Molliex, A., et al. (2013). Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. Nature *495*,

467–473.

Kitao, S., Segref, A., Kast, J., Wilm, M., Mattaj, I.W., and Ohno, M. (2008). A Compartmentalized Phosphorylation/Dephosphorylation System That Regulates U snRNA Export from the Nucleus. Mol. Cell. Biol. 28, pp. 487–497.

Koga, M., Hayashi, M., and Kaida, D. (2015). Splicing inhibition decreases phosphorylation level of Ser2 in Pol II CTD. Nucleic Acids Res. 43, 8258–8267.

Kowarz, E., Löscher, D., and Marschalek, R. (2015). Optimized Sleeping Beauty transposons rapidly generate stable transgenic cell lines. Biotechnol. J. 10, 647–653.

Krämer, A., Grüter, P., Gröning, K., and Kastner, B. (1999). Combined biochemical and electron microscopic analyses reveal the architecture of the mammalian U2 snRNP. J. Cell Biol. 145, 1355–1368.

Krishnamurthy, S., He, X., Reyes-Reyes, M., Moore, C., and Hampsey, M. (2004). Ssu72 is an RNA polymerase II CTD phosphatase. Mol. Cell. 14, pp. 387-39.

Kufel, J., and Grzechnik, P. (2019). Small Nucleolar RNAs Tell a Different Tale. Trends Genet. 35, pp. 104–117.

Kwak, H., and Lis, J.T. (2013). Control of Transcriptional Elongation. Annu. Rev. Genet. 47, 483–508.

Kzhyshkowska, J., Rusch, A., Wolf, H., and Dobner, T. (2003). Regulation of transcription by the heterogeneous nuclear ribonucleoprotein E1B-AP5 is mediated by complex formation with the novel bromodomain-containing protein BRD7. Biochem. J. *371*, 385–393.

Lafarga, M., Tapia, O., Romero, A.M., and Berciano, M.T. (2017). Cajal bodies in neurons. RNA Biol. 14, 712–725.

Lafontaine, D.L.J., Bousquet-Antonelli, C., Henry, Y., Caizergues-Ferrer, M., and Tollervey, D. (1998). The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. Genes Dev. 12, pp. 527–537.

Lagier-Tourenne, C., Polymenidou, M., and Cleveland, D.W. (2010). TDP-43 and FUS/TLS: Emerging roles in RNA processing and neurodegeneration. Hum. Mol. Genet. 19, R46–R64.

Lagier-Tourenne, C., Polymenidou, M., Hutt, K.R., Vu, A.Q., Baughn, M., Huelga, S.C., Clutario, K.M., Ling, S.-C., Liang, T.Y., Mazur, C., et al. (2012). Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. Nat. Neurosci. *15*, 1488–1497.

Lancaster, A.K., Nutter-Upham, A., Lindquist, S., and King, O.D. (2014). PLAAC: A web and command-line application to identify proteins with prion-like amino acid composition. Bioinformatics. 30, 2501–2502.

Lee, Y.B., Chen, H.J., Peres, J.N., Gomez-Deza, J., Attig, J., Štalekar, M., Troakes, C., Nishimura, A.L., Scotter, E.L., Vance, C., et al. (2013). Hexanucleotide repeats in ALS/FTD form length-dependent RNA Foci, sequester RNA binding proteins, and are neurotoxic. Cell Rep. 5, pp. 1178–1186.

Lefebvre, S., Bürglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M., et al. (1995). Identification and characterization of a spinal muscular atrophy-determining gene. Cell. 80, 155–165.

Lemm, I., Girard, C., Kuhn, A.N., Watkins, N.J., Schneider, M., Bordonné, R., and Lührmann, R. (2006). Ongoing U snRNP Biogenesis Is Required for the Integrity of Cajal Bodies. Mol. Biol. Cell. 17, pp. 3221-3231.

Li, J., and Gilmour, D.S. (2011). Promoter proximal pausing and the control of gene expression. Curr. Opin. Genet. Dev. 21, pp. 231-235.

Li, Y.R., King, O.D., Shorter, J., and Gitler, A.D. (2013). Stress granules as crucibles of ALS pathogenesis. J. Cell Biol. 201, 361–372.

Licht, K., Medenbach, J., Lührmann, R., Kambach, C., and Bindereif, A. (2008). 3'-cyclic phosphorylation of U6 snRNA leads to recruitment of recycling factor p110 through LSm proteins. RNA. 14, pp. 1532–1538.

