The Role of hnRNPU, hnRNPUL1 and FUS in Transcription and Splicing

By:

Ang Li

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

School of Biosciences

University of Sheffield

Sheffield, United Kingdom

September 2023

i. Acknowledgments

I first would like to thank my supervisors Prof. Stuart Wilson and Dr. Ian Sudbery for their guidance and help throughout this project. Thanks to my colleagues and friends Ivo, Carmen, Josh, Vicky, Jack, Mikayla, Chiara, Helen, Cristina and Justin for helping me and making the lab a fun place to work in. I would like to thank my advisors Dr. Dan Bose and Dr. Emma Thompson for their inspirations and support, and Dr. Phil Mitchell, Dr. David Turton, and Dr. Nick Van Hateran for their instructions and teaching. Lastly, I would like thank my parents and my families for supporting me emotionally and financially throughout this challenging yet fulfilling time of my life.

ii. Abstract

hnRNPU, hnRNPUL1 and FUS are nuclear RNA-binding proteins that associate closely with each other and the transcription and splicing machinery. hnRNPU has been linked to a neurodevelopmental disorder, whereas hnRNPUL1 and FUS are associated with amyotrophic lateral sclerosis (ALS). Their common connection to neuronal functions implies a potential regulatory network at the root of RNA biogenesis. In this study, we found that hnRNPU is a global repressor of both RNA polymerase II (PolII) transcription and co-transcriptional splicing. hnRNPU interacts with U2 snRNP, possibly repressing its splicing functions and this feeds back on PolII activities. Through its splicing repression, hnRNPU regulates NR4A1, a repressor of immediate early response genes. We established that hnRNPUL1 is a regulator of U7 snRNA levels and histone mRNA biogenesis, and that hnRNPUL1 loss induces global reduction in PolII transcripts, apart from immediate response genes, yet exerts limited effects on splicing. We further discovered that rapid loss of FUS resulted in a bias of transcription towards AT-rich genes, which could be explained by its association with U1 snRNP, and potentially its regulation of the U1 telescripting mechanism. Together our data lead to a model of the regulatory network formed by hnRNPU, hnRNPUL1 and FUS that finely balances PolII transcription, splicing and expression of genes.

iii. Table of Contents

iv. List of Figures

v. Abbreviations

AID	Auxin inducible degron
ALS	Amyotrophic lateral sclerosis
ATP	Adenosine triphosphate
3-IAA	3-indole acetic acid
5-Ph-IAA	5-phenyl acetic acid
cDNA	Complementary DNA
ChIP	Chromatin immunoprecipitation
CTD	C-terminal domain
CUT&RUN	Cleavage under targets & release using nuclease
Dox	Doxycycline
DNA	Deoxyribonucleic acid
eCLIP	enhanced cross-linking and immunoprecipitation
CDK	Cyclin dependent kinase
hnRNP	heterogeneous nuclear ribonucleoprotein
iCLIP	Individual-nucleotide resolution UV crosslinking and immunoprecipitation
IEG	Immediate early gene
IGV	Integrative Genome Viewer
IP	immunoprecipitation
Log2 FC	log2 fold change
lncRNA	Long non-coding RNA
mNET-seq	mammalian native elongating transcript sequencing
mRNA	Messenger RNA
RNA-Seq	RNA sequencing
NEXT	Nuclear exosome targeting
NTP	Nucleoside triphosphate
CPA	Cleavage and polyadenylation complex
PCR	Polymerase chain reaction
PolyA	Poly-adenosine
PROMPT	Promoter upstream transcript
qRT-PCR	Reverse transcription and quantitative PCR
RD	Replication-dependent
RNA	Ribonucleic acid RNA
CB	Cajal body
rRNA	Ribosomal RNA
scaRNA	Small cajal body-specific RNA
HLB	Histone locus body
Ser2P	Phosphorylated Serine-2
Ser5P	Phosphorylated Serine-5
Ser7P	Phosphorylated Serine-7
siRNA	Small interfering RNA
snRNA	Small nuclear RNA
snRNP	Small nuclear ribonucleoprotein
TSS	Transcription start site
TTS	Transcription termination site
TT-Seq	Transient transcriptome sequencing (v/v)

I. Introduction

I.1. Nascent PolII Transcripts are regulated by spliceosome activities

In non-plant eukaryotes, RNAs are transcribed by three multiprotein RNA polymerases (RNA Pol). PolI transcribes the majority of ribosomal RNAs (rRNAs), PolIII transcribes transfer RNAs (tRNAs), the 5.8s rRNA and the U6 snRNA, and PolII transcribes messenger RNAs (mRNAs) and the remaining non-coding RNAs (ncRNAs), such as long non-coding RNAs (lncRNAs), small nucleolar RNAs (snoRNAs) and the majority of small nuclear RNAs (snRNAs) (reviewed by Cramer et al., 2008).

The C-terminus domain (CTD) of Rpb1, the largest subunit of PolII, contains conserved tandem repeats of the heptapeptide $Y_1S_2P_3T_4S_5P_6S_7$, that are post-translationally modified during each phase of transcription (Jasnovidova and Stefl, 2012). These modifications alter the structure of Rpb1 CTD and its interactions with various transcription factors (TFs), which couple PolII transcription to co-transcriptional processing, such as splicing and polyadenylation (PolyA), and regulate the diversity and accuracy of PolII transcripts according to cellular signals, genomic contexts and epigenetic markers.

I.1.1. PolII Initiation, Pausing and Elongation

PolII transcription takes place in several steps, which are partially marked by dynamic phosphorylation of serines (Ser) 2, 5 and 7 of the Rpb1 CTD heptapeptide. A standard model of PolII transcription of a eukaryotic gene starts with an unmodified PolII binding at the promoter region, where it forms the preinitiation complex (PIC) with the Mediator complex and General Transcription Factors (GTFs) TFII-D, A, B, F, E, and H (Greber and Nogales, 2019). Phosphorylation of Rpb1 CTD heptapeptide Ser5 (PolII Ser5P) by cyclin dependent kinase (CDK) 7 within TFIIH evicts Mediator from PolII and initiates transcription.

At 20 to 120 bases downstream of the transcription start site (TSS) of metazoan genes, PolII is paused and stabilized by the negative elongation factors NELF and DSIF (Chen, Smith and Shilatifard, 2018). Phosphorylation of Rpb1 CTD heptapeptide Ser2 (PolII Ser2P), NELF, and DSIF by CDK9 in the positive transcription elongation factor b (p-TEFb) complex release PolII from NELF to start productive elongation of the transcript. The phosphorylated DSIF becomes a positive elongation factor that stabilizes elongating PolII.

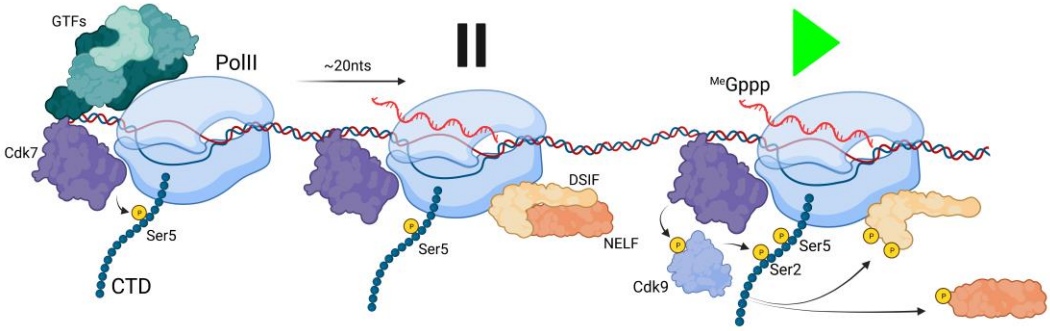

Figure I.1.1. Initiation, pausing and release of PolII elongation

A schematic of assembly and phosphorylation of major factors during the initiation (left), pausing (middle) and release of PolII transcription. PolII first assembles on the promoter of a gene, where general transcription factors (GTFs) are recruited. Cdk7 of the TFIIH complex phosphorylates PolII CTD Ser5 residues and initiates transcription. PolII is paused at 20 to 120 nucleotides (nts) downstream and stabilised by unphosphorylated NELF and DSIF. Cdk9, part of the pTEFb complex, phosphorylates PolII, NELF and DSIF, which evicts NELF and releases PolII and DSIF into productive elongation.

I.1.2. Co-transcriptional splicing

Some eukaryote protein-coding genes and long non-coding RNAs contain introns. The process of splicing removes introns from the premature RNA (pre-RNA) transcripts and join the exons together to generate the mature RNA.

In the well-studied model of the major spliceosome of U1, U2, U4, U5 and U6 small nuclear ribonucleoproteins (snRNPs), the five-prime splice site (5′SS) of the nascent transcribed intron is first bound by U1 snRNP, then the three-prime splice site (3′ss) branch point sequence (BPS) is bound by the U2 snRNP, forming pre-spliceosome complex A (reviewed by Herzel et al, 2017). Complex A then recruits U4/U5/U6 tri-snRNP to form the pre-catalytic spliceosome complex B. Structural rearrangement of complex B releases U1 and U4 snRNPs, and configures the remaining U2, U5 and U6 snRNPs into the activated complex B (B^{act}). Prp2-ATPase activity triggers B^{act} into a catalytic form (B*) for Step I catalysis, which breaks the 5′SS and joins it to the BPS adenosine to form a branched intron lariat-containing complex C. Complex C is activated by Prp16-ATPase to a catalytic form (C*) for Step II catalysis, where the 3′OH of the 5′ exon attacks the first nucleotide of the 3′ exon. The attack cleaves the intron lariat from the joint exon-exon to form post spliceosome complex P. Lastly, Prp43 ATPase activity removes the intron lariat from complex P for degradation and recycles spliceosome components for the next splicing event (Figure I.1.2).

Apart from the major spliceosome, eukaryotic splicing could also be mediated by the minor spliceosome, consisting U11, U12, U4atac, U5 and U6atac snRNPs (Tarn & Steitz, 1996). While the general principle and mechanisms of actions of the major and minor spliceosome are similar, they prefer different 5′ and 3′ splice sites (Tarn & Steitz, 1996). This arose from the different 5′ base-pairing sequences between U1 and U11 snRNAs, and the different 3′ base-pairing sequence between U2 and U12 snRNAs. The minor spliceosome-specific introns are only a small fraction of introns in the eukaryotic genomes, and most of these genes carry only one U12-dependent intron (Turunen et al., 2012).

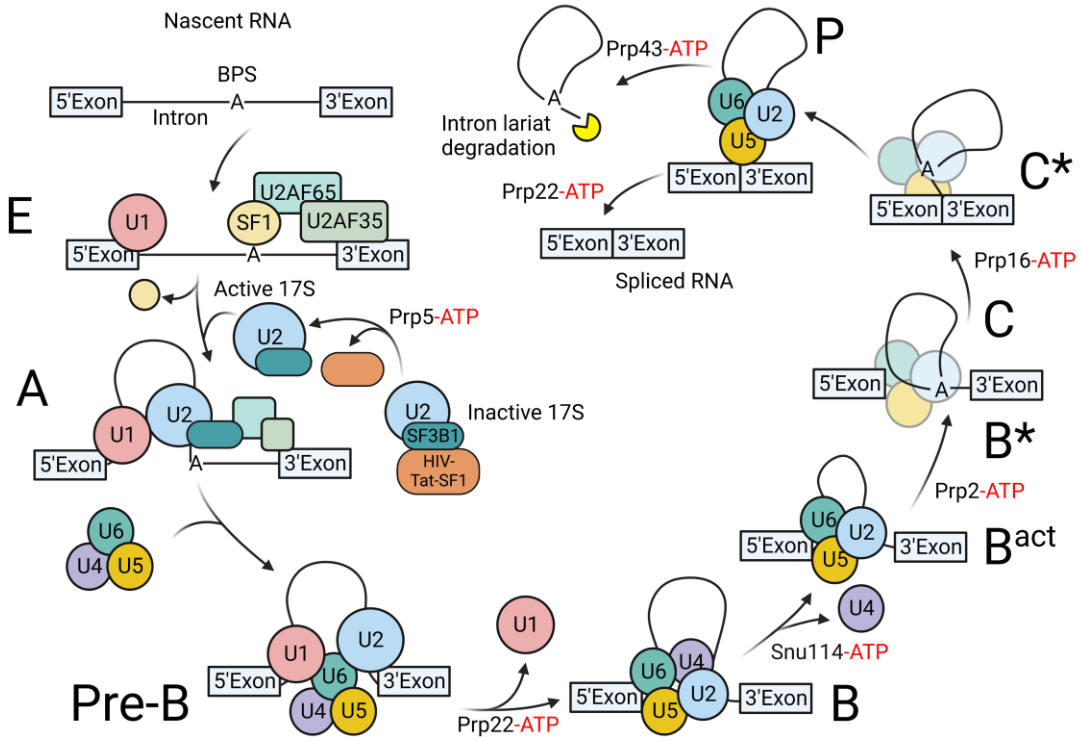

Figure I.1.2. U2-dependent splicing.

An illustration of the assembly and activation of each complex and ATPase involved in each step of the canonical U2 snRNP-dependent splicing of a typical intron, based on an illustration of a splicing cycle by Bai et al., (2018) with mechanisms of the activation and BPS-recognition of human 17S U2-snRNP-and U2 auxiliary factors presented by Rino et al. (2008) and Zhang et al. (2020). Each snRNP is abbreviated to its corresponding snRNA (U1, U2, U4, U5 and U6). The name of each splicing complex is labelled.

I.1.3. U1 snRNP and U2 snRNP regulate nascent RNA transcription and processing

Apart from mediating splicing, U1 and U2 snRNPs play wider roles in regulating the transcription and post-transcriptional processing of nascent transcripts. While the classical model suggests that U1, U2, U4, U5 and U6 snRNPs act stochiometrically in major spliceosomes, U1 snRNP is present in considerable excess to U4 and U6 snRNP in HeLa cells (Baserga & Steitz, 1993). Interestingly, in functional knockdown (KD) studies of snRNPs with antisense morpholino oligonucleotides (AMO), loss of U1 snRNP promoted premature cleavage and polyadenylation (PCPA) of nascent PolII transcripts at early intronic polyadenylation sites (PAS), while this effect was not observed globally following inhibition of U2 snRNP by spliceostatin (SSA) (Kaida et al., 2010). The same research team led by Gideon Dreyfuss later established the interaction between U1 snRNP and the cleavage and polyadenylation factors (CPAFs), that showed the disruption of this interaction by U1 snRNP-5′ splice site base-pairing (So et al., 2019). This protective role of nascent RNA from PCPA by U1 snRNP was coined "U1 snRNP Telescripting", and it was found to be more prominent in longer genes in human, as longer introns are, by random chance, likely to be more susceptible to the presence of a premature PAS (Venters et al., 2019).

In addition, recent studies showed that interactome of U1 snRNP overlaps greatly with that of PolII (Chi et al., 2018), and identified in-vivo association between CTD Ser5 phosphorylated PolII and the spliceosome during transcription at 5′ splice sites (Nojima et al., 2018). The structure of the complex of transcribing PolII associated with U1 snRNP has been solved by Cryo-electron microscopy (Cryo-EM), which revealed the direct contact of PolII components RBP12 and RBP2 with U1 snRNP component U1-70K (Zhang et al., 2021). U1 snRNP was also found to tether non-coding RNA, such as anchoring metastasis Associated Lung Adenocarcinoma Transcript 1 (MALAT1) to nascent promoter-proximal transcripts, through base-pairing of specific sequence elements of the lncRNA and the 5′ of U1 snRNA (Yin et al., 2020).

The roles of U2 snRNP in regulating transcription has been evidenced as well. An *in vivo* study showed that AMO KD of U2 snRNP impaired the 3′ expression of several cell proliferation-relevant genes, such as CDK6 and c-Myc (Koga et al., 2014). A later study of the SSA inhibition of the SF3B component of U2 snRNP showed that it led to PCPA of MALAT1 and accumulation of a short form of MALAT1 in cytoplasm

(Yoshimoto et al., 2021), which the authors attributed to the reduced availability of active U1 snRNP in spliceosome following U2 snRNP inhibition. However, in contrast to the SSA-based U2 snRNP inhibition, specific inhibition of the branch point sequence (BPS) recognizing SF3B1 subunit by pladienolide B (PlaB) resulted in global accumulation of paused promoter-proximal PolII, reduced recruitment of pTEFb and a reduced PolII transcription rate at the start of genes (Caizzi et al., 2021). This suggests that U2 snRNP activity may directly feedback positively to the pause-release and elongation of PolII, which would further enhance transcription activation (Gressel et al., 2019).

I.1.4. SF3B1 conformation changes by TAT-SF1 eviction activates U2 snRNP

As U2 snRNP is present throughout the splicing cycle, dynamic changes in its components and configuration are indicative of splicing complex formation and splicing progression. Early studies isolated two forms of U2 snRNP from HeLa cell: an inactive 12S complex composed of U2 snRNA, Sm proteins, U2A′ and U2B’’; and an *in vitro* splicing-capable 17S complex composed of SF3a and SF3b complexes in addition to the 12S complex (Behrens, Galisson, Legrain and Lührmann, 1993; Brosi, Hauri and Krämer, 1993). Mutations in SF3B1, a subunit of SF3b, disrupted U2 snRNP recognition and catalysis of BPS, which relies on the correct conformation of U2 snRNA branchpoint-interacting stem-loop (BSL) (Perriman and Ares, 2010; Darman et al., 2015). Structural studies revealed that U2 snRNA BSL is confined by a closed SF3B1 HEAT domain in the active spliceosome, while this domain is found in an open conformation in isolated SF3b (Cretu et al., 2016; Kastner, Will, Stark and Lührmann, 2019). Recent Cryo-electromagnetic microscopy (Cryo-EM) studies of the 17S U2 snRNP structure showed that HIV-TAT-SF1, a factor that associates with U2 snRNP yet not found in the active spliceosome, restricts the SF3B1 HEAT domain in the inactive open conformation, and its eviction is likely essential for assembly of U2 snRNP into the pre-spliceosome (Zhang et al., 2020).

I.2. Nuclear RNPs mediate RNA biogenesis

The eukaryotic cell nucleus is a membrane-bound organelle that serves as the production hub of RNA from chromosomal DNA templates. In metazoans and plants, the cell nucleus contains membraneless bodies, such as the nucleolus, Cajal bodies , histone locus bodies (HLBs) and nuclear paraspeckles, that modulate the biogenesis, stability, and export of different RNA classes. These granules are formed by scaffolding proteins and RNAs, as well as their substrate RNAs or ribonucleoproteins (RNPs). The formation and dynamics of these granules are indicative of RNA transcription, processing and export.

I.2.1. SnRNP biogenesis mediated by Cajal bodies

In the human genome, each snRNA is encoded by multiple copies of genes, where their transcription is driven by a promoter region defined by a distal sequence element (DSE) and a proximal sequence element (PSE) (reviewed in Fischer et al., 2011). The major spliceosomal U1, U2, U4, U5 and minor spliceosomal U4, U4atac and U5 snRNAs are transcribed by PolII, and their transcription termination is signalled by a 3 ′ box, where the megadalton Integrator complex mediates endolytic cleavage and subsequent 3′ processing (Baillat et al., 2005). In contrast, U6, U6atac and 7SK snRNAs are transcribed by PolIII, where their transcription initiation requires a 5′ TATA-box and the termination is signalled by a stretch of template encoded uridines (Reddy et al., 1987; Zieve et al., 1977). PolII transcribed snRNAs are co-transcriptionally capped with a 5′ m7Gppp cap, which is recognised by the Cap-binding complex (CBC), the snRNA export factors PHAX, Ars2 and CRM1 (Fornerod et al., 1997; Hallais et al., 2013). In the cytoplasm, the survival of motor neuron (SMN) complex, formed by the SMN and Gemin proteins, assembles seven core Sm proteins onto snRNAs in an ATP-dependent manner, which stabilised the snRNP (Pellizzoni et al., 2002). The SMN-bound snRNA undergoes 5′ processing by the methyltransferase Tgs1, which converts the m7G cap into a 2,2,7-trimethylguanosine (TMG) (Mouaikel et al., 2003), as well as 3′ exonucleolytic degradation that leaves one to two nucleotides after the mature snRNA (Dahlberg et al., 1990). The TMG cap of Sm-association snRNA is recognized by snurportin1, which interacts with importin β and imports the snRNP back into the nucleus (Huber et al., 2002). In the nucleus, the premature snRNPs are localised to Cajal Bodies to undergo final modifications, such as 2′-O-ribose methylation and pseudouridylation, and assembled with their specific protein components into fully

mature snRNPs (Jady & Kiss, 2001; Strzelecka et al., 2010).

The Cajal Body (CB), or coiled body, was first described by Raymon y Cajal in 1903 as silver-stained coiled nuclear bodies in proximity to nucleoli (Cajal, 1910). Nowadays CBs are known as nuclear protein-RNA granules that contain nucleolar components, such as fibrillarin and nucleolin, and can be distinctly marked by the accumulation of coilin and small Cajal body-specific RNAs (scaRNAs) (Raška et al., 1991; Darzacq, 2002). CBs are likely formed by the liquid-liquid phase separation of RNA-binding proteins, which is mediated by the low-complexity domains of its components, such as coilin and Nopp140 (Courchaine et al., 2022).

CBs are involved in both the initial and final stages of snRNP formation. CBs associate with snRNA gene loci in interphase cells, and are enriched in factors required for snRNA transcription and processing, such as Integrator and SMN complexes (Carmo-Fonseca et al., 1992; Mahmoudi et al., 2010; Takata et al., 2012). They also associate with snRNP-specific components, such as U2B″ of the U2 snRNP, which are integrated at the final snRNP formation stage (Smith et al., 1995). Further, CBs associate with snRNPs independently of transcription activities, and mature snRNPs are observed to recycle through Cajal Bodies (Schul et al., 1998; Staněk et al., 2008). Functional studies indicated that loss of SMN proteins by siRNA KD disrupts CBs as well as levels of snRNAs (Girard, 2006). The evidence to date indicates that CBs are critical for snRNA biogenesis.

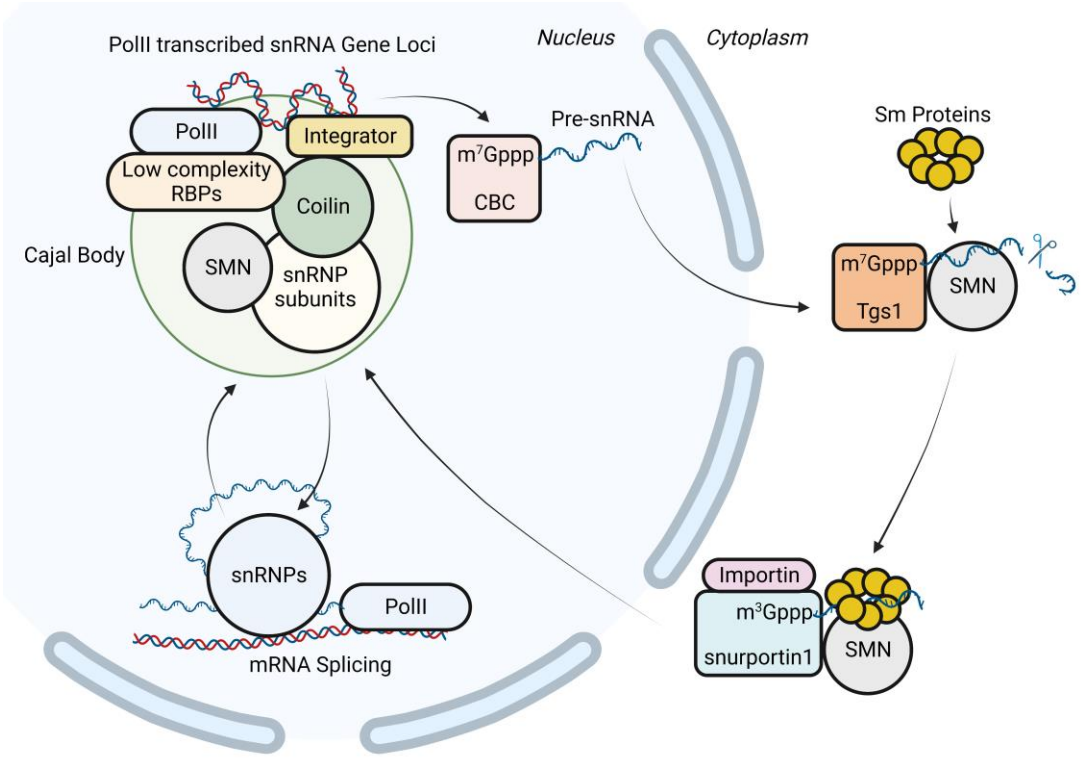

Figure I.2.1. PolII-transcribed SnRNP biogenesis and recycling through Cajal bodies.
A sketch of the biogenesis of PolII transcribed snRNAs (U1, U2, U4, U5, U4atac and maturation of corresponding snRNPs. The presented Cajal body elements are only a subset of its components that are involved in snRNP biogenesis. Note that U3 and U6 snRNA undergo different processing than illustrated here. Cajal bodies form in proximity to snRNA gene loci, and its PolII and integrator component transcribe, process and cap the nascent snRNA transcript with a m^7G cap, which is recognised by the CBC complex and exported into the cytoplasm. In the cytoplasm, the 3′ end of the snRNA is cleaved, SMN proteins assemble Sm ring proteins on the snRNA, and Tgs1 modifies the m^7G cap to a TMG (m^3G) cap. The TMG cap is recognised by snurportin, which recruits importin β to transport the semi-assembled snRNP back into nucleus. SnRNPs undergo final assembly in Cajal bodies with their specific subunits. Mature snRNPs might be recycled through the Cajal body.

I.2.2. U7 snRNP is required for replication-dependent histone mRNA biogenesis

Histones are abundant nuclear proteins that pack DNA into nucleosomes, which is typically 146 base-pairs of DNA wrapped around octamers of 2 copies of each of H2A, H2B, H3 and H4 histones, and stabilized by the linker histone H1 (McGinty & Tan, 2014). These canonical, intronless histone genes are expressed in a replication-dependent (RD) manner during the S-phase of cell division to package newly synthesized DNA (reviewed in Marzluff et al., 2008). Several variant histones, such as H2A.X and macroH2A, are constitutively expressed independently of replication and they serve specialized functions in modifying chromatin structure and regulating gene expression (reviewed in Talbert & Henikoff, 2016).

In the human genome, replication-dependent histones are arranged into several clusters: the large HIST1 locus on chromosome 6 that carries 55 RD histone genes, and the HIST2, HIST3 and HIST4 loci on chromosome 1 carry 12, 4 and 1 RD histone genes, respectively (Marzluff et al., 2002). The clusters of histone genes are packed into Histone Locus Bodies (HLB), which are Cajal Body-like nuclear RNP complexes (Nizami et al., 2010). Although the HLBs can overlap Cajal Bodies and may share some functional components (Suzuki et al., 2022), it is distinctly marked by the presence of histone processing factors, such as nuclear protein at the ataxia-telangiectasia locus (NPAT), FLICE-associated huge protein (FLASH) and the U7 snRNP (Yang et al., 2014).

The transcription of RD histones is activated by the cyclin E/cdk2 phosphorylation of NPAT upon entry into S-phase (Zhao et al., 1998; Zhao et al., 2000). The nascent RD histone pre-mRNA does not undergo polyadenylation as PolII is slowed down at the 3′ stem loop and does not typically reach a PAS (Suzuki et al., 2022). This 3′ stem loop is recognised by stem loop binding protein (SLBP), which interacts with FLASH and forms a docking platform for U7 snRNP to base-pair with the histone downstream element (HDE) (Dominski et al., 1999; Suzuki et al., 2022). Histone cleavage complex (HCC), which contains the endonuclease CPSF73, is recruited to U7 snRNP and mediates a cleavage between the stem-loop and HDE (Sun et al., 2020). This cleaved histone mRNA is exported into the cytoplasm with SLBP, where SLBP interacts with SLIP1 to recruit translation factors eIF4G and eIF3 to initiate S-phase histone synthesis (Cakmakci et al., 2008).

While RD histones are not typically polyadenylated, disruption of the histone 3′ processing mechanisms may result in extended elongation of PolII and usage of downstream PAS for some histone genes. This has been shown in a mutant study of fused in sarcoma (FUS) gene, where the U7 snRNP levels or localization were disrupted, and increased polyadenylation or extension and decreased expression of RD histones were observed (Gadgil et al., 2021). Additionally, hnRNPUL1 has been shown to associate with the U7 snRNP and play a role in repression of RD histone gene transcription in cell cycle arrested cells (Ideue et al., 2012).

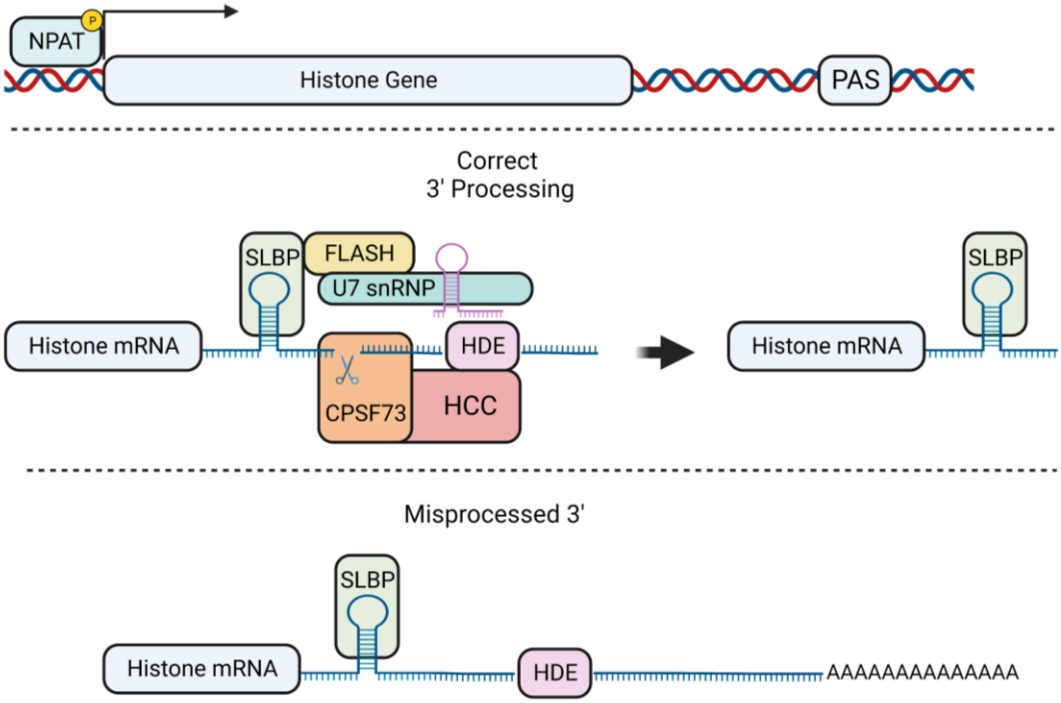

Figure I.2.2. 3′ processing and mis-processing of RD-histone genes

A schematic showing the fates of RD histone pre-mRNA transcripts transcribed from a NPAT-activated loci with downstream poly-adenylation site (PAS) sequence (top panel). Correct 3′ processing of the RD histone pre-mRNA relies on binding of SLBP-FLASH-U7 snRNP complexes to the 3′ stem-loop and the HDE sequence, and the recruitment of HCC complex to mediate a cleavage by the CPSF73 subunit after the stem-loop to produce a cleaved RD histone mRNA (centre panel). Failure in assembly of these complexes results in transcription of the PAS and addition of poly(A) tail to the transcript (bottom panel).

I.2.3. Nuclear paraspeckles regulates Alu-rich mRNA export

Nuclear paraspeckles are nuclear RNA-protein bodies found in differentiated mammalian cells in the periphery of splicing speckles, which are enriched in splicing factors and nuclear mRNAs (Fox et al., 2002; Hutchinson et al., 2007). Paraspeckles are marked by the presence of paraspeckle proteins (PSPs) PSP1, SFPQ/PSF and NONO/P54^{nrb} (Fox et al., 2002), as well as its core long non-coding RNA (lncRNA) nuclear-enriched autosomal transcript 1 (NEAT1) (Sunwoo et al., 2008). There are two major isoforms of NEAT1: the shorter and polyadenylated NEAT1_1, and the longer, RNase P cleaved and 3′ triple-helix stabilized NEAT1_2, both produced from the same gene loci (Guru et al., 1997; Sunwoo et al., 2008). NEAT1_2 is the dominant isoform that forms paraspeckles, where its 5′ and 3′ regions associate with PSPs and other nuclear RBPs, such as FUS and hnRNPH3, to promote the LLPS-based formation of an outer "shell", which wraps NEAT1_2 central region in its "core" (West et al., 2016; Taiana et al., 2020).

The best characterised cellular function of nuclear paraspeckles is the nuclear retention of Alu element-rich and adenosine-to-inosine (A-to-I) edited RNA. A-to-I editing of nascent nuclear RNA is facilitated by the adenosine deaminase acting on RNA (ADAR) enzyme, which recognises double-stranded RNAs formed by inverted repeat elements, such as the primate specific Alu repeats (Kim et al., 2004). Zhang & Carmichael (2001) showed that microinjected double-stranded RNAs are A-to-I edited and retained in the nucleus of *Xenopus* oocyte. This retention was not rescued by splicing, which could facilitate export of spliced transcripts through the TREX mRNA export pathway (Viphakone et al., 2019). They further demonstrated that A-to-I edited RNAs are recognised by a complex containing paraspeckle-associated NONO/P54^{nrb} and SFPQ/PSF (Zhang & Carmichael, 2001). Further, the nuclear retention of A-to-I edited RNA is not present in human embryonic stem cells (hESCs), and appears only upon differentiation, which induces the expression of NEAT1 and formation of paraspeckles (Chen & Carmichael, 2009). Therefore, fully functional NEAT1 is required for nuclear paraspeckle formation and functions in A-to-I edited RNA retention.

I.3. Hetergeneous Nuclear Ribonucleoproteins

Heterogeneous nuclear ribonucleoproteins (hnRNPs) are a family of at least 20 nuclear RNA-binding proteins (RBPs) that interact with heterogenous nuclear RNAs (hnRNAs) (Dreyfuss et al., 1993, reviewed in Geuens et al., 2016). This family of proteins contains a variety of RNA-binding domains, such as the structured RNA Recognition Motif (RRM) and the intrinsically disordered RGG-repeat boxes (Geuens et al., 2016).

I.3.1. hnRNPU

hnRNPU, or Scaffold Attachment Factor A (SAF-A) is one of the largest members of hnRNPs. It was first isolated in the less soluble DNA-bound fraction of the human cell nucleus, and showed high affinity for the DNA scaffold/matrix attachment regions (S/MAR) (Dreyfuss et al, 1993; Göhring & Fackelmayer, 1997). The 120kDa peptide contains a DNA-binding SAP domain, a SPRY domain, an ATPase-like domain, and an RNA-binding RGG-box domain (Kiledjian and Dreyfuss, 1992; Aravind and Koonin, 2000; Figure I.3.1.A). The complete 3D structure of hnRNPU is yet to be solved using X-crystallography, as the C-terminal RGG domain of hnRNPU is of low-complexity and intrinsically disordered (Järvelin et al., 2016). However, a high confidence structural prediction of the SPRY-PNK domain has been produced by AlphaFold (AlphaFoldDB Q00839), and a structure of the SAP domain of the related protein hnRNPUL1 has been solved with NMR (PDB: 1ZRJ). Using the structure of the SAP domain, Sharp et al. (2020) demonstrated that phosphorylation of hnRNPU SAP domain DNA-contacting serines (S14, S26) by Aurora-B kinase during cell division disrupted hnRNPU DNA-binding. This in turn leads to eviction of chromatin-bound RNAs which are normally tethered there by hnRNPU and this allows chromatid condensation and separation. Further, Serine 59 of hnRNPU was identified as a target of DNA-dependent phosphatase kinase in response to DNA double strand breaks (Berglund & Clarke, 2009), suggesting a role of hnRNPU in DNA-damage repair.

The high-confidence Alphafold 2.0 predicted structure of hnRNPU SPRY-ATPase domain revealed high structural homology for the ATPase domain with the kinase domain of mammalian polynucleotide kinase phosphatase (Yonchev et al., 2023). The ATPase-like domain of hnRNPU preserved the spatial structures of the P-loop, Walker A and Walker B motifs, which are essential for NTP-binding in the PNKP substrate binding pocket (Shalaeva et al., 2018). An *in vivo* study by Nozawa et al. (2017)

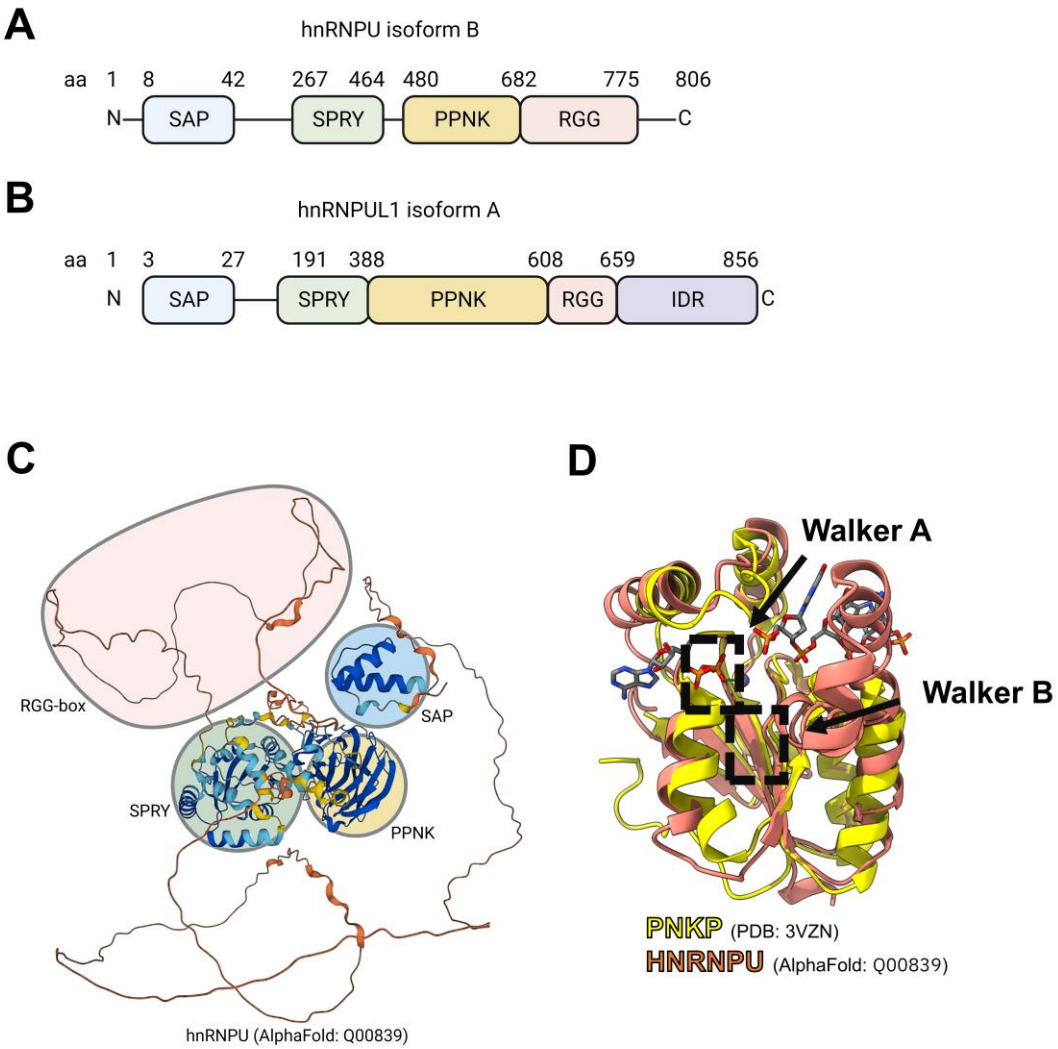

Figure I.3.1. HnRNPU and hnRNPUL1 structures and similarities to PNKP

A, Schematic of the domain structure of hnRNPU isoform B. The amino acid residue (aa) number of the start, junction and ending positions of SAP, SPRY, PPNK and RGG-box domains and the protein are annotated.

B, Schematic of the sequence structure of hnRNPUL1 isoform A. The amino acid number of the start, the junctions and the end positions of each domain and the protein is annotated as in **A**.

C, AlphaFold predicted structure of a monomer of hnRNPU (Uniprot/AlphaFold: Q00839). The known functional domains are circled and annotated. The ribbon representation of the protein structure is coloured according to the confidence of the prediction of the structure, where yellow-orange represents low confidence (<70%) and light-blue and blue represents high confidence (>70%).

D, Superposition of the predicted structure of the hnRNPU PPNK domain and the resolved structure of the PNKP kinase domain, with an ATP and a single-strand DNA substrate of PNKP positioned according to the PNKP structure. Positions of the Walker A and the Walker B motifs are labelled with dashed rectangles and arrows.

revealed that oligomerisation of hnRNPU was required for decompaction and stability of chromosomes, while this was disrupted by a loss-of-function mutation in the hnRNPU Walker A motif, but not by a mutation in the Walker B motif. They also reported that both Walker A and Walker B motifs were required for a weak ATPase activity of a fragment of hnRNPU purified from *E.coli*. However, we did not observe ATPase activity from wild type hnRNPU after stringent purification and removal of a contaminant HSP70 ATPase which copurified from *E.coli*. Furthermore, when we purified full length human hnRNPU directly from human cells we failed to detect any ATPase activity, despite using a sensitive thin layer chromatography assay with radiolabelled ATP (Wilson Lab, unpublished).

The low-complexity hnRNPU RGG-box consists of a core RGGDFRGGAPGDRGG tri-RGG sequence, which is essential for RNA-binding of hnRNPU (Kiledjian & Dreyfuss, 1992), in addition to several di-RGG repeats. The RGG-box is required for hnRNPU binding to XIST RNA and the inactivated X-chromosome (Helbig & Fackelmayer, 2003), where hnRNPU is believed to facilitate X-chromosome silencing.

In addition to its structure-related roles in regulating chromatin condensation and stability, functional analyses suggested that hnRNPU is a potential PolII transcription regulator, yet the literature presents a mixed picture regarding whether it is a transcription inhibitor or co-activator. hnRNPU was found to co-localise to chromatin looping factors RAD21 and CTCF binding sites, and its depletion disrupted TAD boundaries definition and altered gene expressions (Fan et al., 2017). The central domains of hnRNPU, consisting the SPRY-PNK-like domains, were reported to inhibit phosphorylation of PolII CTD by TFIIH, whose kinase activity is mediated by CDK7 (Kim & Nikodem, 1999; Glover-Cutter et al., 2009). On the other hand, the C-terminal sequences of hnRNPU was found to interact with actin and PCAF histone acetyletransferase (HAT) to activate PolII transcription (Obrdlik et al., 2008). This was further supported by the report of VEGF transactivation by extracellular vesicles containing hnRNPU (Wang et al., 2022).

HnRNPU was implicated as a regulator of splicing as well, both directly and indirectly. It was found in only one of the two large-scale mass-spectrometry (MS) assays of purified spliceosomes (Rappsilber et al., 2002; Zhou et al., 2002), suggesting that it is not a core component of the spliceosome. However, its depletion resulted in global

alteration of alternative-splicing patterns, particularly of increased skipped-exon inclusions and reduced intron retention, as well as increased proportions of splicing-capable 17S U2 snRNPs (Huelga et al., 2012; Xiao et al., 2012). In the Wilson Lab, Dr. Yonchev also identified enrichment of hnRNPU across introns of active genes in published eCLIP datasets, suggesting a possible direct role of hnRNPU in modulating intron excision (Yonchev, 2021).

Heterozygous mutation of the hnRNPU gene or deletions at the 1q44 locus, which houses the hnRNPU gene, were reported to cause latent neural and musculoskeletal developmental, facial malformation and seizure, termed as the hnRNPU syndrome (Yates et al., 2017). hnRNPU has been implicated in splicing regulation, where it appears to exert negative control over U2 snRNP. Its depletion using RNA interference led to increased levels of active 17S U2 snRNP, increased formation of Cajal bodies and widespread changes in alternative splicing of pre-mRNAs (Xiao et al., 2012). Evidence from a hnRNPU knock-out (KO) model in mice suggested that hnRNPU depletion resulted in mis-splicing of key factors in Akt signalling pathway, which lead to reduced muscle growth and regeneration (Bagchi et al., 2020). Splicing regulation by hnRNPU was shown to be essential for the survival of neuronal progenitors and neurodevelopment of mice cortex (Sapir et al., 2022). Lastly, a recent screen across all hnRNPs revealed that mutations of hnRNPU were significantly associated with neurodevelopment disease, and disease-related mutations were identified across all functional domains of the protein (Gillentine et al., 2021). The accumulating evidence is positioning hnRNPU as a key regulator of muscle and neuronal development.

I.3.2. HnRNPU-like 1

HnRNPU-like 1 (hnRNPUL1), also known as E1B-55kDa-associated-protein 5 (E1B-AP5), has 56% peptide sequence homology to hnRNPU and a highly similar structure arrangement (Gabler et al., 1998). It contains a SAP domain, a SPRY and a PNK-like domain and a differently arranged RGG-box with several di-RGG repeats (Thandapani et al., 2013). In addition, hnRNPUL1 contains a C-terminal glutamine-proline-tyrosine-rich and prion-like intrinsically disordered region (IDR), which resembles the IDRs found in FET proteins (Figure I.3.1A; King et al., 2012).

HnRNPUL1 was originally identified as an interactor with adenovirus protein E1B-55kDa, where it was found to interfere with the transcription regulation by E1B-55kDa and p53, and its overexpression resulted in mRNA accumulation in the cytoplasm (Gabler et al., 1998). Later hnRNPUL1 was found to interact with nuclear export factor NXF1, further implying its roles in RNA export (Bachi et al., 2000).

The existing literature implicated hnRNPUL1 in the regulation of RNA PolII transcription of different RNA species. It was found to work with BRD7 to activate transcription at glucocorticoid-responsive promoters and inhibit them when the complex was disrupted (Kzhyshkowska et al., 2003). The interaction between hnRNPUL1 and U7 snRNP was found to repress replication-dependent histone gene transcriptions in cell-cycle arrested cells (Ideue et al., 2012). hnRNPUL1 was also found in the interactome of U1 snRNP, which included numerous PolII transcription factors and ALS-SMA causative genes (Chi et al., 2018a; Chi et al., 2018b). Large-scale ChIP analysis of nuclear RNA binding proteins (RBPs) revealed that hnRNPUL1 was enriched over transcription start sites, and interestingly, that hnRNPUL1 was the most enriched RBP over snRNA genes (Xiao et al., 2019). As hnRNPUL1 was observed to interact with Ars2, a protein that stimulates snRNA 3′ processing and determine the fate of snRNA transcripts together with CBC-NEXT and CBC-PHAX complexes (Hallais et al., 2013; Giacometti et al., 2017), hnRNPUL1 might be an important regulator of snRNA transcription and processing.

HNRNPUL1 also plays roles in DNA damage repair. Polo et al. (2012) found that hnRNPUL1 and hnRNPUL2 were excluded from DNA double strand break (DSB) sites in cells undergoing active transcription, while they were recruited to DNA DSB ends in transcription-inhibited cells by the MRN complex to facilitate resection of DNA and

repair of the break. The hotspots of hnRNPUL1 recruitment in response to DNA damage was found to overlap those of PARP1 (Hong et al., 2013), and it was later found that hnRNPUL1 recruitment to DNA DSB ends requires arginine methylation of hnRNPUL1 RGG-box by the PRMT1 methyltransferase (Gurunathan et al., 2015).

Accumulating evidence are suggesting that hnRNPUL1 is essential for neuronal functions. hnRNPUL1 was ranked 10th in a mathematical prediction of prion-like RNA-binding proteins that could potentially cause amyotrophic lateral sclerosis (ALS), a prevalent motor neuron degenerative disease (Li et al., 2013). hnRNPUL1 was reported to interact with ALS-causative proteins FUS, EWSR1 and TAF15 in co-immunoprecipitation followed by mass-spectrometry (Chi et al., 2018) and a mutation in hnRNPUL1 was identified in an ALS clinical case in a patient mutation screen (Cooper-Knock et al., 2017). Finally, the recent hnRNP screen in neurodevelopmental disease also listed hnRNPUL1 as a candidate gene, and phenotypes of hnRNPUL1 mutations moderately correlated to those of mutations in the related proteins, hnRNPU and hnRNPU-like 2 (Gillentine et al., 2021). Recent work from the Wilson laboratory has shown that hnRNPUL1 is mutated in a large cohort of ALS patients and plays a role in the 3' processing of snRNAs in conjunction with the Integrator complex (Yonchev et al, submitted).

I.3.3. FUS

FUS/TLS (Fused in Sarcoma/ translocated in liposarcoma) or FUS belongs to the ALS-causative FET protein family, together with EWSR1 and TAF15 (Law, 2006). FUS is implicated in every stage of gene expression, from transcription to translation, as well as in maintaining chromatin structure and mediating DNA damage repair.

FUS contains an N-terminal disordered SGQY-rich prion-like intrinsically disordered region (IDR) (Figure I.3.2.). This domain was first found fused to CHOP and ERG DNA binding domains to form fusion proteins, which are causative of myxloid liposarcoma and myeloid leukemia, respectively (Crozat et al., 1993; Zinszner et al., 1994; Pérez-Losada et al., 2000; Ichikawa et al., 1999). The FUS IDR in these fusion proteins is believed to act as a transcription activator that enhances tumour proliferation and progression (Zinszner et al., 1994; Kwon, et al., 2013). The IDR of FUS was reported to form amyloid-like fibrils in vitro through liquid-liquid phase separation (LLPS) (Kato et al., 2012; Lee et al., 2020). The formation of FUS condensates is controlled by phosphorylation of its IDR and the structure and properties are affected by pathogenic mutations in this domain (Patel et al., 2015; Murray et al., 2017). The LLPS properties of FUS IDR may partially explain its transcription activation potential, as it could form liquid condensates with the PolII CTD in a phosphorylation-controlled manner (Kwon, et al., 2013).

The central region of FUS is composed of three RGG-boxes with an RNA recognition motif (RRM) and a zinc finger (ZNF) domain lying in between them. *In vivo* crosslinking immuneprecipitation (CLIP) assay and *in vitro* binding data suggested that these core RNA-binding domains (RBDs) of FUS bind GU-rich RNA in a length-dependent manner, without an identifiable specific consensus-sequence (Masuda et al., 2015; Wang et al., 2015). More recently, the structure of the FUS ZNF domain bound to a ‘GGU’ core RNA sequence adjacent to a stem loop recognised by a FUS RRM has been resolved by NMR spectroscopy (Loughlin et al., 2019), and *in vitro* binding assays showed that the FUS RGG boxes have high affinity for G-quadruplex RNA secondary structures, for example, the telomeric repeat-containing RNA (Wang et al., 2015).

The C-terminus of FUS is a PY-rich nuclear localisation signal conserved between FUS and EWSR1 (Lagier-Tourenne et al., 2010). The nuclear import of FUS is mediated by transportin, which recognises the arginine methylation within the FUS NLS (Dormann

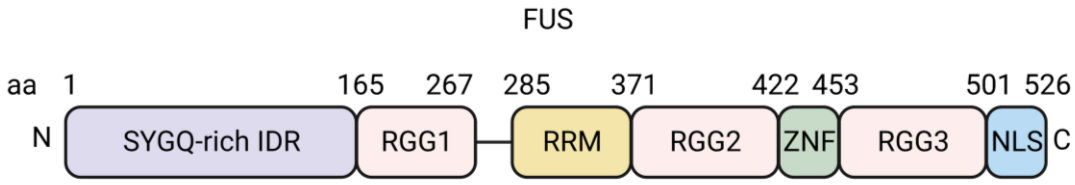

Figure I.3.2. Sequence structure of FUS.

Schematic of the domain structure of FUS. The amino acid residue (aa) number of the start, junction and ending positions of the SYGQ-rich IDR, the three RGG-boxes, the RRM, ZNF and NLS domains are annotated.

et al., 2012). Numerous ALS-causative mutations were reported in the FUS NLS (Vance et al., 2013). Mimicking these mutations in laboratory cells resulted in accumulation of FUS and mislocalization of U1 snRNA in the cytoplasm, as well as reduced Cajal Body counts and disrupted snRNA levels (Yu et al., 2015). ALS-related mutations of FUS also disrupts U7 snRNP localization to the HLBs, which resulted in misprocessing and downregulation of histone mRNAs (Raczynska et al., 2015). Further, Akiyama et al. (2018) studied two ALS-patient mutations in the FUS-NLS in neurons, where they found accumulation of an immediate early gene product FosB in the neurons and aberrant axon branching of mutated neurons.

In addition to its roles in regulating RNA biogenesis and processing, FUS is also involved in facilitating DNA damage repair and maintaining genome stability. The central RGG, RRM and zinc finger domains of FUS bind both single-stranded and double-stranded DNAs, though at a lower affinity than their interaction with RNAs (Wang et al., 2015). *In vivo* evidence showed that FUS KO cells accumulated γH2AX DNA damage marker, and DNA damage induced FUS phosphorylation by DNA-PK and cytoplasmic translocation (Deng et al., 2014). *In vitro* evidence also showed that the activity of the DNA damage sensor PARP1 induced compartmentalization of FUS and poly(ADP-ribose) (PAR) chains at DNA damage sites (Singatulina et al., 2019). The interactions between FUS RGG domain and TERRA further suggests a potential role of FUS in regulating telomeres, which are chromosome terminal DNA components associated with genome replication fidelity and ageing (Takahama et al., 2013).

I.4. Immediate Early Genes Modulate Neuronal Activity and Cancer development

Immediate early genes (IEGs) are primary response genes induced shortly after stimuli, such as hypertonic stress, growth factors and tumour promoters (Sharp et al., 1991, Cochran et al., 1984; Greenberg & Ziff, 1984; Herschman, 1991). Their expression is typically followed by expression of delayed primary response genes that poise the cells to respond to the stimuli (Tullai et al., 2007). The potential roles of immediate early genes in activating gene expression were evidenced in early studies of c-FOS and c-JUN, where they were described as retrovirus-induced transactivating oncogenes (Rauscher et al., 1988). Numerous IEGs, including c-FOS and c-JUN families and members of activating transcription factors (ATFs), were identified as leucine zipper peptide components of the heterogeneous dimeric activator protein 1 (AP-1) transcription factor (Chinenov & Kerppola, 2001), which acts on different promoter response elements depending on its composition (Hai & Curran, 1991).

IEGs are important factors in regulating physiological activities, such as memory retrieval, T-cell immune response and cancer transformation and proliferation. Studies in mice models discovered that fear memory retrieval activates expression of IEGs EGR1 and c-FOS in the hippocampus and amygdala (Rosen et al., 1998; Hall et al., 2001). Furthermore, EGR1 is required for hippocampal long-term potentiation, a process that facilitates brain synaptic response and long-term memory formation (Lynch et al., 1983; Jones et al., 2001). AP-1 family IEG expression might also be associated with T-cell activities and T-cell leukaemia transformation, as c-FOS was observed to be activated in T-cells by proliferation stimuli, where its expression lasted for a day (Clark et al., 2011), and c-FOS and c-JUN family IEGs are activated in the T-cells transformed by human T cell leukemia virus type 1 (HTLV-1) , induced by the HTLV-1 Tax protein (Fujii et al., 2000). Lastly, while multiple human cancer cell types exhibit elevated levels of c-FOS and c-JUN (Neyns et al., 1996; Kharman-Biz et al., 2013), evidence exist for both promotion and downregulation of tumour growth by overexpression of c-FOS and c-JUN, indicating that the effect of AP-1 family IEG activities on tumourogenesis is dependent on the tumour background and upregulated AP-1 components (Herschman, 1991; Shaulian, 2010; Kharman-Biz et al., 2013).

The stress stimulated expression of IEGs is dependent on their transcription activation. Stimulus-induced kinase cascades result in phosphorylation and dimerisation of c-FOS and c-JUN into active AP-1. This can feedback on the transcription initiation of IEGs,

as promoters of mammalian c-FOS, c-JUN and ATF family IEGs contain AP-1 binding sites, such as the serum response element (SRE), 12-O-tetradecanoyl-phorbol-13-acetate-response element (TRE) or cAMP-responsive element (CRE) (van Dam et al., 1993; Cavigelli et al., 1995). The pause-release of PolII at IEGs might be regulated by the RN7SK snRNP, as a recent study revealed that the UV-inducible expression of early response genes are poised by the RN7SK/HEXIM/p-TEFb complex, which releases p-TEFb to phosphorylate and release paused PolII upon stimuli (reviewed in Zhou et al., 2012; C. Quaresma et al., 2016; Studniarek et al., 2021). The duration of IEG expression might be controlled by negative feedback from NR4A1 expression (Guo et al., 2021). NR4A1 is a IEG induced at around 1 hour post-stimuli, slower than the AP-1 family IEGs, which are induced at around 30 minutes. NR4A1 acts as a DNA-binding transcriptional roadblock across EGR1, c-JUN, c-FOS family gene bodies, thereby downregulating their expression (Guo et al., 2021).

I.5. Auxin Inducible Degron allows conditional knock-out of essential proteins.

In vivo studies of proteins are often assisted by observing the effects of depletion of the protein of interest in laboratory culture. However, long-term knock-out (KO) of proteins may introduce secondary effects, and is not feasible for studying essential proteins. A common alternative method is to knockdown protein level by RNA interference (RNAi), which uses small interfering RNA (siRNA) to deplete the mRNA. However, RNAi suffers from off-target effects and incomplete silencing, as well as disruption to both the protein and the RNA, which may complicate interpretation of the results. In addition, sufficient RNAi depletion usually takes 48 to 96 hours, allowing secondary effects to manifest through cell cycles. To address the draw-backs of permanent protein KO and RNAi, an ideal protein-KO system should achieve conditional, and preferably reversible, protein depletion within a cell cycle, while requiring minimal disruptions to the genome and transcriptome.

In 2009, Prof. Masato Kanemaki's team (Nishimura et al., 2009) presented an auxin-inducible degron (AID) system that can conditionally and rapidly deplete proteins in a wide variety of non-plant eukaryotic systems. Auxin is a family of phylohormones that regulates the light-sensitive growth of plants (Gomes & Scortecci, 2021). It triggers the degradation of AID peptide sequence-containing IAA/AUX transcription repressors through recruiting the plant SKP-CUL1-F-box-TIR1 complex (SCF-TIR1) and E3 ubiquitin ligase to polyubiquitinate the target proteins, which then leads to their proteolysis by the proteosome (Mazzucotelli et al., 2006; Tan et al., 2007). In non-plant eukaryotes, both the SCF complex and E3 ligase are conserved as part of the protein degradation machinery, but the TIR1 protein is absent (Nishimura et al., 2009). The adaptation of AID protein degron to endogenous human proteins was made possible with gene editing through CRISPR-Cas9 guided homology-directed repair, where a copy of *Oryzo Sativa* TIR1 (OsTIR1) cDNA is incorporated into the AAVS1 locus on chromosome 19, and a full length or minimal AID peptide cDNA from the plant IAA17 protein is added to the amino (N) or carboxy (C) terminal coding sequence of both genomic alleles of the protein of interest. These modifications lead to expression of TIR1 protein to form SCF-TIR1 complex and addition of AID sequence to the protein of interest, which allows conditional auxin-induced polyubiquitination and depletion of the protein of interest in as short as 30 minutes (Nishimura, et al. 2009).

However, the initial AID system had several downsides. 3-indoleacetic acid (3-IAA), the auxin used to induce the degradation, was found to trigger minor transcriptome changes in human cell culture at its working concentration range of 10 μM to 1 mM, and may be oxidised into toxic derivatives (Nishimura, et al. 2009). Numerous human proteins were reported to suffer from basal degradation in the OsTIR1-IAA17 AID system, potentially as a result of auxin-like molecules in human cell metabolites (Natsume et al., 2016). These draw-backs limited the use of the AID system on dose-sensitive proteins and restricted its use in whole organisms.

Several solutions have been proposed by different research groups, including alternating the TIR1 and IAA17 AID sequences to their paralogues or homologues, introducing natural regulators of the TIR1 protein, and chemical modifications of the auxin molecules to improve the system's specificity and sensitivity (Li et al., 2019; Sathyan et al., 2019; Zhang et al., 2021). In 2020, Prof. Kanemaki's team presented their AID2 sytem, developed using the "bump-and-hole" strategy (Yesbolatova et al., 2020). In brief, they created a "hole" in the OsTIR1 3-IAA binding pocket with the F74G mutation, and derived a high-specificity auxin ligand, 5-Ph-IAA, with a phenyl "bump" to fit the mutated binding pocket. This system promised low levels of basal degradation and lowered the working concentration range of auxin from 1-10 nM, minimizing side effects. This solution was demonstrated as an effective conditional knock-out system not just for cell cultures, but also for whole organism models such as mice (Yesbolatova et al., 2020).