Lin, Y., Protter, D.S.W., Rosen, M.K., and Parker, R. (2015). Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. Mol. Cell. 60, 208–219.

Liu, S., Rauhut, R., Vornlocher, H.P., and Lührmann, R. (2006). The network of protein-protein interactions within the human U4/U6.U5 tri-snRNP. RNA. 12, pp. 1418-1430.

Liu, S., Li, P., Dybkov, O., Nottrott, S., Hartmuth, K., Lührmann, R., Carlomagno, T., and Wahl, M.C. (2007). Binding of the human Prp31 Nop domain to a composite RNA-protein platform in U4 snRNP. Science 316, pp. 115-120.

Liu, Z., Chen, O., Wall, J.B.J., Zheng, M., Zhou, Y., Wang, L., Ruth Vaseghi, H., Qian, L., and Liu, J. (2017). Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. Sci. Rep. 7, 2193.

Lorson, C.L., Strasswimmer, J., Yao, J.M., Baleja, J.D., Hahnen, E., Wirth, B., Le, T., Burghes, A.H.M., and Androphy, E.J. (1998). SMN oligomerization defect correlates with spinal muscular atrophy severity. Nat. Genet. 19, pp. 63-66

Lorson, C.L., Hahnen, E., Androphy, E.J., and Wirth, B. (1999). A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. Proc. Natl. Acad. Sci. U. S. A. 96, pp. 6307–6311.

Lubas, M., Andersen, P.R., Schein, A., Dziembowski, A., Kudla, G., and Jensen, T.H. (2015). The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. Cell Rep. 10, 178–192.

Lund, E., and Dahlberg, J.E. (1984). True genes for U1 small nuclear RNA. Copy number, polymorphism, and methylation. J. Biol. Chem. 259, pp. 2013–2021.

Lykke-Andersen, S., Chen, Y., Ardal, B.R., Lilje, B., Waage, J., Sandelin, A., and Jensen, T.H. (2014). Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. Genes Dev. 28, pp. 2498–2517.

Lykke-Andersen, S., Ardal, B.K., Hollensen, A.K., Damgaard, C.K., and Jensen, T.H. (2018). Box C/D snoRNP Autoregulation by a cis-Acting snoRNA in the NOP56 Pre-

mRNA. Mol. Cell. 72, 99–111.

Machyna, M., Heyn, P., and Neugebauer, K.M. (2013). Cajal bodies: Where form meets function. Wiley Interdiscip. Rev. RNA. 4, pp. 17-34.

Machyna, M., Kehr, S., Straube, K., Kappei, D., Buchholz, F., Butter, F., Ule, J., Hertel, J., Stadler, P.F., and Neugebauer, K.M. (2014). The coilin interactome identifies hundreds of small noncoding RNAs that traffic through cajal bodies. Mol. Cell. 56, 389–399.

Macias, S., Cordiner, R.A., Gautier, P., Plass, M., and Cáceres, J.F. (2015). DGCR8 Acts as an Adaptor for the Exosome Complex to Degrade Double-Stranded Structured RNAs. Mol. Cell. 60, pp. 873-885.

Mackenzie, I.R.A., Rademakers, R., and Neumann, M. (2010). TDP-43 and FUS in amyotrophic lateral sclerosis and frontotemporal dementia. Lancet Neurol. *9*, 995–1007.

Mahajan, M.C., Narlikar, G.J., Boyapaty, G., Kingston, R.E., and Weissman, S.M. (2005). Heterogeneous nuclear ribonucleoprotein C1/C2, MeCP1, and SWI/SNF form a chromatin remodeling complex at the -globin locus control region. Proc. Natl. Acad. Sci. 102, pp. 15012-15017.

Mahmoudi, S., Henriksson, S., Weibrecht, I., Smith, S., Söderberg, O., Strömblad, S., Wiman, K.G., and Farnebo, M. (2010). wrap53 is essential for cajal body formation and for targeting the survival of motor neuron complex to cajal bodies. PLoS Biol. 8, e1000521.

Malewicz, M., Kadkhodaei, B., Kee, N., Volakakis, N., Hellman, U., Viktorsson, K., Leung, C.Y., Chen, B., Lewensohn, R., van Gent, D.C., et al. (2011). Essential role for DNA-PK-mediated phosphorylation of NR4A nuclear orphan receptors in DNA double-strand break repair. Genes Dev. 25, 2031–2040.

Maruyama, H., Morino, H., Ito, H., Izumi, Y., Kato, H., Watanabe, Y., Kinoshita, Y., Kamada, M., Nodera, H., Suzuki, H., et al. (2010). Mutations of optineurin in

amyotrophic lateral sclerosis. Nature. 465, pp. 223–226.