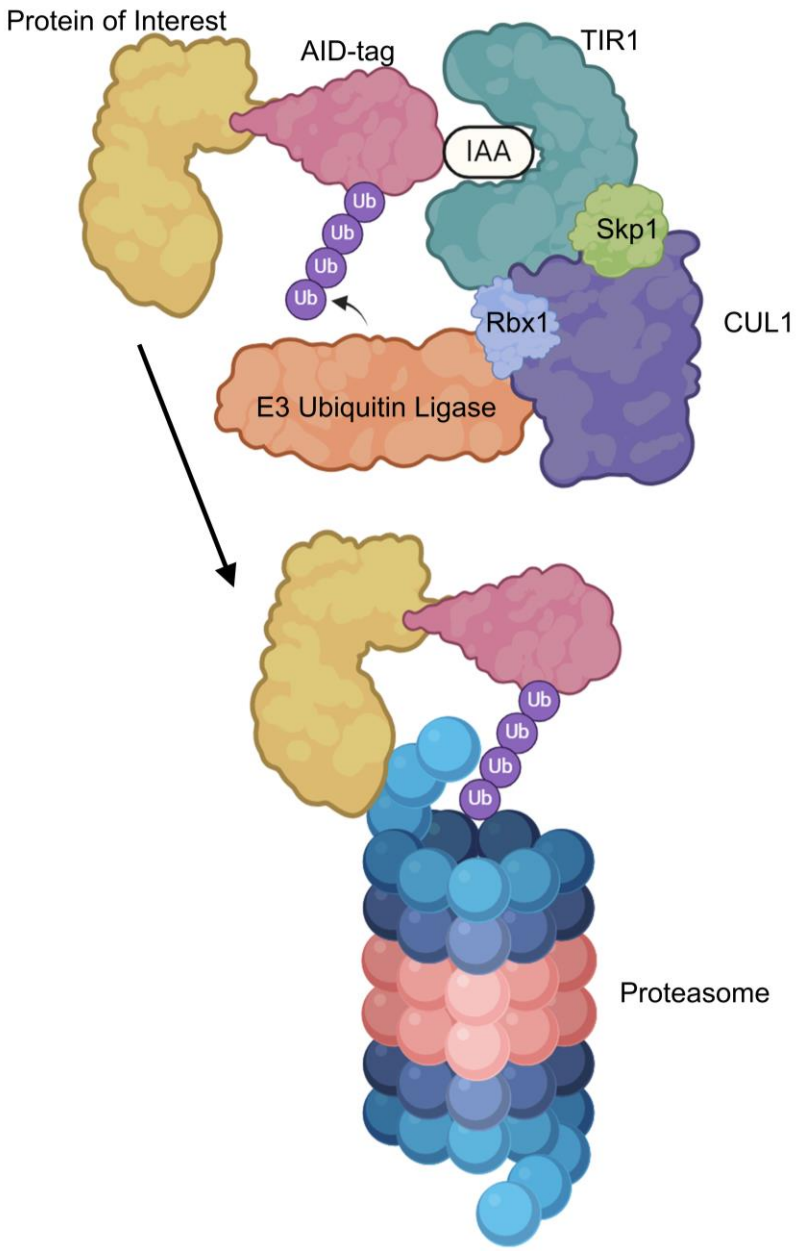

Figure I.5.1. Degradation of AID-tagged proteins

A schematic of the degradation of AID-tagged protein of interest induced by TIR1 recognition. Upon indole acetic acid (IAA or auxin) binding, TIR1 recognises AID-tagged proteins and recruits Skp1-CUL1-Rbx1 (SCF) complex and E3 ubiquitin ligase. E3 ubituiqin ligase polyubiquitinates the AID-tagged protein. The polyubiquitin chain is then recognised by proteasomes, which degrades the AID-tagged protein.

I.6. Aims of this study

Existing evidence in reports and the Wilson Lab indicated that hnRNPU and hnRNPUL1 are proteins with similarly arranged sequence and structural elements with potentially contradicting roles in RNA biogenesis, yet neither of their roles and mechanisms of action have been fully elucidated. We have also observed connections and correlations between their functions and the ALS-associated FET family hnRNP FUS. All three hnRNPs have been associated with human disease and investigating and comparing their cellular functions may provide insight in the roles and mechanisms of the hnRNP regulatory network in hnRNA biogenesis and manifestation of hnRNP-associated disease.

Since all three hnRNPs have been shown to associate with the PolII transcription mechanism and the spliceosome subunits, this study will focus on elucidating their roles in regulating transcription and splicing of RNA PolII transcripts using the AID protein knock-out (KO) system. AID systems for hnRNPU and hnRNPUL1 have been previously established in the Wilson Lab by Dr. Catherine Heath and Dr. Llewelyn Griffith, respectively (Heath, 2018; Griffith, 2019), and a FUS-AID2 system was established in this investigation. We analysed the transcription and co-transcription splicing effects of knocking out hnRNPU, hnRNPUL1 and FUS using transient transcriptome sequencing (TT-Seq).

For hnRNPU, the TT-Seq results were compared with existing fractionation RNA sequencing (RNA-Seq) data to explore possible explanations of its transcription effects. Since hnRNPU was reported to regulate the maturation of U2 snRNP, protein-interaction studies were also used to explore the possible roles of hnRNPU in regulating U2 snRNP activity and its feedback on PolII transcription and splicing. For hnRNPUL1, a northern blot, a dual luciferase assay and the TT-Seq was used to confirm and support existing observations made by Dr. Ivaylo Yonchev and Dr. Carmen Apostol (Yonchev, 2021; Apostol, 2023) and exploring possible correlations between the roles of hnRNPU and hnRNPUL1 in transcription. For FUS, TT-Seq and total RNA-Seq was conducted at an early timepoint (6-hour) of FUS-AID2 depletion to reveal direct effects of its loss. The early transcription and gene expression of FUS were compared to a late-time point (48-hour) total RNA-Seq as well as existing eCLIP data from the ENCODE project to elucidate direct roles of FUS in binding and regulating nascent RNA biogenesis.

II. Material and Methods

II.1. Materials

II.1.1. Cell Lines

Cell Line	Description
HCT116	Human colorectal tumour cell line.
HNRNPU-AID	HCT116 cells constitutively expressing Oryza Sativa TIR1 protein (OsTIR1) and C-terminal FLAG-AID-tagged hnRNPU.
HNRNPUL1-AID	HCT116 cells expressing doxycycline-inducible OsTIR1 and C-terminal FLAG-AID-tagged hnRNPUL1.
T7-HNRNPU-WT/ΔWA/ΔWB	HNRNPU-AID cells expressing T7-tagged hnRNPU wild type (WT)/ Walker A mutant (ΔWA)/ Walker B mutant (ΔWB) proteins.
HCT116-AID2	HCT116 cells constitutively expressing modified OsTIR1(F74G) protein for the AID2 system.
FUS-AID2	HCT116-AID2 cells with C-terminal mAID-tagged FUS.
HNRNPUL1-AID2	HCT116-AID2 cells with C-terminal mAID-tagged hnRNPUL1.
HEK293T	Human embryonic kidney cell line.
Lymphoblastoids	Peripheral B-lymphocytes derived from hnRNPU syndrome patients (HB3261 and HB3280) or control sample (SC3504), transformed by Epstein-Barr Virus (EBV).

II.1.2. Antibiotics

Antibiotic	Stock Concentration	Working Concentration	Solvent
Penicillin	10,000 IU/mL	100 IU/mL	Water
Streptomycin	10,000 µg/mL	100 µg/mL	Water
Hygromycin	50 mg/mL	150 µg/mL	Water
Blasticidin	10 mg/mL	10 µg/mL	Water
Neomycin	50 mg/mL	800 µg/mL	Water
Puromycin	1 mg/mL	1 µg/mL	Water

II.1.3. Protein expression inducing and AID depletion reagents

Reagent	Stock Concentration	Working Concentration	Solvent
3-IAA	250 mM	50 µM	Ethanol
5-Ph-IAA	1 mM	2 nM	Ethanol
Doxycyclin	1 mg/mL	1 µg/mL	Ethanol

II.1.4. Metabolic labelling and Biotinylation reagents

Reagent	Stock Concentration	Solvent
4SU	500 mM	DMSO
4TU	1 M	DMSO

Reagent	Stock Concentration	Solvent
Biotin-HPDP	500 µg/mL	DMF
MTSEA-Biotin-XX	200 µg/mL	DMF

II.1.5. Miscellaneous Buffers and Solutions

Buffer	Solutes	pH	Solvent
10X PBS	1.37 M NaCl 27 mM KCl 43 mM NaH_2PO_4 14.7 mM KH_2PO_4	7.4	Water
5x TBE	4.4 M Tris-HCl 4.4 M Boric Acid 100 mM EDTA, pH 8.0	-	Water

II.1.6. Immunofluorescence Buffers

Buffer	Solutes	pH	Solvent
Fixing solution	1x PBS 4% (w/v) formaldehyde	-	Water
Permeabilisation solution	1x PBS 0.5% (v/v) Triton-X 100	-	Water
Blocking solution	1xPBS 1% (w/v) BSA	-	Water

II.1.7. Protein IP Buffers

Buffer	Solutes	pH	Solvent
IP Lysis Buffer	50 mM HEPES-NaOH, pH 7.5 100 mM NaCl 0.5% (v/v) Triton X-100 1 mM EDTA pH 8.0 10% (v/v) Glycerol 1 mM DTT 1x protease inhibitors	-	Water
Arginine Elution Buffer	1M Arginine-HCl	3.5	Water

II.1.8. Subcellular Fractionation Buffers

Buffer	Solutes	pH	Solvent
Sucrose Wash Buffer	10 mM Tris HCl, pH 8.0 500 mM sucrose 10% (v/v) glycerol 3 mM $CaCl_2$ 2 mM $MgCl_2$	-	Water

Buffer	Solutes	pH	Solvent
Sucrose Lysis Buffer	Sucrose Wash Buffer supplemented with: 0.5% (v/v) Triton-X 100 1 mM DTT 1x protease inhibitors	-	Water
NRB	20 mM HEPES-NaOH, pH 7.5 75 mM NaCl 1 mM DTT 50% (v/v) glycerol	-	Water
NUN	20 mM HEPES-NaOH, pH 7.5 300 mM NaCl 1 M urea 1% (v/v) NP-40 1 mM DTT 1x protease inhibitors	-	Water
Buffer A	10 mM HEPES-NaOH, pH 7.5 10 mM KCl 10% (v/v) glycerol 4 mM $MgCl_2$ 1 mM DTT	-	Water
RIPA	50 mM HEPES-NaOH, pH 7.5 150 mM NaCl 10% (v/v) glycerol 1% (v/v) NP40 0.1% (w/v) SDS 0.5% (w/v) sodium deoxycholate 1 mM DTT	-	Water

II.1.9. SDS-PAGE and Western Blot Buffers

Buffer	Solutes	pH	Solvent
4x SDS-PAGE Loading Buffer	200 mM Tris-HCl, pH 6.8 1% Bromophenol blue 10% (w/v) SDS 50% (v/v) glycerol	-	Water
4x Stacking Gel Buffer	500 mM Tris-HCl, pH 6.8 0.15% SDS	-	Water
4x Resolving Gel Buffer	1.5 M Tris-HCl, pH 8.8 0.15% SDS	-	Water
SDS-PAGE Running Buffer	25 mM Tris 250 mM Glycine 0.1% SDS	-	Water
10x TBST	20 mM Tris-HCl, pH 7.6 137 mM NaCl 2% (v/v) Tween-20	-	Water
Blocking Solution	1x TBST 5% (w/v) Skimmed milk powder	-	Water

II.1.10. Northern Blot Buffers

Buffer	Solutes	pH	Solvent
10x SSPE	1.5 M NaCl 90 mM NaH_2PO_4 10 mM EDTA, pH 8.0	7.4	Water
50 x Denhardt's Solution	1% (w/v) Ficoll 1% (w/v) Polyvinylpyrrolidone 1% (w/v) bovine serum albumin	-	Water
Hybridization Buffer	6 x SSPE 5 x Denhardt's Salt 0.2% (w/v) SDS	-	Water
Stripping buffer	0.1 x SSPE 0.1% SDS (w/v)	-	Water
2x Sample Loading Buffer	20 mM EDTA, pH 8.0 0.05% (w/v) Bromophenol Blue 0.05% (w/v) Xylene Cyanol	-	Formamide
12% Urea-PAGE gel	0.5 x TBE 8M Urea 12% Accugel Acrylamide 19:1 2% TEMED 0.1% (w/v) ammonium persulfate	-	Water

II.1.11. 4SU-Pulse Chase Buffers

Buffer	Solutes	pH	Solvent
10 x TE Buffer	10mM EDTA 100mM Tris-HCl pH 7.5	-	Water
1 x MPG	1 x TE Buffer supplemented with: 1 M NaCl 0.1% IGEPAL (v/v)	-	Water
1 x MPG w/o NaCl	1 x TE Buffer supplemented with: 0.1% IGEPAL (v/v)	-	Water
0.1x MPG	0.1 x TE Buffer supplemented with: 100mM NaCl 0.1% IGEPAL (v/v)	-	Water
0.1x MPG w/o IGEPAL	0.1 x TE Buffer supplemented with: 100mM NaCl	-	Water
0.1 x MPG with 0.5% β-Met	0.1 x MPG without IGEPAL 5% β-Mercaptoethanol (v/v)	-	Water
10 x Pulse Chase Biotinylation Buffer	200 mM NaOAc 1% SDS 10 mM EDTA	-	Water

II.1.12. TT_{chem}-Seq buffers

Buffer	Solutes	pH	Solvent
Biotinylation Buffer	833 mM Tri-HCL, pH 7.4 83 mM EDTA, pH 8.0	-	Water
Pull-down Wash Buffer	100 mM Tris-HCl, pH 7.4 10 mM EDTA, pH 8.0 1 M NaCl 0.1% Tween-20 (v/v)	-	Water
Elution Buffer	100mM DTT	-	Water

II.1.13. Primary antibodies

Target	Source	Origin	Clonality
HIV-Tat-SF1	Santa Cruz (sc-514351)	Mouse	Monoclonal
Ssrp1	Biolegend (10D1)	Mouse	Monoclonal
Histone H3	Abcam (ab1791)	Rabbit	Polyclonal
Tubulin	Sigma (T5168)	Mouse	Monoclonal
hnRNPU	Abcam (3G6)	Mouse	Monoclonal
hnRNPUL1	In house	Rabbit	Polyclonal
Total Pol II	MBL Life Science (MABI0601)	Mouse	Monoclonal
Ser2P Pol II	MBL Life Science (MABI0602)	Mouse	Monoclonal
Ser5P Pol II	MBL Life Science (MABI0603)	Mouse	Monoclonal
U2A′	Santa Cruz (sc-393804)	Mouse	Monoclonal
SF3B1	Santa Cruz (sc-514655)	Mouse	Monoclonal
LARP7	Bethyl (A303-723A)	Rabbit	Polyclonal
SF3B3	Bethyl (A302-508A)	Rabbit	Polyclonal
NR4A1	Novus Bio (NB100-56745)	Rabbit	Polyclonal
EGR1	ProteinTech (22008-1-AP)	Rabbit	Polyclonal
CDK9	Abcam (ab76320)	Rabbit	Polyclonal
CDK7	Abcam (ab137716)	Rabbit	Polyclonal
FUS	Santa Cruz (sc-47711)	Mouse	Monoclonal
Coilin	Bethyl (A303-760A)	Rabbit	Polyclonal
U1A	Abcam (ab155054)	Rabbit	Polyclonal
SF3B3	Bethyl (A302-508A)	Rabbit	Polyclonal
FLAG	Sigma (F3165)	Mouse	Monoclonal
OsTIR1	MBL Life Science (PD048)	Rabbit	Polyclonal
T7-tag	Abcam (ab9138)	Goat	Polyclonal

II.1.14. Secondary Antibodies

Target and modifications	Source	Origin	Clonality
Anti-mouse IgG, HRP Conjugate	Promega (W4021)	Goat	Monoclonal
Anti-rabbit IgG, HRP Conjugate	Promega (W4011)	Goat	Monoclonal
Anti-mouse IgG, Alexa 555	Invitrogen (A-21422)	Goat	Polyclonal
Anti-rabbit IgG, Alexa 488	Invitrogen (A-11088)	Goat	Polyclonal
Anti-goat IgG, HRP Conjugate	Abcam (ab6741)	Rabbit	Polyclonal

II.1.15. Plasmids

Plasmid	Source
5xGAL4-TATA-luciferase	Addgene 46756
CDK1as_T2A_Zeo	Addgene 118596
Gal4-VP16	Addgene 71728
pCDNA5.0-3xFLAG-hnRNPU_WA	Wilson Lab
pCDNA5.0-3xFLAG-hnRNPU_WT	Wilson Lab
pCDNA5-3xFLAG-FUS	Wilson Lab
pCDNA5-3xFLAG-GFP	Wilson Lab
pCDNA5-3xFLAG-hnRNPU_WB	Wilson Lab
pCDNA5-3xFLAG-hnRNPUL1_WT	Wilson Lab
pCMV26-3xFLAG-GAL4-FUS_IDR	This Thesis
pCMV26-3xFLAG-GAL4-hnRNPU_SPRY-RGG	This Thesis
pCMV26-3xFLAG-GAL4-hnRNPUL1_IDR	This Thesis
pCMV26-3xFLAG-GAL4-hnRNPUL1_PPNK-IDR	This Thesis
pCMV26-3xFLAG-GAL4-hnRNPUL1_RGG-IDR	This Thesis
pCMV26-3xFLAG-GAL4-hnRNPUL1_SPRY-IDR	This Thesis
pCMV26-3xFLAG-GAL4-VP16	This Thesis
pCMV-T7-SB100	Addgene 34879
pGL4.75[hRluc/CMV]	Promega E6931
pMGS46	Addgene 126580
pMK287	Addgene 72825
pMK381	Addgene 140536
pUC18-FUS-AID-HygR	This Thesis
pUC18-FUS-AID-NeoR	This Thesis
pUC18-FUS-mAID-HygR	This Thesis
pUC18-FUS-mAID-NeoR	This Thesis
pUC18-hnRNPU-AID-HygR	Wilson Lab
pUC18-hnRNPUL1-AID-HygR	Wilson Lab
pUC18-hnRNPUL1-AID-NeoR	This Thesis
pUC18-hnRNPUL1-mAID-HygR	This Thesis
pUC18-hnRNPUL1-mAID-NeoR	This Thesis
px330-sgAAVS1	Addgene 85802
px330-sgFUS-C	This Thesis
Px330-sgHNRNPUL1	Wilson Lab
pX330-U6-Chimeric_BB-CBh-hSpCas9	Addgene 42230
SB-T7-hnRNPU_ΔWA-T2A-ZeoR	This Thesis
SB-T7-hnRNPU_ΔWB-T2A-ZeoR	This Thesis
SB-T7-hnRNPU_WT-T2A-ZeoR	This Thesis

II.1.16. Cloning PCR Primers

Target	Forward primer Sequence	Reverse primer Sequence
mAID for FUS HDR	ACAGGATCGCAGGGAGAGGCCGTATAAGGAGAAGAGTGCTTGTCCTAAAG	GGGAAAAGTTAGTGGCCCCTGATCCTTTATACATCCTCAAATCGATTTTCCTCAAG
mAID for hnRNPUL1 HDR	AGGGTGGCACAAGTACACAGAAGGAGAAGAGTGCTTGTCCTAAAG	
AID for FUS HDR	ATCGCAGGGAGAGGCCGTATGGGAGCGGTAGTGGCATGATGGG	CTGGGAAGCCAGGCTAATTATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATG
FUS LHA	GCGTATCACGAGGCCCTTTCGTCGGCCACACTGTTGGGGTCAGATTTAGCC	CCCATCATGCCACTACCGCTCCCATACGGCCTCTCCCTGCGATCCTGTCTGTG
FUS RHA	GGTTTGTCCAAACTCATCAATGTATCTTATAATTAGCCTGGCTTCCCAGGTTCTGGAACAG	GTCATCACCGAAACGCGCGATAACTCATTGGCTGCAACCCCTTCCCC
pUC18	TCGCGCGTTTCGGTGATGAC	GACGAAAGGGCCTCGTGATACGC
FUS_HDR-P2A-drugR backbone	GGATCAGGGGCCACTAACTTTTCCC	ATACGGCCTCTCCCTGCGATCCTGTCTGTG
hnRNPUL1 HDR-P2A-drugR backbone		CTGTGTACTTGTGCCACCCT
FUS-N’ IDR	GCCTCAAACGATTATACCCAACAAGCA	ACCAAATTTATTGAAGCCACCACG
hnRNPU SPRY-RGG	GAAGAGAACAAGTATAGCAGAGCC	TCAATAATATCCTTGGTGATAATGCTGAC
hnRNPUL1 SPRY-IDR	GAGGATAGGAGGGGCCGCTCTCCT	CTACTGTGTACTTGTGCCACCCTG
hnRNPUL1 PPNK-IDR	CCCTACTGTTCTGTCCTCCCGGGG	
hnRNPUL1 RGG-IDR	GAAAAGCGCTTTGACAACCGAGGT	
hnRNPUL1 IDR	TTCAACCGCAGCGGAGGTGGTGGC	
FLAG-Gal4	AAGGATGACGATGACAAGCTTGCAATGAAGCTACTGTCTTCTATCGAA	CGATACAGTCAACTGTCTTTGACC
Gal4-DBD backbone	TAGGGGGCGCGACCGGACCC	GGTCAAAGACAGTTGACTGTATCG
FUS sgRNA	CACCGAAAAGCTGTTCCAGAACCTG	AAACCAGGTTCTGGAACAGCTTTTC

Target	Forward primer Sequence	Reverse primer Sequence
T7-hnRNPU	AAACAGCCATGGCTAGCATGACTGGTGGACAGCAAATGGGTATGAGTTCCTCGCCTGTTAATG	CTCCACGTCGCCGCAGGTCAGCAGGCTGCCGCGGCCCTCATAATATCCTTGGTGATAATGCTG
SB-T2A-ZeoR vector	CAGCATTATCACCAAGGATATTATGAGGGCCGCGGCAGCCTGCTGACCTGCGGCGACGTGGAG	GGCTGTTTCCGGAATTCCAGCACACTGG

II.1.17. qPCR Primers

Target	Forward primer Sequence	Reverse primer Sequence
18S rRNA	GTGGAGCGATTTGTCTGGTT	CGGACATCTAAGGGCATCAC
MALAT1	CTAGGACTGAGGAGCAAGCG	CTCGCTCCTTCCTGGAATCC
NEAT1_2	TTAGAATGGTTATCTGAGGAAGTGGC	CAGATCGGATTTGACCAACAAAATGG
FIRRE	CTGGCAGCAGAGACTAAGGT	GTGTTTGCAAGCCAGGTACA
HIST1H1C	AGTGCTGCTAAGGCTGTGAA	TTCGCCTATTTCTTCTTGGG
HIST2H2BE	GTACACCAGCTCCAAGTGAGTC	AAGAGCCTTTGGAGTCAAG
HIST1H3H	CAGGACTTCAAGACCGAC	ATGATAGTCACCCGCTTG
HIST2H2AC	CGGCGTCTTGCCTAACAT	TTTGCCTTGTGGTGACTCTC
U5 snRNA	TCTGGTTTCTCTTCAGATCGCA	GAGTTGTTCCTCTCCACGGA
MAT2A Exon-Intron	AATGCCAAACTGGCAGAACT	TGGATCTTCTGTGCTGATGTAA
MAT2A Exon-Exon		CTCTGATGGGAAGCACAGC
hnRNPH3 Exon-Intron	GGTTGCAGCAAAGAGGAAAT	CCAAAATCAAATGGGTTTTACA
hnRNPH3 Exon-Exon		CTGTGCTTCTCCCCTGGTAGTCC
SAT1 Exon-Intron	CCTTTTACCACTGCCTGGTT	TTCGGCCTGTGTAGTCAGTG
SAT1 Exon-Exon		CAATCCACGGGTCATAGGTA

II.1.18. Northern Blot Probes

Target	Sequence
U6 snRNA	AAAATATGGAACGCTTCACGAA
U7 snRNA	AAAGCCTACTAGACAAATTCTAAAA

II.2. Cloning and cell line generation

II.2.1. Plasmid cloning

Apart from px330-based Cas9-single guide RNA (sgRNA) plasmids, all plasmid construction in this project used Gibson assembly with NEBuilder® HiFi DNA Assembly Master Mix (NEB Cat. # E2621L) with DNA fragments amplified using the polymerase chain reaction (PCR) and Q5® High-Fidelity DNA Polymerase (NEB Cat. # M0491L) as per manufacturer's instructions. Px330-based sgRNA plasmids were constructed by ligating BbsI digested pX330-U6-Chimeric_BB-CBh-hSpCas9 vector (Addgene #42230) with annealed double-stranded DNA oligonucleotides encoding the targeting sgRNA. Cloned plasmids were transformed into chemically competent DH5α *E. coli* and plated on antibiotic agar plates for selection. Viable *E. coli* colonies were tested by colony PCR where applicable, and positive colonies were cultured and extracted with a Miniprep plasmid kit (Qiagen Cat. #27104) for the Sanger sequencing service at Source Bioscience. Colonies carrying plasmids of correct sequences were then cultured for Midi plasmid extraction (Qiagen Cat. #12945) and downstream transfections.

II.2.2. Transfection methods

Chemical transfection mix was prepared by vortex mixing 500 to 1000 ng plasmid DNA per ml of cell culture media with 3 x DNA mass of polyethyleneimine (PEI) in 10% cell culture media volume of warm, un-supplemented DMEM (Merck Cat. #D0822). The transfection mix was incubated at room temperature for 10 minutes before evenly dispensing into cell culture by pipetting and swirling the cell culture simultaneously. Cells were typically transfected for 48 to 72 hours before harvesting for analysis or reseeding into antibiotic-supplemented media for gene-editing selections.

Electroporation transfection was conducted using the NEON electroporation system (Invitrogen Cat. #MPK5000) per manufacturer's instructions and data sheets.

II.2.3. Constructing FLAG-Gal4 DBD fused plasmids

DNA fragments encoding the peptides of interest (FUS N-terminal IDR, hnRNPU

SPRY-RGG and various hnRNPUL1 truncations per described in Chapter IV.2) were PCR amplified from existing plasmids and inserted into the CMV-driven Gal4 DNA-binding Domain (DBD) backbone vector, which was PCR amplified from Gal4-VP16 plasmid (Addgene #71728). The Gal4-fused constructs were then migrated into pCMV26-3xFLAG backbone by PCR amplification and Gibson Assembly.

II.2.4. Establishing hnRNPU-AID-T7-HNRNPU wild type / Walker A mutant / Walker B mutant cells

T2A-Zeocin resistance (T2A-ZeoR) backbone from CDK1as_T2A_Zeo (Addgene #118596) and existing hnRNPU WT, WA, and WB cDNAs in pCDNA5 vectors were amplified and assembled using NEBuilder Gibson Assembly Mix. T7 tag was incorporated into the 5′ forward primer of hnRNPU cDNAs and the 3′ reverse primer of the T2A-ZeoR backbone. This creates seamless T7-hnRNPU (WT/WA/WB) connected by a T2A to a zeocin resistance gene and together flanked by Sleeping Beauty transposon sequences (Figure III.10.1A).

These transposon plasmids were each co-transfected with a vector expressing Sleeping Beauty transposase (pCMV-T7-SB100, Addgene #34879) using the NEON electroporation system into hnRNPU-AID cells in a 6-well plate at a molar ratio of 5 transposon vector to 1 transposase vector. 72 hours after transfection, the cells were transferred into 15cm dishes. The cells were initially selected with 300μg/ml zeocin for three days, then refreshed with maintenance-level 100μg/ml zeocin for three weeks. Healthy colonies, which do not appear to contain swollen or deformed cells under a light microscope, were picked with cloning discs and grown in 24 wells for RIPA-benzonase lysis and verified for T7-hnRNPU expression by Western blotting with both anti-T7 and anti-hnRNPU antibodies. The colony with the closest level of hnRNPU expression to wild type HCT116 hnRNPU levels was chosen for each cell line.

II.2.5. Establishing the HCT116-AID2 system

pMK381 (AAVS1 CMV-OsTIR1F74G, Addgene #140536) was co-transfected chemically using polyethyleneimine (PEI) with px330-sgAAVS1 (Addgene #85802) into HCT116 cells at a molar ratio of 2: 1 in 6-well plates. The cells were cultured for 72 hours before reseeding each well into a 10cm dish containing puromycin for selection. After two weeks, six viable colonies were picked with cloning discs into 24 well plates for assaying with an anti-OsTIR1 antibody in Western blots. Only two

colonies expressed TIR1(F74G), colony 3 expressed higher levels of TIR1 than colony 2 and was maintained and kept as HCT116-AID2 cells (Figure IV.6.1A).

II.2.6. Constructing hnRNPUL-mAID tagging plasmids

The guiding RNA plasmid (px330-sgHNRNPUL1) for the hnRNPUL1-AID2 system was constructed by Dr. Llewllyn Griffith when he established the hnRNPUL1-AID system (Griffith, 2019). The homology-directed repair (HDR) template plasmids of tagging hnRNPUL1 with mini-AID (mAID) were generated by Gibson assembly of homology arm backbones containing hygromycin (Hyg) and neomycin (Neo) resistance, which was PCR amplified from existing hnRNPUL1-AID HDR templates made by Dr. Griffith, and the mAID sequence, amplified from pMK287 (Addgene #72825).

II.2.7. Constructing FUS-AID and FUS-mAID tagging plasmids

For the guiding the genome editing in the FUS improved AID and AID2 systems, a single-guide RNA for Cas9 is designed in Benchling to target as close to the STOP codon in FUS genomic DNA as possible, then purchased as two strands of oligonucleotides to be annealed and ligated into a BbsI restriction digested px330 vector to obtain the px330-sgFUS-C plasmid.

To tag both alleles of FUS with full-length AID or mAID, around 1000bps up and downstream homology DNA sequences flanking FUS STOP codon (hereafter referred to as “FUS homology arms”) were amplified from HCT116 genomic DNA. pUC18 backbone and AID-P2A-Hygromycin/Neomycin DNA sequences were amplified from existing hnRNPU-AID tagging homology repair plasmids. pUC18, AID-P2A-resistance, and homology arms sequences were first assembled into pUC18-FUS-AID-P2A-resistance homology repair plasmid using Gibson Assembly. From this vector, pUC18-FUS homology arms-P2A-resistance backbone was amplified with PCR and assembled with mAID sequence using Gibson Assembly again to obtain pUC18-FUS-mAID-P2A-Hygromycin/Neomycin resistance HDR template plasmids.

II.2.8. Generating mAID-tagged AID2 cell lines for FUS and hnRNPUL1

The corresponding px330-guide RNA plasmid was co-transfected with both pUC18-Homology-mAID-P2A-Hyg and Neo resistance HDR template plasmids at a molar ratio of 1:2:2, using PEI, into HCT116-AID cells in 6 well plates (Figure IV.6.1B). Transfected cells are cultured for 72 hours and reseeded into media containing both

hygromycin and neomycin for selection. After three weeks of selection, only two colonies survived. They were picked into 24 wells for culturing. Then the colonies were treated with either DMSO or 2nM 5-Ph-IAA and lysed in RIPA buffer for Western blot alongside reference HCT116 lysate to verify knock out of hnRNPUL1 or FUS.

II.3. Laboratory Methods

If not specified otherwise, all procedures were conducted at room temperature, natural humidity and sea-level air pressure under conventional in-doors lighting.

II.3.1. Whole Cell Protein extraction

Adherent cells were washed with PBS and lysed directly on the culture surface with RIPA lysis buffer for 10 minutes on ice. Then the cells were detached by shaking, collected into a 1.5 ml microfuge tube and homogenized by pipetting 20 times. Suspension cells were collected into a 1.5 ml microfuge tube, pelleted by centrifugation at 400 x g for 3 minutes and lysed by resuspension in RIPA lysis buffer and homogenized by pipetting 20 times. Homogenised cells were treated with 250 U/ml benzonase (Merck, Cat. #E1014-25KU) for 30 minutes at room temperature, centrifuged at 17,000 x g for 10 minutes at 4℃ and the supernatant was collected as the whole cell extract (WCL).

II.3.2. Subcellular Fractionation

For each condition per experiment, 2 x 15-cm dishes of 70% to 90% confluent cells were harvested with trypsin and collected into a 15 ml falcon tube. The cells were pelleted by centrifuging at 300 x g for 3 minutes, then rinsed with PBS and drained, and the pellet volume was estimated and noted as V_{cell}. The cell pellet was resuspended by inversion in 7 x V_{cell} of cold Sucrose lysis buffer supplemented with protease inhibitor cocktail, 1 μM DTT and 1:400 (v/v) Ribosafe RNase inhibitor on ice for 5 minutes. The lysed cells were centrifuged at 500 x g for 5 minutes at 4℃ and the supernatant was kept as the cytoplasmic fraction. The pelleted cell nuclei were washed twice with 1 ml of cold Sucrose wash buffer by gentle resuspension and centrifuged at 400 x g for 3 minutes. Then the nuclei pellet was gently resuspended in 3 x V_{cell} of cold NRB buffer and lysed by mixing in 3 x V_{cell} of cold NUN buffer and incubating on ice for 5 minutes with one inversion every minute. The extracted cell nuclei were centrifuged at 1,200 x g for 5 minutes at 4℃ and the supernatant was kept as the nucleoplasm fraction. The pellet containing chromatin was washed with 1 ml of Buffer

A and centrifuged at 1,200 x g for 5 minutes at 4℃. To extract chromatin-fraction proteins, the chromatin pellet was resuspended in 1 x V_{cell} of RIPA lysis buffer, treated with 2500 U/ml benzonase for 45 minutes at room temperature, centrifuged at 17,000 x g for 10 minutes at 4℃ and the supernatant was kept as the chromatin protein extract. To extract chromatin-bound RNA, the chromatin pellet was resuspended in 1 x V_{cell} of buffer A and 3 x V_{cell} of TRI-reagent LS (Merck Cat. #T3934) and denatured by rotating at 1000 rpm at 70℃ for 15 minutes before carrying out TRI-based RNA extraction as described in II.3.6 below.

II.3.3. Western Blot

Protein extracts were denatured in 1x SDS sample loading buffer at 95℃ for 5 minutes and resolved in 10% stacked SDS-PAGE mini-gel at 20 mA per gel with pre-stained protein ladder (ThermoFisher Cat. #26619). Resolved proteins were transferred to nitrocellulose membrane (SERVA Cat. #71224.01) in Trans-Blot Turbo transfer device (Bio-Rad Cat. #1704150) in 10 ml of 1x SDS-PAGE transfer buffer at 1 A for 45 minutes per gel. The membrane was then washed in water, stained with Ponceau stain, cut by protein ladder sizes to strips containing each protein of interest, destained with TBS and blocked in TBS with 5% skimmed milk overnight. The blocked membrane strips were incubated with primary antibodies diluted in TBST milk (1x TBS with 0.2% Tween-20 and 5% skimmed milk) for 1 hour, washed with TBST for 5 minutes and incubated with secondary antibodies diluted in TBST milk for 1 hour. Finally, the membrane strips were washed with TBST for 5 minutes, rinsed with distilled water, incubated in ECL reagents for 30 seconds, and reassembled on a ChemiDoc imaging system (Bio-Rad Cat. #12009077) for auto-optimal visualisation.

II.3.4. Co-immunoprecipitation (Co-IP)

For each condition, 40 µl Protein G magnetic beads were washed in IP lysis buffer and blocked overnight by rotating in 300µl IP lysis buffer containing 1% BSA with 5µg of antibody at 4℃. 2 to 4 15-cm dishes of 70% to 90% confluent cells were rinsed with PBS and harvested with a cell scrapper. Cells were pelleted by centrifuge at 400 x g for 3 minutes, drained, and lysed with 3 x estimated cell volume cold IP lysis buffer, which was supplemented with 1 x protease inhibitors cocktail, 1µM DTT and other supplements as indicated per experiment. Cell nuclei were sheared by syringing lysed cells 10 times through a 0.6-mm aperture sterile needle. Cell lysate concentration was

determined with Bradford assay and were adjusted and equalized between conditions to 10 to 20 mg proteins per ml with IP lysis buffer. Crude cell lysate may be subject to DNase, RNase or Benzonase treatments per experiment, and were cleared up in a 5-minute 17,000 x g centrifuge at 4℃. A small aliquot of lysate supernatant was diluted 10-fold in IP lysis buffer as a 10% Input sample. Antibody-bound beads were washed with 1 ml IP lysis buffer on magnetic tube rack, and 1oaded with an equal volume of cell lysate supernatant between conditions. Cell lysate was immunoprecipitated by rotating for 2 hours at 4℃. The magnetic beads were washed twice with 1 ml cold IP lysis buffer on a magnetic tube rack and eluted by incubating for 3 minutes in 55 µl Arginine Elution Buffer and neutralized with 5 µl pH 8.8 Tris-HCl. Input protein extracts and Co-IP eluents were analyzed with Western Blot to identify enrichment in co-immunoprecipitation.

II.3.5. Immunofluorescence and foci counting

24-wells of 10% to 30% confluent cells seeded on coverslips were washed with warm PBS, fixed in 500 µl fixing solution for 30 minutes, washed three times with PBS, permeabilised with 500 µl permeabilization solution for 5 minutes and blocked with 500 µl blocking solution for 1 hour at 37℃. Blocking solution was removed, and each well was incubated with 0.2 to 1 µg primary antibody diluted in 300 µl blocking solution for 1 hour at 37℃ or overnight at 4℃. Then, the coverslips were washed three times with PBS and incubated with 0.5 µg of fluorescently labelled secondary antibodies diluted in 300 µl of blocking solution per well for 1 hour at 37℃. The secondary antibody was removed and the coverslip was washed three times with PBS. Excess PBS was removed by tapping the coverslips on paper towel, and coverslips were mounted downwards onto droplets of DAPI-supplemented anti-fade solution on glass slides. Lastly, the coverslips were air dried in dark for 10 minutes and sealed around the edges with nail varnish. The fluorescence signals were visualised and captured on a wide field fluorescence microscope with 10 x ocular lenses and a 40 x air or 60 x oil-immersion objective lens. Foci signals, such as that of coilin, were counted for each cell in ImageJ.

II.3.6. RNA Isolation

Cell pellet or cell extract were homogenised in TRI-reagents (Merck Cat. #T9424) and chloroform extracted and ethanol precipitated with glycogen into a crude RNA pellet as per manufacturer's instructions. The crude RNA pellet was resuspended in RNase-

free H_2O and treated with 80 U/ml Turbo DNase (ThermoFisher Cat. #AM2239) and 1:50 Ribosafe RNase inhibitor (Bioline Cat. #BIO-65028) for 45 minutes at 37°C. The DNase-digested RNA was extracted with pH 4.5 acidic phenol-chloroform, and precipitated with 70% ethanol, 100 µM NaOAc and 5 µg of glycogen at -20°C for 2 to 16 hours. The precipitated RNA was pelleted by centrifuging at 17,000 x g for 30 minutes at 4°C, washed with 80% ethanol and air-dried. The dried RNA pellet was resuspended in RNase-free H_2O for downstream analysis such as RT-qPCR or Northern blot.

II.3.7. Reverse transcription-quantitative PCR (RT-qPCR)

0.2 to 1 µg of RNA resuspended in H_2O was reverse transcribed using a High-Capacity cDNA Reverse Transcription Kit (ThermoFisher Cat. #4368814). The cDNA was diluted 10-fold with nuclease-free H_2O for qPCR quantification of non-ribosomal RNA targets and was further diluted 100-fold to quantify ribosomal RNA (rRNA). The qPCR was performed in duplicate using SensiMix™ SYBR® Hi-ROX Kit (Meridian Bioscience Cat. #QT605-05), and the take-off cycles (Ct) were quantified on a Rotor-Gene Q qPCR device (Qiagen Cat. #9001620). Comparative cDNA concentrations or difference in take-off cycles (ΔCt) were calculated in the Rotor-Gene Q software and normalized to control conditions according to the experiment.

II.3.8. Northern blot

2 µg of RNA per sample was resolved on a 10-well 8M Urea-8% polyacrylamide denaturing mini-gel cast with 1 x TBE, and transferred onto an N+ Hybond membrane at 0.2 A for 2 hours in 0.5 x TBE. The Hybond membrane was air-dried, cross-linked under 0.2 joules of UV and blocked in 20 ml Hybridization Buffer for 1 hour at 37°C. 5 pmole of oligonucleotide probe was labelled with 2 µCi of γ-^{32}P ATP (PerkinElmer Cat. # BLU502Z250UC) using polynucleotide kinase (PNK) for 30 minutes at 37 °C. The labelled probe was diluted with 1 ml of Hybridization Buffer, denatured for 5 minutes at 65 °C, and topped up to 5 ml in a syringe with Hybridization Buffer. The labelled probe was filtered through a 20 µm filter and added to the blocked membrane, and the membrane was hybridised for 2 to 16 hours at 37 °C. After hybridization, the probe mix was decanted, and the membrane was rinsed three times with 6 x SSPE and washed in 6 x SSPE for 1 hour at 37 °C. The membrane was dried and exposed to a blanked phosphorimaging screen for 3 to 16 hours, and the phosphorimgaing screen was visualised in a GE Typhoon Phosphorimager. The membrane was stripped by

incubating with Stripping Buffer for 10 minutes at 37 ℃, blocked again and re-probing for a different target.

II.3.9. 4-Thiouridine Pulse-Chase

4 thiouridine (4SU) pulse chase was performed based on the procedure and timing of Bresson et al. (2015) with modifications based on optimisations suggested by Rädle et al. (2013) and Schwalb et al. (2016). In the dark, 2 x 15-cm dishes of 50%-70% cells were labelled with 5 μM of 4SU for 1 hour, rinsed with warm PBS, resupplied with fresh media and incubated for 1 hour to wash out 4SU. The media was refreshed again, and cells were lysed with cold TRI reagent on-plate at 0, 30, 60 and 120 minutes after a second media refresh. RNA was extracted from homogenized TRI cell lysate as described in II.3.6 above, and the purified RNA was resuspended in ddH_2O. 50 μg of RNA was labelled with 0.2 mg/ml Biotin-HPDP in 1 x Pulse-Chase Biotinylation buffer with 50% N,N-dimethylformamide (DMF) by rotating at 700 rpm for 3 hours at 25℃. The biotinylated RNA was captured using μMAC magnetic streptavidin beads (Miltenyi Biotec Cat. #130-091-287) and washed sequentially with 500 μl of 0.1 x MPG, 0.1 x MPG without IGEPAL preheated at 55℃, 1 x MPG twice, 0.1 x MPG, 0.1 x MPG without NaCl, and finally 0.1 x MPG. Captured RNA was eluted twice with 200 μl of 65℃ preheated 0.1 x MPG with 0.5% β-Met. The biotinylated RNA was sequentially acidic phenol chloroform extracted twice, ethanol precipitated, washed with 80% ethanol twice and air-dried. The RNA pellet was resuspended in 15 μl H_2O, quantified using Qubit RNA HS kit (ThermoFisher Cat. # Q32852), and analyzed with RT-qPCR.

II.3.10. Transient transcriptome sequencing (TT-Seq)

TT-Seqs were performed primarily based on the TT_{chem}-Seq protocol developed by Gregersen et al. (2020), with modifications based on the original TT-Seq protocol developed by Schwalb et al. (2016) and in-house optimisations. In brief, the following procedures were performed in dark for each condition: two 15-cm dishes of 50% to 70% confluent cells were labelled with 500 nM 4SU for 5 minutes, rinsed with warm PBS twice, lysed and homogenized on-plate with 1.5 ml cold TRI-reagent. Each plate of cell lysate was collected into a pre-chilled 2 ml microfuge tube. The TRI-reagent-RNA mix was chloroform-extracted and ethanol precipitated. Each tube of RNA pellet was resuspended in 200 μl H_2O, combined into one tube and acidic phenol extracted using Phase Lock Gel Heavy (VWR Cat. #733-2478) tubes, then ethanol precipitated again. After resuspending in 200 μl H_2O, RNA concentration was measured using Qubit RNA

BR Assay Kit (ThermoFisher Cat. # Q10210). For each sample, 100 µg of isolated 4SU-labelled total RNA was spiked in with 2µg of 5-minute 1 mM 4TU labeled BY4712 Saccharomyces *cerevisiae* total RNA and topped up to 100 µl with H_2O. The RNA mix was fragmented by the addition of 20µl of 1M NaOH and incubation on ice for 30 minutes, followed by quenching with 180 µl of 1M pH6.8 Tris-HCl and immediately purified with acidic phenol-chloroform extraction and ethanol precipitation. Fragmented RNA was resuspended in 196 µl H_2O, and biotinylated with 100 µg/ml Biotin-MTSEA XX (Biotium Cat. #90066) in 1% Biotinylation buffer with 50% DMF for 30 minutes at 25 ℃. The biotinylated RNA was sequentially acidic phenol-extracted three times, and ethanol-precipitated followed by two 80% ethanol washes. The biotinylation was verified using streptavidin-HRP conjugate dot-blot. Then, the biotinylated RNA was isolated using µMAC magnetic streptavidin beads (Miltenyi Biotec Cat. #130-091-287) and washed twice with 55 ℃ Pull-down Wash Buffer. The captured RNA was eluted twice with freshly constituted Elution Buffer, acidic phenol-chloroform purified, ethanol precipitated, resuspended in 20 µl H_2O and quantified using Qubit RNA HS kit (ThermoFisher Cat. # Q32852). The RNA fragment size was quality checked on a TapeStation and prepared into TT-Seq libraries using NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB Cat. #E7760). The library was multiplexed and sent for paired-end 150 bases sequencing service by Novogene at 120M reads per library. After sequencing, Novogene conducted basic contamination and sequencing quality controls, and provided demultiplexed sequencing data in the fastq format.

II.3.11. Dual-Luciferase Assay

FLAG-Gal4-fused constructs were co-transfected with a plasmid carrying Gal4-specific promoter driven firefly luciferase (5xGAL4-TATA-luciferase, Addgene #46756) and a plasmid carrying CMV promoter driven human codon-optimised Renilla luciferase (pGL4.75, Promega Cat. #E6931). The quantity of plasmids used were optimised to equalise test protein levels determined by Western blots. A dual luciferase assay was conducted using the Dual-Glo Luciferase Assay Kit (Promega Cat. #E2920) as per manufacturer's instructions. Firefly luciferase reporter activities were normalized to the co-transfected Renilla luciferase activity, and the comparative transcription activation potential of each Gal4-fused construct was calculated as the relative firefly luciferase activity ratio against the Gal4-DBD negative control construct.

II.4. Bioinformatic Analysis

II.4.1. Packages and Environment

Package	Language/Shell	Version
bedops	Bash	2.4.39
bedtools	Bash	2.30.0
deeptools	Bash	3.4.3
gawk	Bash	5.1.0
samtools	Bash	1.15.1
STAR	Bash	2.7.10
subread	Bash	2.0.1
ucsc-bedgraphtobigwig	Bash	377
umi_tools	Bash	1.1.2
pandas	python	1.1
DESeq2	R	1.34.0
DEXSeq	R	1.40.0
R	R	4.1.2
Tidyverse	R	1.3.0
IGV	Windows	2.16.2

II.4.2. Scripts

Script	Github Address
Bam2bedgraph2bigwig	*https://github.com/AngLi9849/RNASeqSnakemake/blob/main2/workflow/rules/bam2bedgraph2bigwig.smk*
DESeq2_Feature_Scale	*https://github.com/AngLi9849/RNASeqSnakemake/blob/main2/workflow/scripts/R/deseq2_feature_scale.R*
Differential_Express	*https://github.com/AngLi9849/RNASeqSnakemake/blob/main2/workflow/rules/differential_express.smk*
Differential_Reports	*https://github.com/AngLi9849/RNASeqSnakemake/blob/main2/workflow/rules/differential_reports.smk*
Differential_Splice	*https://github.com/AngLi9849/RNASeqSnakemake/blob/main2/workflow/rules/differential_splice.smk*
Metagene	*https://github.com/AngLi9849/RNASeqSnakemake/blob/main2/workflow/rules/metagene.smk*
Process_Alignment	*https://github.com/AngLi9849/RNASeqSnakemake/blob/main2/workflow/rules/process_alignment.smk*
QC	*https://github.com/AngLi9849/RNASeqSnakemake/blob/main2/workflow/rules/qc.smk*
Reference	*https://github.com/AngLi9849/RNASeqSnakemake/blob/main2/workflow/rules/reference.one_trs.smk*
Scale_Factors	*https://github.com/AngLi9849/RNASeqSnakemake/blob/main2/workflow/rules/scale_factors.smk*
STAR_Align	*https://github.com/AngLi9849/RNASeqSnakemake/blob/main2/workflow/rules/star_align.smk*
Trim	*https://github.com/AngLi9849/RNASeqSnakemake/blob/main2/workflow/rules/trim.smk*
Utilities	*https://github.com/AngLi9849/RNASeqSnakemake/blob/main2/workflow/rules/utilities.smk*
Validate_Features	*https://github.com/AngLi9849/RNASeqSnakemake/blob/main2/workflow/rules/validate_features.gene_only.smk*

II.4.3. Configuration YAML and Tables

Configuration	Github Address
General Configs	*https://github.com/AngLi9849/RNASeqSnakemake/blob/main2/config/config.yaml*
Experiments	*https://github.com/AngLi9849/RNASeqSnakemake/blob/main2/config/experiments.tsv*
Features	*https://github.com/AngLi9849/RNASeqSnakemake/blob/main2/config/features.tsv*
Gene_Sets	*https://github.com/AngLi9849/RNASeqSnakemake/blob/main2/config/gene_sets.tsv*
Groups	*https://github.com/AngLi9849/RNASeqSnakemake/blob/main2/config/groups.tsv*
Protocols	*https://github.com/AngLi9849/RNASeqSnakemake/blob/main2/config/protocols.tsv*
References	*https://github.com/AngLi9849/RNASeqSnakemake/blob/main2/config/references.tsv*
Samples	*https://github.com/AngLi9849/RNASeqSnakemake/blob/main2/config/samples.tsv*

II.4.4. Snakemake Pipeline

A Snakemake pipeline was constructed to integrate, standardise and conditionally pipeline adaptor trimming, quality control, reference genome and annotation fetching, annotation filtering, read alignment, bigwig and matrix generation, differential expression and differential splicing analysis and data visualisation in R. The pipeline is available for reviewing at https://github.com/AngLi9849/RNASeqSnakemake.

II.4.5. Total and fractionation RNA-Seq mapping and normalisation

Raw fastq files containing paired-end RNA reads were adapter trimmed using cutadapt package and filtered by a minimum quality score of 20 and a minimum read length of 18 (scripted in "Trim"). Trimmed reads were aligned to Ensembl human GRCh38 genome and release 104 annotations using STAR with default parameters and "—multimap_max 100" to allow appropriate snRNA mapping (scripted in "STAR_Align"). Primary alignments in the mapped bam files were filtered and deduplicated using samtools (scripted in "Process_Alignment"). Reads overlapping gene bodies or sub-features were counted using featurecount with parameters "-s 2 -p –minOverlap 10 -M -O" (scripted in "Process_Alignment"), and size factors were generated in R package DESeq2 using gene body count table for downstream normalization (scripted in "Scale_Factors").

II.4.6. TT-Seq mapping and normalisation

Similar to RNA-Seq processing, TT-Seq reads were adaptor-trimmed, quality-checked

and aligned to a combined human GRCh38 and yeast R64-1-1 genome using STAR. PCR and optical duplicates were removed and the primary alignments were analysed and normalised to size-factors generated in DESeq2 using spike-in yeast gene body read counts. Due to the time-sensitive and paired nature of the 4SU-labelling procedure, TT-Seq data was analysed in a replicate-paired fashion, where "replicate" was part of the design formula in both differential expression analysis using DESeq2 and differential splicing analysis using DEXSeq and matrices of log2 fold changes for heatmap plotting were calculated between corresponding control and treatment replicates.

II.4.7. eCLIP data mapping and normalisation

Paired-end, reverse-strand Input and eCLIP read fastq files generated by Nostrand et al. (2016) were retrieved from ENCODE project data base (configured in "Samples", and scripted in "Trim"). The downloaded reads were already adaptor-trimmed and UMI-extracted. UMI-codes were moved to the end of read names using a custom awk script for compatibility with umi-tools package (scripted in "Trim"). Similar to RNA-Seq alignment, eCLIP reads were aligned to the human GRCh38 genome using STAR with 100 multimappers allowed. The alignment bam files were deduplicated using umi-tools package "dedup" command with parameters "--method=unique --random-seed 1 --spliced-is-unique" (scripted in "Process_Alignment"). eCLIP single-nucleotide signals were defined by the 5′ end of the sequenced RNA fragment (Nostrand et al., 2016), which is the first base of read 2 in each pair. Therefore, read 2 from UMI-deduplicated alignments were isolated in samtools using parameter "-F 64", and counted using featurecount command with the additional parameter "—read2pos 5" across features for size factors generation based on gene body counts and differential enrichment analysis across features in DESeq2 (scripted in "Differential_Express"). The per-base coverage of eCLIP signals was extracted from the read 2 bam file into strand-specific bedgraph files using bedtools package "genomecov" command with additonal parameter "-5" for downstream bigwig and matrix generation (scripted in "Bam2bedgraph2bigwig").

II.4.8. Model transcript selection and gene feature and properties extraction

Cell line-specific model transcripts were selected from Ensembl human GRCh38 genome release 104 annotation using Salmon transcriptome mapping and custom awk and bash scripts (scripted in "Reference" and "Validate_Features"). In brief, annotated

transcripts with matching biotype to the parent gene were considered "confident". Confident transcripts were first categorised into different "splice margins" according to their first 5′ and last 3′ splice sites, which defined the most probable splice variants within them. The confident transcripts were ranked according to their exon numbers and spliced transcript length within a splice margin, and the highest ranked transcripts was used as the representative transcript of its splice margin. Confident transcripts were mapped with total RNA-Seq data or cytoplasmic RNA-Seq data of the given cell line in Salmon. The total reads of a splice margin were summed to give its usage. The splice margin with the highest usage within each gene was selected as the model splice margin, and its representative transcript was selected as the model transcript. Features of this transcript, including exons, introns, UTRs and CDS segments, were ordered and assigned forward and reverse indices, and used as the model structure of the gene. The feature indices were used to identify and extract sub-features such as first exons and first introns in downstream processing (scripted in "Features"). Each gene was further categorised according to its Ensembl annotated biotype and the number of exons in its model transcript into mono-exon and multi-exon genes. Length of each feature was calculated based on its annotation, and the GC-content of features was obtained using the bedtools 'nuc' command (scripted in "Features").

II.4.9. Read depth visualisation and Matrix generation

Read alignment in bam files were first converted to per-base raw read depth data in the bedgraph format using "genomecov" command of the bedtools package with parameters "-bga -split" (scripted in "Bam2bedgraph2bigwig"). Bedgraph files were normalised with size factors generated in DESeq2 and converted to bigwig files using bedGraphToBigWig command line tool (scripted in "Bam2bedgraph2bigwig"). Bigwigs were visualised in Integrative Genome Viewer (IGV) according to the in-built Ensembl human GRCh38 genome and release 94 annotations. The bigwig files were used to generate mean read-depth matrix across features using deeptools 'computeMatrix' command (scripted in "Metagene"). Log2 fold change (log2FC) matrices were calculated from the mean read depth matrix using a custom python script with pandas data frame computation functions (scripted in "Metagene"). The number of bins computed for each feature was defined in a table with feature-specific parameters (configured in "Features").

II.4.10. Metagene and Heatmap visualisation

Metagenes were generated from matrix data in R by calculating the median value per bin in the subset of matrix that represents the feature group of interest (scripted in "Differential_Reports"). The medians were plotted against the binned position along the feature using geom_smooth options in ggplot command of the ggplot2 package. Heat maps were generated by plotting the log2FC matrix of features ranked by the property of interest, such as GC content, using geom_raster options in ggplot2. Adaptive log2FC colour scales were calculated according the maximum absolute log2FC in the matrix and centered around white (log2FC = 0), with blue representing decreases (log2 FC < 0) and orange-red representing increases (log2FC > 0).

II.4.11. Differential Expression and Enrichment Analysis

Differential expression or enrichment analysis were performed in the R package DESeq2 using read count tables generated from featurecount and size factors generated for each experiment types as mention above (scripted in "Differential_Express"). The design formula was "~condition" for all experiments other than TT-Seqs, which had the design formula of "~condition + replicate" to address its paired nature. The listData data frame, which contained the differential statistics such as log2FoldChanges and adjusted p-values, was extracted from the result DESeqDataSet object for plotting dot plots such as MA plot, Volcano plot and correlation plots. Reads per kilobase per million (RPKM) of each feature in the given data set were calculated from the total read count extracted using samtools "index" utilities, the baseMean count and the feature annotated length with the formula:

$$RPKM = \frac{(Feature\ Read\ Count)\ \times\ 1{,}000{,}000}{(Total\ Read\ Count)\ \times\ (Feature\ Length)}$$

II.4.12. Spliced and unspliced read filtering and splice junction read counting

Spliced and unspliced reads were filtered from bam files with samtools and custom awk script, where alignments that contained the skipping operation "N" in the column 6 CIGAR string were extracted as the spliced reads, and those that did not contain "N" were extracted as unspliced reads (scripted in "Differential_Splice"). Splice junctions annotated in Ensembl genome that were skipped in at least 1 unique alignment of an analysed total or fractionation RNA-Seq data set were extracted from the STAR alignment output 'SJ.out.tab' with custom awk script, and their 5′ and 3′ splice sites

were isolated and assigned to their parent genes. Using featurecount command with parameters "-M -O", spliced and unspliced reads were counted for each gene when they overlapped by at least 1 and 4bp, respectively, out of 6 bases centred around a splice site.

II.4.13. Differential Splicing ratio analysis

Differential splicing ratio analysis was conducted in the R package DEXSeq based on spliced and unspliced reads (scripted in "Differential_Splice"). The spliced reads count table was assigned as 'countData' and the unspliced read count table was assigned as the 'alternativeCountData', with the underlining implication that the calculated log2 fold changes would reflect the changes in relative ratio of spliced reads in all splice site reads. Similar to differential expression analysis, the listData data frame from the result DEXSeqDataSet object was extracted for plotting dot plots.

II.4.14. Dot plotting for MA, Volcano and Correlation plots

MA (log2FC against mean read count), volcano (P-value against log2FC), GC-log2FC correlation plots and cross-experiment log2FC correlation plots were plotted in ggplot function of the ggplot2 package (scripted in "Differential_Reports"). Dots were plotted using 'geom_point' options, and names of features of the greatest significance or particular interest were labelled using 'geom_text_repel' options of the ggrepel package. In MA, volcano and GC-log2FC plots, dots representing significant (adjusted p-value < 0.05) increases were coloured orange, and those representing significant decreases were coloured blue. In cross-experiment correlation plots, dots may be coloured according to their GC-content in a yellow-violet colour-scale. Lines of regression and its R-values and p-values were computed and drawn using 'stat_cor' options included in the tidyverse package

II.4.15. Public Datasets

Dataset	Source	Reference ID
FUS eCLIP in HepG2	ENCODE	ENCSR464OSH
FUS eCLIP in K562	ENCODE	ENCSR069EVH
hnRNPU eCLIP in HepG2	ENCODE	ENCSR240MVJ
hnRNPU eCLIP in K562	ENCODE	ENCSR520BZQ

III. hnRNPU loss upregulates transcription and splicing

Post-translational modifications of RNA Polymerase II (PolII) C-terminal domain (CTD) are hallmarks of the transcription process and its regulation. Early studies revealed that the middle section of hnRNPU (aa269 to aa536, SPRY to start of Walker B motif) is an inhibitor of CDK7 mediated TFIIH phosphorylation of PolII CTD and transcription elongation (Kim & Nikodem, 1999), and that the C-terminal region of hnRNPU (aa550 to aa823 of hnRNPU) forms a complex with PolII, actin and transcription activating PCAF (Obrdlik et al., 2008), which was shown to be a histone acetyltransferase which binds enhancers (Krumm et al., 1998). Investigation of existing BioChIP (Fan et al., 2017) and eCLIP (Van Nostrand et al., 2016) datasets in the Wilson Lab by Dr. Ivaylo Yochev showed that, relatively to gene bodies, hnRNPU preferentially associated with the promoter region of genes in mouse cells and is enriched around the 5′ end of RNAs in K562 human cells (Yonchev, 2021). In addition, hnRNPU loss was shown to disrupt chromatin scaffolds and histone marker patterns (Fan et al., 2017; Liang et al., 2021), and may lead to diffusion of regulatory chromatin-bound non-coding RNAs (ncRNA), such as Xist and FIRRE (Hasegawa et al., 2010; Hacisuleyman et al., 2014). Together these data suggest that hnRNPU regulates PolII transcription both directly at the start of genes as well as indirectly by shaping chromatin-associated RNA (caRNA) and the epigenetic landscape.

Many hnRNPs were observed to modulate splicing and their loss results in misprocessing of pre-mRNAs (Mayeda et al., 1993; Cao et al., 2012; West et al., 2019 for examples). Xiao et al (2012) found that hnRNPU knock down (KD) increased inclusion of typically skipped exons, such as SMN2 exon 7, potentially as a result of increased Cajal Bodies (CB) and formation of 17S U2 snRNP. Disregulation of splicing may be a crucial factor in neurodevelopmental delays in humans caused by hnRNPU mutations or deletions (Yates et al., 2017; Taylor et al., 2022). Conditional knock out (KO) studies in mice heart and muscle showed that hnRNPU deletion disrupted pre-mRNA splicing and resulted in dilated cardiomyopathy and adult on-set myopathy, respectively (Ye et al., 2015; Bagchi et al., 2020). A recent study also showed that mutant alleles of hnRNPU in mouse cortices disrupted mRNA processing and induced cortical cell deaths (Sapir et al., 2022). Further understanding of the scope and extent of splicing disruption due to loss of hnRNPU function could lead to better knowledge of the underlying mechanisms and potential treatment for patients.

Our particular domain of interest in hnRNPU is the AAA+ ATPase-like (aa480 to aa624) domain. It forms part of the TFIIH-inhibiting fragment (Krumm et al., 1998) and was reported to bind ATP and catalyse ATP hydrolysis (Nozawa et al., 2017). This domain shares high structural and sequence homology between hnRNPU and the related protein hnRNPUL1 and together with the adjacent SPRY domain, was reported as a novel RNA-binding domain in a large scale RBP motif screen (Panhale et al., 2019).

Dr Carmen Apostol and Dr. Ivaylo Yonchev in the Wilson Lab have been researching the biochemical and cellular functions of the ATPase-like domain in hnRNPUL1. Dr Apostol found high structural homology between this domain and the kinase domain of mammalian polynucleotide kinase phosphatase (PNKP), and by showing its binding to GTP, she confirmed that its affinity is not ATP-specific, like that of PNKP (Apostol, 2023). She was also able to transform hnRNPUL1 into a polynucleotide kinase by introducing PNKP-assimilating mutation N456D. Hence, we renamed the ATPase-like domains of hnRNPU-proteins as a Pseudo Polynucleotide Kinase (PPNK) domain.

Dr Apostol further identified complementing surfaces with opposing surface charges between the SPRY and the PPNK domains of hnRNPUL1. She has also shown that SPRY-PPNK domains can simultaneously bind ADP and monophosphorylated 5′ end of uncapped RNA, but not capped RNA. Importantly, Dr Apostol discovered that the ATP-hydrolysis activity of hnRNPUL1, purified from E.Coli, disappeared after removing contaminant HSP70 from the purified protein. Together, this suggested that the SPRY-PPNK domain of hnRNPUL1 is an NTP and uncapped RNA binding domain that does not hydrolyse the substrate on its own. Lastly, Dr Apostol found that loss-of-function mutation at the hnRNPUL1 Walker A motif deprived SPRY-PPNK's NTP-binding ability, while loss-of-function mutation at Walker B motif improved that.

Dr. Ivaylo Yonchev showed that Walker A and Walker B motif mutations enhanced chromatin-association and PolII interaction of hnRNPUL1 (Yonchev, 2021). Through CUT&RUN sequencing, he further identified that these hnRNPUL1 mutants enriched over chromatin structural elements such as enhancer-looping element CTCF-binding sites, and he observed RNA-dependent association between CTCF and both wild type and Walker A mutant hnRNPUL1.

In this chapter, we used subcellular fractionation and high-throughput sequencing techniques to investigate the transcription and splicing effects of knocking out hnRNPU

in an auxin-inducible degron (AID) cell line established by Dr. Catherine Heath. We quantified transcription and nascent RNA splicing changes in hnRNPU KO cells by the combination of TT-Seq and chromatin RNA-seq. We also investigated the splicing level changes in hnRNPU syndrome patient derived lymphoblastoid cells by RNA sequencing. Lastly, we explored the possible molecular links between the splicing upregulation in hnRNPU KO and U2 snRNP activities, in particular that of U2 snRNP stem-loop recognition subunit SF3B1.

We then took our existing understanding of SPRY-PPNK domains and Walker A and B mutants of hnRNPUL1 to investigate hnRNPU, by re-introducing copies of the Walker A and Walker B mutants of hnRNPU in hnRNPU-AID KO cells using Sleeping Beauty transposon system. We assayed the chromatin-association of these mutants by subcellular fractionation, their protein-protein interactions by CoIP and whether they revert the increased chromatin RNA splicing in hnRNPU-AID KO by RT-qPCR.

III.1. CDKs hyperphosphorylate PolII following hnRNPU-AID KO

As hnRNPU is essential for cell survival, earlier loss-of-function studies of this protein have been mostly restricted to RNAi knockdowns and mutations. However, hnRNPU is notoriously difficult to deplete by RNAi (Valente & Goff, 2006), with studies achieving only partial knockdown of the protein. To address this limitation, Dr. Catherine Heath from the Wilson Lab established an hnRNPU-AID cell line in HCT116 cells to conditionally knockout (KO) hnRNPU with 3-IAA treatment. The hnRNPU-AID cell line has a lower basal level of hnRNPU compared with HCT116, and it fully depletes hnRNPU after 24 hours of treatment with 500μM 3IAA (this condition is referred to as "hnRNPU-AID KO" hereafter). (Figure III.1.1A)

In line with the observation by Krumm et al (1998) that hnRNPU inhibited CDK7 activities, we observed increases in RNA PolII Ser2 and Ser5 phosphorylation levels following hnRNPU-AID KO compared to wild-type HCT116 cells. These hyperphosphorylation events could be reverted by treatment with the CDK9 inhibitor DRB or the CDK7 inhibitor THZ1 (Figure III.1.1B).

As PolII is largely nuclear and chromatin associated, we questioned whether the differential extractability of PolII rather than global changes accounted for the changes in phosphorylation levels. HCT116 and hnRNPU-AID KO cells were assayed again with the more denaturing RIPA lysis buffer and treated with benzonase to release chromatin engaged PolII. In hnRNPU-AID KO cells, PolII hyperphosphorylation was still observed in RIPA lysed samples without benzonase treatment, which we anticipate corresponds to PolII which is free or loosely associated with chromatin. In contrast, release of the chromatin engaged PolII using benzonase revealed Ser2P PolII levels were similar and there was only a modest increase in Ser5P PolII levels (Figure III.1.1C). The levels of corresponding kinases CDK9 and CDK7 remained similar between HCT116 and hnRNPU-AID KO cells, and the levels of a hnRNPU sister protein hnRNPUL1 increased slightly, while levels of FUS, an ALS-related HNRNP, decreased noticeably. We also investigated potential changes in protein distribution in subcellular fractionations, where we observed significant enrichment of PolII Ser5P in the hnRNPU-AID KO nucleoplasm fraction and a modest increase for Ser2P (Figure III.1.1D). Together these results indicate that following hnRNPU loss there is a greater proportion of Ser2 and Ser5 phosphorylated PolII in the fraction releasable without

destroying the chromatin whereas the levels of Ser2P PolII in the tightly bound chromatin fraction are unchanged and only modestly increased for Ser5P PolII. Whether these effects on PolII phosphorylation and its biochemical fractionation properties are a consequence of direct inhibition of the CDKs by hnRNPU or an indirect effect are not clear from these experiments.

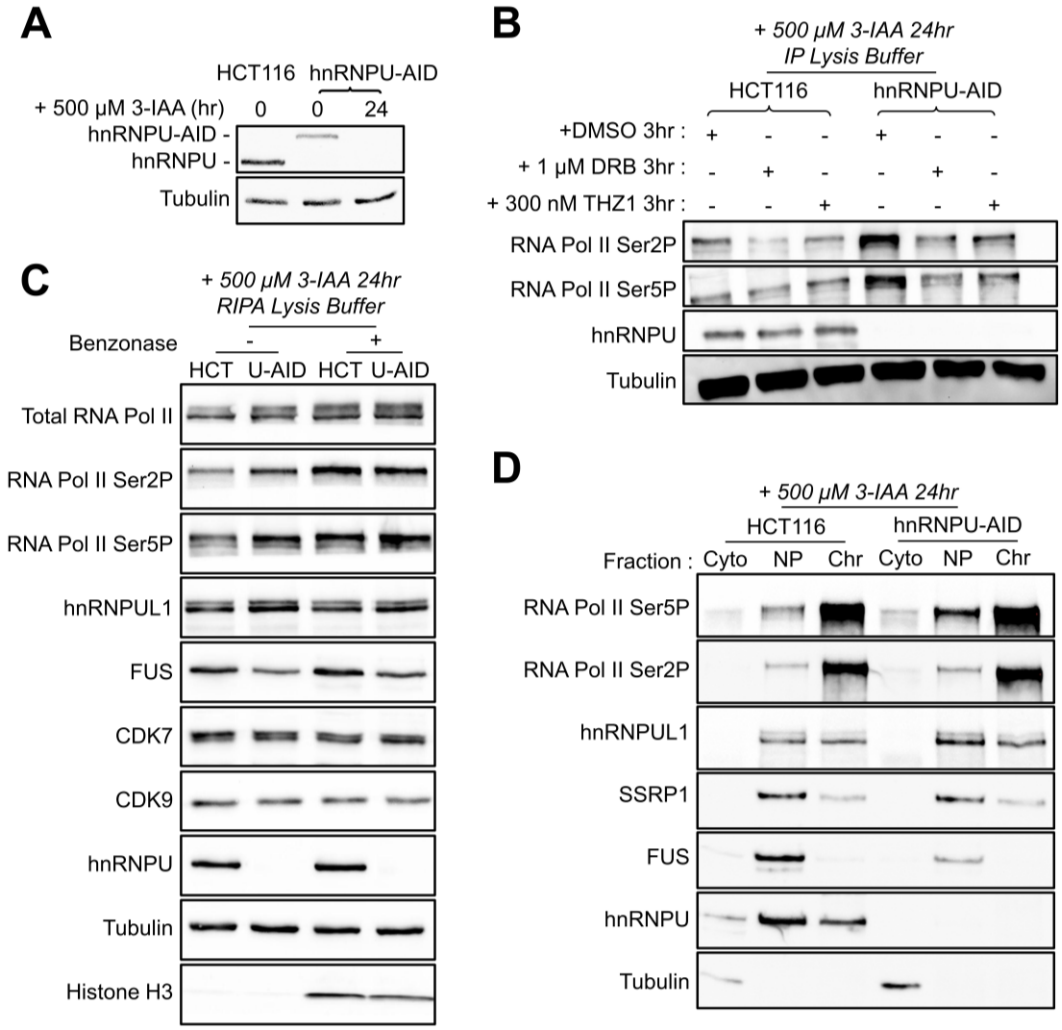

Figure III.1.1. 3-IAA treatment depleted hnRNPU-AID in HCT116-hnRNPU-AID cells and altered PolII phosphorylation levels.

A, Western blot of hnRNPU levels in hnRNPU-AID cells before and after treatment with 500μM 3-IAA for 24 hours compared to untreated HCT116.

B, Western blot of PolII Ser2P and Ser5P levels in HCT116 and hnRNPU-AID cells treated with 3-IAA and either DMSO, DRB or THZ1, and lysed in IP lysis buffer with 1× PhosSTOP phosphatase inhibitor.

C, Western blots of the total, Ser2P and Ser5P PolII, CDK7, CDK9, hnRNPUL1 and FUS in HCT116 (HCT) and hnRNPU-AID (U-AID) cells treated with 3-IAA, lysed in RIPA lysis buffer with PhosSTOP phosphatase inhibitor and treated with or without benzonase.

D, Western blot of PolII Ser2P, Ser5P, hnRNPUL1 and FUS in subcellular fractions (Cyto – cytoplasm, NP – nucleoplasm, Chr – chromatin) of HCT116 and hnRNPU-AID cells treated with 3-IAA.

In **A**, **B** and **C**, tubulin was probed for loading control, and in **C**, histone H3 was probed to prove benzonase digestion of chromatin. In **D**, tubulin is the cytoplasm marker, SSRP1 is the nucleoplasm marker and enrichment of PolII Ser2P is the chromatin marker.

Results of **A** and **C** were repeated in at least three biological replicates, and results of **B** and **D** were repeated in two biological replicates.

III.2. Basal hnRNPU-AID degradation reduced several chromatin ncRNA levels

Dr. Catherine Heath isolated total RNA from subcellular fractions of 24-hour auxin-treated HCT116 and hnRNPU-AID cells for RNA-Seq. She observed lower levels of lncRNAs such as NEAT1, MALAT1 and FIRRE, as well as U1 snRNA in the chromatin fraction of hnRNPU-AID cells in comparison to those in HCT116 (Heath, 2018). (Figure III.2.1A).

During an attempt at differential salt extraction of hnRNPU-AID cells, we noticed that most hnRNPU-AID depletes within 6 hours of 3-IAA treatment (Figure III.2.1B). This inspired investigation into ncRNA chromatin-association at early time-points of hnRNPU-AID 3-IAA treatment. Hence, two biological replicates of chromatin RNA from 0 to 3 hours 3-IAA treated HCT116 and hnRNPU-AID cells were isolated and tested for U1 snRNA (RNU1) and NEAT1 levels by qRT-PCR (Figure III.2.1C), normalized to 18S rRNA.

The results showed that chromatin U1 snRNA and NEAT1 RNA levels were lower in hnRNPU-AID cells than in HCT116 cells even without treatment with 3-IAA. During the 3-hour 3-IAA treatment, chromatin U1 snRNA levels reduced slightly while NEAT1 levels stayed within the margin of error in hnRNPU-AID cells. From this, we concluded that the observed 24-hour 3-IAA treated hnRNPU-AID KO effects were, at least partially, long-term effects of hnRNPU reduction. Therefore, investigations of the hnRNPU-AID cell line were continued at 24 hours of 3-IAA treatment as a model of long-term hnRNPU KO, consistent with existing data sets in the Wilson Lab.

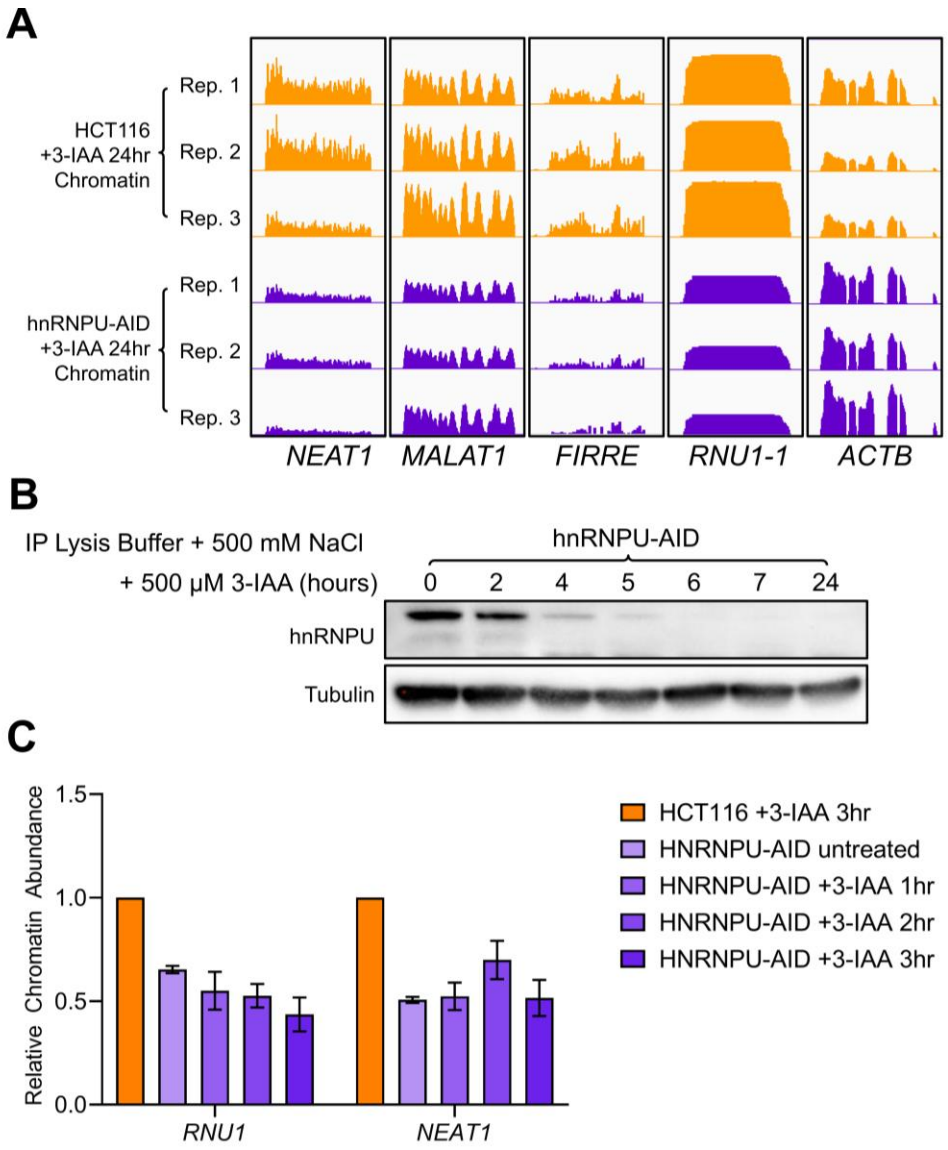

Figure III.2.1. Decreased chromatin lncRNA and U1 snRNA levels in hnRNPU-AID compared to HCT116 cells.

A, Normalised aligned read coverage visualised in IGV at NEAT1, MALAT1, FIRRE, RNU1-1 and ACTB loci in each replicate of chromatin RNA-seq of HCT116 and hnRNPU-AID cells treated with 3-IAA for 24 hours. hnRNPU-AID KO chromatin RNA-seq contained lower read depth at NEAT1, MALAT1, FIRRE and RNU1-1 relative to HCT116 chromatin, compared to typical normaliser loci protein-coding gene ACTB.

B, Western blot of high-salt IP lysis buffer extracted hnRNPU in hnRNPU-AID cells at 0, 2, 4, 5, 6, 7 and 24 hours after 3-IAA treatment, tubulin shown as loading control.

C, RT-qPCR of chromatin-fraction U1 snRNA (RNU1) and NEAT1 RNA in HCT116 and hnRNPU-AID cells in a 3-hour 3-IAA treatment time-course. Error bars represent ranges of two biological replicates.

III.3. hnRNPU regulates PolII interactions with other proteins

Hyperphosphorylation of PolII CTD in hnRNPU-AID KO cells led us to hypothesise that the protein interactome of PolII was altered by the absence of hnRNPU. Hence, we assayed proteins that co-immunoprecipitated with total Rpb1, the main subunit of PolII in DNase-digested whole cell lysate of HCT116 and hnRNPU-AID KO cells (Figure III.3.1A). We used a panspecific monoclonal Rpb1 antibody (MABI-601) to avoid bias for specific phosphorylated forms of PolII.

Rpb1 association with all assayed proteins decreased in hnRNPU-AID KO cell lysate, which included transcription initiation factors FUS and LARP7 (marker of 7SK RNP), PolII Ser2 kinase CDK9, and splicing factor proteins SF3B3 (marker of active U2 snRNP) and U1A (core component of U1 snRNP). The reduction in total PolII interaction with these transcription and splicing factors were reproduced by Dr. Ivaylo Yonchev in a hnRNPU-AID KO chromatin fraction total PolII co-IP, however, he did observe that integrator complex proteins INTS3 and INTS11 associated similarly with PolII with or without hnRNPU (Wilson lab unpublished, data not shown). This may suggest that hnRNPU is required for PolII association with transcription initiation factors and splicing factors, while not essential for its association with termination factors such as the Integrator complex.

Nozawa et al. (2017) observed that α-amanitin inhibition of PolII released hnRNPU from the insoluble nuclear fraction. We wondered if association of hnRNPU and related proteins with chromatin would be affected by functional blockage of PolII transcription in a similar manner. We fractionated two biological replicates of HCT116 cells treated with either DMSO or an irreversible DNA-binding transcription blocker Actinomycin D (ActD) and assayed for hnRNPU, hnRNPUL1 and FUS distributions by Western blotting (Figure III.3.1B).

The levels and distributions of hnRNPUL1 and FUS stayed similar across HCT116 fractions with or without ActD treatment. The typical marker of nucleoplasm SSRP1 enriched on the chromatin following ActD treatment, in contrast hnRNPU levels reduced in the chromatin fraction as did the Ser5 PolII and to a lesser extent Ser2 PolII levels. Thus, the levels of hnRNPU in the chromatin fraction correlate with the amount of ongoing transcription by PolII, which is consistent with the ability of hnRNPU to associate with PolII and nascent RNA.

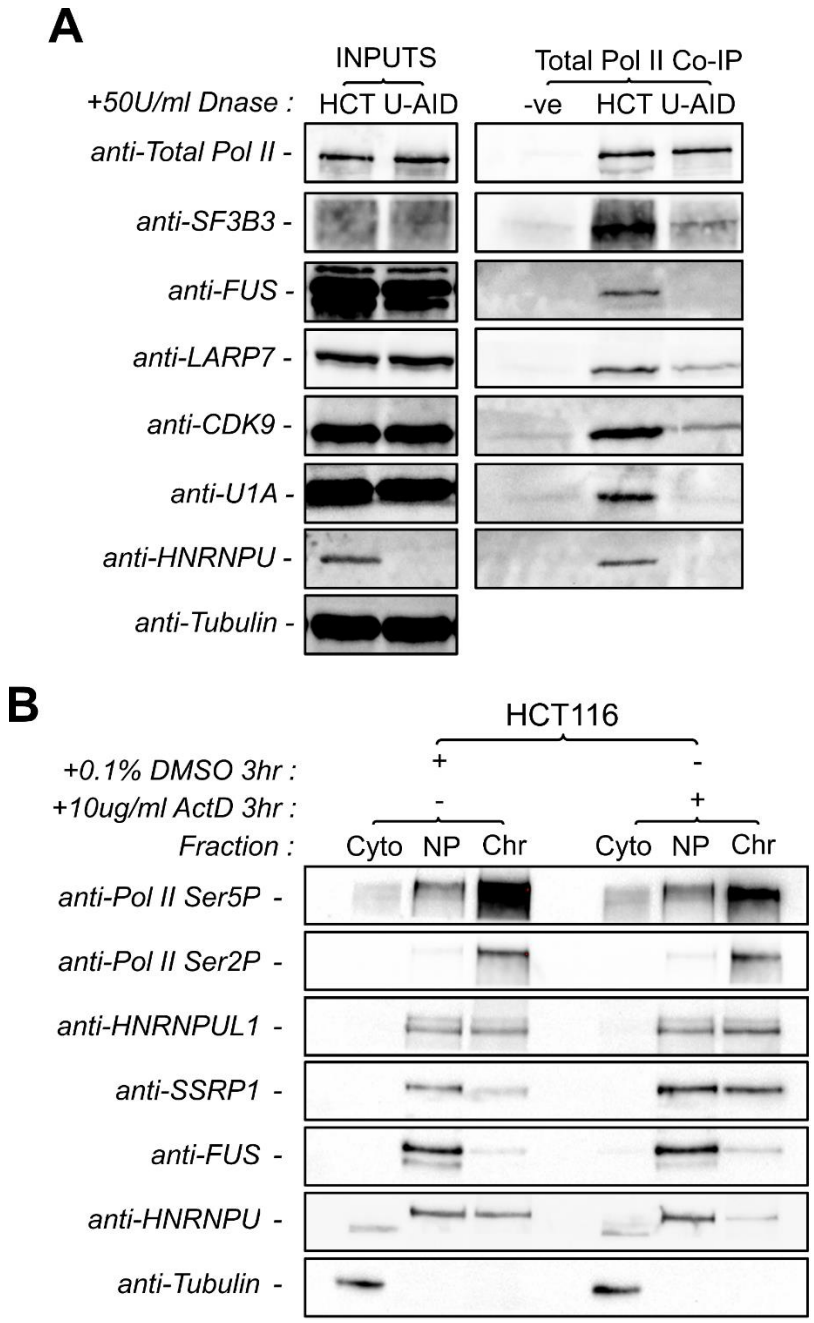

Figure III.3.1. Disruption of hnRNPU and PolII affect each other's behaviour.

A, Western blot showing decreases in levels of SF3B3, FUS, LARP7, CDK9 and U1A in Rpb1 (labelled as PolII) co-immunoprecipitation (co-IP) in 3-IAA treated hnRNPU-AID compared to HCT116 cell lysates. Negative control (-ve) sample was HCT116 lysate immunoprecipitated with a non-specific control monoclonal IgG. Total PolII (RPB1) is shown as loading control for co-IP samples, tubulin is the loading control of inputs. Results were repeated in three biological replicates.

B, Western blot of 0.1% DMSO and 10μg/ml ActD treated HCT116 cells showing relative reduction of hnRNPU compared to SSRP1 and hnRNPUL1 in ActD treated chromatin. In this case, tubulin signals marked cytoplasm, FUS and SSRP1 together marked nucleoplasm and PolII Ser2P and Ser5P enrichment marked chromatin. Results were repeated in two biological replicates.

III.4. PolII Transcription increased globally following hnRNPU-AID KO

The hyperphosphorylation of PolII CTD following hnRNPU-AID KO led us to question whether PolII transcription increased following hnRNPU depletion. Dr. Ivaylo Yonchev have previously assayed levels of PolII-bound nascent RNA using mNET-Seq, where he observed increase levels of anti-sense promoter upstream transcripts (PROMPTs) and the first 100 to 3000 nts of sense genic RNAs in hnRNPU-AID KO cells (Yonchev, 2021). This suggested that PolII was more released from the pause site into active elongation, and transcription was upregulated in both directions from the promoter. To assay the transcription output we decided to use TT-Seq with spike in controls to allow accurate comparisons between datasets.

We followed the TT_{chem}-seq protocol (Gregersen et al., 2020) since it gave more consistent RNA shearing using NaOH rather than sonication used in the original TT-Seq protocol (Schwalb et al., 2016). TT_{chem}-seq also required less sample owing to the use of the more reactive MTSEA-Biotin-XX than the original proposed HPDP-biotin, and ease of preparation and better complexity of 4TU-labeled yeast RNA spike in compared to reverse-transcribed spike-in RNAs (Gregersen et al., 2020). However, we limited 4SU labelling time to 5 minutes as proposed in the original TT-Seq protocol to reduce side effects from post-transcriptional processing (Schwalb et al., 2016). The 15-minute 500μM 4TU-labeled spike-in yeast RNA was prepared using FAE buffer RNA extraction method, proposed by Shedlovskiy et al. (2017), from yeast strain BY4721, a kind gift from Dr. Philip Mitchell. The final purified 4SU-labelled RNA was prepared into strand-specific library using NEBNext® Ultra™ II Directional RNA Library Prep Kit (NEB#7765). The libraries were sequenced by Novogene at a depth of 120M paired 150-base reads per sample.

TT-Seq reads were adaptor-trimmed, quality-checked and aligned to Ensembl Homo *Sapiens* GRCh38 genome sequence and release 104 annotations using STAR (Dobin et al., 2013). Unspliced TT-Seq alignment counts were subject to paired differential expression analysis to compare transcript levels of genes, normalised with size factors generated in DESeq2 using spike-in yeast gene read counts, categorised by exon number and biotypes (Figure III.4.1A). Amongst sufficiently evidenced genes, transcription levels of 29% protein coding genes and 12% multi-exon lncRNA genes increased significantly following hnRNPU-AID KO, while that of 6% protein-coding and 5% multi-exon lncRNA genes decreased significantly. This suggested that more

RNA products were indeed produced by the increased PolII pause-release and elongation observed in hnRNPU-AID KO mNET-seq by Dr. Yonchev.

Metagene analyses were conducted on multi-exon gene bodies and both strands of PROMPT-transcription start site (TSS) regions with at least 10 mapped reads and no overlaps with other expressed genes within 1500bps up and down-stream on the same strand (Figure III.4.1B and C). PROMPTs that overlapped with only antisense lncRNA genes were included as well. TT-Seq read coverage across the multi-exon protein-coding gene bodies was higher in hnRNPU-AID KO cells than in HCT116, with greater increases towards the 3′ end (Figure III.4.1B). The read coverage of PROMPT and TSS regions were slightly higher on average in hnRNPU-AID KO cells (Figure III.4.1C) as well. The metagene profiles confirmed the bidirectional increase of PolII transcription Dr. Yonchev observed in mNET-seq and suggested there was increased PolII elongation following hnRNPU-AID KO.

In addition, we investigated whether properties of the genes played a part in the transcription upregulation. A significant negative correlation between log2 fold changes in unspliced TT-Seq read count and gene GC content was identified (Figure III.4.1D). A similar trend was also observed in a read depth log2 fold change heatmap against protein-coding gene GC content (Figure III.4.1E). These data demonstrate that hnRNPU loss preferentially leads to transcriptional upregulation of AT-rich genes.

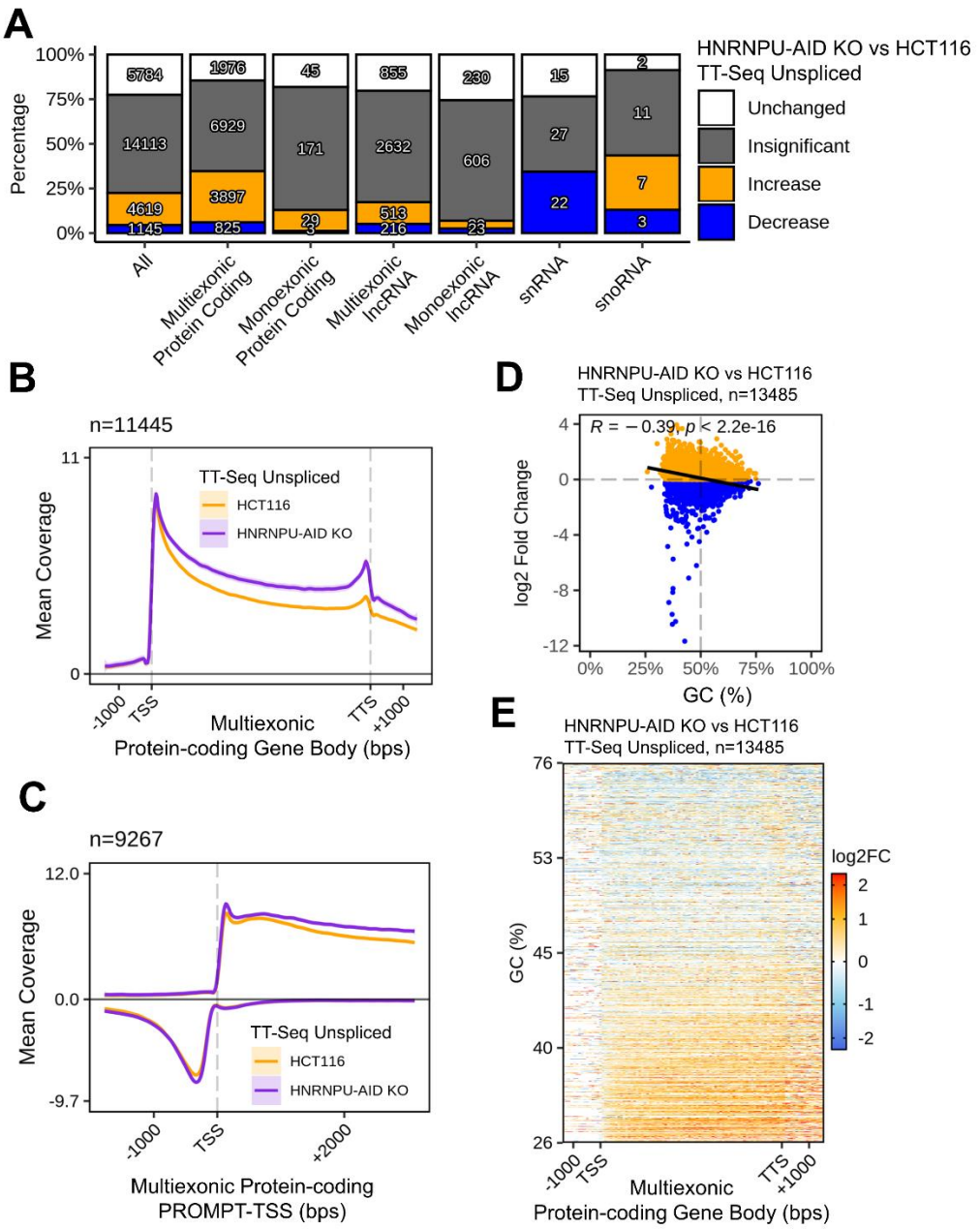

Figure III.4.1. hnRNPU-AID KO vs HCT116 TT-Seq differential expression and metagene analysis.

A, Stacked percentage bar chart of genes with differential unspliced aligned reads in each category of biotypes and exon counts. Labelled numbers are gene counts in each bar sector.

B, Mean unspliced read depth across non-overlapping multi-exon protein-coding gene bodies and 1500 bps up and down-stream of the same strand, included 11445 detected genes.

C, Mean unspliced read coverage over 2000bps upstream and 3000bps downstream on both strands of TSS of multi-exon protein-coding genes, included 9267 genes.

D, Dot plot of log2 fold change in unspliced read count of multi-exon protein-coding genes (y-axis) against their GC content (x-axis). A linear regression model is fitted and shown in black with its R and p values. Increases are shown in orange and decreases in blue.

E, Heatmap of binned log2 fold change (log2FC, heat colour scale as shown) in read coverage along multi-exon protein-coding genes against their GC contents (y-axis).

III.5. hnRNPU-AID KO downregulated transcription of snRNA genes and IEGs

Despite the global increase in transcription of protein-coding and lncRNAs, we noticed that transcription at snRNA loci and immediate early response genes (IEGs) decreased in hnRNPU-AID KO cells.

Downregulation of snRNA was identified as part of global differential TT-Seq analysis (Figure III.5.1A). A metagene plot showed that transcription reduced across snRNA gene loci and downstream regions (Figure III.5.1A). A closer inspection of the snRNA differences with an MA plot showed that majority of significantly downregulated snRNA were RNVU1 loci rather than well-studied spliceosome snRNAs such as RNU1, 2, 4, 5 or 6 (Figure III.5.1B,C). Manual inspection of known curated PolII transcribed snRNA loci revealed that transcription of RNU1 and RNU4 loci slightly increased while that of RNU5 loci decreased (Figure III.5.1D). Interestingly, we observed a common relative reduction in hnRNPU-AID KO TT-Seq read depth downstream of these loci. This could suggest that hnRNPU loss increased nascent snRNA cleavage and processing, and hence destabilised the downstream cleaved pre-snRNA transcripts.

We then identified downregulation of a few IEGs by manual inspection of IGV bigwig tracks, as they are highly expressed and indicative of cell stress levels. To further understand the global effects of hnRNPU-AID KO on IEGs, we extracted a list of immediate-early response genes identified by Tullai et al (2007) and isolated changes in their unspliced TT-Seq read counts from the global differential analysis. Out of 28 sufficiently evidenced IEGs in HCT116 cells, transcription of 11 significantly decreased and only 1 significantly increased (F3) (Figure III.5.2A). Metagene analysis of 26 non-overlapping IEGs showed a decrease in transcription across IEG loci and greater decreases were observed at the beginning of these genes (Figure III.5.2B). Based on a recent report that transcription activation of these stress-response genes is poised by 7SK snRNP (Studniarek et al., 2021), we inspected the read depth of RN7SK in Dr. Heath's fractionation RNA-seq, and noticed that RN7SK levels reduced from the chromatin fraction of hnRNPU-AID KO cells, and increased in the cytoplasm (Figure III.5.2C). This could suggest that hnRNPU loss released RN7SK from chromatin into cytoplasm, which resulted in less 7SK snRNP-driven transcription activation of IEG loci.

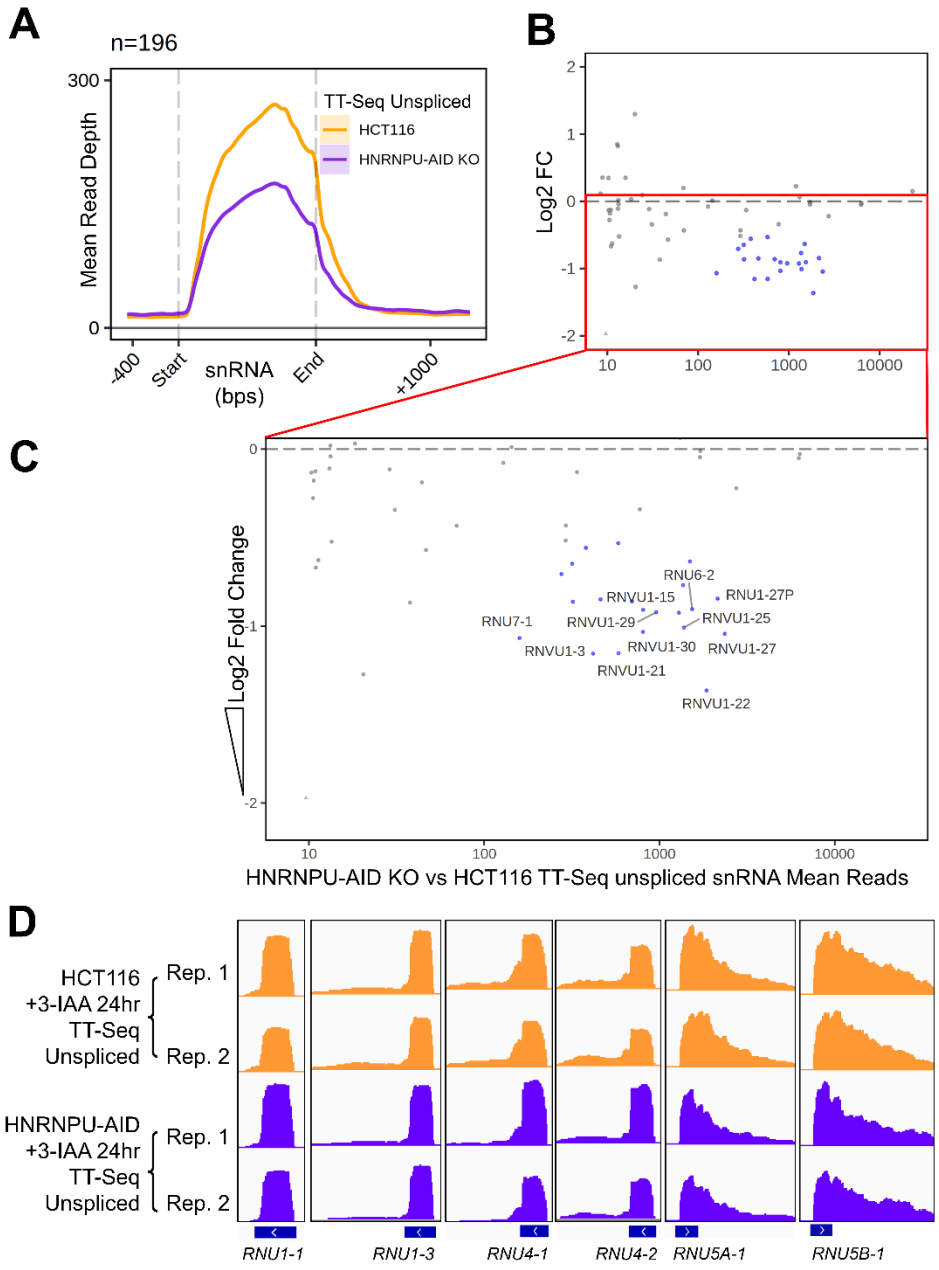

Figure III.5.1. hnRNPU-AID KO downregulated transcription across snRNA loci.
A, Metagene profile of mean unspliced read depth in HCT116 and hnRNPU-AID KO TT-Seq over 196 Ensembl annotated snRNA loci, showing reduced transcription across snRNA and downstream regions in hnRNPU-AID KO cells.

B and **C**, MA plot showing decreased read counts across snRNA loci in hnRNPU-AID KO TT-Seq, with their log2 fold change (y-axis) plotted against their mean read counts (x-axis). **C** is zoomed from **B** on the downregulated snRNA loci, and the most significantly decreased are labelled with their names, showing the majority are RNVU1 loci.

D, TT-Seq read depth along examples of curated PolII transcribed snRNA loci in the two replicates of TT-Seq, visualised and group-autoscaled in IGV. Note that read depth declined more steeply downstream of the mature snRNA gene loci in hnRNPU-AID KO compared to HCT116 cells.

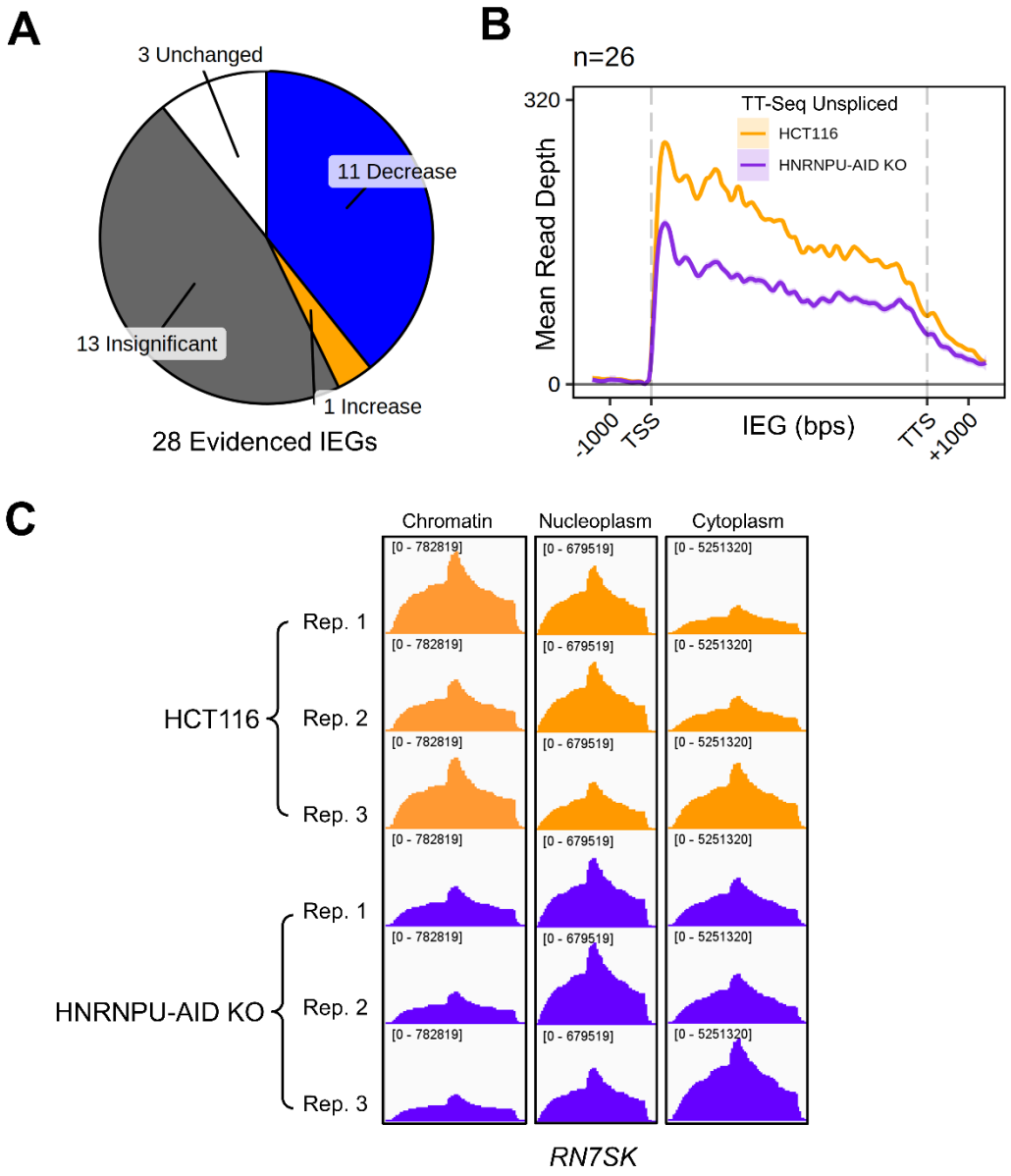

Figure III.5.2. hnRNPU-AID KO downregulated transcription of IEGs.

A, Pie chart showing differential analysis of IEGs of hnRNPU-AID KO against HCT116 TT-Seq, with minimally 1 read mapped. Genes were considered significantly changed with adjusted p <= 0.1 and considered unchanged with adjusted p > 0.95.

B, Metagene plot of unspliced TT-Seq read depth over IEGs and 1500bps up and down-stream in HCT116 and hnRNPU-AID KO cells. 26 IEGs that did not overlap with any other genes were considered. Each gene locus is divided into 1000 bins and read depth are calculated, normalized and averaged for each bin. A smoothen moving average line was plotted for each condition, coloured as indicated in keys.

C, Levels of RN7SK in three replicates of chromatin, nucleoplasm and cytoplasm RNA-seq in hnRNPU-AID KO cells, performed by Dr. Catherine Heath. Read depths are visualised and group-autoscaled in IGV, normalized to size factors generated by gene body read counts in DESeq2. Read depth is displayed in each track to demonstrate fair group-autoscaling.

In addition, a recent report suggested that NR4A1, one of the IEGs, serves as a master negative regulator of typical growth response IEGs, such as EGR1 and FOS, by binding directly to their gene loci and blocking their transcriptions (Guo et al., 2021). In the fractionation RNA-seq data of hnRNPU-AID KO cells, we observed increased levels of spliced NR4A1 transcript in both nucleoplasm and cytoplasm, but reduced levels of nascent transcripts in the chromatin (Figure III.5.3A). We then assessed the protein levels of NR4A1 and EGR1, where we observed upregulation of NR4A1 and downregulation of EGR1 proteins in hnRNPU-AID KO cells (Figure III.5.3B). This is consistent with the increased splicing and export of NR4A1 mRNA and the theory that upregulation of NR4A1 inhibited transcription at IEG gene loci (Figure III.5.3A; Guo et al., 2021). The retention of transcribed, unspliced NR4A1 mRNA on the chromatin mediated by hnRNPU is reminiscent of its ability to tether lncRNAs such as XIST to chromatin. Interestingly this unprocessed pre-mRNA retains the capacity to be processed since loss of hnRNPU leads to increased levels of NR4A1 processed mRNA in both the nucleoplasm and cytoplasm.

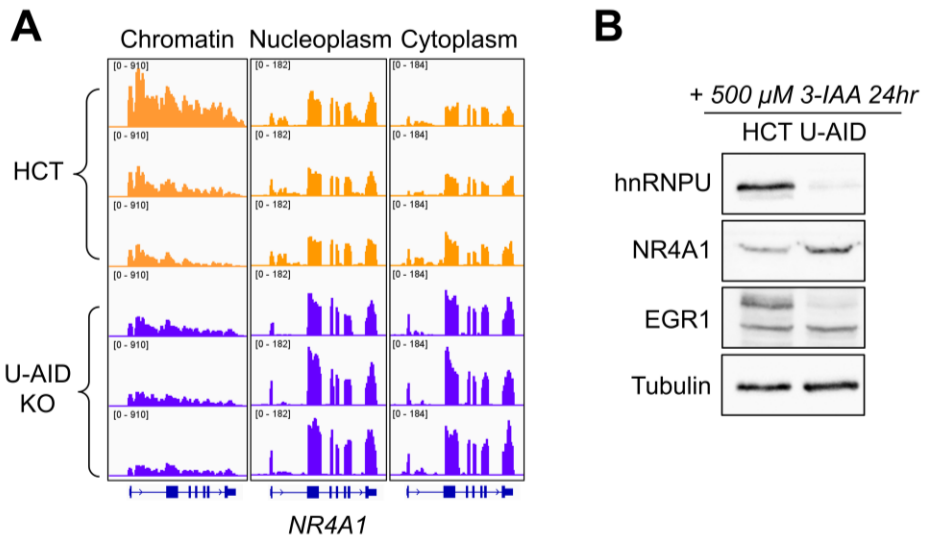

Figure III.5.3. hnRNPU-AID KO upregulated NR4A1 and suppressed EGR1

A, Read-depth across NR4A1 transcript in three replicates of chromatin, nucleoplasm and cytoplasm RNA-seq in HCT116 (HCT) and hnRNPU-AID KO (U-AID KO) cells, performed by Dr. Catherine Heath. Read depths are visualised and group-autoscaled in IGV, normalized to size factors generated by gene body read counts in DESeq2. Read depth is displayed in each track to demonstrate fair group-autoscaling.

B, Western blot showing demonstrating the upregulation of NR4A1 and the downregulation of EGR1 proteins in 3-IAA treated HCT116 (HCT) and hnRNPU-AID (U-AID) cells. HnRNPU was probed to demonstrate its depletion and tubulin is shown as loading control.

III.6. Nascent RNA splicing increased globally following hnRNPU-AID KO

We noticed increased contrast between exonic and intronic read coverage in hnRNPU-AID KO chromatin RNAs (example genes SAT1 and HNRNPH3 shown in Figure III.6.1A) in Dr. Heath's subcellular fractionation RNA-seq datasets. This suggested that pre-mRNAs were spliced more efficiently in hnRNPU-AID KO cells. Hence, we reanalysed the dataset bioinformatically to evaluate the extent and scope of splicing.

A metagene analysis of chromatin RNA-seq read coverage across expressed protein-coding gene introns confirmed the greater contrast between the mean read coverage of flanking exons and that of the intron in hnRNPU-AID KO chromatin RNA indicative of more efficient splicing (Figure III.6.1B). A heatmap of read coverage log2 fold change showed that exon-intron contrast increased globally across introns in hnRNPU-AID KO chromatin RNA, regardless of their length (Figure III.6.1C).

We calculated splicing ratios as the proportion of spliced reads in all reads across splice sites (Figure III.6.1D). The extent and confidence levels in changes of splicing ratios were computed using the DEXSeq package as the change in relative occurrence of reads derived from spliced RNA in all splice site reads. Splicing ratios were calculated both on the per biotype basis (denoted "total splicing ratio" of the biotype) and per gene basis.

Relative to the HCT116 chromatin RNA, total and per-gene splicing ratios of lncRNA and protein-coding genes were higher in hnRNPU-AID KO chromatin RNA by approximately 30% to 50% on average (Figure III.6.2A). Total and per-gene splicing ratios of lncRNAs increased to a greater extent in hnRNPU-AID KO cells than those of protein-coding genes. Notably, this corresponded to the higher enrichment of hnRNPU eCLIP signals over lncRNAs observed by Dr. Yonchev from ENCODE data (Yonchev, 2021).

Per gene differential analysis showed that splicing ratios significantly increased in 50% (n=5561) of 11113 sufficiently evidenced genes and decreased in only 0.16% (n=18) of hnRNPU-AID KO chromatin RNAs (Figure III.6.2B). The splicing ratios increased regardless of gene GC content (Figure III.6.2C). A volcano plot showed that increases in per-gene splicing ratios were greater and more confident than decreases (Figure III.6.3). Interestingly, we found that changes in splicing ratios of well-expressed and partially released chromatin lncRNAs negatively correlated with changes in their

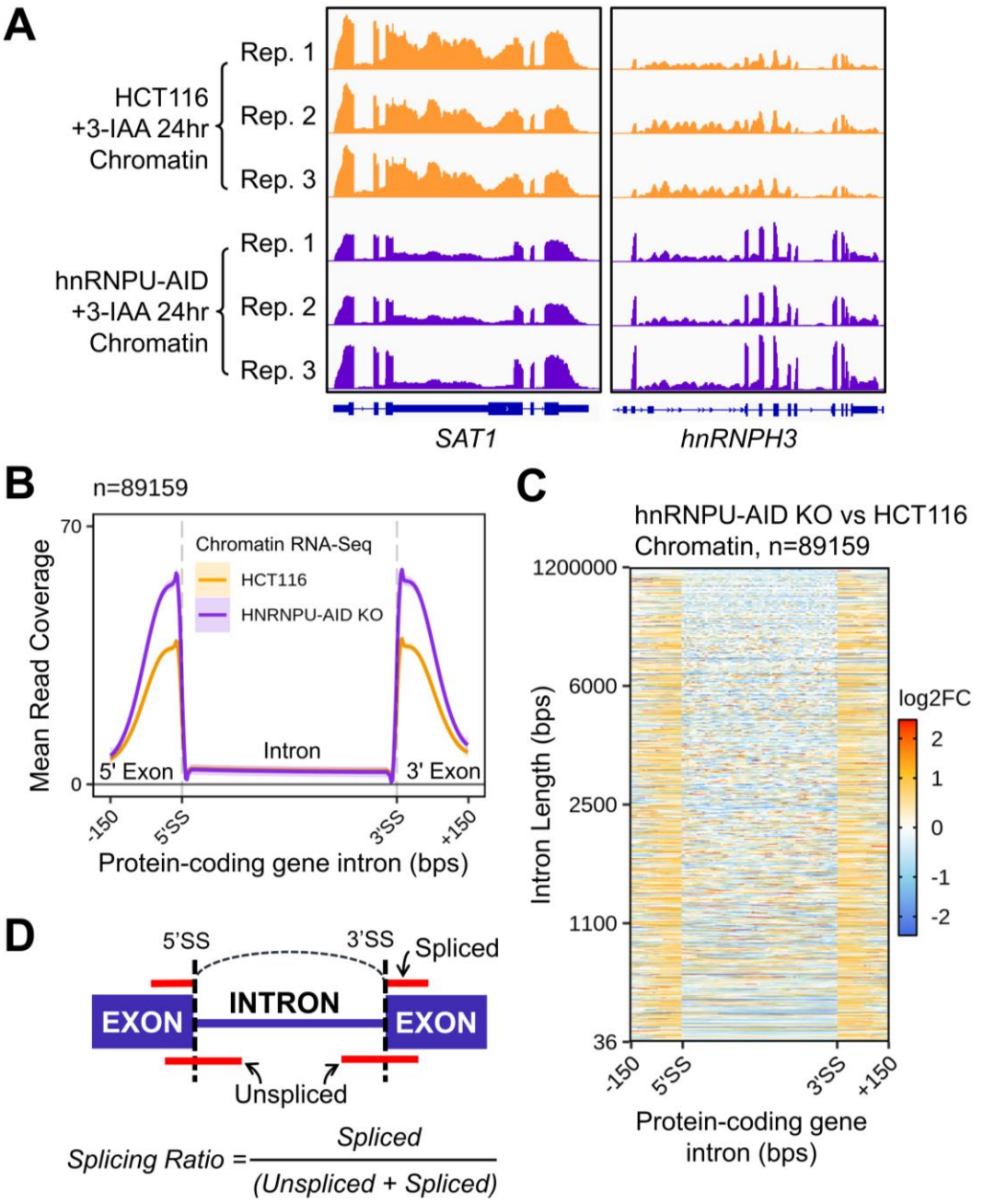

Figure III.6.1. Increased intron-exon contrast suggested upregulated splicing in hnRNPU-AID chromatin RNA-seq.

A, IGV visualisation of normalised read coverage per replicate of chromatin RNA-seq across example genes SAT1 and HNRNPH3 showing increased chromatin RNA splicing in hnRNPU-AID KO cells. Structure of the genes are shown below the read coverage profiles.

B, Metagene plot of mean read depth across protein-coding gene introns and 150 nts up and down-stream showing increase exon-intron constrast in hnRNPU-AID KO chromatin RNA. Read depth was computed from protein-coding gene introns mapped with at least 10 reads, divided into 1000 bins. 5′ splice sites (5′SS) and 3′ splice sites (3′SS) of introns are marked.

C, Heatmap of mean read coverage log2 fold change (log2FC, heat scale as shown) across protein-coding gene introns against intron lengths (y-axis). Read depths on flanking exons relative to the intron in hnRNPU-AID KO chromatin increased across introns of all lengths.

D, Schematics showing sampling of spliced and unspliced reads across splice sites (5′ and 3′SS) and calculation of splicing ratios

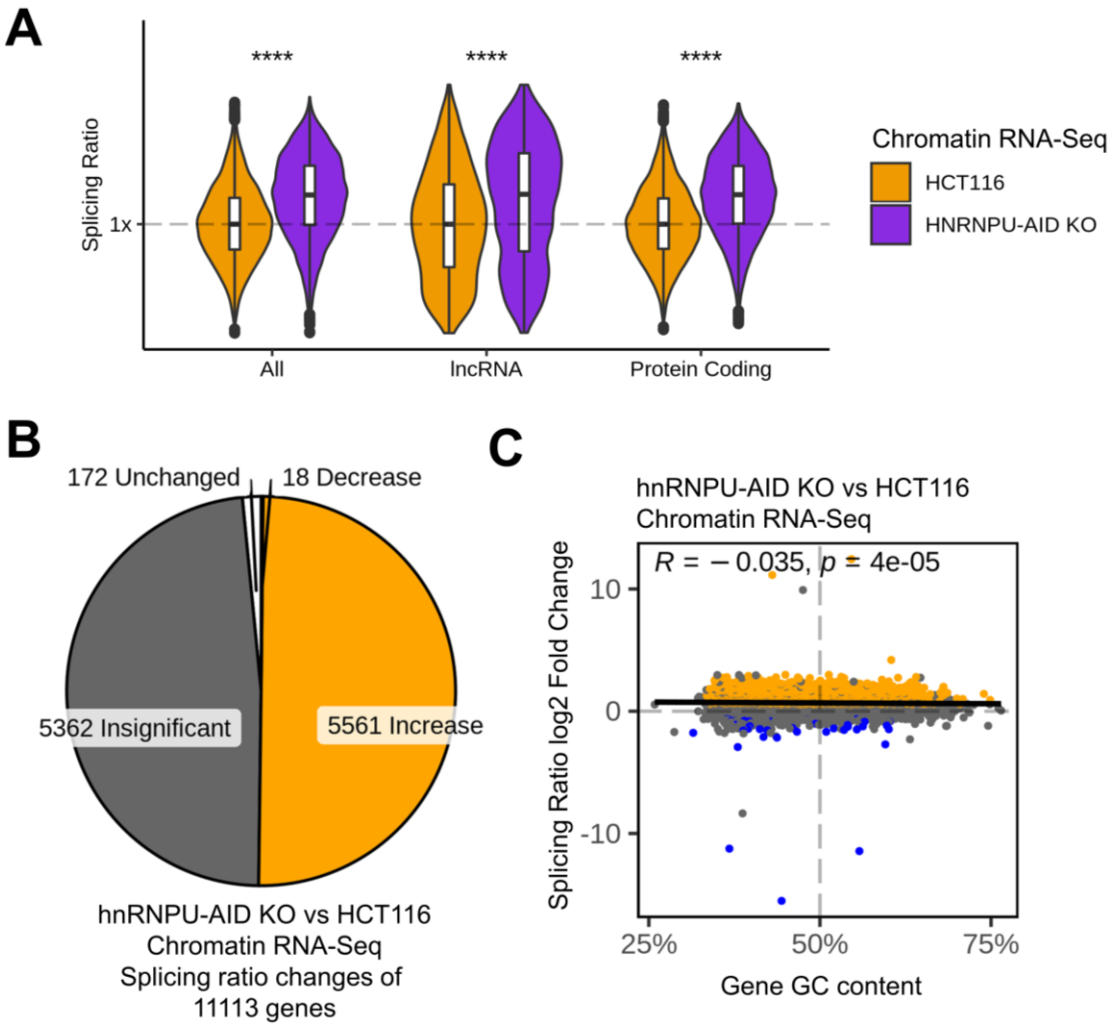

Figure III.6.2. Chromatin fraction RNA splicing ratios increases in hnRNPU-AID KO cells more than in HCT116 cells.