Matera, A.G., and Shpargel, K.B. (2006). Pumping RNA: nuclear bodybuilding along the RNP pipeline. Curr. Opin. Cell Biol. 18, 317–324.

Matera, A.G., and Wang, Z. (2014). A day in the life of the spliceosome. Nat. Rev. Mol. Cell Biol. 15, 108–121.

Matte, A., and Delbaere, L.T. (2006). ATP-binding Motifs. In Encyclopedia of Life Sciences, (ELS). Chichester: John Wiley & Sons, Ltd.

McDonald, K.K., Aulas, A., Destroismaisons, L., Pickles, S., Beleac, E., Camu, W., Rouleau, G.A., and Velde, C. Vande (2011). TAR DNA-binding protein 43 (TDP-43) regulates stress granule dynamics via differential regulation of G3BP and TIA-1. Hum. Mol. Genet. 20, pp. 1400–1410.

McKeegan, K.S., Debieux, C.M., Boulon, S., Bertrand, E., and Watkins, N.J. (2007). A Dynamic Scaffold of Pre-snoRNP Factors Facilitates Human Box C/D snoRNP Assembly. Mol. Cell. Biol. 27, pp. 6782-6793.

McNamara, R.P., McCann, J.L., Gudipaty, S.A., and D'Orso, I. (2013). Transcription factors mediate the enzymatic disassembly of promoter-bound 7sk snrnp to locally recruit p-tefb for transcription elongation. Cell Rep. 5, 1256–1268.

McNamara, R.P., Bacon, C.W., and D'Orso, I. (2016). Transcription elongation control by the 7SK snRNP complex: Releasing the pause. Cell Cycle. 15, 2115–2123

Meier, U.T. (2017). RNA modification in Cajal bodies. RNA Biol. 0, pp. 1-8

Meister, G., Eggert, C., Bühler, D., Brahms, H., Kambach, C., and Fischer, U. (2001). Methylation of Sm proteins by a complex containing PRMT5 and the putative U snRNP assembly factor pICln. Curr. Biol. 11, 1990–1994.

Melton, A.A., Jackson, J., Wang, J., and Lynch, K.W. (2007). Combinatorial Control of Signal-Induced Exon Repression by hnRNP L and PSF. Mol. Cell. Biol. 27, pp. 6972–6984

Misteli, T. (2001). The concept of self-organization in cellular architecture. J. Cell Biol. 155, pp. 181-185.

Mouaikel, J., Narayanan, U., Verheggen, C., Matera, A.G., Bertrand, E., Tazi, J., and Bordonne, R. (2003). Interaction between the small-nuclear-RNA cap hypermethylase and the spinal muscular atrophy protein, survival of motor neuron. EMBO Rep. 4, 616–622

Mroczek, S., and Dziembowski, A. (2013). U6 RNA biogenesis and disease association. Wiley Interdiscip. Rev. RNA. 4, 581–592.

Murphy, S., Yoon, J.B., Gerster, T., and Roeder, R.G. (2015). Oct-1 and Oct-2 potentiate functional interactions of a transcription factor with the proximal sequence element of small nuclear RNA genes. Mol. Cell. Biol. 12, 3247–3261.

Naganuma, T., Nakagawa, S., Tanigawa, A., Sasaki, Y.F., Goshima, N., and Hirose, T. (2012). Alternative 3'-end processing of long noncoding RNA initiates construction of nuclear paraspeckles. TL - 31. EMBO J. *31 VN-r*, 4020–4034.

Natsume, T., Kiyomitsu, T., Saga, Y., and Kanemaki, M.T. (2016). Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors. Cell Rep. 15, pp. 210-218.

Nesic, D. (2004). A role for Cajal bodies in the final steps of U2 snRNP biogenesis. J. Cell Sci. 117, pp. 4423–4433.

Neumann, M., Sampathu, D.M., Kwong, L.K., Truax, A.C., Micsenyi, M.C., Chou, T.T., Bruce, J., Schuck, T., Grossman, M., Clark, C.M., et al. (2006). Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science 314, pp. 130–133.

Nishimoto, Y., Nakagawa, S., Hirose, T., Okano, H.J., Takao, M., Shibata, S., Suyama, S., Kuwako, K.I., Imai, T., Murayama, S., et al. (2013). The long non-coding RNA nuclear-enriched abundant transcript 1-2 induces paraspeckle formation in the motor neuron during the early phase of amyotrophic lateral sclerosis. Mol. Brain. 6, 31.

Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T., and Kanemaki, M. (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. Nat. Methods. 6, pp. 917-922.

Nojima, T., Gomes, T., Grosso, A.R.F., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, N.J. (2015). Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. Cell. 161, 526–540.

Nojima, T., Rebelo, K., Gomes, T., Grosso, A.R., Proudfoot, N.J., and Carmo-Fonseca, M. (2018). RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-transcriptional Splicing. Mol. Cell. 72, pp. 369-379.

Novotný, I., Blažíková, M., Staněk, D., Herman, P., and Malinsky, J. (2011). In vivo kinetics of U4/U6·U5 tri-snRNP formation in Cajal bodies. Mol. Biol. Cell. 22, 513–523.

Novotný, I., Malinová, A., Stejskalová, E., Matějů, D., Klimešová, K., Roithová, A., Švéda, M., Knejzlík, Z., and Staněk, D. (2015). SART3-Dependent Accumulation of Incomplete Spliceosomal snRNPs in Cajal Bodies. Cell Rep. 10, pp. 429–440.

Nozawa, R.S., Boteva, L., Soares, D.C., Naughton, C., Dun, A.R., Buckle, A., Ramsahoye, B., Bruton, P.C., Saleeb, R.S., Arnedo, M., et al. (2017). SAF-A Regulates Interphase Chromosome Structure through Oligomerization with Chromatin-Associated RNAs. Cell. 169, 1214–1227.

O'Reilly, D., Kuznetsova, O. V., Laitem, C., Zaborowska, J., Dienstbier, M., and Murphy, S. (2014). Human snRNA genes use polyadenylation factors to promote efficient transcription termination. Nucleic Acids Res. 42, pp. 264–275.

Onodera, O., Ishihara, T., Shiga, A., Ariizumi, Y., Yokoseki, A., and Nishizawa, M. (2014). Minor splicing pathway is not minor any more: Implications for the pathogenesis of motor neuron diseases. Neuropathology. 34, 99–107.

Palacios, I. (1997). Nuclear import of U snRNPs requires importin beta. EMBO J. 16, 6783–6792.

Panhale, A., Richter, F.M., Ramírez, F., Shvedunova, M., Manke, T., Mittler, G., and

Akhtar, A. (2019). CAPRI enables comparison of evolutionarily conserved RNA interacting regions. Nat. Commun. 10, 2682.

Pellizzoni, L., Baccon, J., Charroux, B., and Dreyfuss, G. (2001). The survival of motor neurons (SMN) protein interacts with the snoRNP proteins fibrillarin and GAR1. Curr. Biol. 11, 1079–1088.

Pellizzoni, L., Yong, J., and Dreyfuss, G. (2002). Essential role for the SMN complex in the specificity of snRNP assembly. Science 298, 1775–1779.

Pena, E., Berciano, M.T., Fernandez, R., Ojeda, J.L., and Lafarga, M. (2001). Neuronal body size correlates with the number of nucleoli and Cajal bodies, and with the organization of the splicing machinery in rat trigeminal ganglion neurons. J. Comp. Neurol. 430, 250–263.

Peterlin, B.M., and Price, D.H. (2006). Controlling the Elongation Phase of Transcription with P-TEFb. Mol. Cell. 23, pp. 297-305.

Piñol-Roma, S., and Dreyfuss, G. (1992). Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. Nature. 355, pp. 730-732.

Piñol-Roma, S., Choi, Y.D., Matunis, M.J., and Dreyfuss, G. (1988). Immunopurification of heterogeneous nuclear ribonucleoprotein particles reveals an assortment of RNA-binding proteins. Genes Dev. 2, 215–227.

Polo, S.E., Blackford, A.N., Chapman, J.R., Baskcomb, L., Gravel, S., Rusch, A., Thomas, A., Blundred, R., Smith, P., Kzhyshkowska, J., et al. (2012). Regulation of DNA-End Resection by hnRNPU-like Proteins Promotes DNA Double-Strand Break Signaling and Repair. Mol. Cell *45*, 505–516.

Pradet-Balade, B., Girard, C., Boulon, S., Paul, C., Azzag, K., Bordonné, R., Bertrand, E., and Verheggen, C. (2011). CRM1 controls the composition of nucleoplasmic pre-snoRNA complexes to licence them for nucleolar transport. EMBO J. 30, 2205–2218.

Puls, I., Jonnakuty, C., LaMonte, B.H., Holzbaur, E.L.F., Tokito, M., Mann, E., Floeter, M.K., Bidus, K., Drayna, D., Oh, S.J., et al. (2003). Mutant dynactin in motor neuron

disease. Nat. Genet. 33, 455–456.