A, Violin and box plots showing upshift of distributions of per-gene chromatin RNA splicing ratios of lncRNA, protein-coding and all expressed genes in hnRNPU-AID KO cells, normalised to the median of HCT116 in each category. Distributions were compared using paired t-test; ****P < 2e-16.

B, Pie chart showing proportions of changes in splicing ratios of all 11113 evidenced genes, with corresponding numbers of genes labelled.

C, Dot plot of splicing ratios log2 fold changes (x-axis) in genes against their GC content (%, y-axis), fitted with linear regression with R and p-values shown.

In **B** and **C**, significant (adjusted P-value < 0.1) increases are shown in orange, decreases in blue and insignificant changes (adjusted P-value >= 0.1) are in grey.

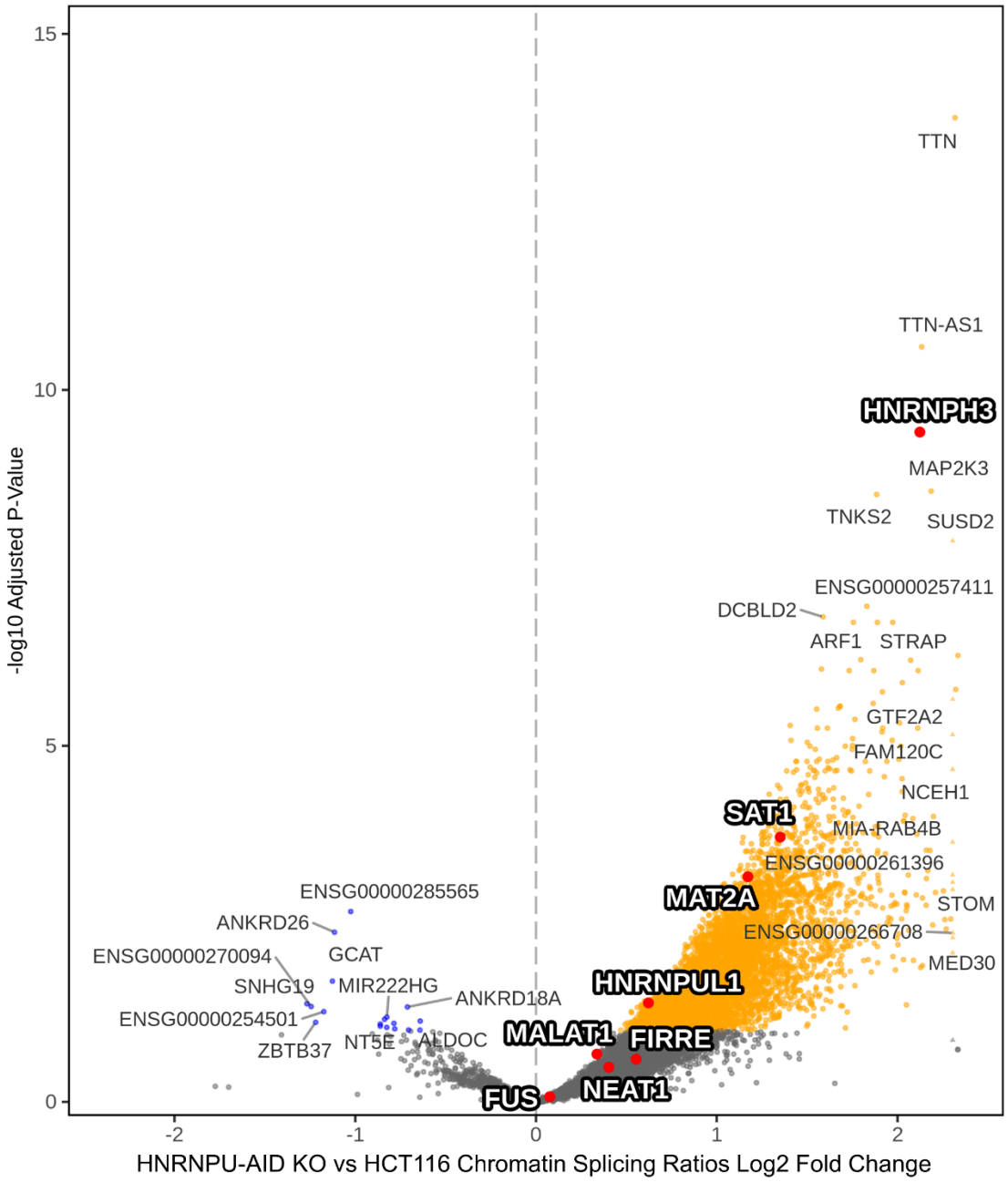

Figure III.6.3. Volcano plot of splicing ratios changes in hnRNPU-AID KO chromatin. Volcano plot of log2 fold changes in gene splicing ratios (x-axis) in hnRNPU-AID KO chromatin against the negative log10 of their adjusted P-values (y-axis). Names of top 20 genes with most significantly increased and decreased splicing ratios are labelled. lncRNAs (NEAT1, MALAT1 and FIRRE) and protein-coding genes (FUS, hnRNPUL1, MAT2A, SAT1 and HNRNPH3) of interest are highlighted in red with names labelled bold and outlined. Significant (adjusted P-value < 0.1) increases are shown in orange, decreases in blue and insignificant changes (adjusted P-value >= 0.1) are in grey.

chromatin levels (Figure III.6.4A), but not with changes in TT-Seq (Figure III.6.4B). These data suggest that increased splicing of chromatin associated RNAs promoted their dissociation from chromatin.

Similar cases of increased splicing were also observed in TT-Seq, hence we analysed co-transcriptional splicing ratios from TT-Seq reads with the same approach. However, due to the nature of TT-Seq, we were mindful of the possibility that spliced reads from highly expressed genes could be carry-over from background total cellular RNAs (Duffy & Simon, 2016).

Global up-regulation of co-transcription splicing following hnRNPU-AID KO was observed in TT-Seq across both lncRNA and protein-coding genes (Figure III.6.5A), similar to that observed in chromatin RNA-seq. Per-gene differential analysis suggested that co-transcription splicing ratios increased significantly in 40% of all evidenced genes, and decreased significantly in only 0.12% (Figure III.6.5B). However, there was no GC bias amongst the genes showing altered splicing (Figure III.6.5C). A volcano plot of TT-Seq splicing ratios also showed more and greater increases than decreases in co-transcriptional splicing (Figure III.6.6).

However, in contrast to the result with chromatin RNA-seq, we did not notice a greater increase in lncRNA splicing ratios in hnRNPU-AID KO TT-Seq (Figure III.6.7A), nor a correlation between the general multi-exon gene splicing ratio changes in chromatin RNA-seq and TT-Seq (Figure III.6.7A). Together these results indicated that the splicing ratios reflected in chromatin RNA-seq might be of chromatin retained RNAs previously transcribed, rather than nascently transcribed RNAs. The increase in splicing ratios of multi-exon genes in hnRNPU-AID KO TT-Seq slightly negatively correlated with the increases in their unspliced TT-Seq read counts (Figure III.6.7B). This could be due to reduced intron reads following better splicing, or a negative feedback between transcription rate and splicing fidelity in yeast as shown by Aslanzadeh et al. (2017).

Interestingly, splicing ratios of lncRNAs NEAT, MALAT1 and FIRRE increased slightly, though insignificantly in both hnRNPU-AID KO sequencing datasets. It is also worth noting that the splicing ratios of hnRNPUL1 increased significantly and that of FUS decreased significantly in hnRNPU-AID KO TT-Seq, in line with the changes in their protein-levels following hnRNPU-AID KO (Figure III.1.1). These specific genes were highlighted in the volcano plots (Figure III.6.3 and Figure III.6.6).

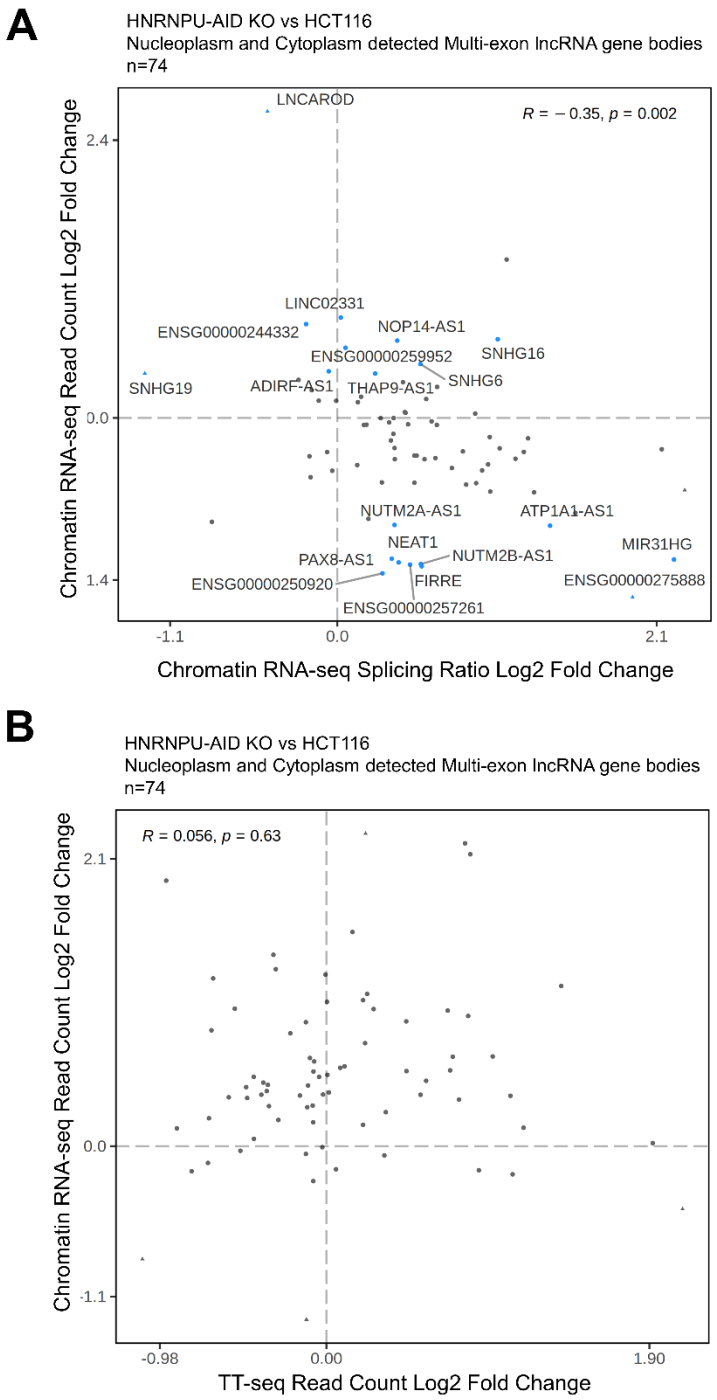

Figure III.6.4. Chromatin lncRNA levels negatively correlate to their splicing ratios

A, Dot plot of detected (mapped with at least 10 reads and 10 reads per kilobase in both nucleoplasm and cytoplasm RNA-seq) and sufficiently evidenced lncRNAs chromatin levels log2 fold changes (FC) against their splicing ratios log2 FC, with a weak but significant negative correlation by Pearson's method (R=-0.35, p=0.002). The top 15% correlated genes on each end of the negative correlation were highlighted and the top 10 are labelled.

B, Dot plot of the same lncRNA genes as in **A**, but showing their chromatin splicing ratio log2FC against TT-Seq unspliced read levels (annotated as "expression") log2 FC. No correlation is detected with Pearson's method (R=0.056, p=0.63).

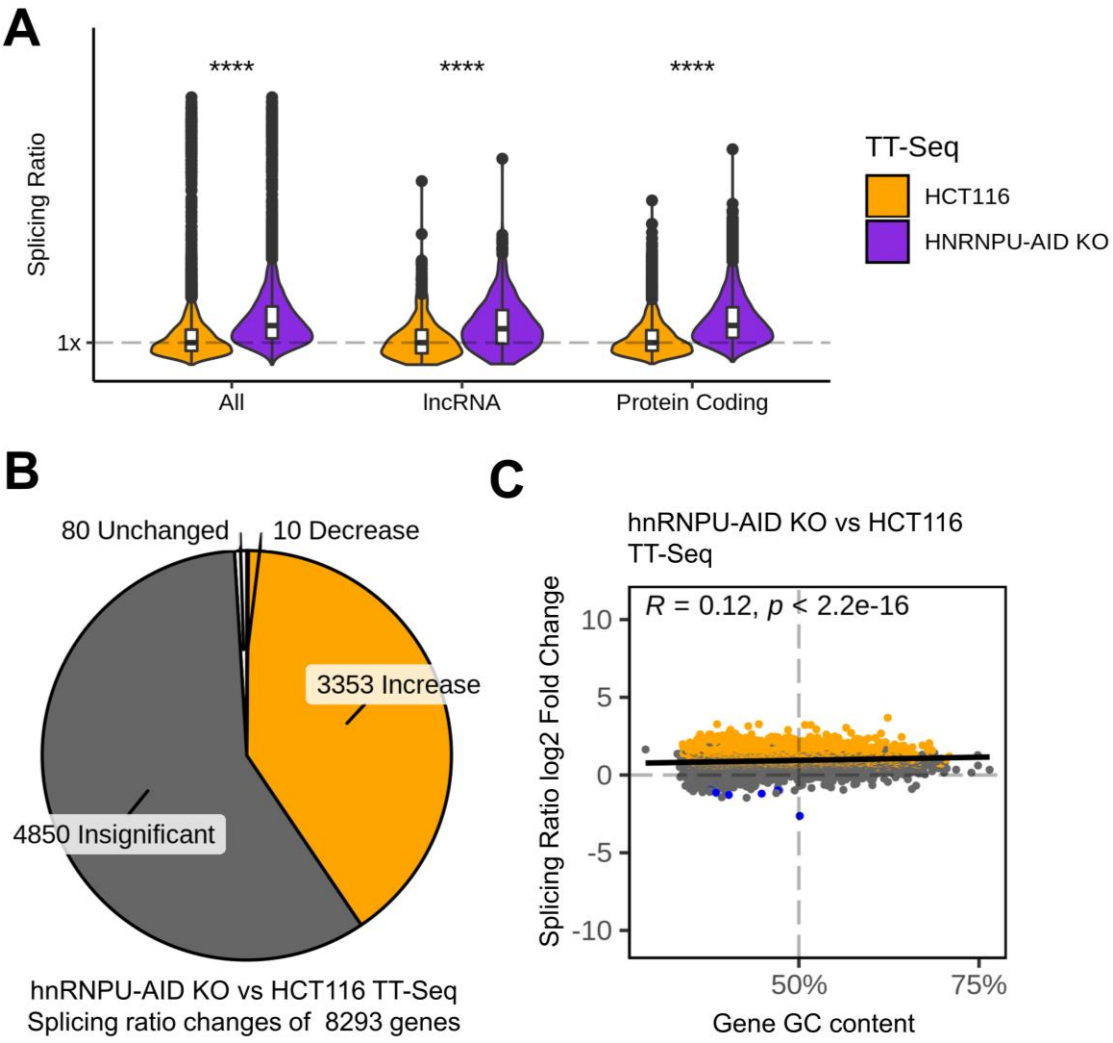

Figure III.6.5. Co-transcription splicing in TT-Seq increased following hnRNPU-AID KO.

A, Violin and box plots showing upshift of distributions of per-gene transient transcript splicing ratios of lncRNA, protein-coding and all expressed genes in hnRNPU-AID KO cells, normalised to the median of HCT116 in each category. Distributions were compared using paired t-test; ****P < 2e-16.

B, Pie chart showing proportions of changes in splicing ratios of all 8293 sufficiently evidenced genes, with corresponding numbers of genes labelled.

C, Dot plot of splicing ratios log2 fold changes (x-axis) in genes against their GC content (%, y-axis), fitted with linear regression with R and p-values shown.

In **B** and **C**, significant (adjusted P-value < 0.1) increases are shown in orange, decreases in blue and insignificant changes (adjusted P-value >= 0.1) are in grey.

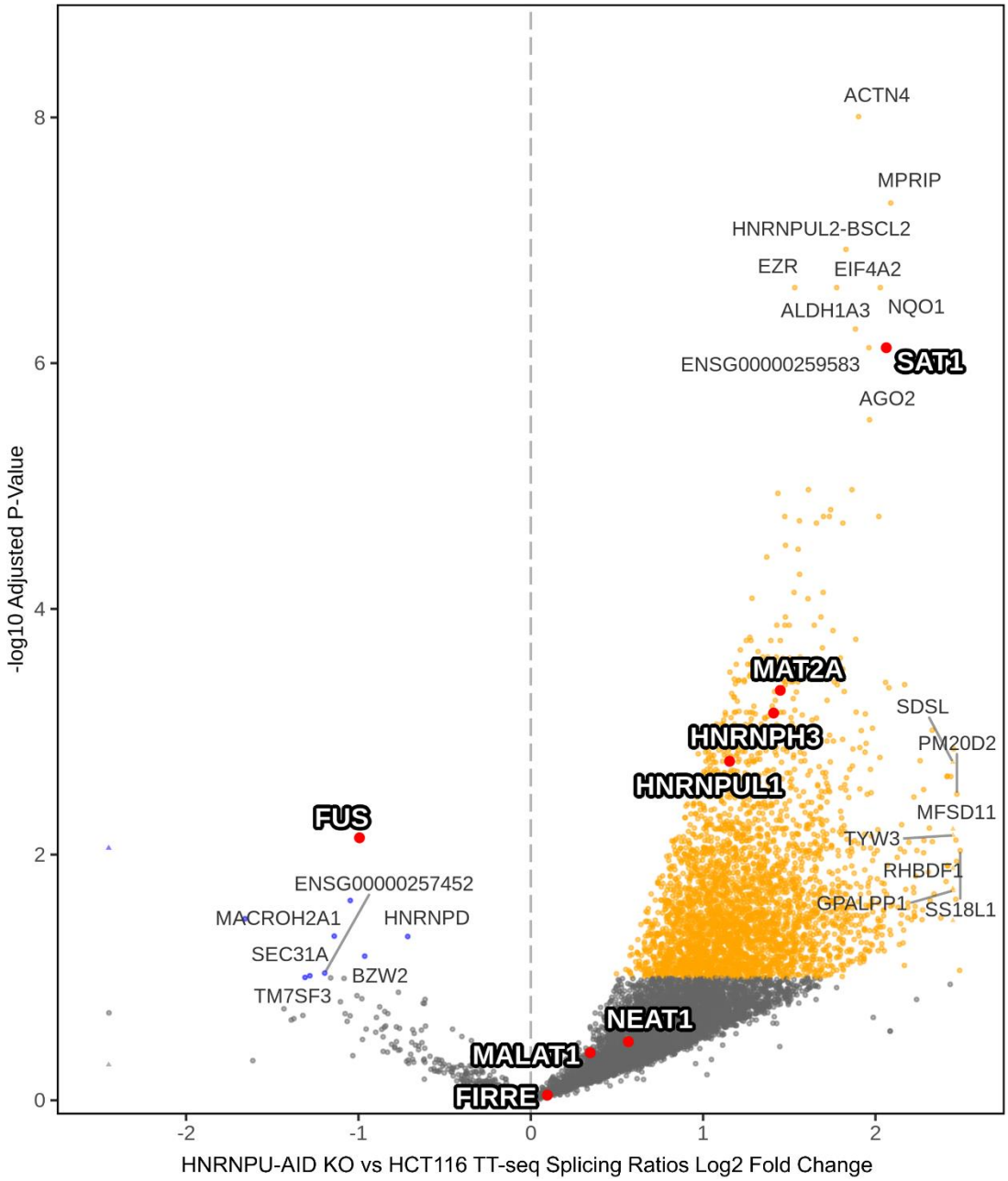

Figure III.6.6.Volcano plot of co-transcriptional splicing changes in hnRNPU-AID KO.

Volcano plot of log2 fold changes in gene splicing ratios (x-axis) in hnRNPU-AID KO TT-Seq against the negative log10 of their adjusted P-values (y-axis). Names of top 20 genes with most significantly increased and decreased splicing ratios are labelled. lncRNAs (NEAT1, MALAT1 and FIRRE) and protein-coding genes (FUS, hnRNPUL1, MAT2A, SAT1 and HNRNPH3) of interest are highlighted in red with names labelled bold and outlined. Significant (adjusted P-value < 0.1) increases are shown in orange, decreases in blue and insignificant changes (adjusted P-value >= 0.1) are in grey.

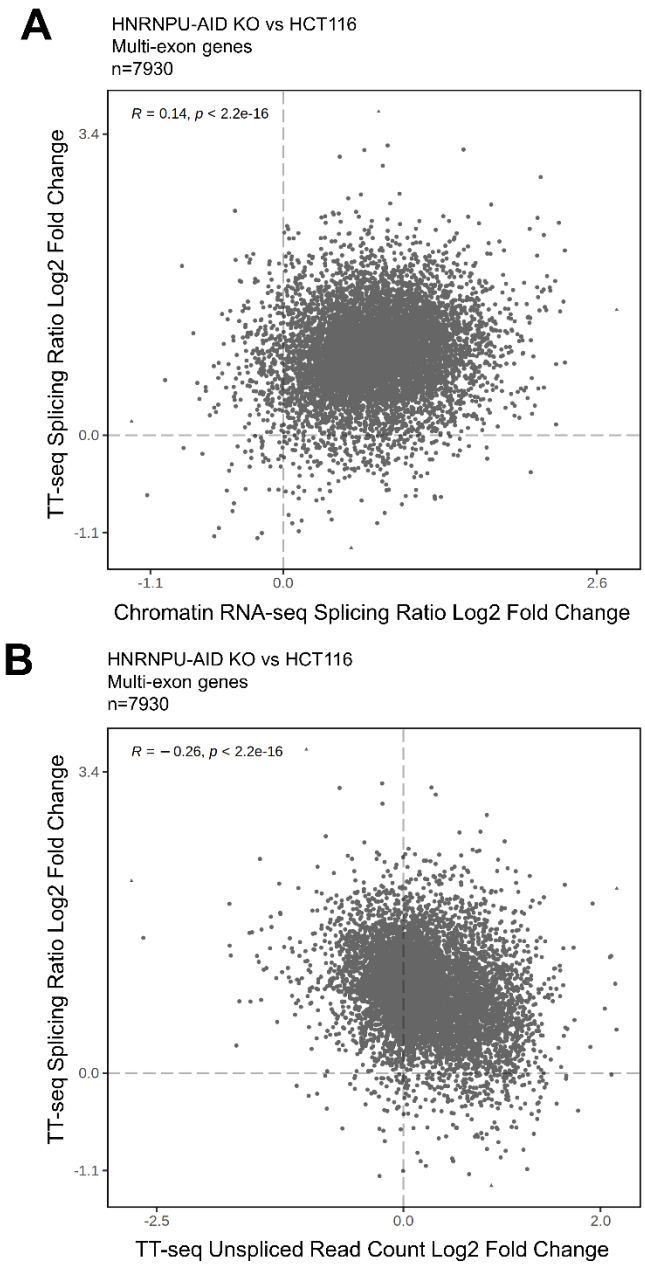

Figure III.6.7. hnRNPU-AID KO TT-Seq splicing ratio changes do not correlate with either chromatin RNA-seq splicing ratio or TT-Seq read count changes.

A, Dot plot of hnRNPU-AID KO vs HCT116 log2 fold changes in multi-exon gene TT-Seq splicing ratio (y-axis) plotted against log2 fold changes in their chromatin RNA-seq splicing ratio (x-axis). They are not correlated according to Pearson's method (R = -0.14, p < 2.2e-16).

B, Dot plot of hnRNPU-AID KO vs HCT116 log2 fold changes in multi-exon gene TT-Seq splicing ratio (y-axis) plotted against log2 fold changes in TT-Seq read count (x-axis). They are insignificantly negatively correlated according to Pearson's method (R = -0.26, p < 2.2e-16).

III.7. hnRNPU-AID KO reduced stabilities of lncRNA FIRRE

The negative correlation between chromatin RNA splicing ratios and levels (Figure III 6.4A) led us to question whether the reduction in chromatin-associated lncRNA levels, notably those of MALAT1, NEAT1 and FIRRE in hnRNPU-AID KO cells, could be a result of reduced stability. This hypothesis was further supported since TT-Seq revealed that transcription levels did not decrease for FIRRE and MALAT1, and not as much as its level reduced in chromatin RNA-seq for NEAT1 (Figure III.2.1A and Figure III.7.1A). Hence, we conducted three replicates of 2-hour 4SU pulse-chase to compare their short-term transcript stabilities in 3-IAA treated HCT116 and hnRNPU-AID cells (Figure III.7.1B).

Figure III.7.1C shows that the stability of FIRRE was noticeably reduced in hnRNPU-AID KO cells, while the decreases in stabilities of NEAT1_2 and MALAT1 transcripts were ambiguous. The reduction in FIRRE RNA stability is consistent with the existing report that chromatin associated RNAs are poorly processed and rapidly degraded by the exosome on departure from the chromatin environment (Schlackow et al., 2017), as well as the earlier observation that XIST dissociates from silenced X-chromosome and undergoes active degradation following hnRNPU knockdown (Hasegawa et al., 2010). Moreover, hnRNPU associates with FIRRE and has been shown to tether it to chromatin (Lu et al. 2017). NEAT1_2 and MALAT1 have half-lives reported longer than 2 hours in some cell lines and may require extended chase time to meaningfully compare their stabilities (Machitani et al., 2020; Amodio et al., 2018).

Following hnRNPU depletion, there is increased splicing of some lncRNAs and we anticipate these spliced lncRNAs evade degradation in the short term (Fig. III6.5A). For many lncRNAs, including MALAT1 (which hosts a small, rarely-used intron), although there is increased splicing, this still only represents a small proportion of the total transcript and therefore total RNA levels for such lncRNAs are frequently downregulated following hnRNPU loss. This is likely to be due to surveillance and degradation by the exosome when such unspliced lncRNAs are released from chromatin. Putting together, these observations may suggest that hnRNPU can tether lncRNAs to chromatin and partially suppress their splicing.

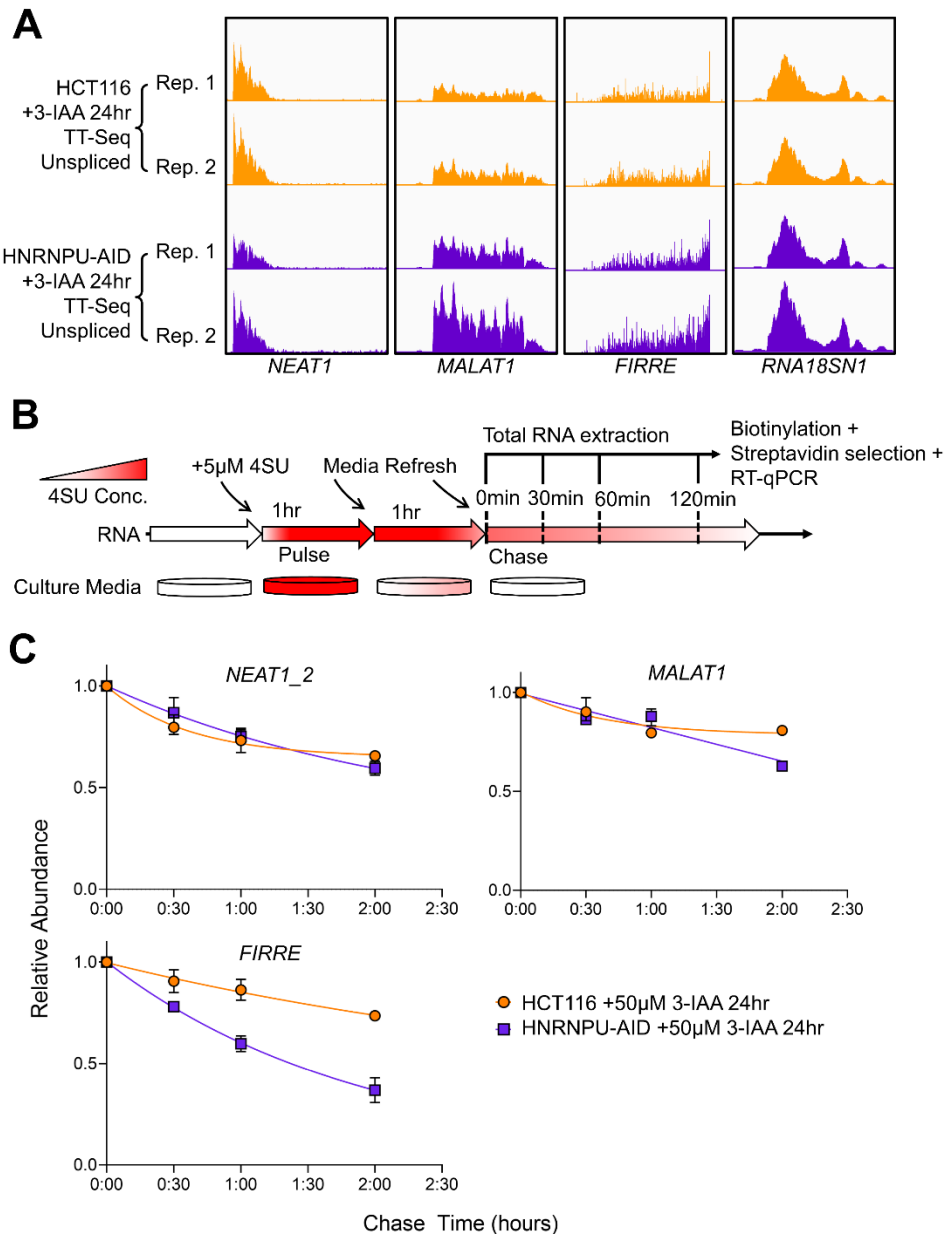

Figure III.7.1. Levels of transcription and stability of lncRNAs NEAT1, MALAT1 and FIRRE in hnRNPU-AID KO cells compared to HCT116 cells.

A, Normalized unspliced read coverage aligned across NEAT1_2, MALAT1, FIRRE and 18S rRNA (RNA18SN1) loci in each replicate hnRNPU-AID KO and HCT116 TT-Seq. IN hnRNPU-AID KO, NEAT1 transient transcript coverage was slightly lower, while MALAT1 and FIRRE coverages were similar to HCT116 cells. 18S rRNA is shown as a normalization example.

B, Schematics of the 2-hour 4SU pulse chase. Cells were treated with 5μM of 4SU for 1hour to label nascently transcribed RNAs (pulse), cultured in refreshed clean media for 1 hour to exhaust cellular 4SU residue. The media was refreshed again, then cells were harvested at 0, 30, 60 and 120 minutes "chase" time points for 4SU labeled RNA purification and RT-qPCR.

C, RT-qPCR of NEAT1_2, MALAT1 and FIRRE in the 2-hour 4SU pulse chase in 24-hour 3-IAA treated HCT116 and hnRNPU-AID cells. For each cell condition, RNA abundance was internally normalized to 18S rRNA then to the chase initial (0hr) time-point. A half-life decay line of fitting is plotted for each target in each condition. Error bars represent anomaly-trimmed range of n=3 biological replicates.

III.8. hnRNPU syndrome patient-derived lymphoblastoid cells showed global upregulation of chromatin RNA splicing

Thanks to the collaboration and help from Dr. Meena Balasubramanian, we obtained two B-lymphoblastoid cell lines derived from blood fibroblasts of hnRNPU syndrome patients, patient 6 with mutation c.G960A and patient 4 with mutation c.23del in Yates et al., 2017. From now on these lymphoblastoid cell lines will be referred to as "HB3261 (960G>A)" and "HB3280 (23del)"). Both cell lines contained a monoallelic mutation that caused a premature STOP codon in the hnRNPU coding sequence (Figure III.8.1A), which we hypothesised would reduce the overall levels of hnRNPU protein.

We received HB3261 (960G>A) first and confirmed the mutation through PCR amplification and Sanger sequencing (Figure III.8.1B). The levels of hnRNPU in this cell line notably decreased compared to a control lymphoblastoid cell line SC3504 (Figure III.8.1C). The extent of reduction in hnRNPU in HB3261 (960G>A) is similar to the basal reduction in hnRNPU-AID cells, hence we hypothesised that the long-term chromatin-associated lncRNA decreases in the hnRNPU-AID cell line were also present in the patient cells, and confirmed that with RT-qPCR on total RNA extracted from HB3261 (960G>A) (Figure III.8.1D).

Upon receipt of HB3280 (23del), we extracted subcellular fractions from three biological replicates of both lymphoblastoid cell lines and a control lymphoblastoid cell line SC3504 (Figure III.8.2A) and isolated the chromatin fraction RNA from two replicates for ribosomal RNA-depleted lncRNA sequencing. Splicing ratios were analysed the same way as for hnRNPU-AID KO chromatin RNA-seq in Chapter III.6.

Total and per-gene chromatin RNA splicing ratios of HB3261 (960G>A) and HB3280 (23del) were higher than those of the control SC3504 cell line (Figure III.8.2B and C). The total levels and distributions of chromatin RNA splicing ratios of the two patient lymphoblastoid cell lines were similar to each other. This increase in splicing ratios corresponded to the similar observations in hnRNPU-AID KO chromatin RNA and TT-Seq. However, we did not observe gene-level correlations between the splicing ratio changes in patient-derived lymphoblastoids and the HCT116-based AID cell line, which is likely due the difference in cell line backgrounds. Overall, these data show that the reduced levels of hnRNPU found in patient cells altered lncRNA levels and RNA splicing as observed in the auxin degron tagged hnRNPU cell line.

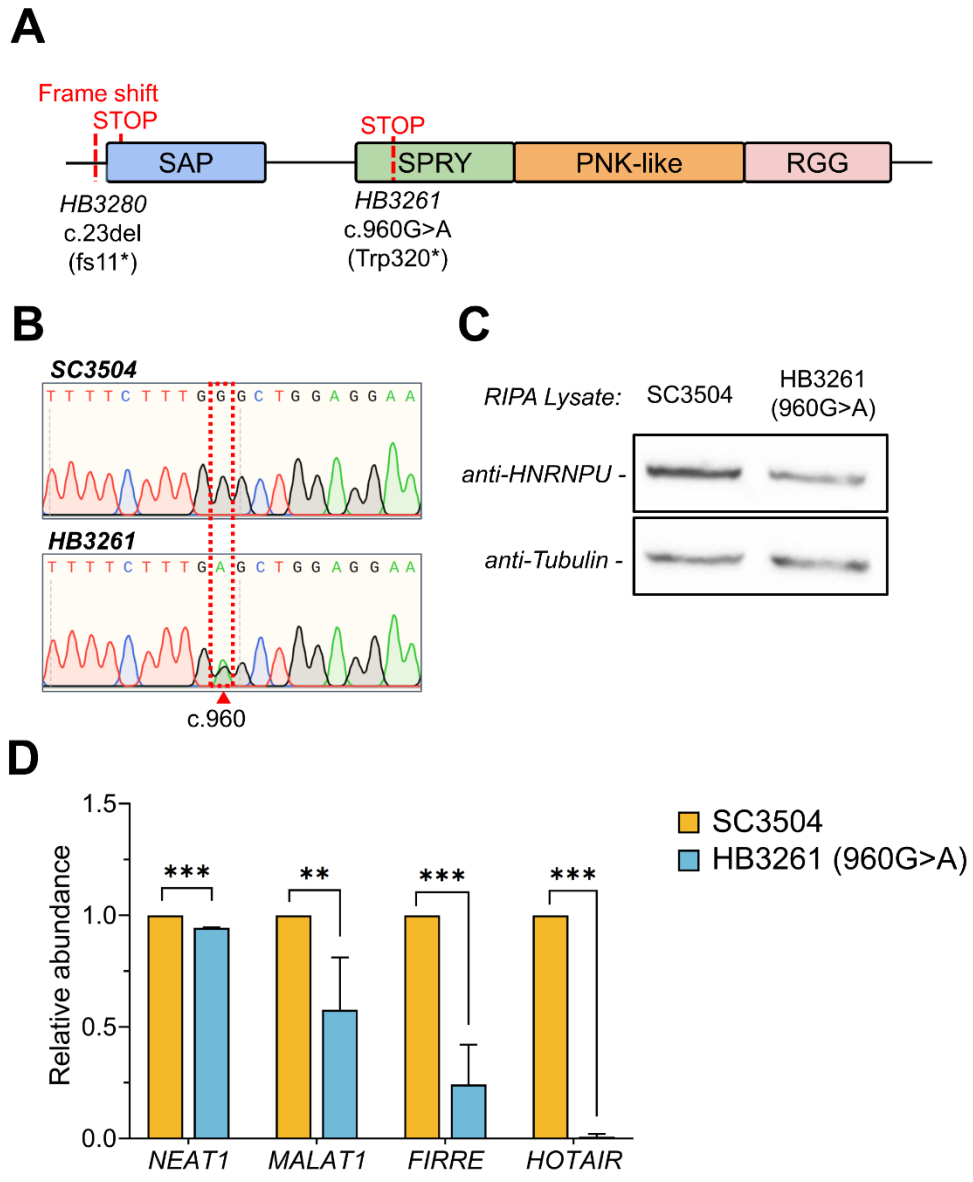

Figure III.8.1. Validation of hnRNPU syndrome patient 6 blood fibroblast-derived lymphoblastoid HB3261 (960G>A).

A, Schematics of the monoallelic mutations in the two hnRNPU syndrome patient cells. HB3280 (23del) carried a frameshift in the coding sequence of hnRNPU that resulted in early STOP codon at frame shifted residue 11, HB3261 (960G>A) carried a point mutation that resulted in early STOP codon at residue Trp320 (or Trp301 of hnRNPU isoform B).

B, Sanger sequencing of PCR amplicon from hnRNPU exon 4. A hybrid of G and A signal is observed at cDNA position 960 in HB3261 cells, which indicates a single-copy mutation that resulted in the early stop codon.

C, Western blot showing hnRNPU protein levels in HB3261 (960G>A) compared to that in SC3504. Cells were lysed in RIPA lysis buffer. Tubulin is shown as loading control.

D, RT-qPCR of lncRNAs NEAT1, MALAT1, FIRRE and HOTAIR in poly(dN)-primed total RNA of HB3261 (960G>A) and SC3504, normalised internally to 18S rRNA levels, then to SC3504 in each replicate. Data is shown as mean ± s.d. of n=3 biological replicates, analysed using two-tailed nonparametric t-tests; $^{**}P < 0.01$, $^{***}P < 0.005$.

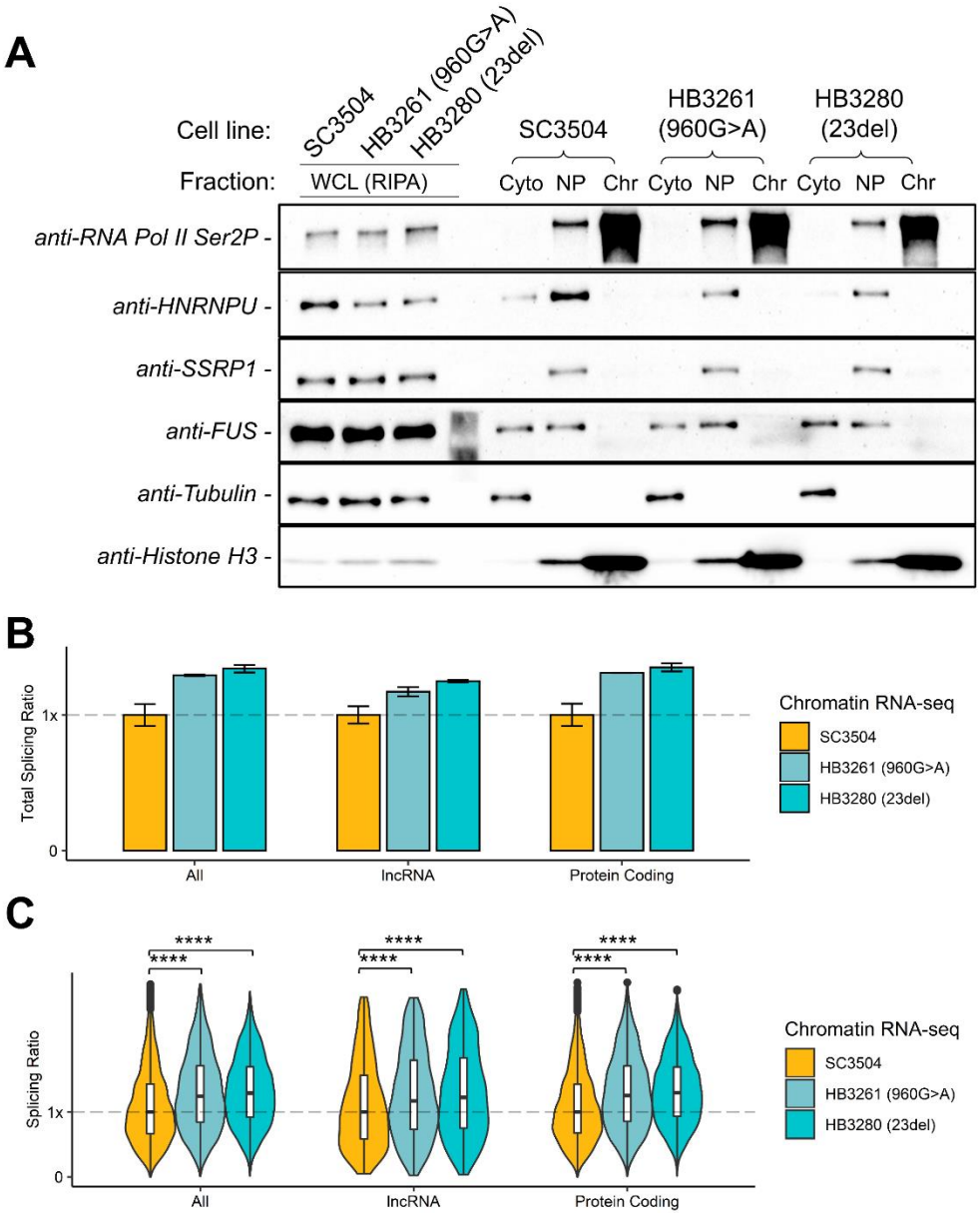

Figure III.8.2. Chromatin RNA splicing ratios were higher in hnRNPU syndrome patient lymphosblastoids than in control SC3504 cells.

A, Western blot of levels of hnRNPU, PolII Ser2 and FUS in whole cell lysate (WCL) by RIPA lysis buffer and subcellular fractions (Cyto – cytoplasm, NP – nucleoplasm, Chr – chromatin) of SC3504, HB3261 (960G>A) and HB3280 (23del) lymphoblastoid cells. Tubulin was used as marker for cytoplasm, SSRP1 for nucleoplasm and histone H3 for chromatin.

B, Bar charts showing mean ± s.d. of total chromatin RNA splicing ratio of lncRNA, protein-coding and all genes (mapped with at least 10 reads) increased in HB3261 (960G>A) and HB3280 (23del) lymphoblastoid cells, compared to the mean of SC3504 in each category.

C, Violin and box plots showing upshift in distributions of per-gene chromatin RNA splicing ratios of lncRNA, protein-coding and all expressed genes, normalised to the median of SC3504 in each category. Distributions were compared using paired t-test, ****P < 2e-16.

III.9. hnRNPU-AID KO Promoted SF3B1 dependent U2 snRNP Activity

Existing reports suggested that hnRNPU knock-down led to upregulated skipped exon inclusion and increased ratios between mature 17S and nascent 12S U2 snRNP (Xiao et al., 2012), and that inhibiting U2 snRNP splice site recognition with pladienolide B (PlaB) disrupted elongation of PolII transcription (Caizzi et al., 2021). It was also observed that phosphorylated PolII Ser5 interacts with spliceosomal intermediates during transcription elongation (Nojima et al., 2018). These observations lead us to speculate that the hyperphosphorylation of PolII Ser5 in hnRNPU-AID KO cells and the increased nascent RNA elongation and splicing were connected by increased spliceosome activities, in particular that of U2 snRNP. We hypothesised that hnRNPU-AID KO upregulated mature 17S U2 snRNP activity, which can be indicated by the dissociation of HIV-Tat-SF1 from 17S U2 snRNP which occurs during the formation of the branch point helix between U2 snRNA and the intronic sequences (marked by U2A′, U2B proteins, and SF3 A and B complexes) (Zhang et al., 2020).

Changes in HIV-Tat-SF1 association with 17S U2 snRNP were investigated by co-immunoprecipitation (co-IP) with a monoclonal anti-U2A′ antibody in DNase treated HCT116 and hnRNPU-AID KO cell lysates. SF3B1 was selected as the 17S U2 snRNP marker, as it is the component specifically inhibited by PlaB.

U2A′ co-immunoprecipitation showed that interactions between HIV-Tat-SF1 and U2A′ were weaker following hnRNPU KO, while that between SF3B1 and U2A′ were largely unchanged (Figure III.9.1A). This suggested that more active mature 17S U2 snRNP was present in hnRNPU-AID KO cells. Notably, U2A′ interaction with PolII Ser5P and GTF2F1 (a GTF component in the PolII pre-initiation complex, PIC) were noticeable only in hnRNPU-AID KO cells. Whilst there was increased Ser5P PolII present in the hnRNPU KO input sample for the immunoprecipitation, this is also reflected in an increased Ser5P PolII association with the U2 snRNP component U2A′. The increased association of U2 snRNP with Ser5P PolII is consistent with the increased splicing activity observed in cells following hnRNPU depletion.

In addition, Dr. Ivaylo Yonchev discovered that hnRNPU and SF3B1 interacted independently of RNA in an over-expression FLAG co-IP of hnRNPU in HEK293T cells (Wilson Lab unpublished data, Figure III.9.1B), in contrast to FUS, HNRNPK and U2A′, which lost their interactions with SF3B1 upon RNase treatment. This showed

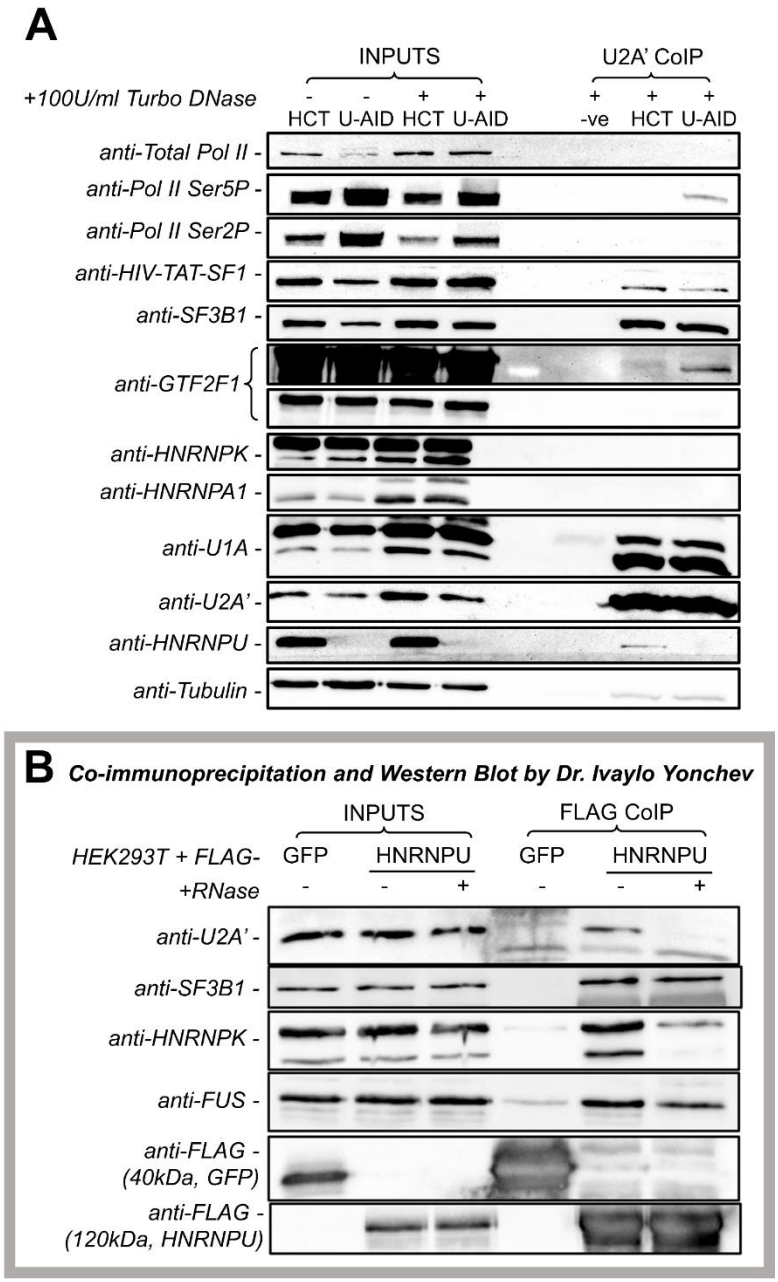

Figure III.9.1.HIV-TAT-SF1 and U2A′ interacted less following hnRNPU-AID KO, and hnRNPU interacted with SF3B1 independent of RNA.

A, Western blot of U2A′ co-IP showing decreased interaction between HIV-TAT-SF1 and U2A′ in hnRNPU-AID KO (U-AID) cells while SF3B1-U2A′ interaction stayed similar. Note that PolII Ser5P and GTF2F1 interactions with U2A′ were only noticeable in hnRNPU-AID KO cells, and hnRNPU co-IP with U2A′ in HCT116 cells was slightly more apparent than that of HNRNP K and A1. Tubulin is shown as input loading control and U2A′ is shown as co-IP loading control.

B, FLAG co-IP and Western blot by Dr. Ivaylo Yonchev showing SF3B1 interact with hnRNPU independent of RNA, in contrast to U2A′, HNRNPK and FUS. Input lysates were HEK293T cells transfected with either FLAG-tagged GFP or hnRNPU, and conditionally treated with RNase. FLAG signals of FLAG-GFP and hnRNPU are shown as input and co-IP loading

control, and equal input loading are seen across U2A′, FUS and SF3B1.that there is a potential direct and specific, protein-protein interaction between hnRNPU and SF3B1.

We further investigated whether PlaB inhibition of SF3B1 would revert the increases in PolII phosphorylation and chromatin RNA splicing ratios in hnRNPU-AID KO cells. Cells were treated with 100nM of PlaB for 3 hours, analysed with Western blot for PolII phosphorylation in either chromatin-intact or benzonase-digested total cell lysate (Figure III.9.2A), and with fractionation followed by RT-qPCR on chromatin RNA for spliced (exon-exon) to unspliced (exon-intron) ratios on selected introns of HNRNPH3 and SAT1 (Figure III.9.2B and C).

PlaB treatment noticeably reduced PolII Ser2 and 5 phosphorylation in benzonase-treated samples, while the effect was less obvious in non-enzyme lysed samples (Figure III.9.2A). With the benzonase treated samples, the excess Ser5 phosphorylation observed following hnRNPU loss was reversed with PlaB treatment, consistent with chromatin engaged Ser5p phosphorylated PolII being the species associated with the spliceosome (Nojima et al., 2018). Noticeably this effect was less obvious in the non-benzonase treated samples, suggesting these samples are enriched for PolII not engaged in co-transcriptional splicing. Whilst the increased Ser2P was observed clearly in the non-benzonase treated samples as reported at the beginning of this chapter, there was no obvious difference in Ser2P levels with or without hnRNPU in benzonase treated samples and in both cases the levels were reduced following PlaB treatment. The suppression of PolII Ser2 and 5 phosphorylation by PlaB in the benzonase treated samples is consistent with the reported repression of PolII elongation by PlaB (Caizzi et al., 2021).

Splicing analysis by RT-qPCR (Figure III.9.2C) showed that PlaB significantly reduced splicing of selected introns from SAT1 and hnRNPH3 in HCT116 and hnRNPU-AID KO cells. In both cases, the enhanced splicing observed following hnRNPU loss was effectively reversed by the PlaB treatment.

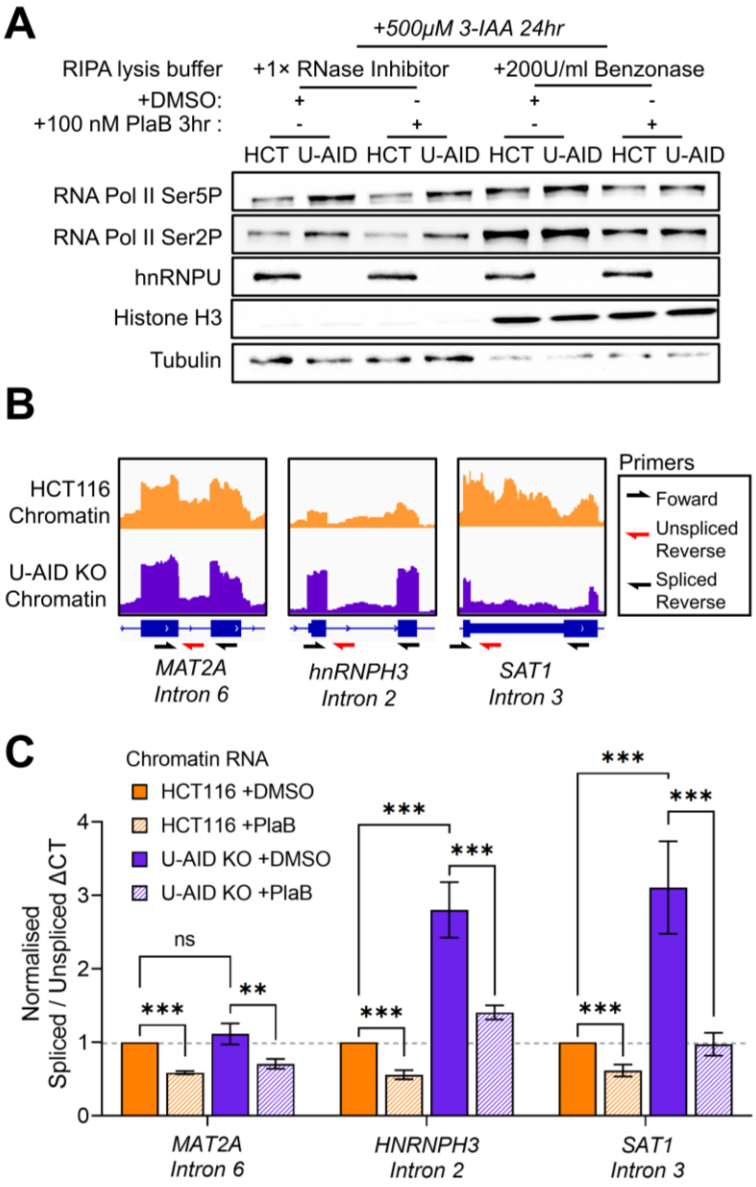

Figure III.9.2. PlaB treatment act oppositely to the upregulation of PolII phosphorylation and chromatin RNA splicing in hnRNPU-AID KO cells.

A, Western blot of PlaB treated HCT116 and hnRNPU-AID (U-AID) KO cell lysates with chromatin either protected with RNase inhibitor or digested by benzonase. PolII Ser2P and Ser5P levels in benzonase-treated samples reduced noticeably in both cell conditions. Loading controls were tubulin and histone H3 for chromatin-intact and depleted samples, respectively. Results were repeated in n=4 biological replicates.

B, Schematics of primer designs for spliced and unspliced transcripts from pre-mRNA of MAT2A intron 6, HNRNPH3 intron 2 and SAT1 intron 3 against their representative HCT116 and hnRNPU-AID (U-AID) KO chromatin RNA-seq read depth visualised in IGV.

C, Relative ΔCT between RT-qPCR of spliced and unspliced transcripts across MAT2A intron 6, HNRNPH3 intron 2 and SAT1 intron 3, normalised to DMSO treated HCT116 in each biological replicate. PlaB treatment reversed the significantly increased splicing of HNRNPH3 intron 2 and SAT1 intron 3 in hnRNPU-AID (U-AID) KO chromatin. MAT2A intron 6 splicing was inhibited as well, despite not significantly upregulated in hnRNPU-AID KO chromatin according to RT-qPCR. Data is shown as mean ± s.d. of n=4 biological replicates, analysed using two-tailed nonparametric t-tests; $^{**}P < 0.01$, $^{***}P < 0.005$.

III.10. Loss-of-function Walker B motif mutation increased hnRNPU chromatin affinity

To further investigate whether the ATP-binding activities of hnRNPU (Nozawa et al, 2017) are required for regulating splicing, we integrated T7-tagged hnRNPU isoform B wild-type (WT), loss-of-function Walker A motif mutant (GKT490-492AAA, denoted "ΔWA") and Walker B motif mutant (D561A, denoted "ΔWB") cDNAs into the hnRNPU-AID cell genome using the Sleeping Beauty transposon (Figure III.10.1A). The resulting cell lines were denoted as "SB-T7-HNRNPU WT/ΔWA/ΔWB"). SB-T7-HNRNPU WT, ΔWA and ΔWB cells had reduced PolII Ser2P levels compared to both HCT116 and hnRNPU-AID cells (Figure III.10.1B).

Dr. Ivaylo Yonchev had observed that ΔWA and ΔWB mutants of hnRNPUL1 are tightly associated with chromatin in a DNA dependent manner, and whole cell lysate did not reflect their true levels (Yonchev, 2021). To address potential issues with extractability of hnRNPU ΔWA and ΔWB mutants, we lysed and fully denatured the cells in SDS sample buffer, and analysed protein levels with Western Blot. Compared to hnRNPU WT and ΔWA, more T7-HNRNPU ΔWB proteins were detected in denatured SDS sample buffer lysates than in benzonase-treated RIPA whole cell lysates (Figure III.10.1B and C). This suggested that T7-HNRNPU ΔWB was relatively insoluble and difficult to extract without denaturation.

Interestingly, in the SDS denatured lysates, levels of FUS proteins in SB-T7-HNRNPU WT and ΔWA cells were similar as in HCT116, while there was less FUS in ΔWB cells, similar to hnRNPU-AID KO cells (Figure III.10.1C). The reversion of FUS levels by T7-HNRNPU WT and ΔWA, while not ΔWB, suggested to us that the less-chromatin associated hnRNPU is likely involved in expression control of the FUS protein.

To understand whether the insolubility of hnRNPU ΔWB is a result of chromatin association, we assayed the levels of T7-HNRNPU mutants in each cellular fraction. The results showed that T7-HNRNPU ΔWB predominantly localised to the chromatin and partially to nucleoplasm (Figure III.10.2A). T7-HNRNPU WT and ΔWA distributed across all cellular fractions, while ΔWA mutant showing a relative enrichment in chromatin than nucleoplasm. This further proved that hnRNPU ΔWB mutant is tightly associated with the insoluble chromatin fraction, while ΔWA mutant slightly improved chromatin affinity of hnRNPU. These results suggested that mutating

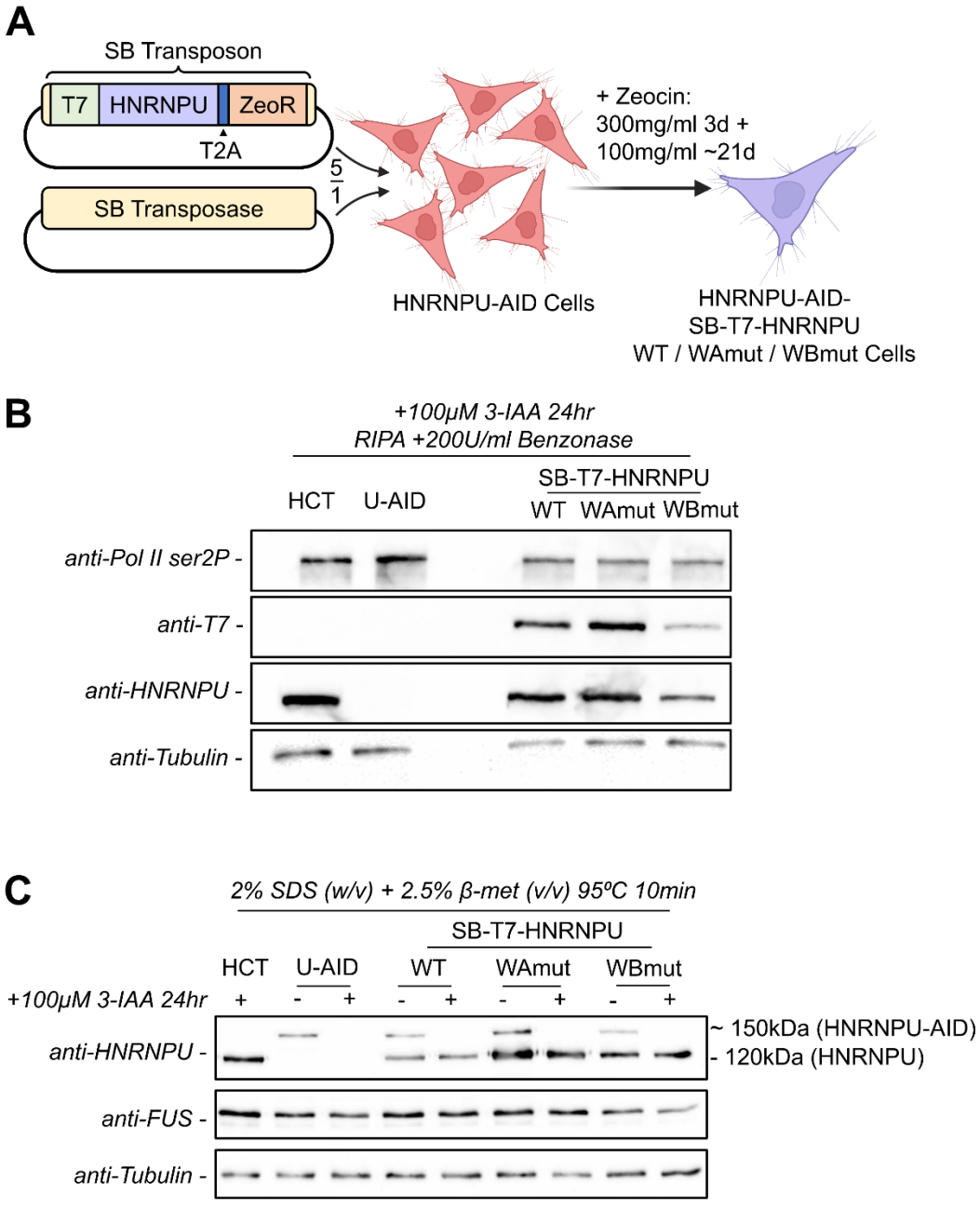

Figure III.10.1. Establishing T7-HNRNPU-Wild Type / Walker A / Walker B cell lines

A, Schematics of Sleeping Beauty (SB) transposon integration and selection of hnRNPU-AID cells complemented with T7-HNRNPU wild type (WT), Walker A mutant (ΔWA) and Walker B mutant (ΔWB) fused to a zeocin resistance (ZeoR) gene through T2A "autolytic" sequence.

B, Western blot of PolII ser2P, T7-tag and hnRNPU signals in 3-IAA treated, RIPA lysed and benzonase digested whole cell lysates of HCT116 (HCT), hnRNPU-AID (U-AID), SB-T7-HNRNPU WT/ΔWA/ΔWB cells.

C, Western blot of hnRNPU and FUS signals HCT116 (HCT), hnRNPU-AID (U-AID), and SB-T7-HNRNPU WT/ΔWA/ΔWB cells conditionally treated with 3-IAA and denatured in SDS sample buffer (composition and denaturation condition shown on top).

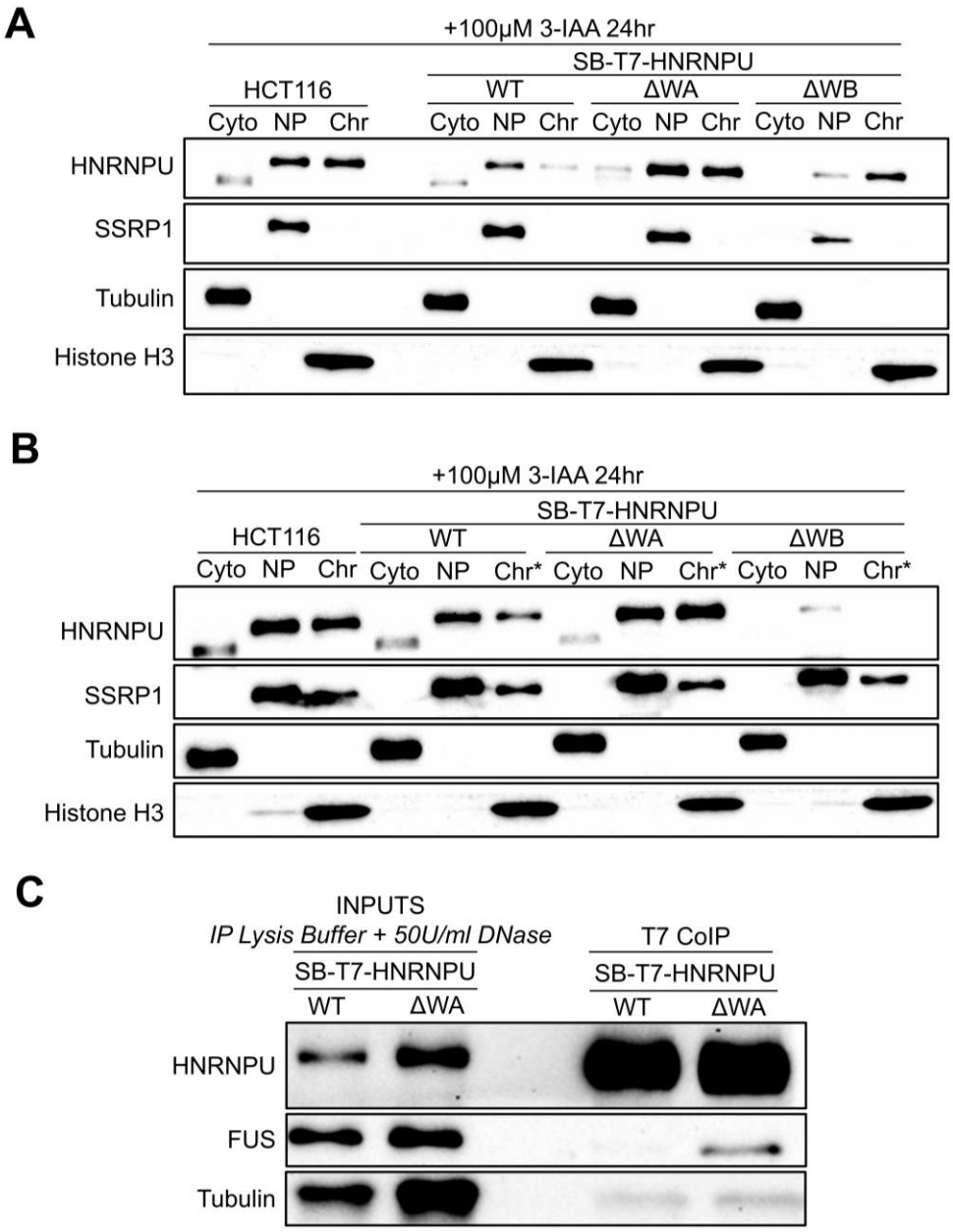

Figure III.10.2. hnRNPU ΔWB is tightly associated with chromatin and is insoluble in non- denaturing buffer

A, Western blot of hnRNPU in subcellular fractions (Cyto: cytoplasm, marked by tubulin; NP: nucleoplasm, marked by SSRP1; Chr: chromatin, marked by histone H3) in HCT116 and SB-T7-HNRNPU WT, ΔWA and ΔWB cells, showing enrichment of T7-HNRNPU ΔWB protein on the chromatin. Results were repeated in n=3 biological replicates.

B, Western blot of same subcellular fractions as in **A,** but chromatin fractions were benzonase digested in IP lysis buffer (Chr). Results were repeated in n=2 biological replicates.

C, Western blot of input and T7-coimmunoprecipitation (Co-IP) of SB-T7-HNRNPU WT and ΔWA cell lysates in IP lysis buffer, treated with 50U/ml DNase. T7 co-IP of T7-HNRNPU ΔWA contained notably more FUS protein than T7-HNRNPU WT. Results were repeated in two biological replicates.

the PPNK domain likely increased the chromatin association of hnRNPU, as is the case for hnRNPUL1 (Yonchev, 2021).

We further investigated whether ΔWA and ΔWB mutations affected the protein interactions of hnRNPU. Prior to this investigation we repeated the subcellular fractionation, but benzonase-digested the chromatin in IP lysis buffer instead of RIPA lysis buffer, to test whether we can extract T7-HNRNPU ΔWB while retaining the native activities of proteins. The results showed that, despite the benzonase digestion, chromatin-bound T7-HNRNPU ΔWB was not soluble in IP-lysis buffer (Figure III.10.2B). Therefore, we only studied protein interactions of T7-HNRNPU WT and ΔWA with co-IP.

As mutations in the Walker A motif of hnRNPUL1 altered its RNA-binding, we hypothesized that it would also affect RNA-mediated interactions of hnRNPU. We decided to conduct Co-IP with anti-T7 tag antibody in DNase digested SB-T7-HNRNPU WT and ΔWA cell lysates with RNase inhibitor. The Co-IP results showed that hnRNPU ΔWA associated with FUS more strongly than T7-HNRNPU WT (Figure III.10.2C). This result replicates what has been observed previously with hnRNPUL1 where the WA mutant is also chromatin restricted (Griffith, 2019) and interacts with FUS more than the WT protein.

III.11. Functional hnRNPU Walker A and B motifs are required to rescue inceased splicing levels in hnRNPU-AID KO

We next investigated whether the T7-HNRNPU ΔWA and ΔWB constructs restored splicing levels in hnRNPU-AID KO cells. Splicing analysis of MAT2A, hnRNPH3 and SAT1 was performed on chromatin RNA of SB-T7-HNRNPU WT, ΔWA and ΔWB cells at the same time as the PlaB treated HCT116 and hnRNPU-AID KO cells in Chapter III.9. Spliced to unspliced RT-qPCR signal ratios across hnRNPH3 exon 2 and SAT1 exon 3 were partially rescued in SB-T7-HNRNPU WT cells from the elevated levels in hnRNPU-AID KO cells, while such rescues were insignificant in SB-T7-HNRNPU ΔWA cells and not observed in ΔWB cells (Figure III.11.1).

Overall the data presented on the global upregulation of nascent splicing in hnRNPU KO cells in Chapter III.6 and the hnRNPU interaction with SF3B1 we observed in Chapter III.9, suggest that wild type hnRNPU is a splicing inhibitor, which potentially interacts with SF3B1 to repress 17S U2 snRNP. Judging by the lack of splicing rescue in SB-T7-HNRNPU ΔWA and ΔWB cells, this inhibition likely requires functional P-loop motifs in the PPNK domain of hnRNPU.

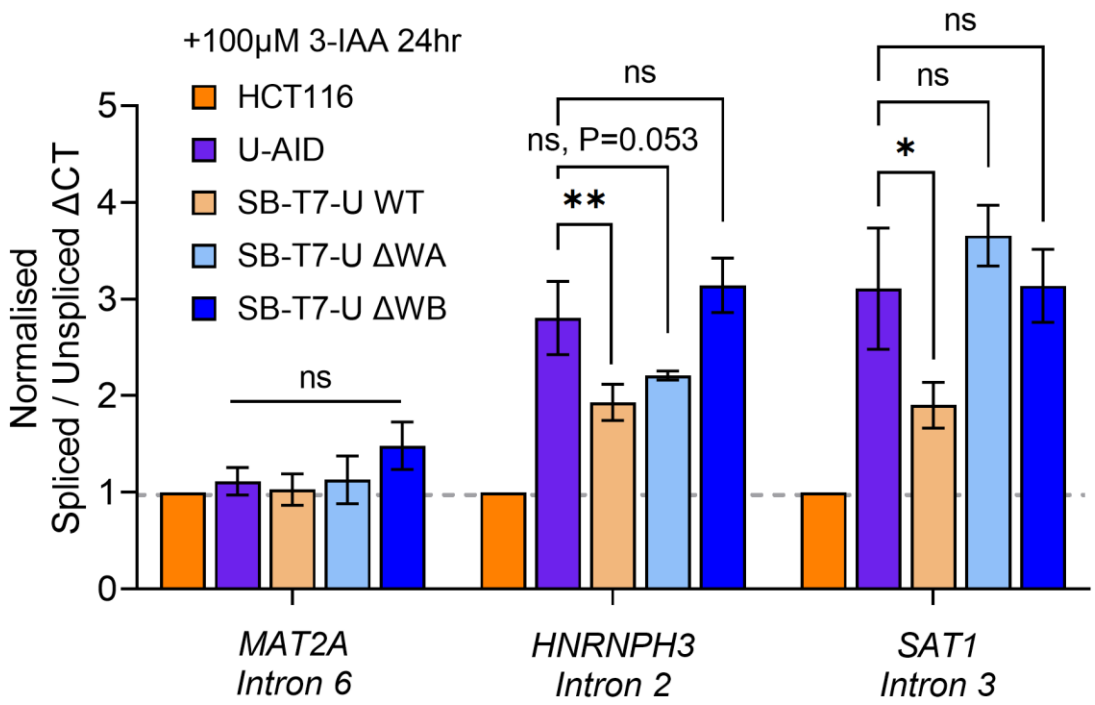

Figure III.11.1. Elevated splicing levels in hnRNPU-AID KO was not significantly rescued by ΔWA and ΔWB mutants.

Bar chart showing ΔCT between spliced to unspliced RT-qPCR signals across MAT2A intron 6, hnRNPH3 intron 2 and SAT1 intron 3 in 3-IAA treated HCT116, hnRNPU-AID (U-AID) and SB-T7-HNRNPU(U) WT, ΔWA and ΔWB cells, normalized to HCT116 samples in each replicate. Splicing across MAT2A intron 6 did not differ significantly between all tested cell lines, while increased splicing in hnRNPU-AID KO cells across hnRNPH3 intron 2 and SAT1 intron 3 were partially rescued in SB-T7-HNRNPU WT cells and not significantly rescued in ΔWA and ΔWB cells. Data are anomaly-trimmed mean ± s.d. of 4 biological replicates, analysed with two-tailed unpaired student's t-test; ns $P > 0.05$, P*$P < 0.05$, **$P < 0.01$.

III.12. Summary

In this chapter, we investigated the transcription and splicing effects of hnRNPU KO in an AID cell line established in the Wilson lab by Dr. Catherine Heath. We found consistent increases in PolII Ser5 phosphorylation and a shift in its nuclear distribution following hnRNPU-AID KO. We also observed that inhibition of PolII transcription dissociated hnRNPU from chromatin. In TT-Seq of hnRNPU-AID KO cells, we observed global upregulation of transcription in multi-exon protein coding genes, with a weak bias towards AT-rich genes, and global downregulation of snRNA gene transcription. Earlier work from Dr Ivaylo Yonchev in the laboratory had also shown that hnRNPU loss leads to widespread transcriptional pause release on protein coding genes. Together these results indicate that hnRNPU globally inhibits RNA polymerase II transcription of protein coding genes and that this effect is in part due to inhibition of transcriptional pause release. These data are consistent with early *in vitro* experiments which showed that depletion of hnRNPU allowed transcription readthrough of the HIV LTR, a well-known transcriptional pause site (Kim & Nikodem, 1999).

We then revisited hnRNPU-AID KO subcellular fraction RNA-Seq data produced by Dr. Heath and found global upregulation of splicing on the chromatin-bound RNAs, which was confirmed in the TT-Seq data. Interestingly, this splicing upregulation spanned across both protein coding and non-coding RNAs with a particularly striking increase in the splicing of lncRNAs. However, many lncRNAs showed reduced overall levels following loss of hnRNPU, suggesting their release from the chromatin environment triggered degradation by the exosome. In the case of the lncRNA FIRRE, it was confirmed that its stability was decreased, despite observing increased splicing. We further showed that in hnRNPU syndrome patients, the premature stop codons found in one allele led to an overall reduced hnRNPU protein level in these patient cells. Consistent with the reduced levels of hnRNPU found in patient cells, we observed a global upregulation in splicing efficiency in patient cells. This perturbation in splicing efficiency may lead to mis-splicing events such as aberrant exon inclusion or skipping, perturbing normal cellular functions, triggering the disease.

Following on from Xiao et al (2012)'s observations that hnRNPU knockdown increased the levels of the mature 17S U2 snRNP, we found that the 17S U2 snRNP had less HIV-

Tat-SF1 associated with it following hnRNPU-AID KO. Since HIV-Tat-SF1 release from U2 snRNP is essential for U2 to form a stable branch point helix interaction with the intron (Zhang et al., 2020), this result is consistent with the general increase in splicing we observe following hnRNPU loss. We also observed RNA-independent interactions between hnRNPU and SF3B1, subunit of 17S U2 snRNP inhibited by HIV-Tat-SF1.

Recent work has shown that the U2 snRNP is required for efficient transcriptional elongation and inhibition of the SF3B1 component of the U2 snRNP by PlaB treatment blocks transcriptional elongation) and reduces the levels of P-TEFb, the PolII Ser2P kinase, associated with genes (Caizzi et al., 2021). We were able to revert the PolII hyperphosphorylation and splicing increases in hnRNPU-AID KO by inhibiting SF3B1 function using PlaB. Together these results indicate that the increased transcription and hyperphosphorylation of PolII we observe following depletion of hnRNPU may be a direct result of its ability to suppress the U2 snRNP and subsequent splicing. Since hnRNPU associates with SF3B1 even in the presence of RNAse it is possible that hnRNPU directly targets SF3B1 as does PlaB, although we cannot rule out the possibility that hnRNPU acts earlier in spliceosome assembly, possibly disrupting the formation of the E complex.

IV. Analysis of roles of hnRNPUL1 in PolII transcription

The proteins hnRNPU-Like 1 (hnRNPUL1) and hnRNPU share similar protein structure and domain sequences, apart from the C-terminus intrinsically disorder region (IDR), which is significantly extended in hnRNPUL1. Earlier investigations of hnRNPUL1 focused on DNA damage repair, as they addressed its association with p53 and Parp1-mediated DNA double-strand breaks repair (Barral et al., 2005; Hong et al., 2013; Polo et al., 2012). A potential role in transcription was noticed more recently, as a report showed transcription activation by fused MEF2D-hnRNPUL1 protein in acute lymphoblastic leukemia (Ohki et al., 2018), and a ChIP-seq study showed that hnRNPUL1 is enriched over snRNA genes (Xiao et al., 2019). In addition, a zebrafish study showed that hnRNPUL1 mutants altered gene expression and splicing, and disrupted skeletal development (Blackwell et al., 2022).

In the Wilson Lab, Dr. Llewllyn Griffith established a doxycycline-inducible AID cell line to conditionally knock out hnRNPUL1 and identified hnRNPUL1 as an essential gene and a regulator of snRNA transcription and processing (Griffith, 2019). He discovered that loss of hnRNPUL1 disrupted Cajal Bodies and mislocalised SART3, both of which are important elements for snRNA and spliceosome maturation. He also found that truncating the low-complexity C-terminal domain (CTD) of hnRNPUL1, composed of an RGG box and IDR, abolished its interactions with PolII and spliceosomal components. By co-immunoprecipitation followed by mass-spectrometry, Dr. Griffith further identified strong interactions between hnRNPUL1 and ALS genes FUS and EWSR1, which were reported to regulate PolII pause-release through their low-complexity domains (Kwon et al., 2013). He also discovered a burst of immediate early response genes (IEG) transcription within 2 hours of hnRNPUL1-AID depletion, suggesting cellular stress induced by hnRNPUL1 loss.

Dr. Ivaylo Yonchev further studied the effects of hnRNPUL1 loss in the AID cell line and identified similar increases in histone RNA misprocessing and enhancer RNA and IEG gene transcription in hnRNPUL1-depleted cells as in FUS knocked down cells (Yonchev, 2021). Carmen Apostol investigated the liquid-liquid phase separation of hnRNPUL1 CTD, and found that, similar to the low complexity domains of FUS, it phase separated on its own as well as co-localized with phase-separated droplets of PolII CTD (Wilson Lab unpublished; Kwon et al., 2013).

In this chapter, we investigated the effect of hnRNPUL1-AID KO on U7 snRNA by northern blotting to investigate the observed increase in histone mRNA misprocessing. We also investigated the transcription activating potential of hnRNPUL1 domains by dual luciferase assays and identified IDR as the sole transcription activator domain of hnRNPUL1. We then investigated the transcription and co-transcriptional splicing effects of hnRNPUL1 loss by TT-Seq, where we found that hnRNPUL1-AID KO downregulated transcription globally and demonstrated opposing effects to hnRNPU-AID KO on snRNAs and IEGs. Lastly, we attempted to resolve basal-degradation observed in the hnRNPUL1-AID cell line by adopting the newly developed AID2 system, but still observed similar reduced basal levels of hnRNPUL1.

IV.1. HNRNPUL1-AID KO down-regulated U7 snRNA levels

The hnRNPUL1-AID cell line was established by stably integrating a copy of doxycycline-inducible TIR1 cDNA into the HCT116 genome at the AAVS1 safe harbour locus and tagging both copies of hnRNPUL1 genomic DNA with full-length IAA-17 AID tag at the C-termus (Griffith, 2019) (Figure IV.1.1A). After activating TIR1 expression with 48-hr doxycycline treatment, it fully knocked out hnRNPUL1-AID within 2-hours of treatment with 50μM 3-IAA (Figure IV.1.1A), this condition is referred to as "HNRNPUL1-AID KO" hereafter. It is worth noting that we observed reduced basal AID-tagged hnRNPUL1 levels before doxycycline and 3-IAA treatment compared to HCT116 cells, similar to the basal reduction of hnRNPU in hnRNPU-AID cell line shown in Chapter III.1.

The expression of replication-dependent (RD) histone genes was previously reported to be repressed by hnRNPUL1 in cells arrested at the G1/S phase of the cell cycle (Ideue et al., 2012). This study also showed that hnRNPUL1 bound U7 snRNA, a result more recently confirmed by analysis of eCLIP data (Wilson lab, unpublished). In this study they, (Ideue et al., 2012), used siRNA to knockdown hnRNPUL1 and suggested that there were no real changes in U7 snRNA levels. However, we were concerned that this result may have been due to inefficient depletion of hnRNPUL1 using RNAi, since we observed significant misprocessing of histone mRNAs leading to their polyadenylation following depletion of hnRNPUL1. Moreover, we observed a strong ChIP-Seq signal downstream of replication dependent histone genes, suggesting a potential role in 3′ processing of these RNAs (Yonchev, 2021). Finally, we had observed reduced transcription of the PolII transcribed spliceosomal snRNAs (U1, U2, U4, U5) following

hnRNPUL1 depletion suggesting hnRNPUL1 might also regulate transcription of U7 snRNA (Griffith, 2019; Yonchev, 2021). Together these data suggested that the defects in histone mRNA 3′ processing we observed might be partly explained by reduced levels of U7 snRNA, as U7 snRNP is required for the appropriate cleavage and post-transcriptional processing of nascent histone mRNAs (Chapter I.2.2).

To investigate U7 snRNA levels we used Northern blots to assay the changes in U7 snRNA upon hnRNPUL1-AID KO, since U7 snRNA was too small to detect using qRT-PCR. Against PolIII-transcribed U6 snRNA, the relative levels of U7 snRNA significantly reduced by 59.3% ± 7.2% in hnRNPUL1-AID KO cells compared to HCT116 (Figure IV.1.1B and C). The reduced levels of U7 snRNA would be expected to contribute to the increased histone mRNA misprocessing we observe following hnRNPUL1 KO. However, we cannot exclude the possibility that hnRNPUL1 also plays a more direct role in histone 3′ end processing since it associates directly with the U7 snRNP.

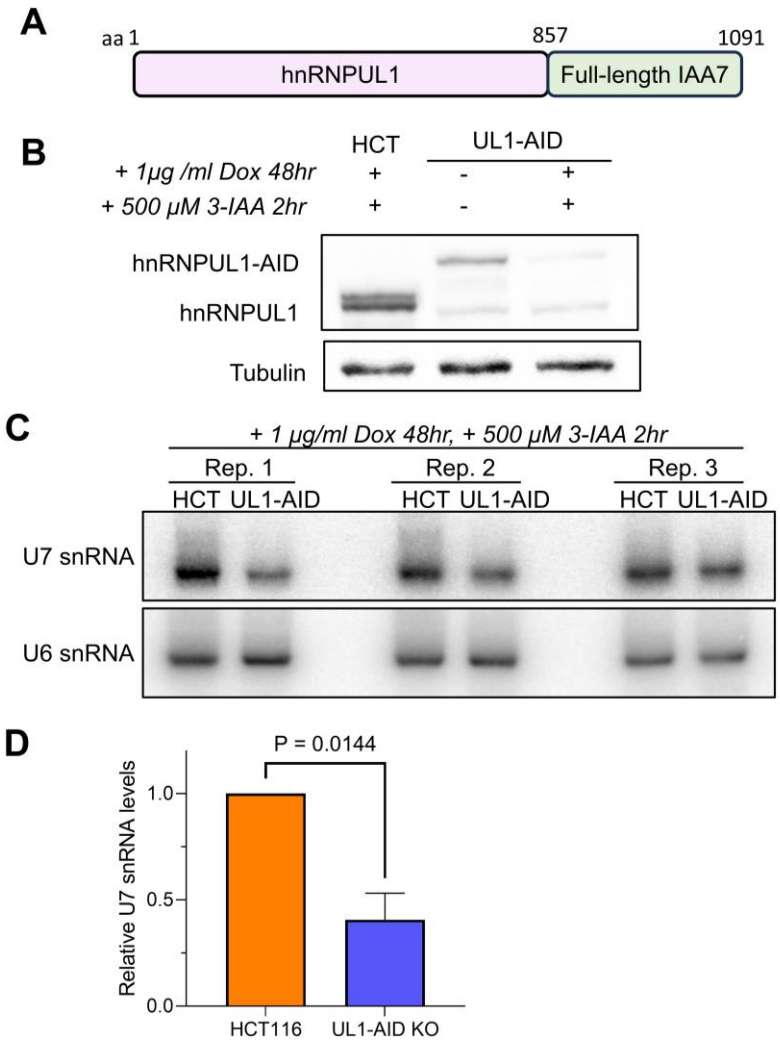

Figure IV.1.1. U7 snRNA levels decreased in doxycycline and 3-IAA induced hnRNPUL1-AID KO cells.

A, A sketch of the hnRNPUL1-AID tagged protein structure. The 857 residue hnRNPUL1 is tagged with a 234 residue full length IAA7 at the C-terminus

B, Western blot of HCT116 (HCT) and hnRNPUL1-AID (UL1-AID) cells conditionally treated with doxycycline (Dox) and 3-IAA. hnRNPUL1-AID is larger (~150kDa) than wild type hnRNPUL1 (~130kDa).

C, Three replicates of northern blots of U7 snRNA relative to U6 snRNA in Dox and 3-IAA treated HCT116 and hnRNPUL1-AID cells. Intensity of U7 snRNA is lower in hnRNPUL1-AID cells

D, Bar chart showing relatively reduced intensities of U7 snRNA normalised to U6 snRNA in HCT116 and hnRNPUL1-AID cells, measured in ImageLab. Data is mean ± s.d. of 4 biological replicates, P-value shown (0.0144) was calculated using unpaired student's t-test.

IV.2. HNRNPUL1 IDR activated PolII transcription

We have previously observed that hnRNPUL1 associates with FUS and both proteins are implicated in histone gene expression (Raczynska, 2015; Griffith, 2019). Both hnRNPUL1 and FUS have low complexity domains that constitute an IDR and RNA binding motifs (RGG boxes in hnRNPUL1, RRM and RGG boxes in FUS). Dr. Apostol showed that they phase separate similarly with PolII CTD in vitro (Apostol, 2023) and Dr. Griffith showed that low complexity domains of hnRNPUL1 are required for its interaction with PolII and FUS *in vivo* (Griffith, 2019). Moreover, the IDR of FUS is well known to have transcription activation potential (Kwon et al, 2013).

In addition, Dr. Ivaylo Yonchev identified enrichment of hnRNPUL1 ΔWA and ΔWB mutants over enhancer loci in CUT&RUN experiments (Yonchev, 2021). In hnRNPUL1-AID KO cells, he observed downregulation of transcription 500bps up and downstream of snRNA gene loci measured by mNET-seq, and reductions in levels of short-lived enhancer RNAs in chromatin RNA-seq (Yonchev, 2021). Hence, we hypothesised that hnRNPUL1 interacted with PolII CTD through its low-complexity domains to activate productive elongation of intronless genes, such as snRNA and enhancer RNAs. Furthermore, several studies have identified chromosomal translocations involving the C-terminal half of hnRNPUL1 fused to the DNA binding domain of the MEF2D transcription factor. These translocations lead to a fusion protein with transcription factor activity which triggers B cell acute lymphoblastic leukaemia (Gu et al, 2016; Yasuda et al., 2016).

We then used a dual-luciferase reporter assay to explore the transcriptional activation potential of hnRNPUL1, where the intronless reporter gene is robust against changes in splicing and only reports on transcription activation. We first constructed Gal4 DNA binding domain (DBD) fused to various sections of hnRNPUL1 (Figure IV.2.1B). Then we tested their potential at activating expression of a Gal4 DBD specific promoter driven intronless firefly luciferase cDNA in a dual-luciferase assay in HEK293T cells, normalised to co-transfected CMV-promoter driven Renilla luciferase. We used hnRNPU SPRY-RGG domains as an IDR-less negative control, and FUS IDR and VP16 as positive controls. The levels of expression of these constructs were verified by Western blot (Figure IV.2.1A), however VP16 appeared hardly extractable using the RIPA + benzonase lysis buffer. The results showed that, amongst all Gal4-HNRNPU/UL1 constructs, only Gal4-HNRNPUL1 IDR activated firefly luciferase

expression by 2.27-fold over the background Gal4 DBD, significant (P=0.0104) but less than positive controls Gal4-FUS IDR (12-fold) and Gal4-VP16 (370 fold) constructs (Figure IV.2.1B). None of the Gal4 DBD-HNRNPU/UL1 constructs that contained SPRY-PNK or RGG domains activated the expression of the luciferase cDNA more than background. These data indicate that hnRNPUL1 IDR in isolation can act as a transcriptional activation domain, as seen in B-ALL patients with MEF2D fusions (Gu et al, 2016). However, in the context of larger constructs including the SPRY-RGG domains this activity is lost.

Considering this result, it is possible that the IDR of hnRNPUL1 mainly drives the association of hnRNPUL1 with PolII CTD and presumably subsequent loading onto nascent RNAs, rather than to enhance transcription. Alternatively, it may act as a transcription factor but only at certain promoters such as those found at snRNA genes and the specific features of these promoters are not recapitulated in the reporter assay we have used here.

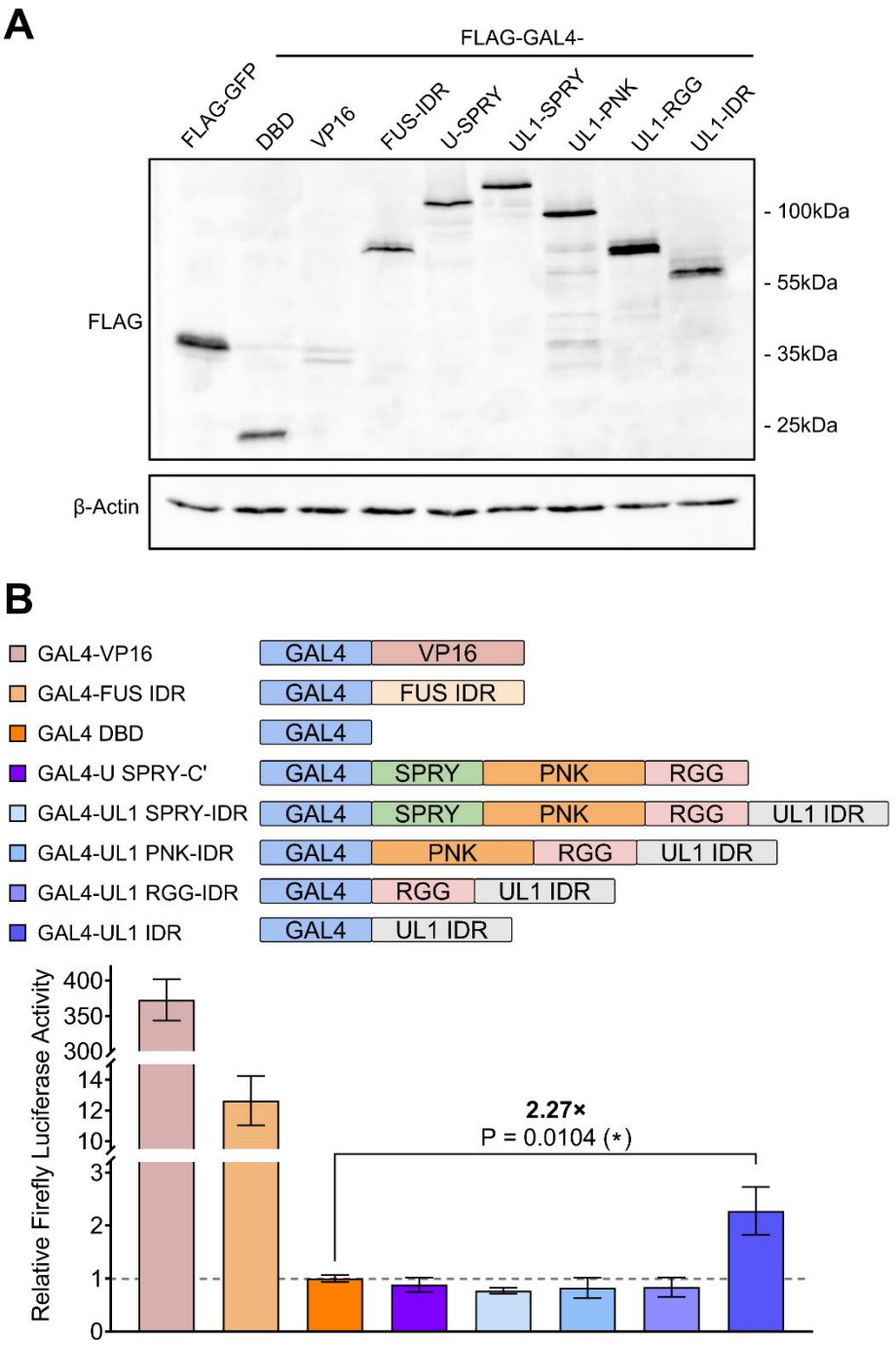

Figure IV.2.1. Gal4-HNRNPUL1 IDR activated firefly luciferase expression.

A, Western blot of whole cell lysates of HEK293T cells transfected with various FLAG tagged Gal4 fused constructs (U-HNRNPU, UL1- hnRNPUL1). FLAG-GFP is used as transfection control and β-actin is probed as the loading control.

B, Levels of firefly luciferase activities in transfected cell lysates, normalized internally to Renilla lucifer ase activity and to the mean of FLAG-Gal4 DBD transfected negative control lysate in each replicate. Domain structures of fused constructs are shown next to corresponding keys. Data is mean ± s.d. of n=4 biological replicates, analysed with unpaired student's t-test; **P* < 0.05.

IV.3. HNRNPUL1-AID KO downregulates PolII transcription globally and alters co-transcriptional splicing

Existing data from the Wilson Lab provided conflicting views for the transcriptional effects of hnRNPUL1-AID KO, as Dr. Griffith showed reduced PolII association at snRNA loci (Griffith, 2019), while the internal-normalized mNET-seq by Dr. Ivaylo Yonchev showed a slight increase in PolII-bound transcripts up and down-stream of snRNAs (Yonchev, 2021). Dr. Ivaylo Yonchev also observed increased splicing of retained introns in a nuclear RNA-seq of hnRNPUL1-AID KO cells, which led us to question whether co-transcription splicing was disrupted. Therefore, we carried out TT-Seq in hnRNPUL1-AID KO cells alongside the hnRNPU-AID KO TT-Seq described in Chapter III.4. We expected to leverage the spike-in based normalization to identify global changes in transcription and co-transcriptional splicing.

TT-Seq reads were processed the same way as described in Chapter III.4. Differential analysis of spike-in normalized unspliced TT-Seq reads showed that hnRNPUL1-AID KO led to downregulated of transcription of all PolII transcribed biotypes, regardless of the number of exons (Figure IV.3.1A). Metagene analysis on multi-exon protein-coding genes showed that transcription was downregulated across their entire length (Figure IV.3.1B). Closer inspection of the PROMPT-TSS region showed that transcription is downregulation in both directions from the TSS (Figure IV.3.1C). We also noticed that, compared to HCT116, hnRNPUL1-AID KO TT-Seq read depth decreased to a greater extent in the initial peak of the TSS metagene. This may indicate a reduction in PolII initiation and paused transcripts, while pause-release and elongation are disrupted to a lesser degree.

The TT-Seq data also revealed that there was reduced transcription of histone genes (Figure IV.3.1D,E), in common with other protein coding genes. Therefore, the effects of hnRNPUL1 loss on histone genes are twofold, firstly we observe reduced transcription as revealed in the TT-Seq analysis and secondly, misprocessing of the histone mRNAs leading to increased polyadenylation (Yonchev, 2021). These phenotypes are similar to those reported for FUS (Raczynska et al., 2015), suggesting the two proteins may function together at histone gene loci, consistent with their ability to interact with each other (Griffith, 2019).

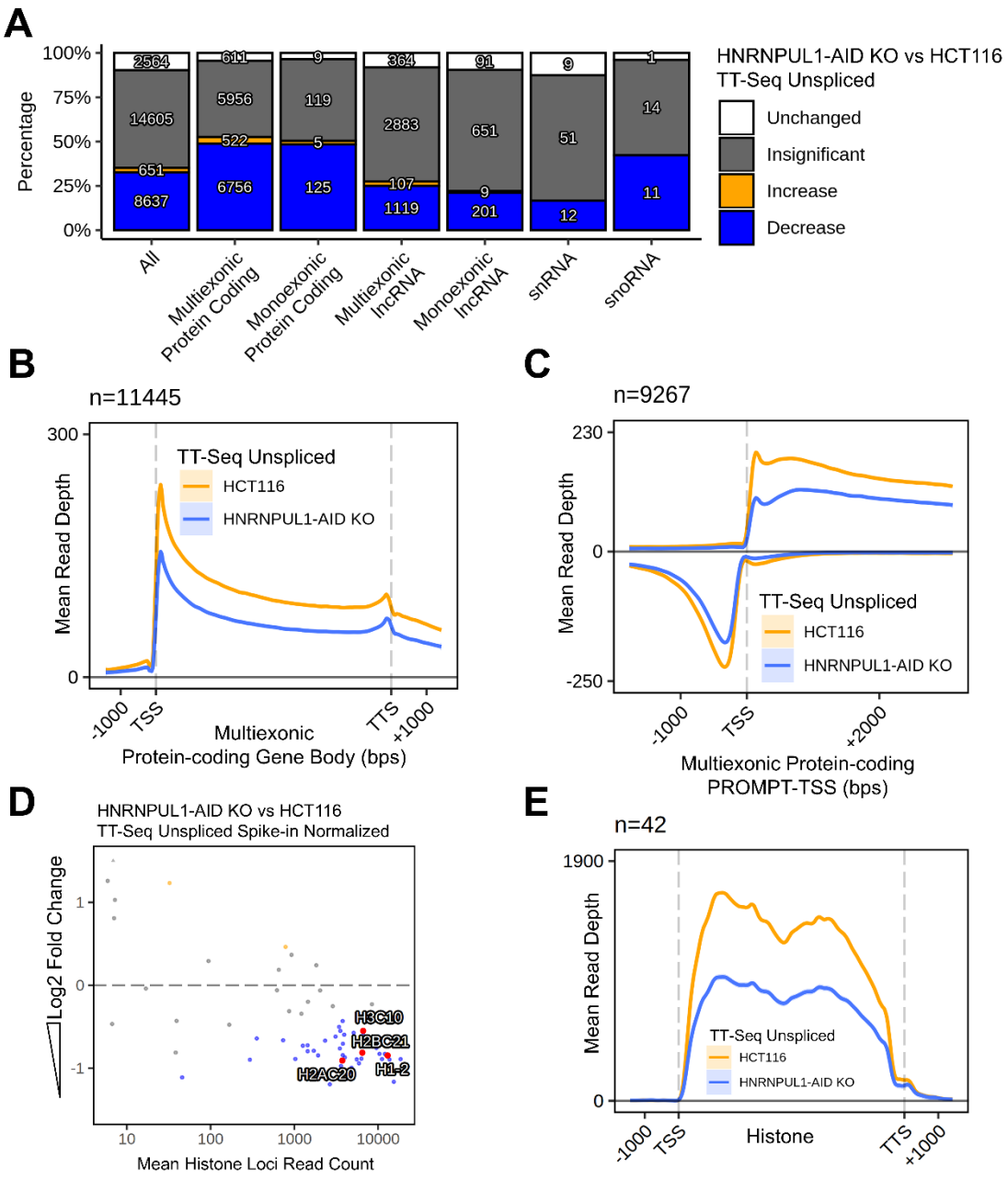

Figure IV.3.1. 48-hr IAA induced hnRNPUL1-AID KO downregulated transcription.

A, Stacked percentage bar chart of genes with differential unspliced aligned reads in each category of biotypes and exon counts. Labelled numbers are gene counts in each bar sector. Genes with changes at least adjusted p-value of 0.1 were considered significant, while changes with adjusted p-value more than 0.95 are considered unchanged.

B, C and E, Mean unspliced read depth of hnRNPUL1-AID KO TT-Seq across 11445 non-overlapping multi-exon protein-coding gene bodies (**B**), 9267 multi-exon protein-coding gene PROMPT-TSS region (**C**, 1500bps upstream and 3000bps downstream of TSS), and 42 histone gene loci (**E**). Decreases in TT-Seq read depth suggested that PolII transcription is globally downregulated in hnRNPUL1-AID KO cells.

D, MA plot of log2 fold changes in hnRNPUL1-AID KO TT-Seq unspliced read count across histone gene loci compared to in HCT116. A few examples of replication dependent histone mRNA genes are highlighted in red and labelled with their Ensembl gene symbols.

We then analysed the changes in co-transcriptional splicing in hnRNPUL1-AID KO cells, as described for hnRNPU-AID KO TT-Seq in Chapter III.6. We only found significant co-transcriptional splicing increases in 101 genes, and significant decreases in only 11 genes, out of 3347 sufficiently evidenced genes (Figure IV.3.2A). This suggested that the increase in co-transcriptional splicing is not global, unlike that caused by hnRNPU-AID KO.

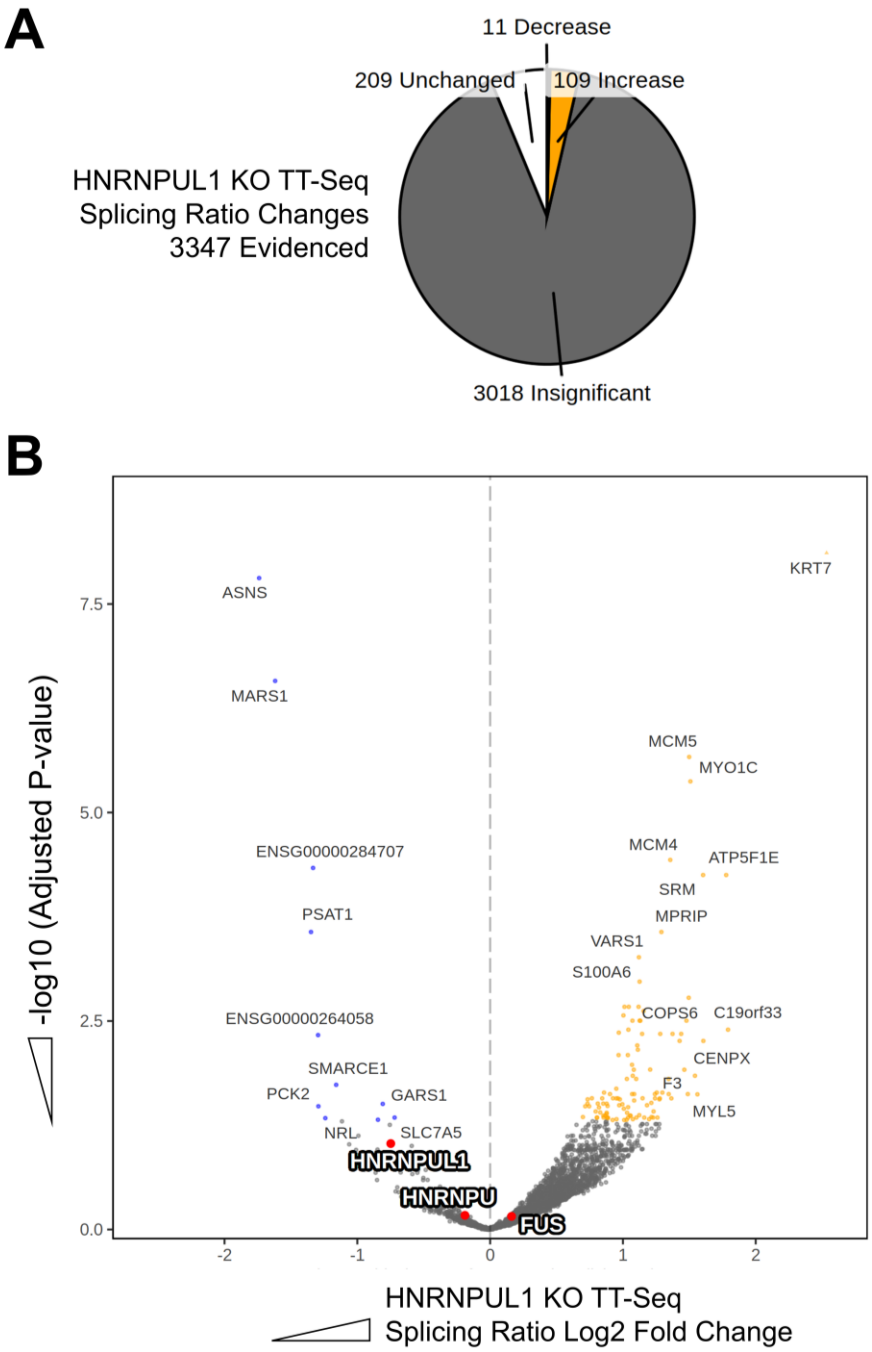

Figure IV.3.2. hnRNPUL1-AID KO upregulated splicing in a small fraction of genes

A, Pie chart showing changed in splicing ratios of 3347 sufficiently evidenced genes. The Majority (3018) of these genes showed insignificant changes (grey, 0.05<P<0.95), 209 were unchanged (white, P>0.95), 101 showed significant upregulation (orange, P<0.05) while only 11 showed significant decreases (blue, P<0.05) in splicing ratios.

B, Volcano plot of splicing changes in individual genes in hnRNPUL1 KO TT-Seq. The most significantly upregulated and downregulated genes were labelled and hnRNPU, hnRNPUL1 and FUS are highlighted.

IV.4. HNRNPUL1-AID KO downregulated processing of snRNA

Dr. Griffith observed hnRNPUL1-AID KO reduced the ratio between mature snRNA levels and nascent snRNA transcripts and concluded that snRNA processing is disrupted by loss of hnRNPUL1 (Griffith, 2019). I validated this observation using the TT-Seq data.

Metagene analysis across snRNA genes showed a reduction in read depth across the mature snRNA region of these genes, but to a lesser extent than that in protein-coding genes (Figure IV.4.1A). We also noticed that the read depth downstream of the mature 3′ end of the snRNA was nearly identical between hnRNPUL1-AID KO and HCT116 TT-Seq. This suggested that the ratio of nascent gene body RNA to downstream RNA was altered, indicative of a 3′ processing defect at snRNA gene loci. This result was consistent with the previous qRT-PCR analysis (Griffith, 2019). Interestingly this result is directly opposite that seen in the hnRNPU-AID KO TT-Seq, suggesting that the two proteins oppose each other's roles in regulating snRNA transcription and processing.

In differential analysis, we noticed that RNU4-1 was the most significantly downregulated amongst all snRNA loci. Read depth across this locus also showed the greater contrast between the annotated mature snRNA and the downstream nascent transcripts in hnRNPUL1-AID KO TT-Seq. This corresponded to Dr. Griffith's observation that, amongst snRNAs, hnRNPUL1 associated the strongest with U4 snRNA in RNA-immunoprecipitation (RIP) assay, and the levels of mature U4 snRNA were the most downregulated in hnRNPUL1-AID KO cells (Griffith, 2019).

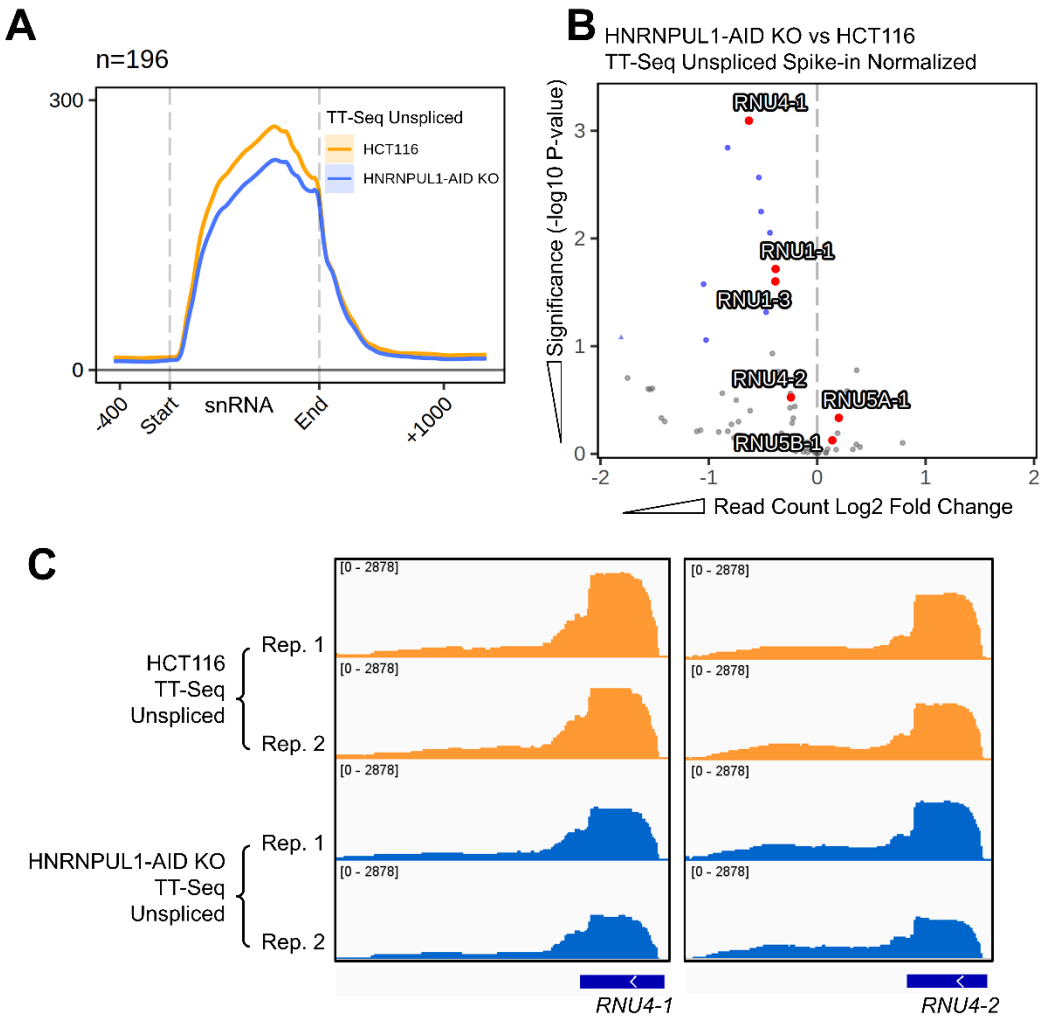

Figure IV.4.1. hnRNPUL1-AID KO reduced pre-snRNA processing.

A, Mean unspliced and spike-in normalized TT-Seq read depth across 500bps upstream and 1500bps downstream 196 snRNA loci in HCT116 and hnRNPUL1-AID KO TT-Seq. Read reduced in the annotated mature snRNA loci in hnRNPUL1-AID KO TT-Seq while stayed similar for 1000bps downstream.

B, Volcano plot showing significance of changes in snRNA genes TT-Seq read counts (y-axis) against the log2 fold changes (x-axis) between hnRNPUL1-AID KO and HCT116 cells. Curated RNU1, RNU4, and RNU5 loci are highlighted in red and labelled, showing RNU4-1 as the most significantly downregulated snRNA gene.

C, TT-Seq read depth across RNU4-1 and RNU4-2 loci in HCT116 and hnRNPUL1-AID KO cells, visualised and group-autoscaled in IGV. The read depths across annonated mature U4 snRNA gene body are lower in hnRNPUL1-AID KO TT-Seq, while downstream (to the left) are similar to HCT116.

IV.5. HNRNPUL1-AID KO upregulated transcription of IEGs

Dr. Griffith and Dr. Yonchev observed increased PolII deposition across immediate early genes NR4A1, EGR1, FOS and FOSB loci and RNA transcripts upon hnRNPUL1-AID KO by ChIP and mNET-seq (Griffith, 2019; Yonchev, 2021). Dr. Yonchev further discovered stronger association between 7SK snRNP and P-TEFb release factors RBM7 and SC-35, which suggested more 7SK snRNP mediated activation of PolII transcription by P-TEFb, a hallmark of a pro-survival response to genotoxic stress (Bugai et al., 2019). These results were consistent with the reported role of hnRNPUL1 in DNA double strand break repair (Polo et al., 2012) and suggested that hnRNPUL1 loss led to a stress response potentially due to defects in DNA damage.

To confirm this result, we manually inspected TT-Seq read depth in IGV, and found that transcription of NR4A1, EGR1, FOS, and FOSB increased in hnRNPUL1-AID KO cells, as we expected. We then extracted the same list of IEGs from Tullai et al (2007) as in Chapter III.5 from our TT-Seq differential analysis and identified transcription upregulation of a few other IEGs, such as EGR3, GEM, and ATF3. The activation of this subgroup of IEGs contradicted the general downregulation of protein-coding gene transcription in hnRNPUL1-AID KO and suggested that hnRNPUL1 loss induced transcription of stress response genes as suggested from earlier analysis by Griffith and Yonchev.

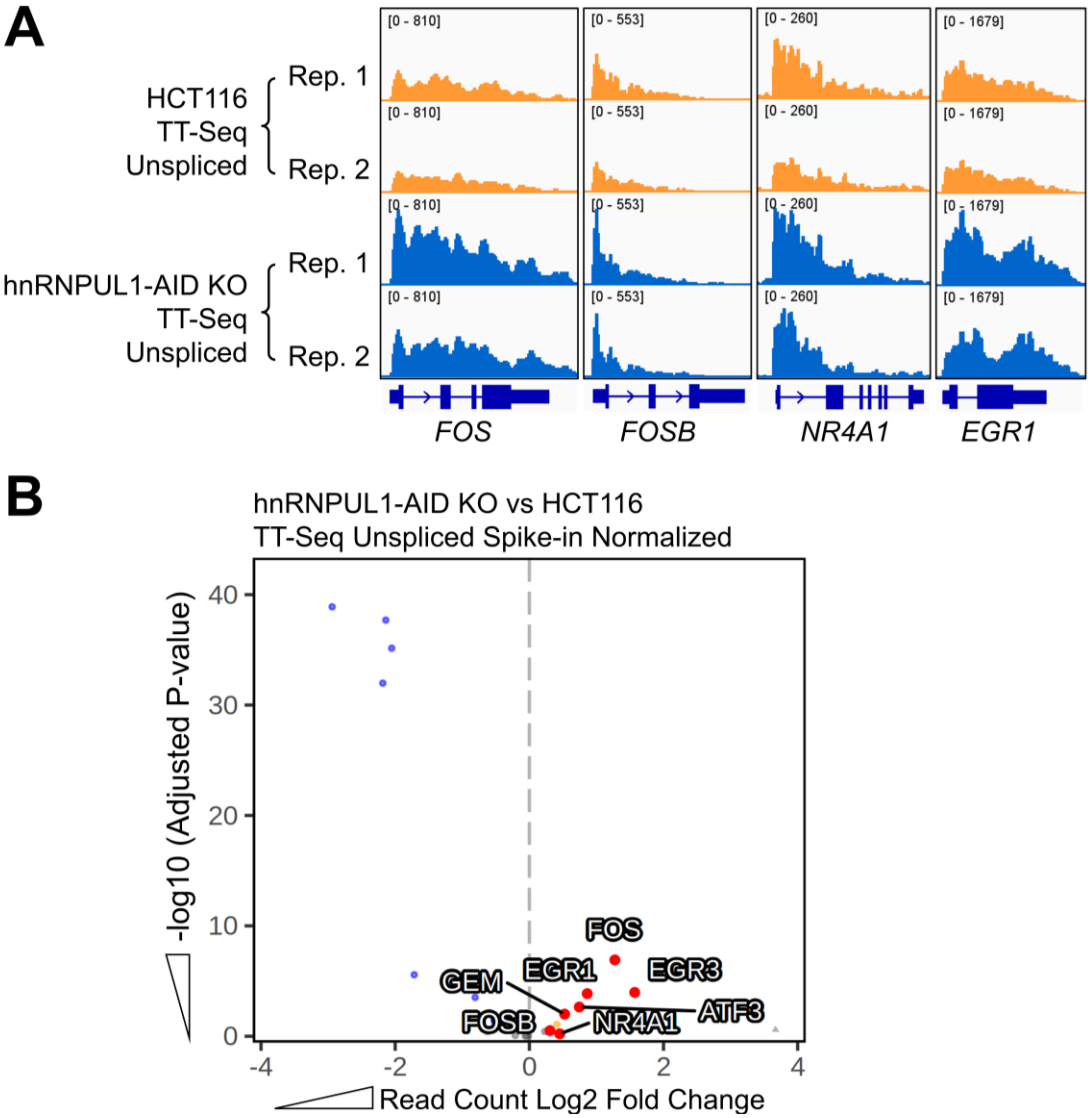

Figure IV.5.1. hnRNPUL1-AID KO upregulated transcription of IEGs.

A, Two biological replicates of hnRNPUL1-AID KO and HCT116 TT-Seq unspliced read depth aross IEG FOS, FOSB, NR4A1 and EGR1, visualised and autoscaled in IGV. Trancription of these genes are shown upregulated in hnRNPUL1-AID KO.

B, Volcano plot showing significance of changes in IEGs (y-axis) against the log2 fold changes (x-axis) of TT-Seq read counts in hnRNPUL1-AID KO against HCT116 cells. Significantly (adjusted P-value <= 0.1 upregulated IEGs were shown as yellow dots and downregulated IEGs were shown in blue. AP-1 family IEGs, including FOS, FOSB, GEM, NR4A1, ATF3, EGR1 and EGR3 are highlighted in red and labelled with gene names.

IV.6. An Attempt at Establishing hnRNPUL1-AID2 cell line

The basal degradation we observed in both hnRNPU-AID and hnRNPUL1-AID cell lines meant that some of the effects we observed maybe due to long term partial depletion of these proteins. Since this effect was most acute for hnRNPUL1, we decided to develop a different degron system which would hopefully minimize the effects of basal degradation. Variants of AID system have been proposed to improve tagging and degradation efficiencies (Natsume et al., 2016; Sathyan et al., 2019). After consideration and comparisons, we selected the AID2 system, which uses a TIR1 protein with a mutated auxin-binding site (F74G) and a redesigned auxin molecule (5-Ph-IAA) to achieve better specificity and sensitivity (Yesbolatova et al., 2020).

Using targeted homology-directed DNA repair (HDR) facilitated by Cas9 and an existing and tested guide RNA targeting the human AAVS1 loci, we integrated CMV promoter driven TIR1(F74G) cDNA into the AAVS1 loci of HCT116 genome. The genetically modified HCT116 cells were selected using puromycin against the co-integrated puromycin-resistant marker. Six viable colonies were tested using Western blot for the expression of TIR1, and two were positive (Figure IV.6.1A). Colony 3 expressed more TIR1 than colony 2 and was maintained as the HCT116-AID2 cell line.

We then used Cas9 and two HDR templates to insert mAID-P2A-antibiotic resistance gene-SV40 poly(A) signal sequences to the 3′ end of hnRNPUL1 coding region of HCT116-AID2 cells. The Cas9 was guided by a single guide RNA designed by Dr. Griffith for the original hnRNPUL1-AID system. The two HDR templates carried resistance genes against hygromycin and neomycin respectively, which co-expressed with hnRNPUL1 and isolated during translation by the P2A sequence (Figure IV.6.1B).

We then selected the successful editing of both hnRNPUL1 alleles using hygromycin and neomycin at the same time and obtained 24 viable colonies. We tested levels of hnRNPUL1 before and after 2nM 5-Ph-IAA treatment in 12 colonies, and found that, despite successful tagging and depletion of hnRNPUL1 in eight of them, the hnRNPUL1-mAID construct has a lower basal levels compared to hnRNPUL1 in wild type HCT116 cells, similar to that of hnRNPUL1-AID in the original doxycycline inducible AID system (Figure IV.6.1C). We maintained the hnRNPUL1-mAID KO colony 13 as the main hnRNPUL1-AID2 colony but decided to pause research on this cell line and search for possible reasons and solutions to the basal reduction.