Quiñones, A., Dobberstein, K.U., and Rainov, N.G. (2003). The egr-1 gene is induced by DNA-damaging agents and non-genotoxic drugs in both normal and neoplastic human cells. Life Sci. 72, pp. 2975-2992.

Raczynska, K.D., Ruepp, M.D., Brzek, A., Reber, S., Romeo, V., Rindlisbacher, B., Heller, M., Szweykowska-Kulinska, Z., Jarmolowski, A., and Schümperli, D. (2015). FUS/TLS contributes to replication-dependent histone gene expression by interaction with U7 snRNPs and histone-specific transcription factors. Nucleic Acids Res. *43*, 9711–9728.

Raker, V.A., Plessel, G., and Lührmann, R. (1996). The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro. EMBO J. 15, pp. 2256-2269.

Ramos, J.A., Zenser, N., Leyser, O., and Callis, J. (2007). Rapid Degradation of Auxin/Indoleacetic Acid Proteins Requires Conserved Amino Acids of Domain II and Is Proteasome Dependent. Plant Cell. 13, 2349–2360.

Raška, I., Ochs, R.L., Andrade, L.E.C., Chan, E.K.L., Burlingame, R., Peebles, C., Gruol, D., and Tan, E.M. (1990). Association between the nucleolus and the coiled body. J. Struct. Biol. 104, pp. 120-127.

Renton, A.E., Chiò, A., and Traynor, B.J. (2014). State of play in amyotrophic lateral sclerosis genetics. Nat. Neurosci. 17, 17–23.

Reyes-Reyes, M., and Hampsey, M. (2007). Role for the Ssu72 C-Terminal Domain Phosphatase in RNA Polymerase II Transcription Elongation. Mol. Cell. Biol. 27, 926–936.

Van Riggelen, J., Yetil, A., and Felsher, D.W. (2010). MYC as a regulator of ribosome biogenesis and protein synthesis. Nat. Rev. Cancer. 10, pp. 301-309.

Romac, J.M., Graff, D.H., and Keene, J.D. (2015). The U1 small nuclear ribonucleoprotein (snRNP) 70K protein is transported independently of U1 snRNP

particles via a nuclear localization signal in the RNA-binding domain. Mol. Cell. Biol. 14, pp. 4662-4670.

Rulten, S.L., Rotheray, A., Green, R.L., Grundy, G.J., Moore, D.A.Q., Gomez-Herreros, F., Hafezparast, M., and Caldecott, K.W. (2014). PARP-1 dependent recruitment of the amyotrophic lateral sclerosis-associated protein FUS/TLS to sites of oxidative DNA damage. Nucleic Acids Res. 42, pp. 307-314.

Saldi, T., Cortazar, M.A., Sheridan, R.M., and Bentley, D.L. (2016). Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing. J. Mol. Biol. 428, pp. 2623-2635.

Sauterer, R.A., Feeney, R.J., and Zieve, G.W. (1988). Cytoplasmic assembly of snRNP particles from stored proteins and newly transcribed snRNA's in L929 mouse fibroblasts. Exp. Cell Res. 176, pp. 344-359.

Sawyer, I.A., Sturgill, D., Sung, M.H., Hager, G.L., and Dundr, M. (2016). Cajal body function in genome organization and transcriptome diversity. BioEssays. 38, 1197–1208.

Schul, W., van Driel, R., and de Jong, L. (1998). Coiled Bodies and U2 snRNA Genes Adjacent to Coiled Bodies Are Enriched in Factors Required for snRNA Transcription. Mol. Biol. Cell. 9, 1025-1036.

Schüller, R., Forné, I., Straub, T., Schreieck, A., Texier, Y., Shah, N., Decker, T.M., Cramer, P., Imhof, A., and Eick, D. (2016). Heptad-Specific Phosphorylation of RNA Polymerase II CTD. Mol. Cell. 61, 305–314.

Schulz, S., Chachami, G., Kozaczkiewicz, L., Winter, U., Stankovic-Valentin, N., Haas, P., Hofmann, K., Urlaub, H., Ovaa, H., Wittbrodt, J., et al. (2012). Ubiquitin-specific protease-like 1 (USPL1) is a SUMO isopeptidase with essential, non-catalytic functions. EMBO Rep. 13, 930–938.