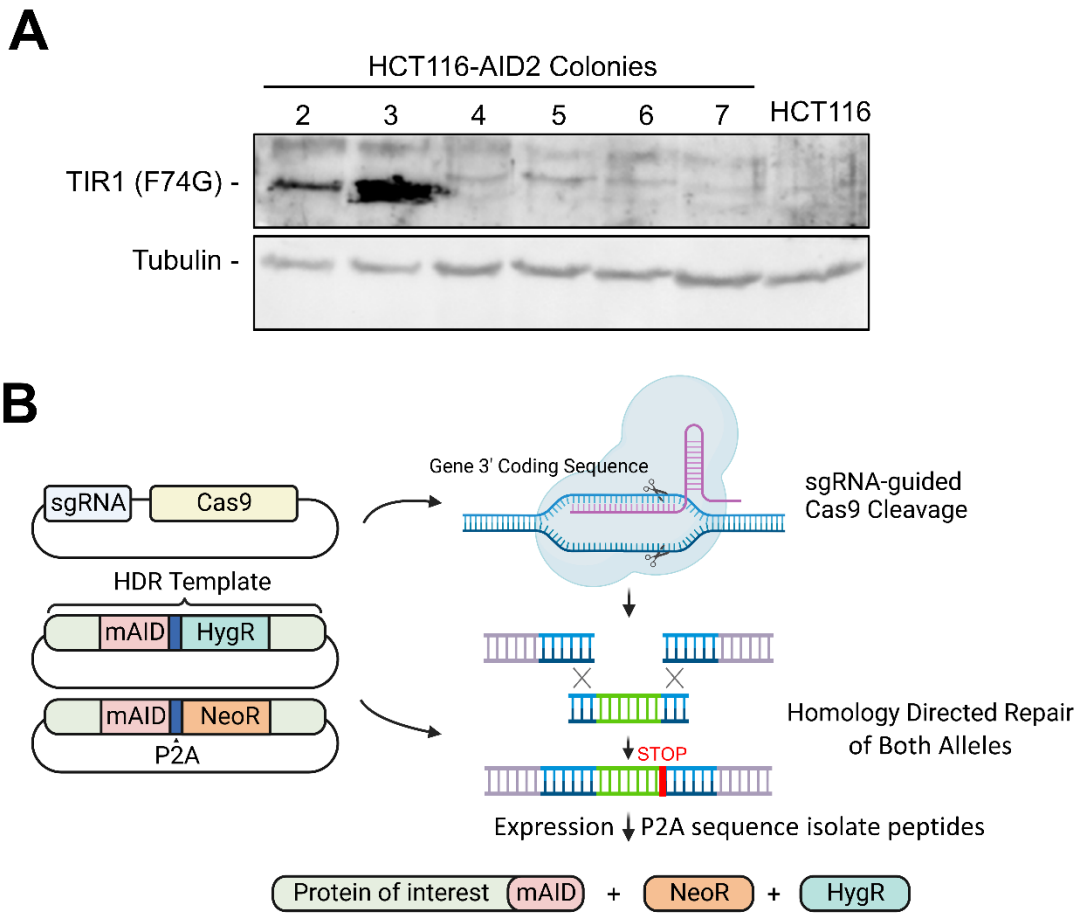

Figure IV.6.1. Establishing hnRNPUL1-AID2 cell line

A, Western blot identifying expression of TIR1 (F74G) in colony 2 and 3 of viable HCT116-AID2 colonies, tubulin was probed as loading control.

B, Schematics of Cas9-HDR editing strategy to tag C-terminal ends of both alleles of hnRNPUL1 with mAID tag and co-express neomycin and hygromycin resistance proteins (NeoR and HygR) through the P2A 'autocleavage' sequence.

C, Western blots of 12 hnRNPUL1-AID2 colonies showing successful tagging and 5-Ph-IAA depletion of all hnRNPUL1-mAID in colony 13, 16, 22, 25, 31, 37, 41 and 46. All successfully edited colonies showed lower basal levels of hnRNPUL1-mAID than levels of hnRNPUL1 in HCT116 cells.

IV.7. Summary

In this chapter, we have established that hnRNPUL1 KO led to reduced levels of U7 snRNA in the cell. Given that U7 snRNP is an essential component of RD histone mRNA 3′ processing, reduced U7 snRNA levels likely contributed to the RD histone mRNA processing defect observed by Dr. Yonchev (Yonchev, 2021). Interestingly, our TT-Seq analysis also revealed reduced transcription of histone genes following hnRNPUL1 loss, in common with loss of FUS (Raczynska, 2015). Since hnRNPUL1 interacts with FUS independently of RNA (Griffith, 2019) and both hnRNPUL1 and FUS interacts with PolII (Kwon et al., 2013; Yonchev, 2021), it is possible that they act co-operatively in the transcription of RD histone genes and maintenance of U7 snRNP.

The TT-Seq of hnRNPUL1 KO cells also revealed a global downregulation of PolII transcription. This was accompanied by the transcriptional upregulation of IEGs, a potential hallmark of cellular stress. Since hnRNPUL1 plays a crucial role in DNA double strand break repair (Polo et al., 2012), the global down-regulation of protein-coding gene transcription and the upregulation of IEG transcription could be a transcriptional response to accumulated DNA damage. Strikingly, the hnRNPUL1 KO effects on IEGs opposed the hnRNPU KO effects where IEGs were down regulated. Therefore, the balance of hnRNPU and hnRNPUL1 activity might be a crucial part of the cellular machinery that regulates the expression of IEGs upon stimuli.

Finally, we attempted to generate a better degron system for hnRNPUL1 using the AID2 system, but we still observed basal degradation of the protein. The AID2 system worked successfully for FUS in the following Chapter V, which suggests that intrinsic properties of hnRNPUL1 might have contributed to its basal degradation upon tagging with the mini-AID tag. A possible solution would be to tag the structured N-terminal sequence of hnRNPUL1 to avoid disrupting the stability of its intrinsically disordered C-terminal domain.

V. Analysis of the role of FUS in transcription and splicing

As Dr. Yonchev had previously identified similarities between depletion of hnRNPUL1 using an hnRNPUL1-AID cell line and knockdown of FUS using siRNA, we considered constructing a conditional knock out system for FUS to study the possible links between the two proteins. We also observed a reduction in FUS protein levels in hnRNPU-AID knockout cells, and we wondered to what extent that contributed to the hnRNPU transcription and splicing effects observed in the hnRNPU-AID cell line following depletion of hnRNPU-AID.

FUS (also known as FUS/TLS) is a well-studied ALS-associated gene that has been implicated in the regulation of transcription and processing of nascent RNAs. It was shown that FUS knockdown led to increased pausing of Ser2 phosphorylated PolII at promoter-proximal sites (Schwartz et al., 2012). This has been partly attributed to the potential of FUS to form liquid-liquid phase separated droplets through its N-terminal IDR, both on its own and with the CTD of PolII (Kwon et al., 2013), as well as to the association between FUS and the polyadenylation machinery that may feedback on PolII pause-release and elongation (Masuda et al., 2015). U1 snRNA has been identified as the major RNA target of FUS (Jutzi et al., 2020). As the spliceosomal U1 snRNP has been widely implied as a transcription regulator in both structural and functional analysis (Review by Venters et al., 2019; Caizzi et al., 2021; Zhang et al., 2021), the co-interactions of FUS with U1 snRNP and PolII has been suggested as a key link between transcription and splicing regulation (Yu & Reed, 2015).

In addition, studies showed that FUS interacts with SMN to facilitate formation of Cajal Bodies, which are key cellular machineries for processing and maturation of snRNAs (Yamazaki et al., 2012). FUS was also shown to interact with U7 snRNP to facilitate its maturation and proper transcription and processing of replication-dependent histone mRNAs (Raczynska et al., 2015). The effects of FUS siRNA knock down on histone misprocessing and snRNA levels have also been verified in the Wilson Lab by Dr. Yonchev (Yonchev, 2021). Mimicking ALS-associated FUS-C terminal mutations in laboratory cell lines lead to abnormal cytoplasmic localisation of FUS and caused similar disruption in Cajal bodies and snRNA levels as FUS knock-down, as well as cytoplasmic enrichment of U1 snRNA (Yamazaki et al., 2012; Yu et al., 2015; Jutzi et al., 2020). This suggested that FUS is localised in the nucleus by its C-terminal domain

and such localisation is important for its cellular functions.

Previous studies of FUS function have been hampered by the use of RNA interference approaches which requires extended incubation periods prior to analysis, typically at least 48 hours. This in turn means that some of the effects seen previously, following depletion of FUS, could be indirect. Too address this issue, we adopted the recently developed AID2 system to establish a fast FUS depletion cell model. We then used TT-Seq and RNA-seq to evaluate the transcription and mRNA expression effects at an early time point of FUS depletion.

V.1. Establishing FUS-AID2 cell line

Initially, we attempted to establish an 'Improved AID' system proposed by Sathyan et al. (2019). In this system, auxin response factor 16 (ARF16), a natural TIR1 regulator, is co-expressed with TIR1 to inhibit its activity until auxin treatment. However, the HCT116-TIR1-ARF16-FUS-AID cell line we established did not effectively degrade FUS with 3-IAA treatment. Therefore, we moved onto the HCT116-AID2 system once it was established.

Similar to the construction of the AID2-hnRNPUL1-mAID cells described in Chaper IV.6, we used a Cas9-gRNA – HDR template system to tag the C-terminus of the two copies of FUS coding sequences in HCT116-AID2 cells with an mAID tag, and selected successfully edited colonies with hygromycin and neomycin. We obtained two viable colonies of HCT116-AID2-FUS-mAID (hereafter referred to as "FUS-AID2") cells, both had similar FUS protein levels to the parental HCT116 cells and fully depleted FUS within 6 hours of 2nM 5-Ph-IAA treatment (Figure VI.2.1.C). Colony 3 was designated as the main FUS-AID2 cell line, as it was derived earlier and more cell stocks of it were available. Of note, we did not observe faster depletion of FUS-mAID with more than 2 nM 5-Ph-IAA (Figure V.1.1C).

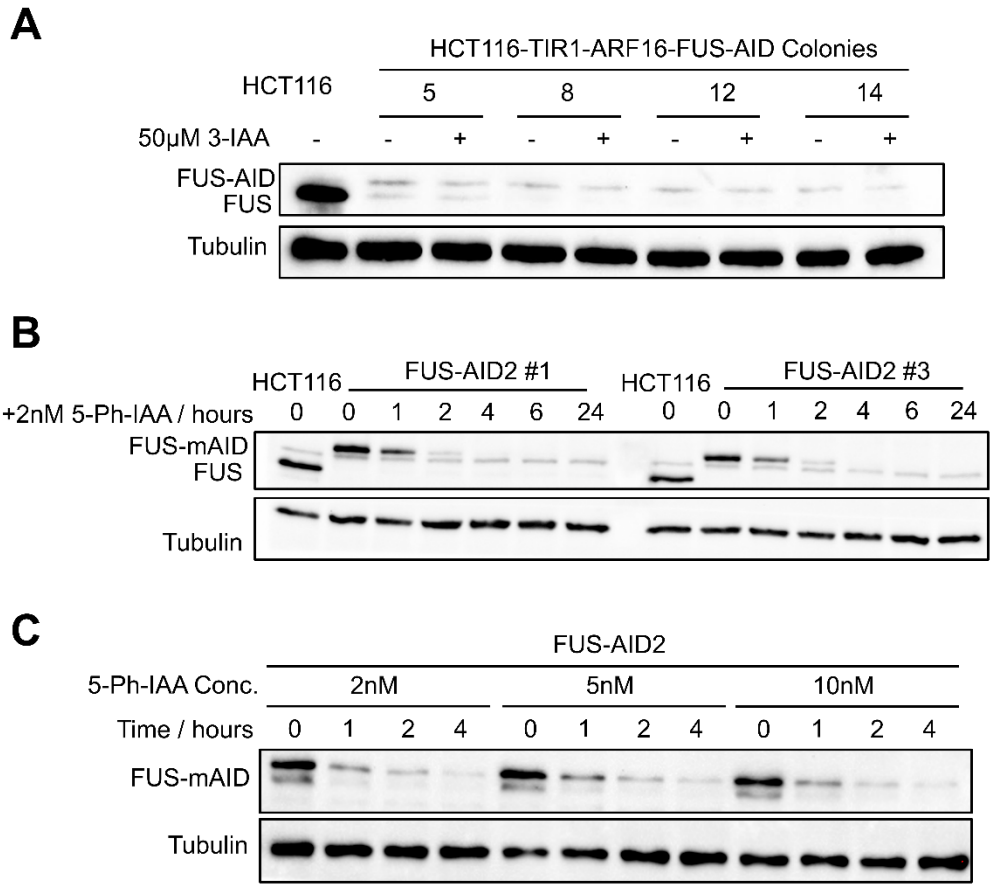

Figure V.1.1. Establishing FUS-AID2 cell line

A, Western blot of four viable HCT116-TIR1-ARF16-FUS-AID colonies, treated with and without 50μM 3-IAA. All colonies showed trace-level of FUS-AID compared to wilde type FUS in HCT116 cells.

B, Western blot showing FUS-mAID depletion within 0 to 24 hours of 2nM 5-Ph-IAA treatment in colony 1 and 3 of viable FUS-AID2 cells, compared to wild type HCT116. FUS-mAID completely depleted at six hours after treatment started.

C, Western blot showing depletion of FUS-mAID in FUS-AID2 cells by treatment of 2nM, 5nM and 10nM of 5-Ph-IAA for 0 to 4 hours. The rate of depletion was not noticeably different between FUS-AID2 cells treated with different concentrations of 5-Ph-IAA.

In **A**, **B** and **C**, tubulin was shown as loading control.

V.2. Validation of FUS functional loss in FUS-AID2 cells

To validate that the addition of mAID tag to FUS did not disrupt its function, we tested the reported FUS knock-down effect on RD histone mRNA processing, Cajal Body (CB) counts and snRNA levels in FUS-AID2 cells within 24 to 96 hours of 5-Ph-IAA treatment, similar to the duration of typical siRNA knockdowns used in past literatures.

First, we assayed CB counts by immunofluorescence of the CB marker coilin in FUS-AID2 cells treated with 2nM 5-Ph-IAA for 6, 24, 48, 72 and 96 hours or solvent-equivalent amount of DMSO (final concentration was 0.001% v/v) and counted the number of coilin foci using ImageJ software. In contrast to the reported complete loss of CB in FUS knock down or knock out cells, we observed gradual reduction of CB counts from 3-8 per cell in the DMSO treated FUS-AID2 cells to 1-5 per cell at 96 hours of 5-Ph-IAA treatment (Figure V.2.1). Interestingly, by statistical analysis, we observed an abrupt reduction of CB counts between 24 and 48 hours of 5-Ph-IAA treatment (Figure V.2.2A), which suggested that the loss of CB upon FUS depletion might be cell-cycle dependent. We also confirmed that the FUS C-terminal nuclear localisation signal was not disrupted by the mAID-tag, as the FUS-mAID immunofluorescence signal was in the DAPI-marked nucleus and characteristically similar to the wild type FUS in HCT116 cells (Figure V.2.1).

We then assayed levels of snRNA and misprocessing of RD histone mRNAs in FUS-AID2 cells. We extracted total RNAs from FUS-AID2 cells treated with either DMSO for 24 hours or 5-Ph-IAA for 24 hours and 72 hours, then reverse transcribed either with poly(dN) primers for qPCR testing of total histone mRNA and snRNA levels, or with poly(dT) primers for qPCR testing of misprocessed and polyadenylated RD histone mRNA levels.

Compared to FUS-AID2 cells treated with DMSO, levels of U1 and U2 snRNAs did not significantly change at 24 hours of 5-Ph-IAA treatment but reduced significantly (~20% for U1 snRNA and ~15% for U2 snRNA) at 72 hours (Figure V.2.2B). Levels of U4 and U5 increased significantly at 24 hours of 5-Ph-IAA but showed no statistically significance in their changes at 72 hours. The reduction of snRNA levels corresponded to decreases in Cajal bodies counts and the two phenomena may be interrelated.

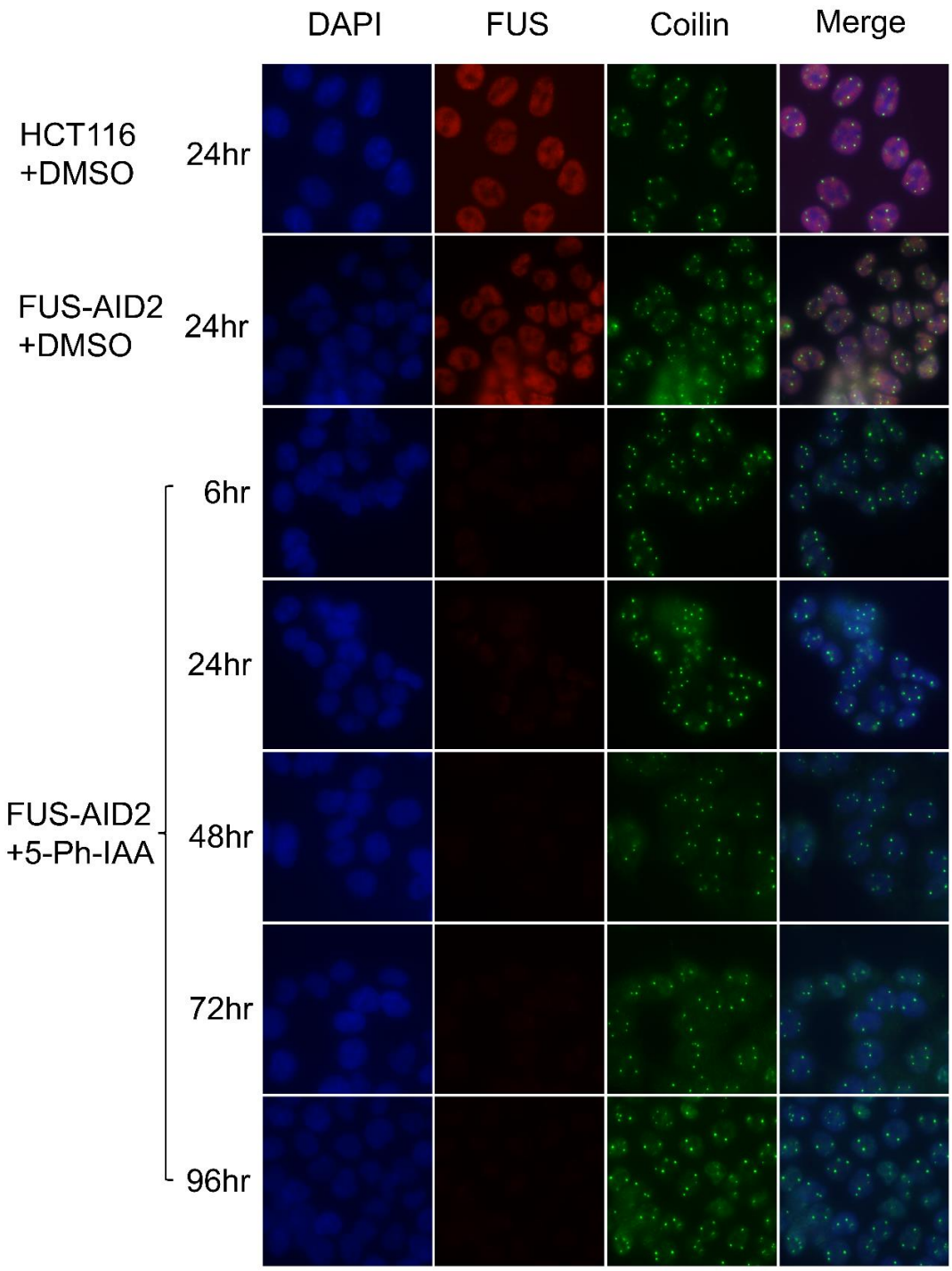

Figure V.2.1. Per-cell Cajal body counts reduced gradually in FUS-AID2 cells after FUS-mAID depletion.

Immunofluorescence of FUS and coilin in HCT116 and FUS-AID2 cells treated with DMSO or 5-Ph-IAA for 6 to 96 hours. Number of coilin foci (Cajal Bodies) per FUS-AID2 cell reduced gradually after 5-Ph-IAA treatment. HCT116 images were taken with 60× oil-immersed object lens, while all FUS-AID2 images were taken with conventional 40× object lens to avoid distortion to the coilin foci. Note that in DMSO treated FUS-AID2 cells, signals of FUS-mAID were localised in the nucleus, similar to wild-type FUS in HCT116 cells.

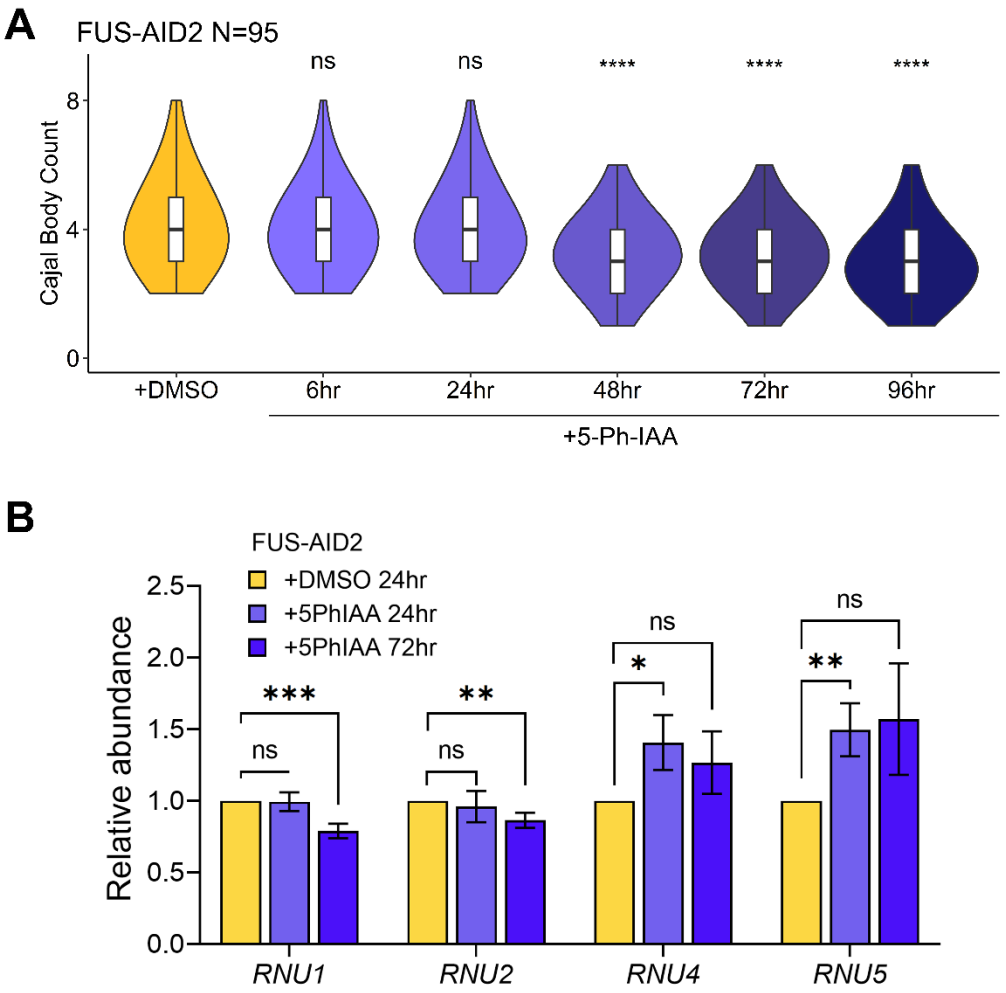

Figure V.2.2. Reduced Cajal Body counts and disrupted snRNA levels in FUS-mAID depleted FUS-AID2 cells.

A, Violin and box plots of coilin foci (Cajal Body) counts per cell in FUS-AID2 cells treated with either DMSO or 5-Ph-IAA for 6 to 96 hours. Number of Cajal Bodies were counted from 95 cells from two biological replicates. Cajal Body counts only reduced significantly after 48 hours of 5-Ph-IAA treatment. Distributions of Cajal Body counts in 5-Ph-IAA treated cells were statistically compared to that in DMSO treated cells by unpaired student's t-test; ns $P > 0.05$, ****$P < 0.001$.

B, Bar charts showing gradual reduction in RNU1 and RNU2 snRNA levels and increases in RNU4 and RNU5 snRNA levels in 24 and 72-hour 5-Ph-IAA treated FUS-AID2 cells. Normalized to 18S rRNA. Data are mean ± s.d. and statistically compared with unpaired student's t-tests; * $P < 0.05$, **$P < 0.01$, ***$P < 0.005$.

We assayed levels of four example RD histone mRNAs in both total and polyadenylated RNAs, normalised to GAPDH mRNA levels in both pools. In 24 and 72 hours 5-Ph-IAA treated FUS-AID2 cells, total mRNA levels of three RD histones significantly reduced (Figure V.2.3B), and the misprocessed polyadenylated mRNA levels of two RD histones, HIST1H1C and HIST2H2BE increased significantly, while the misprocessed levels of HIST1H2AC and HIST1H3H decreased significantly but to a lesser extent than their total levels (Figure V.2.3A). We calculated the apparent misprocessing ratios of these four RD histone mRNAs as the normalised ratio between their polyadenylated levels and total levels, and found that three out of the four tested RD histone genes had significantly higher mean misprocessing ratios after 48 and 72 hours of 5-Ph-IAA treatment compared with DMSO treated FUS-mAID cells (Figure V.2.3C). This confirmed that the RD histone mRNA processing related functions of FUS are likely preserved in FUS-AID2 cells.

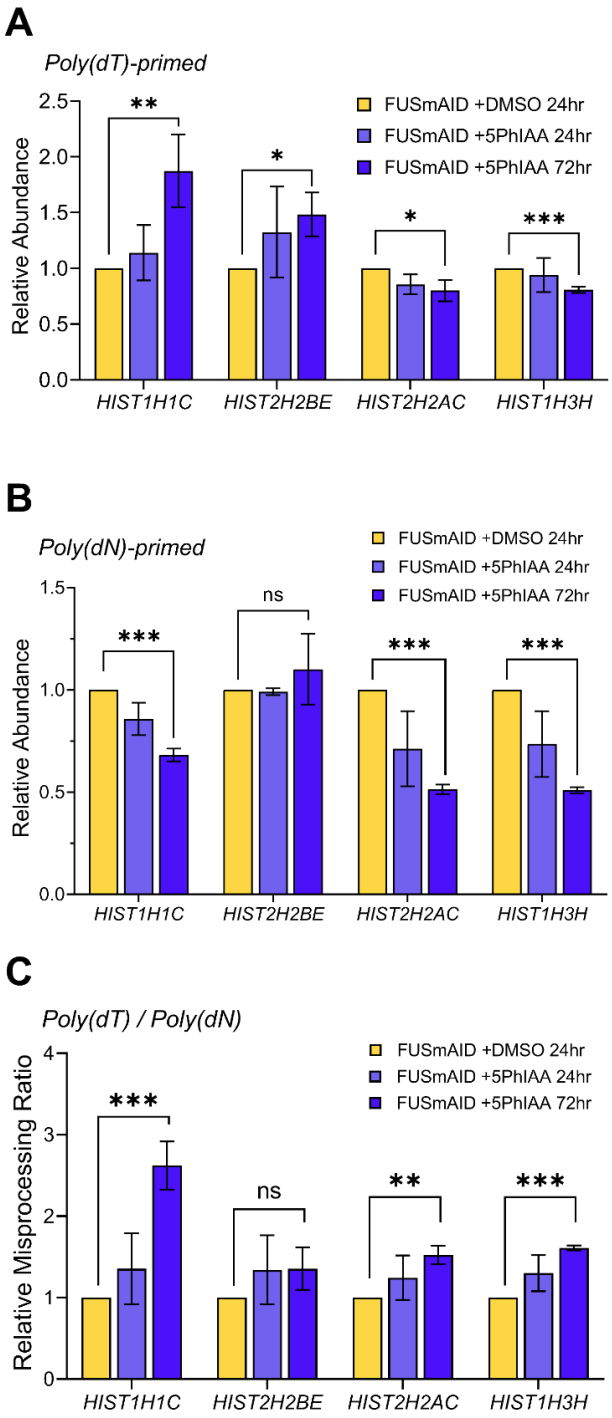

Figure V.2.3. RT-qPCR of histone mRNA misprocessing in HCT116-AID2-FUS-mAID cells.

A, **B**, RT-qPCR of histone mRNAs in poly(dT) (**A**) or poly(dN) (**B**) primed total cDNA prepared from HCT116-AID2-FUS-mAID cells treated with 0.001% DMSO for 24 hours or 2nM 5-Ph-IAA for 24 or 72 hours, normalized to GAPDH mRNA and compared to the DMSO treated sample in each replicate.

C, Relative misprocessing ratios calculated from poly(dT) and poly(dN)-primed histone mRNA abundance. Data are mean ± s.d. of n=3 biological replicates, analysed using two-tailed nonparametric t-tests; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.005$.

V.3. Reduced transcription of GC-rich genes following FUS depletion

The functional validation of the FUS-mAID2 cells reassured us to further explore its early effects on transcription and splicing. Three biological replicates of TT-Seq were performed on FUS-AID2 cells treated with either DMSO or 5-Ph-IAA for 6 hours. TT-Seq data were analysed using the same methods as described for hnRNPU-AID KO cells in Chapter III.4

At first glance, differential analysis did not reveal any significant changes in transcription at 6 hours of 5-Ph-IAA induced FUS-mAID depletion. The TT-Seq metagene profiles of unspliced reads in 5-Ph-IAA and DMSO treated FUS-AID2 cells were similar as well. We then inspected the TSS regions more closely using a genome browser, where we noticed decreased levels of TT-Seq read depths in sense directions of protein-coding gene TSSs. Metagene analysis of PROMPT-TSS region validated this observation (Figure V.3.1C,D), where we saw reduced TSS proximity sense TT-Seq signals in both multi-exon and mono-exon protein-coding genes, but not in lncRNAs.

We interrogated the TT-Seq data further using heatmaps of log2 fold coverage changes against properties of multi-exon protein coding genes (Figure V.3.2). Despite insignificant overall changes in gene body transcription, we observed FUS-mAID depletion selectively reduced transcription of genes with higher GC content and increased transcription of AT rich genes (Figure V.3.2). Interestingly, this GC-biased transcriptional effect started at approximately 500bps downstream of the TSS.

As FUS was reported as a strong interactor with the 5′SS defining U1 snRNP, we suspected the transcriptional effects caused by depletion of FUS were connected with the presence of the first 5′ splice site of genes. We therefore analysed the TT-Seq read depth changes at the first exon and intron to determine the element that led to the GC-based transcription effect bias following FUS-mAID depletion.

The GC content of first exons did not correlate with changes in their transcription levels after FUS-mAID depletion, while GC content and transcription levels of the first intron negatively correlated better than those of gene bodies (Figure V.3.2A). Gene body elements after the first intron (internal and last introns, exons) also showed moderate to good negative correlations between their GC content and transcription levels, suggesting that the early effects on transcription triggered by GC bias within the first intron were propagated down the gene body.

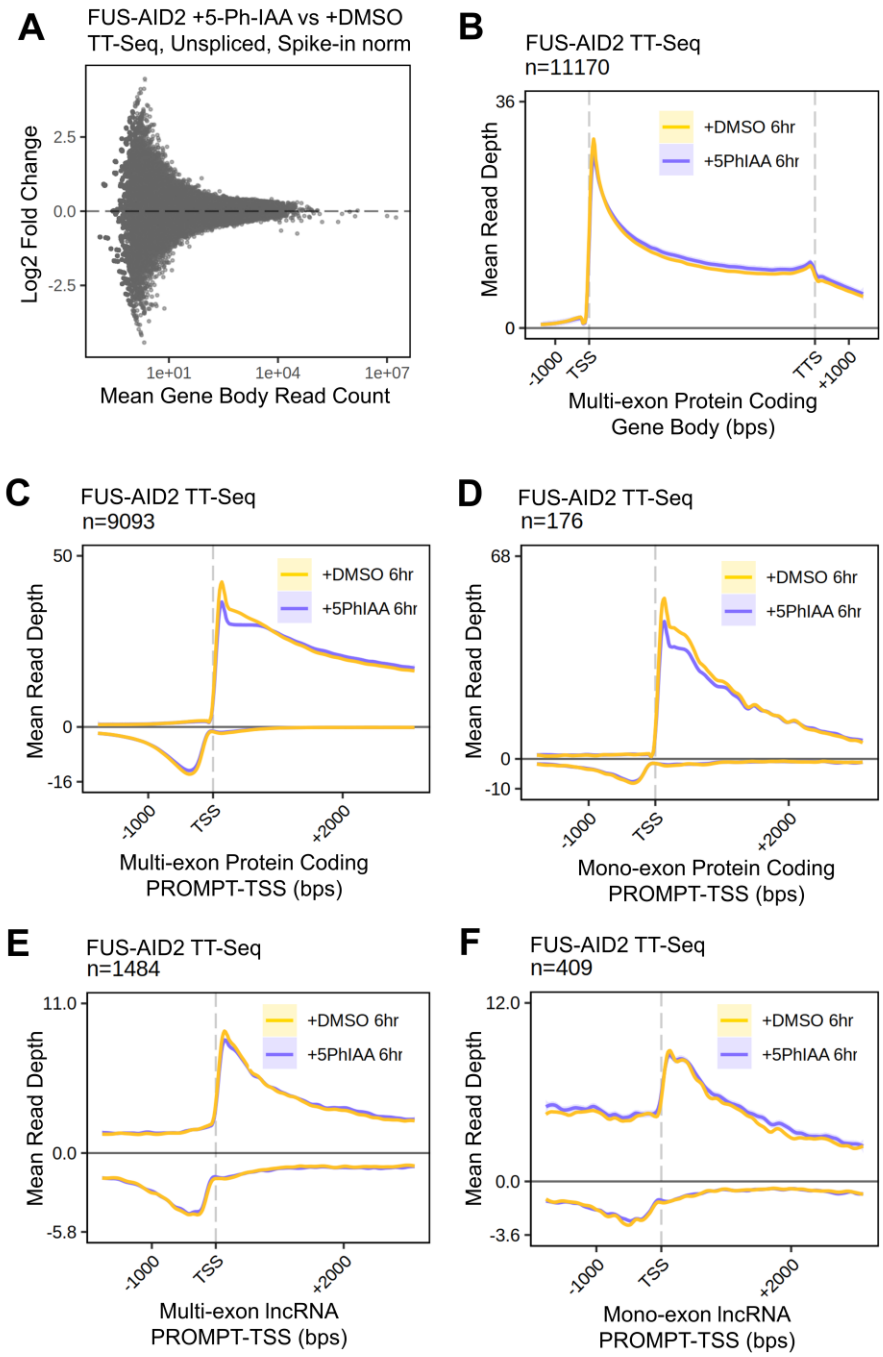

Figure V.3.1. 6-hour FUS-mAID depletion caused minor transcription reduction at the protein-coding TSS-proximal region

A, MA plot showing all transcription level changes in FUS-AID2 +5-Ph-IAA cells were not significant and marked in grey.

B, Metagene plot showing similar average read depth across multi-exon protein-coding gene bodies in FUS-AID2 6-hour 5-Ph-IAA and DMSO treated TT-Seqs.

C, D, E and F, Metagene plots showing reduced read depth at TSS-proximal region in 6-hour 5-Ph-IAA treated FUS-AID2 cells across (**C**) multi-exon and (**D**) mono-exon protein coding genes, and unchanged read depths across (**E**) multi-exon and (**F**) mono-exon lncRNA genes.

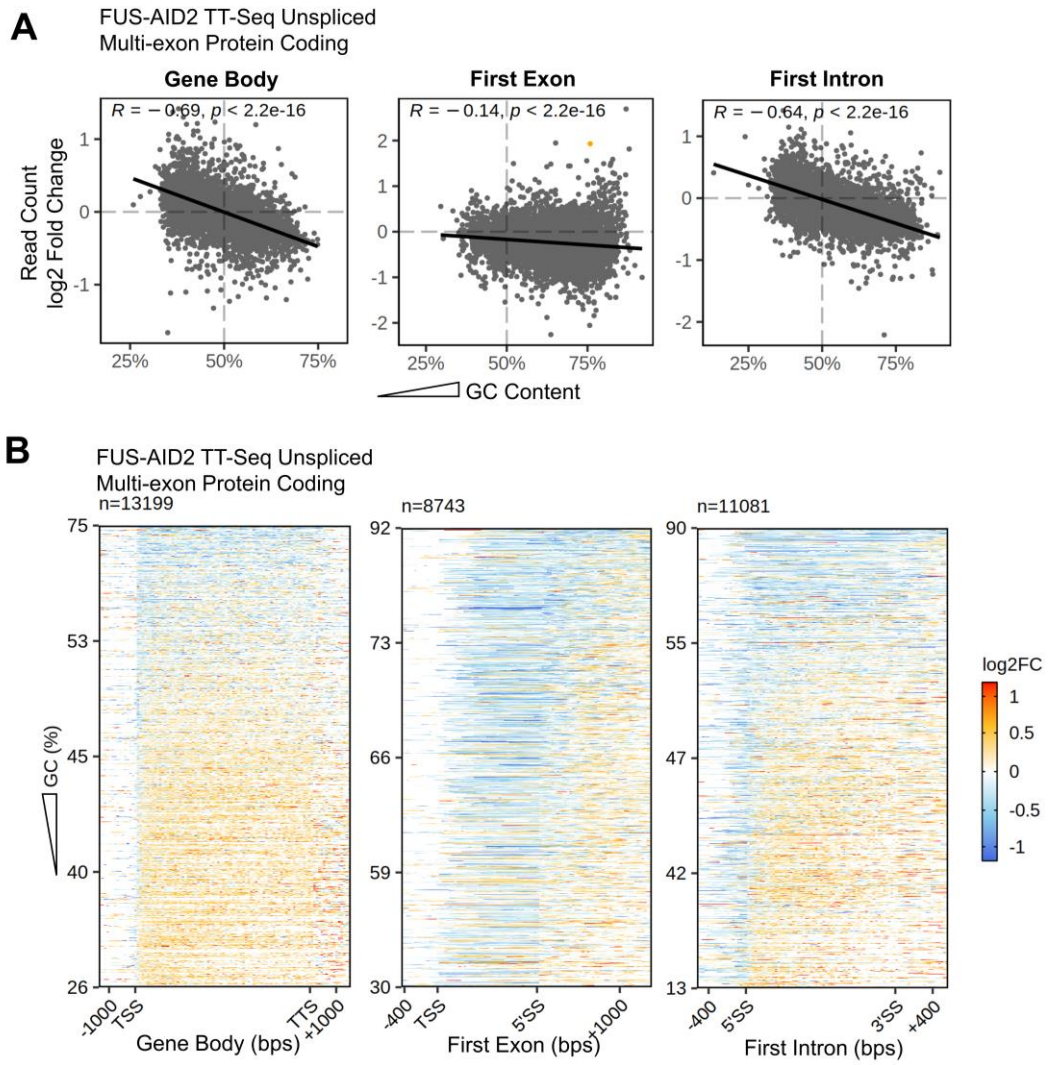

Figure V.3.2. Transcription changes in 6-hour treated FUS-AID2 TT-Seq were gene and first intron GC-biased.

A, Scatter plot of unspliced FUS-AID2 TT-Seq read count changes across gene bodies (including introns) (left), first exons (centre) and first introns (right) of multi-exon protein-coding genes, with at least 10 reads aligned, against their respective GC-contents. Negative GC correlations were observed in gene body and first intron read count changes, while no GC correlation was found in first exon read count changes.

B, Heatmap of log2 fold changes (log2FC) across 100 bins along each feature analysed in **A** at corresponding positions, ranked by their GC contents (y axis, high to low, up to down). Heat scales by log2FC is shown to the right. Biases observed in **A** could be observed in heatmaps as well with a clear transition point visible at the first 5′SS in first exon and first intron heatmaps.

V.4. FUS enriches over GC-rich first introns and downstream transcripts

Dr Yonchev identified enrichment of FUS over intronic regions in published eCLIP datasets by Gene Yeo Lab (Yonchev, 2021; Van Nostrand et al., 2016). As our TT-Seq data revealed first-intron GC-content dependent transcription effects immediately after FUS-mAID KO, we wondered if FUS is enriched over first introns, in particular, at the first 5′SS in the eCLIP datasets, and whether the GC content of the first intron affected its enrichment.

We retrieved raw FUS eCLIP and input sequencing data from HepG2 and K562 cell lines deposited by Gene Yeo Lab on ENCODE data base and aligned them to Ensembl Homo Sapiens GRCh38 genome using STAR. PCR and optical duplicates were removed using the integrated unique molecule identifiers (UMIs). Reads are normalised to size factors generated using gene body read counts in DESeq2, processed to the 5′ end single nucleotide position of the sequenced RNA fragment and analysed using log2 fold enrichment metagene and heatmaps to identify enrichment patterns across multi-exon protein-coding genes.

Metagene analyses showed that, in both K562 and HepG2 cells, FUS was enriched over the TSS-proximal and around TTS of gene transcripts (Figure V.4.1A and Figure V.4.2A). However, FUS did not enrich across the body of first exons, instead it enriched over the first 5′SS and 5′ region of first introns. The 5′ FUS enrichment over first introns is more noticeable than those over internal and last introns.

Log2 fold enrichment heatmaps and DESeq2-based differential enrichment analysis against GC content showed that FUS is preferentially enriched over GC-rich first introns in both cell lines (Figure V.4.1B and Figure V.4.2A). Both analyses also supported the FUS enrichment over GC-rich gene transcripts in K562 data, while only the heatmap supported this in HepG2 data. FUS enrichment over first protein-coding exons were not correlated to their GC content in either cell line.

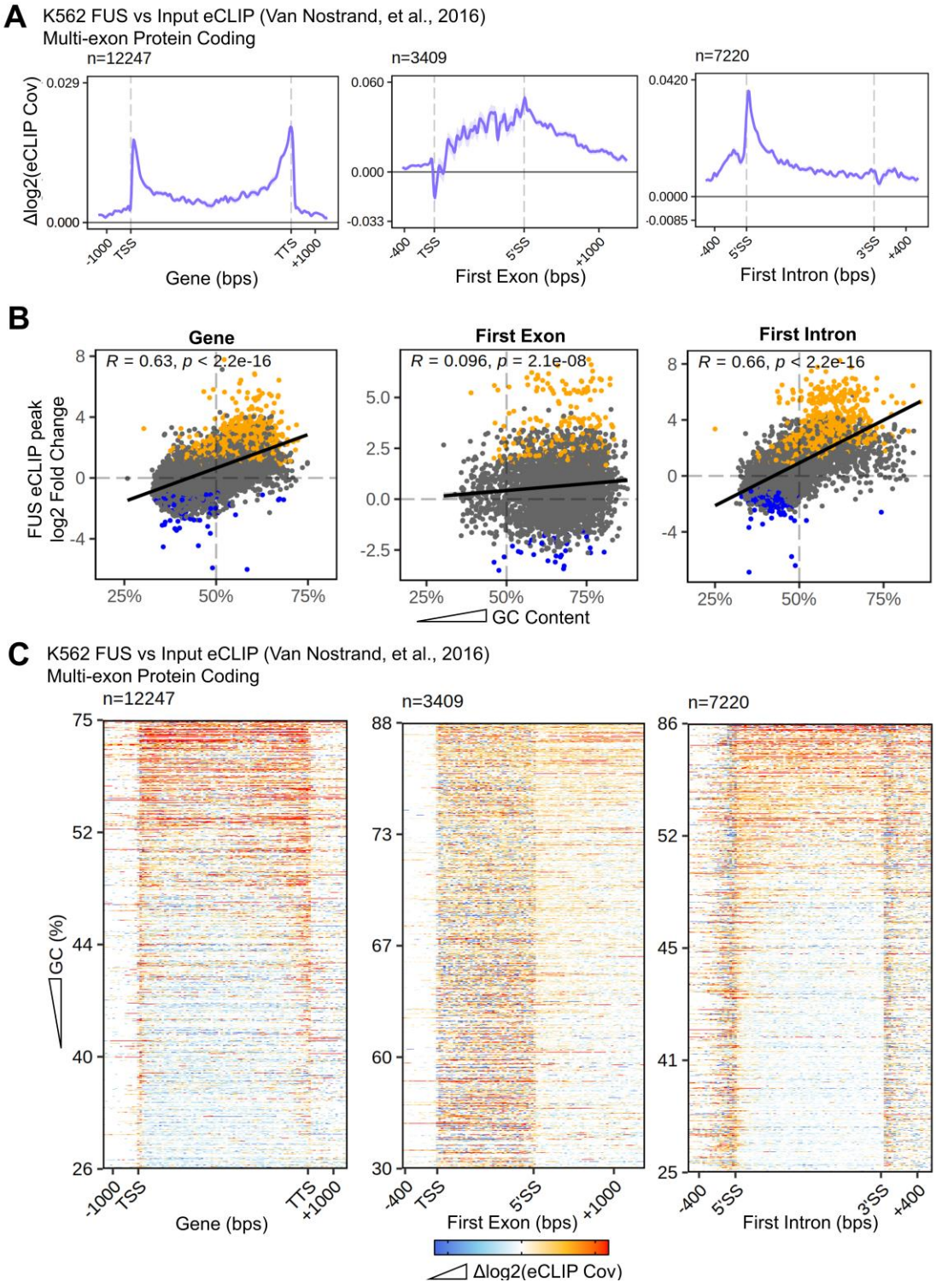

Figure V.4.1. K562 FUS eCLIP data showed FUS enrichment on GC-rich first introns and genes

A, **B** and **C**, (**A**) Metagene plots, (**B**) dot plot and (**C**) heatmap across gene body (left), first exon (centre) and first intron (right), showing log2 fold change of FUS eCLIP peak counts in K562 cells along each feature and against their GC contents. FUS was shown to prefer both GC-rich genes and first introns in K562 cells, but not GC-biased for the first exon.

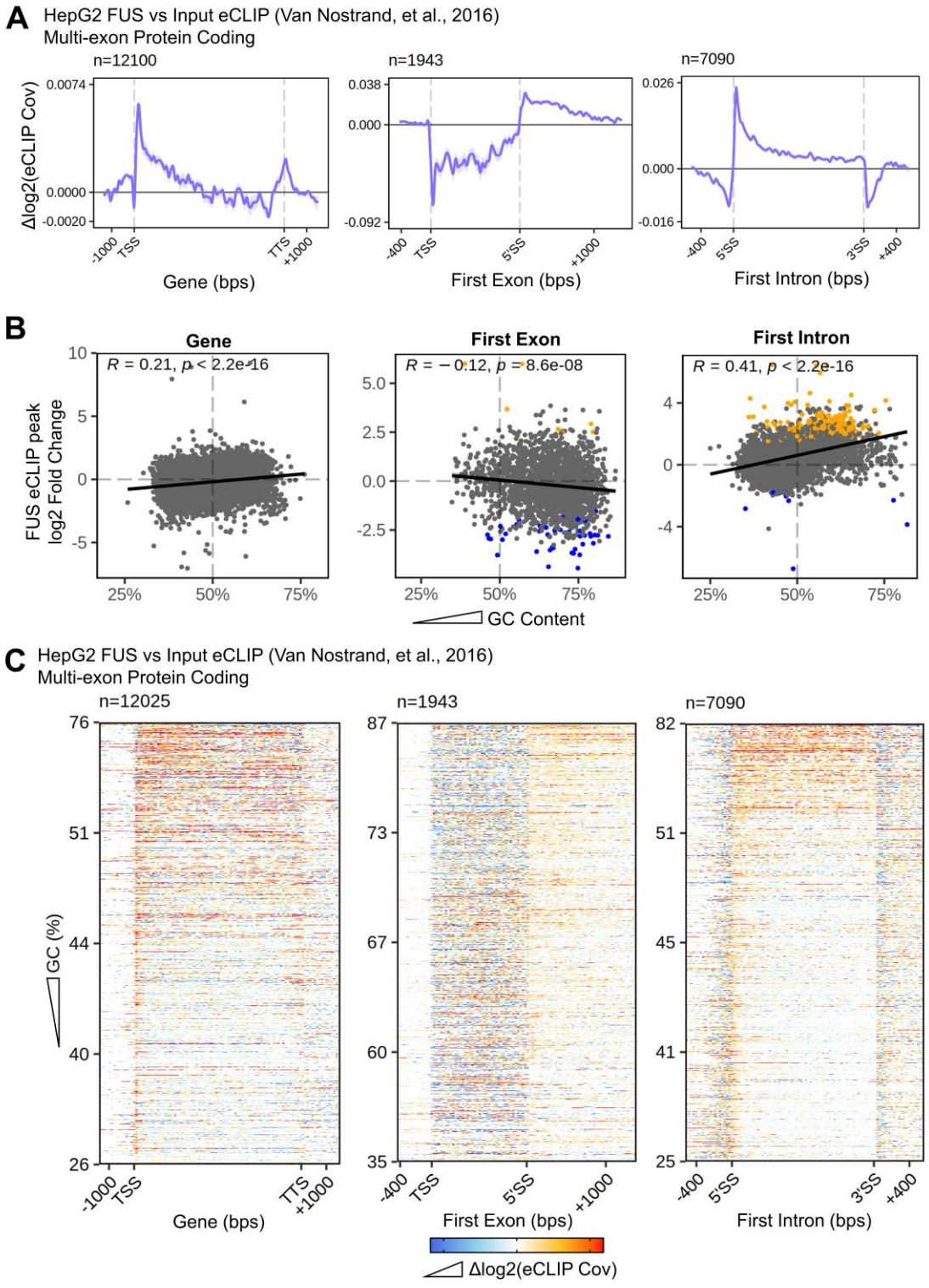

Figure V.4.2. HepG2 FUS eCLIP data showed FUS enrichment on GC-rich first introns

A, **B** and **C**, (**A**) Metagene plots, (**B**) dot plots and (**C**) heatmaps across gene body (left), first exon (centre) and first intron (right), showing log2 fold change of FUS eCLIP peak counts in HepG2 cells along each feature and against their GC contents. Heatmaps suggested FUS enriched on both GC-rich genes and introns in HepG2 cells, but dot plots only supported such observation in first introns.

V.5. FUS-mAID KO leads to reduced co-transcriptional splicing of several splicing factors

While inspecting the coverage of FUS-AID2 TT-Seq data in genome viewer, we observed a small and inconsistent relative reduction in protein-coding exon signals in +5-Ph-IAA data compared to those in +DMSO data. We wondered if FUS loss resulted in reduced splicing of specific gene groups, hence we analysed the changes in co-transcriptional splicing using spliced and unspliced reads in FUS-mAID KO TT-Seq as described for hnRNPU-AID KO data in Chapter III.

The overall co-transcriptional splicing levels were lower in +5-Ph-IAA than in +DMSO FUS-AID2 TT-Seq, but reductions in most individual genes were not considered significant. However, we noticed that splicing reduction of U1 snRNP core protein SNRPA (encoding U1A) was the most significant. We also observed a few other splicing-related genes, including SF3B2, SRSF9, LSM4 and FUS itself were in the top list of genes whose splicing was impaired.

However, we did not observe significant enrichment of FUS eCLIP signals over SNRPA and SRSF9, the genes that were most significantly affected in splicing (Figure V.5.2A). In addition, we did not observe a reduction in U1A protein levels within 24 or 72 hours of FUS-mAID KO either (Figure V.5.2B). This indicated that, despite the changes in splicing levels, the protein expression of U1A was likely unaffected in the short-term of FUS loss. This could be a result of the long half-life of U1 snRNP and slow turn-over rate of the associated U1A proteins (Sauterer et al., 1988), alternatively it may be because U1A autoregulates its expression through regulation of polyadenylation of its mRNA (Gunderson et al., 1994).

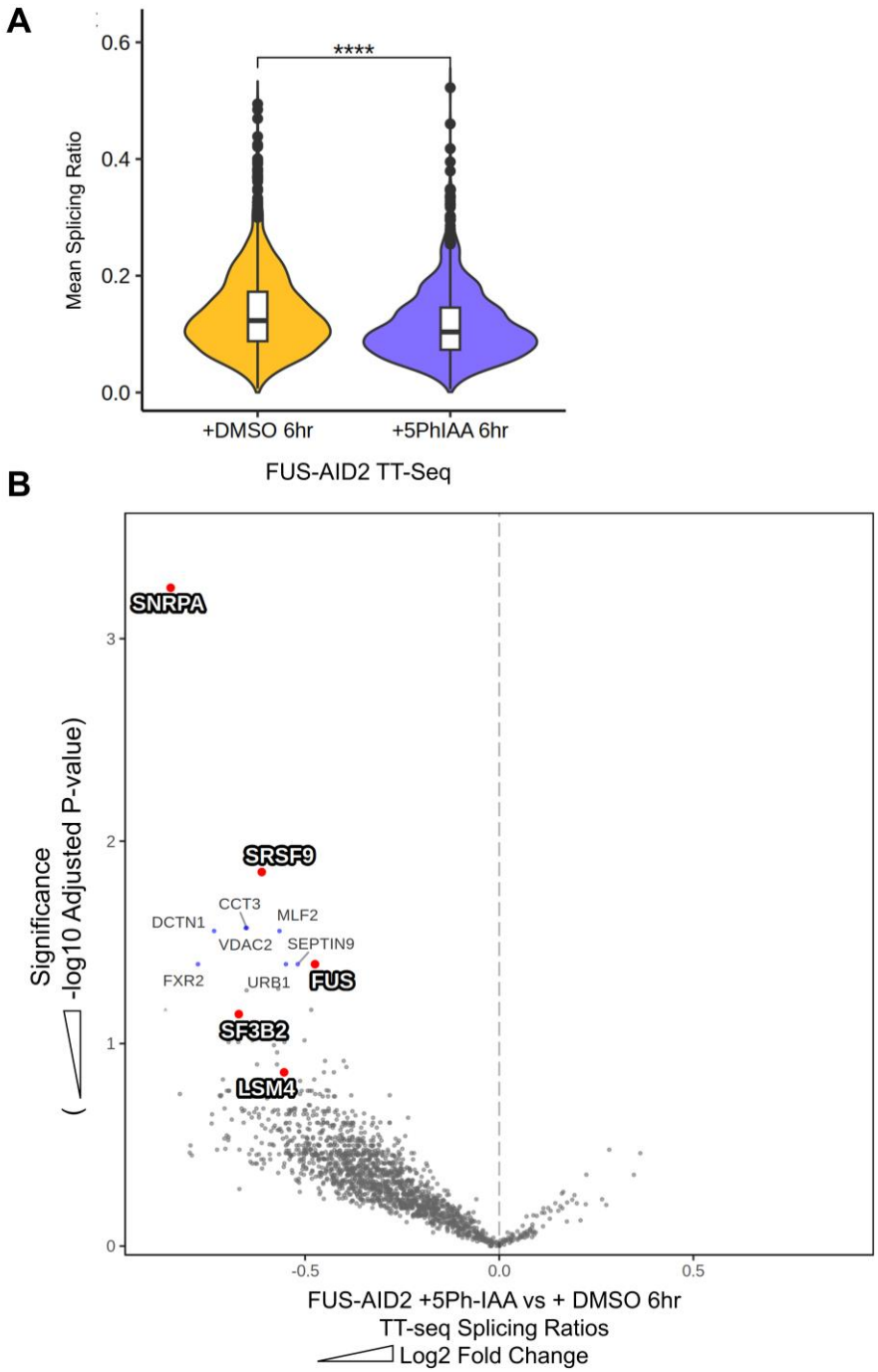

Figure V.5.1. Co-transcriptional splicing ratio changes in FUS-mAID 6hr KO cells.
A, Violin plot showing distributions of TT-Seq splicing ratio across all multi-exon genes in FUS-AID2 cells treated with either 6 hours of DMSO or 5-Ph-IAA. Distributions were statistically compared by unpaired student's t-test; ****$P < 0.001$.

B, Volcano plot showing TT-Seq splicing ratio log2 fold changes (x-axis) of protein-coding genes in 6-hour 5-Ph-IAA treated FUS-AID2 cells against the significance of the change (-log10 P-value, y-axis). The most significantly downregulated genes are named and marked in blue, and splicing-related genes are highlighted in red.

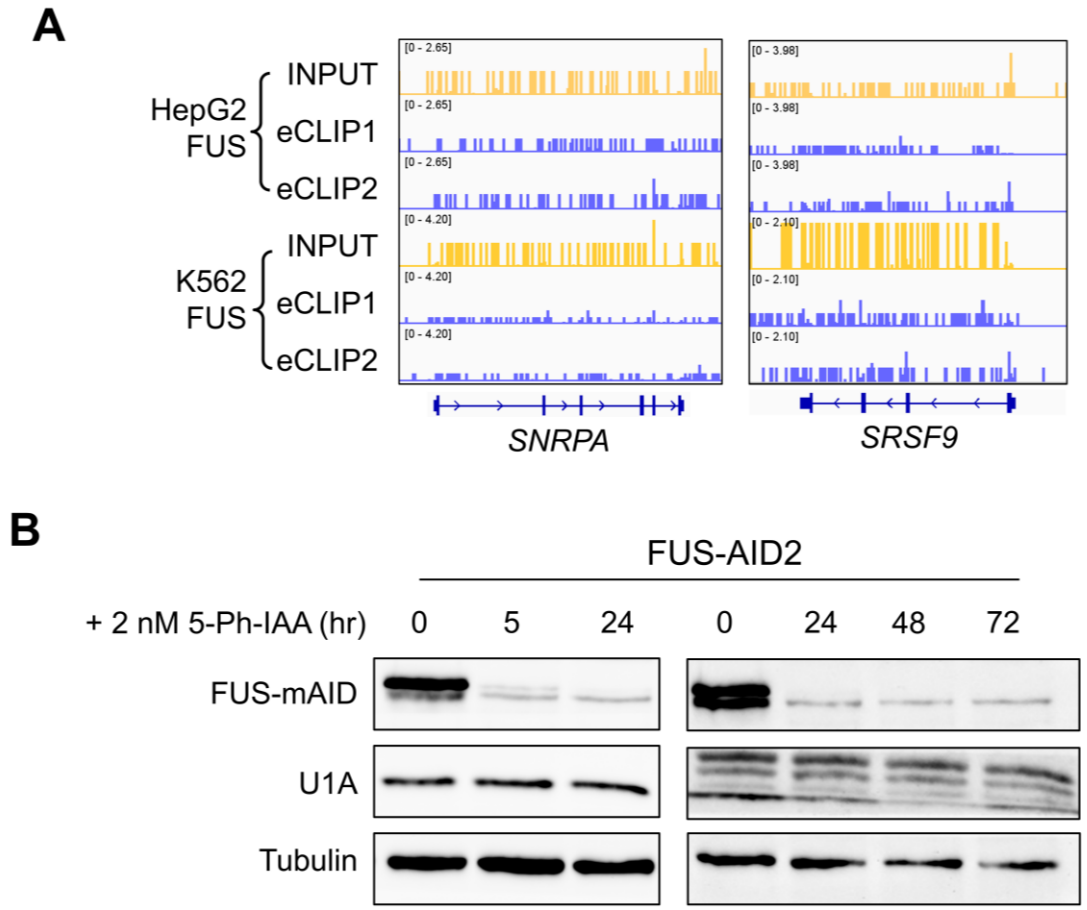

Figure V.5.2. FUS does not enrich over U1A and SRSF9 genes, and FUS KO does not reduce U1A protein levels in 72 hours.

A, IGV visualisation of eCLIP signal depth in HepG2 and K562 FUS INPUT and eCLIP datasets across U1A (SNRPA) and SRSF9 genes. The gene structure is shown at the bottom, and the read depth within each data set has been normalised and group-autoscaled, as indicated by signal coverage depth scale on the top left corner of each track window.

B, Western blot of U1A protein levels in FUS-AID2 cells treated with 5-Ph-IAA in a 24-hour time-course (left) and a 72-hour time course (right). FUS was probed to show its depletion and tubulin was probed for loading control. The 24-hour time course was repeated in two biological replicates and the 72-hour time course was repeated in three biological replicates.

V.6. FUS-mAID KO early and late expression changes corresponded to early transcription and splicing effects

In addition to the TT-Seq of 5-Ph-IAA and DMSO treated FUS-AID2 cells, we also conducted rRNA subtracted total RNA sequencing of the corresponding sets of 6-hour and 48-hour treated FUS-AID2 cells. The RNA-seq data were aligned to the human genome, processed to remove PCR and optical duplicates and multimappers. The per-gene read counts were then internally normalised using DESeq2-generated size factors for differential expression analysis of genes.