Schwartz, J.C., Ebmeier, C.C., Podell, E.R., Heimiller, J., Taatjes, D.J., and Cech, T.R. (2012). FUS binds the CTD of RNA polymerase II and regulates its phosphorylation at

Ser2. Genes Dev. 26, pp. 2690-2695.

Schwartz, J.C., Podell, E.R., Han, S.S.W., Berry, J.D., Eggan, K.C., and Cech, T.R. (2014). FUS is sequestered in nuclear aggregates in ALS patient fibroblasts. Mol. Biol. Cell. 25, pp. 2571-2578.

Shang, Y., and Huang, E.J. (2016). Mechanisms of FUS mutations in familial amyotrophic lateral sclerosis. Brain Res. 1647, 65–78.

Sharma, V., Khurana, S., Kubben, N., Abdelmohsen, K., Oberdoerffer, P., Gorospe, M., and Misteli, T. (2015). A BRCA1-interacting lncRNA regulates homologous recombination. EMBO Rep. 16, 1520–1534.

Shelkovnikova, T.A., Robinson, H.K., Troakes, C., Ninkina, N., and Buchman, V.L. (2014). Compromised paraspeckle formation as a pathogenic factor in FUSopathies. Hum. Mol. Genet. *23*, 2298–2312.

Skourti-Stathaki, K., and Proudfoot, N.J. (2014). A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes Dev. 28, 1384–1396.

Sleeman, J.E., and Lamond, A.I. (1999). Newly assembled snRNPs associate with coiled bodies before speckles, suggesting a nuclear snRNP maturation pathway. Curr. Biol. 9, 1065–1074.

Smith, E., Lin, C., and Shilatifard, A. (2011a). The super elongation complex (SEC) and MLL in development and disease. Genes Dev. 25, pp. 661-672.

Smith, E.R., Lin, C., Garrett, A.S., Thornton, J., Mohaghegh, N., Hu, D., Jackson, J., Saraf, A., Swanson, S.K., Seidel, C., et al. (2011b). The Little Elongation Complex Regulates Small Nuclear RNA Transcription. Mol. Cell. 44, pp. 954-965.

Smith, J.R., Maguire, S., Davis, L.A., Alexander, M., Yang, F., Chandran, S., ffrench-Constant, C., and Pedersen, R.A. (2008). Robust, Persistent Transgene Expression in Human Embryonic Stem Cells Is Achieved with AAVS1-Targeted Integration. Stem Cells. 26, 496–504.

Søgaard, T.M.M., and Svejstrup, J.Q. (2007). Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. J. Biol. Chem. 282, pp. 14113-14120.

Soutourina, J. (2018). Transcription regulation by the Mediator complex. Nat. Rev. Mol. Cell Biol. 19, 262–74.

Staněk, D. (2017). Cajal bodies and snRNPs - friends with benefits. RNA Biol. 14, pp. 1-9.

Staněk, D., and Neugebauer, K.M. (2004). Detection of snRNP assembly intermediates in Cajal bodies by fluorescence resonance energy transfer. J. Cell Biol. 166, 1015–1025.

Staněk, D., Rader, S.D., Klingauf, M., and Neugebauer, K.M. (2003). Targeting of U4/U6 small nuclear RNP assembly factor SART3/p110 to Cajal bodies. J. Cell Biol. 160, 505–516.

Suzuki, T., Izumi, H., and Ohno, M. (2010). Cajal body surveillance of U snRNA export complex assembly. J. Cell Biol. 190, 603–612.

Takahashi, H., Takigawa, I., Watanabe, M., Anwar, D., Shibata, M., Tomomori-Sato, C., Sato, S., Ranjan, A., Seidel, C.W., Tsukiyama, T., et al. (2015). MED26 regulates the transcription of snRNA genes through the recruitment of little elongation complex. Nat. Commun. 6, 5941.

Takata, H., Nishijima, H., Maeshima, K., and Shibahara, K. -i. (2012). The integrator complex is required for integrity of Cajal bodies. J. Cell Sci. 125, 166–175.

Tapia, O., Bengoechea, R., Palanca, A., Arteaga, R., Val-Bernal, J.F., Tizzano, E.F., Berciano, M.T., and Lafarga, M. (2012). Reorganization of Cajal bodies and nucleolar targeting of coilin in motor neurons of type i spinal muscular atrophy. Histochem. Cell Biol. 137, 657–667.

Tardiff, D.F., and Rosbash, M. (2006). Arrested yeast splicing complexes indicate stepwise snRNP recruitment during in vivo spliceosome assembly. RNA. 12, 968–979.

Taylor, J.P., Brown, R.H., and Cleveland, D.W. (2016). Decoding ALS: From genes to mechanism. Nature. 539, 197–206.

Thandapani, P., O'Connor, T.R., Bailey, T.L., and Richard, S. (2013). Defining the RGG/RG Motif. Mol. Cell. 50, 613–623.

Tollervey, J.R., Curk, T., Rogelj, B., Briese, M., Cereda, M., Kayikci, M., König, J., Hortobágyi, T., Nishimura, A.L., Župunski, V., et al. (2011). Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. Nat. Neurosci. 14, pp. 452-458.

Tsuiji, H., Iguchi, Y., Furuya, A., Kataoka, A., Hatsuta, H., Atsuta, N., Tanaka, F., Hashizume, Y., Akatsu, H., Murayama, S., et al. (2013). Spliceosome integrity is defective in the motor neuron diseases ALS and SMA. EMBO Mol. Med. 5, pp. 221-234.

Turner, B.J., Bäumer, D., Parkinson, N.J., Scaber, J., Ansorge, O., and Talbot, K. (2008). TDP-43 expression in mouse models of amyotrophic lateral sclerosis and spinal muscular atrophy. BMC Neurosci. 9 pp. 104.

Tyc, K., and Steitz, J.A. (1989). U3, U8 and U13 comprise a new class of mammalian snRNPs localized in the cell nucleolus. EMBO J. 8, pp. 3113-3119.

Tycowski, K.T., Shu, M. Di, Kukoyi, A., and Steitz, J.A. (2009). A Conserved WD40 Protein Binds the Cajal Body Localization Signal of scaRNP Particles. Mol. Cell. 34, pp. 47-57.

Vance, C., Rogelj, B., Hortobágyi, T., De Vos, K.J., Nishimura, A.L., Sreedharan, J., Hu, X., Smith, B., Ruddy, D., Wright, P., et al. (2009). Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. Science 323, pp. 1208-1211.

Voss, K., Forné, I., Descostes, N., Hintermair, C., Schüller, R., Maqbool, M.A., Heidemann, M., Flatley, A., Imhof, A., Gut, M., et al. (2015). Site-specific methylation and acetylation of lysine residues in the C-terminal domain (CTD) of RNA polymerase

II. Transcription. 6, 91–101.

Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The Spliceosome: Design Principles of a Dynamic RNP Machine. Cell. 136, 701–718.

Walsh, M.J., Cooper-Knock, J., Dodd, J.E., Stopford, M.J., Mihaylov, S.R., Kirby, J., Shaw, P.J., and Hautbergue, G.M. (2015). Invited Review: Decoding the pathophysiological mechanisms that underlie RNA dysregulation in neurodegenerative disorders: A review of the current state of the art. Neuropathol. Appl. Neurobiol. 41, pp. 109–134.

Wang, Q., Sawyer, I.A., Sung, M.H., Sturgill, D., Shevtsov, S.P., Pegoraro, G., Hakim, O., Baek, S., Hager, G.L., and Dundr, M. (2016). Cajal bodies are linked to genome conformation. Nat. Commun. 7, pp. 10966.

Wang, W.Y., Pan, L., Su, S.C., Quinn, E.J., Sasaki, M., Jimenez, J.C., MacKenzie, I.R.A., Huang, E.J., and Tsai, L.H. (2013). Interaction of FUS and HDAC1 regulates DNA damage response and repair in neurons. Nat. Neurosci. 16, pp. 1383-1391.

Wang, Y., Wang, F., Wang, R., Zhao, P., and Xia, Q. (2015). 2A self-cleaving peptide-based multi-gene expression system in the silkworm Bombyx mori. Sci. Rep. 5:16273.

Wani, S., Hirose, Y., and Ohkuma, Y. (2014). Human RNA polymerase II-associated protein 2 (RPAP2) interacts directly with the RNA polymerase II subunit Rpb6 and participates in pre-mRNA 3'-end formation. Drug Discov. Ther.

Webster, C.P., Smith, E.F., Bauer, C.S., Moller, A., Hautbergue, G.M., Ferraiuolo, L., Myszczynska, M.A., Higginbottom, A., Walsh, M.J., Whitworth, A.J., et al. (2016). The C9orf72 protein interacts with Rab1a and the ULK1 complex to regulate initiation of autophagy. EMBO J. 35, 1656–1676.

White, R.J. (2011). Transcription by RNA polymerase III: More complex than we thought. Nat. Rev. Genet. 12, 459–463.