We did not observe any significant expression changes between 6-hour 5-Ph-IAA and DMSO treated FUS-AID2 cells, while at 48 hours of treatments approximately 20% of differentially expressed genes (DEGs) were found in 13545 genes detected. At both time points we observed strong negative correlations between the expression changes and the gene GC content, similar to the transcription change GC-bias observed in 6-hour treated TT-Seq (Figure V.6.1).

We also noticed that the splicing-related genes with minor reduced co-transcriptional splicing levels in 6-hour 5-Ph-IAA treated FUS-AID2 TT-Seq showed reduced mRNA levels at 48 hours of treatment, although none of the changes were significant at a 5% FDR (Figure V.6.1B), suggesting that splicing disruption upon FUS-mAID KO may correspond with reduced transcript production at a later time point.

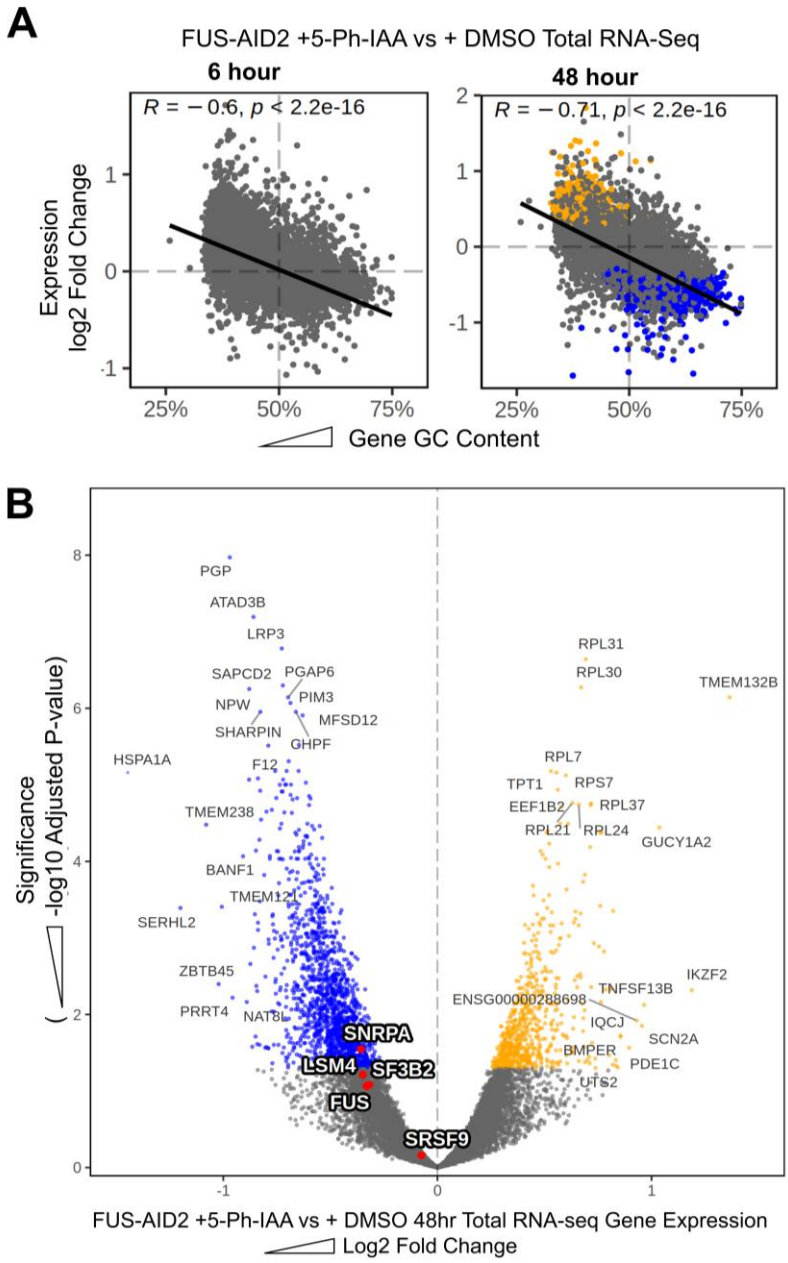

Figure V.6.1. RNA expression changes in FUS-mAID 6hr and 48hr KO cells.
A, Dot plots showing log2 fold changes (y-axis) of total RNA expression changes in FUS-AID2 cells at 6 (left) and 48 (right) hours of 5-Ph-IAA treatment against the gene GC contents (x-axis). Significant (P<0.05) upregulated genes are in blue and downregulated are in blue.

B, Volcano plot showing total RNA expression log2 fold changes (x-axis) of protein-coding genes in 48-hour 5-Ph-IAA treated FUS-AID2 cells against the significance of the change (-log10 P-value, y-axis). Significant (P<0.05) upregulated genes are in blue and downregulated are in blue. The most significantly changed genes are named, and the splicing-related genes noted in TT-Seq splicing ratio analysis are highlighted in red.

V.7. FUS-mAID KO affected ribosomal subunit expression

While quality checking and inspecting the 6-hour treated FUS-AID2 TT-Seq data in genome viewer, we noticed that the read depth across a particular region in one of the 28S ribosomal RNA (rRNA) loci (RNA28SN-1) was consistently higher in +5-Ph-IAA compared to +DMSO data (Figure V.7.1). We initially believed this was either an alignment artifact due to the repetitive nature of rRNAs, or a stress response to FUS loss. However, in the 48-hour treated total RNA-Seq data, we noticed that several ribosomal protein subunit genes were the most significantly increased DEGs. This hinted wider effects of FUS loss on ribosomal components.

We identified that the differential expression of ribosomal subunit protein genes in 48-hour treated FUS-AID2 cells were negatively correlated to their gene GC-content. We also observed similar negative GC correlations in their transcription and expression level changes at 6-hour treated TT-Seq and total RNA-Seq data sets. However, we did not observe enrichment of FUS across these genes or ribosomal RNAs in the published eCLIP data sets.

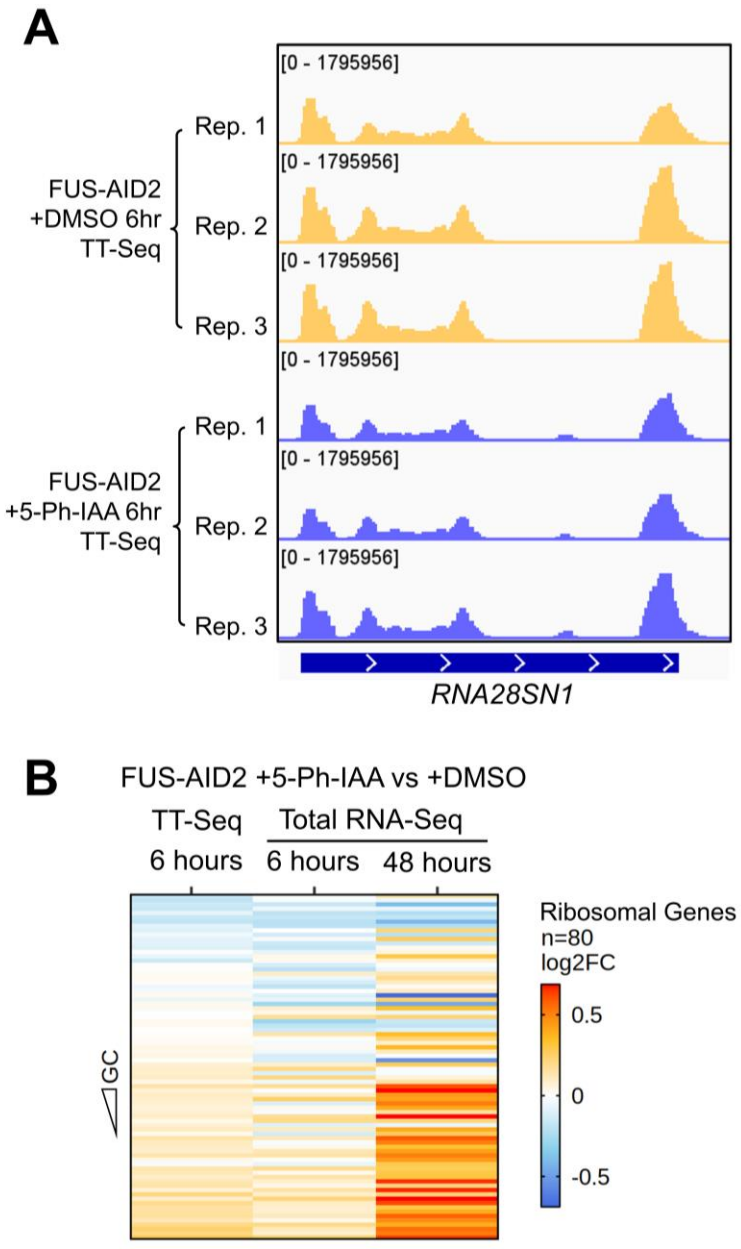

Figure V.7.1. Ribosomal RNA and protein gene expression are altered in FUS-mAID KO cells.

A, IGV tracks showing TT-Seq read depth across RNA28SN1 loci in FUS-AID2 cells treated with either DMSO or 5-Ph-IAA for 6 hours. Read depths are normalised using spike-in generated scale factors and auto-scaled, as shown on top left corner of each track.

B, Heatmap showing the log2 fold changes (log2FC) of ribosomal protein genes against their GC content in 6-hour 5-Ph-IAA treated FUS AID2 TT-Seq and total RNA-Seq, as well as in the 48-hour 5-Ph-IAA treated FUS-AID2 total RNA-Seq.

V.8. FUS-mAID KO transcription effects partially correlated to hnRNPU-AID KO effects, not to hnRNPUL1-AID KO effects

As described in Chapter III, we observed reduced FUS protein levels and a bias of further increased AT-rich gene transcription in hnRNPU-AID KO cells (Figure III.1.1 and Figure III.4.1). Dr. Yonchev also noticed strong enrichment of hnRNPU on nascent transcripts of the FUS gene in eCLIP data (Yonchev, 2021). As we observed similar GC-based transcription bias in rapid 6-hour FUS-mAID KO, we wondered if there was a correlation between the transcription effects in hnRNPU-AID KO cells and FUS-AID2 +5-Ph-IAA 6-hour cells.

A correlation analysis between gene body TT-Seq read changes between hnRNPU-AID KO TT-Seq and FUS-AID2 6-hour treated showed that their transcription changes moderately positively correlate with each other. When the GC content of the gene bodies are considered in this correlation, the GC rich-genes (purple) and AT-rich genes (yellow) showed distinct trends,where the correlations within AT-rich genes showed better conformity to the observed line of regression (Figure V.8.1A). This might suggest that the transcription change bias of AT-rich genes in hnRNPU-AID KO was partially related to reduced FUS levels, while the transcription changes in GC-rich genes were caused by other effects.

We were also interested in the possible correlations between the global transcription downregulation in hnRNPUL1-AID KO cell and FUS-mAID KO transcription effects, due to the strong interaction between FUS and hnRNPUL1 and similarities between FUS and hnRNPUL1 loss identified by Dr. Griffith and Dr. Yonchev, as mentioned in Chapter IV.

We identified a weak positive correlation between hnRNPUL1-AID KO transcription effects with FUS-mAID KO transcription effects, which meant we could not confidently explain the general transcription downregulation in hnRNPUL1-AID KO cells by functional disruption of FUS (Figure V.8.1B).

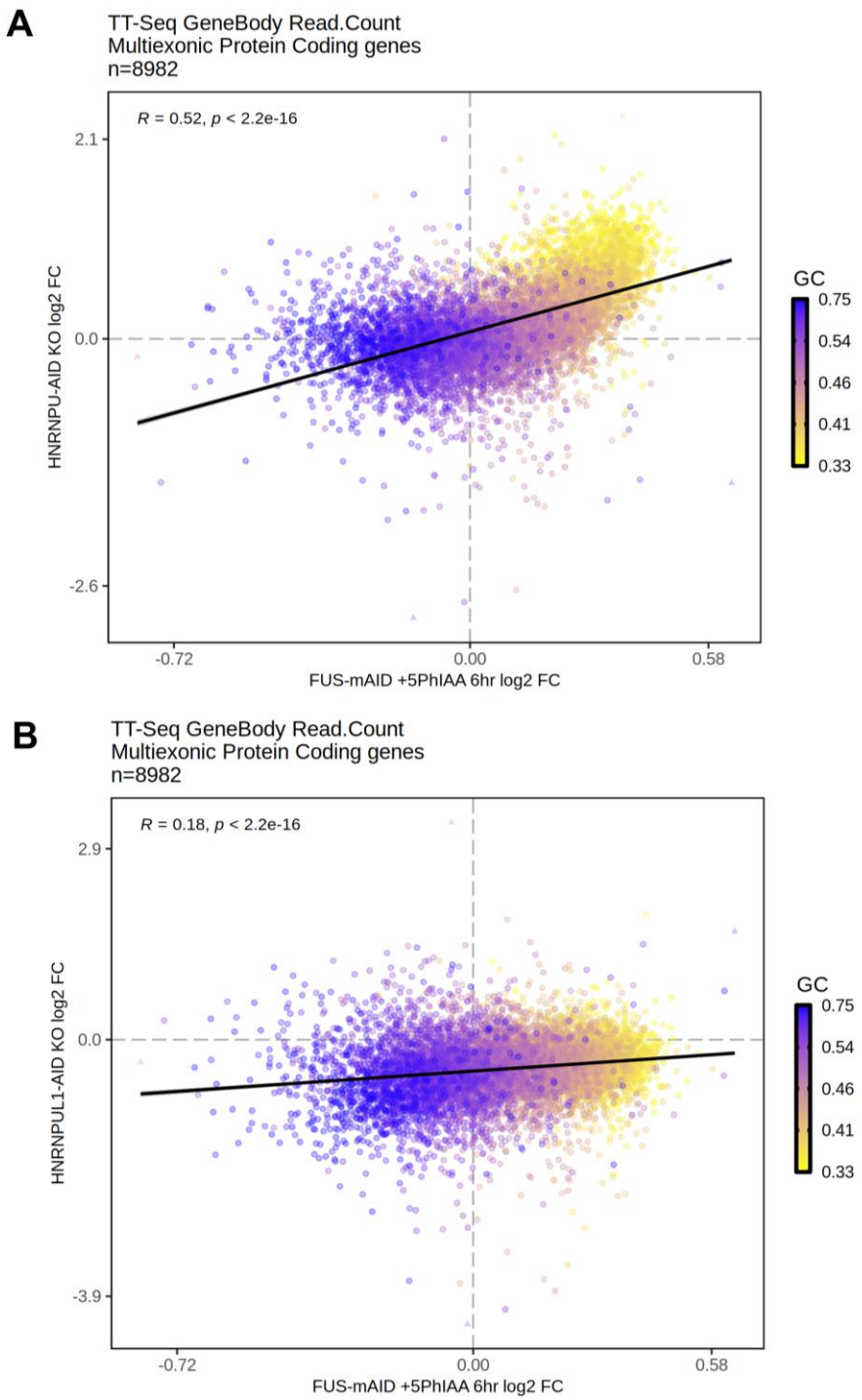

Figure V.8.1. Correlations between FUS-mAID KO, hnRNPU-AID KO and hnRNPUL1-AID KO TT-Seq

A, Dot plot showing log2 fold changes of unspliced TT-Seq read count across 8982 multi-exon protein-coding genes in hnRNPU-AID KO cells (y-axis) against that in FUS-mAID 6-hour KO cells (x-axis). Gene GC contents are represented in a blue to yellow (high to low) scale.

B, Unspliced TT-Seq read count changes of hnRNPUL1-AID KO cells against FUS-mAID 6-hour KO cells, plotted using the same method as in **A**.

V.9. Summary

In this Chapter, we established and validated a FUS-AID2 degron system to explore the early and direct transcriptional, splicing and gene expression effects of FUS functional loss. We discovered that FUS loss in 6 hours resulted in transcription changes post first 5′SS of multi-exon genes, which negatively correlated to the GC-content of gene bodies and first introns. The GC-based transcription effects were accompanied by minor changes in mRNA levels at 6 hours of FUS depletion and more significant changes at 48 hours. We also observed changes in ribosomal protein gene transcription and mRNA levels that correlated well with their GC content, suggesting that the GC-biased transcription effects of FUS loss may result in cascading effects to other cellular machineries.

FUS RNA binding profiles in published eCLIP datasets showed enrichment preferentially across GC-rich genes and first introns, particular over their 5′SSs. This suggests that FUS binding to GC-rich intronic 5′SSs may facilitate PolII transcription to proceed across introns in GC-rich genes, especially the first 5′SS, while loss of FUS might have slowed down or stalled transcription at such sites. Interestingly, GC rich regions of genes, including exons, tend to have slower transcriptional elongation rates and U1 snRNP can stimulate transcriptional elongation (Mimoso & Adelman, 2023). Therefore FUS may help to maintain optimal elongation rates across GC rich introns by driving recruitment of U1 snRNP. However, in the absence of complementary PRO-seq data to match our TT-Seq data, we cannot easily infer transcriptional elongation rates, so it is difficult to determine the precise role of FUS in transcriptional elongation. The canonical 5′SS defining spliceosomal U1 snRNA was identified as a prevalent interactor of FUS (Jutzi et al., 2020) and as an important regulator of PolII productive elongation, particularly across AT rich introns (Mimoso & Adelman, 2023). Further investigations of the interactions between FUS, U1 snRNP and PolII at the 5′SSs of multi-exon protein-coding genes may provide insights into the transcription and splicing roles of both FUS and U1 snRNP, and the mechanisms of transcription-splicing coupling in mammalians systems.

Despite the well-characterised interactions between FUS and U1 snRNP, we only observed minor and insignificant overall downregulation of co-transcriptional splicing at the early 6-hour time point of FUS-mAID depletion. However, multiple splicing factors were identified in the genes whose splicing was most disrupted, including the

genes encoding SNRPA (encoding U1A, U1 snRNP constitutive subunit), SF3B2 (17S U2 snRNP marker SF3B subunit), and LSM4 (U6 snRNP subunit). The expression levels of these splicing factor reduced slightly at 48 hours of FUS-mAID depletion, matching the extent of their earlier co-transcription splicing reduction. This observation might imply that FUS facilitated splicing of splicing factors, likely indirectly, and long-term loss of FUS function may have resulted in subtle and wide-spread disruption of the splicing machinery.

Lastly, we observed moderate correlations between the FUS-mAID 6-hour depletion transcription changes with those in hnRNPU-AID KO cells, in particular within AT-rich genes. This meant that part of the observed transcription increase bias towards AT-rich genes might be a result of the reduced levels of FUS proteins in hnRNPU-AID KO cells. We did not observe correlations between FUS-mAID KO and hnRNPUL1-AID KO transcription effects, which meant that the global transcription downregulation in hnRNPUL1-AID KO cells could not simply be attributed to the observed interactions and similarities between hnRNPUL1 and FUS.

VI. Discussion

VI.1. HnRNPU regulates the splicing-transcription feedback loop

In Chapter III, we inspected the effects of HNRNPU KO on splicing and transcription using subcellular fractionation RNA-Seq and TT-Seq. We demonstrated that HNRNPU loss resulted in global upregulation of transcription and splicing of chromatin associated transcripts, including nascent transcripts (Figure III.4.1, Figure III.6.2 and Figure III.6.6). However, we did not find significant per-gene correlation between the transcription and splicing effects of hnRNPU KO (Figure III.6.7). This meant that these two effects are likely not directly associated at per-gene level. This could be explained in two ways, either that the splicing and transcription regulation by hnRNPU are direct, but independent of each other, or that one of these two effects was caused by secondary global effects of the other.

It is well established that splicing and PolII transcription feeds back positively on each other, where inhibition of splicing factors such as SC35 and U2 snRNP disrupts PolII transcription, and phosphorylation of PolII CTD stimulates splicing (Guo et al., 2019). However, the global correlation and causality between splicing and transcription has not been fully established at per-gene levels, hence we cannot deduce which mechanism was affected primarily by hnRNPU KO based on existing models of splicing and transcription.

On the other hand, we observed that blocking of transcription by ActD dissociated hnRNPU from chromatin (Figure III.3.1), and Nozawa et al (2017) observed that both RNase and PolII-inhibitor amanitin exerted similar effects on hnRNPU. Our preliminary testing showed that inhibition of splicing by PlaB does not cause shift in hnRNPU subcellular fraction distribution. This implied that co-transcriptional and chromatin-associated actions of hnRNPU act downstream of active transcription but upstream of active splicing. In addition, we have identified that hnRNPU KO increased proportion of BPS-recognising active 17S U2 snRNP compared to HIV-Tat-SF1-associated inactive U2 snRNP, and that hnRNPU may interact with SF3B1 independent of either DNA or RNA (Figure III.9.1). Lastly, we showed that Pladienolide B inhibition of U2 snRNP could rescue the elevated splicing levels in hnRNPU KO cells (Figure III.9.2).

The association between hnRNPU and U2 snRNP could extend beyond SF3B1. The interactome mass spectrometry study conducted by Dr. Griffith using FLAG-tagged hnRNPU (Griffith, 2019) identified several SF3B subunits as well as a significant enrichment of U2A′ (SNRPA1) with hnRNPU (Figure VI.1.1B). However, chromatin-interacting splicing factors might be under-represented in this dataset, as the cell lysate was not DNase treated prior to co-immunoprecipitation. Further evidence of hnRNPU-U2 snRNP interaction is present in an interactome study of U2 auxillary factor 65 (U2AF65, or U2AF2) during T-cell activation (Whisenant et al., 2015). HnRNPU was highly enriched in this data set, similar by ranking with the classical U2AF65 pre-splicing BPS-site competitor hnRNPC and interactor U2AF35 (Thickman et al., 2006; Zarnack et al., 2013) (Figure VI.1.1A). The association of hnRNPU with both U2AF65 and SF3B1 suggests that it may act to inhibit splicing during spliceosomal E-complex formation when U2AF65 associates with SF3B1 (Das et al., 2000; and Galardi et al., 2022).

In conclusion, we speculate that hnRNPU associates with chromatin through nascent transcripts, where it may interfere with poised 17S U2 snRNP directly, and possibly inhibit its activation and BPS recognition. This means that hnRNPU would stall U2 snRNP before formation of the spliceosomal A complex, reducing the turnover rate of U2 snRNP and its global positive feedback on PolII pause-release and elongation velocity, as suggested by Caizzi et al (2021). This model is consistent with our observation of the global increase in splicing and transcription, and increased processing of mRNAs such as NR4A1 following hnRNPU KO (Figure III.4.1, Figure III.5.3 and Figure III.6.2). This is also consistent with our previous analysis of the hnRNPU eCLIP data, where hnRNPU is enriched over intronic RNA (Yonchev, 2021).

This proposed model of action of hnRNPU might facilitate our understanding of hnRNPU-related neuro-development disorder, as hnRNPU loss affected splicing and expression of neural development and function-associated genes, such as SMN, FUS and immediate early genes EGR1 and FOS (Figure III.1.1, Figure III.5.2, Figure III.5.3, Figure III.6.3 and Figure III.8.2). The transcription inhibition of IEGs caused by hnRNPU loss might worth further investigation, as loss of AP1 transcription factor family IEGs could lead to cascading effects in neuronal response to stimuli, such as their growth and response to memory retrieval (Tuvikene et al., 2016; Silva et al., 2018).

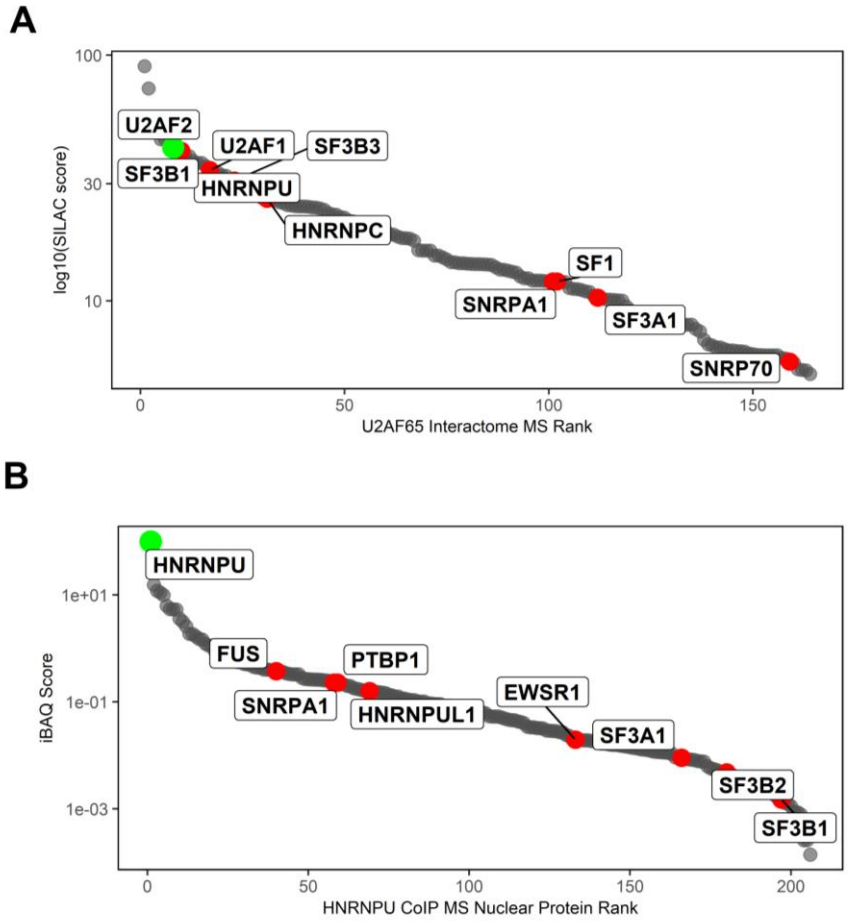

Figure VI.1.1. U2AF65 and hnRNPU Mass spectrometry data.
A). The interactome ranking of U2AF65 in the mass-spectrometry (MS) data of Whisenant et al., (2015), where the log10 transformed SILAC score (y-axis) of each protein was plotted against their SILAC score ranking (x-axis). hnRNPU, hnRNPC, SF3 subunits, U2AF35 (U2AF1), U2A′ (SNRPA1), SF1 and U1-70K (SNRP70) were highlighted in red, while the bait U2AF65 (U2AF2) was highlighted in green.

B). The interactome ranking of nuclear, non-histone and non-ribosomal proteins in the FLAG-hnRNPU co-IP MS data generated by Dr. Griffith (2019). Normalized iBAQ scores (y-axis, log10 transformed) of each protein was plotted against their iBAQ score ranking (x-axis). hnRNPUL1, FUS, EWSR1, SF3 subunits, U2A′ (SNRPA1) are highlighted in red, while the bait hnRNPU was highlighted in green.

It is also worth noting that FUS is one of the few exceptions in the splicing effect of hnRNPU KO, where its pre-mRNA splicing level reduced in both chromatin RNA-Seq and TT-Seq (Figure III.6.3, Figure III.6.6), and this was accompanied by reduced protein levels (Figure III.1.1). As FUS is a well-established ALS-associated gene (Vance et al., 2013; Sun et al., 2015; An et al., 2019), it is possible that disruption of FUS biogenesis may also contribute to symptoms in hnRNPU syndrome patients.

VI.2. HnRNPU's roles in maintaining the chromatin structure and its associated RNA scaffolds

While the possible U2 snRNP-associated splicing and transcription effects explained the majority of our observations of hnRNPU KO on protein-coding genes, it does not fully explain the dissociation of chromatin-bound, typically intronless non-coding RNAs, such as U1 snRNA, 7SK, MALAT1, NEAT1 and FIRRE (Figure III.2.1, Figure III.5.2 and Figure III.7.1). We speculate that U1 and 7SK snRNAs might be turned over faster and released into nucleoplasm following global upregulation of PolII pause-release and splicing, as U1 snRNP is released from the spliceosome upon formation of activated complex B (reviewed in Herzel et al, 2017), and both 7SK and U1 snRNA were reported to associate with PolII at promoter proximal regions to facilitate PolII release and elongation (Venters et al., 2019; Studniarek et al., 2021). However, the loss of the other chromatin-associated RNAs would need further explanation.

Since hnRNPU contains both DNA and RNA-binding motifs, it has been speculated as a DNA-RNA tethering protein since its discovery (Kiledjian and Dreyfuss, 1992). This possible RNA-tethering role of hnRNPU was used to explain the Xist dissociation from silenced X-chromosome in hnRNPU knocked down XX female mammalian cells (Hasegawa et al. 2010), despite those later findings triggered debate over the necessity of hnRNPU in this process (Kolpa et al., 2016). Chromatin-associated lncRNAs, such as FIRRE and C_0T-1 RNA, was reported to defuse from their nuclear foci following hnRNPU loss or mutations (Lu et al., 2017; Kolpa et al., 2021), and hnRNPU's nuclear distribution correlated significantly with the nuclear RNA distribution, but not with the DAPI-stained DNA (Creamer et al., 2021). These findings indicated that hnRNPU might be directly tethering chromatin-associated non-coding RNA.

On the other hand, the reduction of chromatin associated MALAT1 following hnRNPU KO hinted an alternative explanation. MALAT1 was reported to associate with promoter proximal genome sites and nascent RNA through base-pairing with U1 snRNA (Engreitz et al., 2014). More recently, a study showed that rapid depletion of U1 snRNP component U1-70K dissociated hundreds of lncRNAs from the chromatin and again revealed that direct U1 snRNP binding to chromatin associated RNAs was integral to their tethering on chromatin and that in many cases this occurred via RNA

polymerase II (Yin et al., 2020). Following hnRNPU KO we see loss of U1 snRNA from the chromatin fraction (Figure III.2.1), together with increased splicing of chromatin associated RNAs, including lncRNAs (Figure III.6.2A)). Therefore we speculate that by preventing progression of the spliceosome beyond E complex formation, through direct inhibition of U2 snRNP, hnRNPU tethers both U1 and U2 snRNP in the chromatin environment together with the RNAs directly bound by U1 and U2 snRNP. Following hnRNPU KO, U2 snRNP inhibition is relieved allowing progression of splicing, ultimately leading to release of U1 snRNP and the chromatin associated RNA, an extreme example of this is exemplified by NR4A1 pre-mRNA where the majority of the transcript is found unspliced and chromatin associated in normal cells and following hnRNPU depletion, an increase in nuclear and cytoplasmic spliced mRNA is observed, though the underlying principals appear to apply widely across both pre-mRNAs and lncRNAs.

The chromatin-associated RNA scaffold or nuclear matrix, which mainly comprises of nascent intronic RNA, has been implicated in maintaining an open chromatin structure (Creamer et al., 2021). Moreover, chromatin-associated non-coding RNA elements, such as enhancers and PROMPTs, contribute to the silencing and expression of gene loci and formation of transcription hubs (reviewed in Li & Fu, 2019). HnRNPU was also implicated in chromatin structure arrangement, as it was reported to interact with similar DNA loci as the transcription loop forming proteins CTCF and RAD21, and its loss alternated boundaries of topologically associated domains and leads to chromatin compaction (Fan et al., 2018; Nozawa e al., 2017). More specifically, Sharp et al. (2020) showed that, in prophase, Aurora-B kinase phosphorylates the hnRNPU SAP domain and evicts chromatin associated hnRNPU and hnRNAs and this eviction facilitates the chromatin condensation for cell division. Nozawa et al. (2017) also demonstrated that the ATP-dependent oligomerisation of hnRNPU promotes the decompaction of chromatin structures. Together these observations indicated that hnRNPU facilitates the decompaction of chromatin, probably through its interaction with a chromatin-associated RNA scaffold. We speculate that this could be a consequence of splicing inhibition caused by hnRNPU-U2 snRNP interaction. Consistent with this model, a recent report showed that splicing inhibition resulted in decompaction of transcribed gene loci (Leidescher et al., 2022). This explanation is also supported by our observation that the Walker A motif mutant failed to recover the increased splicing

caused by hnRNPU KO (Figure III.11.1), and this mutant did not rescue the chromatin compaction following hnRNPU siRNA knock down in the earlier study of Nozawa et al. (2017). Therefore, we propose that the rigid polymers produced by hnRNPs decorating intronic RNA, which have been shown to drive transcription loop formation and a more open chromatin structure (Leidescher et al., 2022), are regulated by hnRNPU which controls the levels of intronic RNA present through splicing inhibition.

VI.3. The opposing activities of hnRNPUL1 and hnRNPU

In Chapter IV, I studied the transcription and splicing effects of hnRNPUL1 KO in a doxycycline inducible AID cell line. Dr. Llewlyn Griffith, Dr. Ivaylo Yonchev in the Wilson Lab have identified increased histone pre-mRNA misprocessing and reduce U1, U2 and U4 snRNA levels in hnRNPUL1 KO cells (Griffith, 2019; Yonchev, 2021). We first verified that U7 snRNA level significantly reduced in hnRNPUL1 KO cells (Figure IV.1.1), in accordance to reported roles of hnRNPUL1 in facilitating U7 snRNP-mediated processing of histone pre-mRNA (Ideue et al., 2012).

Through a Gal4 DNA binding domain fusion reporter assay, we have identified the IDR of hnRNPUL1 as a weak transcription activator, yet truncations containing other domains of the protein would inhibit its activating potential (Figure IV.2.1). This corresponded well to the LLPS between PolII CTD and hnRNPUL1 IDR observed by Dr. Carmen Apostol (Apostol, 2023). We then conducted TT-Seq, and found that hnRNPUL1 KO resulted in global downregulation of PolII transcription (Figure IV.3.1) of protein coding genes, while exerting limited effects on co-transcriptional splicing (Figure IV.3.2). In the TT-Seq data, we also discovered that hnRNPUL1 KO stimulated transcription of IEGs and reduced 3′ cleavage and processing of PolII transcribed pre-snRNAs (Figure IV.5.1 and Figure IV.4.1). These transcription-associated effects of hnRNPUL1 KO are opposite to the transcription effects of hnRNPU KO (Figure III.4.1, Figure III.5.1 and Figure III.5.2). However, we did not find direct per-gene level negative correlation between these effects, in fact, at per-gene levels we observed a weak positive correlation (Figure VI.3.1), a possible result of Pearson's spurious correlation due to the shared HCT116 control condition. This indicated that hnRNPUL1 and hnRNPU might act opposite to each other in regulating the global activity of PolII, yet they may act in concert to drive transcription of specific genes.

An antagonist relationship between hnRNPU and hnRNPUL1 has been noted in their functions in DNA-damage repair, where hnRNPU was reported to be evicted from DNA-damage site while hnRNPUL1 was shown to be recruited to DNA double strand breaks (Polo et al. 2012; Britton et al., 2014). Dr. Yonchev have also noted that while hnRNPU associated with CTCF/RAD21 binding sites, hnRNPUL1 wild type protein did not, but its Walker A and Walker B mutants did (Yonchev, 2021). This implied that hnRNPUL1 might be excluded from hnRNPU binding loci until certain modifications or conformation changes take place.

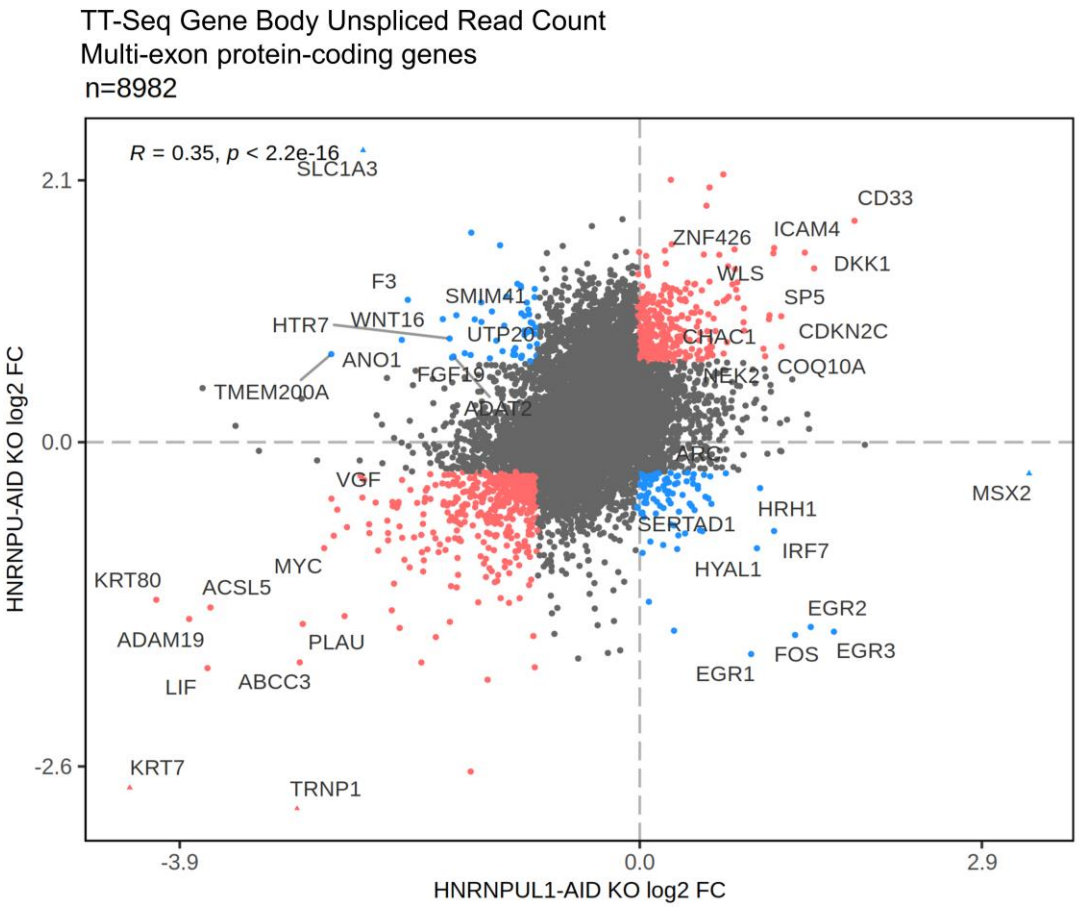

Figure VI.3.1. Correlation of hnRNPU and hnRNPUL1 KO transcription effects
A dot plot of log2 fold changes (log2 FC) in hnRNPU-AID and hnRNPUL1-AID KO TT-Seq unspliced read counts across 8982 multi-exon protein-coding genes, compared to HCT116 wild type TT-Seq. The R-values and p-value of a slightly positive (R=0.35) Pearson correlation are shown on the top left corner. The 10% most positively correlated genes are highlighted in red, while the 10% most negatively correlated genes are highlighted in blue. Note that FOS, EGR1, EGR2 and EGR3 are amongst the most negatively correlated genes.

VI.4. FUS-U1 snRNP axis in GC-content biased transcription regulation

In Chapter V, we established an AID2 system to study the effects of rapid depletion of FUS on transcription and splicing (Figure V.1.1), which binds both hnRNPU and hnRNPUL1. We verified that the FUS-AID2 cell line could reproduce the disruptions to Cajal bodies, snRNA level and histone mRNA processing at late time points (24 to 72 hours), as reported in FUS mutant studies and tested by Dr. Yonchev using FUS knock-down cells (Gadgil et al., 2021; Yonchev, 2021). Using TT-Seq, we showed that rapid KO of FUS in 6 hours did not dramatically disrupt the global PolII-driven transcription (Figure V.3.1). However, the transcription alterations caused by FUS rapid loss correlated negatively with the GC content of the first intron of the gene and took place mainly after the first 5′SS (Figure V.3.2). We showed that the GC-content biased transcription changes caused by FUS KO were also reflected in total RNA expression using RNA-Seq (Figure V.6.1).

Since GC rich regions in pre-mRNA tend to have lower transcriptional elongation rates, U1 snRNP promotes elongation (Mimoso & Adelman, 2023) and FUS is a prominent U1 snRNP interactor, our data suggest that FUS may directly recruit U1 snRNP to GC rich introns to promote their elongation. Consistent with this model, we found that FUS is enriched over the 5’ SS of GC-rich first introns in both K562 and HepG2 cells in ENCODE eCLIP data (Figure V.4.1 and Figure V.4.2). This enrichment over the 5’ SS is in agreement with the well-recognised interaction between FUS and U1 snRNA (Yu et al., 2015; Jutzi et al., 2020). Intriguingly, the preferentially enrichment of FUS over GC-rich introns coincides with its affinity towards GU-rich RNA-motifs (Wang et al., 2015), as well the high-affinity of the FUS RGG box for the G-quadruplex RNA secondary structure (McAninch & Mihailescu, 2016; Yagi et al., 2018; Imperatore et al., 2020). The high G-affinity of FUS raises the possibility that GC rich intronic RNA sequences form G-quadruplexes upon transcription and recruit or stabilise U1 snRNA-associated FUS to form phase-separated hubs of transcription factors and promote transcription at the 5′ SS.

The high affinity between U1 snRNP and FUS reported by Yu & Reed, 2015 also raises the possibility that FUS regulates the U1 telescripting of multi-exon PolII transcripts. Multi-exon genes and their introns, both coding and non-coding, shows a negative correlation between length and GC-content (Figure VI.4.1). This makes longer AT-rich genes and introns more susceptible to the occurrence of premature PAS, which are AU-

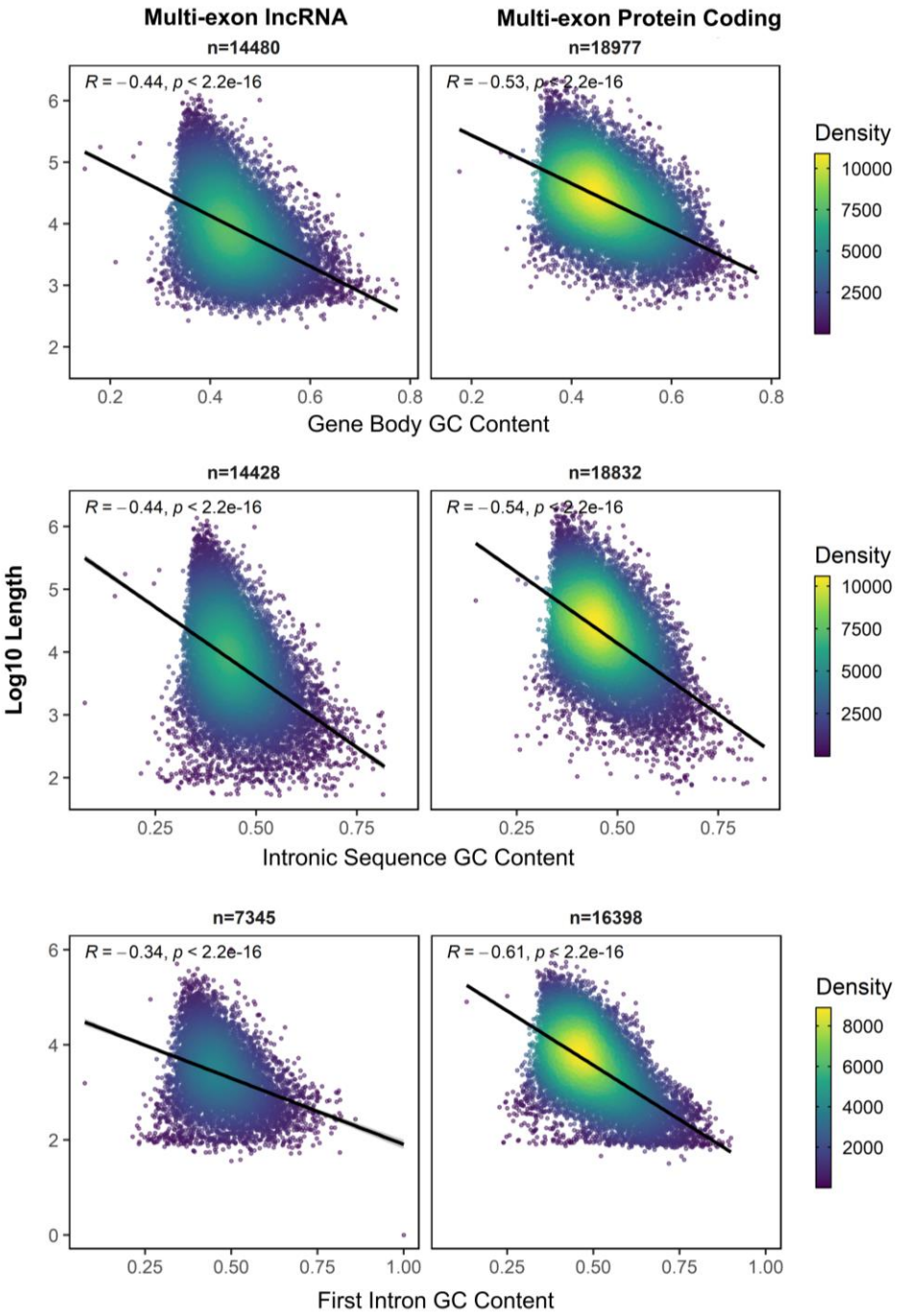

Figure VI.4.1. GC-length correlations of multi-exon gene bodies, intronic sequences and first introns.

GC content (x-axis) of HCT116 multi-exon lncRNA (left) and protein-coding (right) gene bodies (top), intronic sequences (middle) and first introns (bottom) negatively correlate with their lengths (y-axis, log10-transformed) to various extent. Gene bodies and introns were based on primary human transcripts validated in FUS-AID2 total RNA-Seq data. Local density of each feature is coloured according to a colour scale as indicated for each feature on the right side. The lines of regression were drawn for each data set, and the R-value and the p-value of the Pearson correlation were labelled on the top-left corners.

rich motifs (Beaudoing et al., 2000). Since FUS shows high affinity for GC-rich first introns (Figure V.4.1; Figure V.4.2), and rapid FUS KO resulted in better transcription across AT-rich genes (Figure V.3.2), we hypothesise that a routine role of FUS might be recruiting U1 snRNP to short, GC-rich first introns and enhancing elongation across them, which, in turn, reduces the availability of U1 snRNP at AT-rich, long genes (Figure VI.4.2). Following FUS KO, U1 snRNP would be redistributed to longer AT rich genes, promoting their transcription. Thus, FUS may be able to regulate the balance of transcription between shorter GC and longer AT rich genes. Longer human genes were suggested to be more associated with early-development functions, while short genes are more involved in immune response and neurodegenerative disease (Lopes et al., 2021). Hence, elucidating the roles of FUS in regulating U1 telescripting might provide insight into transcriptome landscape shaping during metazoan development and neuron maintenance.

Since the well-characterised loss of Cajal body in FUS mutants or knock-down (Yamazaki et al., 2012; Yu et al., 2015; Jutzi et al., 2020; Yonchev, 2021) only started occurring after 48 hours of FUS depletion in our AID2 system (Figure V.2.1 and Figure V.2.2), we speculate that the early and late effects of FUS loss are dissimilar and may be cell-cycle dependent. In addition, existing reports suggested that ALS-causing NLS-mutations of FUS lead to its cytoplasmic accumulation, where it interacts with mature RNA transcripts and ribosomes, and disrupts protein synthesis (Sephton et al., 2014; López-Erauskin et al., 2020; Sévigny et al., 2020). Since FUS KO altered the expression of ribosomal subunit mRNA in correspondence to the gene GC-content biased transcription effects (Figure V.7.1), we speculate that the loss of FUS from the nucleus may result in secondary effects on ribosomal dynamics and assembly, and this could contribute to the ALS pathology as well.

The global landscape of co-transcriptional splicing was not significantly disrupted by 6-hour FUS KO, however, it is interesting that the splicing ratios of the pre-mRNA of several splicing factors were downregulated slightly yet significantly (Figure V.5.1), and in particular that of the U1A gene (SNRPA). We could not identify significant changes in protein levels of U1A up to 72 hours of FUS KO (Figure V.5.2), but this effect might alter the long-term expression of U1A in a FUS-deficient cell line and manifest into further U1 snRNP-related side effects that may contribute to FUS-related disease. The lack of reduced U1A protein levels at 72 hours following FUS depletion

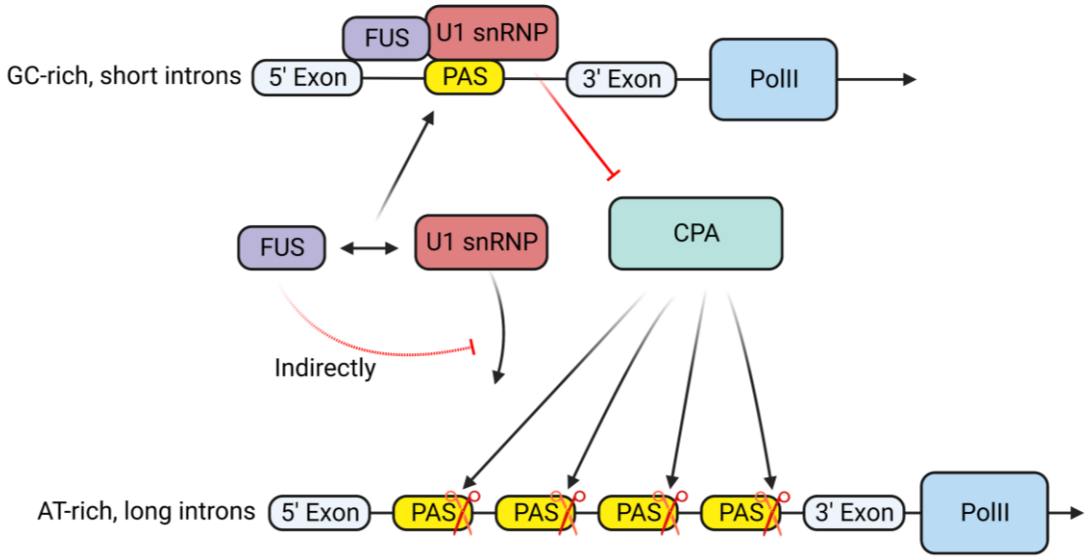

Figure VI.4.2. A model of FUS regulation of U1 telescripting and GC-bias in transcription.

Schematic of the possible model of FUS's function in regulating U1-telescripting and intron GC content bias in transcription. FUS, which has high affinity for G-rich sequences, recruits U1 snRNP to GC-rich introns, which tend to be shorter and less-prone to presence of premature PAS. U1 snRNP then inhibits the recognition and cleavage by the cleavage and polyadenylation (CPA) complex, promoting the completion and appropriate processing of the transcript. This, in turn, reduces the availability of U1 snRNP to the AT-rich, longer and more PAS-prone introns, leaving them susceptible to CPA-driven premature termination.

and subsequent inhibition of its splicing might also be explained by the ability of U1A to autoregulate its expression through regulation of its polyadenylation (Gunderson et al., 1994)..

It is worth noting that GC-bias in transcription and RNA expression analysis could be a result of bias in read depth. In the human transcriptome, the GC-content of genes correlate negatively with their length, and in our data sets of TT-Seq and RNA-Seq, the GC-content of genes correlate positively with their read depth. While this bias would not alter the correlation of the log2 fold changes calculated, as the bias would be present in both control and KO conditions, the conclusions and discussions drawn here for FUS KO-related GC-bias in transcription would need further experimental validation and cross-comparison with other hnRNPs to minimise the possibility of a false discovery.

VI.5. The regulatory network of hnRNPU, UL1 and FUS and neuron functions

HnRNPU, UL1, FUS and multiple other hnRNPs are associated with neuron-related disease in clinical cases. While this investigation has not provided all the answers to the functions of hnRNPU, hnRNPUL1 and FUS, a summary of the connections between these three proteins, particularly their functions related to splicing and transcription, may help to establish a corner of the regulatory network formed by hnRNPs, and how it may impact and facilitate neuron functions.

Firstly, hnRNPU might be regulating and co-operating with FUS to facilitate feedback of splicing machinery on PolII transcription and AP-1 IEG expression. We observed that hnRNPU KO decreased the expression of FUS (Figure III.1.1C). In the eCLIP datasets, both hnRNPU and hnRNPUL1 were highly enriched across the retained introns of FUS, and these introns were spliced more poorly in hnRNPU KO cells (Figure VI.5.1). This implies that hnRNPU promotes the expression of FUS by binding and facilitating the splicing of its retained introns. This would suggest that patients with impaired hnRNPU expression may suffer from side effects of increased retention of FUS introns, and possibly reduced FUS protein levels, which could be tested by assaying levels of FUS in patient fibroblasts. We also discovered that hnRNPU might be regulating U2 snRNP through direct interactions (Figure III.9.1B), and FUS is a well-established interactor and potential regulator of the U1 snRNP (Yu et al., 2015). This implied a possible hnRNPU-FUS axis in regulating U1-U2 snRNP functions in splicing and transcription. hnRNPU-FUS axis is further evidenced by the partial correlation between hnRNPU and FUS KO transcription effects on AT-rich genes (Figure V.8.1A). In addition, we identified that hnRNPU KO downregulated AP-1 family IEG expression by promoting splicing of the repressor NR4A1 (Figure III.5.3), while FUS mutations and depletions were reported to promote IEG expression (Yonchev, 2021). Hence, hnRNPU and FUS may balance AP-1 IEG expression, where hnRNPU promotes FUS and AP-1 IEG expression, and FUS represses AP-1 IEGs.

Secondly, hnRNPU and hnRNPUL1 might antagonize each other's activities. We observed that hnRNPUL1 protein was slightly upregulated in hnRNPU KO cells. (Figure III.1.1C). Our TT-Seq analyses revealed that hnRNPU KO and hnRNPUL1 KO cells had opposite effects on global PolII transcription, and their effects on IEG transcription showed the reverse trend, but still in opposition (Figure III.4.1; Figure IV.3.1; Figure III.5.2; Figure IV.5.1). In addition, the hnRNPU and hnRNPUL1-

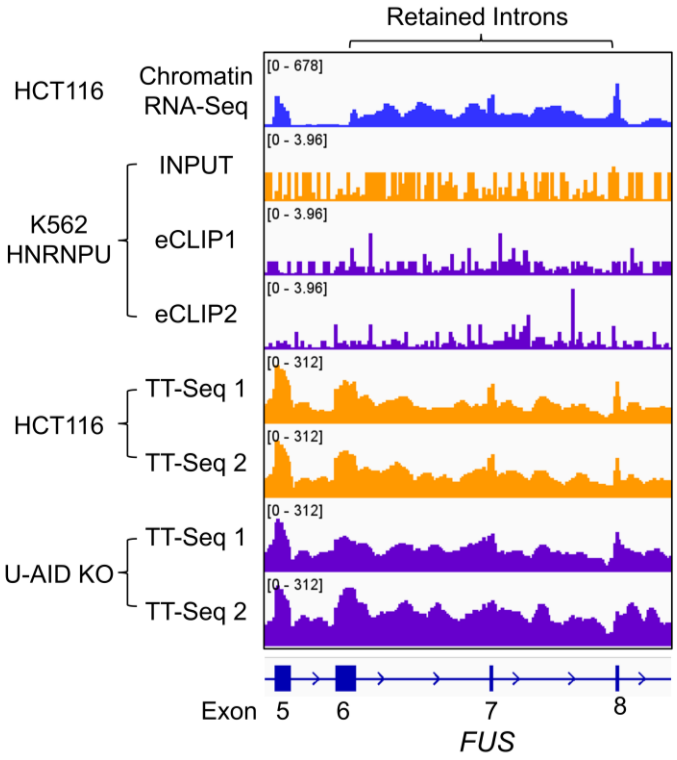

Figure VI.5.1. hnRNPU binds and regulates the retained introns of FUS

IGV-visualised signal depth of HCT116 chromatin RNA-Seq reads, eCLIP single-nucleotide signals in K562 hnRNPU INPUT and eCLIP, and HCT116 and hnRNPU-AID (U-AID) KO TT-Seq reads across FUS gene. hnRNPU showed enrichment across these introns and these introns were spliced worse co-transcriptionally in hnRNPU-AID KO cells. The intron retention was evidenced in the HCT116 chromatin RNA-Seq data. Signal depth of each dataset were normalized and group-autoscaled, as indicated by the signal depth value labelled on top left of each track lane.

enriched retained introns of FUS were spliced less efficiently in hnRNPU KO cells, while they appeared to have been spliced better in hnRNPUL1 KO cells (Yonchev, 2021). Since hnRNPUL1 is one of the top interactors of hnRNPU in our mass spectrometry data, it is possible that hnRNPU and hnRNPUL1's activities oppose each other through direct interactions and interference, as well as through indirect regulation of transcription and expression.

Finally, Dr. Griffith and Dr. Yonchev have established that hnRNPUL1 closely interacts with FUS, and both proteins are important in maintaining Cajal bodies and snRNA expression (Griffith, 2019; Yonchev, 2021). HnRNPUL1 and FUS also regulate U7 snRNP functions and the appropriate processing of RD histone mRNAs (Figure V.2.3) (Ideue et al., 2012; Gadgil et al., 2021; Yonchev, 2021), and long-term of loss of both proteins could upregulate AP-1 family IEGs, such as EGR1 and Fos (Griffith, 2019; Yonchev, 2021). In addition, both hnRNPUL1 and FUS contain intrinsically disordered domains that may interact with PolII CTD through LLPS (Kwon et al, 2013; Apostol, 2023). The similarities between their functions as well as their association with PolII and each other implied that they may be acting co-operatively in regulating the formation of Cajal bodies and the transcription and expression of snRNAs, RD histone mRNAs and IEGs. We did not observe the hnRNPUL1 KO-induced global downregulation of transcription in TT-Seq following FUS KO, but this could be a result of the long-term basal degradation in hnRNPUL1-AID cells, which caused the PolII transcription downregulation as a result of accumulated DNA-damage.

Interestingly, hnRNPU, hnRNPUL1 and FUS all affected expression of IEGs. Since the AP-1 family IEGs are crucial to neuron functions, the roles of IEG expression changes in neuronal disease associated with these proteins might be worth further investigation. There might also be a connection between the roles of these proteins in regulating RD histone gene expression and their associated neuropathy. The turnover and post-translational modifications of histone proteins have been shown to play crucial roles in modelling neuronal transcription landscapes (Maze et al., 2012; Maze et al., 2015). Lastly, we have also identified significant downregulation of RNA expression of a glutamate receptor subunit, GRIN2B, in both hnRNPU and hnRNPUL1 KO cells (Wilson Lab, unpublished). This gene has been associated with neurodevelopmental symptoms that overlaps with those of hnRNPU neurodevelopment disorder (Sabo et al.,

2023). Hence, further verification of the expression of this gene in patient conditions might inspire treatments for hnRNPU syndrome patients.

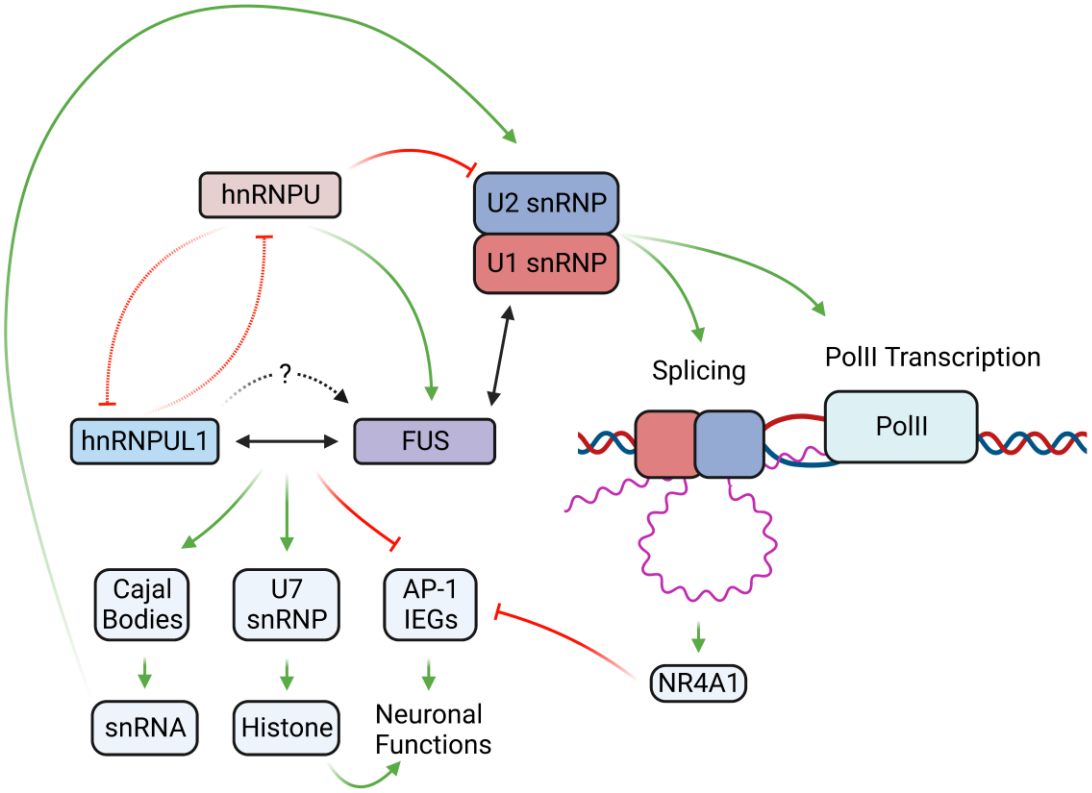

Figure VI.5.2. The regulatory network of hnRNPU, UL1 and FUS

A schematic of the potential regulatory network formed by hnRNPU, hnRNPUL1 and FUS and their effects on other mechanisms. hnRNPU facilitates the expression of FUS and inhibits U2 snRNP, while FUS interacts and regulates U1 snRNP activities. U1 and U2 snRNPs both positively regulate splicing and transcription. Splicing upregulation would promote NR4A1 expression, and lead to suppression of IEGs. hnRNPU and hnRNPUL1 evidently opposes each other's activities, while hnRNPUL1 and FUS associates with each other and act co-operatively in maintaining snRNA biogenesis by Cajal bodies and RD histone mRNA biogenesis by U7 snRNP. hnRNPUL1 and FUS also act to suppress the expression of IEGs, although the mechanisms of this suppression as well as the effects of hnRNPUL1 on FUS remain unclear at the current moment. Both IEGs and histone expression facilitates the appropriate development and functions of neurons.