Will, C.L., Urlaub, H., Achsel, T., Gentzel, M., Wilm, M., and Lührmann, R. (2002). Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p

homologue and an SF3b DEAD-box protein. EMBO J. 21, pp. 4978-4988.

Woo, J.S., Imm, J.H., Min, C.K., Kim, K.J., Cha, S.S., and Oh, B.H. (2006). Structural and functional insights into the B30.2/SPRY domain. EMBO J. 25, 1353–1363.

Xiao, R., Tang, P., Yang, B., Huang, J., Zhou, Y., Shao, C., Li, H., Sun, H., Zhang, Y., and Fu, X.D. (2012). Nuclear Matrix Factor hnRNP U/SAF-A Exerts a Global Control of Alternative Splicing by Regulating U2 snRNP Maturation. Mol. Cell *45*, 656–668.

Yamamoto, J., Hagiwara, Y., Chiba, K., Isobe, T., Narita, T., Handa, H., and Yamaguchi, Y. (2014). DSIF and NELF interact with Integrator to specify the correct post-transcriptional fate of snRNA genes. Nat. Commun. 5, 4263

Yamazaki, T., Chen, S., Yu, Y., Yan, B., Haertlein, T.C., Carrasco, M. a., Tapia, J.C., Zhai, B., Das, R., Lalancette-Hebert, M., et al. (2012). FUS-SMN Protein Interactions Link the Motor Neuron Diseases ALS and SMA. Cell Rep. *2*, 799–806.

Yoshimura, A., Fujii, R., Watanabe, Y., Okabe, S., Fukui, K., and Takumi, T. (2006). Myosin-Va Facilitates the Accumulation of mRNA/Protein Complex in Dendritic Spines. Curr. Biol. 16, 2345–2351

Yu, Y., and Reed, R. (2015). FUS functions in coupling transcription to splicing by mediating an interaction between RNAP II and U1 snRNP. Proc. Natl. Acad. Sci. 112, 8608–8613.

Yu, Y., Chi, B., Xia, W., Gangopadhyay, J., Yamazaki, T., Winkelbauer-Hurt, M.E., Yin, S., Eliasse, Y., Adams, E., Shaw, C.E., et al. (2015). U1 snRNP is mislocalized in ALS patient fibroblasts bearing NLS mutations in FUS and is required for motor neuron outgrowth in zebrafish. Nucleic Acids Res. *43*, 3208–3218.

Yu, Y.T., Shu, M. Di, and Steitz, J.A. (1998). Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. EMBO J. 17, 5783–5795.

Zaborowska, J., Egloff, S., and Murphy, S. (2016). The pol II CTD: New twists in the tail. Nat. Struct. Mol. Biol. 23, 771–777.

Zaric, B., Chami, M., Rémigy, H., Engel, A., Ballmer-Hofer, K., Winkler, F.K., and Kambach, C. (2005). Reconstitution of two recombinant LSm protein complexes reveals aspects of their architecture, assembly, and function. J. Biol. Chem. 280, 16066–16075.

Zhang, D.W., Mosley, A.L., Ramisetty, S.R., Rodríguez-Molina, J.B., Washburn, M.P., and Ansari, A.Z. (2012). Ssu72 phosphatase-dependent erasure of phospho-Ser7 marks on the RNA polymerase II C-terminal domain is essential for viability and transcription termination. J. Biol. Chem. 287, 8541–8551.

Zhang, R., So, B.R., Li, P., Yong, J., Glisovic, T., Wan, L., and Dreyfuss, G. (2011). Structure of a key intermediate of the SMN complex reveals Gemin2's crucial function in snRNP assembly. Cell. 146, 384–395.

Zhang, T., Wu, Y.C., Mullane, P., Ji, Y.J., Liu, H., He, L., Arora, A., Hwang, H.Y., Alessi, A.F., Niaki, A.G., et al. (2018). FUS Regulates Activity of MicroRNA-Mediated Gene Silencing. Mol. Cell. 69, pp. 787–801.

Zhang, Z., Lotti, F., Dittmar, K., Younis, I., Wan, L., Kasim, M., and Dreyfuss, G. (2008). SMN Deficiency Causes Tissue-Specific Perturbations in the Repertoire of snRNAs and Widespread Defects in Splicing. Cell. 133, pp. 585-600.

Zhao, M., Kim, J.R., van Bruggen, R., and Park, J. (2018). RNA-Binding Proteins in Amyotrophic Lateral Sclerosis. Mol. Cells. 41, 818–829.

Zlotorynski, E. (2017). Transcription: Paused means poised. Nat. Rev. Mol. Cell Biol. 18, 404-405.