VI.6. Evaluation of transcription and splicing analyses

This investigation into the transcription and splicing functions utilised established protocols for protein and RNA studies, such as subcellular fractionation, RNA-Seq and TT-Seq, which provided a solid basis for comparing and reproducing works presented here. However, due to the currently limited understanding of the human genome and transcriptome architecture and the ever-updating human genome annotations, improvements could be made on the transcription and splicing analyses presented here.

The majority of bioinformatic differential and metagene analysis conducted on transcription and abundance of gene features in this study was based on the selected set of most prevalent transcripts of genes as identified by Salmon. This approach simplifies the structure of each gene, however, it may have introduced noise and background from the less prevalent transcripts, which may carry separate transcription start sites, splice sites and termination sites. This may result in shift of metagene peaks and under- or over-representation of intronic reads. My manual inspection of transcript annotations and adjustments on parameters have ruled out the mis-selection of transcripts of highly expressed genes. However, for lowly expressed genes, the transcript used in this study needs to be carefully inspected for its accuracy as a representation as the gene structure.

The splicing analyses were mostly conducted through the DEXSeq package, where the only splice site reads, based on the alignment information from STAR, were considered. DEXSeq was designed as an exon usage analysis tool, but I found it mathematically suitable for computing the significance and changes in ratios between any two negative binomially distributed transcriptomic counting values. The other, and better recognised option was the rmats package. Despite that rmats package is better statistically formulated and has more functions, its output is heavily dependent on the accuracy of the transcriptome annotation provided. Unfortunately, within the time frame of this project, I was not able to rule out the rmats consideration of aberrantly included exons as evidence for intron retention, and hence the lack of intron retention analysis in this investiagtion. Additional work on filtering and correcting the transcriptome annotations, based on experimental data for each cell line, would be required to fully utilise the power and accuracy of rmats for genome wide splicing analysis that may provide confident and significant targets for wet-lab validations.

VII. Reference

Akiyama, T. *et al.* (2018) 'Aberrant axon branching via Fos-B dysregulation in FUS-ALS motor neurons', *SSRN Electronic Journal* [Preprint]. doi:10.2139/ssrn.3294755.

Amodio, N. *et al.* (2018) 'Malat1: A druggable long non-coding RNA for targeted anti-cancer approaches', *Journal of Hematology & Oncology*, 11(1). doi:10.1186/s13045-018-0606-4.

An, H. *et al.* (2019) 'Als-linked FUS mutations confer loss and gain of function in the nucleus by promoting excessive formation of dysfunctional paraspeckles', *Acta Neuropathologica Communications*, 7(1). doi:10.1186/s40478-019-0658-x.

Aslanzadeh, V. *et al.* (2017) 'Transcription rate strongly affects splicing fidelity and cotranscriptionality in budding yeast', *Genome Research*, 28(2), pp. 203–213. doi:10.1101/gr.225615.117.

Bachi, A. *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(1), pp. 136–158. doi:10.1017/s1355838200991994.

Bagchi, D. *et al.* (2020) 'Adult-onset myopathy with constitutive activation of Akt following the loss of HNRNP-U', *iScience*, 23(7), p. 101319. doi:10.1016/j.isci.2020.101319.

Bai, R. *et al.* (2018) 'Structures of the fully assembled *saccharomyces cerevisiae* spliceosome before activation', *Science*, 360(6396), pp. 1423–1429. doi:10.1126/science.aau0325.

Baillat, D. *et al.* (2005) 'Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II', *Cell*, 123(2), pp. 265–276. doi:10.1016/j.cell.2005.08.019.

Barral, P.M. *et al.* (2005) 'The interaction of the HNRNP family member E1B-AP5 with P53′, *FEBS Letters*, 579(13), pp. 2752–2758. doi:10.1016/j.febslet.2005.03.095.

Baserga, S. (1993) ‘The Diverse World of Small Ribonucleoproteins’, in J. Steitz (ed.) *The RNA World*. CSH Monographs, pp. 359–381.

Beaudoing, E. *et al.* (2000) ‘Patterns of variant polyadenylation signal usage in human genes’, *Genome Research*, 10(7), pp. 1001–1010. doi:10.1101/gr.10.7.1001.

Behrens, S.E. *et al.* (1993) ‘Evidence that the 60-kDa protein of 17S U2 small nuclear ribonucleoprotein is immunologically and functionally related to the yeast PRP9 splicing factor and is required for the efficient formation of prespliceosomes.’, *Proceedings of the National Academy of Sciences*, 90(17), pp. 8229–8233. doi:10.1073/pnas.90.17.8229.

Berglund, F.M. and Clarke, P.R. (2009) ‘HNRNP-U is a specific DNA-dependent protein kinase substrate phosphorylated in response to DNA double-strand breaks’, *Biochemical and Biophysical Research Communications*, 381(1), pp. 59–64. doi:10.1016/j.bbrc.2009.02.019.

Berglund, J.A., Abovich, N. and Rosbash, M. (1998) ‘A cooperative interaction between U2AF65 and MBBP/SF1 facilitates branchpoint region recognition’, *Genes & Development*, 12(6), pp. 858–867. doi:10.1101/gad.12.6.858.

Biorender (no date) *BioRender*. Available at: https://biorender.com/ (Accessed: 23 January 2023).

Blackwell, D.L. *et al.* (2022) ‘HNRNPUL1 controls transcription, splicing, and modulates skeletal and limb development in vivo’, *G3 Genes|Genomes|Genetics*, 12(5). doi:10.1093/g3journal/jkac067.

Bresson, S.M. *et al.* (2015) ‘Canonical poly(a) polymerase activity promotes the decay of a wide variety of mammalian nuclear RNAS’, *PLOS Genetics*, 11(10). doi:10.1371/journal.pgen.1005610.

Britton, S. *et al.* (2014) ‘DNA damage triggers Saf-A and RNA biogenesis factors exclusion from chromatin coupled to R-loops removal’, *Nucleic Acids Research*, 42(14), pp. 9047–9062. doi:10.1093/nar/gku601.

Brosi, R., Hauri, H.P. and Krämer, A. (1993) ‘Separation of splicing factor SF3 into

two components and purification of SF3a activity', *Journal of Biological Chemistry*, 268(23), pp. 17640–17646. doi:10.1016/s0021-9258(19)85380-2.

Bugai, A. et al. (2019) 'P-tefb activation by RBM7 shapes a pro-survival transcriptional response to genotoxic stress', Molecular Cell, 74(2). doi:10.1016/j.molcel.2019.01.033.

C. Quaresma, A.J., Bugai, A. and Barboric, M. (2016) 'Cracking the control of RNA polymerase II elongation by 7sk snrnp and p-tefb', *Nucleic Acids Research*, 44(16), pp. 7527–7539. doi:10.1093/nar/gkw585.

Caizzi, L. *et al.* (2021) 'Efficient RNA polymerase II pause release requires U2 snRNP function', *Molecular Cell*, 81(9), pp. 1920-1934.e9. doi:10.1016/j.molcel.2021.02.016.

Cajal, R. y S. (1910) 'El núcleo de las células piramidales del cerebro humano y de algunos mamiferos', *Trab. Lab. Invest. Biol.*, 8, pp. 27–62.

Cakmakci, N.G. *et al.* (2008) 'Slip1, a factor required for activation of histone mrna translation by the stem-loop binding protein', *Molecular and Cellular Biology*, 28(3), pp. 1182–1194. doi:10.1128/mcb.01500-07.

Cao, W. *et al.* (2012) 'Control of alternative splicing by forskolin through hnRNP K during neuronal differentiation', *Nucleic Acids Research*, 40(16), pp. 8059–8071. doi:10.1093/nar/gks504.

Carmo-Fonseca, M. *et al.* (1992) 'Transcription-dependent colocalization of the U1, U2, U4/U6, and U5 snrnps in coiled bodies', *Journal of Cell Biology*, 117(1), pp. 1–14. doi:10.1083/jcb.117.1.1.

Cavigelli, M. *et al.* (1995) 'Induction of c-FOS expression through JNK-mediated TCF/elk-1 phosphorylation.', *The EMBO Journal*, 14(23), pp. 5957–5964. doi:10.1002/j.1460-2075.1995.tb00284.x.

Chapman, R.D. *et al.* (2008) 'Molecular evolution of the RNA polymerase II CTD', *Trends in Genetics*, 24(6), pp. 289–296. doi:10.1016/j.tig.2008.03.010.

Chen, F.X., Smith, E.R. and Shilatifard, A. (2018) 'Born to run: control of transcription

elongation by RNA polymerase II', *Nature Reviews Molecular Cell Biology*, 19(7), pp. 464–478. doi:10.1038/s41580-018-0010-5.

Chen, L.-L. and Carmichael, G.G. (2009) 'Altered nuclear retention of mrnas containing inverted repeats in human embryonic stem cells: Functional role of a nuclear noncoding RNA', *Molecular Cell*, 35(4), pp. 467–478. doi:10.1016/j.molcel.2009.06.027.

Chi, B., O'Connell, Jeremy D, *et al.* (2018) 'The neurodegenerative diseases ALS and SMA are linked at the molecular level via the ASC-1 Complex', *Nucleic Acids Research*, 46(22), pp. 11939–11951. doi:10.1093/nar/gky1093.

Chi, B., O'Connell, Jeremy D., *et al.* (2018) 'Interactome analyses revealed that the U1 snRNP machinery overlaps extensively with the RNAP II machinery and contains multiple ALS/SMA-causative proteins', *Scientific Reports*, 8(1). doi:10.1038/s41598-018-27136-3.

Chinenov, Y. and Kerppola, T.K. (2001) 'Close encounters of many kinds: Fos-jun interactions that mediate transcription regulatory specificity', *Oncogene*, 20(19), pp. 2438–2452. doi:10.1038/sj.onc.1204385.

Clark, C.E., Hasan, M. and Bousso, P. (2011) 'A role for the immediate early gene product C-FOS in imprinting T cells with short-term memory for signal summation', *PLoS ONE*, 6(4). doi:10.1371/journal.pone.0018916.

Cochran, B.H. *et al.* (1984) 'Expression of the c-*fos* gene and of an *fos*-related gene is stimulated by platelet-derived growth factor', *Science*, 226(4678), pp. 1080–1082. doi:10.1126/science.6093261.

Cooper-Knock, J. *et al.* (2017) 'Targeted genetic screen in amyotrophic lateral sclerosis reveals novel genetic variants with synergistic effect on clinical phenotype', *Frontiers in Molecular Neuroscience*, 10. doi:10.3389/fnmol.2017.00370.

Courchaine, E. *et al.* (2022) 'The coilin N-terminus mediates multivalent interactions between coilin and NOPP140 to form and maintain Cajal Bodies', *Nature Communications*, 13(1). doi:10.1038/s41467-022-33434-2.

Cramer, P. *et al.* (2008) 'Structure of Eukaryotic RNA Polymerases', *Annual Review of Biophysics*, 37(1), pp. 337–352. doi:10.1146/annurev.biophys.37.032807.130008.

Creamer, K.M., Kolpa, H.J. and Lawrence, J.B. (2021) 'Nascent RNA scaffolds contribute to chromosome territory architecture and counter chromatin compaction', *Molecular Cell*, 81(17). doi:10.1016/j.molcel.2021.07.004.

Cretu, C. *et al.* (2016) 'Molecular Architecture of SF3b and Structural Consequences of Its Cancer-Related Mutations', *Molecular Cell*, 64(2), pp. 307–319. doi:10.1016/j.molcel.2016.08.036.

Crozat, A. *et al.* (1993) 'Fusion of chop to a novel RNA-binding protein in human myxoid liposarcoma', *Nature*, 363(6430), pp. 640–644. doi:10.1038/363640a0.

Curran, T. and Franza, B.R. (1988) 'Fos and jun: The AP-1 connection', *Cell*, 55(3), pp. 395–397. doi:10.1016/0092-8674(88)90024-4.

Dahlberg, J.E. *et al.* (1990) 'Formation of the 3′ end of U1 snrna', *Molecular Biology Reports*, 14(2–3), pp. 161–162. doi:10.1007/bf00360457.

Darman, R.B. *et al.* (2015) 'Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3′ Splice Site Selection through Use of a Different Branch Point', *Cell Reports*, 13(5), pp. 1033–1045. doi:10.1016/j.celrep.2015.09.053.

Darzacq, X. (2002) 'Cajal body-specific small nuclear RNAS: A novel class of 2'-O-methylation and pseudouridylation guide RNAS', *The EMBO Journal*, 21(11), pp. 2746–2756. doi:10.1093/emboj/21.11.2746.

Das, R., Zhou, Z. and Reed, R. (2000) 'Functional Association of U2 snrnp with the ATP-independent spliceosomal complex E', Molecular Cell, 5(5), pp. 779–787. doi:10.1016/s1097-2765(00)80318-4.

Deng, Q. *et al.* (2014) 'FUS is phosphorylated by DNA-PK and accumulates in the cytoplasm after DNA damage', *Journal of Neuroscience*, 34(23), pp. 7802–7813. doi:10.1523/jneurosci.0172-14.2014.

Dobin, A. *et al.* (2013) 'Star: Ultrafast universal RNA-seq aligner', *Bioinformatics*,

29(1), pp. 15–21. doi:10.1093/bioinformatics/bts635.

Dominski, Z. *et al.* (1999) 'Stem-loop binding protein facilitates 3′-end formation by stabilizing U7 snrnp binding to histone pre-mrna', *Molecular and Cellular Biology*, 19(5), pp. 3561–3570. doi:10.1128/mcb.19.5.3561.

Dormann, D. *et al.* (2012) 'Arginine methylation next to the PY-NLS modulates transportin binding and nuclear import of Fus', *The EMBO Journal*, 31(22), pp. 4258–4275. doi:10.1038/emboj.2012.261.

Engreitz, J.M. *et al.* (2014) 'RNA-RNA interactions enable specific targeting of noncoding RNAS to nascent Pre-mRNAs and chromatin sites', *Cell*, 159(1), pp. 188–199. doi:10.1016/j.cell.2014.08.018.

Fan, H. *et al.* (2017) 'The nuclear matrix protein HNRNPU maintains 3D genome architecture globally in mouse hepatocytes', *Genome Research*, 28(2), pp. 192–202. doi:10.1101/gr.224576.117.

Fisher, R.P. (2018) 'CDK7: A kinase at the core of transcription and in the crosshairs of cancer drug discovery', *Transcription*, 10(2), pp. 47–56. doi:10.1080/21541264.2018.1553483.

Fornerod, M. *et al.* (1997) 'CRM1 is an export receptor for leucine-rich nuclear export signals', *Cell*, 90(6), pp. 1051–1060. doi:10.1016/s0092-8674(00)80371-2.

Fox, A.H. *et al.* (2002) 'Paraspeckles: A Novel Nuclear Domain', *Current Biology*, 12(1), pp. 13–25. doi:10.1016/s0960-9822(01)00632-7.

Fujii, M. *et al.* (2000) 'Activation of oncogenic transcription factor AP-1 in T cells infected with human T cell leukemia virus type 1', *AIDS Research and Human Retroviruses*, 16(16), pp. 1603–1606. doi:10.1089/08892220050193029.

Gabler, S. *et al.* (1998) 'E1B 55-kilodalton-associated protein: A cellular protein with RNA-binding activity implicated in nucleocytoplasmic transport of adenovirus and cellular mrnas', *Journal of Virology*, 72(10), pp. 7960–7971. doi:10.1128/jvi.72.10.7960-7971.1998.

Gadgil, A. *et al.* (2021) 'Als-linked Fus mutants affect the localization of U7 snRNP

and replication-dependent histone gene expression in human cells', *Scientific Reports*, 11(1). doi:10.1038/s41598-021-91453-3.

Geuens, T., Bouhy, D. and Timmerman, V. (2016) 'The HNRNP family: Insights into their role in health and disease', *Human Genetics*, 135(8), pp. 851–867. doi:10.1007/s00439-016-1683-5.

Giacometti, S. *et al.* (2017) 'Mutually exclusive CBC-containing complexes contribute to RNA fate', *Cell Reports*, 18(11), pp. 2635–2650. doi:10.1016/j.celrep.2017.02.046.

Gillentine, M.A. *et al.* (2021) 'Rare deleterious mutations of HNRNP genes result in shared neurodevelopmental disorders', *Genome Medicine*, 13(1). doi:10.1186/s13073-021-00870-6.

Girard, C. (2006) 'Depletion of SMN by RNA interference in Hela cells induces defects in Cajal Body Formation', *Nucleic Acids Research*, 34(10), pp. 2925–2932. doi:10.1093/nar/gkl374.

Glover-Cutter, K. *et al.* (2009) 'TFIIH-associated CDK7 kinase functions in phosphorylation of C-terminal domain SER7 residues, promoter-proximal pausing, and termination by RNA polymerase II', *Molecular and Cellular Biology*, 29(20), pp. 5455–5464. doi:10.1128/mcb.00637-09.

Gomes, G.L. and Scortecci, K.C. (2021) 'Auxin and its role in plant development: Structure, signalling, regulation and Response Mechanisms', *Plant Biology*, 23(6), pp. 894–904. doi:10.1111/plb.13303.

Gozani, O., Potashkin, J. and Reed, R. (1998) 'A potential role for U2AF-SAP 155 interactions in recruiting U2 snrnp to the Branch Site', *Molecular and Cellular Biology*, 18(8), pp. 4752–4760. doi:10.1128/mcb.18.8.4752.

Greber, B.J. and Nogales, E. (2019) 'The Structures of Eukaryotic Transcription Pre-initiation Complexes and Their Functional Implications', *Subcellular Biochemistry*, pp. 143–192. doi:10.1007/978-3-030-28151-9_5.

Greenberg, M.E. and Ziff, E.B. (1984) 'Stimulation of 3T3 cells induces transcription

of the c-FOS proto-oncogene', *Nature*, 311(5985), pp. 433–438. doi:10.1038/311433a0.

Gregersen, L.H., Mitter, R. and Svejstrup, J.Q. (2020) 'Using TTchem-seq for profiling nascent transcription and measuring transcript elongation', *Nature Protocols*, 15(2), pp. 604–627. doi:10.1038/s41596-019-0262-3.

Gressel, S., Schwalb, B. and Cramer, P. (2019) 'The pause-initiation limit restricts transcription activation in human cells', Nature Communications, 10(1). doi:10.1038/s41467-019-11536-8.

Griffith, L. (2019) *An investigation of the cellular functions of HNRNPUL1*. thesis. The University of Sheffield.

Gunderson, S.I. et al. (1994) 'The human U1A snrnp protein regulates polyadenylation via a direct interaction with poly(a) polymerase', Cell, 76(3), pp. 531–541. doi:10.1016/0092-8674(94)90116-3.

Guo, H. *et al.* (2021) 'NR4A1 regulates expression of immediate early genes, suppressing replication stress in cancer', *Molecular Cell*, 81(19). doi:10.1016/j.molcel.2021.09.016.

Guo, Y.E. et al. (2019) 'Pol II phosphorylation regulates a switch between transcriptional and splicing condensates', Nature, 572(7770), pp. 543–548. doi:10.1038/s41586-019-1464-0.

Guru, S.C. *et al.* (1997) 'A 2.8-MB clone contig of the multiple endocrine neoplasia type 1 (MEN1) region at 11q13′, *Genomics*, 42(3), pp. 436–445. doi:10.1006/geno.1997.4783.

Gurunathan, G. *et al.* (2015) 'Arginine methylation of HNRNPUL1 regulates interaction with NBS1 and recruitment to sites of DNA damage', *Scientific Reports*, 5(1). doi:10.1038/srep10475.

Göhring, F. and Fackelmayer, F.O. (1997) 'The Scaffold/matrix attachment region binding protein hnrnp-U (SAF-A) is directly bound to chromosomal DNA *in vivo*: a chemical cross-linking study', *Biochemistry*, 36(27), pp. 8276–8283.

doi:10.1021/bi970480f.

Hacisuleyman, E. *et al.* (2014) 'Topological organization of multichromosomal regions by the long intergenic noncoding RNA firre', *Nature Structural & Molecular Biology*, 21(2), pp. 198–206. doi:10.1038/nsmb.2764.

Hai, T. and Curran, T. (1991) 'Cross-family dimerization of transcription factors fos/jun and ATF/CREB alters DNA binding specificity.', *Proceedings of the National Academy of Sciences*, 88(9), pp. 3720–3724. doi:10.1073/pnas.88.9.3720.

Hall, J., Thomas, K.L. and Everitt, B.J. (2001) 'Cellular imaging of*zif268*expression in the hippocampus and amygdala during contextual and cued fear memory retrieval: Selective activation of hippocampal CA1 neurons during the recall of Contextual Memories', *The Journal of Neuroscience*, 21(6), pp. 2186–2193. doi:10.1523/jneurosci.21-06-02186.2001.

Hallais, M. *et al.* (2013) 'CBC–ARS2 stimulates 3′-end maturation of multiple RNA families and favors cap-proximal processing', *Nature Structural & Molecular Biology*, 20(12), pp. 1358–1366. doi:10.1038/nsmb.2720.

Hasegawa, Y. *et al.* (2010) 'The matrix protein hnRNP U is required for chromosomal localization of Xist RNA', *Developmental Cell*, 19(3), pp. 469–476. doi:10.1016/j.devcel.2010.08.006.

Heath, C.G. (2018) *The localisation of non-coding RNA in the human cell*. thesis. The University of Sheffield.

Helbig, R. and Fackelmayer, F.O. (2003) 'Scaffold attachment factor A (SAF-A) is concentrated in inactive X chromosome territories through its RGG domain', *Chromosoma*, 112(4), pp. 173–182. doi:10.1007/s00412-003-0258-0.

Herschman, H.R. (1991) 'Primary response genes induced by growth factors and tumor promoters', *Annual Review of Biochemistry*, 60(1), pp. 281–319. doi:10.1146/annurev.bi.60.070191.001433.

Herzel, L. *et al.* (2017) 'Splicing and transcription touch base: co-transcriptional

spliceosome assembly and function', *Nature Reviews Molecular Cell Biology*, 18(10), pp. 637–650. doi:10.1038/nrm.2017.63.

Hong, Z. *et al.* (2013) 'The role of HNRPUL1 involved in DNA damage response is related to PARP1', *PLoS ONE*, 8(4). doi:10.1371/journal.pone.0060208.

Huber, J., Dickmanns, A. and Lührmann, R. (2002) 'The importin-β binding domain of SNURPORTIN1 is responsible for the RAN- and energy-independent nuclear import of spliceosomal U snrnps in vitro', *Journal of Cell Biology*, 156(3), pp. 467–479. doi:10.1083/jcb.200108114.

Huelga, S.C. *et al.* (2012) 'Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by HNRNP proteins', *Cell Reports*, 1(2), pp. 167–178. doi:10.1016/j.celrep.2012.02.001.

Hutchinson, J.N. *et al.* (2007) 'A screen for nuclear transcripts identifies two linked noncoding RNAS associated with SC35 splicing domains', *BMC Genomics*, 8(1). doi:10.1186/1471-2164-8-39.

Ichikawa, H. *et al.* (1999) 'Dual transforming activities of the FUS (TLS)-ERG leukemia fusion protein conferred by two N-terminal domains of FUS (TLS)', *Molecular and Cellular Biology*, 19(11), pp. 7639–7650. doi:10.1128/mcb.19.11.7639.

Ideue, T. *et al.* (2012) 'U7 small nuclear ribonucleoprotein represses histone gene transcription in cell cycle-arrested cells', *Proceedings of the National Academy of Sciences*, 109(15), pp. 5693–5698. doi:10.1073/pnas.1200523109.

Ielasi, F.S. *et al.* (2022) 'Human histone pre-mrna assembles histone or canonical mrna-processing complexes by overlapping 3′-end sequence elements', *Nucleic Acids Research*, 50(21), pp. 12425–12443. doi:10.1093/nar/gkac878.

Imperatore, J.A. *et al.* (2020) 'FUS recognizes g quadruplex structures within neuronal mrnas', *Frontiers in Molecular Biosciences*, 7. doi:10.3389/fmolb.2020.00006.

Jady, B.E. and Kiss, T. (2001) 'A small nucleolar guide RNA functions both in 2'-O-ribose methylation and pseudouridylation of the U5 spliceosomal RNA', *The*

EMBO Journal, 20(3), pp. 541–551. doi:10.1093/emboj/20.3.541.

Jasnovidova, O. and Stefl, R. (2012) 'The CTD code of RNA polymerase II: a structural view', *Wiley Interdisciplinary Reviews: RNA*, 4(1), pp. 1–16. doi:10.1002/wrna.1138.

Jones, M.W. *et al.* (2001) 'A requirement for the immediate early gene Zif268 in the expression of late LTP and long-term memories', *Nature Neuroscience*, 4(3), pp. 289–296. doi:10.1038/85138.

Jutzi, D. *et al.* (2020) 'Aberrant interaction of FUS with the U1 snrna provides a molecular mechanism of FUS induced amyotrophic lateral sclerosis', *Nature Communications*, 11(1). doi:10.1038/s41467-020-20191-3.

Järvelin, A.I. *et al.* (2016) 'The new (dis)order in RNA regulation', *Cell Communication and Signaling*, 14(1). doi:10.1186/s12964-016-0132-3.

Kaida, D. *et al.* (2010) 'U1 SNRNP protects pre-mrnas from premature cleavage and polyadenylation', *Nature*, 468(7324), pp. 664–668. doi:10.1038/nature09479.

Kastner, B. *et al.* (2019) 'Structural Insights into Nuclear pre-mRNA Splicing in Higher Eukaryotes', *Cold Spring Harbor Perspectives in Biology*, 11(11), p. a032417. doi:10.1101/cshperspect.a032417.

Kato, M. *et al.* (2012) 'Cell-free formation of RNA granules: Low complexity sequence domains form dynamic fibers within hydrogels', *Cell*, 149(4), pp. 753–767. doi:10.1016/j.cell.2012.04.017.

Kharman-Biz, A. *et al.* (2013) 'Expression of activator protein-1 (AP-1) family members in breast cancer', *BMC Cancer*, 13(1). doi:10.1186/1471-2407-13-441.

Kiledjian, M. and Dreyfuss, G. (1992) 'Primary structure and binding activity of the hnrnp U protein: Binding RNA through RGG box.', *The EMBO Journal*, 11(7), pp. 2655–2664. doi:10.1002/j.1460-2075.1992.tb05331.x.

Kim, D.D.Y. *et al.* (2004) 'Widespread RNA editing of embedded *alu* elements in the human transcriptome', *Genome Research*, 14(9), pp. 1719–1725.

doi:10.1101/gr.2855504.

Kim, M.K. and Nikodem, V.M. (1999) ‘HnRNP u inhibits carboxy-terminal domain phosphorylation by TFIIH and represses RNA polymerase II elongation’, *Molecular and Cellular Biology*, 19(10), pp. 6833–6844. doi:10.1128/mcb.19.10.6833.

King, O.D., Gitler, A.D. and Shorter, J. (2012) ‘The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease’, *Brain Research*, 1462, pp. 61–80. doi:10.1016/j.brainres.2012.01.016.

Koga, M. *et al.* (2014) ‘U2 snRNP is required for expression of the 3′ end of genes’, *PLoS ONE*, 9(5). doi:10.1371/journal.pone.0098015.

Kolpa, H.J. *et al.* (2021) ‘Saf-A mutants disrupt chromatin structure through dominant negative effects on RNAs associated with chromatin’, *Mammalian Genome*, 33(2), pp. 366–381. doi:10.1007/s00335-021-09935-8.

Kolpa, H.J., Fackelmayer, F.O. and Lawrence, J.B. (2016) ‘Saf-A requirement in anchoring Xist RNA to chromatin varies in transformed and primary cells’, *Developmental Cell*, 39(1), pp. 9–10. doi:10.1016/j.devcel.2016.09.021.

Krumm, A. *et al.* (1998) ‘Long-distance transcriptional enhancement by the histone acetyltransferase PCAF’, *Proceedings of the National Academy of Sciences*, 95(23), pp. 13501–13506. doi:10.1073/pnas.95.23.13501.

Kwon, I. *et al.* (2013) ‘Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains’, *Cell*, 155(5), pp. 1049–1060. doi:10.1016/j.cell.2013.10.033.

Kzhyshkowska, J. *et al.* (2003) ‘Regulation of transcription by the heterogeneous nuclear ribonucleoprotein E1B-AP5 is mediated by complex formation with the novel bromodomain-containing protein BRD7’, *Biochemical Journal*, 371(2), pp. 385–393. doi:10.1042/bj20021281.

Lagier-Tourenne, C., Polymenidou, M. and Cleveland, D.W. (2010) ‘TDP-43 and FUS/TLS: Emerging roles in RNA processing and neurodegeneration’, *Human*

Molecular Genetics, 19(R1). doi:10.1093/hmg/ddq137.

Law, W.J. (2006) 'TLS, EWS and TAF15: A model for transcriptional integration of gene expression', *Briefings in Functional Genomics and Proteomics*, 5(1), pp. 8–14. doi:10.1093/bfgp/ell015.

Lee, M. *et al.* (2020) 'Molecular structure and interactions within amyloid-like fibrils formed by a low-complexity protein sequence from Fus', *Nature Communications*, 11(1). doi:10.1038/s41467-020-19512-3.

Leidescher, S. *et al.* (2020) *Spatial organization of transcribed eukaryotic genes* [Preprint]. doi:10.1101/2020.05.20.106591.

Li, S. *et al.* (2019) 'An efficient auxin-inducible degron system with low basal degradation in human cells', *Nature Methods*, 16(9), pp. 866–869. doi:10.1038/s41592-019-0512-x.

Li, X. and Fu, X.-D. (2019) 'Chromatin-associated RNAS as facilitators of functional genomic interactions', *Nature Reviews Genetics*, 20(9), pp. 503–519. doi:10.1038/s41576-019-0135-1.

Li, Y.R. *et al.* (2013) 'Stress granules as crucibles of ALS pathogenesis', *Journal of Cell Biology*, 201(3), pp. 361–372. doi:10.1083/jcb.201302044.

Liang, Y. *et al.* (2021) 'HNRNPU promotes the progression of hepatocellular carcinoma by enhancing cdk2 transcription', *Experimental Cell Research*, 409(1), p. 112898. doi:10.1016/j.yexcr.2021.112898.

Loerch, S. *et al.* (2019) 'The pre-mrna splicing and transcription factor TAT-SF1 is a functional partner of the spliceosome sf3b1 subunit via a U2AF homology motif interface', *Journal of Biological Chemistry*, 294(8), pp. 2892–5793. doi:10.1074/jbc.ra118.006764.

Lopes, I. *et al.* (2021) 'Gene size matters: An analysis of gene length in the human genome', *Frontiers in Genetics*, 12. doi:10.3389/fgene.2021.559998.

Loughlin, F.E. *et al.* (2019) 'The solution structure of FUS bound to RNA reveals a bipartite mode of RNA recognition with both sequence and shape specificity',

Molecular Cell, 73(3). doi:10.1016/j.molcel.2018.11.012.

Lu, Y. *et al.* (2017) 'The NF-KB–responsive long noncoding RNA firre regulates posttranscriptional regulation of inflammatory gene expression through interacting with hnRNPU', *The Journal of Immunology*, 199(10), pp. 3571–3582. doi:10.4049/jimmunol.1700091.

Lynch, G. *et al.* (1983) 'Intracellular injections of EGTA block induction of hippocampal long-term potentiation', *Nature*, 305(5936), pp. 719–721. doi:10.1038/305719a0.

López-Erauskin, J. *et al.* (2020) 'ALS/FTD-linked mutation in FUS suppresses intra-axonal protein synthesis and drives disease without nuclear loss-of-function of Fus', *Neuron*, 106(2), p. 354. doi:10.1016/j.neuron.2020.04.006.

Machitani, M., Taniguchi, I. and Ohno, M. (2020) 'ARS2 regulates nuclear paraspeckle formation through 3′-end processing and stability of NEAT1 long noncoding RNA', *Molecular and Cellular Biology*, 40(4). doi:10.1128/mcb.00269-19.

Mahmoudi, S. *et al.* (2010) 'Wrap53 is essential for cajal body formation and for targeting the survival of motor neuron complex to Cajal Bodies', *PLoS Biology*, 8(11). doi:10.1371/journal.pbio.1000521.

Maji, D., Grossfield, A. and Kielkopf, C.L. (2019) 'Structures of SF3B1 reveal a dynamic achilles heel of spliceosome assembly: Implications for cancer-associated abnormalities and drug discovery', *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1862(11–12), p. 194440. doi:10.1016/j.bbagrm.2019.194440.

Marzluff, W.F. *et al.* (2002) 'The human and mouse replication-dependent histone genes', *Genomics*, 80(5), pp. 487–498. doi:10.1006/geno.2002.6850.

Marzluff, W.F., Wagner, E.J. and Duronio, R.J. (2008) 'Metabolism and regulation of canonical histone mrnas: Life without a poly(a) tail', *Nature Reviews Genetics*, 9(11), pp. 843–854. doi:10.1038/nrg2438.

Masuda, A. *et al.* (2015) 'Position-specific binding of FUS to nascent RNA regulates

mrna length', *Genes & Development*, 29(10), pp. 1045–1057. doi:10.1101/gad.255737.114.

Matera, A.G. (1999) 'Nuclear bodies: Multifaceted subdomains of the Interchromatin Space', *Trends in Cell Biology*, 9(8), pp. 302–309. doi:10.1016/s0962-8924(99)01606-2.

Mayeda, A., Helfman, D.M. and Krainer, A.R. (1993) 'Modulation of exon skipping and inclusion by heterogeneous nuclear ribonucleoprotein A1 and pre-mrna splicing factor SF2/ASF.', *Molecular and Cellular Biology*, 13(5), pp. 2993–3001. doi:10.1128/mcb.13.5.2993.

Maze, I. *et al.* (2015) 'Critical role of histone turnover in neuronal transcription and plasticity', *Neuron*, 87(1), pp. 77–94. doi:10.1016/j.neuron.2015.06.014.

Maze, I., Noh, K.-M. and Allis, C.D. (2012) 'Histone regulation in the CNS: Basic principles of epigenetic plasticity', *Neuropsychopharmacology*, 38(1), pp. 3–22. doi:10.1038/npp.2012.124.

Mazzucotelli, E. *et al.* (2006) 'The E3 ubiquitin ligase gene family in plants: Regulation by degradation', *Current Genomics*, 7(8), pp. 509–522. doi:10.2174/138920206779315728.

McAninch, D. and Mihailescu, M.-R. (2016) 'Fused in sarcoma (FUS) targets neuronal g quadruplex containing mrnas.', *Biophysical Journal*, 110(3). doi:10.1016/j.bpj.2015.11.1322.

McGinty, R.K. and Tan, S. (2014) 'Nucleosome structure and function', *Chemical Reviews*, 115(6), pp. 2255–2273. doi:10.1021/cr500373h.

Mimoso, C.A. and Adelman, K. (2023) 'U1 snrnp increases RNA Pol II elongation rate to enable synthesis of long genes', Molecular Cell, 83(8). doi:10.1016/j.molcel.2023.03.002.

Mouaikel, J. *et al.* (2003) 'Interaction between the small-nuclear-RNA cap hypermethylase and the spinal muscular atrophy protein, survival of motor neuron', *EMBO reports*, 4(6), pp. 616–622. doi:10.1038/sj.embor.embor863.

Murray, D.T. *et al.* (2017) 'Structure of FUS protein fibrils and its relevance to self-assembly and phase separation of low-complexity domains', *Cell*, 171(3). doi:10.1016/j.cell.2017.08.048.

Natsume, T. *et al.* (2016a) 'Rapid protein depletion in human cells by auxin-inducible degron tagging with short homology donors', *Cell Reports*, 15(1), pp. 210–218. doi:10.1016/j.celrep.2016.03.001.

Natsume, T. *et al.* (2016b) 'Rapid protein depletion in human cells by auxin-inducible degron tagging with short homology donors', *Cell Reports*, 15(1), pp. 210–218. doi:10.1016/j.celrep.2016.03.001.

Neugebauer, K.M. (2017) 'Special focus on the cajal body', *RNA Biology*, 14(6), pp. 669–670. doi:10.1080/15476286.2017.1316928.

Neyns, B. *et al.* (1996) 'Expression of the jun family of genes in human ovarian cancer and normal ovarian surface epithelium. Oncogene.', *Oncogene*, 12(6), pp. 1247–1257.

Nizami, Z., Deryusheva, S. and Gall, J.G. (2010) 'The Cajal body and histone locus body', *Cold Spring Harbor Perspectives in Biology*, 2(7). doi:10.1101/cshperspect.a000653.

Nojima, T. *et al.* (2018) 'RNA polymerase II phosphorylated on CTD serine 5 interacts with the spliceosome during co-transcriptional splicing', *Molecular Cell*, 72(2). doi:10.1016/j.molcel.2018.09.004.

Nozawa, R.-S. *et al.* (2017) 'SAF-A regulates interphase chromosome structure through oligomerization with chromatin-associated RNAS', *Cell*, 169(7). doi:10.1016/j.cell.2017.05.029.

Obrdlik, A. *et al.* (2008) 'The histone acetyltransferase PCAF associates with actin and hnrnp U for RNA polymerase II transcription', *Molecular and Cellular Biology*, 28(20), pp. 6342–6357. doi:10.1128/mcb.00766-08.

Ohki, K. *et al.* (2018) 'Clinical and molecular characteristics of *mef2d* fusion-positive B-cell precursor acute lymphoblastic leukemia in childhood, including a novel

translocation resulting in *MEF2D-hnrnph1* gene fusion', *Haematologica*, 104(1), pp. 128–137. doi:10.3324/haematol.2017.186320.

Panhale, A. *et al.* (2019) 'Capri enables comparison of evolutionarily conserved RNA interacting regions', *Nature Communications*, 10(1). doi:10.1038/s41467-019-10585-3.

Patel, A. *et al.* (2015) 'A liquid-to-solid phase transition of the ALS protein Fus accelerated by disease mutation', *Cell*, 162(5), pp. 1066–1077. doi:10.1016/j.cell.2015.07.047.

Pellizzoni, L., Yong, J. and Dreyfuss, G. (2002) 'Essential role for the SMN complex in the specificity of snRNP Assembly', *Science*, 298(5599), pp. 1775–1779. doi:10.1126/science.1074962.

Perriman, R. and Ares, M. (2010) 'Invariant U2 snRNA Nucleotides Form a Stem Loop to Recognize the Intron Early in Splicing', *Molecular Cell*, 38(3), pp. 416–427. doi:10.1016/j.molcel.2010.02.036.

Polo, S.E. *et al.* (2012) 'Regulation of DNA-end resection by HNRNPU-like proteins promotes DNA double-strand break signaling and Repair', *Molecular Cell*, 45(4), pp. 505–516. doi:10.1016/j.molcel.2011.12.035.

Puvvula, P.K. and Moon, A.M. (2021) 'Novel cell-penetrating peptides derived from scaffold-attachment- factor a inhibits cancer cell proliferation and survival', *Frontiers in Oncology*, 11. doi:10.3389/fonc.2021.621825.

Pérez-Losada, J. *et al.* (2000) 'Liposarcoma initiated by FUS/TLS-CHOP: The FUS/TLS domain plays a critical role in the pathogenesis of liposarcoma', *Oncogene*, 19(52), pp. 6015–6022. doi:10.1038/sj.onc.1204018.

Raczynska, K.D. *et al.* (2015) 'FUS/TLS contributes to replication-dependent histone gene expression by interaction with U7 snRNPs and histone-specific transcription factors', *Nucleic Acids Research* [Preprint]. doi:10.1093/nar/gkv794.

Rappsilber, J. *et al.* (2002) 'Large-scale proteomic analysis of the human spliceosome',

Genome Research, 12(8), pp. 1231–1245. doi:10.1101/gr.473902.

Rauscher, F.J. *et al.* (1988) 'Fos-associated protein P39 is the product of the *jun* proto-oncogene', *Science*, 240(4855), pp. 1010–1016. doi:10.1126/science.3130660.

Raška, I. *et al.* (1991) 'Immunological and ultrastructural studies of the nuclear coiled body with autoimmune antibodies', *Experimental Cell Research*, 195(1), pp. 27–37. doi:10.1016/0014-4827(91)90496-h.

Reddy, R. *et al.* (1987) 'The capped U6 small nuclear RNA is transcribed by RNA polymerase III.', *Journal of Biological Chemistry*, 262(1), pp. 75–81. doi:10.1016/s0021-9258(19)75890-6.

Rino, J. *et al.* (2008) 'Splicing factors SF1 and U2AF associate in extraspliceosomal complexes', *Molecular and Cellular Biology*, 28(9), pp. 3045–3057. doi:10.1128/mcb.02015-07.

Rosen, J.B. *et al.* (1998) 'Immediate-early gene expression in the amygdala following footshock stress and Contextual Fear Conditioning', *Brain Research*, 796(1–2), pp. 132–142. doi:10.1016/s0006-8993(98)00294-7.

Rädle, B. *et al.* (2013) 'Metabolic labeling of newly transcribed RNA for high resolution gene expression profiling of RNA synthesis, processing and decay in cell culture', *Journal of Visualized Experiments* [Preprint], (78). doi:10.3791/50195.

Sabo, S.L. *et al.* (2023) 'GRIN2B-related neurodevelopmental disorder: Current understanding of pathophysiological mechanisms', *Frontiers in Synaptic Neuroscience*, 14. doi:10.3389/fnsyn.2022.1090865.

Sapir, T. *et al.* (2022) 'Heterogeneous nuclear ribonucleoprotein U (HNRNPU) safeguards the developing mouse cortex', *Nature Communications*, 13(1). doi:10.1038/s41467-022-31752-z.

Sathyan, K.M. *et al.* (2019) *An improved auxin-inducible degron system preserves native protein levels and enables rapid and specific protein depletion* [Preprint].

doi:10.1101/585927.

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', *Experimental Cell Research*, 176(2), pp. 344–359. doi:10.1016/0014-4827(88)90336-9.

Schlackow, M. *et al.* (2017) 'Distinctive patterns of transcription and RNA processing for human lincrnas', *Molecular Cell*, 65(1), pp. 25–38. doi:10.1016/j.molcel.2016.11.029.

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', *Molecular Biology of the Cell*, 9(5), pp. 1025–1036. doi:10.1091/mbc.9.5.1025.

Schwalb, B. *et al.* (2016) 'TT-Seq maps the human transient transcriptome', *Science*, 352(6290), pp. 1225–1228. doi:10.1126/science.aad9841.

Schwartz, J.C. *et al.* (2012) 'FUS binds the CTD of RNA polymerase II and regulates its phosphorylation at SER2', *Genes & Development*, 26(24), pp. 2690–2695. doi:10.1101/gad.204602.112.

Sephton, C.F. *et al.* (2014) 'Activity-dependent fus dysregulation disrupts synaptic homeostasis', *Proceedings of the National Academy of Sciences*, 111(44). doi:10.1073/pnas.1406162111.

Shalaeva, D.N. *et al.* (2018) 'Evolution of cation binding in the active sites of P-loop nucleoside triphosphatases in relation to the basic catalytic mechanism', *eLife*, 7. doi:10.7554/elife.37373.

Sharp, F. *et al.* (1991) 'C-FOS mrna, fos, and fos-related antigen induction by hypertonic saline and stress', *The Journal of Neuroscience*, 11(8), pp. 2321–2331. doi:10.1523/jneurosci.11-08-02321.1991.

Sharp, J.A. *et al.* (2020) 'Cell division requires RNA eviction from condensing chromosomes', *Journal of Cell Biology*, 219(11). doi:10.1083/jcb.201910148.

Shaulian, E. (2010) 'AP-1 — the jun proteins: Oncogenes or tumor suppressors in

disguise?', *Cellular Signalling*, 22(6), pp. 894–899. doi:10.1016/j.cellsig.2009.12.008.

Shedlovskiy, D., Shcherbik, N. and Pestov, D.G. (2017) 'One-step hot formamide extraction of RNA from *saccharomyces cerevisiae*', *RNA Biology*, 14(12), pp. 1722–1726. doi:10.1080/15476286.2017.1345417.

Silva, B.A., Burns, A.M. and Gräff, J. (2018) 'A cfos activation map of remote fear memory attenuation', *Psychopharmacology*, 236(1), pp. 369–381. doi:10.1007/s00213-018-5000-y.

Singatulina, A.S. *et al.* (2019) 'PARP-1 activation directs FUS to DNA damage sites to form PARG-reversible compartments enriched in damaged DNA', *Cell Reports*, 27(6). doi:10.1016/j.celrep.2019.04.031.

Smith, K.P. *et al.* (1995) 'U2 and U1 snrna gene loci associate with coiled bodies', *Journal of Cellular Biochemistry*, 59(4), pp. 473–485. doi:10.1002/jcb.240590408.

So, B.R. *et al.* (2019) 'A complex of U1 snrnp with cleavage and polyadenylation factors controls telescripting, regulating mrna transcription in human cells', *Molecular Cell*, 76(4). doi:10.1016/j.molcel.2019.08.007.

Staněk, D. *et al.* (2008) 'Spliceosomal small nuclear ribonucleoprotein particles repeatedly cycle through Cajal Bodies', *Molecular Biology of the Cell*, 19(6), pp. 2534–2543. doi:10.1091/mbc.e07-12-1259.

Stewart, M. (2006) 'Faculty opinions recommendation of rules for nuclear localization sequence recognition by karyopherin beta 2.', *Faculty Opinions – Post-Publication Peer Review of the Biomedical Literature* [Preprint]. doi:10.3410/f.1033926.388067.

Strzelecka, M. *et al.* (2010) 'Coilin-dependent snrnp assembly is essential for zebrafish embryogenesis', *Nature Structural & Molecular Biology*, 17(4), pp. 403–409. doi:10.1038/nsmb.1783.

Studniarek, C. *et al.* (2021) 'The 7SK/P-TEFb snRNP controls ultraviolet radiation-

induced transcriptional reprogramming', *Cell Reports*, 35(2), p. 108965. doi:10.1016/j.celrep.2021.108965.

Sun, S. *et al.* (2015) 'Als-causative mutations in FUS/TLS confer gain and loss of function by altered association with SMN and U1-snrnp', *Nature Communications*, 6(1). doi:10.1038/ncomms7171.

Sun, Y. *et al.* (2020) 'Structure of an active human histone pre-mrna 3′-end processing machinery', *Science*, 367(6478), pp. 700–703. doi:10.1126/science.aaz7758.

Sunwoo, H. *et al.* (2008) '*men ε/β* nuclear-retained non-coding RNAS are up-regulated upon muscle differentiation and are essential components of Paraspeckles', *Genome Research*, 19(3), pp. 347–359. doi:10.1101/gr.087775.108.

Suzuki, H. *et al.* (2022) 'The 3′ pol II pausing at replication-dependent histone genes is regulated by mediator through Cajal Bodies' association with histone locus bodies', *Nature Communications*, 13(1). doi:10.1038/s41467-022-30632-w.

Sévigny, M. *et al.* (2020) 'FUS contributes to mtor-dependent inhibition of translation', *Journal of Biological Chemistry*, 295(52), pp. 18459–18473. doi:10.1074/jbc.ra120.013801.

Taiana, E. *et al.* (2020) 'LncRNA NEAT1 in Paraspeckles: A structural scaffold for cellular DNA damage response systems?', *Non-Coding RNA*, 6(3), p. 26. doi:10.3390/ncrna6030026.

Takahama, K. *et al.* (2013) 'Regulation of telomere length by G-quadruplex telomere DNA- and Terra-binding protein TLS/FUS', *Chemistry & Biology*, 20(3), pp. 341–350. doi:10.1016/j.chembiol.2013.02.013.

Takata, H. *et al.* (2012) 'The integrator complex is required for integrity of Cajal Bodies', *Journal of Cell Science*, 125(1), pp. 166–175. doi:10.1242/jcs.090837.

Talbert, P.B. and Henikoff, S. (2016) 'Histone variants on the move: Substrates for chromatin dynamics', *Nature Reviews Molecular Cell Biology*, 18(2), pp. 115–126. doi:10.1038/nrm.2016.148.

Tan, X. *et al.* (2007) *Mechanism of auxin perception by the TIR1 ubiquitin ligase*

[Preprint]. doi:10.2210/pdb2p1q/pdb.

Tarn, W.-Y. and Steitz, J.A. (1996) ‘A novel spliceosome containing U11, U12, and U5 snrnps excises a minor class (at–AC) intron in vitro’, *Cell*, 84(5), pp. 801–811. doi:10.1016/s0092-8674(00)81057-0.

Taylor, J. *et al.* (2022) ‘Expanding the phenotype of HNRNPU-related neurodevelopmental disorder with emphasis on seizure phenotype and review of literature’, *American Journal of Medical Genetics Part A*, 188(5), pp. 1497–1514. doi:10.1002/ajmg.a.62677.

Thandapani, P. *et al.* (2013) ‘Defining the RGG/RG motif’, *Molecular Cell*, 50(5), pp. 613–623. doi:10.1016/j.molcel.2013.05.021.

Thickman, K.R. *et al.* (2006) ‘Multiple u2af65 binding sites within SF3B155: Thermodynamic and spectroscopic characterization of protein–protein interactions among pre-mrna splicing factors’, *Journal of Molecular Biology*, 356(3), pp. 664–683. doi:10.1016/j.jmb.2005.11.067.

Tullai, J.W. *et al.* (2007) ‘Immediate-early and delayed primary response genes are distinct in function and genomic architecture’, *Journal of Biological Chemistry*, 282(33), pp. 23981–23995. doi:10.1074/jbc.m702044200.

Turunen, J.J. *et al.* (2012) ‘The significant other: Splicing by the minor spliceosome’, *Wiley Interdisciplinary Reviews: RNA*, 4(1), pp. 61–76. doi:10.1002/wrna.1141.

Tuvikene, J. *et al.* (2016) ‘AP-1 transcription factors mediate BDNF-positive feedback loop in cortical neurons’, *The Journal of Neuroscience*, 36(4), pp. 1290–1305. doi:10.1523/jneurosci.3360-15.2016.

Valente, S.T. and Goff, S.P. (2006) ‘Inhibition of HIV-1 gene expression by a fragment of hnRNP U’, *Molecular Cell*, 23(4), pp. 597–605. doi:10.1016/j.molcel.2006.07.021.

van Dam, H. *et al.* (1993) ‘Heterodimer Formation of CJUN and ATF-2 is responsible for induction of c-JUN by the 243 amino acid adenovirus E1A protein.’, *The EMBO Journal*, 12(2), pp. 479–487. doi:10.1002/j.1460-2075.1993.tb05680.x.

Van Nostrand, E.L. *et al.* (2016) 'Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced clip (eclip)', *Nature Methods*, 13(6), pp. 508–514. doi:10.1038/nmeth.3810.

Vance, C. *et al.* (2013) 'Als mutant fus disrupts nuclear localization and sequesters wild-type fus within cytoplasmic stress granules', *Human Molecular Genetics*, 22(13), pp. 2676–2688. doi:10.1093/hmg/ddt117.

Venters, C.C. *et al.* (2019) 'U1 snrnp telescripting: Suppression of premature transcription termination in introns as a new layer of gene regulation', *Cold Spring Harbor Perspectives in Biology*, 11(2). doi:10.1101/cshperspect.a032235.

Viphakone, N. *et al.* (2019) 'Co-transcriptional loading of RNA export factors shapes the human transcriptome', *Molecular Cell*, 75(2). doi:10.1016/j.molcel.2019.04.034.

Wang, H. *et al.* (2022) 'Heterogeneous nuclear ribonucleoprotein U-actin complex derived from extracellular vesicles facilitates proliferation and migration of human coronary artery endothelial cells by promoting RNA polymerase II transcription', *Bioengineered*, 13(5), pp. 11469–11486. doi:10.1080/21655979.2022.2066754.

Wang, X., Schwartz, J.C. and Cech, T.R. (2015a) 'Nucleic acid-binding specificity of human FUS protein', *Nucleic Acids Research*, 43(15), pp. 7535–7543. doi:10.1093/nar/gkv679.

Wang, X., Schwartz, J.C. and Cech, T.R. (2015b) 'Nucleic acid-binding specificity of human FUS protein', *Nucleic Acids Research*, 43(15), pp. 7535–7543. doi:10.1093/nar/gkv679.

West, J.A. *et al.* (2016) 'Structural, super-resolution microscopy analysis of Paraspeckle Nuclear Body Organization', *Journal of Cell Biology*, 214(7), pp. 817–830. doi:10.1083/jcb.201601071.

West, K.O. *et al.* (2019) 'The splicing factor hnRNP M is a critical regulator of innate immune gene expression in macrophages', *Cell Reports*, 29(6).

doi:10.1016/j.celrep.2019.09.078.

Whisenant, T.C. *et al.* (2015) 'The activation-induced assembly of an RNA/protein interactome centered on the splicing factor U2AF2 regulates gene expression in human CD4 T cells', *PLOS ONE*, 10(12). doi:10.1371/journal.pone.0144409.

Wickham, H. *et al.* (2019) 'Welcome to the Tidyverse', *Journal of Open Source Software*, 4(43), p. 1686. doi:10.21105/joss.01686.

Wu, Q. and Krainer, A.R. (1999) 'At-AC pre-mrna splicing mechanisms and conservation of minor introns in voltage-gated ion channel genes', *Molecular and Cellular Biology*, 19(5), pp. 3225–3236. doi:10.1128/mcb.19.5.3225.

Wu, Z., Nicoll, M. and Ingham, R.J. (2021) 'AP-1 family transcription factors: A diverse family of proteins that regulate varied cellular activities in classical hodgkin lymphoma and alk+ ALCL', *Experimental Hematology & Oncology*, 10(1). doi:10.1186/s40164-020-00197-9.

Xiao, R. *et al.* (2012) 'Nuclear Matrix Factor hnRNP U/SAF-A Exerts a Global Control of Alternative Splicing by Regulating U2 snRNP Maturation', *Molecular Cell*, 45(5), pp. 656–668. doi:10.1016/j.molcel.2012.01.009.

Yagi, R., Miyazaki, T. and Oyoshi, T. (2018) 'G-quadruplex binding ability of TLS/FUS depends on the β-spiral structure of the RGG domain', *Nucleic Acids Research*, 46(12), pp. 5894–5901. doi:10.1093/nar/gky391.

Yamazaki, T. *et al.* (2012) 'FUS-SMN protein interactions link the motor neuron diseases ALS and SMA', *Cell Reports*, 2(4), pp. 799–806. doi:10.1016/j.celrep.2012.08.025.

Yang, L. *et al.* (2014) 'Self-assembled FUS binds active chromatin and regulates gene transcription', *Proceedings of the National Academy of Sciences*, 111(50), pp. 17809–17814. doi:10.1073/pnas.1414004111.

Yang, X. *et al.* (2014) 'A conserved interaction that is essential for the biogenesis of histone locus bodies', *Journal of Biological Chemistry*, 289(49), pp. 33767–33782. doi:10.1074/jbc.m114.616466.

Yasuda, T. *et al.* (2016) 'Recurrent DUX4 fusions in B cell acute lymphoblastic leukemia of adolescents and young adults', *Nature Genetics*, 48(5), pp. 569–574. doi:10.1038/ng.3535.

Yates, T.M. *et al.* (2017) 'De novo mutations in HNRNPU result in a neurodevelopmental syndrome', *American Journal of Medical Genetics Part A*, 173(11), pp. 3003–3012. doi:10.1002/ajmg.a.38492.

Ye, J. *et al.* (2015) 'HnRNP U protein is required for normal pre-mrna splicing and postnatal heart development and function', *Proceedings of the National Academy of Sciences*, 112(23). doi:10.1073/pnas.1508461112.

Yesbolatova, A. *et al.* (2020) 'The auxin-inducible degron 2 technology provides sharp degradation control in yeast, mammalian cells, and mice', *Nature Communications*, 11(1). doi:10.1038/s41467-020-19532-z.

Yin, Y. *et al.* (2020) 'U1 snrnp regulates chromatin retention of noncoding RNAS', *Nature*, 580(7801), pp. 147–150. doi:10.1038/s41586-020-2105-3.

Yonchev, I.D. (2021) *Characterisation of the roles of hnRNPU and hnRNPUL1 in gene expression*. thesis. The University of Sheffield.

Yoshimoto, R. *et al.* (2021) 'Spliceostatin a interaction with SF3B limits U1 snrnp availability and causes premature cleavage and polyadenylation', *Cell Chemical Biology*, 28(9). doi:10.1016/j.chembiol.2021.03.002.

Yu, Y. and Reed, R. (2015) 'FUS functions in coupling transcription to splicing by mediating an interaction between RNAP II and U1 snRNP', *Proceedings of the National Academy of Sciences*, 112(28), pp. 8608–8613. doi:10.1073/pnas.1506282112.

Yu, Y. *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 Research*, 43(6), pp. 3208–3218. doi:10.1093/nar/gkv157.

Zarnack, K. *et al.* (2013) 'Direct competition between hnrnp C and U2AF65 protects the transcriptome from the exonization of alu elements', *Cell*, 152(3), pp. 453–

466. doi:10.1016/j.cell.2012.12.023.

Zhang, S. *et al.* (2021) *Structure of a transcribing RNA polymerase II-U1 snRNP complex* [Preprint]. doi:10.2210/pdb7b0y/pdb.

Zhang, X.-R. *et al.* (2021) *An improved auxin-inducible degron system for fission yeast* [Preprint]. doi:10.1101/2021.07.19.452993.

Zhang, Z. and Carmichael, G.G. (2001) 'The fate of dsrna in the nucleus: A p54nrb-Containing Complex Mediates the Nuclear Retention of Promiscuously A-to-I Edited RNAs', *Cell*, 106(4), pp. 465–476. doi:10.1016/s0092-8674(01)00466-4.

Zhang, Z. *et al.* (2020) 'Molecular architecture of the human 17S U2 snRNP', *Nature*, 583(7815), pp. 310–313. doi:10.1038/s41586-020-2344-3.

Zhao, J. *et al.* (1998) 'Expression of NPAT, a novel substrate of cyclin E-CDK2, promotes S-phase entry', *Genes & Development*, 12(4), pp. 456–461. doi:10.1101/gad.12.4.456.

Zhao, J. *et al.* (2000) 'NPAT links Cyclin E–cdk2 to the regulation of replication-dependent histone gene transcription', *Genes & Development*, 14(18), pp. 2283–2297. doi:10.1101/gad.827700.

Zhou, Q., Li, T. and Price, D.H. (2012) 'RNA polymerase II elongation control', *Annual Review of Biochemistry*, 81(1), pp. 119–143. doi:10.1146/annurev-biochem-052610-095910.

Zhou, Z. *et al.* (2002) 'Comprehensive proteomic analysis of the human spliceosome', *Nature*, 419(6903), pp. 182–185. doi:10.1038/nature01031.

Zieve, G., Benecke, B.J. and Penman, S. (1977) 'Synthesis of two classes of small RNA species in vivo and in vitro', *Biochemistry*, 16(20), pp. 4520–4525. doi:10.1021/bi00639a029.

Zinszner, H., Albalat, R. and Ron, D. (1994) 'A novel effector domain from the RNA-binding protein TLS or EWS is required for oncogenic transformation by chop.', *Genes & Development*, 8(21), pp. 2513–2526. doi:10.1101/gad.8.21.2513.